NIST Special Publication 800-82

Revision 2

Guide to Industrial Control Systems (ICS) Security

Supervisory Control and Data Acquisition (SCADA) systems, Distributed Control Systems (DCS), and other control system configurations such as Programmable Logic Controllers (PLC)

Keith Stouffer
Intelligent Systems Division
Engineering Laboratory

Victoria Pillitteri
Suzanne Lightman
Computer Security Division
Information Technology Laboratory

Marshall Abrams
The MITRE Corporation

Adam Hahn
Washington State University

This publication is available free of charge from:
http://dx.doi.org/10.6028/NIST.SP.800-82r2

May 2015

U.S. Department of Commerce
Penny Pritzker, Secretary

National Institute of Standards and Technology
Willie May, Under Secretary of Commerce for Standards and Technology and Director

Table of Contents

List of Appendices

List of Figures

List of Tables

Executive Summary

This document provides guidance for establishing secure industrial control systems (ICS). These ICS, which include supervisory control and data acquisition (SCADA) systems, distributed control systems (DCS), and other control system configurations such as Programmable Logic Controllers (PLC) are often found in the industrial control sectors. ICS are typically used in industries such as electric, water and wastewater, oil and natural gas, transportation, chemical, pharmaceutical, pulp and paper, food and beverage, and discrete manufacturing (e.g., automotive, aerospace, and durable goods.) SCADA systems are generally used to control dispersed assets using centralized data acquisition and supervisory control. DCS are generally used to control production systems within a local area such as a factory using supervisory and regulatory control. PLCs are generally used for discrete control for specific applications and generally provide regulatory control. These control systems are vital to the operation of the U.S. critical infrastructures that are often highly interconnected and mutually dependent systems. It is important to note that approximately 90 percent of the nation's critical infrastructures are privately owned and operated. Federal agencies also operate many of the ICS mentioned above; other examples include air traffic control and materials handling (e.g., Postal Service mail handling.) This document provides an overview of these ICS and typical system topologies, identifies typical threats and vulnerabilities to these systems, and provides recommended security countermeasures to mitigate the associated risks.

Initially, ICS had little resemblance to traditional information technology (IT) systems in that ICS were isolated systems running proprietary control protocols using specialized hardware and software. Many ICS components were in physically secured areas and the components were not connected to IT networks or systems. Widely available, low-cost Internet Protocol (IP) devices are now replacing proprietary solutions, which increases the possibility of cybersecurity vulnerabilities and incidents. As ICS are adopting IT solutions to promote corporate business systems connectivity and remote access capabilities, and are being designed and implemented using industry standard computers, operating systems (OS) and network protocols, they are starting to resemble IT systems. This integration supports new IT capabilities, but it provides significantly less isolation for ICS from the outside world than predecessor systems, creating a greater need to secure these systems. The increasing use of wireless networking places ICS implementations at greater risk from adversaries who are in relatively close physical proximity but do not have direct physical access to the equipment. While security solutions have been designed to deal with these security issues in typical IT systems, special precautions must be taken when introducing these same solutions to ICS environments. In some cases, new security solutions are needed that are tailored to the ICS environment.

Although some characteristics are similar, ICS also have characteristics that differ from traditional information processing systems. Many of these differences stem from the fact that logic executing in ICS has a direct effect on the physical world. Some of these characteristics include significant risk to the health and safety of human lives and serious damage to the environment, as well as serious financial issues such as production losses, negative impact to a nation’s economy, and compromise of proprietary information. ICS have unique performance and reliability requirements and often use operating systems and applications that may be considered unconventional to typical IT personnel. Furthermore, the goals of safety and efficiency sometimes conflict with security in the design and operation of control systems.

ICS cybersecurity programs should always be part of broader ICS safety and reliability programs at both industrial sites and enterprise cybersecurity programs, because cybersecurity is essential to the safe and reliable operation of modern industrial processes. Threats to control systems can come from numerous sources, including hostile governments, terrorist groups, disgruntled employees, malicious intruders, complexities, accidents, and natural disasters as well as malicious or accidental actions by insiders. ICS security objectives typically follow the priority of availability and integrity, followed by confidentiality.

Possible incidents an ICS may face include the following:

- Blocked or delayed flow of information through ICS networks, which could disrupt ICS operation.
- Unauthorized changes to instructions, commands, or alarm thresholds, which could damage, disable, or shut down equipment, create environmental impacts, and/or endanger human life.
- Inaccurate information sent to system operators, either to disguise unauthorized changes, or to cause the operators to initiate inappropriate actions, which could have various negative effects.
- ICS software or configuration settings modified, or ICS software infected with malware, which could have various negative effects.
- Interference with the operation of equipment protection systems, which could endanger costly and difficult-to-replace equipment.
- Interference with the operation of safety systems, which could endanger human life.

Major security objectives for an ICS implementation should include the following:

- **Restricting logical access to the ICS network and network activity.** This may include using unidirectional gateways, a demilitarized zone (DMZ) network architecture with firewalls to prevent network traffic from passing directly between the corporate and ICS networks, and having separate authentication mechanisms and credentials for users of the corporate and ICS networks. The ICS should also use a network topology that has multiple layers, with the most critical communications occurring in the most secure and reliable layer.
- **Restricting physical access to the ICS network and devices.** Unauthorized physical access to components could cause serious disruption of the ICS's functionality. A combination of physical access controls should be used, such as locks, card readers, and/or guards.
- **Protecting individual ICS components from exploitation.** This includes deploying security patches in as expeditious a manner as possible, after testing them under field conditions; disabling all unused ports and services and assuring that they remain disabled; restricting ICS user privileges to only those that are required for each person's role; tracking and monitoring audit trails; and using security controls such as antivirus software and file integrity checking software where technically feasible to prevent, deter, detect, and mitigate malware.
- **Restricting unauthorized modification of data.** This includes data that is in transit (at least across the network boundaries) and at rest.
- **Detecting security events and incidents.** Detecting security events, which have not yet escalated into incidents, can help defenders break the attack chain before attackers attain their objectives. This includes the capability to detect failed ICS components, unavailable services, and exhausted resources that are important to provide proper and safe functioning of the ICS.
- **Maintaining functionality during adverse conditions.** This involves designing the ICS so that each critical component has a redundant counterpart. Additionally, if a component fails, it should fail in a manner that does not generate unnecessary traffic on the ICS or other networks, or does not cause another problem elsewhere, such as a cascading event. The ICS should also allow for graceful degradation such as moving from "normal operation" with full automation to "emergency operation" with operators more involved and less automation to "manual operation" with no automation.

- **Restoring the system after an incident.** Incidents are inevitable and an incident response plan is essential. A major characteristic of a good security program is how quickly the system can be recovered after an incident has occurred.

To properly address security in an ICS, it is essential for a cross-functional cybersecurity team to share their varied domain knowledge and experience to evaluate and mitigate risk to the ICS. The cybersecurity team should consist of a member of the organization's IT staff, control engineer, control system operator, network and system security expert, a member of the management staff, and a member of the physical security department at a minimum. For continuity and completeness, the cybersecurity team should consult with the control system vendor and/or system integrator as well. The cybersecurity team should coordinate closely with site management (e.g., facility superintendent) and the company's Chief Information Officer (CIO) or Chief Security Officer (CSO), who in turn, accepts complete responsibility and accountability for the cybersecurity of the ICS, and for any safety incidents, reliability incidents, or equipment damage caused directly or indirectly by cyber incidents. An effective cybersecurity program for an ICS should apply a strategy known as "defense-in-depth," layering security mechanisms such that the impact of a failure in any one mechanism is minimized. Organizations should not rely on "security by obscurity."

In a typical ICS this means a defense-in-depth strategy that includes:

- Developing security policies, procedures, training and educational material that applies specifically to the ICS.
- Considering ICS security policies and procedures based on the Homeland Security Advisory System Threat Level, deploying increasingly heightened security postures as the Threat Level increases.
- Addressing security throughout the lifecycle of the ICS from architecture design to procurement to installation to maintenance to decommissioning.
- Implementing a network topology for the ICS that has multiple layers, with the most critical communications occurring in the most secure and reliable layer.
- Providing logical separation between the corporate and ICS networks (e.g., stateful inspection firewall(s) between the networks, unidirectional gateways).
- Employing a DMZ network architecture (i.e., prevent direct traffic between the corporate and ICS networks).
- Ensuring that critical components are redundant and are on redundant networks.
- Designing critical systems for graceful degradation (fault tolerant) to prevent catastrophic cascading events.
- Disabling unused ports and services on ICS devices after testing to assure this will not impact ICS operation.
- Restricting physical access to the ICS network and devices.
- Restricting ICS user privileges to only those that are required to perform each person's job (i.e., establishing role-based access control and configuring each role based on the principle of least privilege).
- Using separate authentication mechanisms and credentials for users of the ICS network and the corporate network (i.e., ICS network accounts do not use corporate network user accounts).

- Using modern technology, such as smart cards for Personal Identity Verification (PIV).
- Implementing security controls such as intrusion detection software, antivirus software and file integrity checking software, where technically feasible, to prevent, deter, detect, and mitigate the introduction, exposure, and propagation of malicious software to, within, and from the ICS.
- Applying security techniques such as encryption and/or cryptographic hashes to ICS data storage and communications where determined appropriate.
- Expeditiously deploying security patches after testing all patches under field conditions on a test system if possible, before installation on the ICS.
- Tracking and monitoring audit trails on critical areas of the ICS.
- Employing reliable and secure network protocols and services where feasible.

The National Institute of Standards and Technology (NIST), in cooperation with the public and private sector ICS community, has developed specific guidance on the application of the security controls in NIST Special Publication (SP) 800-53 Revision 4, *Security and Privacy Controls for Federal Information Systems and Organizations* [22], to ICS.

While many controls in Appendix F of NIST SP 800-53 are applicable to ICS as written, many controls require ICS-specific interpretation and/or augmentation by adding one or more of the following to the control:

- ICS Supplemental Guidance provides organizations with additional information on the application of the security controls and control enhancements in Appendix F of NIST SP 800-53 to ICS and the environments in which these specialized systems operate. The Supplemental Guidance also provides information as to why a particular security control or control enhancement may not be applicable in some ICS environments and may be a candidate for tailoring (i.e., the application of scoping guidance and/or compensating controls). ICS Supplemental Guidance does not replace the original Supplemental Guidance in Appendix F of NIST SP 800-53.
- ICS Enhancements (one or more) that provide enhancement augmentations to the original control that may be required for some ICS.
- ICS Enhancement Supplemental Guidance that provides guidance on how the control enhancement applies, or does not apply, in ICS environments.

The most successful method for securing an ICS is to gather industry recommended practices and engage in a proactive, collaborative effort between management, the controls engineer and operator, the IT organization, and a trusted automation advisor. This team should draw upon the wealth of information available from ongoing federal government, industry groups, vendor and standards organizational activities listed in Appendix D—.

2. Overview of Industrial Control Systems

Industrial control system (ICS) is a general term that encompasses several types of control systems, including supervisory control and data acquisition (SCADA) systems, distributed control systems (DCS), and other control system configurations such as Programmable Logic Controllers (PLC) often found in the industrial sectors and critical infrastructures. An ICS consists of combinations of control components (e.g., electrical, mechanical, hydraulic, pneumatic) that act together to achieve an industrial objective (e.g., manufacturing, transportation of matter or energy). The part of the system primarily concerned with producing the output is referred to as the process. The control part of the system includes the specification of the desired output or performance. Control can be fully automated or may include a human in the loop. Systems can be configured to operate open-loop, closed-loop, and manual mode. In open-loop control systems the output is controlled by established settings. In closed-loop control systems, the output has an effect on the input in such a way as to maintain the desired objective. In manual mode the system is controlled completely by humans. The part of the system primarily concerned with maintaining conformance with specifications is referred to as the controller (or control). A typical ICS may contain numerous control loops, Human Machine Interfaces (HMIs), and remote diagnostics and maintenance tools built using an array of network protocols. ICS control industrial processes are typically used in electrical, water and wastewater, oil and natural gas, chemical, transportation, pharmaceutical, pulp and paper, food and beverage, and discrete manufacturing (e.g., automotive, aerospace, and durable goods) industries.

ICS are critical to the operation of the U.S. critical infrastructures that are often highly interconnected and mutually dependent systems. It is important to note that approximately 85 percent of the nation's critical infrastructures are privately owned and operated[1]. Federal agencies also operate many of the industrial processes mentioned above as well as air traffic control. This section provides an overview of SCADA, DCS, and PLC systems, including typical topologies and components. Several diagrams are presented to depict the network topology, connections, components, and protocols typically found on each system to facilitate the understanding of these systems. These examples only attempt to identify notional topology concepts. Actual implementations of ICS may be hybrids that blur the line between DCS and SCADA systems. Note that the diagrams in this section do not focus on securing ICS. Security architecture and security controls are discussed in Section 5 and Section 6 of this document respectively.

2.1 Evolution of Industrial Control Systems

Many of today's ICS evolved from the insertion of IT capabilities into existing physical systems, often replacing or supplementing physical control mechanisms. For example, embedded digital controls replaced analog mechanical controls in rotating machines and engines. Improvements in cost-and performance have encouraged this evolution, resulting in many of today's "smart" technologies such as the smart electric grid, smart transportation, smart buildings, and smart manufacturing. While this increases the connectivity and criticality of these systems, it also creates a greater need for their adaptability, resilience, safety, and security.

Engineering of ICS continues to evolve to provide new capabilities while maintaining the typical long lifecycles of these systems. The introduction of IT capabilities into physical systems presents emergent behavior that has security implications. Engineering models and analysis are evolving to address these emergent properties including safety, security, privacy, and environmental impact interdependencies.

[1] http://www.dhs.gov/critical-infrastructure-sector-partnerships (last updated April 2014)

2.2 ICS Industrial Sectors and Their Interdependencies

Control systems are used in many different industrial sectors and critical infrastructures, including manufacturing, distribution, and transportation.

2.2.1 Manufacturing Industries

Manufacturing presents a large and diverse industrial sector with many different processes, which can be categorized into *process-based* and *discrete-based* manufacturing.

The *process-based* manufacturing industries typically utilize two main processes [1]:

- **Continuous Manufacturing Processes.** These processes run continuously, often with transitions to make different grades of a product. Typical continuous manufacturing processes include fuel or steam flow in a power plant, petroleum in a refinery, and distillation in a chemical plant.
- **Batch Manufacturing Processes.** These processes have distinct processing steps, conducted on a quantity of material. There is a distinct start and end step to a batch process with the possibility of brief steady state operations during intermediate steps. Typical batch manufacturing processes include food manufacturing.

The *discrete-based* manufacturing industries typically conduct a series of steps on a single device to create the end product. Electronic and mechanical parts assembly and parts machining are typical examples of this type of industry.

Both process-based and discrete-based industries utilize the same types of control systems, sensors, and networks. Some facilities are a hybrid of discrete and process-based manufacturing.

2.2.2 Distribution Industries

ICS are used to control geographically dispersed assets, often scattered over thousands of square kilometers, including distribution systems such as water distribution and wastewater collection systems, agricultural irrigation systems, oil and natural gas pipelines, electrical power grids, and railway transportation systems.

2.2.3 Differences between Manufacturing and Distribution ICS

While control systems used in manufacturing and distribution industries are very similar in operation, they are different in some aspects. Manufacturing industries are usually located within a confined factory or plant-centric area, when compared to geographically dispersed distribution industries. Communications in manufacturing industries are usually performed using local area network (LAN) technologies that are typically more reliable and high speed as compared to the long-distance communication wide-area networks (WAN) and wireless/RF (radio frequency) technologies used by distribution industries. The ICS used in distribution industries are designed to handle long-distance communication challenges such as delays and data loss posed by the various communication media used. The security controls may differ among network types.

2.2.4 ICS and Critical Infrastructure Interdependencies

The U.S. critical infrastructure is often referred to as a "system of systems" because of the interdependencies that exist between its various industrial sectors as well as interconnections between business partners [8] [9]. Critical infrastructures are highly interconnected and mutually dependent in

complex ways, both physically and through a host of information and communications technologies. An incident in one infrastructure can directly and indirectly affect other infrastructures through cascading and escalating failures.

Both the electrical power transmission and distribution grid industries use geographically distributed SCADA control technology to operate highly interconnected and dynamic systems consisting of thousands of public and private utilities and rural cooperatives for supplying electricity to end users. Some SCADA systems monitor and control electricity distribution by collecting data from and issuing commands to geographically remote field control stations from a centralized location. SCADA systems are also used to monitor and control water, oil and natural gas distribution, including pipelines, ships, trucks, and rail systems, as well as wastewater collection systems.

SCADA systems and DCS are often networked together. This is the case for electric power control centers and electric power generation facilities. Although the electric power generation facility operation is controlled by a DCS, the DCS must communicate with the SCADA system to coordinate production output with transmission and distribution demands.

Electric power is often thought to be one of the most prevalent sources of disruptions of interdependent critical infrastructures. As an example, a cascading failure can be initiated by a disruption of the microwave communications network used for an electric power transmission SCADA system. The lack of monitoring and control capabilities could cause a large generating unit to be taken offline, an event that would lead to loss of power at a transmission substation. This loss could cause a major imbalance, triggering a cascading failure across the power grid. This could result in large area blackouts that could potentially affect oil and natural gas production, refinery operations, water treatment systems, wastewater collection systems, and pipeline transport systems that rely on the grid for electric power.

2.3 ICS Operation and Components

The basic operation of an ICS is shown in Figure 2-1 [2]. Some critical processes may also include safety systems. Key components include the following:

A typical ICS contains numerous control loops, human interfaces, and remote diagnostics and maintenance tools built using an array of network protocols on layered network architectures. A control loop utilizes sensors, actuators, and controllers (e.g., PLCs) to manipulate some controlled process. A sensor is a device that produces a measurement of some physical property and then sends this information as controlled variables to the controller. The controller interprets the signals and generates corresponding manipulated variables, based on a control algorithm and target set points, which it transmits to the actuators. Actuators such as control valves, breakers, switches, and motors are used to directly manipulate the controlled process based on commands from the controller.

Operators and engineers use human interfaces to monitor and configure set points, control algorithms, and to adjust and establish parameters in the controller. The human interface also displays process status information and historical information. Diagnostics and maintenance utilities are used to prevent, identify, and recover from abnormal operation or failures.

Sometimes these control loops are nested and/or cascading –whereby the set point for one loop is based on the process variable determined by another loop. Supervisory-level loops and lower-level loops operate continuously over the duration of a process with cycle times ranging on the order of milliseconds to minutes.

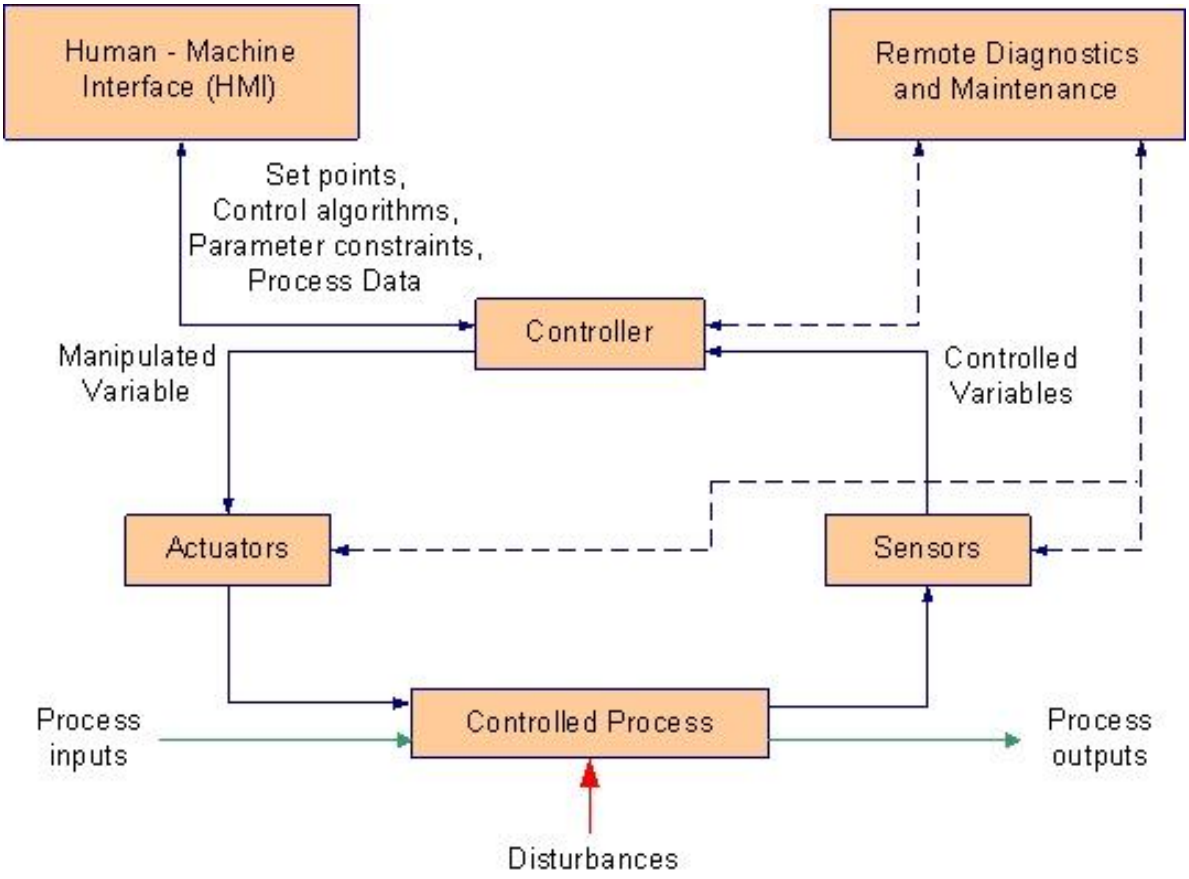

Figure 2-1. ICS Operation

To support subsequent discussions, this section defines key ICS components that are used in control and networking. Some of these components can be described generically for use in SCADA systems, DCS and PLCs, while others are unique to one. The Glossary of Terms in Appendix B— contains a more detailed listing of control and networking components. Additionally, Figure 2-5 and Figure 2-6 show SCADA implementation examples; Figure 2-7 shows a DCS implementation example and Figure 2-8 shows a PLC implementation example that incorporates these components.

2.3.1 ICS System Design Considerations

While Section 2.3 introduced the basic components of an ICS, the design of an ICS, including whether a SCADA, DCS, or PLC-based topologies are used depends on many factors. This section identifies key factors that drive design decisions regarding the control, communication, reliability, and redundancy properties of the ICS. Because these factors heavily influence the design of the ICS, they will also help determine the security needs of the system.

- **Control Timing Requirements.** ICS processes have a wide range of time-related requirements, including very high speed, consistency, regularity, and synchronization. Humans may not be able to reliably and consistently meet these requirements; automated controllers may be necessary. Some systems may require the computation to be performed as close to the sensor and actuators as possible to reduce communication latency and perform necessary control actions on time.

- **Geographic Distribution.** Systems have varying degrees of distribution, ranging from a small system (e.g., local PLC-controlled process) to large, distributed systems (e.g., oil pipelines, electric power grid). Greater distribution typically implies a need for wide area (e.g., leased lines, circuit switching, and packet switching) and mobile communication.
- **Hierarchy.** Supervisory control is used to provide a central location that can aggregate data from multiple locations to support control decisions based on the current state of the system. Often a hierarchical/centralized control is used to provide human operators with a comprehensive view of the entire system.
- **Control Complexity.** Often control functions can be performed by simple controllers and preset algorithms. However, more complex systems (e.g., air traffic control) require human operators to ensure that all control actions are appropriate to meet the larger objectives of the system.
- **Availability.** The system's availability (i.e., reliability) requirements are also an important factor in design. Systems with strong availability/up-time requirements may require more redundancy or alternate implementations across all communication and control.
- **Impact of Failures.** The failure of a control function could incur substantially different impacts across domains. Systems with greater impacts often require the ability to continue operations through redundant controls, or the ability to operate in a degraded state. The design needs to address these requirements.
- **Safety.** The system's safety requirements area also an important factor in design. Systems must be able to detect unsafe conditions and trigger actions to reduce unsafe conditions to safe ones. In most safety-critical operations, human oversight and control of a potentially dangerous process is an essential part of the safety system.

2.3.2 SCADA Systems

SCADA systems are used to control dispersed assets where centralized data acquisition is as important as control [3] [4]. These systems are used in distribution systems such as water distribution and wastewater collection systems, oil and natural gas pipelines, electrical utility transmission and distribution systems, and rail and other public transportation systems. SCADA systems integrate data acquisition systems with data transmission systems and HMI software to provide a centralized monitoring and control system for numerous process inputs and outputs. SCADA systems are designed to collect field information, transfer it to a central computer facility, and display the information to the operator graphically or textually, thereby allowing the operator to monitor or control an entire system from a central location in near real time. Based on the sophistication and setup of the individual system, control of any individual system, operation, or task can be automatic, or it can be performed by operator commands.

Typical hardware includes a control server placed at a control center, communications equipment (e.g., radio, telephone line, cable, or satellite), and one or more geographically distributed field sites consisting of Remote Terminal Units (RTUs) and/or PLCs, which controls actuators and/or monitors sensors. The control server stores and processes the information from RTU inputs and outputs, while the RTU or PLC controls the local process. The communications hardware allows the transfer of information and data back and forth between the control server and the RTUs or PLCs. The software is programmed to tell the system what and when to monitor, what parameter ranges are acceptable, and what response to initiate when parameters change outside acceptable values. An Intelligent Electronic Device (IED), such as a protective relay, may communicate directly to the control server, or a local RTU may poll the IEDs to collect the data and pass it to the control server. IEDs provide a direct interface to control and monitor equipment and sensors. IEDs may be directly polled and controlled by the control server and in most

cases have local programming that allows for the IED to act without direct instructions from the control center. SCADA systems are usually designed to be fault-tolerant systems with significant redundancy built into the system. Redundancy may not be a sufficient countermeasure in the face of malicious attack.

Figure 2-2 shows the components and general configuration of a SCADA system. The control center houses a control server and the communications routers. Other control center components include the HMI, engineering workstations, and the data historian, which are all connected by a LAN. The control center collects and logs information gathered by the field sites, displays information to the HMI, and may generate actions based upon detected events. The control center is also responsible for centralized alarming, trend analyses, and reporting. The field site performs local control of actuators and monitors sensors (Note that sensors and actuators are only shown in Figure 2-5). Field sites are often equipped with a remote access capability to allow operators to perform remote diagnostics and repairs usually over a separate dial up modem or WAN connection. Standard and proprietary communication protocols running over serial and network communications are used to transport information between the control center and field sites using telemetry techniques such as telephone line, cable, fiber, and radio frequency such as broadcast, microwave and satellite.

SCADA communication topologies vary among implementations. The various topologies used, including point-to-point, series, series-star, and multi-drop [5], are shown in Figure 2-3.

Point-to-point is functionally the simplest type; however, it is expensive because of the individual channels needed for each connection. In a series configuration, the number of channels used is reduced; however, channel sharing has an impact on the efficiency and complexity of SCADA operations. Similarly, the series-star and multi-drop configurations' use of one channel per device results in decreased efficiency and increased system complexity.

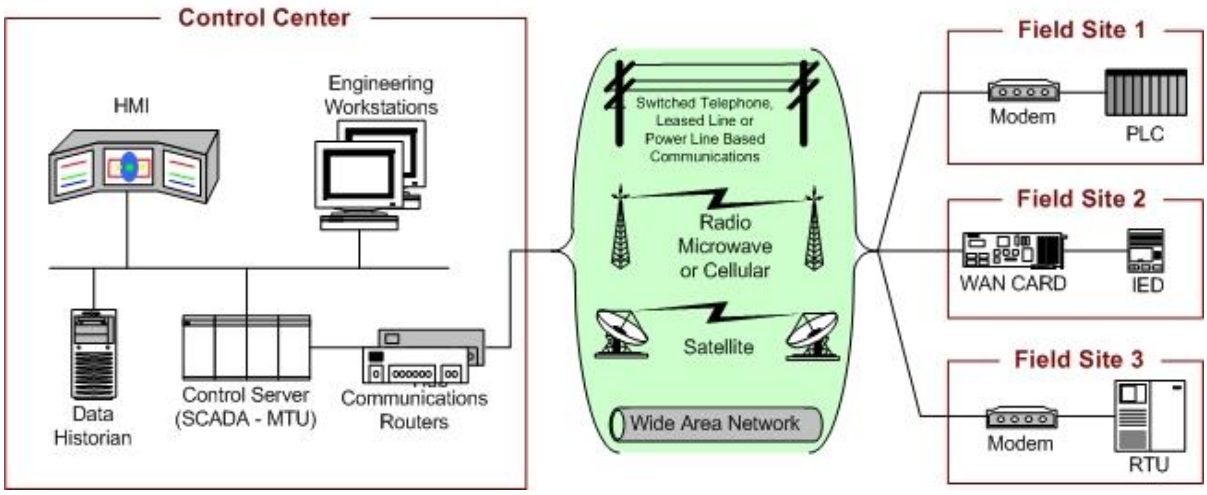

Figure 2-2. SCADA System General Layout

The four basic topologies Figure 2-3 can be further augmented using dedicated devices to manage communication exchanges as well as message switching and buffering. Large SCADA systems containing hundreds of RTUs often employee a sub-control server to alleviate the burden on the primary server. This type of topology is shown in Figure 2-4.

Figure 2-5 shows an example of a SCADA system implementation. This particular SCADA system consists of a primary control center and three field sites. A second backup control center provides redundancy in the event of a primary control center malfunction. Point-to-point connections are used for all control center to field site communications, with two connections using radio telemetry. The third field site is local to the control center and uses the WAN for communications. A regional control center resides above the primary control center for a higher level of supervisory control. The corporate network has access to all control centers through the WAN, and field sites can be accessed remotely for troubleshooting and maintenance operations. The primary control center polls field devices for data at defined intervals (e.g., 5 seconds, 60 seconds) and can send new set points to a field device as required. In addition to polling and issuing high-level commands, the control server also watches for priority interrupts coming from field site alarm systems.

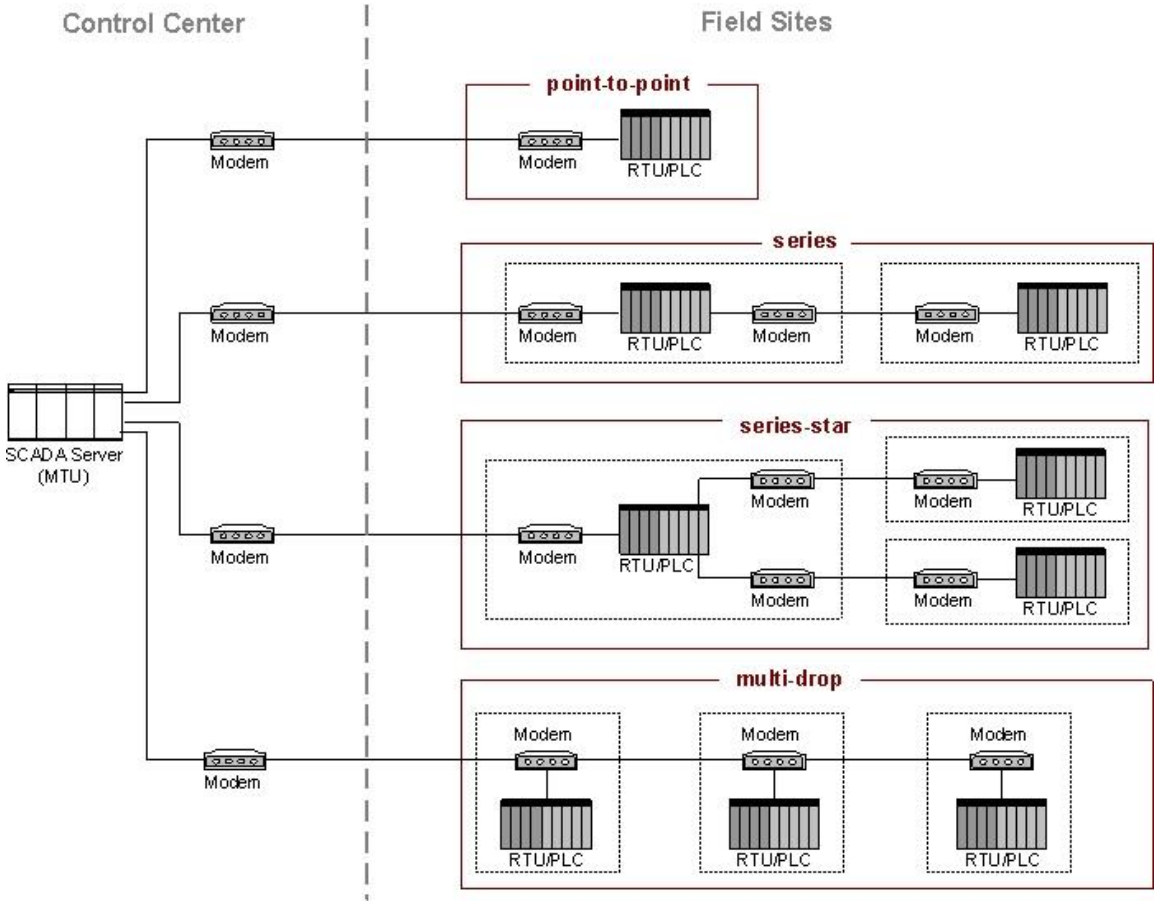

Figure 2-3. Basic SCADA Communication Topologies

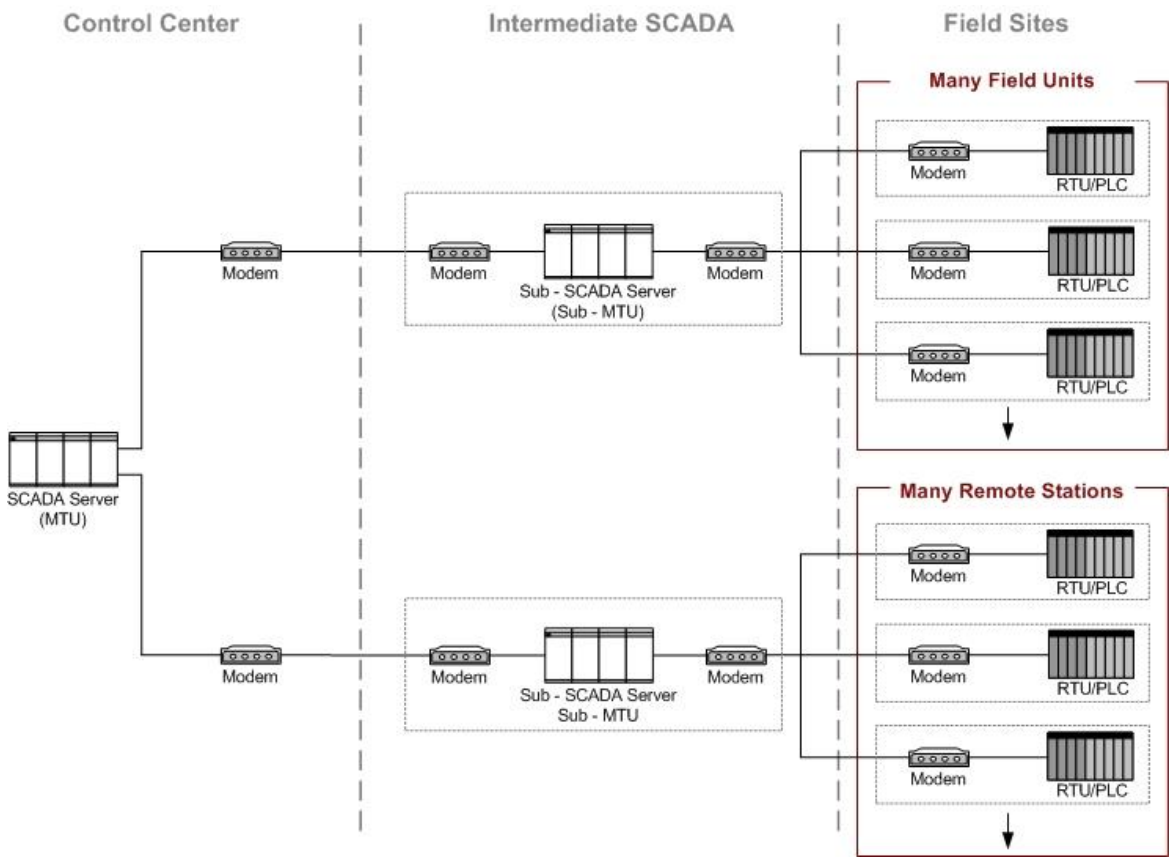

Figure 2-4. Large SCADA Communication Topology

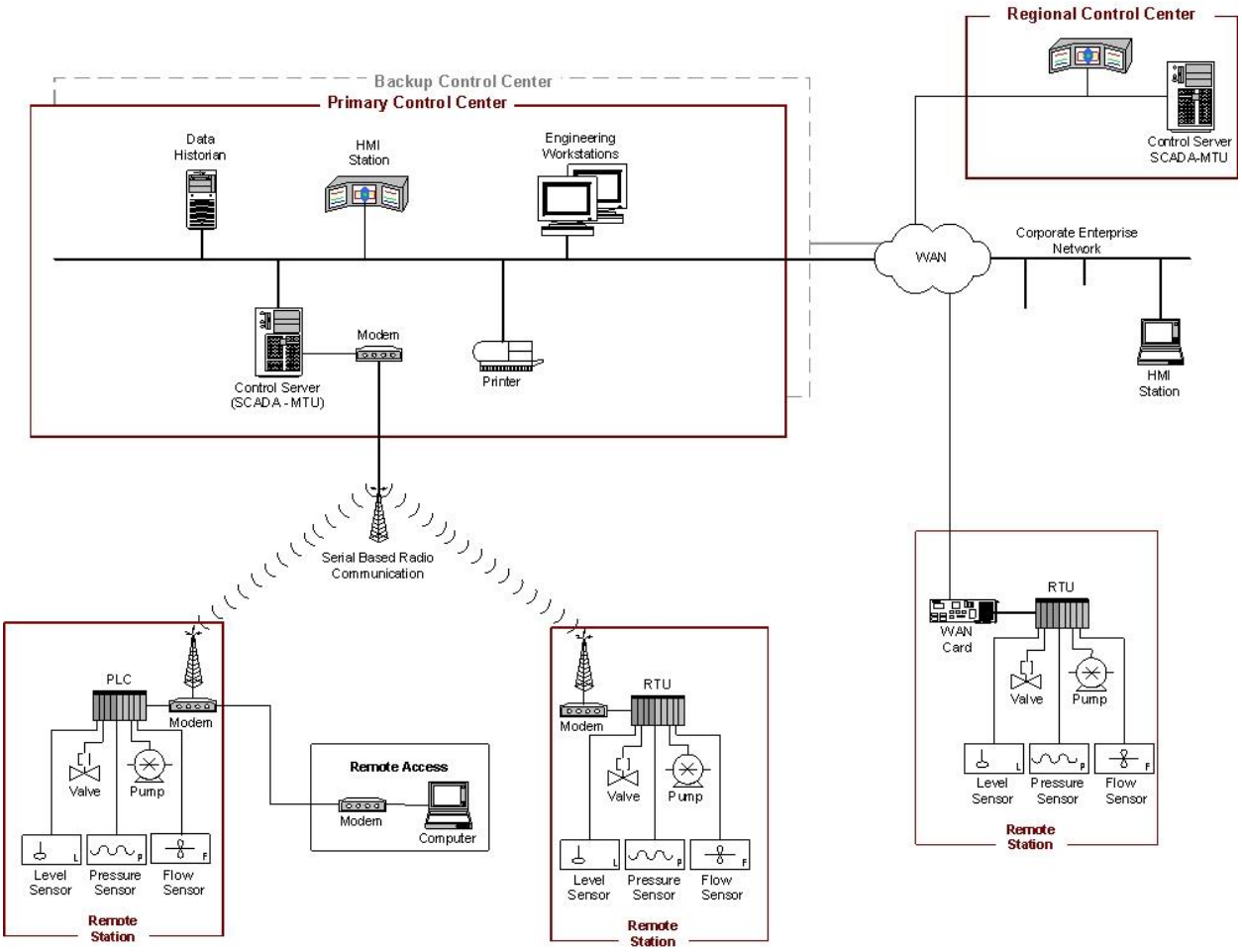

Figure 2-5. SCADA System Implementation Example (Distribution Monitoring and Control)

Figure 2-6 shows an example implementation for rail monitoring and control. This example includes a rail control center that houses the SCADA system and three sections of a rail system. The SCADA system polls the rail sections for information such as the status of the trains, signal systems, traction electrification systems, and ticket vending machines. This information is also fed to operator consoles at the HMI station within the rail control center. The SCADA system also monitors operator inputs at the rail control center and disperses high-level operator commands to the rail section components. In addition, the SCADA system monitors conditions at the individual rail sections and issues commands based on these conditions (e.g., stopping a train to prevent it from entering an area that has been determined to be flooded or occupied by another train based on condition monitoring).

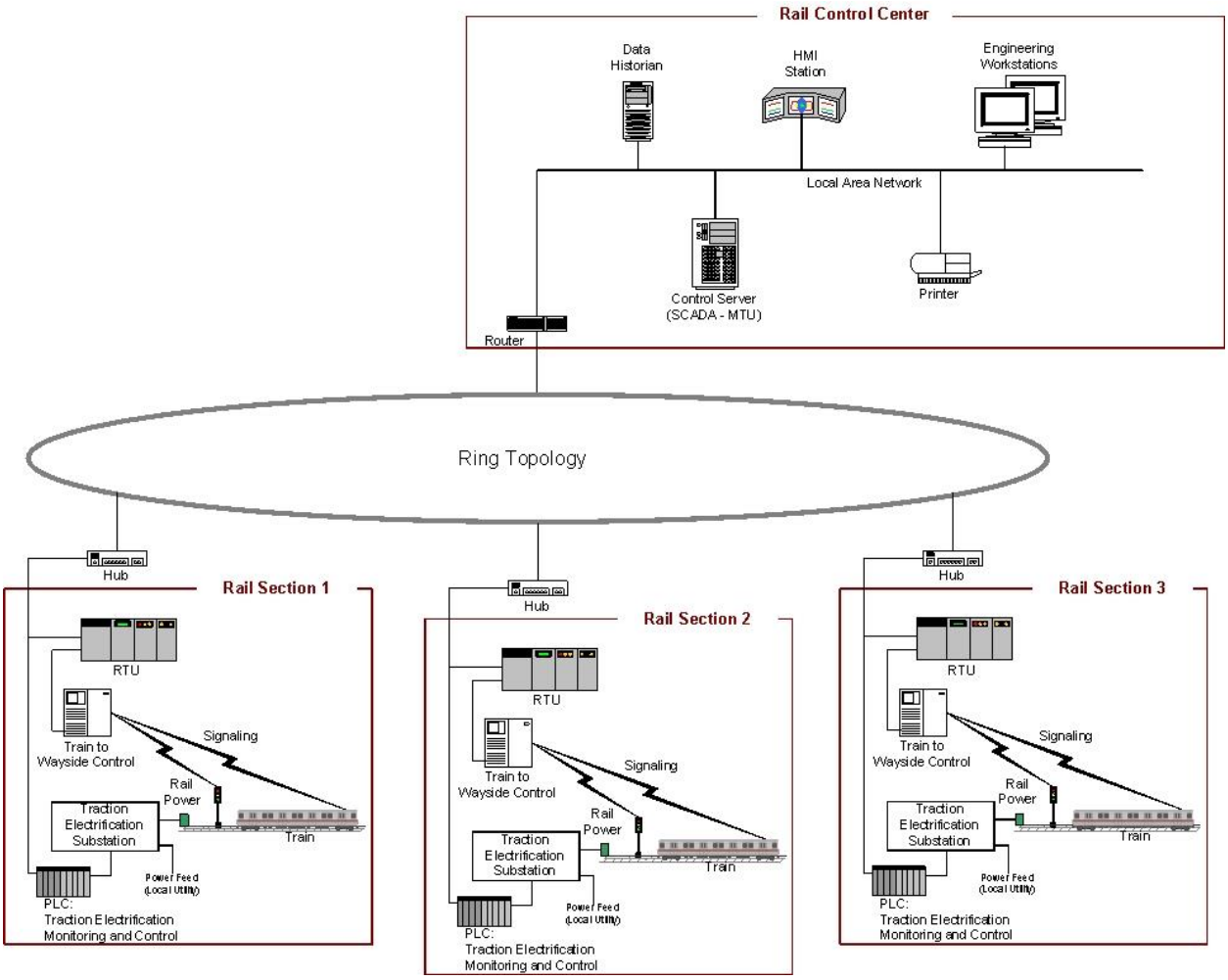

Figure 2-6. SCADA System Implementation Example (Rail Monitoring and Control)

2.3.3 Distributed Control Systems

DCS are used to control production systems within the same geographic location for industries such as oil refineries, water and wastewater treatment, electric power generation plants, chemical manufacturing plants, automotive production, and pharmaceutical processing facilities. These systems are usually process control or discrete part control systems.

DCS are integrated as a control architecture containing a supervisory level of control overseeing multiple, integrated sub-systems that are responsible for controlling the details of a localized process. A DCS uses a centralized supervisory control loop to mediate a group of localized controllers that share the overall tasks of carrying out an entire production process [6]. Product and process control are usually achieved by deploying feedback or feedforward control loops whereby key product and/or process conditions are automatically maintained around a desired set point. To accomplish the desired product and/or process tolerance around a specified set point, specific process controllers, or more capable PLCs, are employed in the field and are tuned to provide the desired tolerance as well as the rate of self-correction during process upsets. By modularizing the production system, a DCS reduces the impact of a single fault on the

overall system. In many modern systems, the DCS is interfaced with the corporate network to give business operations a view of production.

An example implementation showing the components and general configuration of a DCS is depicted in Figure 2-7. This DCS encompasses an entire facility from the bottom-level production processes up to the corporate or enterprise layer. In this example, a supervisory controller (control server) communicates to its subordinates via a control network. The supervisor sends set points to and requests data from the distributed field controllers. The distributed controllers control their process actuators based on control server commands and sensor feedback from process sensors.

Figure 2-7 gives examples of low-level controllers found on a DCS system. The field control devices shown include a PLC, a process controller, a single loop controller, and a machine controller. The single loop controller interfaces sensors and actuators using point-to-point wiring, while the other three field devices incorporate fieldbus networks to interface with process sensors and actuators. Fieldbus networks eliminate the need for point-to-point wiring between a controller and individual field sensors and actuators. Additionally, a fieldbus allows greater functionality beyond control, including field device diagnostics, and can accomplish control algorithms within the fieldbus, thereby avoiding signal routing back to the PLC for every control operation. Standard industrial communication protocols designed by industry groups such as Modbus and Fieldbus [7] are often used on control networks and fieldbus networks.

In addition to the supervisory-level and field-level control loops, intermediate levels of control may also exist. For example, in the case of a DCS controlling a discrete part manufacturing facility, there could be an intermediate level supervisor for each cell within the plant. This supervisor would encompass a manufacturing cell containing a machine controller that processes a part and a robot controller that handles raw stock and final products. There could be several of these cells that manage field-level controllers under the main DCS supervisory control loop.

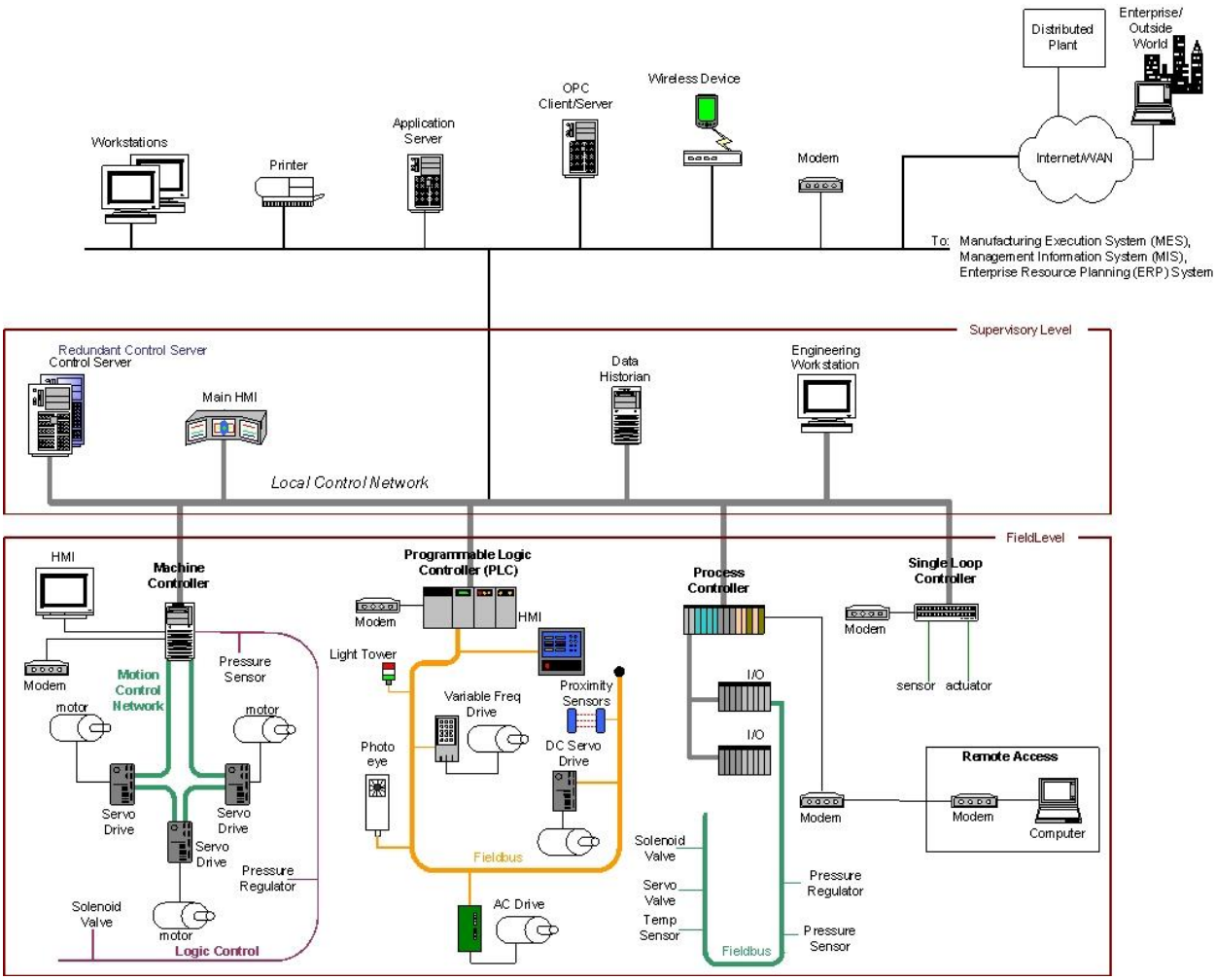

Figure 2-7. DCS Implementation Example

2.3.4 Programmable Logic Controller Based Topologies

PLCs are used in both SCADA and DCS systems as the control components of an overall hierarchical system to provide local management of processes through feedback control as described in the sections above. In the case of SCADA systems, they may provide the same functionality of RTUs. When used in DCS, PLCs are implemented as local controllers within a supervisory control scheme.

In addition to PLC usage in SCADA and DCS, PLCs are also implemented as the primary controller in smaller control system configurations to provide operational control of discrete processes such as automobile assembly lines and power plant soot blower controls These topologies differ from SCADA and DCS in that they generally lack a central control server and HMI and, therefore, primarily provide closed-loop control without direct human involvement. PLCs have a user-programmable memory for storing instructions for the purpose of implementing specific functions such as I/O control, logic, timing, counting, three mode proportional-integral-derivative (PID) control, communication, arithmetic, and data and file processing.

Figure 2-8 shows control of a manufacturing process being performed by a PLC over a fieldbus network. The PLC is accessible via a programming interface located on an engineering workstation, and data is stored in a data historian, all connected on a LAN.

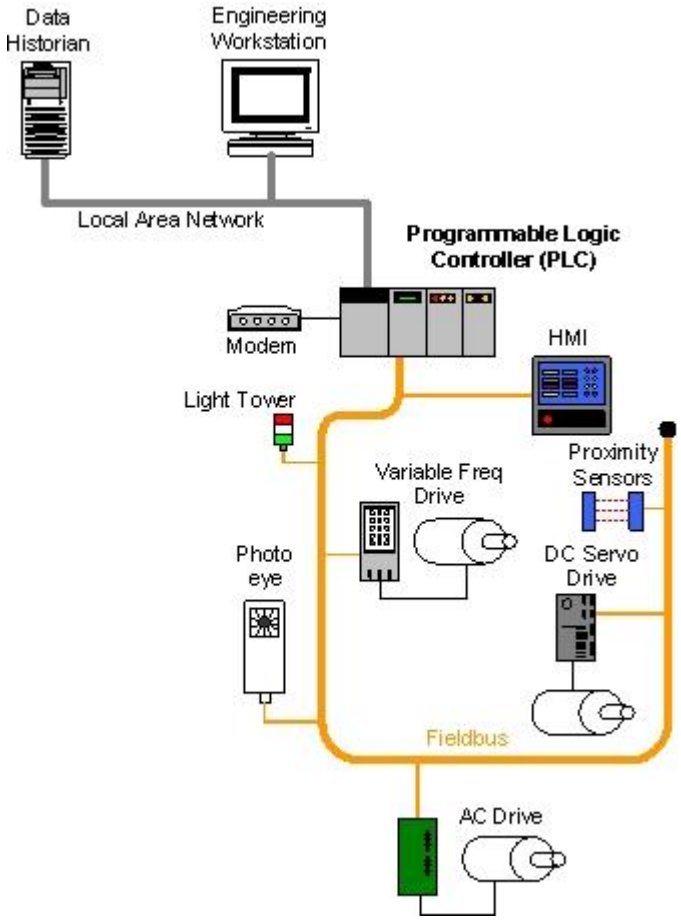

Figure 2-8. PLC Control System Implementation Example

2.4 Comparing ICS and IT Systems Security

ICS control the physical world and IT systems manage data. ICS have many characteristics that differ from traditional IT systems, including different risks and priorities. Some of these include significant risk to the health and safety of human lives, serious damage to the environment, and financial issues such as production losses, and negative impact to a nation's economy. ICS have different performance and reliability requirements, and also use operating systems and applications that may be considered unconventional in a typical IT network environment. Security protections must be implemented in a way that maintains system integrity during normal operations as well as during times of cyber attack [17].

Initially, ICS had little resemblance to IT systems in that ICS were isolated systems running proprietary control protocols using specialized hardware and software. Widely available, low-cost Ethernet and Internet Protocol (IP) devices are now replacing the older proprietary technologies, which increases the possibility of cybersecurity vulnerabilities and incidents. As ICS are adopting IT solutions to promote corporate connectivity and remote access capabilities, and are being designed and implemented using industry standard computers, operating systems (OS) and network protocols, they are starting to resemble IT systems. This integration supports new IT capabilities, but it provides significantly less isolation for ICS from the outside world than predecessor systems, creating a greater need to secure these systems. While security solutions have been designed to deal with these security issues in typical IT systems, special precautions must be taken when introducing these same solutions to ICS environments. In some cases, new security solutions are needed that are tailored to the ICS environment.

The environments in which ICS and IT systems operate are constantly changing. The environments of operation include, but are not limited to: the threat space; vulnerabilities; missions/business functions; mission/business processes; enterprise and information security architectures; information technologies; personnel; facilities; supply chain relationships; organizational governance/culture; procurement/acquisition processes; organizational policies/procedures; organizational assumptions, constraints, risk tolerance, and priorities/trade-offs).

The following lists some special considerations when considering security for ICS:

- **Timeliness and Performance Requirements.** ICS are generally time-critical, with the criterion for acceptable levels of delay and jitter dictated by the individual installation. Some systems require reliable, deterministic responses. High throughput is typically not essential to ICS. In contrast, IT systems typically require high throughput, and they can typically withstand some level of delay and jitter. For some ICS, automated response time or system response to human interaction is very critical. Some ICS are built on real-time operating systems (RTOS), where real-time refers to timeliness requirements. The units of real-time are very application dependent and must be explicitly stated.
- **Availability Requirements.** Many ICS processes are continuous in nature. Unexpected outages of systems that control industrial processes are not acceptable. Outages often must be planned and scheduled days or weeks in advance. Exhaustive pre-deployment testing is essential to ensure high availability (i.e., reliability) for the ICS. Control systems often cannot be easily stopped and started without affecting production. In some cases, the products being produced or equipment being used is more important than the information being relayed. Therefore, the use of typical IT strategies such as rebooting a component, are usually not acceptable solutions due to the adverse impact on the requirements for high availability, reliability and maintainability of the ICS. Some ICS employ redundant components, often running in parallel, to provide continuity when primary components are unavailable.

- **Risk Management Requirements.** In a typical IT system, data confidentiality and integrity are typically the primary concerns. For an ICS, human safety and fault tolerance to prevent loss of life or endangerment of public health or confidence, regulatory compliance, loss of equipment, loss of intellectual property, or lost or damaged products are the primary concerns. The personnel responsible for operating, securing, and maintaining ICS must understand the important link between safety and security. Any security measure that impairs safety is unacceptable.

- **Physical Effects.** ICS field devices (e.g., PLC, operator station, DCS controller) are directly responsible for controlling physical processes. ICS can have very complex interactions with physical processes and consequences in the ICS domain that can manifest in physical events. Understanding these potential physical effects often requires communication between experts in control systems and in the particular physical domain.

- **System Operation.** ICS operating systems (OS) and control networks are often quite different from IT counterparts, requiring different skill sets, experience, and levels of expertise. Control networks are typically managed by control engineers, not IT personnel. Assumptions that differences are not significant can have disastrous consequences on system operations.

- **Resource Constraints.** ICS and their real time OSs are often resource-constrained systems that do not include typical contemporary IT security capabilities. Legacy systems are often lacking resources common on modern IT systems. Many systems may not have desired features including encryption capabilities, error logging, and password protection. Indiscriminate use of IT security practices in ICS may cause availability and timing disruptions. There may not be computing resources available on ICS components to retrofit these systems with current security capabilities. Adding resources or features may not be possible.

- **Communications.** Communication protocols and media used by ICS environments for field device control and intra-processor communication are typically different from most IT environments, and may be proprietary.

- **Change Management.** Change management is paramount to maintaining the integrity of both IT and control systems. Unpatched software represents one of the greatest vulnerabilities to a system. Software updates on IT systems, including security patches, are typically applied in a timely fashion based on appropriate security policy and procedures. In addition, these procedures are often automated using server-based tools. Software updates on ICS cannot always be implemented on a timely basis. These updates need to be thoroughly tested by both the vendor of the industrial control application and the end user of the application before being implemented. Additionally, the ICS owner must plan and schedule ICS outages days/weeks in advance. The ICS may also require revalidation as part of the update process. Another issue is that many ICS utilize older versions of operating systems that are no longer supported by the vendor. Consequently, available patches may not be applicable. Change management is also applicable to hardware and firmware. The change management process, when applied to ICS, requires careful assessment by ICS experts (e.g., control engineers) working in conjunction with security and IT personnel.

- **Managed Support.** Typical IT systems allow for diversified support styles, perhaps supporting disparate but interconnected technology architectures. For ICS, service support is sometimes via a single vendor, which may not have a diversified and interoperable support solution from another vendor. In some instances, third-party security solutions are not allowed due to ICS vendor license and service agreements, and loss of service support can occur if third party applications are installed without vendor acknowledgement or approval.

- **Component Lifetime.** Typical IT components have a lifetime on the order of 3 to 5 years, with brevity due to the quick evolution of technology. For ICS where technology has been developed in many cases for very specific use and implementation, the lifetime of the deployed technology is often in the order of 10 to 15 years and sometimes longer.
- **Component Location.** Most IT components and some ICS are located in business and commercial facilities physically accessible by local transportation. Remote locations may be utilized for backup facilities. Distributed ICS components may be isolated, remote, and require extensive transportation effort to reach. Component location also needs to consider necessary physical and environmental security measures.

Table 2-1 summarizes some of the typical differences between IT systems and ICS.

Table 2-1. Summary of IT System and ICS Differences

Category	Information Technology System	Industrial Control System
Performance Requirements	Non-real-time Response must be consistent High throughput is demanded High delay and jitter may be acceptable Less critical emergency interaction Tightly restricted access control can be implemented to the degree necessary for security	Real-time Response is time-critical Modest throughput is acceptable High delay and/or jitter is not acceptable Response to human and other emergency interaction is critical Access to ICS should be strictly controlled, but should not hamper or interfere with human-machine interaction
Availability (Reliability) Requirements	Responses such as rebooting are acceptable Availability deficiencies can often be tolerated, depending on the system's operational requirements	Responses such as rebooting may not be acceptable because of process availability requirements Availability requirements may necessitate redundant systems Outages must be planned and scheduled days/weeks in advance High availability requires exhaustive pre-deployment testing
Risk Management Requirements	Manage data Data confidentiality and integrity is paramount Fault tolerance is less important – momentary downtime is not a major risk Major risk impact is delay of business operations	Control physical world Human safety is paramount, followed by protection of the process Fault tolerance is essential, even momentary downtime may not be acceptable Major risk impacts are regulatory non-compliance, environmental impacts, loss of life, equipment, or production
System Operation	Systems are designed for use with typical operating systems Upgrades are straightforward with the availability of automated deployment tools	Differing and possibly proprietary operating systems, often without security capabilities built in Software changes must be carefully made, usually by software vendors, because of the specialized control algorithms and perhaps modified hardware and software involved
Resource Constraints	Systems are specified with enough resources to support the addition of third-party applications such as security solutions	Systems are designed to support the intended industrial process and may not have enough memory and computing resources to support the addition of security capabilities

Category	Information Technology System	Industrial Control System
Communications	Standard communications protocols Primarily wired networks with some localized wireless capabilities Typical IT networking practices	Many proprietary and standard communication protocols Several types of communications media used including dedicated wire and wireless (radio and satellite) Networks are complex and sometimes require the expertise of control engineers
Change Management	Software changes are applied in a timely fashion in the presence of good security policy and procedures. The procedures are often automated.	Software changes must be thoroughly tested and deployed incrementally throughout a system to ensure that the integrity of the control system is maintained. ICS outages often must be planned and scheduled days/weeks in advance. ICS may use OSs that are no longer supported
Managed Support	Allow for diversified support styles	Service support is usually via a single vendor
Component Lifetime	Lifetime on the order of 3 to 5 years	Lifetime on the order of 10 to 15 years
Components Location	Components are usually local and easy to access	Components can be isolated, remote, and require extensive physical effort to gain access to them

In summary, the operational and risk differences between ICS and IT systems create the need for increased sophistication in applying cybersecurity and operational strategies. A cross-functional team of control engineers, control system operators and IT security professionals needs to work closely to understand the possible implications of the installation, operation, and maintenance of security solutions in conjunction with control system operation. IT professionals working with ICS need to understand the reliability impacts of information security technologies before deployment. Some of the OSs and applications running on ICS may not operate correctly with commercial-off-the-shelf (COTS) IT cybersecurity solutions because of specialized ICS environment architectures.

2.5 Other Types of Control Systems

Although this guide provides guidance for securing ICS, other types of control systems share similar characteristics and many of the recommendations from this guide are applicable and could be used as a reference to protect such systems against cybersecurity threats. For example, although many building, transportation, medical, security and logistics systems use different protocols, ports and services, and are configured and operate in different modes than ICS, they share similar characteristics to traditional ICS [18]. Examples of some of these systems and protocols include:

Other Types of Control Systems

- Advanced Metering Infrastructure.
- Building Automation Systems.
- Building Management Control Systems.
- Closed-Circuit Television (CCTV) Surveillance Systems.
- CO2 Monitoring.
- Digital Signage Systems.
- Digital Video Management Systems.
- Electronic Security Systems.
- Emergency Management Systems.

- Energy Management Systems.
- Exterior Lighting Control Systems.
- Fire Alarm Systems.
- Fire Sprinkler Systems.
- Interior Lighting Control Systems.
- Intrusion Detection Systems.
- Physical Access Control Systems.
- Public Safety/Land Mobile Radios.
- Renewable Energy Geothermal Systems.
- Renewable Energy Photo Voltaic Systems.
- Shade Control Systems.
- Smoke and Purge Systems.
- Vertical Transport System (Elevators and Escalators).
- Laboratory Instrument Control Systems.
- Laboratory Information Management Systems (LIMS).

Protocols/Ports and Services

- Modbus: Master/Slave - Port 502.
- BACnet[2]: Master/Slave - Port 47808.
- LonWorks/LonTalk[3]: Peer to Peer - Port 1679.
- DNP3: Master/Slave – Port 19999 when using Transport Layer Security (TLS), Port 20000 when not using TLS.
- IEEE 802.x - Peer to Peer.
- ZigBee - Peer to Peer.
- Bluetooth – Master/Slave.

The security controls provided in Appendix G— of this guide are general and flexible enough be used to evaluate other types of control systems, but subject matter experts should review the controls and tailor them as appropriate to address the uniqueness of other types of control systems. There is no "one size fits all," and the risks may not be the same, even within a particular group. For example, a building has many different sub-systems such as building automation, fire alarm, physical access control, digital signage, CCTV, etc. Critical life safety systems such as the fire alarm and physical access control systems may drive the impact level to be a "High," while the other systems will usually be "Low." An organization might decide to evaluate each sub-system individually, or decide to use an aggregated approach. The control systems evaluation should be coupled to the Business Impact, Contingency Plan, and Incident Response Plan to ensure organizational critical functions and operations can be recovered and restored as defined by the organizations Recovery Time Objectives.

[2] http://www.bacnet.org/
[3] http://en.wikipedia.org/wiki/LonWorks

3. ICS Risk Management and Assessment

3.1 Risk Management

Organizations manage risk every day in meeting their business objectives. These risks may include financial risk, risk of equipment failure, and personnel safety risk, to name just a few. Organizations must develop processes to evaluate the risks associated with their business and to decide how to deal with those risks based on organizational priorities and both internal and external constraints. This management of risk is conducted as an interactive, ongoing process as part of normal operations. Organizations that use ICS have historically managed risk through good practices in safety and engineering. Safety assessments are well established in most sectors and are often incorporated into regulatory requirements. Information security risk management is an added dimension that can be complementary. The risk management process and framework outlined in this section can be applied to any risk assessment including both safety and information security.

A risk management process should be employed throughout an organization, using a three-tiered approach to address risk at the (i) organization level; (ii) mission/business process level; and (iii) information system level (IT and ICS). The risk management process is carried out seamlessly across the three tiers with the overall objective of continuous improvement in the organization's risk-related activities and effective inter-tier and intra-tier communication among all stakeholders having a shared interest in the mission/business success of the organization.

This section focuses primarily on ICS considerations at the information system level, however, it is important to note that the risk management activities, information, and artifacts at each tier impact and inform the other tiers. Section 6 extends the concepts presented here to the control family level and provides ICS-specific recommendations to augment security control families. Throughout the following discussion of risk management, ICS considerations will be highlighted and the impact that these considerations have on the risk management process will be discussed.

For more information on multi-tiered risk management and the risk management process, refer to NIST Special Publication 800-39, *Managing Information Security Risk: Organization, Mission and Information System View* [20]. NIST Special Publication 800-37 Revision 1, *Guide for Applying the Risk Management Framework to Federal Information Systems: A Security Life Cycle Approach* [21], provides guidelines for applying the Risk Management Framework to federal information systems to include conducting the activities of security categorization,[4] security control selection and implementation, security control assessment, information system authorization,[5] and security control monitoring. NIST Special Publication 800-30, *Guide for Conducting Risk Assessments*, provides a step-by-step process for organizations on: (i) how to prepare for risk assessments; (ii) how to conduct risk assessments; (iii) how to communicate risk assessment results to key organizational personnel; and (iv) how to maintain the risk assessments over time [79].

[4] FIPS 199 provides security categorization guidance for non-national security systems [15]. CNSS Instruction 1253 provides similar guidance for national security systems.

[5] Security authorization is the official management decision given by a senior organizational official to authorize operation of an information system and to explicitly accept the risk to organizational operations and assets, individuals, other organizations, and the Nation based on the implementation of an agreed-upon set of security controls.

3.2 Introduction to the Risk Management Process

As shown in Figure 3-1, the risk management process has four components: *framing*, *assessing*, *responding* and *monitoring*. These activities are interdependent and often occur simultaneously within an organization. For example, the results of the monitoring component will feed into the framing component. As the environment in which organizations operate is always changing, risk management must be a continuous process where all components have on-going activities. It is important to remember that these components apply to the management of any risk whether information security, physical security, safety or financial.

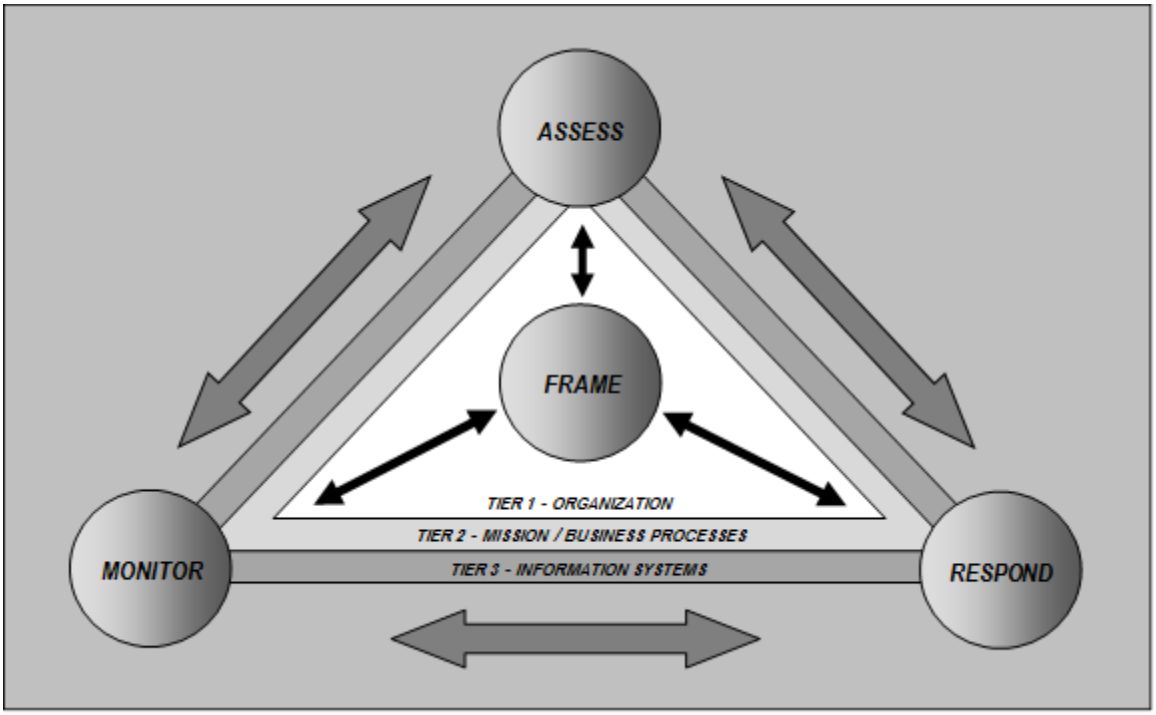

Figure 3-1. Risk Management Process Applied Across the Tiers

The *framing component* in the risk management process consists of developing a framework for the risk management decisions to be made. The level of risk that an organization is willing to accept is its *risk tolerance* [21, p.6].

The framing component should include review of existing documentation, such as prior risk assessments. There may be related activities; such as community wide disaster management planning that also should be considered since they impact the requirements that a risk assessment must consider.

ICS-specific Recommendations and Guidance

For operators of ICS, safety is the major consideration that directly affects decisions on how systems are engineered and operated. Safety can be defined as "freedom from conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment."[6] Part of the framing component for an ICS organization is determining how these requirements interact with information security. For example, if safety requirements conflict with good security practice, how will the organization decide between the two priorities? Most ICS operators would answer that safety is the main consideration – the framing component makes such assumptions explicit so that there is agreement throughout the process and the organization.
Another major concern for ICS operators is the availability of services provided by the ICS. The ICS may be part of critical infrastructure (for example, water or power systems), where there is a significant need for continuous and reliable operations. As a result, ICS may have strict requirements for availability or for recovery. Such assumptions should be developed and stated in the framing component. Otherwise, the organization may make risk decisions that result in unintended consequences on those who depend on the services provided.

The physical operating environment is another aspect of risk framing that organizations should consider when working with ICS. ICS often have specific environmental requirements (e.g., a manufacturing process may require precise temperature), or they may be tied to their physical environment for operations. Such requirements and constraints should be explicitly stated in the framing component so that the risks arising from these constraints can be identified and considered.

Assessing risk requires that organizations identify their threats and vulnerabilities, the harm that such threats and vulnerabilities may cause the organization and the likelihood that adverse events arising from those threats and vulnerabilities may actually occur.

ICS-specific Recommendations and Guidance

The DHS National Cybersecurity & Communications Integration Center (NCCIC)[7] serves as a centralized location where operational elements involved in cybersecurity and communications reliance are coordinated and integrated. The Industrial Control Systems Cyber Emergency Response Team (ICS-CERT)[8] collaborates with international and private sector Computer Emergency Response Teams (CERTs) to share control systems-related security incidents and mitigation measures. ICS-CERT works to reduce risks within and across all critical infrastructure sectors by partnering with law enforcement agencies and the intelligence community and coordinating efforts among Federal, state, local, and tribal governments and control systems owners, operators, and vendors.

When assessing the potential impact to an organization's mission from a potential ICS incident, it is important to incorporate the effect on the physical process/system, impact on dependent systems/processes, and impact on the physical environment among other possibilities. In addition, the potential impact on safety should always be considered.

[6] MIL-STD-882E, *Standard Practice – System Safety*, Department of Defense (DoD), May 11, 2012, https://acc.dau.mil/CommunityBrowser.aspx?id=683694
[7] http://www.dhs.gov/about-national-cybersecurity-communications-integration-center
[8] https://ics-cert.us-cert.gov/

The *responding component* is based on the concept of a consistent organization-wide response to the *identification* of risk. Response to identification of risk (as opposed to the response to an incident) requires that organizations first identify possible courses of actions to address risk, evaluate those possibilities in light of the organization's risk tolerance and other considerations determined during the framing step, and choose the best alternative for the organization. The response component includes the implementation of the chosen course of action to address the identified risk: *acceptance, avoidance, mitigation, sharing, transfer,* or any combination of those options[9].

ICS-specific Recommendations and Guidance

For ICS, available risk responses may be constrained by system requirements, potential adverse impact on operations, or even regulatory compliance regimes. An example of risk sharing is when utilities enter into agreements to "loan" line workers in an emergency, which reduces the duration of the effect of an incident to acceptable levels.

Monitoring is the fourth component of the risk management activities. Organizations must monitor risk on an on-going basis including: the implementation of chosen risk management strategies; the changes in the environment that may affect the risk calculation; and, the effectiveness and efficiency of risk reduction activities. The activities in the monitoring component impact all the other components.

3.3 Special Considerations for Doing an ICS Risk Assessment

The nature of ICS means that when an organization does a risk assessment, there may be additional considerations that do not exist when doing a risk assessment of a traditional IT system. Because the impact of a cyber incident in an ICS may include both physical and digital effects, risk assessments need to incorporate those potential effects. This section will provide a more in-depth examination of the following:

- Impacts on safety and use of safety assessments.
- Physical impact of a cyber incident on an ICS, including the larger physical environment; effect on the process controlled, and the physical effect on the ICS itself.
- The consequences for risk assessments of non-digital control components within an ICS.

3.3.1 Safety within an ICS Information Security Risk Assessment

The culture of safety and safety assessments is well established within the majority of the ICS user community. Information security risk assessments should be seen as complementary to such assessments though the assessments may use different approaches and cover different areas. Safety assessments are concerned primarily with the physical world. Information security risk assessments primarily look at the digital world. However, in an ICS environment, the physical and the digital are intertwined and significant overlap may occur.

It is important that organizations consider all aspects of risk management for safety (e.g., risk framing, risk tolerances), as well as the safety assessment results, when carrying out risk assessments for information security. The personnel responsible for the information security risk assessment must be able

[9] For additional information on accepting, avoiding, mitigating, sharing, or transferring risk, refer to NIST Special Publication 800-39 [20].

to identify and communicate identified risks that could have safety implications. Conversely, the personnel charged with safety assessments must be familiar with the potential physical impacts and their likelihood developed by the information security risk assessment process.

3.3.2 Potential Physical Impacts of an ICS Incident

Evaluating the potential physical damage from a cyber incident should incorporate: i) how an incident could manipulate the operation of sensors and actuators to impact the physical environment; ii) what redundant controls exist in the ICS to prevent an impact; and iii) how a physical incident could emerge based on these conditions. A physical impact could negatively impact the surrounding world through multiple means, including the release of hazardous materials (e.g., pollution, crude oil), damaging kinetic forces (e.g., explosions), and exposure to energy sources (e.g., electricity, steam). The physical incident could negatively impact the ICS and supporting infrastructure, the various processes performed by the ICS, or the larger physical environment. An evaluation of the potential physical impacts should include all parts of an ICS, beginning with evaluating the potential impacts on the set of sensor and actuators. Each of these domains will be further explored below.

Evaluating the impact of a cyber incident on the physical environment should focus on potential damage to human safety, the natural environment, and other critical infrastructures. Human safety impacts should be evaluated based on whether injury, disease, or death is possible from a malfunction of the ICS. This should incorporate any previously performed safety impact assessments performed by the organization regarding both employees and the general public. Environmental impacts also may need to be addressed. This analysis should incorporate any available environmental impact assessments performed by the organization to determine how an incident could impact natural resources and wildlife over the short or long term. In addition, it should be noted that ICS may not be located within a single, controlled location and can be distributed over a wide physical area and exposed to uncontrolled environments. Finally, the impact on the physical environment should explore the extent to which an incident could damage infrastructures external to the ICS (e.g., electric generation/delivery, transportation infrastructures, and water services).

3.3.3 Impact of Physical Disruption of an ICS Process

In addition to the impact on the physical environment, the risk assessment should also evaluate potential effects to the physical process performed by the ICS under consideration, as well as other systems. An incident that impacts the ICS and disrupts the dependent process may cause cascading impacts into other related ICS processes and the general public's dependence on the resulting products and services. Impact to related ICS processes could include both systems and processes within the organization (e.g., a manufacturing process that depends on the process controlled by the system under consideration) or systems and processes external to the organization (e.g., a utility selling generated energy to a nearby plant).

A cyber incident can also negatively impact the physical ICS under consideration. This type of impact primarily includes the physical infrastructure of the plant (e.g., tanks, valves, motors), along with both the digital and non-digital control mechanisms (e.g., cables, PLCs, pressure gauge). Damage to the ICS or physical plant may cause either short or long term outages depending on the degree of the incident. An example of a cyber incident impacting the ICS is the Stuxnet malware, which caused physical damage to the centrifuges as well as disrupting dependent processes.

3.3.4 Incorporating Non-digital Aspects of ICS into Impact Evaluations

The impacts on the ICS cannot be adequately determined by focusing only on the digital aspects of the system, as there are often non-digital mechanisms available that provide fault tolerance and prevent the ICS from acting outside of acceptable parameters. Therefore, these mechanisms may help reduce any negative impact that a digital incident on the ICS might have and must be incorporated into the risk assessment process. For example, ICS often have non-digital control mechanisms that can prevent the ICS from operating outside of a safe boundary, and thereby limit the impact of an attack (e.g., a mechanical relief pressure valve). In addition, analog mechanisms (e.g., meters, alarms) can be used to observe the physical system state to provide operators with reliable data if digital readings are unavailable or corrupted. Table 3-1 provides a categorization of non-digital control mechanisms that could be available to reduce the impact of an ICS incident.

Table 3-1. Categories of Non-Digital ICS Control Components

System Type	Description
Analog Displays or Alarms	Non-digital mechanisms that measure and display the state of the physical system (e.g., temperature, pressure, voltage, current) and can provide the operator with accurate information in situations when digital displays are unavailable or corrupted. The information may be provided to the operator on some non-digital display (e.g., thermometers, pressure gauges) and through audible alarms.
Manual Control Mechanisms	Manual control mechanisms (e.g., manual valve controls, physical breaker switches) provide operators with the ability to manually control an actuator without relying on the digital control system. This ensures that an actuator can be controlled even if the control system is unavailable or compromised.
Analog Control Systems	Analog control systems use non-digital sensors and actuators to monitor and control a physical process. These may be able to prevent the physical process from entering an undesired state in situations when the digital control system is unavailable or corrupted. Analog controls include devices such as regulators, governors, and electromechanical relays.

Determination of the potential impact that a cyber incident may have on the ICS should incorporate analysis of all non-digital control mechanisms and the extent to which they can mitigate potential negative impacts to the ICS. There are multiple considerations when considering the possible mitigation effects of non-digital control mechanisms, such as:

- Non-digital control mechanisms may require additional time and human involvement to perform necessary monitoring or control functions and these efforts may be substantial. For example, such mechanisms may require operators to travel to a remote site to perform certain control functions. Such mechanisms may also depend on human response times, which may be slower than automated controls.

- Manual and analog systems may not provide monitoring or control capabilities with the same degree of accuracy and reliability as the digital control system. This may present risk if the primary control system is unavailable or corrupted due to reduced quality, safety, or efficiency of the system. For example, a digital/numeric protection relay provides more accuracy and reliable detection of faults than analog/static relays, therefore, the system maybe more likely to exhibit a spurious relay tripping if the digital relays are not available.

3.3.5 Incorporating the Impact of Safety Systems

Safety systems may also reduce the impact of a cyber incident to the ICS. Safety systems are often deployed to perform specific monitoring and control functions to ensure the safety of people, the environment, process, and ICS. While these systems are traditionally implemented to be fully redundant with respect to the primary ICS, they may not provide complete redundancy from cyber incidents, specifically from a sophisticated attacker. The impact of the implemented security controls on the safety system should be evaluated to determine that they do not negatively impact the system.

3.3.6 Considering the Propagation of Impact to Connected Systems

Evaluating the impact of an incident must also incorporate how the impact from the ICS could *propagate* to a connected ICS or physical system. An ICS may be interconnected with other systems, such that failures in one system or process can easily cascade to other systems either within or external to the organization. Impact propagation could occur due to both physical and logical dependencies. Proper communication of the results of risk assessments to the operators of connected or interdependent systems and processes is one way to mitigate such impacts.

Logical damage to an interconnected ICS could occur if the cyber incident propagated to the connected control systems. An example could be if a virus or worm propagated to a connected ICS and then impacted that system. Physical damage could also propagate to other interconnected ICS. If an incident impacts the physical environment of an ICS, it may also impact other related physical domains. For example, the impact could result in a physical hazard which degrades nearby physical environments. Additionally, the impact could also degrade the common shared dependencies (e.g., power supply), or result in a shortage of material needed for a later stage in an industrial process.

5. ICS Security Architecture

When designing a network architecture for an ICS deployment, it is usually recommended to separate the ICS network from the corporate network. The nature of network traffic on these two networks is different: Internet access, FTP, email, and remote access will typically be permitted on the corporate network but should not be allowed on the ICS network. Rigorous change control procedures for network equipment, configuration, and software changes may not be in place on the corporate network. If ICS network traffic is carried on the corporate network, it could be intercepted or be subjected to DoS or Man-in-the-Middle attacks [5.14]. By having separate networks, security and performance problems on the corporate network should not be able to affect the ICS network.

Practical considerations, such as cost of ICS installation or maintaining a homogenous network infrastructure, often mean that a connection is required between the ICS and corporate networks. This connection is a significant security risk and should be protected by boundary protection devices. If the networks must be connected, it is strongly recommended that only minimal (single if possible) connections be allowed and that the connection is through a firewall and a DMZ. A DMZ is a separate network segment that connects directly to the firewall. Servers containing the data from the ICS that needs to be accessed from the corporate network are put on this network segment. Only these systems should be accessible from the corporate network. With any external connections, the minimum access should be permitted through the firewall, including opening only the ports required for specific communication. The following sections elaborate on these architectural considerations. The ICS-CERT recommended practices working group provides additional guidance as recommended practices[11].

5.1 Network Segmentation and Segregation

This section addresses partitioning the ICS into security domains and separating the ICS from other networks, such as the corporate network, and presents illustrative security architecture. Operational risk analysis should be performed to determine critical parts of each ICS network and operation and help define what parts of the ICS need to be segmented. Network segmentation involves partitioning the network into smaller networks. For example, one large ICS network is partitioned into multiple ICS networks, where the partitioning is based on factors such as management authority, uniform policy and level of trust, functional criticality, and amount of communications traffic that crosses the domain boundary. Network segmentation and segregation is one of the most effective architectural concepts that an organization can implement to protect its ICS. Segmentation establishes security domains, or enclaves, that are typically defined as being managed by the same authority, enforcing the same policy, and having a uniform level of trust. Segmentation can minimize the method and level of access to sensitive information, ICS communication and equipment configuration, and can make it significantly more difficult for a malicious cyber adversary and can contain the effects of non-malicious errors and accidents. A practical consideration in defining a security domain is the amount of communications traffic that crosses the domain boundary, because domain protection typically involves examining boundary traffic and determining whether it is permitted.

The aim of network segmentation and segregation is to minimize access to sensitive information for those systems and people who don't need it, while ensuring that the organization can continue to operate effectively. This can be achieved using a number of techniques and technologies depending on the network's architecture and configuration.

[11] ICS-CERT recommended practices may be found at http://ics-cert.us-cert.gov/Recommended-Practices.

Traditionally, network segmentation and segregation is implemented at the gateway between domains. ICS environments often have multiple well-defined domains, such as operational LANs, control LANs, and operational DMZs, as well as gateways to non-ICS and less trustworthy domains such as the Internet and the corporate LANs. When insider attacks, social engineering, mobile devices, and other vulnerabilities and predisposing conditions discussed in Appendix C— are considered, protecting domain gateways is prudent and worth considering.

Network segregation involves developing and enforcing a ruleset controlling which communications are permitted through the boundary. Rules typically are based on source and destination identity and the type or content of the data being transferred.

When implementing network segmentation and segregation correctly you are minimizing the method and level of access to sensitive information. This can be achieved using a variety of technologies and methods. Depending on the architecture and configuration of your network, some of the common technologies and methods used include:

- Logical network separation enforced by encryption or network device-enforced partitioning.
  - Virtual Local Area Networks (VLANS).
  - Encrypted Virtual Private Networks (VPNs) use cryptographic mechanisms to separate traffic combined on one network.
  - Unidirectional gateways restrict communications between connections to a single direction, therefore, segmenting the network.

- Physical network separation to completely prevent any interconnectivity of traffic between domains.

- Network traffic filtering which can utilize a variety of technologies at various network layers to enforce security requirements and domains.
  - Network layer filtering that restricts which systems are able to communicate with others on the network based on IP and route information.
  - State-based filtering that restricts which systems are able to communicate with others on the network based on their intended function or current state of operation.
  - Port and/or protocol level filtering that restricts the number and type of services that each system can use to communicate with others on the network.
  - Application filtering that commonly filters the content of communications between systems at the application layer. This includes application-level firewalls, proxies, and content-based filter.

Some vendors are making products to filter ICS protocols at the application level which they market as ICS firewalls.

Regardless of the technology chosen to implement network segmentation and segregation, there are four common themes that implement the concept of defense-in-depth by providing for good network segmentation and segregation:

- Apply technologies at more than just the network layer. Each system and network should be segmented and segregated, where possible, from the data link layer up to and including the application layer.
- Use the principles of least privilege and need-to-know. If a system doesn't need to communicate with another system, it should not be allowed to. If a system needs to talk only to another system on a specific port or protocol and nothing else–or it needs to transfer a limited set of labeled or fixed-format data, it should be restricted as such.

- Separate information and infrastructure based on security requirements. This may include using different hardware or platforms based on different threat and risk environments in which each system or network segment operates. The most critical components require more strict isolation from other components. In addition to network separation, the use of virtualization could be employed to accomplish the required isolation.

- Implement whitelisting[12] instead of blacklisting; that is, grant access to the known good, rather than denying access to the known bad. The set of applications that run in ICS is essentially static, making whitelisting more practical. This will also improve an organization's capacity to analyze log files.

5.2 Boundary Protection

Boundary protection devices control the flow of information between interconnected security domains to protect the ICS against malicious cyber adversaries and non-malicious errors and accidents. Transferring information between systems representing different security domains with different security policies introduces risk that such transfers violate one or more domain security policies. Boundary protection devices are key components of specific architectural solutions that enforce specific security policies. Organizations can isolate ICS and business system components performing different missions and/or business functions. Such isolation limits unauthorized information flows among system components and also provides the opportunity to deploy greater levels of protection for selected components. Separating system components with boundary protection mechanisms provides the capability for increased protection of individual components and more effective control of information flows between those components.

Boundary protection controls include gateways, routers, firewalls, guards, network-based malicious code analysis and virtualization systems, intrusion detection systems (networked and host-based), encrypted tunnels, managed interfaces, mail gateways, and unidirectional gateways (e.g., data diodes). Boundary protection devices determine whether data transfer is permitted, often by examining the data or associated metadata.

Network and ICS security architects must decide which domains are to be permitted direct communication, the policies governing permitted communication, the devices to be used to enforce the policy, and the topology for provisioning and implementing these decisions, which are typically based on the trust relationship between domains. Trust involves the degree of control that the organization has over the external domain (e.g., another domain in the same organization, a contracted service provider, the Internet).

Boundary protection devices are arranged in accordance with organizational security architecture. A common architectural construct is the demilitarized zones (DMZ), a host or network segment inserted as a "neutral zone" between security domains. Its purpose is to enforce the ICS domain's information security policy for external information exchange and to provide external domains with restricted access while shielding the ICS domain from outside threats.

Additional architectural considerations and functions that can be performed by boundary protection devices for inter-domain communications include:

[12] A **whitelist** is a list or register of those that are being provided a particular privilege, service, mobility, access or recognition. Only those on the list will be accepted, approved or recognized (i.e., permitted). Whitelisting is the reverse of blacklisting, the practice of identifying those that are denied, unrecognized, or ostracized (i.e., prohibited).

- Denying communications traffic by default and allowing communications traffic by exception (i.e., deny all, permit by exception). A deny-all, permit-by-exception communications traffic policy ensures that only those connections which are approved are allowed. This is known as a white-listing policy.

- Implementing proxy servers that act as an intermediary for external domains' requesting information system resources (e.g., files, connections, or services) from the ICS domain. External requests established through an initial connection to the proxy server are evaluated to manage complexity and to provide additional protection by limiting direct connectivity.

- Preventing the unauthorized exfiltration of information. Techniques include, for example, deep packet inspection firewalls and XML gateways. These devices verify adherence to protocol formats and specification at the application layer and serve to identify vulnerabilities that cannot be detected by devices operating at the network or transport layers. The limited number of formats, especially the prohibition of free form text in email, eases the use of such techniques at ICS boundaries.

- Only allowing communication between authorized and authenticated source and destinations address pairs by one or more of the organization, system, application, and individual.

- Extending the DMZ concept to other separate subnetworks is useful, for example, in isolating ICS to prevent adversaries from discovering the analysis and forensics techniques of organizations.

- Enforcing physical access control to limit authorized access to ICS components.

- Concealing network addresses of ICS components from discovery (e.g., network address not published or entered in domain name systems), requiring prior knowledge for access.

- Disabling control and troubleshooting services and protocols, especially those employing broadcast messaging, which can facilitate network exploration.

- Configuring boundary protection devices to fail in a predetermined state. Preferred failure states for ICS involve balancing multiple factors including safety and security.

- Configuring security domains with separate network addresses (i.e., as disjoint subnets).

- Disabling feedback (e.g., non-verbose mode) to senders when there is a failure in protocol validation format to prevent adversaries from obtaining information.

- Implementing one-way data flow, especially between different security domains.

- Establishing passive monitoring of ICS networks to actively detect anomalous communications and provide alerts.

5.3 Firewalls

Network firewalls are devices or systems that control the flow of network traffic between networks employing differing security postures. In most modern applications, firewalls and firewall environments are discussed in the context of Internet connectivity and the UDP/IP protocol suite. However, firewalls have applicability in network environments that do not include or require Internet connectivity. For example, many corporate networks employ firewalls to restrict connectivity to and from internal networks servicing more sensitive functions, such as the accounting or human resource departments. Firewalls can

further restrict ICS inter-subnetwork communications between functional security subnets and devices. By employing firewalls to control connectivity to these areas, an organization can prevent unauthorized access to the respective systems and resources within the more sensitive areas. There are three general classes of firewalls:

- **Packet Filtering Firewalls.** The most basic type of firewall is called a packet filter. Packet filter firewalls are essentially routing devices that include access control functionality for system addresses and communication sessions. The access control is governed by a set of directives collectively referred to as a rule set. In their most basic form, packet filters operate at layer 3 (network) of the Open Systems Interconnection (OSI), ISO/IEC 7498 model. This type of firewall checks basic information in each packet, such as IP addresses, against a set of criteria before forwarding the packet. Depending on the packet and the criteria, the firewall can drop the packet, forward it, or send a message to the originator. This type of firewall can offer a high level of security, but could result in overhead and delay impacts on network performance.

- **Stateful Inspection Firewalls.** Stateful inspection firewalls are packet filters that incorporate added awareness of the OSI model data at layer 4 (transport). Stateful inspection firewalls filter packets at the network layer, determine whether session packets are legitimate, and evaluate the contents of packets at the transport layer (e.g., TCP, UDP) as well. Stateful inspection keeps track of active sessions and uses that information to determine if packets should be forwarded or blocked. It offers a high level of security and good performance, but it may be more expensive and complex to administer. Additional rule sets for ICS applications may be required.

- **Application-Proxy Gateway Firewalls.** This class of firewalls examines packets at the application layer and filters traffic based on specific application rules, such as specified applications (e.g., browsers) or protocols (e.g., FTP). Firewalls of this type can be very effective in preventing attacks on the remote access and configuration services provided by ICS components. They offer a high level of security, but could have overhead and delay impacts on network performance, which can be unacceptable in an ICS environment. NIST SP 800-41 Revision 1, *Guidelines on Firewalls and Firewall Policy* [85], provides general guidance for the selection of firewalls and the firewall policies.

In an ICS environment, firewalls are most often deployed between the ICS network and the corporate network [34]. Properly configured, they can greatly restrict undesired access to and from control system host computers and controllers, thereby improving security. They can also potentially improve a control network's responsiveness by removing non-essential traffic from the network. When properly designed, configured, and maintained, dedicated hardware firewalls can contribute significantly to increasing the security of today's ICS environments.

Firewalls provide several tools to enforce a security policy that cannot be accomplished locally on the current set of process control devices available in the market, including the ability to:

- Block all communications with the exception of specifically enabled communications between devices on the unprotected LAN and protected ICS networks. Blocking can be based on, for example, source and destination IP address pairs, services, ports, state of the connection, and specified applications or protocols supported by the firewall. Blocking can occur on both inbound and outbound packets, which is helpful in limiting high-risk communications such as email.

- Enforce secure authentication of all users seeking to gain access to the ICS network. There is flexibility to employ varying protection levels of authentication methods including simple passwords, complex passwords, multi-factor authentication technologies, tokens, biometrics and smart cards. Select the particular method based upon the vulnerability of the ICS network to be protected, rather than using the method that is available at the device level.

- Enforce destination authorization. Users can be restricted and allowed to reach only the nodes on the control network necessary for their job function. This reduces the potential of users intentionally or accidentally gaining access to and control of devices for which they are not authorized, but adds to the complexity for on-the-job-training or cross-training employees.
- Record information flow for traffic monitoring, analysis, and intrusion detection.
- Permit the ICS to implement operational policies appropriate to the ICS but that might not be appropriate in an IT network, such as prohibition of less secure communications like email, and permitted use of easy-to-remember usernames and group passwords.
- Be designed with documented and minimal (single if possible) connections that permit the ICS network to be severed from the corporate network, should that decision be made, in times of serious cyber incidents.

Other possible deployments include using either host-based firewalls or small standalone hardware firewalls in front of, or running on, individual control devices. Using firewalls on an individual device basis can create significant management overhead, especially in change management of firewall configurations, however this practice will also simplify individual configuration rulesets.

There are several issues that must be addressed when deploying firewalls in ICS environments, particularly the following:

- The possible addition of delay to control system communications.
- The lack of experience in the design of rule sets suitable for industrial applications. Firewalls used to protect control systems should be configured so they do not permit either incoming or outgoing traffic by default. The default configuration should be modified only when it is necessary to permit connections to or from trusted systems to perform authorized ICS functions.

Firewalls require ongoing support, maintenance, and backup. Rule sets need to be reviewed to make sure that they are providing adequate protection in light of ever-changing security threats. System capabilities (e.g., storage space for firewall logs) should be monitored to make sure that the firewall is performing its data collection tasks and can be depended upon in the event of a security violation. Real-time monitoring of firewalls and other security sensors is required to rapidly detect and initiate response to cyber incidents.

5.4 Logically Separated Control Network

The ICS network should, at a minimum, be logically separated from the corporate network on physically separate network devices. Based on the ICS network configuration, additional separation needs to be considered for Safety Instrumented Systems and Security Systems (e.g., physical monitoring and access controls, doors, gates, cameras, VoIP, access card readers) that are often either part of the ICS network or utilize the same communications infrastructure for remote sites. When enterprise connectivity is required:

- There should be documented and minimal (single if possible) access points between the ICS network and the corporate network. Redundant (i.e., backup) access points, if present, must be documented.
- A stateful firewall between the ICS network and corporate network should be configured to deny all traffic except that which is explicitly authorized.
- The firewall rules should at a minimum provide source and destination filtering (i.e., filter on media access control [MAC] address), in addition to TCP and User Datagram Protocol (UDP) port filtering and Internet Control Message Protocol (ICMP) type and code filtering.

An acceptable approach to enabling communication between an ICS network and a corporate network is to implement an intermediate DMZ network. The DMZ should be connected to the firewall such that specific (restricted) communication may occur between only the corporate network and the DMZ, and the ICS network and the DMZ. The corporate network and the ICS network should not communicate directly with each other. This approach is described in Sections 5.5.4 and 5.5.5. Additional security may be obtained by implementing a Virtual Private Network (VPN) between the ICS and external networks.

5.5 Network Segregation

ICS networks and corporate networks can be segregated to enhance cybersecurity using different architectures. This section describes several possible architectures and explains the advantages and disadvantages of each. Please note that the intent of the diagrams in Section 5.5 is to show the placement of firewalls to segregate the network. Not all devices that would be typically found on the control network or corporate network are shown. Section 5.6 provides guidance on a recommended defense-in-depth architecture.

5.5.1 Dual-Homed Computer/Dual Network Interface Cards (NIC)

Dual-homed computers can pass network traffic from one network to another. A computer without proper security controls could pose additional threats. To prevent this, no systems other than firewalls should be configured as dual-homed to span both the control and corporate networks. All connections between the control network and the corporate network should be through a firewall. This configuration provides no security improvement and should not be used to bridge networks (e.g., ICS and corporate networks).

5.5.2 Firewall between Corporate Network and Control Network

By introducing a simple two-port firewall between the corporate and control networks, as shown in Figure 5-1, a significant security improvement can be achieved. Properly configured, a firewall significantly reduces the chance of a successful external attack on the control network.

Unfortunately, two issues still remain with this design. First, if the data historian resides on the corporate network, the firewall must allow the data historian to communicate with the control devices on the control network. A packet originating from a malicious or incorrectly configured host on the corporate network (appearing to be the data historian) would be forwarded to individual PLCs/DCS.

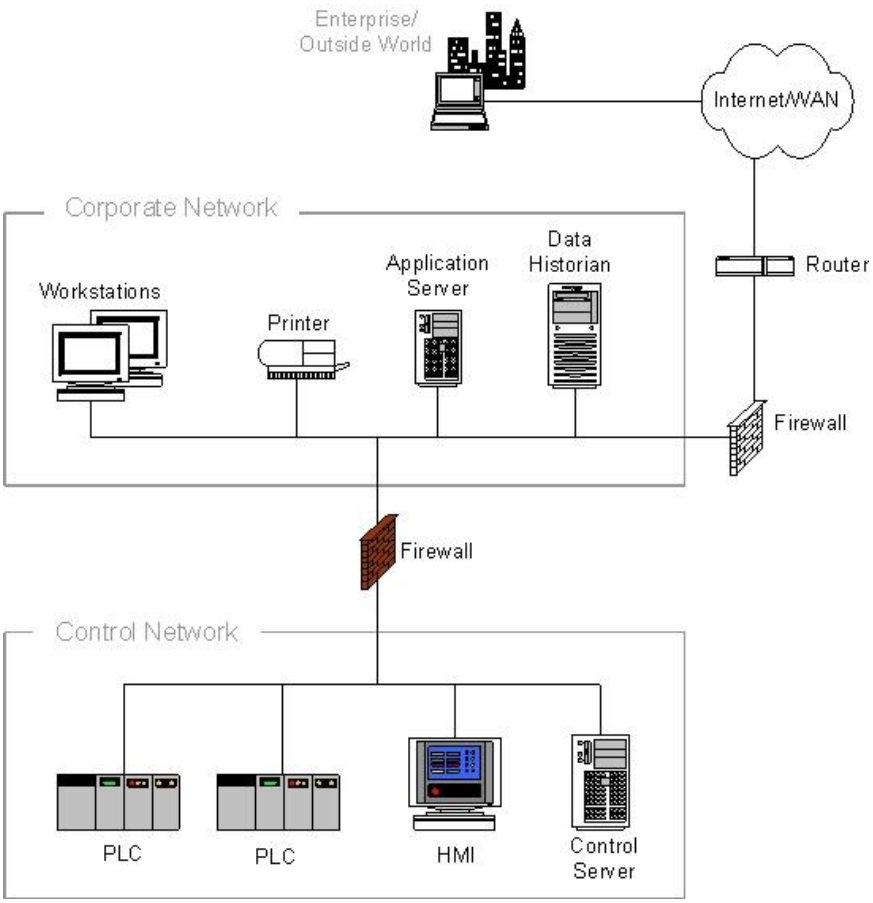

Figure 5-1. Firewall between Corporate Network and Control Network

If the data historian resides on the control network, a firewall rule must exist that allows all hosts from the enterprise to communicate with the historian. Typically, this communication occurs at the application layer as Structured Query Language (SQL) or Hypertext Transfer Protocol (HTTP) requests. Flaws in the historian's application layer code could result in a compromised historian. Once the historian is compromised, the remaining nodes on the control network are vulnerable to a worm propagating or an interactive attack.

Another issue with having a simple firewall between the networks is that spoofed packets can be constructed that can affect the control network, potentially permitting covert data to be tunneled in allowed protocols. For example, if HTTP packets are allowed through the firewall, then Trojan horse software accidentally introduced on an HMI or control network laptop could be controlled by a remote entity and send data (such as captured passwords) to that entity, disguised as legitimate traffic.

In summary, while this architecture is a significant improvement over a non-segregated network, it requires the use of firewall rules that allow direct communications between the corporate network and control network devices. This can result in possible security breaches if not very carefully designed and monitored [35].

5.5.3 Firewall and Router between Corporate Network and Control Network

A slightly more sophisticated design, shown in Figure 5-2, is the use of a router/firewall combination. The router sits in front of the firewall and offers basic packet filtering services, while the firewall handles the more complex issues using either stateful inspection or proxy techniques. This type of design is very popular in Internet-facing firewalls because it allows the faster router to handle the bulk of the incoming packets, especially in the case of DoS attacks, and reduces the load on the firewall. It also offers improved defense-in-depth because there are two different devices an adversary must bypass [35].

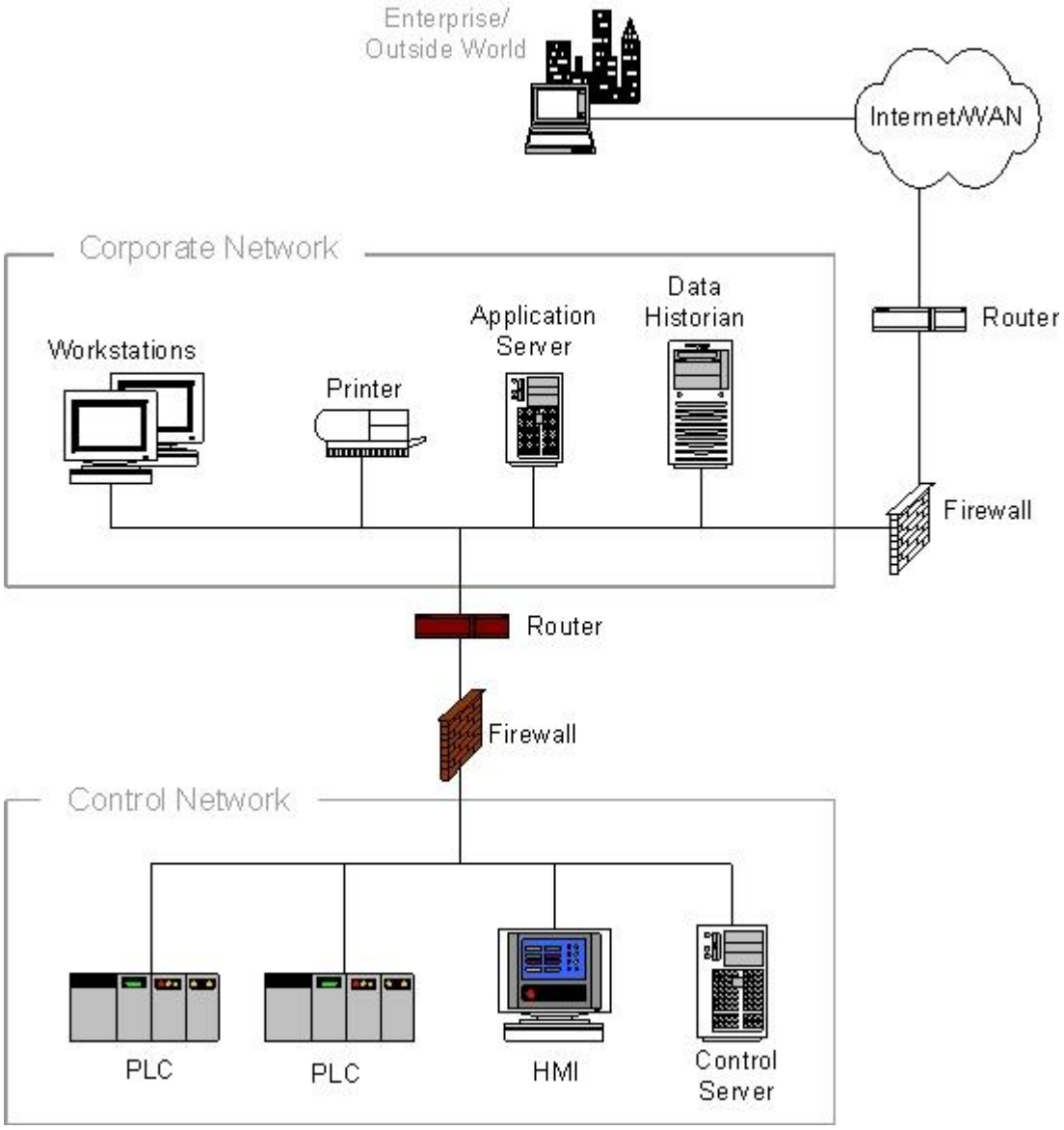

Figure 5-2. Firewall and Router between Corporate Network and Control Network

5.5.4 Firewall with DMZ between Corporate Network and Control Network

A significant improvement is the use of firewalls with the ability to establish a DMZ between the corporate and control networks. Each DMZ holds one or more critical components, such as the data historian, the wireless access point, or remote and third party access systems. In effect, the use of a DMZ-capable firewall allows the creation of an intermediate network.

Creating a DMZ requires that the firewall offer three or more interfaces, rather than the typical public and private interfaces. One of the interfaces is connected to the corporate network, the second to the control network, and the remaining interfaces to the shared or insecure devices such as the data historian server or wireless access points on the DMZ network. Implementing continuous ingress and egress traffic monitoring on the DMZ is recommended. Additionally, firewall rulesets that only permit connections between the control network and DMZ that are initiated by control network devices are recommended. Figure 5-3 provides an example of this architecture.

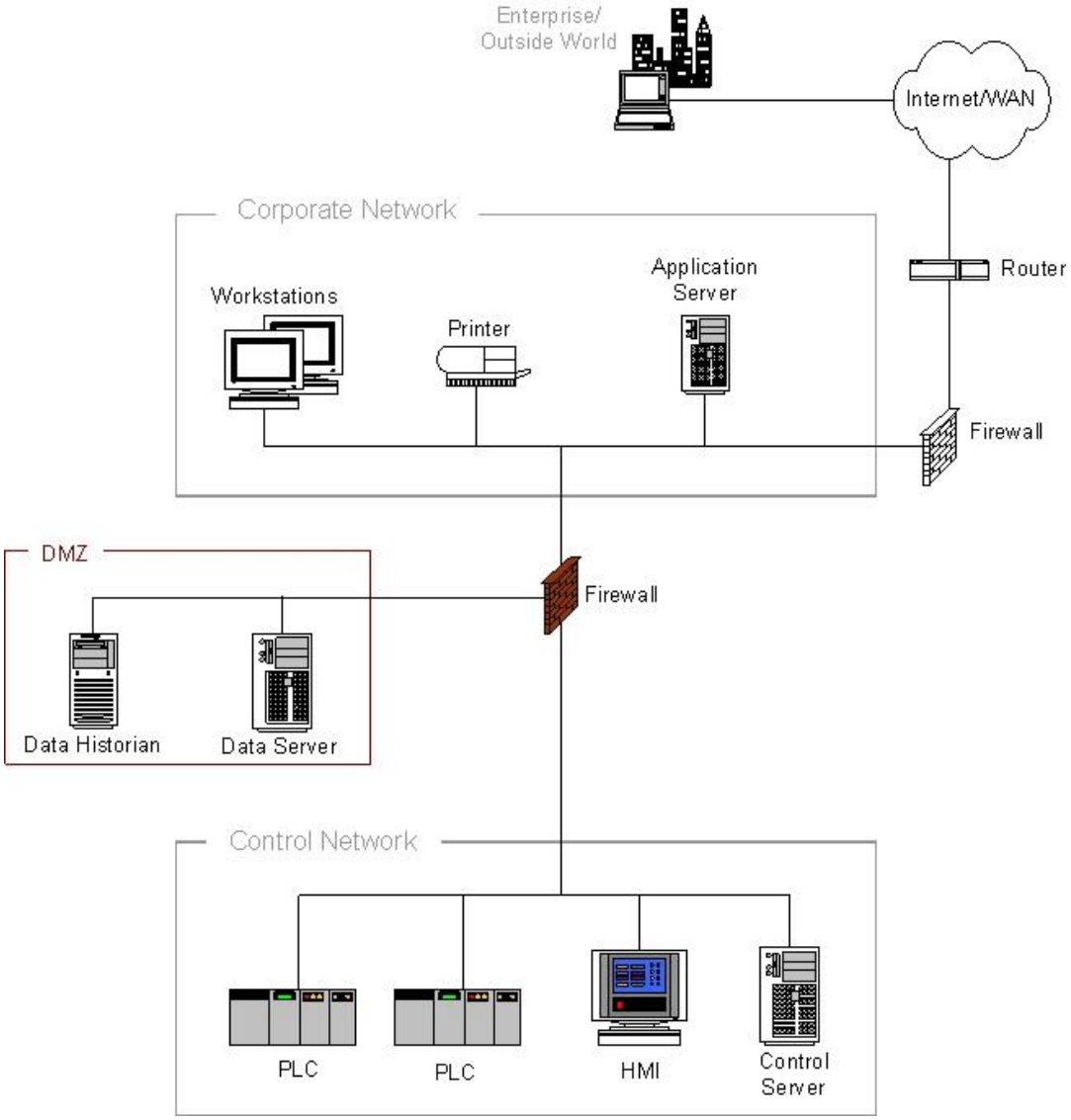

Figure 5-3. Firewall with DMZ between Corporate Network and Control Network

By placing corporate-accessible components in the DMZ, no direct communication paths are required from the corporate network to the control network; each path effectively ends in the DMZ. Most firewalls can allow for multiple DMZs, and can specify what type of traffic may be forwarded between zones. As Figure 5-3 shows, the firewall can block arbitrary packets from the corporate network from entering the control network, and can also regulate traffic from the other network zones including the control network. With well-planned rule sets, a clear separation can be maintained between the control network and other networks, with little or no traffic passing directly between the corporate and control networks.

If a patch management server, an antivirus server, or other security server is to be used for the control network, it should be located directly on the DMZ. Both functions could reside on a single server. Having patch management and antivirus management dedicated to the control network allows for controlled and secure updates that can be tailored for the unique needs of the ICS environment. It may also be helpful if the antivirus product chosen for ICS protection is not the same as the antivirus product used for the corporate network. For example, if a malware incident occurs and one antivirus product cannot detect or stop the malware, it is somewhat likely that another product may have that capability.

The primary security risk in this type of architecture is that if a computer in the DMZ is compromised, then it can be used to launch an attack against the control network via application traffic permitted from the DMZ to the control network. This risk can be greatly reduced if a concerted effort is made to harden and actively patch the servers in the DMZ and if the firewall ruleset permits only connections between the control network and DMZ that are initiated by control network devices. Other concerns with this architecture are the added complexity and the potential increased cost of firewalls with several ports. For more critical systems, however, the improved security should more than offset these disadvantages [35].

5.5.5 Paired Firewalls between Corporate Network and Control Network

A variation on the firewall with a DMZ solution is to use a pair of firewalls positioned between the corporate and ICS networks, as shown in Figure 5-4. Common servers such as the data historian are situated between the firewalls in a DMZ-like network zone sometimes referred to as a Manufacturing Execution System (MES) layer. As in the architectures described previously, the first firewall blocks arbitrary packets from proceeding to the control network or the shared historians. The second firewall can prevent unwanted traffic from a compromised server from entering the control network, and prevent control network traffic from impacting the shared servers.

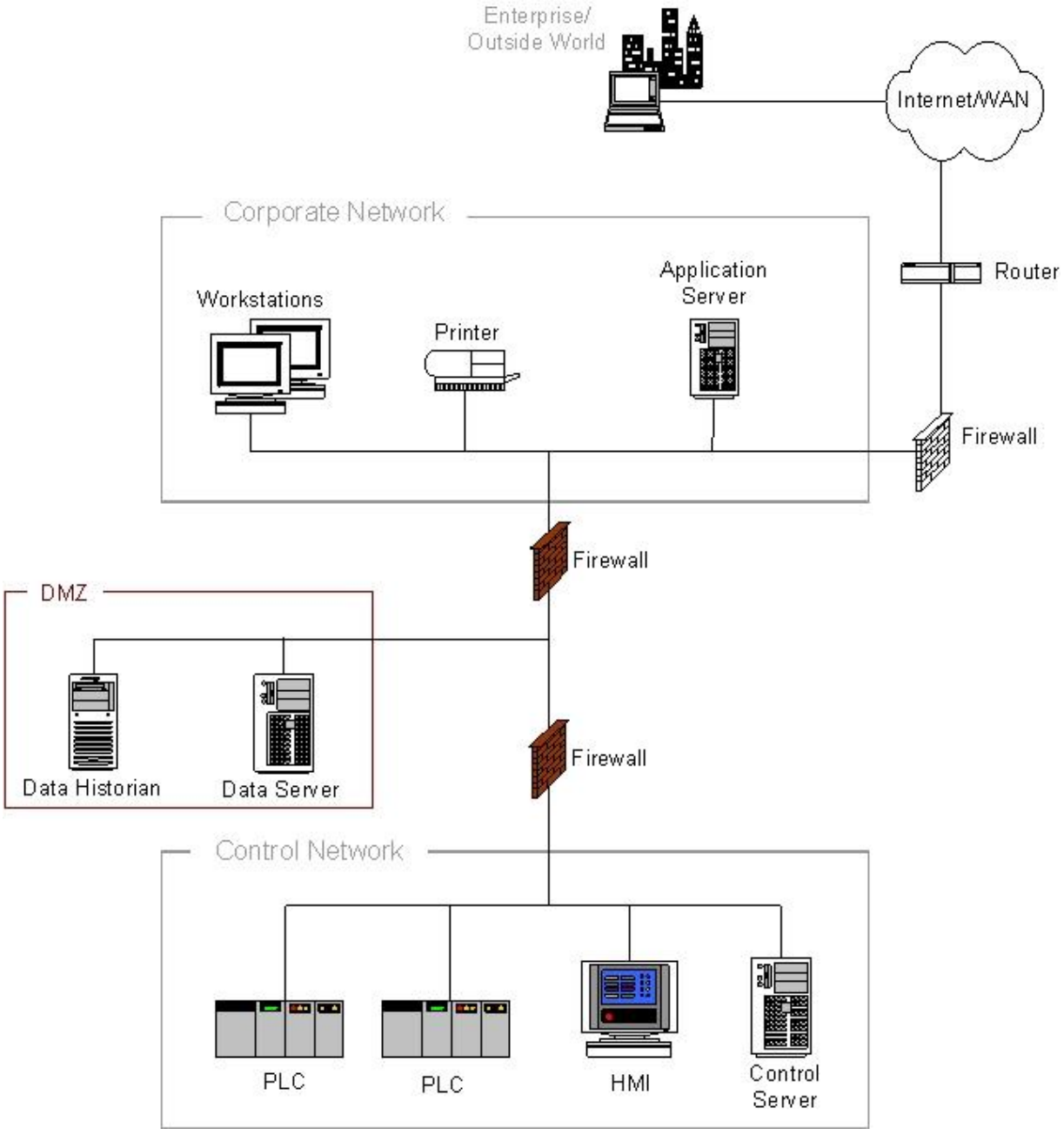

Figure 5-4. Paired Firewalls between Corporate Network and Control Network

If firewalls from two different manufacturers are used, then this solution may offer an advantage. It also allows the control group and the IT group to have clearly separated device responsibility because each can manage a firewall on its own, if the decision is made within the organization to do so. The primary disadvantage with two-firewall architectures is the increased cost and management complexity. For environments with stringent security requirements or the need for clear management separation, this architecture has some strong advantages.

5.5.6 Network Segregation Summary

In summary, dual-homed computers generally not provide suitable isolation between control networks and corporate networks. The two-zone solutions (no DMZ) are not recommended because they provide only weak protection. If used, they should only be deployed with extreme care. The most secure, manageable, and scalable control network and corporate network segregation architectures are typically based on a system with at least three zones, incorporating one or more DMZs.

5.6 Recommended Defense-in-Depth Architecture

A single security product, technology or solution cannot adequately protect an ICS by itself. A multiple layer strategy involving two (or more) different overlapping security mechanisms, a technique also known as defense-in-depth, is desired so that the impact of a failure in any one mechanism is minimized. A defense-in-depth architecture strategy includes the use of firewalls, the creation of demilitarized zones, intrusion detection capabilities along with effective security policies, training programs, incident response mechanisms and physical security. In addition, an effective defense-in-depth strategy requires a thorough understanding of possible attack vectors on an ICS. These include:

- Backdoors and holes in network perimeter.
- Vulnerabilities in common protocols.
- Attacks on field devices.
- Database attacks.
- Communications hijacking and ‘man-in-the-middle’ attacks.
- Spoofing attacks.
- Attacks on privileged and/or shared accounts.

Figure 5-5 shows an ICS defense-in-depth architecture strategy that has been developed by the DHS Control Systems Security Program (CSSP) NCCIC/ICS-CERT Recommended Practices committee[13] as described in the *Control Systems Cyber Security: Defense in Depth Strategies* [36] document. Additional supporting documents that cover specific issues and associated mitigations are also included on the site.

The *Control Systems Cyber Security: Defense in Depth Strategies* document provides guidance and direction for developing defense-in-depth architecture strategies for organizations that use control system networks while maintaining a multi-tiered information architecture that requires:

- Maintenance of various field devices, telemetry collection, and/or industrial-level process systems.
- Access to facilities via remote data link or modem.
- Public facing services for customer or corporate operations.

[13] Information on the CSSP Recommended Practices is located at http://ics-cert.us-cert.gov/Recommended-Practices

This strategy includes firewalls, the use of demilitarized zones and intrusion detection capabilities throughout the ICS architecture. The use of several demilitarized zones in Figure 5-5 provides the added capability to separate functionalities and access privileges and has proved to be very effective in protecting large architectures comprised of networks with different operational mandates. Intrusion detection deployments apply different rule-sets and signatures unique to each domain being monitored.

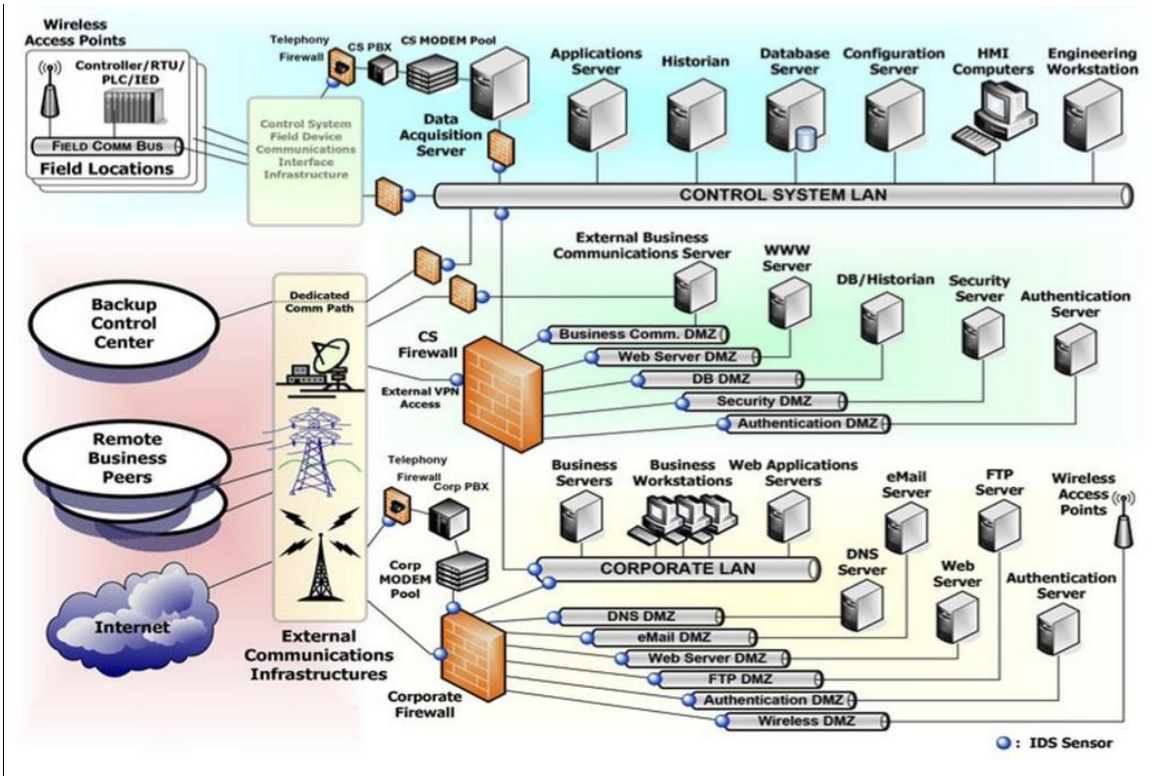

Figure 5-5. CSSP Recommended Defense-In-Depth Architecture

5.7 General Firewall Policies for ICS

Once the defense-in-depth architecture is in place, the work of determining exactly what traffic should be allowed through the firewalls begins. Configuring the firewalls to deny all except for the traffic absolutely required for business needs is every organization's basic premise, but the reality is much more difficult. Exactly what does "absolutely required for business" mean and what are the security impacts of allowing that traffic through? For example, many organizations considered allowing SQL traffic through the firewall as required for business for many data historian servers. Unfortunately, the SQL vulnerability was also the target for the Slammer worm [Table C-8. Example Adversarial Incidents]. Many important protocols used in the industrial world, such as HTTP, FTP, OPC/DCOM, EtherNet/IP, and Modbus/TCP, have significant security vulnerabilities.

The remaining material in this section summarizes some of the key points from the Centre for the Protection of National Infrastructure's (CPNI) *Firewall Deployment for SCADA and Process Control Networks: Good Practice Guide* [35].

When installing a single two-port firewall without a DMZ for shared servers (i.e., the architecture described in Section 5.5.2), particular care needs to be taken with the rule design. At a minimum, all rules

should be stateful rules that are both IP address and port (application) specific. The address portion of the rules should restrict incoming traffic to a very small set of shared devices (e.g., the data historian) on the control network from a controlled set of addresses on the corporate network. Allowing any IP addresses on the corporate network to access servers inside the control network is not recommended. In addition, the allowed ports should be carefully restricted to relatively secure protocols such as Hypertext Transfer Protocol Secure (HTTPS). Allowing HTTP, FTP, or other unsecured protocols to cross the firewall is a security risk due to the potential for traffic sniffing and modification. Rules should be added to deny hosts outside the control network from initiating connections with hosts on the control network. Rules should only allow devices internal to the control network the ability to establish connections outside the control network.

On the other hand, if the DMZ architecture is being used, then it is possible to configure the system so that no traffic will go directly between the corporate network and the control network. With a few special exceptions (noted below), all traffic from either side can terminate at the servers in the DMZ. This allows more flexibility in the protocols allowed through the firewall. For example, Modbus/TCP might be used to communicate from the PLCs to the data historian, while HTTP might be used for communication between the historian and enterprise clients. Both protocols are inherently insecure, yet in this case they can be used safely because neither actually crosses between the two networks. An extension to this concept is the idea of using "disjoint" protocols in all control network to corporate network communications. That is, if a protocol is allowed between the control network and DMZ, then it is explicitly **not** allowed between the DMZ and corporate network. This design greatly reduces the chance of a worm such as Slammer actually making its way into the control network, because the worm would have to use two different exploits over two different protocols.

One area of considerable variation in practice is the control of outbound traffic from the control network, which could represent a significant risk if unmanaged. One example is Trojan horse software that uses HTTP tunneling to exploit poorly defined outbound rules. Thus, it is important that outbound rules be as stringent as inbound rules.

Example outbound rules include:

- Outbound traffic through the control network firewall should be limited to essential communications only and should be limited to authorized traffic originating from DMZ servers.
- All outbound traffic from the control network to the corporate network should be source and destination-restricted by service and port.

In addition to these rules, the firewall should be configured with outbound filtering to stop forged IP packets from leaving the control network or the DMZ. In practice this is achieved by checking the source IP addresses of outgoing packets against the firewall's respective network interface address. The intent is to prevent the control network from being the source of spoofed (i.e., forged) communications, which are often used in DoS attacks. Thus, the firewalls should be configured to forward IP packets only if those packets have a correct source IP address for the control network or DMZ networks. Finally, Internet access by devices on the control network should be strongly discouraged.

In summary, the following should be considered as recommended practice for general firewall rule sets:

- The base rule set should be deny all, permit none.
- Ports and services between the control network environment and the corporate network should be enabled and permissions granted on a specific case-by-case basis. There should be a documented business justification with risk analysis and a responsible person for each permitted incoming or outgoing data flow.
- All "permit" rules should be both IP address and TCP/UDP port specific, and stateful if appropriate.
- All rules should restrict traffic to a specific IP address or range of addresses.
- Traffic should be prevented from transiting directly from the control network to the corporate network. All traffic should terminate in the DMZ.
- Any protocol allowed between the control network and DMZ should explicitly NOT be allowed between the DMZ and corporate networks (and vice-versa).
- All outbound traffic from the control network to the corporate network should be source and destination-restricted by service and port.
- Outbound packets from the control network or DMZ should be allowed only if those packets have a correct source IP address that is assigned to the control network or DMZ devices.
- Control network devices should not be allowed to access the Internet.
- Control networks should not be directly connected to the Internet, even if protected via a firewall.
- All firewall management traffic should be carried on either a separate, secured management network (e.g., out of band) or over an encrypted network with multi-factor authentication. Traffic should also be restricted by IP address to specific management stations.
- All firewall policies should be tested periodically.
- All firewalls should be backed up immediately prior to commissioning.

These should be considered only as guidelines. A careful assessment of each control environment is required before implementing any firewall rule sets.

5.8 Recommended Firewall Rules for Specific Services

Beside the general rules described above, it is difficult to outline all-purpose rules for specific protocols. The needs and recommended practices vary significantly between industries for any given protocol and should be analyzed on an organization-by-organization basis. The Industrial Automation Open Networking Association (IAONA) offers a template for conducting such an analysis [37], assessing each of the protocols commonly found in industrial environments in terms of function, security risk, worst case impact, and suggested measures. Some of the key points from the IAONA document are summarized in this section. The reader is advised to consult this document directly when developing rule sets.

5.8.1 Domain Name System (DNS)

Domain Name System (DNS) is primarily used to translate between domain names and IP addresses. For example, a DNS could map a domain name such as *control.com* to an IP address such as *192.168.1.1*. Most Internet services rely heavily on DNS, but its use on the control network is relatively rare at this time. In most cases there is little reason to allow DNS requests out of the control network to the corporate network and no reason to allow DNS requests into the control network. DNS requests from the control network to DMZ should be addressed on a case-by-case basis. Local DNS or the use of host files is recommended.

5.8.2 Hypertext Transfer Protocol (HTTP)

HTTP is the protocol underlying Web browsing services on the Internet. Like DNS, it is critical to most Internet services. It is seeing increasing use on the plant floor as well as an all-purpose query tool. Unfortunately, it has little inherent security, and many HTTP applications have vulnerabilities that can be exploited. HTTP can be a transport mechanism for many manually performed attacks and automated worms.

In general, HTTP should not be allowed to cross from the public/corporate to the control network. If web-based technologies are absolutely required, the following best practices should be applied:

- Control access to web-based services on the physical or network layer using white-listing;
- Apply access control to both source and destination;
- Implement authorization to access the service on the application layer (instead of physical or network-layer checks);
- Implement service using only the necessary technologies (e.g., scripts are used only if they are required);
- Check service according to known application security practices;
- Log all attempts of service usage ; and
- Use HTTPS rather than HTTP, and only for specific authorized devices.

5.8.3 FTP and Trivial File Transfer Protocol (TFTP)

FTP and Trivial File Transfer Protocol (TFTP) are used for transferring files between devices. They are implemented on almost every platform including many SCADA systems, DCS, PLCs, and RTUs, because they are very well known and use minimum processing power. Unfortunately, neither protocol was created with security in mind; for FTP, the login password is not encrypted, and for TFTP, no login is required at all. Furthermore, some FTP implementations have a history of buffer overflow vulnerabilities. As a result, all TFTP communications should be blocked, while FTP communications should be allowed for outbound sessions only or if secured with additional token-based multi-factor authentication and an encrypted tunnel. More secure protocols, such as Secure FTP (SFTP) or Secure Copy (SCP), should be employed whenever possible.

5.8.4 Telnet

The telnet protocol defines an interactive, text-based communications session between a client and a host. It is used mainly for remote login and simple control services to systems with limited resources or to systems with limited needs for security. It is a severe security risk because all telnet traffic, including passwords, is unencrypted, and it can allow a remote individual considerable control over a device. It is recommended to use the Secure Shell (SSH) protocol [5.8.6] for remote administration. Inbound telnet

sessions from the corporate to the control network should be prohibited unless secured with token-based multi-factor authentication and an encrypted tunnel. Outbound telnet sessions should be allowed only over encrypted tunnels (e.g., VPN) to specific authorized devices.

5.8.5 Dynamic Host Configuration Protocol (DHCP)

DHCP is used on IP networks for dynamically distributing network configuration parameters, such as IP addresses for interfaces and services. The base DHCP includes no mechanism for authenticating servers and clients. Rogue DHCP servers can provide incorrect information to clients. Unauthorized clients can gain access to server and cause exhaustion of available resources (e.g., IP addresses). To prevent this, it is recommended to use static configuration instead of dynamic address allocation, which should be the typical configuration for ICS devices. If dynamic allocation is necessary, it is recommended to enable DHCP snooping to defend against rogue DHCP servers, Address Resolution Protocol (ARP) and IP spoofing. The DHCP servers should be placed in the same network segment as configured equipment (e.g., on the router). DHCP relaying is not recommended.

5.8.6 Secure Shell (SSH)

SSH allows remote access to a device. It provides secure authentication and authorization based on cryptography. If remote access is required to the control network, SSH is recommended as the alternative to telnet, rlogin, rsh, rcp and other insecure remote access tools.

5.8.7 Simple Object Access Protocol (SOAP)

SOAP is an XML-based format syntax to exchange messages. Traffic flows related to SOAP-based services should be controlled at the firewall between corporate and ICS network segments. If these services are necessary, deep-packet inspection and/or application layer firewalls should be used to restrict the contents of messages.

5.8.8 Simple Mail Transfer Protocol (SMTP)

SMTP is the primary email transfer protocol on the Internet. Email messages often contain malware, so inbound email should not be allowed to any control network device. Outbound SMTP mail messages from the control network to the corporate network are acceptable to send alert messages.

5.8.9 Simple Network Management Protocol (SNMP)

SNMP is used to provide network management services between a central management console and network devices such as routers, printers, and PLCs. Although SNMP is an extremely useful service for maintaining a network, it is very weak in security. Versions 1 and 2 of SNMP use unencrypted passwords to both read and configure devices (including devices such as PLCs), and in many cases the passwords are well known and cannot be changed. Version 3 is considerably more secure but is still limited in use. SNMP V1 & V2 commands both to and from the control network should be prohibited unless they are over a separate, secured management network, whereas SNMP V3 commands may be able to be sent to the ICS using the security features inherent to V3.

5.8.10 Distributed Component Object Model (DCOM)

DCOM is the underlying protocol for OLE for Process Control (OPC). It utilizes Microsoft's Remote Procedure Call (RPC) service which, when not patched, has many vulnerabilities. These vulnerabilities were the basis for the Blaster worm[14] exploits. In addition, OPC, which utilizes DCOM, dynamically opens a wide range of ports (1024 to 65535) that can be extremely difficult to filter at the firewall. This protocol should only be allowed between control network and DMZ networks and explicitly blocked between the DMZ and corporate network. Also, users are advised to restrict the port ranges used by making registry modifications on devices using DCOM.

5.8.11 SCADA and Industrial Protocols

SCADA and industrial protocols, such as Modbus/TCP, EtherNet/IP, IEC 61850, ICCP and DNP3[15], are critical for communications to most control devices. Unfortunately, many of these protocols were designed without security built in and do not typically require any authentication to remotely execute commands on a control device. These protocols should only be allowed within the control network and not allowed to cross into the corporate network.

5.9 Network Address Translation (NAT)

Network address translation (NAT) is a service where IP addresses used on one side of a network device can be mapped to a different set on the other side on an as-needed basis. It was originally designed for IP address reduction purposes so that an organization with a large number of devices that occasionally needed Internet access could get by with a smaller set of assigned Internet addresses.

To do this, most NAT implementations rely on the premise that not every internal device is actively communicating with external hosts at a given moment. The firewall is configured to have a limited number of outwardly visible IP addresses. When an internal host seeks to communicate with an external host, the firewall remaps the internal IP address and port to one of the currently unused, more limited, public IP addresses, effectively concentrating outgoing traffic into fewer IP addresses. The firewall must track the state of each connection and how each private internal IP address and source port was remapped onto an outwardly visible IP address/port pair. When returning traffic reaches the firewall, the mapping is reversed and the packets forwarded to the proper internal host.

For example, a control network device may need to establish a connection with an external, non-control network host (for instance, to send a critical alert email). NAT allows the internal IP address of the initiating control network host to be replaced by the firewall; subsequent return traffic packets are remapped back to the internal IP address and sent to the appropriate control network device. More specifically, if the control network is assigned the private subnet 192.168.1.xxx and the Internet network expects the device to use the corporate assigned addresses in the range 192.6.yyy.zzz, then a NAT firewall will substitute (and track) a 192.6.yyy.zzz source address into every outbound IP packet generated by a control network device.

Producer-consumer protocols, such as EtherNet/IP and Foundation Fieldbus, are particularly troublesome because NAT does not support the multicast-based traffic that these protocols need to offer their full services.

[14] http://en.wikipedia.org/wiki/Blaster_%28computer_worm%29

[15] IEEE 1815-2012, *IEEE Standard for Electric Power Systems Communications—Distributed Network Protocol (DNP3),*) incorporates DNP3 Secure Authentication version 5 (DNP3-SAv5) which provides strong application layer authentication with remote security credential management. See https://standards.ieee.org/findstds/standard/1815-2012.html.

In general, while NAT offers some distinct advantages, its impact on the actual industrial protocols and configuration should be assessed carefully before it is deployed. Furthermore, certain protocols are specifically broken by NAT because of the lack of direct addressing. For example, OPC requires special third-party tunneling software to work with NAT.

5.10 Specific ICS Firewall Issues

In addition to the issues with firewalls and ICS already discussed, there are some additional problems that need to be examined in more detail. The rest of this section discusses three specific areas of concern: the placement of data historians, remote access for ICS support, and multicast traffic.

5.10.1 Data Historians

The existence of shared control network/corporate network servers such as data historians and asset management servers can have a significant impact on firewall design and configuration. In three-zone systems the placement of these servers in a DMZ is relatively straightforward, but in two-zone designs the issues become complex. Placing the historian on the corporate side of the firewall means that a number of insecure protocols, such as Modbus/TCP or DCOM, must be allowed through the firewall and that every control device reporting to the historian is exposed to the corporate side of the network. On the other hand, putting the historian on the control network side means other equally questionable protocols, such as HTTP or SQL, must be allowed through the firewall, and there is now a server accessible to nearly everyone in the organization sitting on the control network.

In general, the best solution is to avoid two-zone systems (no DMZ) and use a three-zone design, placing the data collector in the control network and the historian component in the DMZ.

5.10.2 Remote Support Access

Another issue for ICS firewall design is user and/or vendor remote access into the control network. Any users accessing the control network from remote networks should be required to authenticate using an appropriately strong mechanism such as token-based authentication. While it is possible for the controls group to set up their own remote access system with multi-factor authentication on the DMZ, in most organizations it is typically more efficient to use existing systems set up by the IT department. In this case a connection through the firewall from the IT remote access server is needed.

Remote support personnel connecting over the Internet or via dialup modems should use an encrypted protocol, such as running a corporate VPN connection client, application server, or secure HTTP access, and authenticate using a strong mechanism, such as a token based multi-factor authentication scheme, in order to connect to the general corporate network. Once connected, they should be required to authenticate a second time at the control network firewall using a strong mechanism, such as a token based multi-factor authentication scheme, to gain access to the control network. Proxy servers can also provide additional capabilities for securing remote support access.

5.10.3 Multicast Traffic

Most industrial producer-consumer (or publisher-subscriber) protocols operating over Ethernet, such as EtherNet/IP and Foundation Fieldbus HSE, are IP multicast-based. The first advantage of IP multicasting is network efficiency; by not repeating the data transmission to the multiple destinations, a significant reduction in network load can occur. The second advantage is that the sending host need not be concerned with knowing every IP address of every destination host listening for the broadcast information. The third, and perhaps most important for industrial control purposes, is that a single multicast message offers

far better capabilities for time synchronization between multiple control devices than multiple unicast messages.

If the source and destinations of a multicast packet are connected with no intervening routers or firewalls between them, the multicast transmission is relatively seamless. However, if the source and destinations are not on the same LAN, forwarding the multicast messages to a destination becomes more complicated. To solve the problem of multicast message routing, hosts need to join (or leave) a group by informing the multicast router on their network of the relevant group ID through the use of the Internet Group Management Protocol (IGMP). Multicast routers subsequently know of the members of multicast groups on their network and can decide whether or not to forward a received multicast message onto their network. A multicast routing protocol is also required. From a firewall administration perspective, monitoring and filtering IGMP traffic becomes another series of rule sets to manage, adding to the complexity of the firewall.

Another firewall issue related to multicasting is the use of NAT. A firewall performing NAT that receives a multicast packet from an external host has no reverse mapping for which internal group ID should receive the data. If IGMP-aware, it could broadcast it to every group ID it knows about, because one of them will be correct, but this could cause serious issues if an unintended control packet were broadcast to a critical node. The safest action for the firewall to take is to drop the packet. Thus, multicasting is generally considered NAT-unfriendly.

5.11 Unidirectional Gateways

Hardware-enforced unidirectional gateways (e.g., data diodes) are increasingly deployed at the boundary between ICS and IT networks, as well as between Safety Instrumented System networks and control networks. Unidirectional gateways are a combination of hardware and software. The hardware permits data to flow from one network to another, but is physically unable to send any information at all back into the source network. The software replicates databases and emulates protocol servers and devices.

5.12 Single Points of Failure

Single points of failure can exist at any level of the ANSI/ISO stack. An example is PLC control of safety interlocks. Because security is usually being added to the ICS environment, an evaluation should be done to identify potential failure points and a risk assessment done to evaluate each point's exposure. Remediation methods can then be postulated and evaluated and a "risk versus reward" determination made and design and implementation done.

5.13 Redundancy and Fault Tolerance

ICS components or networks that are classified as critical to the organization have high availability requirements. One method of achieving high availability is through the use of redundancy. Additionally, if a component fails, it should fail in a manner that does not generate unnecessary traffic on the ICS, or does not cause another problem elsewhere, such as a cascading event.

The control system should have the ability to execute an appropriate fail-safe process upon the loss of communications with the ICS or the loss of the ICS itself. The organization should define what "loss of communications" means (e.g., 500 milliseconds, 5 seconds, 5 minutes, etc. without communications). The organization should then, based on potential consequences, define the appropriate fail-safe process for their industry.

Backups should be performed using the "backup-in-depth" approach, with layers of backups (e.g., local, facility, disaster) that are time-sequenced such that rapid recent local backups are available for immediate use and secure backups are available to recover from a massive security incident. A mixture of backup/restore approaches and storage methods should be used to ensure that backups are rigorously produced, securely stored, and appropriately accessible for restoration.

5.14 Preventing Man-in-the-Middle Attacks

A man-in-the-middle attack requires knowledge of the protocol being manipulated. The Address Resolution Protocol (ARP) man-in-the-middle attack is a popular method for an adversary to gain access to the network flow of information on a target system. This is performed by attacking the network ARP cache tables of the controller and the workstation machines. Using the compromised computer on the control network, the adversary poisons the ARP tables on each host and informs them that they must route all their traffic through a specific IP and hardware address (i.e., the adversary's machine). By manipulating the ARP tables, the adversary can insert their machine between the two target machines and/or devices.

The ARP man-in-the-middle attack works by initiating gratuitous ARP commands to confuse each host (i.e., ARP poisoning). These ARP commands cause each of the two target hosts to use the MAC address of the adversary as the address for the other target host. When a successful man-in-the-middle attack is performed, the hosts on each side of the attack are unaware that their network data is taking a different route through the adversary's computer.

Once an adversary has successfully inserted their machine into the information stream, they now have full control over the data communications and could carry out several types of attacks. One possible attack method is the replay attack. In its simplest form, captured data from the control/HMI is modified to instantiate activity when received by the device controller. Captured data reflecting normal operations in the ICS could be played back to the operator as required. This would cause the operator's HMI to appear to be normal and the attack will go unobserved. During this replay attack the adversary could continue to send commands to the controller and/or field devices to cause an undesirable event while the operator is unaware of the true state of the system.

Another attack that could be carried out with the man-in-the-middle attack is sending false messages to the operator, and could take the form of a false negative or a false positive. This may cause the operator to take an action, such as flipping a breaker, when it is not required, or it may cause the operator to think everything is fine and not take an action when an action is required. The adversary could send commands to the operator's console indicating a system change, and when the operator follows normal procedures and attempts to correct the problem, the operator's action could cause an undesirable event. There are variations of the modification and replay of control data which could impact the operations of the system.

Protocol manipulation and the man-in-the-middle attack are among the most popular ways to manipulate insecure protocols, such as those found in control systems. However, there are mitigation techniques [38] that can be applied to secure the systems through MAC address locking, static tables, encryption, authentication, and monitoring.

- **MAC Address Locking** - The ARP man-in-the-middle attack requires the adversary to be connected to the local network or have control of a local computer on the network. Port security, also called MAC address locking, is one method to secure the physical connection at the end of each port on a network switch. High-end corporate class network switches usually have some kind of option for MAC address locking. MAC address locking is very effective against a rogue individual looking to physically plug into the internal network. Without port security, any open network jack on the wall

could be used as an avenue onto the corporate network. Port security locks a specific MAC address to a specific port on a managed switch. If the MAC address does not match, the communication link is disabled and the intruder will not be able to achieve their goal. Some of the more advanced switches have an auto resetting option, which will reset the security measure if the original MAC is returned to the port.

Although port security is not attacker proof, it does add a layer of added security to the physical network. It also protects the local network from employees plugging un-patched and out-of-date systems onto the protected network. This reduces the number of target computers a remote adversary can access. These security measures not only protect against attacks from external networks but provide added physical protection as well.

- **Static Tables** – An ICS network that stays relatively static could attempt to implement statically coded ARP tables. Most operating systems have the capability to statically code all of the MAC addresses into the ARP table on each computer. Statically coding the ARP tables on each computer prevents the adversary from changing them by sending ARP reply packets to the victim computer. While this technique is not feasible on a large and/or dynamic corporate network, the limited number of hosts on an ICS network could be effectively protected this way.

- **Encryption** - As a longer-term solution, systems should be designed to include encryption between devices in order to make it very difficult to reverse engineer protocols and forge packets on control system networks. Encrypting the communications between devices would make it nearly impossible to perform this attack. Protocols that provide strong authentication also provide resilience to man-in-the-middle attacks. The impact of encryption on network and operational performance needs to be considered.

- **Authentication -** Protocols with strong authentication provide resilience to man-in-the-middle attacks.

- **Monitoring** - Monitoring for ARP poisoning provides an added layer of defense. There are several programs available (e.g., ARPwatch) that can monitor for changing MAC addresses through the ARP packets.

5.15 Authentication and Authorization

An ICS may contain a large number of systems, each of which must be accessed by a variety of users. Performing the authentication and authorization of these users presents a challenge to the ICS. Managing these user's accounts can be problematic as employees are added, removed, and as their roles change. As the number of systems and users grow, the process of managing these accounts becomes more complicated.

The authentication of a user or system is the process of verifying the claimed identity. Authorization, the process of granting the user access privileges, is determined by applying policy rules to the authenticated identity and other relevant information[16]. Authorization is enforced by some access control mechanism. The authentication process can be used to control access to both systems (e.g. HMIs, field devices, SCADA servers) and networks (e.g., remote substations LANs).

Authentication and authorization can be performed either in a distributed or centralized approach. With distributed authentication and authorization, every system performs these steps on their own. Each system is responsible for storing its own set of user accounts, credentials, and roles and performing the identification and authentication of the user. This approach typically does not require any additional infrastructure. However, this approach is problematic in that it does not scale well as the size of the system increases. For example, if a user leaves the organization, the corresponding user account must be removed from each system individually.

In contrast to the distributed approach, centralized authentication and authorization systems are commonly used to manage larger number of users and accounts. A centralized approach utilizes some central authentication system (e.g., Microsoft Active Directory, Lightweight Directory Access Protocol (LDAP) to store all accounts and manage the authentication and authorization of all individuals and systems. An authentication protocol (e.g., Kerberos, RADIUS, TACACS+) is then used to communicate data between the authentication server and the system performing authentication.

While a centralized approach provides substantially improved scalability, it also presents numerous additional concerns that may impact its use in ICS environments. The following considerations apply:

- Authentication servers create a single system that is responsible for managing all system accounts and must be highly secured.
- The authentications server system requires high availability because its failure may prevent users from authenticating to a system during an emergency. Redundancy may be required.
- Some clients may cache user credentials locally to ensure that users can still be authenticated in the absence of the server. Caching may only be available for users that have recently authenticated. Caching also introduces complications for revocation.
- Networks used to support the authentication protocol must be reliable and secure to ensure authentication attempts are not hindered.

[16] In general, authorization to perform a set of operations is determined by evaluating attributes associated with the subject, object, requested operations, and, in some cases, environment conditions against policy, rules, or relationships that describe the allowable operations for a given set of attributes. For further information see NIST SP 800-162, *Guide to Attribute Based Access Control (ABAC) Definition and Considerations*, at http://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.SP.800-162.pdf.http://nvlpubs.nist.gov/nistpubs/specialpublications/NIST.sp.800-162.pdf

5.15.1 ICS Implementation Considerations

While centralized authentication and authorization servers are commonly used in an IT environment, there are many challenges to integrating them into ICS. While authentication servers and protocols integrate with many commodity IT products (e.g., Microsoft Windows, Linux, Oracle), often ICS may utilize their own application-specific accounts and authentication mechanisms that were not designed to interface with third party servers and protocols. This limits the adoption of such mechanism in an ICS environment. Older network devices and most field devices do not support any mechanisms to integrated with a centralized authentication system.

5.16 Monitoring, Logging, and Auditing

The security architecture of an ICS must also incorporate mechanisms to monitor, log, and audit activities occurring on various systems and networks. Monitoring, logging, and auditing activities are imperative to understanding the current state of the ICS, validating that the system is operating as intended, and that no policy violations or cyber incidents have hindered the operation of the system. Network security monitoring is valuable to characterize the normal state of the ICS, and can provide indications of compromised systems when signature-based technologies fail. Additionally, strong system monitoring, logging, and auditing is necessary to troubleshoot and perform any necessary forensic analysis of the system[17].

5.17 Incident Detection, Response, and System Recovery

Incidents are inevitable and incident detection, response, and system recovery plans are essential. Major characteristics of a good security program are how soon after an incident has occurred that the incident can be detected and how quickly a system can be recovered after an incident has been detected. Incident response in ICS is closely aligned to disaster recovery, specifically to address the stringent uptime requirements of ICS. Incident Responders must be trained for ICS-specific scenarios, as normal methods of recovering IT systems may not apply to ICS.

[17] For further information see NIST SP 800-94, *Guide to Intrusion Detection and Prevention Systems (IDPS)* [55].