IoT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things

David Hanes, CCIE No. 3491

Gonzalo Salgueiro, CCIE No. 4541

Patrick Grossetete

Robert Barton, CCIE No. 6660, CCDE No. 2013:6

Jerome Henry, CCIE No. 24750

Cisco Press

800 East 96th Street

Indianapolis, Indiana 46240 USA

IoT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things

David Hanes, CCIE No. 3491
Gonzalo Salgueiro, CCIE No. 4541
Patrick Grossetete
Robert Barton, CCIE No. 6660, CCDE No. 2013:6
Jerome Henry, CCIE No. 24750

Published by:
Cisco Press
800 East 96th Street
Indianapolis, IN 46240 USA

Printed in the United States of America

First Printing June 2017

Library of Congress Control Number: 2017937632

ISBN-13: 978-1-58714-456-1

ISBN-10: 1-58714-456-5

Warning and Disclaimer

This book is designed to provide information about the core technologies that make up the Internet of Things, IoT. Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied.

The information is provided on an “as is” basis. The authors, Cisco Press, and Cisco Systems, Inc. shall have neither liability nor responsibility to any person or entity with respect to any loss or damages arising from the information contained in this book or from the use of the discs or programs that may accompany it.

The opinions expressed in this book belong to the author and are not necessarily those of Cisco Systems, Inc.

Trademark Acknowledgments

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Cisco Press or Cisco Systems, Inc., cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

Special Sales

For information about buying this title in bulk quantities, or for special sales opportunities (which may include electronic versions; custom cover designs; and content particular to your business, training goals, marketing focus, or branding interests), please contact our corporate sales department at corpsales@pearsoned.com or (800) 382-3419.

For government sales inquiries, please contact governmentsales@pearsoned.com.

For questions about sales outside the U.S., please contact intlcs@pearson.com.

Feedback Information

At Cisco Press, our goal is to create in-depth technical books of the highest quality and value. Each book is crafted with care and precision, undergoing rigorous development that involves the unique expertise of members from the professional technical community.

Readers' feedback is a natural continuation of this process. If you have any comments regarding how we could improve the quality of this book, or otherwise alter it to better suit your needs, you can contact us through email at feedback@ciscopress.com. Please make sure to include the book title and ISBN in your message.

We greatly appreciate your assistance.

Editor-in-Chief: Mark Taub

Product Line Manager: Brett Bartow

Business Operation Manager, Cisco Press: Ronald Fligge

Executive Editor: Mary Beth Ray

Managing Editor: Sandra Schroeder

Development Editor: Eleanor Bru

Project Editor: Mandie Frank

Copy Editor: Kitty Wilson

Technical Editors: Robb Henshaw, Samuel Pasquier

Editorial Assistant: Vanessa Evans

Cover Designer: Chuti Prasertsith

Composition: codeMantra

Indexer: Cheryl Lenser

Proofreader: Sasirekha Durairajan

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

About the Authors

David Hanes, CCIE No. 3491, is a Technical Leader specializing in IoT and working in Cisco Technical Services as part of the Cloud Support Technical Assistance Center (TAC). With experience in the incubation of new technologies, he is currently leading the TAC support effort for Cisco's IoT cloud solutions. He also has technical expertise in the areas of collaboration and cognitive computing.

David has multiple patents issued and pending in the areas of IoT and collaboration. He is an active participant in the SIP Forum and in the IETF as an RFC contributor and author. David has written and contributed to various industry publications and white papers and is a coauthor of the Cisco Press book *Fax, Modem, and Text for IP Telephony*. He has spoken at industry and technical conferences worldwide and has been honored as a Hall of Fame speaker by Cisco Live.

Since joining Cisco in 1997, David has worked as a TAC engineer for the WAN, WAN Switching, and Multiservice Voice teams; as a team lead for the Multiservice Voice team; as an escalation engineer covering a variety of VoIP technologies; and as a field trial support engineer. Prior to working at Cisco, David was a systems engineer for Sprint, where he gained his first computer networking experience working on the Frame Relay and X.25 protocols. He holds a degree in electrical engineering from North Carolina State University.

Gonzalo Salgueiro, CCIE No. 4541, is a Principal Engineer in Technical Services, working on several emerging technologies and the services opportunities they offer. Gonzalo has spent more than 20 years at Cisco, establishing himself as a subject matter expert, innovator, and industry thought leader in various technologies, including Collaboration, ML/AI, Cloud, and IoT.

Gonzalo is an established member of numerous industry organizations and is a regular presenter and distinguished speaker at a variety of technical industry conferences and Cisco events around the world. He currently holds various industry leadership roles, including serving as a member of the Board of Directors of the SIP Forum, co-chair of the INSIPID and SIPBRANDY IETF working groups, member of the IoT Directorate in the IETF, and co-chair of the WebRTC Task Group, IPv6 Task Group, and FoIP Task Group in the SIP Forum. He is an active contributor to various industry organizations and standardization activities.

Gonzalo co-authored the Cisco Press book *Fax, Modem, and Text for IP Telephony*. He has also co-authored 24 IETF RFCs, 4 IEEE papers, 4 ITU contributions, and numerous industry and academic research papers on a variety of different technical topics. He is also coinventor of 65+ patents (issued and pending) and has contributed to various interop and open source development efforts. Gonzalo received a master's degree in physics from the University of Miami.

Patrick Grossetete is a Distinguished Engineer, Technical Marketing, working on field communication architecture and design (IEEE 802.15.4g/e RF, IEEE 1901.2a PLC, LoRaWAN, IPv6, 6LoWPAN, RPL, …) in the Cisco Internet of Things Connected Group.

He joined Cisco through its acquisition of Arch Rock, where he was Director of Product Management and Customer Solutions, focusing on IPv6-based wireless sensor network technology for smart grid, energy, and environmental optimization applications.

Previously, Patrick led a product management team at Cisco, responsible for a suite of Cisco IOS software technologies, including IPv6 and IP Mobility. Patrick regularly speaks at conferences and industry events, including the IPv6 Forum, which he joined in 1999 as a Cisco representative. Patrick also acts as reviewer on European Commission–sponsored projects, including GEANT and ENVIROFI.

Patrick is coauthor of the books *Global IPv6 Strategies* and *Deploying IPv6 Networks*, published by Cisco Press, as well as several white papers, such as *Unified Field Area Network Architecture for Distribution Automation* (2014) and *IPv6 Architecture for Field Area Networks* (2012). In June 2003, he received the IPv6 Forum Internet Pioneer Award at the San Diego Summit, and he is an IPv6 Forum Fellow. Before his days at Cisco and Arch Rock, he worked at Digital Equipment Corporation as a consulting engineer and was involved with network design and deployment. He received a degree in computer science from the Control Data Institute, Paris, France.

Rob Barton, CCIE No. 6660 (R&S and Security), CCDE No. 2013:6, is a Principal Systems Engineer working in Cisco's Digital Transformation and Innovation organization. Rob is a registered professional engineer (P.Eng) and has worked in the IT industry for more than 20 years, the last 17 of which have been at Cisco. Rob graduated from the University of British Columbia with a degree in engineering physics, where he specialized in computer and radio communications. Rob's areas of interest include wireless communications, IPv6, IoT, and industrial control systems. Rob coauthored the Cisco Press book *End-to-End QoS,* 2nd edition. He resides in Vancouver, Canada, with his wife and two children.

Jerome Henry, CCIE No. 24750, is a Principal Engineer in the Enterprise Infrastructure and Solutions Group at Cisco systems. Jerome has more than 15 years' experience teaching technical Cisco courses in more than 15 countries and 4 languages, to audiences ranging from bachelor's degree students to networking professionals and Cisco internal system engineers. Focusing on his wireless and networking experience, Jerome joined Cisco in 2012. Before that time, he was consulted and taught heterogeneous networks and wireless integration with the European Airespace team, which was later acquired by Cisco to become their main wireless solution. He then spent several years with a Cisco Learning partner, developing networking courses and working on training materials for emerging technologies.

Jerome is a certified wireless networking expert (CWNE No. 45) and has developed multiple Cisco courses and authored several wireless books and video courses. Jerome is also a member of the IEEE, where he was elevated to Senior Member in 2013, and also participates with Wi-Fi Alliance working groups, with a strong focus on IoT and low power. With more than 10,000 hours in the classroom, Jerome was awarded the IT Training Award Best Instructor silver medal. He is based in Research Triangle Park, North Carolina.

Chapter Contributors

The authors would like to thank the following people for their content contributions and industry expertise in the following chapters:

Security (Chapter 7):

Robert Albach, Senior Product Manager, Cisco Industrial Security Portfolio

Rik Irons-McLean, Energy Solutions Architecture Lead, Cisco

Data Analytics (Chapter 8):

Brian Sak, CCIE No. 14441, Technical Solutions Architect, Big Data Analytics, Cisco

Kapil Bakshi, Distinguished Systems Engineer, Big Data and Cloud Computing, US Public Sector

Manufacturing (Chapter 9):

Brandon Lackey, Technology Entrepreneur and Co-Founder, DronePilots Network

Ted Grevers, Engineering Manager, Cisco IoT Vertical Solutions

Oil and Gas (Chapter 10):

Willy Fotso Guifo, Senior Manager IoT Services, Global Business Lead for the Oil and Gas Vertical, Cisco

Dimitrios Tasidis, Solutions Architect, IoT Services Technical Lead for the Oil and Gas Vertical, Cisco

Smart and Connected Cities (Chapter 12):

Munish Khetrapal, Director of Business Development, Smart + Connected Cities group, Cisco

Prachi Goel, Program Analyst, Smart + Connected Cities Solutions Management, Cisco

Mining (Chapter 14):

Lyle Tanner, Customer Solutions Architect, Cisco

Public Safety (Chapter 15):

Kevin Holcomb, Technical Marketing Engineer, IoT Vertical Solutions, Cisco

Kevin McFadden, Vertical Solutions Architect, Cisco

About the Technical Reviewers

Robb Henshaw is the head of Global Communications, IoT Cloud, at Cisco Jasper. Robb was previously the senior director of Global Communications for Jasper, a global IoT platform leader that was acquired by Cisco in March 2016. Prior to working at Jasper, Rob spent 15 years establishing and running global communications programs for mobile and wireless companies, including Airespace (acquired by Cisco), Proxim Wireless, and SugarSync (acquired by J2 Global).

Samuel Pasquier is head of Product Management for the IoT Connectivity portfolio in the Enterprise Networking Group (ENG) at Cisco. Based in San Jose, California, he is responsible for working closely with both the sales team and the engineering team to develop and execute the product strategy and roadmap for the entire IoT Connectivity portfolio, as well as the Cisco fog computing software solution (IOx).

Samuel has been with Cisco Systems since 2004. He spent six years on the Catalyst 6500 engineering team as a software technical leader. He designed, implemented, and sustained several key infrastructure features on the platform. Thanks to that previous experience, Samuel has a very deep understanding of software development and architecture. He was the product line manager for the Catalyst 6500/6800 and led the team of product managers that completed a full refresh of the Catalyst 6500 Portfolio with the launch of Catalyst 6800.

Prior to his current role, Samuel was leading the team of product managers defining the roadmap for the Catalyst Fixed Access Switches Portfolio (Catalyst 2960-X, 3750, 3650, and 3850).

Samuel holds a master's degree in computer science from *Ecole pour l'informatique et les techniques avancees* (EPITA) in France. He is a regular speaker at network industry events.

Dedications

From David Hanes:

To my loving wife, Holly, my best friend, inspiration, and biggest fan, whose unconditional love, support, and selflessness are the foundation for all the successes and achievements in my life, and to my amazing children, Haley, Hannah, and Kyle, who are true joys and blessings and ensure that my life never has a dull moment.

I would also like to dedicate this book to my wonderful parents—to my Dad for instilling in me a love of learning and always challenging me to push to new heights, and to my Mom for providing the encouragement and confidence to reach those heights.

From Gonzalo Salgueiro:

This book is dedicated to my family. First and foremost, to my loving wife and best friend, Becky, who is my inspiration and makes every day a dream come true. It's always been you—then, now, and forever. To our four amazing children: Alejandro, Sofia, Gabriela, and Mateo. They fill my life with immeasurable joy, wonder, and purpose. I love you all.

I also dedicate this book to my parents, Alberto and Elena, who are my inspiration and to whom I owe everything for all that I am today. I simply don't have the words to express my eternal gratitude for all you have sacrificed and given on my behalf.

Finally, I dedicate this book to my grandmother, Nelida, whom we tragically lost this past year. Only recently have I realized the extent to which my life is filled with the fingerprints of her kindness and grace.

From Patrick Grossetete:

To Mya and all new generations that will live with the Internet of Things.

From Rob Barton:

First, I would like to dedicate this book to my beautiful wife, Loretta. You have been my biggest supporter and cheerleader and have taught me so much over the past 20 years. Without your unending encouragement (not to mention your extreme patience and willingness to share my time with another book project), this undertaking would never have happened. Thanks to your loving support, I find myself in a position where I can pass on what I have learned to the next generation of Internet engineers. 너무 사랑해요.

I also want to dedicate this book to my parents, Richard and Peggy. It’s thanks to mom and dad that I began to develop my interest in science and engineering, beginning with my first Lego set when I was four years old, down to the many hours Dad and I spent discussing physics and calculus and watching *The Mechanical Universe* together when I was a teenager.

I love you all so much!

From Jerome Henry:

This book is dedicated to my children, Ines and Hermes. Your curious minds stimulate my research every day.

Acknowledgments

We would like to acknowledge the following people, who helped us complete this project.

Michael Boland: Michael, your suggestions and industry expertise in mining were an incredible asset in the development of this book. We truly appreciate your time, thoughtfulness, and clear suggestions!

Maik Seewald: Maik, you are a true Internet engineer! Your assistance and suggestions helped clarify the utilities industry's direction on standards and protocols, helping make our book as up-to-date as possible.

Rick Geiger: Rick, you are one of the top utility experts in the world. Thanks for your help and suggestions on the utilities chapter.

Ken Batke: Thanks for your contribution and guidance of the utilities section. Your feedback was much appreciated.

Mahyar Khosravi: Mahyar, thanks for lending us your expertise in the oil and gas industry and all the feedback and suggestions to the oil and gas chapter.

Dave Cronberger: Huge thanks go out to Dave for his excellent guidance and feedback on the book. Dave, you're never shy to tell someone the straightforward advice they need and bring them back on track. Your reviews and feedback on the manufacturing section were invaluable!

Kyle Connor and Barb Rigel: Thanks, Kyle and Barb, for your very thorough reviews and feedback on the transportation section. Your guidance and insights on the state of this quickly transforming industry were an incredible help to our team.

Eleanor Bru: Ellie was our development editor on this project, and it was an absolute pleasure working with her. Thanks for your dedication and patience and for making the review process go so well.

Mary Beth Ray: We'd also like to send out a big thanks to Mary Beth for keeping us on track and supporting us all the way through this challenging project and ensuring that we had everything we needed to bring the book to completion.

Mandie Frank: As our project editor, Mandie tied up all the loose ends and helped us put the finishing touches on this book. Thanks, Mandie, for all your help in guiding us to the finish line!

Kitty Wilson: As our copyeditor, we thank you for the long hours you contributed in making the final edits to our text. Your meticulous attention to detail is greatly appreciated.

We would also like to acknowledge our management teams for their unwavering support and understanding as we worked long hours to meet the many deadlines associated with a publication such as this.

David would like to give a special thanks to Marty Martinez, Rob Taylor, Jitendra Lal, and Shane Kirby, for their guidance and support.

Gonzalo would like to thank the TS leadership team, especially Marc Holloman, Marty Martinez, and Tom Berghoff. This book wouldn't have been possible without their unwavering encouragement, belief, and support.

Patrick thanks the product management and engineering team from the IoT Connected Business Unit for their support and tireless efforts in making IoT products real.

Jerome would like to give a special thanks to Matt MacPherson, for sharing his wisdom and vision on the convergence of enterprise, 5G, and IoT.

Rob would also like to express his thanks to his leadership team, including Dr. Rick Huijbregts and Bernadette Wightman, for their strong support and backing during this project. Your focus on innovation is at the heart of the digital transformation revolution and has inspired much of the wisdom woven throughout the pages of this book.

Contents at a Glance

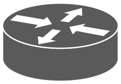
Contents

Reader Services

Register your copy at www.ciscopress.com/title/ISBN for convenient access to downloads, updates, and corrections as they become available. To start the registration process, go to www.ciscopress.com/register and log in or create an account.* Enter the product ISBN 9781587144561 and click Submit. Once the process is complete, you will find any available bonus content under Registered Products.

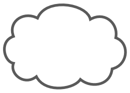
* Be sure to check the box saying that you would like to hear from us to receive exclusive discounts on future editions of this product.

Icons Used in This Book

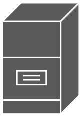


Router

Switch



Cloud



File/
Application Server

Command Syntax Conventions

The conventions used to present command syntax in this book are the same conventions used in the IOS Command Reference. The Command Reference describes these conventions as follows:

- **Boldface** indicates commands and keywords that are entered literally as shown. In actual configuration examples and output (not general command syntax), boldface indicates commands that are manually input by the user (such as a **show** command).
- *Italic* indicates arguments for which you supply actual values.
- Vertical bars (|) separate alternative, mutually exclusive elements.
- Square brackets ([ ]) indicate an optional element.
- Braces ({ }) indicate a required choice.
- Braces within brackets ([{ }]) indicate a required choice within an optional element.

Foreword

Greetings from the past. I am writing this foreword in what is for you the bygone technological era of February 2017. Back then (or now, to me), most cars still had human drivers. We still needed traffic lights, and most of those lights ran on timers, completely blind to the traffic on the streets. As I write this, most residential utility meters are mechanical, and utility workers have to walk from house to house to get readings. The vast majority of toasters can't tweet.

I joined Cisco in 2013 and became the company's Internet of Things leader in 2015. The scope and velocity of the technological change my team sees is immense—so much so that book forewords can have a short shelf life.

But we can prepare for the changes and opportunities that are coming at us. We will have to use different tools from the ones we used to build the current Internet. We need a rock-solid understanding of the fundamentals of the Internet of Things: Where we are today, the challenges we face, and where those opportunities lie. Cisco's most knowledgeable engineers and top technical talent wrote this book so we could build toward this future together.

Where Things Are

I expect this book to be a useful tool for you, even if you don't pick it up until 2020, when the number of "Internet of Things" (if we still call it that) devices might have reached 50 billion, from a paltry 6.4 billion in 2016. Manufacturing plants will be smarter and more efficient than they've ever been, thanks to their capabilities to process, share, and react to sensor information and other data. Complex machines like cars will be comprehensively metered, down to the component level, with their massive data streams fanning out into vast analytics systems that serve life-safety, ecological, and financial services—and even the manufacturing plants that made them—in real time. The things will become so smart—tractors, teacups, tape measures—that the product companies will be transformed into services companies.

It will have been the biggest technology transition the world has ever seen.

Currently, the networking protocols to collect and collate and analyze and transmit that data are still evolving—fast. We have a robust and stable Internet, but it was built to connect people and general-purpose computers, not billions of highly specialized devices sending out constant streams of machine data.

Our global network is designed to mimic point-to-point connectivity, and it is, for the most part, neutral to the devices that connect to it and to the types of data they are designed to send and receive. Currently, several companies, including Cisco, are coming up with ways to add a layer of mediation between the billions of devices coming online and the data and analytical warehouses that will be the repositories of their data for business and other applications. (We call this layer "the edge," for now.)

Since a lot of the data and telemetry that devices create will need to be sent wirelessly, we're also doing what we can to improve the reliability and speed of data transfer, as well as to lower its latency and the power it takes to send each bit. There are several emerging wireless standards in this race. And in a few years, there will still be several—because different types of devices and applications will need different things from their wireless systems. Currently, the mobile carriers are the big players that are being joined by the largest consumers of data services, like the energy and transportation companies. The next few years are going to see a lot of competition and innovation as old and new companies compete to be the transporters of all this information.

We're also working to make sure that IoT devices themselves can strengthen the security of the networks they use. Right now (in your past), the network itself has very limited knowledge of what types of data it should be sending and what it should not be. Devices can get hijacked to attack other devices—or the network itself. By the time you read this, I am confident that this security problem along with other IoT challenges, such as scalability and interoperability issues, will be closer to getting solved. This book will help us get there. It is an educational resource that captures the fundamentals of IoT in a coherent and comprehensive manner. IoT is poised to change our world, and this book provides the necessary foundation for understanding and navigating the shifting IoT landscape.

The Adoption Curve

From my vantage point in 2017, it's clear we have a lot of work ahead of us to make the Internet of Things into a fabric that all businesses can easily connect to. I'm sure it's going to get done, though. And soon. I know this because we're building the tools ourselves here at Cisco and because I talk all the time to business leaders and entrepreneurs who are betting their companies on IoT-powered processes.

Building IoT solutions, keeping them safe, making them inexpensive and maintainable, and processing and profiting from the data they generate are all enormous opportunities. My team's job is to make all these jobs easier for you, and it all starts with education—ours and yours.

— **Rowan Trollope**, SVP and GM of IoT and Applications Groups, Cisco

Introduction

A major technology shift is happening in our world, and it is centered around the Internet of Things (IoT). The IoT is all about connecting the unconnected. Most of the objects in our current world are not connected to a computer network, but that paradigm is rapidly changing. Previously unconnected objects that are all around us are being provided with the ability to communicate with other objects and people, which in turn drives new services and efficiencies in our daily lives. This is the basic premise behind IoT and illustrates why some theorize that it will be as transformative as the Industrial Revolution.

We, the authors of this book, have decades of computer networking experience, much of it focused on IoT and related technologies. Our combined experience with IoT ranges from early product deployments and testing, to network design, to implementation, training, and troubleshooting. This experience allowed us to take a pragmatic approach to writing on this subject and distill the essential elements that form the foundation or fundamentals for this topic. This book embodies principal elements that you need for understanding IoT from both a technical perspective and an industry point of view.

This book leverages a three-part approach for teaching the fundamentals of IoT. Part I provides a high-level overview of IoT and what you need to know from a design perspective. Part II takes you through the technical building blocks of IoT, including the pertinent technologies and protocols. Finally, Part III steps you through common industry use cases so you can see how IoT is applied in the real world.

To successfully work in the IoT area, you must have a fundamental understanding of IoT principles and use cases. This book provides this knowledge in a logical format that makes it not only a great general resource for learning about IoT now but also a handy reference for more specific IoT questions you may have in the future.

Who Should Read This Book?

This book was written for networking professionals looking for an authoritative and comprehensive introduction to the topic of IoT. It is focused on readers who have networking experience and are looking to master the essential concepts and technologies behind IoT and how they are applied, resulting in basic proficiency. Therefore, readers should have a basic understanding of computer networking concepts and be familiar with basic networking terminology. Readers may be advanced-level networking students or hold titles or positions such as network operator, administrator, and manager; network designer or architect; network engineer; network technician; network analyst or consultant; and network database administrator.

How This Book Is Organized

Part I, "Introduction to IoT"

Part 1 helps you make sense of the IoT word. This word has often been misused and can cover multiple realities. This first part of the book helps you understand what exactly IoT is and provides an overview of the landscape of smart objects, from those that control telescope mirrors with hundreds of actions per seconds, to those that send rust information once a month. This part also shows you how IoT networks are designed and constructed.

Chapter 1, "What Is IoT?"

This chapter provides an overview of the history and beginnings of IoT. This chapter also examines the convergence of operational technology (OT) and informational technology (IT) and provides a reference model to position IoT in the general network landscape.

Chapter 2, "IoT Network Architecture and Design"

Multiple standards and industry organizations have defined specific architectures for IoT, including ETSI/oneM2M and the IoT World Forum. This chapter compares those architectures and suggests a simplified model that can help you articulate the key functions of IoT without the need for vertical-specific elements. This chapter also guides you through the core IoT functional stack and the data infrastructure stack.

Part II, "Engineering IoT Networks"

Once you understand the IoT landscape and the general principles of IoT networks, Part II takes a deep dive into IoT network engineering, from smart objects and the network that connects them to applications, data analytics, and security. This part covers in detail each layer of an IoT network and examines for each layer the protocols in place (those that have been there for a long time and new protocols that are gaining traction), use cases, and the different architectures that define an efficient IoT solution.

Chapter 3, "Smart Objects: The 'Things' in IoT"

Smart objects can be of many types, from things you wear to things you install in walls, windows, bridges, trains, cars, or streetlights. This chapter guides you through the different types of smart objects, from those that simply record information to those that are programmed to perform actions in response to changes.

Chapter 4, "Connecting Smart Objects"

Once you deploy smart objects, they need to connect to the network. This chapter guides you through the different elements you need to understand to build a network for IoT: connection technologies, such as 802.15.4, 802.15g, 802.15e 1901.2a, 802.11ah, LoRaWAN, NB-IoT, and other LTE variations; wireless bands and ranges; power considerations; and topologies.

Chapter 5, "IP as the IoT Network Layer"

Early IoT protocols did not rely on an OSI network layer. This chapter shows you how, as IoT networks now include millions of sensors, IP has become the protocol of choice for network connectivity. This chapter also details how IP was optimized, with enhancements like 6LoWPAN, 6TiSCH, and RPL, to adapt to the low-power and lossy networks (LLNs) where IoT usually operates.

Chapter 6, "Application Protocols for IoT"

Smart objects need to communicate over the network with applications to report on environmental readings or receive information, configurations, and instructions. This chapter guides you through the different common application protocols, from MQTT, CoAP, and SCADA to generic and web-based protocols. This chapter also provides architecture recommendations to optimize your IoT network application and communication efficiency.

Chapter 7, "Data and Analytics for IoT"

Somewhere in a data center or in the cloud, data coming from millions of sensors is analyzed and correlated with data coming from millions of others. *Big data* and *machine learning* are keywords in this world. This chapter details what big data is and how machine learning works, and it explains the tools used to make intelligence of large amount of data and to analyze in real time network flows and streams.

Chapter 8, "Securing IoT"

Hacking an IoT smart object can provide very deep access into your network and data. This chapter explains the security practices for IT and OT and details how security is applied to an IoT environment. This chapter also describes tools to conduct a formal risk analysis on an IoT infrastructure.

Part III, "IoT in Industry"

Once you know how to architect an IoT network, Part III helps you apply that knowledge to key industries that IoT is revolutionizing. For each of the seven verticals covered in this part, you will learn how IoT can be used and what IoT architecture is recommended to increase safety, operational efficiency, and user experience.

Chapter 9, "Manufacturing"

Any gain in productivity can have a large impact on manufacturing, and IoT has introduced a very disruptive change in this world. This chapter explains connected manufacturing and data processing for this environment, and it details the architecture and components of a converged factory, including IACS and CPwE. This chapter also examines the process automation protocols, including EtherNet/IP, PROFINET, and Modbus/TCP.

Chapter 10, "Oil and Gas"

Oil and gas are among the most critical resources used by modern society. This chapter shows how IoT is massively leveraged in this vertical to improve operational efficiency. This chapter also addresses the sensitive topic of OT security and provides architectural recommendations for IoT in the oil and gas world.

Chapter 11, "Utilities"

Utility companies provide the services that run our cities, businesses, and entire economy. IoT in this vertical, and the ability to visualize and control energy consumption, is critical for the utility companies and also for end users. This chapter guides you through the GridBlocks reference model, the substation and control systems, and the FAN GridBlocks, to help you understand the smart grid and how IoT is used in this vertical.

Chapter 12, "Smart and Connected Cities"

Smart and connected cities include street lighting, smart parking, traffic optimization, waste collection and management, and smart environment. These various use cases are more and more being combined into organized citywide IoT solutions where data and smart objects serve multiple purposes. This chapter discusses the various IoT solutions for smart and connected cities.

Chapter 13, "Transportation"

This chapter talks about roadways, rail, mass transit, and fleet management. You will learn how IoT is used to allow for communication between vehicles and the infrastructure through protocols like DSRC and WAVE and how IoT increases the efficiency and safety of the transportation infrastructure.

Chapter 14, "Mining"

The mining industry is often described as "gigantic vehicles moving gigantic volumes of material." IoT is becoming a key component in this world to maintain competiveness while ensuring safety. From self-driving haulers to radar-guided 350-metric-ton shovels, this chapter shows you the various use cases of IoT in mining. This chapter also suggests an architectural IoT strategy for deploying smart objects in an ever-changing and often extreme environment.

Chapter 15, "Public Safety"

The primary objective of public safety organizations is to keep citizens, communities, and public spaces safe. These organizations have long been at the forefront of new technology adoption, and IoT has become a key component of their operations. This chapter describes the emergency response IoT architecture and details how public safety operators leverage IoT to better exchange information and leverage big data to respond more quickly and efficiently to emergencies.

Chapter 8

Securing IoT

It is often said that if World War III breaks out, it will be fought in cyberspace. As IoT brings more and more systems together under the umbrella of network connectivity, security has never been more important. From the electrical grid system that powers our world, to the lights that control the flow of traffic in a city, to the systems that keep airplanes flying in an organized and efficient way, security of the networks, devices, and the applications that use them is foundational and essential for all modern communications systems. Providing security in such a world is not easy. Security is among the very few, if not the only, technology disciplines that must operate with external forces continually working against desired outcomes. To further complicate matters, these external forces are able to leverage traditional technology as well as nontechnical methods (for example, physical security, operational processes, and so on) to meet their goals. With so many potential attack vectors, information and cybersecurity is a challenging, but engaging, topic that is of critical importance to technology vendors, enterprises, and service providers alike.

Information technology (IT) environments have faced active attacks and information security threats for many decades, and the incidents and lessons learned are well-known and documented. By contrast, operational technology (OT) environments were traditionally kept in silos and had only limited connection to other networks. Thus, the history of cyber attacks on OT systems is much shorter and has far fewer incidents documented. Therefore, the learning opportunities and the body of cataloged incidents with their corresponding mitigations are not as rich as in the IT world. Security in the OT world also addresses a wider scope than in the IT world. For example, in OT, the word *security* is almost synonymous with *safety*. In fact, many of the industrial security standards that form the foundation for industrial IoT security also incorporate equipment and personnel safety recommendations.

It is for these reasons that this chapter focuses on the core principles of securing OT environments. IT security is a vast domain with many books dedicated to its various aspects. An exhaustive treatment of the subject is simply not possible in one chapter, so we instead focus on OT security and the elements of IT security that are fundamental to OT security. In addition, the industry-specific chapters in Part III, “IoT in Industry,” discuss the application of security to specific industry verticals.

This chapter provides a historical perspective of OT security, how it has evolved, and some of the common challenges it faces. It also details some of the key differences between securing IT and OT environments. Finally, this chapter explores a number of practical steps for creating a more secure industrial environment, including best practices in introducing modern IT network security into legacy industrial environments.
It includes the following sections:

- **A Brief History of OT Security:** This section provides an overview of how OT environments have evolved and the impact that the evolution has had on securing operational networks.
- **Common Challenges in OT Security:** This section provides a synopsis of different security challenges in operational environments, including legacy systems and insecure protocols and assets.
- **How IT and OT Security Practices and Systems Vary:** This section provides a comparison between the security practices in enterprise IT environments and operational industrial environments.
- **Formal Risk Analysis Structures: OCTAVE and FAIR:** This section provides a holistic view of securing an operational environment and a risk assessment framework that includes the people, processes, and vendor ecosystem components that make up a control system.
- **The Phased Application of Security in an Operational Environment:** This section provides a description of a phased approach to introducing modern network security into largely preexisting legacy industrial networks.

A Brief History of OT Security

To better understand the current situation in industrial environments, it is important to differentiate between assumptions and realities. Few topics in information technology inspire more fear, uncertainty, or doubt than cybersecurity. This chapter is therefore limited to incidents and data sources from official sources rather than public media reports or uncorroborated third-party accounts.

More than in most other sectors, cybersecurity incidents in industrial environments can result in physical consequences that can cause threats to human lives as well as damage to equipment, infrastructure, and the environment. While there are certainly traditional IT-related security threats in industrial environments, it is the physical manifestations and impacts of the OT security incidents that capture media attention and elicit broad-based public concern.

One example of a reported incident where physical damage was caused by a cybersecurity attack is the Stuxnet malware that damaged uranium enrichment systems in Iran. Another example is an event that damaged a furnace in a German smelter. In both incidents, multiple steps led to the undesirable outcomes. Many of the security policies and mitigation procedures that were in place went unheeded; however, if properly implemented, they could have impeded or possibly stopped the attacks entirely. For example, Stuxnet is thought to have been deployed on USB memory sticks up to two years before it was finally identified and discovered.

In addition to physical damage, operational interruptions have occurred in OT environments due to cybersecurity incidents. For example, in 2000, the sewage control system of Maroochy Shire in Queensland, Australia, was accessed remotely, and it released 800,000 liters of sewage into the surrounding waterways. In 2015, the control systems of the Ukrainian power distribution operator Kyiv Oblenergo were remotely accessed by attackers, causing an outage that lasted several hours and resulted in days of degraded service for thousands of customers. In both cases, known mitigation techniques could have been applied to detect the attacks earlier or block the ability to hijack production systems and affect service.

Historically, attackers were skilled individuals with deep knowledge of technology and the systems they were attacking. However, as technology has advanced, tools have been created to make attacks much easier to carry out. To further complicate matters, these tools have become more broadly available and more easily obtainable. Compounding this problem, many of the legacy protocols used in IoT environments are many decades old, and there was no thought of security when they were first developed. This means that attackers with limited or no technical capabilities now have the potential to launch cyber attacks, greatly increasing the frequency of attacks and the overall threat to end operators. It is, however, a common misconception that attackers always have the advantage and that end operators lack effective defensive capabilities. An important advantage for operators is the fact that they are far more familiar with their environment and have a better understanding of their processes, and can thus leverage multiple technologies and capabilities to defend their networks against attack. This is critical as networks will continue to face ever-evolving and changing methods of attack that will be increasingly difficult to defend against and respond to.

Communication networks, both local and geographically dispersed, have been used in industrial environments for decades. For example, remote monitoring of substations in utilities and communications between semi-autonomous systems in manufacturing are long-standing examples of such OT networks. These OT-specific communication systems have typically been standalone and physically isolated from the traditional IT enterprise networks in the same companies. While it follows the traditional logic of "security through obscurity," this form of network compartmentalization has led to the independent evolution of IT and OT networks, with interconnections between the environments strictly segregated and monitored.

The isolation between industrial networks and the traditional IT business networks has been referred to as an "air gap," suggesting that there are no links between the two. While there are clearly examples of such extreme isolation in some industries, it is actually not an accurate description of most IoT networks today. Broadly speaking, there is a varying amount of interconnection between OT and IT network environments, and many interdependencies between the two influence the level of interconnection.

In addition to the policies, regulations, and governance imposed by the different industrial environments, there is also a certain amount of end-user preference and deployment-specific design that determines the degree of isolation between IT and OT environments. While some organizations continue to maintain strict separation, others are starting to allow certain elements of interconnection. One common example of this is the use of Ethernet and IP to transport control systems in industrial environments.
As much as IT and OT networks are still operated and managed separately in a good portion of the world, the prevailing trend is to consolidate networks based on IT-centric technologies such as TCP/IP, Ethernet, and common APIs.

This evolution of ever-increasing IT technologies in the OT space comes with the benefits of increased accessibility and a larger base of skilled operators than with the nonstandard and proprietary communication methods in traditional industrial environments. The challenges associated with these well-known IT standards is that security vulnerabilities are more widely known, and abuse of those systems is often easier and occurs on a much larger scale. This accessibility and scale makes security a major concern, particularly because many systems and devices in the operational domain were never envisioned to run on a shared, open standards–based infrastructure, and they were not designed and developed with high levels of built-in security capabilities.

Projects in industrial environments are often capital intensive, with an expected life span that can be measured in decades. Unlike in IT-based enterprises, OT-deployed solutions commonly have no reason to change as they are designed to meet specific (and often single-use) functions, and have no requirements or incentives to be upgraded. A huge focus and priority in OT is system uptime and high availability, so changes are typically only made to fix faults or introduce new system capabilities in support of that goal. As a result, deployed OT systems often have slower development and upgrade cycles and can quickly become out of sync with traditional IT network environments. The outcome is that both OT technologies and the knowledge of those looking after those operational systems have progressed at a slower pace than their IT counterparts.

Most of the industrial control systems deployed today, their components, and the limited associated security elements were designed when adherence to published and open standards were rare. The proprietary nature of these systems meant that threats from the outside world were unlikely to occur and were rarely addressed. There has, however, been a growing trend whereby OT system vulnerabilities have been exposed and reported. This increase is depicted in Figure 8-1, which shows the history of vulnerability disclosures in industrial control systems (ICSs) since 2010. While the number of reports has been increasing over the past years, it is likely that there are still many others that are not reported or discovered.

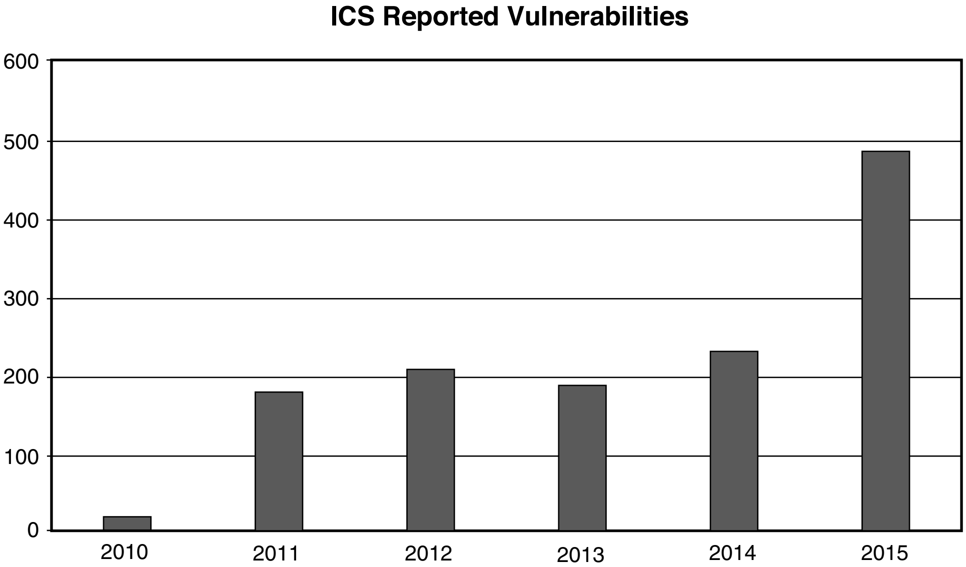


Figure 8-1 *History of Vulnerability Disclosures in Industrial Control Systems Since 2010 (US Industrial Control Systems Cyber Emergency Response Team (ICS-CERT) https://ics-cert.us-cert.gov).*

Given the slow rate of change and extended upgrade cycles of most OT environments, the investment in security for industrial communication and compute technologies has historically lagged behind the investment in securing traditional IT enterprise environments.

Common Challenges in OT Security

The security challenges faced in IoT are by no means new and are not limited to specific industrial environments. The following sections discuss some of the common challenges faced in IoT.

Erosion of Network Architecture

Two of the major challenges in securing industrial environments have been initial design and ongoing maintenance. The initial design challenges arose from the concept that networks were safe due to physical separation from the enterprise with minimal or no connectivity to the outside world, and the assumption that attackers lacked sufficient knowledge to carry out security attacks. In many cases, the initial network design is sound and even follows well-defined industrial best practices and standards, such as the Purdue Model for Control Hierarchy that was introduced in Chapter 2, “IoT Network Architecture and Design.” The challenge, and the biggest threat to network security, is standards and best practices either being misunderstood or the network being poorly maintained. In fact, from a security design perspective, it is better to know that communication paths are insecure than to not know the actual communication paths. It is more common that, over time, what may have been a solid design to begin with is eroded through ad hoc updates and individual changes to hardware and machinery without consideration for the broader network impact. This kind of organic growth has led to miscalculations of expanding networks and the introduction of wireless communication in a standalone fashion, without consideration of the impact to the original security design. These uncontrolled or poorly controlled OT network evolutions have, in many cases, over time led to weak or inadequate network and systems security.

There is a wide variety in secured network designs within and across different industries. For example, power utilities have a strong history of leveraging modern technologies for operational activities, and in North America there are regulatory requirements in place from regulatory authorities, such as North American Electric Reliability Corporation's (NERC's) Critical Infrastructure Protection (CIP), discussed in greater detail in Chapter 11, "Utilities"), to implement secure network connectivity and control with reasonably prescriptive actions. By contrast, in other industries, there are often no legislative requirements or compliance policies, which has resulted in widespread differences in security capabilities.

In many industries, the control systems consist of packages, skids, or components that are self-contained and may be integrated as semi-autonomous portions of the network. These packages may not be as fully or tightly integrated into the overall control system, network management tools, or security applications, resulting in potential risk.

Pervasive Legacy Systems

Due to the static nature and long lifecycles of equipment in industrial environments, many operational systems may be deemed legacy systems. For example, in a power utility environment, it is not uncommon to have racks of old mechanical equipment still operating alongside modern intelligent electronic devices (IEDs). In many cases, legacy components are not restricted to isolated network segments but have now been consolidated into the IT operational environment. From a security perspective, this is potentially dangerous as many devices may have historical vulnerabilities or weaknesses that have not been patched and updated, or it may be that patches are not even available due to the age of the equipment.

Beyond the endpoints, the communication infrastructure and shared centralized compute resources are often not built to comply with modern standards. In fact, their communication methods and protocols may be generations old and must be interoperable with the oldest operating entity in the communications path. This includes switches, routers, firewalls, wireless access points, servers, remote access systems, patch management, and network management tools. All of these may have exploitable vulnerabilities and must be protected.

Insecure Operational Protocols

Many industrial control protocols, particularly those that are serial based, were designed without inherent strong security requirements. Furthermore, their operation was often within an assumed secure network. In addition to any inherent weaknesses or vulnerabilities, their operational environment may not have been designed with secured access control in mind.

Industrial protocols, such as supervisory control and data acquisition (SCADA) (refer to Chapter 6, "Application Protocols for IoT"), particularly the older variants, suffer from common security issues. Three examples of this are a frequent lack of authentication between communication endpoints, no means of securing and protecting data at rest or in motion, and insufficient granularity of control to properly specify recipients or avoid default broadcast approaches. These may not be as critical in self-contained systems, but between zones or on longer network segments, such as a WAN (particularly a public WAN), they may be significant considerations.

The structure and operation of most of these protocols is often publicly available. While they may have been originated by a private firm, for the sake of interoperability, they are typically published for others to implement. Thus, it becomes a relatively simple matter to compromise the protocols themselves and introduce malicious actors that may use them to compromise control systems for either reconnaissance or attack purposes that could lead to undesirable impacts in normal system operation.

The following sections discuss some common industrial protocols and their respective security concerns. Note that many have serial, IP, or Ethernet-based versions, and the security challenges and vulnerabilities are different for the different variants.

Modbus

Modbus is commonly found in many industries, such as utilities and manufacturing environments, and has multiple variants (for example, serial, TCP/IP). It was created by the first programmable logic controller (PLC) vendor, Modicon, and has been in use since the 1970s. It is one of the most widely used protocols in industrial deployments, and its development is governed by the Modbus Organization. For more details on Modbus, refer to Chapter 6.

The security challenges that have existed with Modbus are not unusual. Authentication of communicating endpoints was not a default operation because it would allow an inappropriate source to send improper commands to the recipient. For example, for a message to reach its destination, nothing more than the proper Modbus address and function call (code) is necessary.

Some older and serial-based versions of Modbus communicate via broadcast. The ability to curb the broadcast function does not exist in some versions. There is potential for a recipient to act on a command that was not specifically targeting it. Furthermore, an attack could potentially impact unintended recipient devices, thus reducing the need to understand the details of the network topology.

Validation of the Modbus message content is also not performed by the initiating application. Instead, Modbus depends on the network stack to perform this function. This could open up the potential for protocol abuse in the system.

DNP3 (Distributed Network Protocol)

DNP3 is found in multiple deployment scenarios and industries. It is common in utilities and is also found in discrete and continuous process systems. Like many other ICS/SCADA protocols, it was intended for serial communication between controllers and simple IEDs. (For more detailed information on DNP3, refer to Chapter 6.)

There is an explicit "secure" version of DNP3, but there also remain many insecure implementations of DNP3 as well. DNP3 has placed great emphasis on the reliable delivery of messages. That emphasis, while normally highly desirable, has a specific weakness from a security perspective. In the case of DNP3, participants allow for unsolicited responses, which could trigger an undesired response. The missing security element here is the ability to establish trust in the system's state and thus the ability to trust the veracity of the information being presented. This is akin to the security flaws presented by Gratuitous ARP messages in Ethernet networks, which has been addressed by Dynamic ARP Inspection (DAI) in modern Ethernet switches.

ICCP (Inter-Control Center Communications Protocol)

ICCP is a common control protocol in utilities across North America that is frequently used to communicate between utilities. Given that it must traverse the boundaries between different networks, it holds an extra level of exposure and risk that could expose a utility to cyber attack.

Unlike other control protocols, ICCP was designed from inception to work across a WAN. Despite this role, initial versions of ICCP had several significant gaps in the area of security. One key vulnerability is that the system did not require authentication for communication. Second, encryption across the protocol was not enabled as a default condition, thus exposing connections to man-in-the-middle (MITM) and replay attacks.

OPC (OLE for Process Control)

OPC is based on the Microsoft interoperability methodology Object Linking and Embedding (OLE). This is an example where an IT standard used within the IT domain and personal computers has been leveraged for use as a control protocol across an industrial network.

In industrial control networks, OPC is limited to operation at the higher levels of the control space, with a dependence on Windows-based platforms. Concerns around OPC begin with the operating system on which it operates. Many of the Windows devices in the operational space are old, not fully patched, and at risk due to a plethora of well-known vulnerabilities. The dependence on OPC may reinforce that dependence. While newer versions of OPC have enhanced security capabilities, they have also opened up new communications modes, which have both positive and negative security potential.

Of particular concern with OPC is the dependence on the Remote Procedure Call (RPC) protocol, which creates two classes of exposure. The first requires you to clearly understand the many vulnerabilities associated with RPC, and the second requires you to identify the level of risk these vulnerabilities bring to a specific network.

International Electrotechnical Commission (IEC) Protocols

The IEC 61850 standard was created to allow vendor-agnostic engineering of power utility systems, which would, in turn, allow interoperability between vendors and standardized communication protocols. Three message types were initially defined: MMS (Manufacturing Message Specification), GOOSE (Generic Object Oriented Substation Event), and SV (Sampled Values). Web services was a fourth protocol that was added later. Here we provide a short summary of each, but for more information on IEC protocols, see Chapter 11:

- **MMS (61850-8.1):** MMS is a client/server protocol that leverages TCP/IP and operates at Layer 3. It provides the same functionality as other SCADA protocols, such as IEC 60870 and Modbus.
- **GOOSE (61850-8.1):** GOOSE is a Layer 2 protocol that operates via multicast over Ethernet. It allows IEDs to exchange data "horizontally," between bays and between substations, especially for interlocking, measurement, and tripping signals.
- **SV (61850-9-2):** SV is a Layer 2 protocol that operates via multicast over Ethernet. It carries voltage and current samples, typically on the process bus, but it can also flow over the station bus.

Both GOOSE and SV operate via a publisher/subscriber model, with no reliability mechanism to ensure that data has been received.

IEC 61850 has several known security deficiencies that could be leveraged by skilled attackers to compromise a control system. Authentication is embedded in MMS, but it is based on clear-text passwords, and authentication is not available in GOOSE or SV. Firmware is typically not signed, which means there is no way to verify its authenticity or integrity. GOOSE and SV have limited message integrity, which makes it relatively easy to impersonate a publisher.

When the standard was first released, there was minimal security capability in these protocols, but this is being addressed by IEC 62351 with the introduction of well-known IT-based security measures, such as certificate exchange.

IEC 60870 is widely used for SCADA telecontrol in Europe, particularly in the power utility industry, and for widely geographically dispersed control systems. Part 5 of the standard outlines the communication profiles used between endpoints to exchange telecontrol messages. 60870-5-101 is the serial implementation profile, 60870-5-104 is the IP implementation profile, and 60870-5-103 is used for protection equipment. Again, in the early iterations of IEC 60870-5, security was lacking. This is now being addressed by IEC 62351, with the 60870-5-7 security extensions work, applicable to 60870-101 and 60870-104.

Other Protocols

At times, discussions about the security of industrial systems are decidedly focused on industrial control protocols as if they were the sum total of what would be observed or considered. This assumption is narrow-minded and problematic on many levels. In fact, it is highly recommended that a security practitioner passively identify all aspects of the traffic traversing the network prior to implementing any kind of controls or security measures therein. Of particular importance are proper accounting, handling, and understanding of the most basic protocols, transport mechanisms, and foundational elements of any network, including ARP, UDP, TCP, IP, and SNMP.

Some specialized environments may also have other background control protocols. For example, many IoT networks reach all the way to the individual sensors, so protocols such as Constrained Application Protocol (CoAP) (see Chapter 6) and Datagram Transport Layer Security (DTLS) are used, and have to be considered separately from a security perspective.

Device Insecurity

Beyond the communications protocols that are used and the installation base of legacy systems, control and communication elements themselves have a history of vulnerabilities. As mentioned earlier in this chapter (see Figure 8-1), prior to 2010, the security community paid little attention to industrial compute, and as a result, OT systems have not gone through the same “trial by fire” as IT systems. Figure 8-2 shows this graphically by simply overlaying the count of industrial security topics presented at the Black Hat security conference with the number of vulnerabilities reported for industrial control systems. The correlation between presentations on the subject of OT security at Black Hat and the number of vulnerabilities discovered is obvious, including the associated slowing of discoveries.

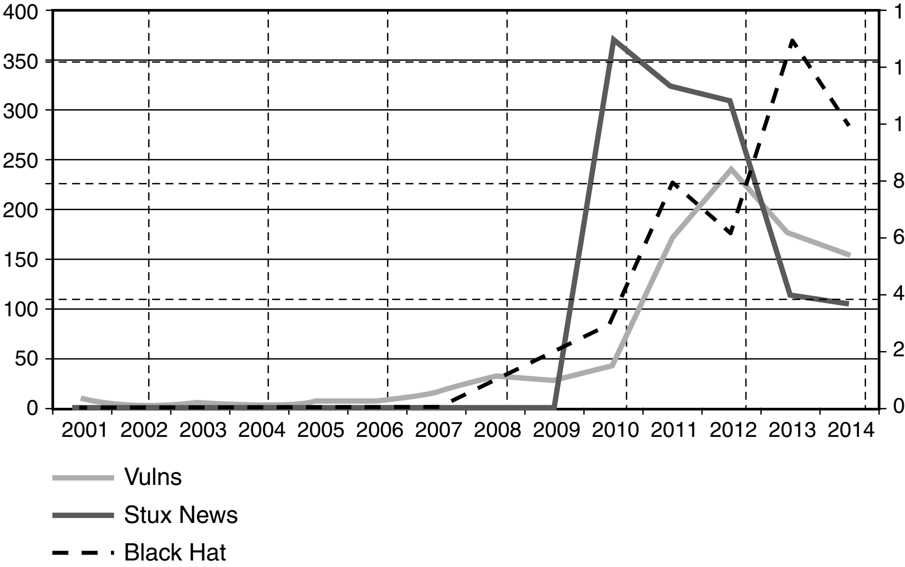


Figure 8-2 *Correlation of Industrial Black Hat Presentations with Discovered Industrial Vulnerabilities (US Industrial Control Systems Cyber Emergency Response Team (ICS-CERT) https://ics-cert.us-cert.gov).*

To understand the nature of the device insecurity, it is important to review the history of what vulnerabilities were discovered and what types of devices were affected. A review of the time period 2000 to 2010 reveals that the bulk of discoveries were at the higher levels of the operational network, including control systems trusted to operate plants, transmission systems, oil pipelines, or whatever critical function is in use.

It is not difficult to understand why such systems are frequently found vulnerable. First, many of the systems utilize software packages that can be easily downloaded and worked against. Second, they operate on common hardware and standard operating systems, such as Microsoft Windows. Third, Windows and the components used within those applications are well known to traditionally IT-focused security researchers. There is little need to develop new tools or techniques when those that have long been in place are sufficiently adequate to breach the target's defenses. For example, Stuxnet, the most famous of the industrial compute-based attacks, was initially successful because it was able to exploit a previously unknown vulnerability in Windows.

The ICS vendor community is also lagging behind IT counterparts with regard to security capabilities and practices, as well as cooperation with third-party security researchers. That said, this situation is beginning to get significant industry focus and is improving through a number of recent initiatives designed to formally address security vulnerability and system testing in the industrial environment. While there are some formal standards, such as ISO/IEC 15408 (Common Criteria), ISO/IEC 19790, and a few others, there remain few formal security testing entities. Beyond formal testing, there is little regulatory enforcement of common criteria that address device security testing.

It was not too long ago that the security research community was viewed as a threat, rather than as a valued and often free service to expose potential dangers. While the situation has improved, operational efforts still significantly lag behind IT-based initiatives, such as bug bounty reward programs and advanced vulnerability preparation programs, along the lines of something like the Microsoft Active Protections Program (MAPP). To go a step further, in the industrial realm, there aren't even parallels to the laws that protect individuals' private data. While many states and countries require notification if an individual's personal and financial data is possibly exposed, outside the electrical utility industry, very few laws require the reporting of incidents that may have put lives at risk.

Dependence on External Vendors

While modern IT environments may be outsourcing business operations or relegating certain processing or storage functions to the cloud, it is less common for the original equipment manufacturers of the IT hardware assets to be required to operate the equipment. However, that level of vendor dependence is not uncommon in some industrial spaces.

Direct and on-demand access to critical systems on the plant floor or in the field are sometimes written directly into contracts or are required for valid product warranties. This has clear benefits in many industries as it allows vendors to remotely manage and monitor equipment and to proactively alert the customer if problems are beginning to creep in. While contracts may be written to describe equipment monitoring and management requirements with explicit statements of what type of access is required and under what conditions, they generally fail to address questions of shared liability for security breaches or processes to ensure communication security.

Such vendor dependence and control are not limited to remote access. Onsite management of non-employees that are to be granted compute and network access are also required, but again, control conditions and shared responsibility statements are yet to be observed.

Security Knowledge

In the industrial operations space, the technical investment is primarily in connectivity and compute. It has seen far less investment in security relative to its IT counterpart. According to the research firm Infonetics, the industrial firewall market in 2015 was only approximately 4% the size of the overall firewall market.

Another relevant challenge in terms of OT security expertise is the comparatively higher age of the industrial workforce. According to a study by the US Bureau of Labor, in North America the average age gap between manufacturing workers and other non-farm workers doubled between 2000 and 2012, and the trend shows no sign of reversing. Simultaneously, new connectivity technologies are being introduced in OT industrial environments that require up-to-date skills, such as TCP/IP, Ethernet, and wireless that are quickly replacing serial-based legacy technologies. The rapid expansion of extended communications networks and the need for an industrial controls-aware workforce creates an equally serious gap in security awareness.

This gap in OT security knowledge is actively being addressed. Education for industrial security environments has grown steadily, particularly in the electrical utility space, where regulations such as NERC CIP (CIP 004) and IEC 62351 (01) require ongoing training.

Due to the importance of security in the industrial space, all likely attack surfaces are treated as unsafe. Unfortunately, considering the potential massive public impact of breaching these systems, there remains a healthy paranoia concerning the connection of IT-centric technologies and external connections, despite the massive amount of investment in security in these areas. Bringing industrial networks up to the latest and most secure levels is a slow process due to deep historical cultural and philosophical differences between OT and IT environments.

How IT and OT Security Practices and Systems Vary

The differences between an enterprise IT environment and an industrial-focused OT deployment are important to understand because they have a direct impact on the security practice applied to them. Some of these areas are touched on briefly earlier in this chapter, and they are more explicitly discussed in the following sections.

The Purdue Model for Control Hierarchy

Regardless of where a security threat arises, it must be consistently and unequivocally treated. IT information is typically used to make business decisions, such as those in process optimization, whereas OT information is instead characteristically leveraged to make physical decisions, such as closing a valve, increasing pressure, and so on. Thus, the operational domain must also address physical safety and environmental factors as part of its security strategy—and this is not normally associated with the IT domain. Organizationally, IT and OT teams and tools have been historically separate, but this has begun to change, and they have started to converge, leading to more traditionally IT-centric solutions being introduced to support operational activities. For example, systems such as firewalls and intrusion prevention systems (IPS) are being used in IoT networks.

As the borders between traditionally separate OT and IT domains blur, they must align strategies and work more closely together to ensure end-to-end security. The types of devices that are found in industrial OT environments are typically much more highly optimized for tasks and industrial protocol-specific operation than their IT counterparts. Furthermore, their operational profile differs as well.

Industrial environments consist of both operational and enterprise domains. To understand the security and networking requirements for a control system, the use of a logical framework to describe the basic composition and function is needed. The Purdue Model for Control Hierarchy, introduced in Chapter 2, is the most widely used framework across industrial environments globally and is used in manufacturing, oil and gas, and many other industries. It segments devices and equipment by hierarchical function levels and areas and has been incorporated into the ISA99/IEC 62443 security standard, as shown in Figure 8-3. For additional detail on how the Purdue Model for Control Hierarchy is applied to the manufacturing and oil and gas industries, see Chapter 9, "Manufacturing," and Chapter 10, "Oil and Gas."

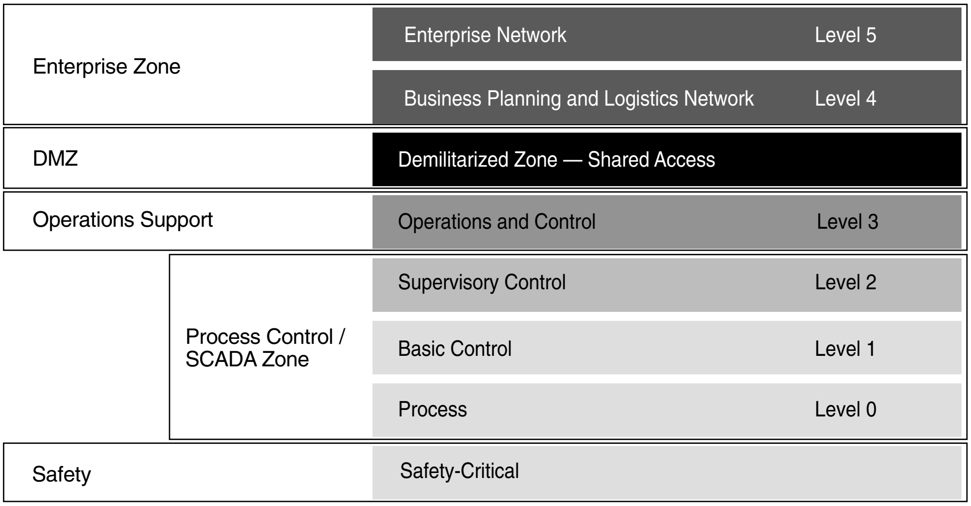


Figure 8-3 *The Logical Framework Based on the Purdue Model for Control Hierarchy*

This model identifies levels of operations and defines each level. The enterprise and operational domains are separated into different zones and kept in strict isolation via an industrial demilitarized zone (DMZ):

- Enterprise zone
  - **Level 5: Enterprise network:** Corporate-level applications such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), document management, and services such as Internet access and VPN entry from the outside world exist at this level.
  - **Level 4: Business planning and logistics network:** The IT services exist at this level and may include scheduling systems, material flow applications, optimization and planning systems, and local IT services such as phone, email, printing, and security monitoring.
- Industrial demilitarized zone
  - **DMZ:** The DMZ provides a buffer zone where services and data can be shared between the operational and enterprise zones. It also allows for easy segmentation of organizational control. By default, no traffic should traverse the DMZ; everything should originate from or terminate on this area.
- Operational zone
  - **Level 3: Operations and control:** This level includes the functions involved in managing the workflows to produce the desired end products and for monitoring and controlling the entire operational system. This could include production scheduling, reliability assurance, systemwide control optimization, security management, network management, and potentially other required IT services, such as DHCP, DNS, and timing.
  - **Level 2: Supervisory control:** This level includes zone control rooms, controller status, control system network/application administration, and other control-related applications, such as human-machine interface (HMI) and historian.
  - **Level 1: Basic control:** At this level, controllers and IEDs, dedicated HMIs, and other applications may talk to each other to run part or all of the control function.
  - **Level 0: Process:** This is where devices such as sensors and actuators and machines such as drives, motors, and robots communicate with controllers or IEDs.
- Safety zone
  - **Safety-critical:** This level includes devices, sensors, and other equipment used to manage the safety functions of the control system.

One of the key advantages of designing an industrial network in structured levels, as with the Purdue model, is that it allows security to be correctly applied at each level and between levels. For example, IT networks typically reside at Levels 4 and 5 and use security principles common to IT networks. The lower levels are where the industrial systems and

IoT networks reside. As shown in Figure 8-3, a DMZ resides between the IT and OT levels. Clearly, to protect the lower industrial layers, security technologies such as firewalls, proxy servers, and IPSs should be used to ensure that only authorized connections from trusted sources on expected ports are being used. At the DMZ, and, in fact, even between the lower levels, industrial firewalls that are capable of understanding the control protocols should be used to ensure the continuous operation of the OT network.

Although security vulnerabilities may potentially exist at each level of the model, it is clear that due to the amount of connectivity and sophistication of devices and systems, the higher levels have a greater chance of incursion due to the wider attack surface. This does not mean that lower levels are not as important from a security perspective; rather, it means that their attack surface is smaller, and if mitigation techniques are implemented properly, there is potentially less impact to the overall system. As shown in Figure 8-4, a review of published vulnerabilities associated with industrial security in 2011 shows that the assets at the higher levels of the framework had more detected vulnerabilities.

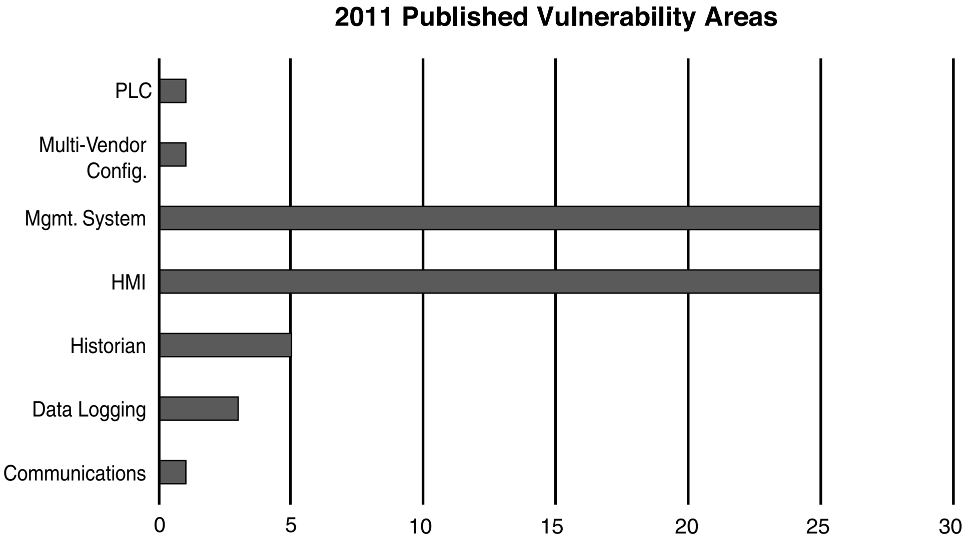


Figure 8-4 *2011 Industrial Security Report of Published Vulnerability Areas (US Industrial Control Systems Cyber Emergency Response Team (ICS-CERT) https://ics-cert.us-cert.gov).*

OT Network Characteristics Impacting Security

While IT and OT networks are beginning to converge, they still maintain many divergent characteristics in terms of how they operate and the traffic they handle. These differences influence how they are treated in the context of a security strategy. For example, compare the nature of how traffic flows across IT and OT networks:

- **IT networks:** In an IT environment, there are many diverse data flows. The communication data flows that emanate from a typical IT endpoint travel relatively far. They frequently traverse the network through layers of switches and eventually make their way to a set of local or remote servers, which they may connect to directly. Data in the form of email, file transfers, or print services will likely all make its way to the central data center, where it is responded to, or triggers actions in more local services, such as a printer. In the case of email or web browsing, the endpoint initiates actions that leave the confines of the enterprise network and potentially travel around the earth.

- **OT networks:** By comparison, in an OT environment (Levels 0–3), there are typically two types of operational traffic. The first is local traffic that may be contained within a specific package or area to provide local monitoring and closed-loop control. This is the traffic that is used for real-time (or near-real-time) processes and does not need to leave the process control levels. The second type of traffic is used for monitoring and control of areas or zones or the overall system. SCADA traffic is a good example of this, where information about remote devices or summary information from a function is shared at a system level so that operators can understand how the overall system, or parts of it, are operating. They can then implement appropriate control commands based on this information.

When IT endpoints communicate, it is typically short and frequent conversations with many connections. The nature of the communications is open, and almost anybody can speak with anybody else, such as with email or browsing. Although there are clearly access controls, most of those controls are at the application level rather than the network level.

In an OT environment, endpoint communication is typically point-to-point, such as a SCADA master to SCADA slave, or uses multicast or broadcast, leveraging a publisher/subscriber type of model. Communication could be TCP or UDP or neither (as in the case of PROFINET, discussed in Chapter 9, “Manufacturing”).

Although network timing in the OT space typically mirrors that of the enterprise with NTP/SNTP used for device clocking against a master time source, a number of use cases require an extremely accurate clock source and extremely accurate time/synchronization distribution, as well as measurable and consistent latency/jitter. Some industrial applications require timing via IEEE 1588, PTP (Precision Time Protocol), so that information from source and destination can be accurately measured and compared at microsecond intervals with communication equipment introducing delays of no more than 50 nanoseconds. Jitter for the sending and receiving of information must also be minimized to ensure correct operation. By way of comparison, in the enterprise space, voice is often considered the highest-priority application, with a typical one-way delay of 150 milliseconds or more. In a number of operational environments for oil and gas, manufacturing, and power utilities, delay must be under 10 microseconds. Security attacks that cause delay, such as denial of service (DoS) attacks, can cause systems to malfunction purely by disrupting the timing mechanism.

IT networks are typically more mature and use up-to-date technologies. These mature modern networking practices are critical to meet the high degree of flexibility required in the IT environment. Virtual networking, virtual workspaces, and virtual servers are commonplace. It is likely that there are a wide variety of device types actively participating in any given network at any one time. Flexible interoperability is thus critical. To achieve interoperability, there is usually minimal proprietary communication activity, and the emphasis is typically on open standards. The movement to IPv6 continues to progress, and higher-order network services, such as quality of service (QoS), are normal as well. Endpoints are not just promiscuous in their communications, but they operate a wide number of applications from a large number of diverse vendors. The open nature of these compute systems means a wide range of protocols are traversing the OT network.

Industrial networks often still rely on serial communication technologies or have mixed serial and Ethernet. This means that not only do many devices lack IP capabilities, but it is not even possible to monitor and secure the serial traffic in the same way you do for IP or Ethernet. In some environments, the network remains very static, meaning a baseline of traffic patterns can be built up and monitored for changes. In static environments, the visibility of devices, protocols, and traffic flows can be managed and secured more easily. However, there is a continued growth of mobile devices and ad hoc connectivity, especially in industries such as transportation and smart cities, as well as a rise in mobile fleet assets across a plethora of other industries. These dynamic and variable networks are much more difficult to baseline, monitor, and secure.

Security Priorities: Integrity, Availability, and Confidentiality

Security priorities are driven by the nature of the assets in each environment. In an IT realm, the most critical element and the target of attacks has been information. In an OT realm, the critical assets are the process participants: workers and equipment. Security priorities diverge based on those differences.

In the IT business world, there are legal, regulatory, and commercial obligations to protect data, especially data of individuals who may or may not be employed by the organization. This emphasis on privacy focuses on the confidentiality, integrity, and availability of the data—not necessarily on a system or a physical asset. The impact of losing a compute device is considered minimal compared to the information that it could hold or provide access to. By way of comparison, in the OT world, losing a device due to a security vulnerability means production stops, and the company cannot perform its basic operation. Loss of information stored on these devices is a lower concern, but there are certainly confidential data sets in the operating environment that may have economic impacts, such as formulations and processes.

In an operational space, the safety and continuity of the process participants is considered the most critical concern. Thus, the goal is the continued uptime of devices and the safety of the people who operate them. The result is to emphasize availability, integrity, and confidentiality. The impact of loss here extends even to loss of life.

Security Focus

Security focus is frequently driven by the history of security impacts that an organization has experienced. In an IT environment, the most painful experiences have typically been intrusion campaigns in which critical data is extracted or corrupted. The result has been a significant investment in capital goods and humanpower to reduce these external threats and minimize potential internal malevolent actors.

In the OT space, the history of loss due to external actors has not been as long, even though the potential for harm on a human scale is clearly significantly higher. The result is that the security events that have been experienced have come more from human error than external attacks. Interest and investment in industrial security have primarily been in the standard access control layers. Where OT has diverged, to some degree, is to emphasize the application layer control between the higher-level controller layer and the receiving operating layer. Later in this chapter you will learn more about the value and risks associated with this approach.

Formal Risk Analysis Structures: OCTAVE and FAIR

Within the industrial environment, there are a number of standards, guidelines, and best practices available to help understand risk and how to mitigate it. IEC 62443 is the most commonly used standard globally across industrial verticals. It consists of a number of parts, including 62443-3-2 for risk assessments, and 62443-3-3 for foundational requirements used to secure the industrial environment from a networking and communications perspective. Also, ISO 27001 is widely used for organizational people, process, and information security management. In addition, the National Institute of Standards and Technology (NIST) provides a series of documents for critical infrastructure, such as the NIST Cybersecurity Framework (CSF). In the utilities domain, the North American Electric Reliability Corporation's (NERC's) Critical Infrastructure Protection (CIP) has legally binding guidelines for North American utilities, and IEC 62351 is the cybersecurity standard for power utilities.

The key for any industrial environment is that it needs to address security holistically and not just focus on technology. It must include people and processes, and it should include all the vendor ecosystem components that make up a control system.

In this section, we present a brief review of two such risk assessment frameworks:

- OCTAVE (Operationally Critical Threat, Asset and Vulnerability Evaluation) from the Software Engineering Institute at Carnegie Mellon University
- FAIR (Factor Analysis of Information Risk) from The Open Group

These two systems work toward establishing a more secure environment but with two different approaches and sets of priorities. Knowledge of the environment is key to determining security risks and plays a key role in driving priorities.

OCTAVE

OCTAVE has undergone multiple iterations. The version this section focuses on is OCTAVE Allegro, which is intended to be a lightweight and less burdensome process to implement. Allegro assumes that a robust security team is not on standby or immediately at the ready to initiate a comprehensive security review. This approach and the assumptions it makes are quite appropriate, given that many operational technology areas are similarly lacking in security-focused human assets. Figure 8-5 illustrates the OCTAVE Allegro steps and phases.

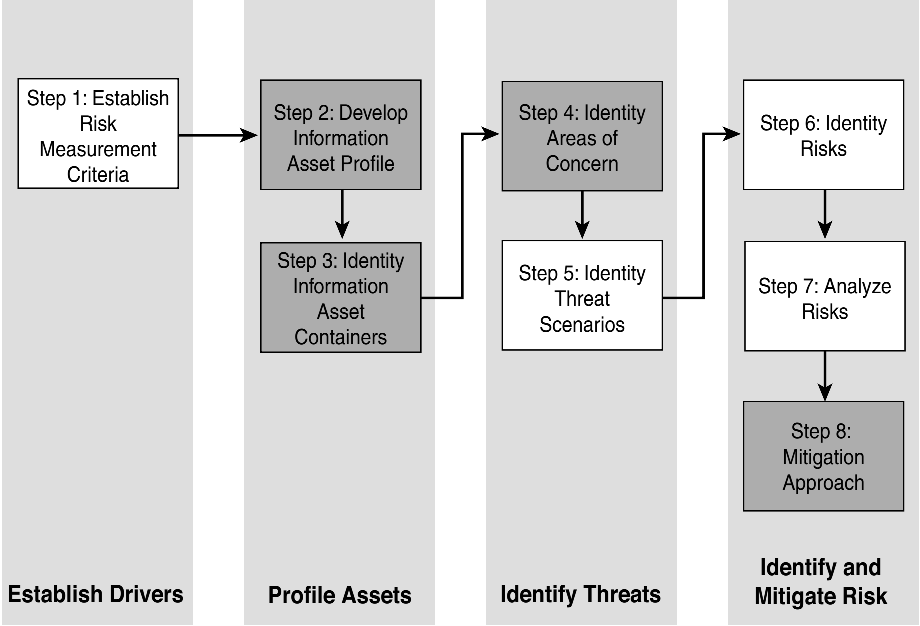

Step 1: Establish Risk Measurement Criteria
Step 2: Develop Information Asset Profile
Step 3: Identity Information Asset Containers
Step 4: Identity Areas of Concern
Step 5: Identity Threat Scenarios
Step 6: Identity Risks
Step 7: Analyze Risks
Step 8: Mitigation Approach
Establish Drivers
Profile Assets
Identify Threats
Identify and Mitigate Risk

Figure 8-5 *OCTAVE Allegro Steps and Phases (see https://blog.compass-security.com/2013/04/lean-risk-assessment-based-on-octave-allegro/).*

The first step of the OCTAVE Allegro methodology is to establish a risk measurement criterion. OCTAVE provides a fairly simple means of doing this with an emphasis on impact, value, and measurement. The point of having a risk measurement criterion is that at any point in the later stages, prioritization can take place against the reference model. (While OCTAVE has more details to contribute, we suggest using the FAIR model, described next, for risk assessment.)

The second step is to develop an information asset profile. This profile is populated with assets, a prioritization of assets, attributes associated with each asset, including owners, custodians, people, explicit security requirements, and technology assets. It is important to stress the importance of process. Certainly, the need to protect information does not disappear, but operational safety and continuity are more critical.

Within this asset profile, process are multiple substages that complete the definition of the assets. Some of these are simply survey and reporting activities, such as identifying the asset and attributes associated with it, such as its owners, custodians, human actors with which it interacts, and the composition of its technology assets. There are, however, judgment-based attributes such as prioritization. Rather than simply assigning an arbitrary ranking, the system calls for a justification of the prioritization. With an understanding of the asset attributes, particularly the technical components, appropriate threat mitigation methods can be applied. With the application of risk assessment, the level of security investment can be aligned with that individual asset.

The third step is to identify information asset containers. Roughly speaking, this is the range of transports and possible locations where the information might reside. This references the compute elements and the networks by which they communicate. However, it can also mean physical manifestations such as hard copy documents or even the people who know the information. Note that the operable target here is information, which includes data from which the information is derived.

In OCTAVE, the emphasis is on the container level rather than the asset level. The value is to reduce potential inhibitors within the container for information operation. In the OT world, the emphasis is on reducing potential inhibitors in the containerized operational space. If there is some attribute of the information that is endemic to it, then the entire container operates with that attribute because the information is the defining element. In some cases this may not be true, even in IT environments. Discrete atomic-level data may become actionable information only if it is seen in the context of the rest of the data. Similarly, operational data taken without knowledge of the rest of the elements may not be of particular value either.

The fourth step is to identify areas of concern. At this point, we depart from a data flow, touch, and attribute focus to one where judgments are made through a mapping of security-related attributes to more business-focused use cases. At this stage, the analyst looks to risk profiles and delves into the previously mentioned risk analysis. It is no longer just facts, but there is also an element of creativity that can factor into the evaluation. History both within and outside the organization can contribute. References to similar operational use cases and incidents of security failures are reasonable associations.

Closely related is the fifth step, where threat scenarios are identified. Threats are broadly (and properly) identified as potential undesirable events. This definition means that results from both malevolent and accidental causes are viable threats. In the context of operational focus, this is a valuable consideration. It is at this point that an explicit identification of actors, motives, and outcomes occurs. These scenarios are described in threat trees to trace the path to undesired outcomes, which, in turn, can be associated with risk metrics.

At the sixth step risks are identified. Within OCTAVE, risk is the possibility of an undesired outcome. This is extended to focus on how the organization is impacted. For more focused analysis, this can be localized, but the potential impact to the organization could extend outside the boundaries of the operation.

The seventh step is risk analysis, with the effort placed on qualitative evaluation of the impacts of the risk. Here the risk measurement criteria defined in the first step are explicitly brought into the process.

Finally, mitigation is applied at the eighth step. There are three outputs or decisions to be taken at this stage. One may be to accept a risk and do nothing, other than document the situation, potential outcomes, and reasons for accepting the risk. The second is to mitigate the risk with whatever control effort is required. By walking back through the threat scenarios to asset profiles, a pairing of compensating controls to mitigate those threat/risk pairings should be discoverable and then implemented. The final possible action is to defer a decision, meaning risk is neither accepted nor mitigated. This may imply further research or activity, but it is not required by the process.

OCTAVE is a balanced information-focused process. What it offers in terms of discipline and largely unconstrained breadth, however, is offset by its lack of security specificity. There is an assumption that beyond these steps are seemingly means of identifying specific mitigations that can be mapped to the threats and risks exposed during the analysis process.

FAIR

FAIR (Factor Analysis of Information Risk) is a technical standard for risk definition from The Open Group. While information security is the focus, much as it is for OCTAVE, FAIR has clear applications within operational technology. Like OCTAVE, it also allows for non-malicious actors as a potential cause for harm, but it goes to greater lengths to emphasize the point. For many operational groups, it is a welcome acknowledgement of existing contingency planning. Unlike with OCTAVE, there is a significant emphasis on naming, with risk taxonomy definition as a very specific target.

FAIR places emphasis on both unambiguous definitions and the idea that risk and associated attributes are measurable. Measurable, quantifiable metrics are a key area of emphasis, which should lend itself well to an operational world with a richness of operational data.

At its base, FAIR has a definition of risk as the probable frequency and probable magnitude of loss. With this definition, a clear hierarchy of sub-elements emerges, with one side of the taxonomy focused on frequency and the other on magnitude.

Loss even frequency is the result of a threat agent acting on an asset with a resulting loss to the organization. This happens with a given frequency called the threat event frequency (TEF), in which a specified time window becomes a probability. There are multiple sub-attributes that define frequency of events, all of which can be understood with some form of measurable metric. Threat event frequencies are applied to a vulnerability. *Vulnerability* here is not necessarily some compute asset weakness, but is more broadly defined as the probability that the targeted asset will fail as a result of the actions applied. There are further sub-attributes here as well.

The other side of the risk taxonomy is the probable loss magnitude (PLM), which begins to quantify the impacts, with the emphasis again being on measurable metrics. The FAIR specification makes it a point to emphasize how ephemeral some of these cost estimates can be, and this may indeed be the case when information security is the target of the discussion. Fortunately for the OT operator, a significant emphasis on operational efficiency and analysis makes understanding and quantifying costs much easier.

FAIR defines six forms of loss, four of them externally focused and two internally focused. Of particular value for operational teams are productivity and replacement loss. Response loss is also reasonably measured, with fines and judgments easy to measure but difficult to predict. Finally, competitive advantage and reputation are the least measurable.

Note The discussion of OCTAVE Allegro and FAIR is meant to give you a grounding in formal risk analysis processes. While there are others, both represent mechanics that can be applied in an OT environment.

The Phased Application of Security in an Operational Environment

It is a security practitioner's goal to safely secure the environment for which he or she is responsible. For an operational technologist, this process is different because the priorities and assets to be protected are highly differentiated from the better-known IT environment. The differences have been discussed at length in this chapter, but many of the processes used by IT security practitioners still have validity and can be used in an OT environment. If there is one key concept to grasp, it is that security for an IoT environment is an ongoing process in which steps forward can be taken, but there is no true finish line.

The following sections present a phased approach to introduce modern network security into largely preexisting legacy industrial networks.

Secured Network Infrastructure and Assets

Given that networks, compute, or operational elements in a typical IoT or industrial system have likely been in place for many years and given that the physical layout largely defines the operational process, this phased approach to introducing modern network security begins with very modest, non-intrusive steps.

As a first step, you need to analyze and secure the basic network design. Most automated process systems or even hierarchical energy distribution systems have a high degree of correlation between the network design and the operational design. It is a basic tenet of ISA99 and IEC 62443 that functions should be segmented into zones (cells) and that communication crossing the boundaries of those zones should be secured and controlled through the concept of conduits. In response to this, it is suggested that a security professional discover the state of his or her network and all communication channels.

Figure 8-6 illustrates inter-level security models and inter-zone conduits in the process control hierarchy.

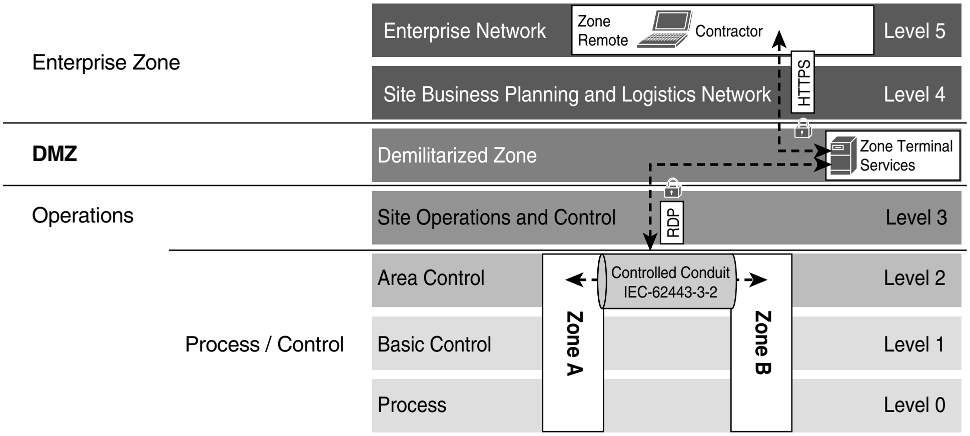


Figure 8-6 *Security Between Levels and Zones in the Process Control Hierarchy Model*

Normal network discovery processes can be highly problematic for older networking equipment. In fact, the discovery process in pursuit of improved safety, security, and operational state can result in degradation of all three. Given that condition, the network discovery process may require manual inspection of physical connections, starting from the highest accessible aggregation point and working all the way down to the last access layer. This discovery activity must include a search for wireless access points. For the sake of risk reduction, any on-wire network mapping should be done passively as much as possible.

It is fair to note that this prescribed process is much more likely to succeed in a smaller confined environment such as a plant floor. In geographically distributed environments, it may not be possible to trace the network, and in such cases, the long-haul connections may not be physical or may be carried by an outside communication provider. For those sections of the operational network, explicit partnering with other entities is required.

A side activity of this network tracing process is to note the connectivity state of the physical connections. This is not just an exercise to see what fiber or cables are in what ports but to observe the use or operational state of other physical connections, such as USB, SD card, alarm channel, serial, or other connections, at each network appliance. For more modern environments where updated networking devices and protocols are used, tools like NetFlow and IPFIX can also be used to discover the network communication paths.

As the network mapping reaches the aggregation point, it is worthwhile to continue to the connected asset level.

Normally, in an IT environment, the very first stage of discovery is focused on assets connected to the network. Assets remain critical, but from an efficiency and criticality perspective, it is generally recommended to find data paths into and between zones (cells) rather than the serial links between devices within a zone. One thing to continually be on the lookout for is the ever-dangerous, unsecured, and often undocumented convenience port.

Any physical port that is not physically locked down or doesn't have an enforceable protection policy is an uncontrolled threat vector.

Once the network is physically mapped, the next step is to perform a connectivity analysis through the switch and router ARP tables and DHCP requests within the network infrastructure. This should help further illuminate connectivity, good or bad, that has occurred. Firewall and network infrastructure data can contribute to understanding what devices are talking to other devices and the traffic paths over which this is done.

At this stage, the network should be reasonably well understood and prepared for secure connectivity.

Modern networking equipment offers a rich set of access control and secured communications capabilities. Starting at the cell/zone level, it is important to ensure that there is a clear ingress/egress aggregation point for each zone. If your communications patterns are well identified, you can apply access control policies to manage who and what can enter those physical portions of the process. If you are not comfortable explicitly controlling the traffic, then begin with alert-only actions. With time, you should be confident enough in your knowledge to apply controls.

At upstream levels, consider traffic controls such as denial of service (DoS) protection, traffic normalization activities, and quality of service (QoS) controls (such as marking and black-holing or rate-limiting scavenger-class traffic). The goal here is to ensure that these aggregated traffic segments are carrying high-priority traffic without impediment.

Network infrastructure should also provide the ability to secure communications between zones via secured conduits (see Figure 8-6). The primary method is encrypted communications in the form of virtual private networks (VPNs). VPNs can come in multiple forms, such as site-to-site, which would be appropriate between a utility substation and a control center, or perhaps in cell-to-cell communications. Remote access controls can be established in more ad hoc situations and utilize the convenience of browser-based VPNs with Secure Sockets Layer (SSL)–based VPNs. If latency concerns are not particularly high, you can use Media Access Control Security (MACSec) hop-by-hop encryption to allow for potential controls and visibility at key junctions.

The next discovery phase should align with the software and configurations of the assets on the network. At this point, the rights and roles of the network administrator may be insufficient to access the required information. Certainly, the network infrastructure and its status are within the network admin's view, but the individual assets likely are not. At this point, organizational cooperation is required for success. For an experienced IT-based network practitioner, this is not an unusual situation. It is very common, especially in larger enterprises, to see a separation of responsibilities and controls between the communications transport and the assets to which they are connected. At the operations level, similar cooperation is required with those responsible for the maintenance of the OT assets.

There are reasonable sources of information describing the configuration state of OT assets. The control systems associated with the processes hold historical data describing what is connected and what those assets are doing. A review of historical data should provide an idea of what assets are present and what operations are being performed on them, and it should identify such things as firmware updates and health status. The volume of data to analyze may be challenging, but if it is organized correctly, it would be valuable for understanding asset operation.

With an initial asset inventory completed, you can initiate a risk analysis based on the network and assets, and determine an initial scope of security needs.

Deploying Dedicated Security Appliances

The next stage is to expand the security footprint with focused security functionality. The goal is to provide visibility, safety, and security for traffic within the network. Visibility provides an understanding of application and communication behavior. With visibility, you can set policy actions that reflect the desired behaviors for inter-zone and conduit security.

While network elements can provide simplified views with connection histories or some kind of flow data, you get a true understanding when you look within the packets on the network. This level of visibility is typically achieved with deep packet inspection (DPI) technologies such as intrusion detection/prevention systems (IDS/IPS). These technologies can be used to detect many kinds of traffic of interest, from simply identifying what applications are speaking, to whether communications are being obfuscated, to whether exploits are targeting vulnerabilities, to passively identifying assets on the network.

With the goal of identifying assets, an IDS/IPS can detect what kind of assets are present on the network. Passive OS identification programs can capture patterns that expose the base operating systems and other applications communicating on the network. The organizationally unique identifier (OUI) in a captured MAC address, which could have come from ARP table exploration, is yet another means of exposure. Coupled with the physical and historical data mentioned before, this is a valuable tool to expand on the asset inventory without having to dangerously or intrusively prod critical systems.

Application-specific protocols are also detectable by IDS/IPS systems. For more IT-like applications, user agents are of value, but traditionally, combinations of port numbers and other protocol differentiators can contribute to identification. Some applications have behaviors that are found only in certain software releases. Knowledge of those differences can help to determine the software version being run on a particular asset.

Within applications and industrial protocols are well-defined commands and, often, associated parameter values. Again, an IDS/IPS can be configured to identify those commands and values to learn what actions are being taken and what associated settings are being changed.

All these actions can be done from a non-intrusive deployment scenario. Modern DPI implementations can work out-of-band from a span or tap. Viewing copies of packets has no impact on traffic performance or latency. It is easily the safest means of getting deep insight into the activities happening on a network.

Visibility and an understanding of network connectivity uncover the information necessary to initiate access control activity. Access control is typically achieved with access control lists (ACLs), which are available on practically all modern network equipment. For improved scalability, however, a dedicated firewall would be preferred. Providing strong segmentation and zone access control is the first step. Access control, however, is not just limited to the typical address and protocol identifiers. Modern firewalls have the ability to discern attributes associated with the user accessing the network, allowing controls to be placed on the "who" element also. In addition, access control can be aligned with applications and application behaviors. Equipped with the right toolset, a modern OT practitioner can ensure that only those operators in a certain user class can initiate any external commands to that particular asset.

Safety is a particular benefit as application controls can be managed at the cell/zone edge through an IDS/IPS. The same technologies that observe the who and what can also manage the values being passed to the target asset. For example, in a manufacturing scenario where a robot operates, there may be an area frequented by workers who are within the potential range of the robot's operation. The range is unique to the physical layout of the cell, and parameter changes could cause physical harm to a plant worker. With an IDS/IPS, the system can detect that a parameter value exceeds the safety range and act accordingly to ensure worker safety.

Safety and security are closely related linguistically (for example, in German, the same word, Sicherheit, can be used for both), but for a security practitioner, security is more commonly associated with threats. Threat identification and protection is a key attribute of IPSs using DPI.

Mature IPSs have thousands of threat identifiers, which address the complete range of asset types where remotely exploitable vulnerabilities are known. In some cases, the nature of the threat identifier is generic enough that it addresses a common technique without having to be associated with a particular application instance of the vulnerability type.

Placement priorities for dedicated security devices vary according to the security practitioner's perception of risk. If visibility is incomplete and concern dictates that further knowledge is necessary prior to creating a proactive defense, the security device should be placed where that gap is perceived. It is important to note that the process of gaining visibility or addressing risk is dynamic. Networks change, and as knowledge is gained, new priorities (either in the form of visible threats or a reduction of gaps) creates new points of emphasis. Given this dynamism, consider the idea that placement of a dedicated security device can change as well. In other words, just because you start with a device in one location does not mean you can't move it later to address security gaps.

Inevitably a decision must be made. Here we discuss some of the relative merits of different placement locations. Placement at the operational cell is likely the most fine-grained deployment scenario. By *fine-grained* we mean that it is the lowest portion of a network that gives network-based access to the lowest level of operational assets. As discussed earlier, the nature of the deployment—out-of-band or in-line—depends on the organization's comfort level for in-line operation and desire to actually exert control. In either case, the industrial security appliance should be attached directly to the switch, which denotes the access point into the cell. This location gives the greatest level of control for safety controls, visibility, and threats. If network design has properly segmented to a single zone entry point, then this is an optimal deployment location. For safety considerations, application control can be exerted to ensure that application changes will not allow for dangerous settings. Threats can be mitigated as they traverse the device, and traffic entering and exiting the cell can be made visible.

A particularly valuable function is enabled if a security device can terminate VPNs in addition to performing deep packet inspection. Secured communication, potentially from a vendor representative outside the organization, can be terminated at the ingress to the device and then inspected. The time cost of the termination would be similar to what would be done on the switch, and then inspection of what that remote user accessing the network is doing is viable. Naturally, any potential threat traffic can be halted as well.

If the zone/cell houses critical infrastructure and remote operation is requisite, a redundant high-availability configuration for both the network and security infrastructure is advised.

For the purposes of pure visibility, hanging off a mirror or span port from the switch would be optimal. For control capabilities, one must be in-line to truly act on undesired traffic. In most cases, the preferred location is upstream of the zone/cell access switch between the aggregation layer and the zone switch. It may be viable to have the security device between the zone assets and the zone access switch as well.

For broader, less detailed levels of control, placement of dedicated security devices upstream of the aggregation switches is the preferred approach. If the network has multiple zones going through the aggregation switch with mostly redundant functionality but with no communication between them, this may be a more efficient point of deployment.

At some point, a functional layer above the lowest zone layer becomes connected to the network, and there should be a device located between those functions and their OT charges in the zones/cells. At that next layer up, there may be HMIs or other lower-level operational tools. For safety considerations, a control point between that layer and the cell is valuable.

At the higher level of the network are a good number of higher-function assets, such as standard network elements (for example, directory servers, network monitoring tools, remote access plus proxy servers, print servers, security control elements). More operationally focused functionality involves elements such as engineering workstations and operations control applications. Depending on the diversity and network topologies at play, these operational structures could be replicated within their own subzones (subnets) at the same level. There may be justification for using a dedicated security device between the subzones, depending on the need to control access, but for the most part, this is a zone that needs controls placed above and below.

Below is where industrial awareness and, potentially, hardware ruggedization is more likely to be needed. With some amount of industrial traffic traversing this layer, a dedicated and security-aware tool would be advisable.

Above this highest level, a dedicated security device with IT-centric threat controls is recommended. If the applications hosted here are similar in nature to those found in IT environments (for example, Windows- or Linux-based applications), this requires common networking infrastructure, web-based access, and so on for proper visibility, control, and protection. Applying such controls to all ingress points (above and below) is important. There should be no assumptions made that an IT-centric threat can only emanate from the IT/enterprise layer above the DMZ. Attackers would not limit themselves to such thinking.

There is evidence that end-of-life OS and software components exist in operational environments. An all-too-common and unfortunate attribute of such systems is that further patching for security vulnerabilities is likely unavailable. To protect those systems after their official end-of-support date, the concept of a "virtual patch" layer may be possible. The idea is that protections for vulnerabilities can be applied through the network path by which these systems communicate. While this is not a substitute for keeping abreast of patching, it may be a mitigation approach that fits your organization's risk acceptance policy.

At the logical edge of the operational space is the DMZ (demilitarized zone)—a security boundary between two diverged compute realms. Assets in this area are meant to bridge communications in a secure fashion between the enterprise's IT realm and the industrial OT realm. Security should be applied both above and below this layer.

Before we leave the second phase of operational security, it is important to reemphasize that security, in whatever location, is an ongoing process. The policies applied and the knowledge gained should never stagnate. Conditions will inevitably change, so security deployments and sometimes networks themselves must change to adapt. Where you place your security enforcement products and the policies they employ must be ready to change with them.

Higher-Order Policy Convergence and Network Monitoring

So far we have focused on very basic concepts that are common and easily implemented by network engineering groups. Finding network professionals with experience performing such functions or even training those without prior experience is not difficult.

Another security practice that adds value to a networked industrial space is convergence, which is the adoption and integration of security across operational boundaries. This means coordinating security on both the IT and OT sides of the organization. Convergence of the IT and OT spaces is merging, or at least there is active coordination across formerly distinct IT and OT boundaries. From a security perspective, the value follows the argument that most new networking and compute technologies coming to the operations space were previously found and established in the IT space. It is expected to also be true that the practices and tools associated with those new technologies are likely to be more mature in the IT space.

There are advanced enterprise-wide practices related to access control, threat detection, and many other security mechanisms that could benefit OT security. As stated earlier, the key is to adjust the approach to fit the target environment.

Several areas are more likely to require some kind of coordination across IT and OT environments. Two such areas are remote access and threat detection. For remote access, most large industrial organizations backhaul communication through the IT network. Some communications, such as email and web browsing, are obvious communication types that are likely to touch shared IT infrastructure. Often vendors or consultants who require some kind of remote access to OT assets also traverse the IT side of the network. Given this, it would be of significant value for an OT security practitioner to coordinate access control policies from the remote initiator across the Internet-facing security layers, through the core network, and to a handoff point at the industrial demarcation and deeper, toward the IoT assets. The use of common access controls and operational conditions eases and protects network assets to a greater degree than having divergent groups creating ad hoc methods. Using location information, participant device security stance, user identity, and access target attributes are all standard functions that modern access policy tools can make use of. Such sophistication is a relatively new practice in industrial environments, and so, if these functions are available, an OT security practitioner would benefit from coordination with his or her IT equivalents.

Network security monitoring (NSM) is a process of finding intruders in a network. It is achieved by collecting and analyzing indicators and warnings to prioritize and investigate incidents with the assumption that there is, in fact, an undesired presence.

The practice of NSM is not new, yet it is not implemented often or thoroughly enough even within reasonably mature and large organizations. There are many reasons for this underutilization, but lack of education and organizational patience are common reasons. To simplify the approach, there is a large amount of readily available data that, if reviewed, would expose the activities of an intruder.

It is important to note that NSM is inherently a process in which discovery occurs through the review of evidence and actions that have already happened. This is not meant to imply that it is a purely postmortem type of activity. If you recognize that intrusion activities are, much like security, an ongoing process, then you see that there is a similar set of stages that an attacker must go through. The tools deployed will slow that process and introduce opportunities to detect and thwart the attacker, but there is rarely a single event that represents an attack in its entirety. NSM is the discipline that will most likely discover the extent of the attack process and, in turn, define the scope for its remediation.

Summary

As industries modernize in pursuit of operational efficiencies, improved safety, and competitive agilities, they must do so securely. Modernization processes frequently initiate greater connectivity in the context of older and highly vulnerable OT assets and processes. Security is a process that must be applied throughout the lifecycle of that change and operation. To achieve security, an organization must be able to define risks and make informed choices about how best to address them.

Fortunately, much of what is available to minimize risks from threats is readily available. Network connectivity can be made secure with the right equipment and policies. Threats from unsafe practices, attacks, and remote access needs can be identified and controlled with dedicated industrial security appliances and practices. With time, there are opportunities to expand risk reduction through convergence and cooperation. Learning from the more extensive and mature security practices and tools in IT environments as well as coordinating layers of defense to protect critical industrial assets are key security enablers for operational environments.

Index

The Cisco Learning Network
Welcome to The Cisco Learning Network