Enterprise Information Protection

by Fred Cohen, Ph.D.

Table of Contents

Front matter

Enterprise Information Protection
is
Copyright © 1977-2008 by Fred Cohen - All Rights Reserved.
ISBN # 1-878109-43-X
Published by Fred Cohen & Associates out of Livermore, CA.
2008-04-28

You may not copy this material or any parts of it without the express written permission of the author.

We are not liable for anything you do with this book or use this book for, including but not limited to choking people with its pages, using the book to trip folks, the kid leaving it on the steps, the pollution it creates when you burn it, or any side effects you may experience in the afterlife from having sworn about certain things I may have said. All of these things; as well as any other things that happen as either a direct or indirect consequence of my having written this book are not my fault, and I take no responsibility and MAKE NO WARRANTY EITHER EXPRESSED OR IMPLIED ABOUT ANYTHING IN THIS BOOK. The law says I have to use large block letters to say that so you are amply warned.

If you are offended by anything I have written, or if you feel your children or parents have been corrupted by it, remember: this warning was given to you, and you could have looked at it before reading the book or buying it.

WARNING! READ THIS PAGE BEFORE BUYING THE BOOK!

That should do it I think...

1 Background and Introduction

Background of the book

This book is really the second edition of several previous books I wrote, all combined into one. These books were created one after the other to form what I then called the CISO ToolKit – aimed at helping CISOs consider things they do – sort of like a notebook to help me and others like me do our tasks better, miss fewer things, and check on ourselves.[1.1]

This book tries to describe the basic structure of information protection and protection programs in enterprises, provide and suggest ways to keep track of what is going on in a real enterprise, and includes questions and endnotes to help make it suitable for use in graduate or advanced undergraduate programs as a text. It is designed to provide clear and concise explanations of key issues in information protection with pictures that allow the material to be presented, referenced, and understood.

Previous books were criticized by the corporate community for their excessive academic influence, while the academic community criticized them for the lack of citations and mathematics. I settled on an approach using endnotes. This approach allows the text to flow without full details or citations while fulfilling the desire to bring additional clarity where necessary, academic propriety where appropriate, and – frankly – thanking those who are responsible for doing the hard work that made this book and its bases possible. We stand on the shoulders of giants, and they should be recognized for their efforts.[1,2]

The endnotes are provided for two purposes; (1) to help flesh out some of the more concise statements that, on their face, may seem simple, but that in fact say a great deal more than what is immediately obvious or apparent, and (2) to provide references, where applicable, to other works that have influenced this work or that are directly cited as sources for select information.

Background of the author

This is not my first book – or even the first one I have written on this subject. In several previous books, detailed information on my background in various subfields are provided, including my experience in information protection, information warfare, digital forensics, computer viruses, deception and counter-deception, and other related fields. Please feel free to read all about me in those books if you have the time or desire.[1,3] For the purposes of this book, I will concentrate on my work related to enterprise information protection architecture and governance, skipping many of the details.

My first real experience in this area was in the 1980s when I started to get involved in businesses and ended up helping to create, build, and run a business that grew to 250 employees in a matter of less than a year. In that business, the Radon testing business, high integrity in analysis results, confidentiality of the information about scientific measurements of individual homes, and availability of systems to take measurements within defined time frames were key factors to success.

If answers were wrong, we would either fail to identify potential sources of cancer or cause people to spend thousands of dollars in needless remediation of non-existent problems, and invalidate scientific research. If we leaked information it could lead to reduction in home values, public scorn, liability, and violate a promise we made to our customers in exchange for using their measurements for scientific research that ultimately changed the way the scientific community understood effects of radiation. Since Radon decays fairly quickly with time and measurement accuracy is closely related to the available quantities of radiation above background levels, loss of measurement capabilities for even a few days could invalidate tens of thousands of results, forcing the business to give free retests and potentially delaying or causing loss of home sales.

As CEO of this small to medium-sized business, I had the ultimate day-to-day decision-making authority, subject to board decisions that were quite often made in the start-up process. Because of the small size, I was also in charge of information technology and implemented most of the controls in place for quite some time. I had essentially all of the risk management responsibility and also architected the information environment, wrote much of the initial code that did calculations, and did lots of other things. For example, I helped design and build the heat pipes that kept a substantial quantity of toluene at the proper temperature to prevent its spontaneous combustion. Availability loss of this system for too long could potentially result in total business loss along with the destruction of a substantial building with people in it.

After my tenure at that business ended (I quit to do other things after diversifying and expanding it into water testing and other similar testing, and the business ultimately failed with the assets sold to another radon testing corporation), I spent many years doing consulting, education, and research for my own and other businesses of all sizes, generally in the area of information protection. This included doing consulting for many of the largest enterprises in the world as well as government agencies, ranging from the extreme technical end of consulting to pure business consulting related to information technology and protection related issues, and on occasion to other business issues.

In the late 1990s I led a research group at Sandia National Laboratories and, along the way, worked on issues related to the year 2000, critical infrastructure protection, information warfare, intelligence, and digital forensics. In the 2000s I worked as an industry analyst for Burton Group for three years helping them to create and define their security and risk management strategies service. After staying for a year longer than originally planned, we parted as friends and I still do consulting through them for their clients as well as through other companies for other companies.

These days, I do research and patent new technologies, do consulting in these areas, work on digital forensics cases, and am starting up a new graduate educational institution, called California Sciences Institute – a non-profit California Public Benefit corporation. And I write books – like this one – to help me keep track of the complex field I work in and to help others learn about it and track their own efforts.

Introduction to the book

This book starts with the structure of information protection at a very high level and rapid pace in Chapter 2, while the graphics provide a lot more detail than the explanations. The goal is to provide a rapid-fire overview of material that a good chief information security officer (CISO) already understands along with simple ways it can be explained at a high level to others.

The rest of the book drills down into the details of the items covered in the overview. The goal is to provide at least two levels of additional detail for everything explained in the Chapter 2. In many cases more levels of detail are provided. This acts as a reference as well as an aide to assuring coverage when reviewing an issue. A more comprehensive set of details relating to that issue can be explored starting with the coverage provided and extending or curtailing as appropriate to the task at hand.

Pictures are used to depict various views of enterprise information protection and to cover the facets described within the book in a more concise fashion. I use many of these pictures in collecting data about clients and describing protection issues to those who want to know what I am doing or why I am doing it.

There are questions at the end of sections, an index at the end of the book, a detailed table of contents, and endnotes at the end of everything else. These are not just to increase the page count and make the book thicker, although they do that. They are designed to provide a useful way to quickly reference things in the book when you are looking for them. I use the book and these end items in discussions all the time. I hope they will be as useful to you as they are to me.

Many readers have commented that there are many complex pictures and diagrams. Perhaps they seem too complex and too busy at first. But I have found them to be very useful in the form they are in because they provide enough drill-down to allow me to apply them while still retaining enough top-level structure to allow overall explanations. The question of how much detail and how readable the pictures are in book form is solved by making all of the graphics available on the all.net Web site.

What people in the information protection business do can be very complex and it usually involves a wide variety of different issues that cross many common boundaries. The pictures are not for teaching novices about the subject matter – they are for helping experts make sure they don't miss things, and to make sure that the relationships at two or three levels of depth are clear. While many people don't, at first, want to see complex pictures, as they gain knowledge in a subject, the simplistic pictures become more of a hindrance than an aide, and instead of saving time and effort, they increase the time and effort. When looking at pictures of all the bones in the human body in an anatomy book, or a payables sheet with aging, or a circuit diagram of a microprocessor or mother board, or the plans of an office building, there are more items shown in less space than in any of the diagrams in this book.

This book is a blue print of information protection for a large complex enterprise. Just as the wiring diagram for a building is complicated because there are a lot of wires doing a lot of things, the blue print of information protection for a global enterprise is complicated because there is a lot to it.

The cover

The cover “art” overviews enterprise information protection. It is the overarching picture to keep in mind when reading this book and thinking about things in the terms the book describes.

Background questions

1. Why are the pictures so complicated in this book?
2. What do you expect to get out of this book?
3. What is your background in information protection?
4. Looking at the endnotes, do you think they will be helpful to your understanding?
5. Looking at the extended table of contents, can you easily find things you are looking for within the book?
6. Go to http://all.net/, look at the security architecture picture, and drill down into the details by using the clickable diagram. Does this provide better pictures than the ones in the book?

2 Enterprise information protection

Enterprise information protection is formed by a combination of governance, activities, and technologies. Enterprise information protection governance has the same basic principles and operates within the same basic structures as other types of enterprise governance. But it has significant unique content, and requires individuals with specific skills and influence in order to be effective.

A systematic comprehensive approach

The systematic comprehensive information protection program ultimately starts with how the business works and ends with assuring proper protection of content and its business utility.[2.1] Oversight defines duties to protect, risk management turns these duties into decisions about risk acceptance, transfer, avoidance, and mitigation, and identifies what to protect and how well. Executive security management then figures out how to protect and uses power and influence within organizations to provide control.

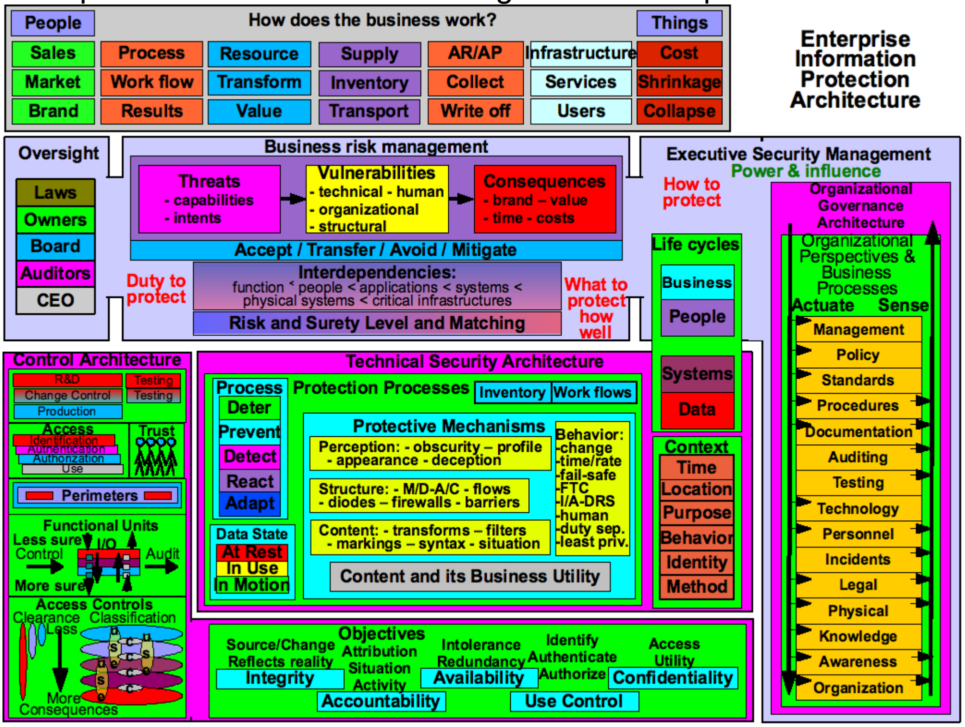


Figure 2-1 – Enterprise Information Protection Governance Model

Organizational issues and business processes drive control architecture[2.2] and interact with technical security architecture to affect the protection processes. These processes ultimately control protective mechanisms that interact directly with content and its business utility to assure that risk is adequately controlled for the needs of the organization.

The architectural model

Figure 2-1 depicts elements of enterprise information protection architecture and how they interact with each other. The presentation here is slanted toward a corporate view in terms of the usage, but essentially all elements are always present.

At the top is the notion of how the "business" works. At a detailed level, this may be codified in terms of process diagrams and associated details such as timeliness requirements, business consequences of information and information technology failures of different sorts, internal and external interdependencies, and so forth. At a higher level it is divided into different common functions, such as sales, marketing, and brand, resources that get transformed and produce value, and so forth. These comprise the basic functions of the organization and the foundation for analysis of the value and import of its function or utility.

Oversight comes from laws, owners, the board of directors or a similar entity, auditors, and the chief executive officer. It produces a set of duties to protect that include legal and regulatory duties, contractual duties, and self-imposed duties. Oversight is also tasked with responsibility for making certain that the duties imposed are carried out and, typically, for making decisions that affect the entire enterprise.

The business risk management function seeks to transform the duties to protect into a set of identified things to protect and surety levels associated with that protection. Surety should be matched to the consequences associated with failures, taking into account the complex nature of these issues. As a side effect of this process, understanding of risks in the form of threats, vulnerabilities, and consequences; event sequences that could induce potentially serious negative consequences; interdependencies and risk aggregation issues; decisions about risk acceptance, avoidance, transfer, and mitigation; and notions of acceptable residual risk are provided to enterprise information protection management for their use and to oversight for their approval.

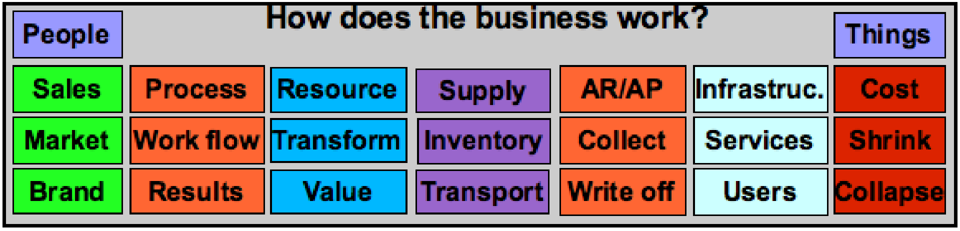
Enterprise information protection management transforms the duty to protect, what to protect, and how well; and the other outcomes of oversight and risk management process; into the actions taken by the organization to implement protection. This is done through the use of power and influence. The Chief Information Security Officer (CISO) or other responsible party tasked with these issues typically has little budget, but their position and standing should provide them with the necessary influence to get the job done, if they know how to apply that influence effectively. Specifically, they should have positional power that grants them access to information required in order to get feedback from the organizational processes they influence and adequate influence to adapt those processes to meet the needs of the organization. If these conditions are not met then the program will fail and the enterprise will suffer the consequences.

The enterprise operates protection through the creation, operation, and adaptation of a control architecture. The control architecture includes structural mechanisms that obtain security objectives through access control models, functional units, perimeters, mechanisms using identification, authentication, and authorization to facilitate use, change control, and other non-architectural mechanisms for specific situations.

The technical security architecture implements technical controls by defining protection processes in the form of defensive processes associated with data states and contexts over life cycles of systems and data; and managing the inventory under control through work flow controls so as to direct, observe, and adapt the protective mechanisms. Those protective mechanisms come in the form of perception, structure, content, and behavior controls that directly contact or protect the content and assure its ongoing business utility.[2.3]

Business modeling

In order to be a useful part of a business, information protection has to meaningfully address business issues. Because the function of information and information technology in a business is to help the business function, it is necessary to understand and describe how the business works to put and keep information protection in context. A simplified view of these issues is shown in Figure 2-2.



How does the business work?

People						Things
Sales	Process	Resource	Supply	AR/AP	Infrastruc.	Cost
Market	Work flow	Transform	Inventory	Collect	Services	Shrink
Brand	Results	Value	Transport	Write off	Users	Collapse

Figure 2-2 – Business Modeling Overview

Different businesses work in different ways. They all have people and things and some sort of marketing and sales function that ties to reputation and good will, often codified in the term “brand”. Businesses have processes that involve work flows to produce results. Most businesses take some sort of resources and transform them to produce value. Most businesses use supplies, have inventory, and transport goods or services through some media. Businesses have accounts receivable and payable, a collection and payment process, and a write off process that form the basis for accounting. Information technology is an example of a service based on an infrastructure provided to users. Businesses also have cost and shrinkage associated with inventory and can collapse if the weights of costs and shrinkage are too high.

While this is obviously a simplification of businesses and how they operate, it can be used as a basis for understanding businesses in terms that allow value and import of business function and process to be clarified. From the standpoint of information protection, this is the key to making sense of business process as it interacts with information technology. Process descriptions and diagrams can be formed to show how a business works, and the interactions with information technology can be mapped through the model to understand how protection failures induce business consequences.

How information supports the business

Information can be thought of as everything of value to a business that is not codified in terms of cash or physical assets. As Jim Schweitzer says in “Protecting Business Information: A Manager's Guide”:[2.7]

> “*... if all information about how to run a business were to be lost, the residual value would probably be the selling price of plants and equipment.*”

While today's world might seem to many to be very different from when Jim wrote his book, this point is still largely true, as is most of the rest of what he says in his book.

Without certain content, no business can function, and protection (assuring the utility) of that content is vital to the success and survival of every business. While much of today's information is codified in information technology, much of it still is not. Much of it is in the minds of people and much of it is on pieces of paper. Regardless of the form that the information takes, its integrity, availability, confidentiality, control over its use, and accountability for it are vital to success.

The questions that have to be answered in order to make reasonable and prudent business decisions with regard to information protection start at understanding:

- What is the content that is important to the business?
- Why is that content important?
- What bad things happen how soon when protection fails?
- Where and in what form is this content?

Unless and until this is understood, efforts to achieve protection are bound to be misdirected and wasteful, and will likely miss the most important things.

The last question is used to differentiate between two large areas of coverage. One way to deal with consequences is, in some cases, to have the information in multiple or different places and forms. The key differentiator in most cases is whether the information lies in paper, fiche, or some other written form; in the minds of people; or in automated information technologies.

How information technology supports business

To the extent that business functions are dependent on automated information technology, failures in that technology can produce business consequences. Failures typically involve loss of integrity, availability, confidentiality, use control, or accountability, at least in the model used in this book. Business consequences depend on the specifics of the business and can range from negligible to business collapse, and beyond if the business, through the society in which it operates, causes indirect effects. These social consequences can also have indirect effects on the business, perhaps causing further business consequences. No business is an island.

Linkage between the model and technology

The purpose of business modeling in the context of this book is to understand and codify the business consequences of information protection failures; and to map those failures and consequences into the information technologies associated with them. At the level of business modeling we are discussing in this book, we look at the as-is state and do not seek to model future business scenarios or approaches, as is done in other sorts of business modeling. The other sorts of business modeling might be well informed to apply the information protection business models to their future scenarios in order to understand the protection implications of those changes and futures.

While most information technology discussions deal with automated information systems, in the general sense, the things discussed in this book and enterprise information protection also deal with paper, FAX machines and other similar technologies, people and other animals, social interactions, telephone calls, and interactions between these things, including without limit scanners, instant messages, mobile computing and communications, and so forth.

The links between a business and information and technology are all within the scope of enterprise information protection and of this book, and business modeling for information protection must encompass this scope if it is to be effective.

Oversight

Duty to protect comes from legal and regulatory mandates, contractual obligations, fiduciary duties to shareholders to retain and grow their value, and self-imposed policies. High-level decision-makers make business decisions that can end their careers if they take too much risk and calamity comes, and that can end their careers if they spend too much time, money, effort, or good will mitigating risk, and calamity never appears averted by it.

In practice, the Chief Executive Officer (CEO) is the person who most sets the direction of the business and ultimately defines the duties to protect. While the board has to approve policies, auditors have to review what goes on, owners express their views periodically, and legal mandates are, in many cases, forcing. The CEO, or in some cases the Chief Operating Officer (COO), has day-to-day responsibility for running the enterprise. For the purposes of this book, we will assume an operational CEO, as opposed to a COO in charge of day-to-day operations (inward facing) with a CEO for external and upward interactions (outward facing) is the operative top-level decision maker.

Because of the key role of the CEO, the CEO usually gets paid more than any other executive, has more liability, for example through Sarbanes Oxley Act requirements, and has direct responsibility for enterprise business risks when they reach the level where other executives dare not or are not permitted to go.

When I was President (and effectively CEO) of the Radon Project, I had this responsibility, as I do for Fred Cohen & Associates and for California Sciences Institute. This means that I am personally responsible for making policy, subject to approval by the board, and for making risk management decisions, subject to override by the board. When I make those decisions, the entire business is directly impacted in many ways. It is my duty to decide what is worth protecting and on what basis to make the judgment calls that are inevitable for any enterprise. It is the duty of those who work for me to help me make good decisions by getting me good information, and to carry out the decisions that I make. I am overseen by the board and others, and I oversee the decisions that others make. That's what oversight is all about.

Business risk management

For enterprise information protection purposes, the risks that are being managed are business risks. That is, there are potentially serious negative consequences associated with decisions about information protection, and those consequences are typically considered in terms of the business. There are also personal consequences in making decisions about what to do and not to do. For example; damage to the reputations of those who work at an institution, damage to the good name of the executives and board members, and the potential for going to jail if laws are broken or regulations are not followed.. These consequences are not typically formalized, even if they do play into the process.

The risk management process and the risk management program that operates the process, if working properly, track and transform the duty to protect into decisions about what to protect and hew well. To do these processes well, it is best to start with the business model.

Ideally, based on the business model and thresholds on risk tolerance and decision-making identified by top management, a business risk management process (1) considers the potentially serious negative consequences associated with failures to assure the utility of content and determines which of them are acceptable, transferable, avoidable, and mitigable, and (2) determines a mixed strategy to optimally accept, transfer, avoid, and mitigate risks. To do this process properly for information and information technology risks, interdependencies, risk aggregations, and matching of surety to risk are normally applied.

The analysis generally starts with the business model results, perhaps in the form of lists of potentially serious negative consequences (by serious, I mean in excess of risk acceptance thresholds), the association of those consequences to protection failures, and the association of those failures with content and systems. Once direct causes of consequences are identified, dependencies of those causes, be they information, systems, people, or other dependencies, are identified. The consequences and indirect dependencies are then analyzed for risk aggregation by identifying common failure modes and causes.

If the consequences are high enough so that a system, content, person, or something else (the *item under consideration*) warrants further investigation, then threats (people, groups, and nature) that have capabilities and intents (except for nature) are analyzed with respect to that item to identify what is commonly called the "design basis threat". This is the threat set used as a basis for the design of protection. The protection designer considers the vulnerabilities of the set of items with respect to the threat set to select protective measures with the goal of minimizing cost plus loss. I will call this the *theoretical view of business risk management* because, while there are a lot of approaches to minimization, the reality is that optimal protection is not understood today and may never be understood or understandable. Figure 2-3 shows this view.

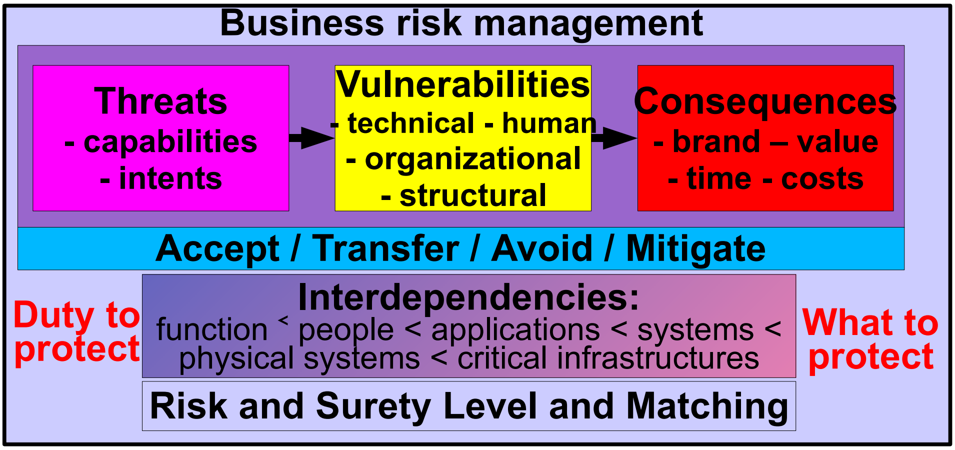


Figure 2-3 – The risk management landscape simplified

In practice, to actually do all of this analysis for every item and get accuracy that is good enough to make optimizing decisions, adds a lot to the cost of the risk management process that has no direct benefit to the business other than the benefit of better informed decisions. If the marginal value of better informed decision doesn't exceed the marginal cost of being better informed, it's better to make a less well informed decision.

Figure 2-4 shows the overall landscape of practical risk management including the process used to reduce analytical costs. The sub-figure at the bottom left addresses the risk management process in light of these limitations.

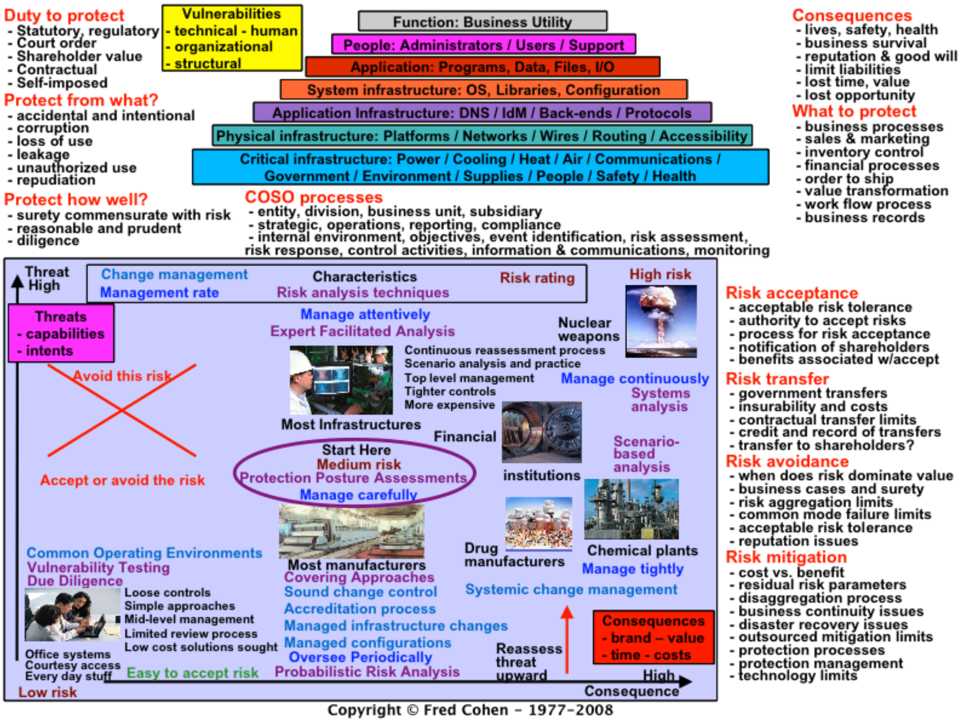


Figure 2-4 – The risk management landscape simplified

To keep the costs of the risk management process within reason, a periodic process usually called an information protection posture assessment (IPPA)[2.6] is typically undertaken. An IPPA starts by creating a nominal business model or, if one exists, using it. Based on the business model, an IPPA uses experts to identify classes of event sequences with potentially serious negative consequences and places them within a two dimensional space. This 2-D space divides event sequences into low, medium, and high consequences on one dimension, and low, medium, and high threats in the other dimension.

Different methodologies for managing risks apply to different places in this space, ranging from due diligence approaches for low threats and low consequences to methods far more comprehensive than an IPPA for high consequences and high threats. Low consequence situations with high threats represent risks that should be avoided because they will cost more to protect then they are worth. As threats get to the medium level, the risk should either be accepted or avoided because it is not worth mitigating at that consequence level; however, transfer may be achievable in that part of the space in some cases. If consequences are considered high and threats are considered low, the threat or consequences assessment should be revisited and it should be assumed that threats are higher than previously identified. The chart also maps management process rate, risk assessment process, change management requirements, risk rating, and other characteristics.

Interdependencies and risk aggregation

The interdependency pyramid in Figure 2-4 shows how business utility depends on complex chains of interdependent information technologies and supporting infrastructures. In order for information technology to operate, all of the dependencies must operate to the level required to service the business function. The technology portion of this issue is increasingly understood by information technology experts, but they often ignore underlying infrastructures required for these functions to operate. In using the interdependency viewpoint, many implicit assumptions are made explicit, leading to more detailed consideration and analysis where appropriate.

Enterprises cannot control all of their dependencies, but they can use redundancy and diversification to reduce the criticality of any individual dependency and thus reduce the aggregation of risk associated with these components. Internationalization, for example, reduces the aggregated risk caused by dependency on any one government being stable, while using physically diverse locations for data centers reduces common mode failures associated with dependencies on critical infrastructure elements.

But how much is enough? The duty to protect

A fundamental question that every enterprise must answer is: How much protection is enough?

- Do we need three redundant data centers?
- Do we need to diversify dependency on operating systems?
- Do we need protection at every layer in the technology picture?
- Do we need to have all of those protective measures?

The answers to these sorts of questions come from a combination of the duty to protect and risk management decisions. Duty to protect comes from legal and regulatory mandates, contractual obligations, fiduciary duties to shareholders to retain and grow their value, and self-imposed policies. High-level decision-makers make business decisions that can end their career if they take too much risk or spend too much reducing risk. That's why they have to make the decisions and manage the risks. However; because different levels of management are granted different decision-making powers, decisions must be made by decision makers at a level suitable to the potentially serious negative consequences involved. Without the information on these consequences, an enterprise cannot determine who has to make what decision, and in many cases, this has led to disaster for top management.[2.4]

Governance, power, and influence

In order to affect desired protections to the level of assurance desired for content and its utility, successful enterprises create governance structures.[2.5] These structures include actuators that cause things to happen, sensors that measure what is going on, and control mechanisms that use power and influence to actuate, sense, and ultimately, control the protection program and the processes it involves. The actuators, sensors, and control mechanisms are combinations of people, processes, and things.

There is usually an individual in charge of the overall information protection program, and we will call that individual the Chief Information Security Officer (CISO). In order for the protection program to be effective, the CISO has to have (1) the power and influence within the enterprise to effectively control the protection program and process, (2) the information and access to find out what is going on within the enterprise, and (3) the knowledge and skills necessary to understand and apply the actuators effectively to get the process and program to meet the duties to protect. Many enterprises have high cost plus loss because top management fails to: (1) understand the role of the CISO, (2) place the CISO properly in governance, (3) provide adequate power and influence for the CISO, or (4) grant the CISO adequate access to information.

Governance typically involves the creation of enterprise controls affecting management, policy, control standards, procedures, documentation, audits, tests, physical and informational technologies, personnel, incident handling, legal issues, knowledge and awareness, and organizational issues. These must cover the life cycles of business, people, systems, and data, and operate and support the control architecture and technical security architecture. It feeds information to and gets guidance from risk management process and oversight, and is ultimately responsible for assuring the utility of content to the business. The CISO is the individual responsible for seeing to it that all of this happens in a reasoned and coordinated fashion.

To be successful, the role of the CISO has to have visibility into and good communications with the HR and legal departments, the CFO and CEO, the CIO and those that work in information technology, facilities management and physical security functions, and business unit owners. The CISO typically interfaces to external law enforcement, legal, investigative, and governmental bodies, works within the greater security community to keep up to date, and interacts with others in that community with whom information technology must interact in the global era of the information age. Without all of these touch points, the CISO will fail to meet the obligations of the role and end up not fulfilling the needs of the enterprise to assure the utility of content.

Because of the enormous scope of this role, placement within the enterprise hierarchy becomes problematic, and in many cases, poor positioning of the CISO leads to excessive cost and loss.

Control architecture

Control architecture may be the most complex thing to understand about enterprise information protection because it is so ephemeral and yet so critical. Control architecture goes directly to how the enterprise thinks about and acts on information protection issues. It may seem like a list of standard concepts from an introductory computer security text, but it really forms the foundations of the field, and the field continues to be rocked by the fact that these foundations are not as well understood or solid in today's environment as most people in the field assume them to be.

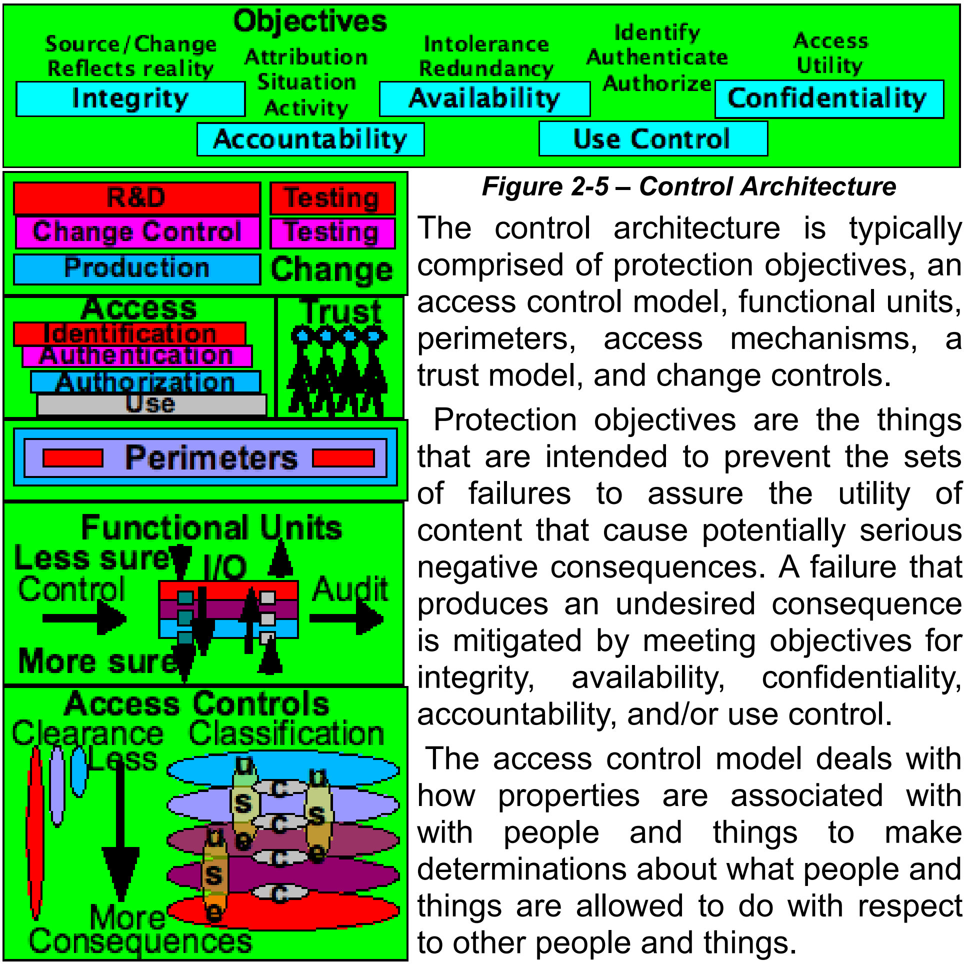


Figure 2-5 – Control Architecture

The control architecture is typically comprised of protection objectives, an access control model, functional units, perimeters, access mechanisms, a trust model, and change controls.

Protection objectives are the things that are intended to prevent the sets of failures to assure the utility of content that cause potentially serious negative consequences. A failure that produces an undesired consequence is mitigated by meeting objectives for integrity, availability, confidentiality, accountability, and/or use control.

The access control model deals with how properties are associated with with people and things to make determinations about what people and things are allowed to do with respect to other people and things.

Functional units take in control requirements, put out audit trails, and implement the mechanisms that assure the utility of content. For example, an intrusion detector, a software component, a guard at a door, a user, and an operating system are functional units that each play a role in providing assurance. Functional units work together to achieve protection and are coordinated for effect.

Perimeters surround or otherwise deal with content to reduce the complexity of control. They are typically semi-permeable with the goal of limiting the things that can affect or be affected by content and the mechanisms that act on content. Functional units are the mechanisms that form and permeate the perimeters.

Access is necessary in order to gain utility from content but it is also the thing that causes utility to be lost. Access is necessary in order to permeate the perimeters. If legitimate access is low friction, it enhances the utility of content, while if it is high friction, it impedes that utility. If illegitimate access is low friction, it reduces the utility of content, while if it is high friction, it enhances the utility of content. Access mechanisms have the task of making legitimate access low friction and the illegitimate access high friction. Typically access includes identification, authentication, authorization, and use controls.

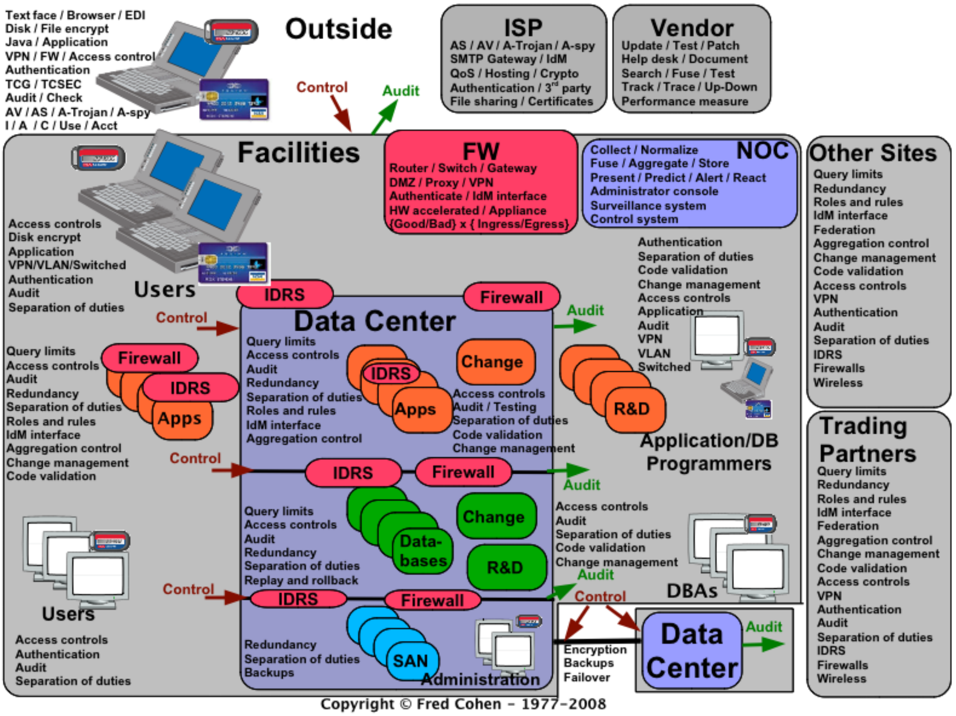
Trust models are usually implicit. Trust is, in essence, the capacity to be harmed by someone or something else. Because this capacity is generally transitive, trust includes interdependencies. When trust exceeds trustworthiness, excessive losses are likely, while when trustworthiness exceeds trust, utility is not fully gained. It is hard to work with others if you do not trust them.[2.8]

Change control is the means by which en enterprise can maintain assurance of the utility of content while allowing for changes over time. Uncontrolled change leads to loss of utility because of errors and omissions associated with the changes and because of the nature of the world in which things decay with time if not controlled. Typically, there is a research and development function where changes are proposed, implemented, and made to work on simulated data in a simulated environment, a testing and change control function where changes are verified as meeting control requirements, and a production environment in which changes are strictly limited and invoked when appropriate.

As a reminder, the control architecture is not the implementation of things that carry out these controls. Rather it is a model of what the controls are, how they work, and how they interact to assure the utility of content. As a model, it is sometimes hard to convince people that such a thing as a control architecture really even exists. And yet, people make assumptions about the control architecture all the time, these assumptions are often wrong, and these wrong assumptions lead to mismatches in implementation and failures in execution. For that reason, a documented and well defined control architecture is important to enterprise protection program success.

Technical security architecture

Figure 2-6 illustrates the logical locations of many of the typical protection mechanisms in an enterprise application architecture.



Copyright © Fred Cohen – 1977–2008

Figure 2-6 – Enterprise Technical Security Architecture Template

In this depiction, users with different systems and protection capabilities interact with applications either locally or remotely. They interact with infrastructure and application elements within facilities perimeters. Within these facilities there are typically physical and logical zones, often including a data center for high-valued information assets and links to other data centers for resiliency and access to additional capabilities. Control and audit paths exist throughout. Internal users use applications for business functions. Application programmers and database administrators do research and development and use change control mechanisms to alter applications, databases, and storage area networks resulting in interfaces, analytical processes, and storage and retrieval associated with applications. Many technologies are associated with information protection throughout this process.

The selection and implementation of protection technologies is a major facet of the risk mitigation process associated with these systems. Without these protection technologies, business functions would be subject to a wide range of attacks. There would be regular failures that would reduce business value to the point where information technologies would have little or no real utility because the cost of failures would outweigh the benefits of use.

On the other hand, when these technologies run amok, costs soar and the utility of content falls because operational complexity goes up, it becomes harder for workers to get their jobs done because they are hindered by all of the protection mechanisms, and the mechanisms fail more often because there are more of them.

Most enterprises have many instances of portions of this set of technologies implementing this or a similar architectural structure. The many different components involved and their different locations within the technology implementation should help to clarify the complexity of selecting from among alternative sets of protection mechanisms. As ultimately implemented, different enterprises use different components in different ways and places to meet different protective needs. And they do so differently for different content and business functions because of the different consequences associated with failures and costs associated with protection. One size does not fit all.

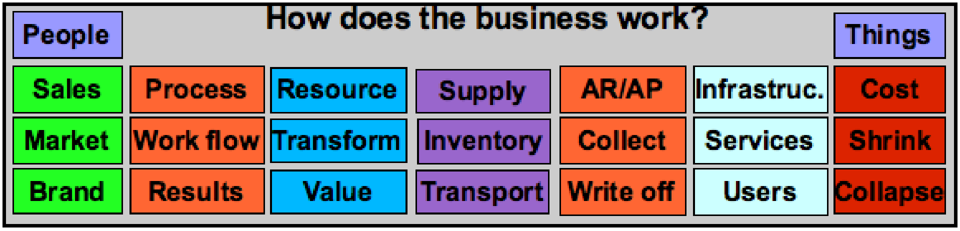
If properly controlled and managed, technologies such as these in an organized structure such as this, implement elements of the control architecture so that the right sets of technologies are in the right places to assure the utility of content to the desired level of certainty against the identified threats. If this is not done through a systematic approach, the result will be a mix of many technologies not matched to the needs, ongoing incidents that ultimately increase costs and losses, and a lack of improvement over time.

Overview questions

1. Why is the enterprise information protection governance picture so complicated?
2. Why is a business model required? Doesn't the domain knowledge of the people in the workplace act as just such a model?
3. How does an enterprise keep track of all the laws in all the countries all over the world and keep them all straight?
4. Can't we use standard probabilistic risk analysis techniques instead of this complicated approach to risk management?
5. Is the CISO supposed to be all things to all people?
6. How can one person understand all of the technology, models, management processes and practices, analysis, oversight requirements and how the business work?
7. Why not just use the CIA model instead of the complicated set of objectives defined here?
8. Why do we need an abstract control architecture level when it is almost never any different than the standard stuff used by everyone already?
9. Isn't security architecture really just the technical part and management responsible for all of the other things identified here?
10. What is this thing called cost plus loss and why can't we know exact numbers for it in advance?
11. What is the fundamental objective of an enterprise information protection program in simple terms?
12. Why are there so many security technologies?
13. Isn't the reliance on security mechanisms outside of the enterprise potentially hazardous to the enterprise?

3 How business works and is modeled

Each "business" is unique in what it does, and yet businesses share some things with each other. Figure 3-1 is intended to start the thinking about what should be modeled within an enterprise as part of the business model used for information protection. But it is, by no means, either comprehensive or definitive.[3.1]



How does the business work?

People						Things
Sales	Process	Resource	Supply	AR/AP	Infrastruc.	Cost
Market	Work flow	Transform	Inventory	Collect	Services	Shrink
Brand	Results	Value	Transport	Write off	Users	Collapse

Figure 3-1 How does the business work?

For example, all businesses involve people and things.

- **People** have to be dealt with in terms of their value in doing things and have to be paid in order to keep working.
- **Things** have inherent value, are inventoried and tracked, and get bought, sold, lost, and stolen.

Because most businesses deal in financial currency, this is certainly an important element of the business modeling process, but the value of most businesses is an order of magnitude or more higher than the inventory value of its assets. This difference is, in one form or another, the information value of the enterprise. Enterprises also value different things. For example, educational institutions are generally non-profit and their main output is graduating students with life-long knowledge that will help them live better and help society prosper. Military enterprises produce the force needed to help exert influence through direct application of power, the potential for force that deters conflicts, and people and skill sets that benefit society as a whole, but they can also produce devastation and large-scale loss of life, liberty, health, and property. Business models are unique to each enterprise and, while some commonalities exist, they are not sufficient to allow a single model to be built today to reflect everything we need to do or know for every enterprise. Like all of information protection, business modeling is something you do, not something you buy.[2.6.1]

Things to consider in the business model

Most businesses can be understood at some level in terms of some common issues outlined here.

Sales, Market, Brand: Brand is a reputational element of the information value of a business and represents a critical factor in sales. Information protection failures tend to harm brand, but claims of security rarely enhance brand substantially. Brand is vital to generation of leads, sales, and ease of success in business. Marketing and the markets that a business operate in dictate to a large extent the aspects of information protection that apply and the tolerance for risk and need for protection. Sales are more directly related to income. All of these also involve business processes that are key to success. Failures in these processes lead to release of critical competitive information, like pricing or customer details, incorrect pricing, inability to process orders, etc.. Any of these may be catastrophic to some businesses.

Process, Work Flow, Results: Business processes are critical to business survival and increasingly they are highly automated. Failed work flows can be highly destructive and cause subtle effects like the ability for unauthorized individuals to cause unauthorized changes to business processes, grant themselves access or monies, disrupt operations, destroy logistics, and otherwise disrupt business operations.

Resources, Transforms, Value: Resources are transformed into value through processes. For example, land is transformed into gold through extraction processes while chemicals are transformed into medicines through chemical processes and raw data is transformed into competitive intelligence through analytical processes. These processes are fundamental to how many businesses operate and failures in theses processes lead to failures in the ability of the enterprise to produce value.

Supply, Inventory, Transport: Many enterprises take supplies of some sort and move them from place to place in order to produce value. Wholesalers and retailers move supplies from suppliers through warehouses and storefronts into consumers or customers while many companies have internal logistics processes that support their operations in one way or another. Disruptions in the supply and logistics process can cause anything from military campaigns to businesses to fall apart.

AR/AP, Collections, Write-offs: With the exception of purely cash businesses, all businesses have accounts payable and receivable, collection processes, and write-offs. These processes are critical to cash flow and business operations as well as profitability and customer relations. Failures in these processes can cause businesses to lose the confidence of their customers, to offend customers, to be stolen from in large quantity, and to be unable to meet payroll or other obligations and go bankrupt. Other elements of the financial systems of businesses are also important in much the same way and are subject to malicious attack for their direct financial value.

Infrastructures, Services, Users: Infrastructure is used in conjunction with services and applications to meet the desires and needs of users. The value of the infrastructure comes in the utility of the services provided to users. If infrastructures or the services they support fail, the harm is in reduction of business utility. These services also support content that may have inherent value, lose value with exposure or time, or otherwise be affected by failures in protection. At the same time the utility is dictated by the ability to use these services.

Cost, Shrinkage, Collapse: Costs and changes in costs and cost structure, shrinkage (loss and theft of inventory), and ultimately collapse of markets or businesses effect enterprises in a wide range of ways.

These and other business functions can be codified in terms of business process diagrams. The elements of the processes diagrams can be associated with failure conditions producing losses as a function of the durations of the failures. Information technology and its role in supporting these business processes can be codified by indicating which processes technology interacts with and how losses of integrity, availability, confidentiality, use control, and accountability can impact those processes. These then are the depictions of the business that help to understand information and information technology related risks from a business perspective.

What's in the business model?

At a simplistic level, a business model for enterprise information protection can help to clarify and codify the answers to some fairly simple questions; What does the business do? How does the business do it? How does the business interact with information? What are the business implications of failures? How does the protection program mitigate the failures with potentially serious negative consequences? Does that mitigation cover what has to be covered?

Every enterprise uses some kind of business model to do this, but few formalize it or really even understand that it is there. The model today is most often in the heads of select individuals, it might include some spread sheets, databases, and text files in computers, and it might be partially on pieces of paper. But because the business model is critical to risk management, this informality leads to many failures in the risk management process leading to serious negative business consequences.

The business models around today typically contain some set of business functions that are considered to be important for one reason or another and some sort of valuation associated with those business functions and their loss to the business. Different people in the enterprise have different models in their minds and, as a result, there is often a mismatch between what management actually values and what technologists think management values. I often hear expressions from workers that indicate that if a particular function fails, the business will disappear by the next morning. Then when I talk to top management, they tell me that it's not really true. I end up discussing this with many different people to try to put together the elephant from the depictions given by the blind people that only “see” the things they happen to touch. [3.2]

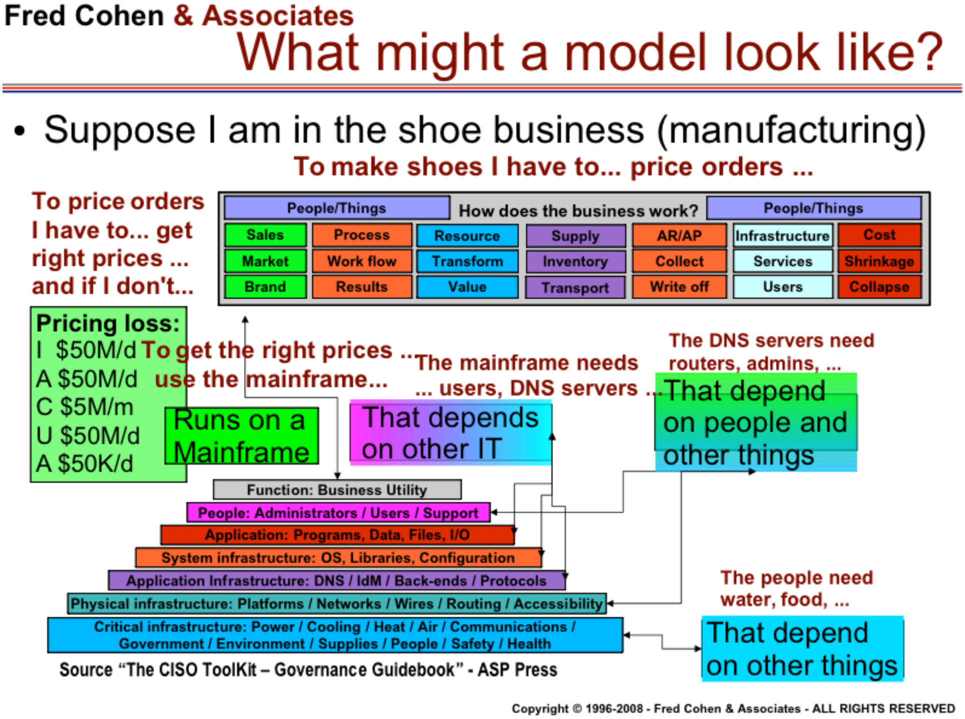
Some things do not belong in a business model. Excessive detail is undesirable because of the cost of generating and updating it and the lack of benefit gained from it. The challenge is how to eliminate the trivial while keeping the important things.

This challenge is best met today by starting at the top of the management hierarchy and understanding the business, how it works, and what it requires in order to function. This is supposed to be done by executive management as part of the Committee Of Sponsoring Organizations (COSO) of the Treadway Commission process described elsewhere and used to meet regulatory compliance requirements of the Sarbanes Oxley Act in the US or similar acts in other jurisdictions.

Excessive details are eliminated by balancing the effort and cost of data collection, entry, analysis, and presentation against the utility of the information to the process. The effort is repeated recursively through an investigative process until all business functions are understood to the level required to separate the important from the trivial in terms of the information functions of the business.

What does a business model look like?

Figure 3-2 is a depiction of a small part of a larger business model associated with an information protection analysis. This example is for a business that manufactures shoes.



Fred Cohen & Associates

What might a model look like?

- Suppose I am in the shoe business (manufacturing)

To make shoes I have to... price orders ...

To price orders I have to... get right prices ... and if I don't...

People/Things			How does the business work?		People/Things	
Sales	Process	Resource	Supply	AR/AP	Infrastructure	Cost
Market	Work flow	Transform	Inventory	Collect	Services	Shrinkage
Brand	Results	Value	Transport	Write off	Users	Collapse

Pricing loss:
I $50M/d
A $50M/d
C $5M/m
U $50M/d
A $50K/d

To get the right prices ... use the mainframe...

Runs on a Mainframe

The mainframe needs ... users, DNS servers ...

That depends on other IT

The DNS servers need routers, admins, ...

That depend on people and other things

Function: Business Utility

People: Administrators / Users / Support

Application: Programs, Data, Files, I/O

System infrastructure: OS, Libraries, Configuration

Application Infrastructure: DNS / IdM / Back-ends / Protocols

Physical infrastructure: Platforms / Networks / Wires / Routing / Accessibility

Critical infrastructure: Power / Cooling / Heat / Air / Communications / Government / Environment / Supplies / People / Safety / Health

The people need water, food, ...

That depend on other things

Source "The CISO ToolKit – Governance Guidebook" - ASP Press

Figure 3-2 – A depiction of part of a business model

The model starts out with some simple concepts. For example, to sell shoes, I have to do some list of things. In this example, one of those things is to price orders for customers. To price orders, I have to do many things, including, getting the right prices. If I don't get the right prices, the consequences are codifiable directly in terms of the business. For example, if I produce the wrong prices, I can lose money by not getting sales I otherwise would have gotten because I charged too much or lose money by charging too little. These translate directly into dollar losses for the business in time frames that are fairly clear. To get the right prices, I need results from a program that runs on a mainframe. If the program is wrong, the database it uses is wrong, the mainframe stops working, and so forth, I get no prices or wrong prices. For the mainframe to run, it needs to have a working domain name system (DNS), routers, power, etc.. Each of the things necessary to sell shoes ends up being codified in terms of content, technology, people, and things.

Business models rarely look like pictures. While pictures like Figure 3-2 are useful for explaining what a business model has in it and how it makes sense in terms of understanding how protection failures lead to business loss, these pictures don't help in the analysis, take time and effort to generate, and are not amenable to automated application, searching, and so forth. But they do show the level at which a business has to be modeled to be meaningful in terms of making reasonable and prudent decisions about protection. Hopefully, the model does a good job of modeling business functions at a gross level, specific key issues that interact with content and technology, and interdependencies. It is even better if the model allows rapid and efficient examination of malicious and accidental events so that business consequences can be readily understood and systematically analyzed.

How to build a business model

A business model is typically built by a team of people. The core team is typically fairly small and uses meetings with the top executives, business owners, people responsible for business consequences, people who understand how things really work because they do these things every day, and people who understand the technology issues and how technology carries out business functions.

The team starts at the top by understanding the business. It is hard to explain how a business is understood unless you have spent your life understanding businesses, in which case you don't need me to explain it to you. But in most enterprises, the CEO and others at that level know what their business is and does, how it works, and they think of it in some manner. There are also not all that many different sorts of businesses in the world. For example, there is a company that supplies food to restaurants all over the US. Some might say they are in the food business. Others might say they are in the distribution business, just like a wholesale book company, because they distribute food. They might also be in the shipping business because they run a fleet of trucks. They might be in the finance business because they finance their customers over the short run through their invoicing and payment processes. They might manufacture some of the food they sell, which would put them in the food processing business. They might grow their own crops and be in the farming business. I haven't ever worked for or studied them, so I don't know what business they are really in. And chances are, neither do most of their information technologists.

To understand their business and present it in terms that are meaningful to the decision makers at the executive level, the team cannot just guess at it. It is necessary for top management to come to an understanding with the team about what businesses they are in, not from an external public relations standpoint, but from an internal operational standpoint. This is necessary so that when the team presents a business consequence to top management, both can understand it clearly in terms of how they think of the business and make reasonable business decisions when the team explains the business consequences of the sorts of failures that can occur.

Based on this understanding, the team can then put information and information technology failures into business context. For a large enterprise, there are may be many consequences and event sequences. As a result, automation can come in handy. This typically includes the use of an inventory system of some sort to track the elements of the business model and a means for doing automated analysis against the inventory.

The inventory serves many other purposes, and a good inventory should include many things not required for the business model, but for the purposes of business modeling, the inventory should include all of the people and things that the business model identifies. From the perspective of the business model, each inventory item should be associated with business processes, consequences of different sorts of failures, interdependencies, and change management processes. It is all the better if the inventory provides linkage into other applications associated with how the business model is used.

How to use the business model

Once a business model is developed, it can be used systematically to answer questions about risks. For example:

- What systems are how important and why?
- How can threats interact with systems to cause harm?
- What is important enough to protect how well?
- What changed or changes when I do this?
- What am I missing and how do I compensate for it?

If the model can be "run" as a simulation, we may also answer:

- What are the SPOFs and what fails when what else fails?
- What happens as this gets overloaded?
- Which of these options will do better?

Ideally, the model is a constantly updated integrated view of enterprise information protection and business operations. But in practice, it is usually a periodically revisited collection of parts pieced together with manual efforts in which specific elements are used for specific analysis. The cost of keeping the model up to date limits its granularity and, as all models, it is a loose approximation.

How deep the detailing and how often changes are made depend on the business consequences, the modeling costs, and the support for modeling by management. Models also provide a basis for measurement in the sense that they define the things that have to be measured in order to make decisions, and this guides and helps management to direct what is measured and what is not. In this sense, the model is also potentially very harmful to the enterprise because there is a potential for risk aggregation when there are modeling errors.

The model is sometimes used to identify the necessary controls, the feedback that makes sense to business consequences, to keep track of decisions and their implications, to act as a place to track changes with time, to justify decisions, to automate, systematize, and enhance analysis, to reduce errors and omissions, and to allow differential analysis of alternatives.

Business modeling questions

1. What's the first step in defining a business model for enterprise information protection architecture?
2. How do you determine what business processes are critical so you can determine what to model?
3. Once you start to model, how deeply do you have to drill in on which areas and how do you decide?
4. How do you map failures of the business into failures in information protection and technology components?
5. What products are available to help build a business model, and what things cannot realistically be provided by a product?
6. How do you determine the consequences of losses of IACUA?
7. How do you put issues of time into the model?
8. How do you model interdependencies?
9. What else do you need to simulate the business in terms of protection failures?
10. How does a business model track changes to the business over time?

4 Oversight

Oversight is the critical governance function provided by top management relating to information protection and it is fundamental to proper operation of a protection program. It is the job of oversight to assure that proper duties to protect are put in place, that the management measures the effectiveness of the protection program in fulfilling those duties, and that management adapts the protection program to meet those duties.

Laws: Laws and regulations define the legally mandated duties to protect associated with jurisdictions. All laws of all jurisdictions in which an enterprise operates have to be considered in order to make prudent determinations about duty to protect.

Owners: The owners are the ones hurt by bad management decisions and they need to assure that their investment is not lost by electing proper boards of directors. For public companies there are regulatory assurances to support the public owners so that they don't have to get involved in the details of selections in order to reasonably protect their investments, but this lack of direct control by owners is often reflected in the frauds we see in the world. Owners of privately held firms are directly responsible for the disposition of their assets and for proper protection and they directly suffer from poor decisions in this regard.

Board: The board of directors is legally and morally responsible to assure that the CEO and other officers are doing their jobs and have the ability to define additional duties to protect in keeping with their responsibilities. They also have oversight responsibility to act on behalf of the shareholders to assure that the shareholder value is protected.

Auditors: Auditors are tasked with providing independent and objective feedback to the shareholders, board of directors, CEO, and others on the effectiveness of the protection program in fulfilling the duties to protect within the risk tolerance parameters set by management.

CEO: The CEO is responsible for day-to-day control over the enterprise, and as part and parcel of this responsibility, for protecting shareholder value, for identifying the duties to protect, for assuring that those duties are carried out, for measuring the performance of those duties to allow adequate control to improve situations that warrant improvement and for keeping costs as low as possible without undertaking inappropriate levels of risk.

In concert, these elements comprise the oversight functions of enterprise information protection and define the duty to protect.

Duty to protect

Individuals in organizations have duties to protect various things. Duties stem from three general areas; externally imposed duties, internally imposed duties, duties, and associated with contracts.

Externally imposed duties

Legal and regulatory mandates are derived from laws, regulations, protective orders, judicial determinations, and ordinances at all jurisdictional levels. There are generally three classes of these duties; (1) those associated with all businesses in jurisdictions, (2) those associated with specific types of businesses involving special duties like public health and safety duties of drug or chemical manufacturers, and (3) fiduciary duties to shareholders by officers to retain and grow the value of shareholder investments.

Internally imposed duties

Companies often decide to protect private information, safety of workers, release of information to third parties, and other similar information or assets beyond the levels imposed by government. When these decisions are codified in any form, including but not limited to normal operating procedures and processes, documented practices, or policies, they obtain the force of a legal obligation. These self-imposed duties can be the basis of legal actions against the corporation. A good example was a privacy policy published on a Web site by a large Internet Service Provider. They didn't follow their self-imposed privacy policy, got sued for disclosure of information, and lost $1M in the process. If they had no such policy they would have had no such duty to protect and would have had no liability. Trade secrets, copyrights, trademarks, and patents are important examples of intellectual property with self-imposed duties to protect.

Contractual duties

Contractual obligations are legally binding obligations voluntarily taken on as part of doing business. They typically include things like safe harbor agreements, confidentiality and non-disclosure agreements, trade secret agreements, licensing agreements for patented or copyrighted material, and almost anything else that the parties wish to codify in a legal agreement as part of the terms and conditions of doing business.

Some well known duties to protect

Some of the best known duties to protect stem from legal cases. Of course these cases are constantly underway with different outcomes in different jurisdictions, and a detailed review of these issues is really the subject of a book in and of itself. However; some of the more obvious duties that exist today are worth outlining here.

Business record retention and disposition

This is a major field in and of itself.[4.1] For a wide range of reasons, ranging from government regulations and laws to fulfilling contract obligations, to securing intellectual property rights, to meeting financial regulations, to assuring employee rights, businesses create and possess records of their operations. These are commonly called “normal business records”, they are typically relied upon for business operations, and they are considered to be reasonably reliable and admissible for legal purposes.

Specific sorts of records of these sorts have specific retention requirements in specific jurisdictions. For example, employment records are normally retained for 4 years after the employee's relationship is terminated, while financial records of banks have to be retained subject to banking regulations, there are requirements to collect and retain records associated with legal matters both underway and anticipated, and so forth. There are emergent businesses including major law and accounting firms that provide tracking of requirements associated with retention, but each enterprise has to do its own internal reviews of requirements to address its needs from a legal standpoint.

Once retention requirements for records, including without limit records contained in emails, system logs, transmissions like instant messages and chat session, telephone records, and so forth) are identified, decisions about how to collect, store, and dispose of these records and when to do so must be made. This includes the meta-data (data about the data such as file system change time, creation times, and so forth) associated with these records. This must meet legal requirements for retaining records related to pending or anticipated legal matters (commonly called legal holds), and disposition of records held elsewhere or for others.

Process has to be defined for carrying out these responsibilities and the process must be put in place and properly managed and executed. The failure to do any of these things may result in legal liability and in recent cases, has resulted in major financial losses.

Disaster recovery and business continuity planning

Disaster recovery is distinct from business continuity, however; they also share a great deal of commonality. Most enterprises have some set of programs that cover recovery from large-scale incidents as well as loss of key personnel.[4.2] To the extent that they don't, they should. The duties to protect should normally include duties associated with these programs, including delegation of authority for the various components of these aspects as they apply to information in its various forms. The information protection function is typically only part of the overall process that includes everything from limiting the simultaneous appearance of multiple key personnel in the same place to assuring that there is enough cash available through loans to recover from a major warehouse fire. Part of this almost always involves information and technology, paper records, and intellectual property, and duties related thereto.

Privacy requirements

Privacy requirements range over a wide spectrum as well. They may include, without limitation;

- Legal limitations on what can be released about employees and others. For example, the Gramm Leach Bliley Act (GLBA) deals with financial records, the Health Information Portability and Accountability Act (HIPAA) deals with medical records, the EU privacy act limits storage of records about individuals, and other similar laws from all over the world impose similar requirements;

- Contractual obligations are created by policies and contracts. For example, posted privacy policies are legally enforceable contracts, written contracts and agreements with others often mandate privacy controls, subcontractors may sign non-disclosure agreements, and so forth;
- Internal management decisions about what is and is not acceptable in the work place;
- HR limitations on information about individuals and who can access which records under what circumstances, including without limit employee evaluations, recommendation letters, sanctions, and internal investigative reports;
- Attorney-client, medical, health, religious, and other privileged matters;
- Video tapes and other closed circuit or related recording devices often have specific protection requirements such as where they can be placed and used and what they can and cannot record or be used for.

There are many forums that discuss privacy issues and this area is becoming a substantial legal specialty.

Financial records and related content controls

For public companies there are specific limitations on controlled financial information such as pending mergers and acquisitions, protection of financial reports prior to official release, and accuracy requirements on earnings expectations. For almost all companies there are other financial numbers, such as price lists or discounts to select customers and cost figures for bulk purchases, that are proprietary and very important to maintain in strictest confidence. Pending contracts are also often quite sensitive. Laws related to financial records are generally covered by the US Federal Trade Commission and similar bodies throughout the World, banking regulations worldwide, and many other mandates.

Intellectual property controls

Specific requirements exist for protection of patents and pre-patent content, including records and notes, copyrighted materials, and trade secrets. These include all manner of marking, storage, publication, and related mandates that must be followed, depending on the enterprise approach to this property, the laws that apply, and the specifics of the values at hand. The duties associated with these typically involve enterprise processes defined by top management.[4.3]

Customer content

Customer content (or other peoples' content) is often protected by companies that provide service to others. In addition to contractual obligations, there may be internally imposed duties associated with how customers and their content are treated. Because of the relationships between companies and the increasing horizontal nature of relationships involving information, this is becoming increasingly important and increasingly specified in duties. [4.4]

Codes of ethics and standards of practice

Many companies have codes of ethics or other similar mandates that are imposed by top management as part of corporate governance. There are also standards of practice associated with various professions and professional societies have codes of ethics that are sometime included in or referenced by enterprise governance. Many of these increasingly address information protection issues.[4.5]

Operational status

In many enterprises, operational status is considered important to the well being of the organization and has impacts on morale as well as the potential to be externally exploited. Production levels are commonly watched in manufacturing contexts and shared internally but considered sensitive for others to see. Workplace accidents are also important in this respect. In many companies and governmental bodies, operations security is vital to the safety and security of the operations and the people involved in them. [4.6]

Investigations underway

Ongoing investigations are often considered to be highly sensitive matters subject to protections of all sorts. This is particularly important because, during an investigative process, many things may be examined that end up not being important. The taint of the investigation itself may be quite harmful to individuals and their careers, produce the perception of conflicts of interest, and have other similar side effects.

Reporting requirements

For regulated industries and transactions there are a wide range of requirements. For example, funds transfers in excess of a few thousand dollars may have to be reported, international shipments normally have customs and other similar requirements, reporting on controlled substances is required for those licensed to deal in or with them, and many industries have other similar or specific reporting or compliance requirements.

Risk tolerance and management

Top management defines risk tolerances and different tolerances and decision-making authorities are assigned to different people and positions. This leads to the creation of authorities for decisions and the process by which decisions of different sorts are made. The duties are ultimately determined by a combination of policies and processes that include exceptions, appeals, and approvals. These lead to different assignments of duties for different situations.

Managing changes in oversight

Given the wide range of issues involved in oversight and its responsibility to define and enforce the duties to protect, it quickly becomes clear that a change management process is required to deal with the sheer volume of duties that may be defined and the changes associated with them across all venues. For example, for every contract, a change in duties may have to be reflected in new peering requirements with outside entities and take years to put in effect.

Laws are still being written in this area and legal cases are underway all of the time that change the landscape in this largely unsettled area of law. Peering agreements between entities and the mandates for things like safe harbor and transitive contract requirements mean that changes in laws elsewhere may have effects on duties of the enterprise at almost any time.

Personnel also change at the top management level, and this translates into different views on the optional duties and different stances on risk tolerance, approaches to risk taking, mitigation, transfer, and acceptance, and levels of assignments to manage the duties themselves.

The world changes and, while many stable countries have relatively stable legal situations, other places in the world are under flux, and internationalization combined with increased globalization and interdependencies, sometimes result in radical changes to the situation that get reflected in rapid changes of select duties. When there is a revolution somewhere in the world, or even a radical change of government, there may be very short notice of substantial change. While this rarely impacts top-level duties to protect, it sometimes does happen, literally, and overnight.

Decisions made by oversight

In addition to top-down mandates, executive management at the enterprise level also requires feedback in order to make reasonable and prudent decisions. Oversight makes a wide range of decisions, many of which indirectly effect the protection program, and some of which completely ignore the protection program while having major impacts upon it.

A common example of a radical change that is often with little or no notice is a merger or acquisition requiring the integration of a new enterprise into the existing enterprise. Unless the CISO is part of the merger and acquisitions team, there may be a legal requirement for concealment of the change from them. And even if they do know in advance, there may be little they can do to prepare for such a change other than let top management know what it might cost and how long it might take. Due diligence often ignores information protection issues and surprises can be stunning.

The need for independent evaluation

As for all decision-makers, the quality of oversight decisions depends heavily on the information they get. One of the fundamental goals of the CISO is to get the best information to top management so they can make the best possible decisions. Indeed this is one of the least understood elements of the role of protection executives and one of the most important things that they do.

- ***Top management is often treated as if they know nothing about the decisions they make.***
- ***Technical people often complain that they are not understood by management.***
- ***But management can only do so much to ask about everything going on within their enterprise.***
- ***In order to make good decisions top management needs good information.***
- ***It is incumbent on those who have the information to make it available and understandable in a form that allows better decisions to be made.***

Whether it is fear, loathing, jealousy, disdain, or a lack of caring that keeps management and workers at all levels from providing top management with true and reliable information, each time the information is withheld or restated, it becomes harder for the executive to make a sound decision.

This is one of the reasons that audit exists, and one of the reasons that consultants are often used by upper management. If it weren't for independent investigations and evaluations by outside experts, many of the most important internal issues would never rise to the surface and gain the attention they need by top management.

This is also an important reason that it is vital for the CISO position to be independent of the CIO, CFO, and others who might cover up their own mistakes or prevent a detailed examination of what is going on at every level within the enterprise. When the CEO does not have at least monthly confidential discussions with the CISO at which the CISO's management chain is not present, the CISO will find themselves forced to shade the truth about what is underway. Where a CISO doesn't have the power to independently investigate anything of import, their management will conceal the truth from them or focus them away from things that top management must know about. And when the CISO works for someone they have to report on, the CISO is likely to be fired unless they “go along to get along”. This is poison for the enterprise and poor oversight.

Oversight questions

1. In light of the complexities associated with duties, what sort of documentation should be expected to be provided by top management to provide clarity surrounding the duties to protect and what should be left to which others to define?
2. Take and example of one type of content, one common set of requirements, one country, or one enterprise and create the definitive list of duties. What were the challenges, what was easy, what resources did you use, and what do you think would be required to do this for a whole enterprise?
3. How would you provide details on the duties to protect to all of the appropriate parties within the enterprise and keep them up to date with changes in this situation?
4. What mechanisms and systems should be in place to handle changes associated with duties to protect and how can these systems help to ripple the necessary changes through the enterprise?
5. How do you let top management know what they need to know to make quality information protection decisions?
6. Why is it important for the CISO to be independent of the CIO, CFO, and others at the executive level and what alternatives are there to an independent CISO?
7. What information is needed to make good decisions about duties to protect?
8. When the duties to protect specify who is authorized to make what level of risk management decision, how can they codify the difference between the purchasing authority of someone who can buy a security device and the risk-related authorities associated with the decision to make the purchase based on the potentially serious negative consequences it is intended to mitigate?

5 Risk management and what to protect

Risk management transforms duty to protect into what to protect and how well to protect it, selects between risk acceptance, transfer, avoidance, and mitigation, and for risk mitigation, attempts to match surety of mitigation with desired risk reduction.

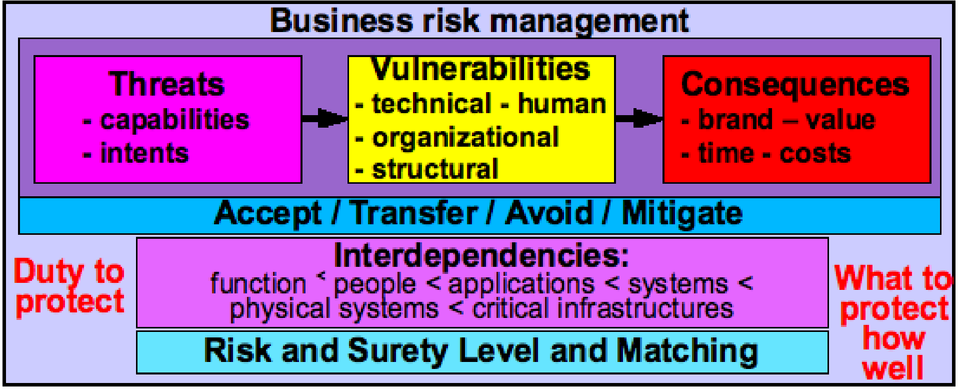


Figure 5-1 An overview of risk management

Risks are generally formed from the combination of threats, vulnerabilities, and consequences. Threats, including nature and accidents as well as individual actors and groups, possibly acting in concert, induce or conspire with sequences of events to exploit vulnerabilities to induce consequences.

Some background on the nature of risk

Risk management is most often concerned with potentially serious negative consequences, however; since consequences may be direct, indirect, intended, unintended, obvious, and/or subtle. Since event sequences may or may not occur at some future time, there are few if any certainties to rely upon in this field. Since experiments cannot generally be repeated in the real world, measurement of risks, even after the fact, cannot be done.

For certain types of risks, such as some types of naturally occurring phenomena, the use of statistics can help to pool risks when more detailed analysis is unavailable or too costly to be justified. For example, every winter there is cold weather in certain locations, and combined with precipitation levels, this brings snow. Without predicting the precise weather at each location, statistics can be used to identify when spring rains will combine with snow melts to bring about flooding and where those floods will occur with what level of severity and how often. These statistical models can then be used to assess a likelihood of homes being inundated with water, the average loss values associated with those homes can be determined based on market prices, and insurance can be sold with a reasonable expectation of making a profit by charging more each year for the insurance than the average expected loss from paying out claims plus expenses associated with running the insurance company. On average the insurance company makes money, while the home owner pays a premium every year to make certain that they don't lose the very large investment in their home, even realizing that on the average, they lose money over the long run in the process.

The statistics used on this type of industry are produced by actuaries who generate actuarial tables that are used to look up the price to charge based on the parameters modeled in the statistics. Parameters that are not relevant are determined to be so by the analysis of large numbers of instances based on the evidence of past experience, while weightings under conditions are generated by those same experiential pieces of evidence.

These models are excellent for situations in which large numbers of samples are available, events are not reasonably predictable on a specific basis but happen with enough frequency to be predictable in the aggregate to within a reasonable degree of accuracy, when those events behave approximately like random stochastic processes or other mathematically modeled occurrences, and in cases where past history is a reliable predictor of future events. Unfortunately, this is often not the case, especially in the information protection field or in astable periods of change.

Many things may go wrong when such models are used because the nature of risks is not always the same as the risks of nature. The underlying phenomena may change, and when these changes occur, old statistics are no longer reliable as indicators. For example, as climate changes occur, the statistical basis for assumptions in the models change and unless the models are updated, they will yield erroneous results.

The information protection business often deals with intentional actors in individual and groups forms, with malicious intent, and who adapt intelligently with respect to defenses. As a result, statistical models used for outcomes of weather phenomena are poor at making predictions about what will occur in the future. While probabilistic risk analysis based on multiplying the likelihood of an event by the consequence of the event to generate an expected loss with time is usable for flood insurance risks to losses of homes, the indirect and hard to anticipate nature of events and consequences and the lack of large numbers of samples of events make even the basic process of gathering data on information-related risks infeasible as a methodology for considering this analysis. Still, there are many who continue to use this approach in making protection-related decisions.

The problem with measuring risks, even after the fact, remains an enormous hurdle to overcome in seeking to perfect the practice of risk management as a scientific discipline. If a protective measure was in place and nothing bad happened, what does that tell us about what would have happened if the measure was not in place? We can be reasonably certain that a large enterprise with no protective measures in place against information-related attack will fail and do so relatively quickly. That seems to imply that the protection program is saving the entire value of the enterprise every day. And of course it is. But does that mean that the budget for protection should be higher, or since the enterprise is still here, that it is high enough? We cannot repeat what happened yesterday to see what would have happened if we spent more or less. But we might be able to simulate it. [5.1]

The methodologies described and discussed here are practical for use in enterprise risk management processes, but the problem remains and will likely always remain that we cannot predict risks reliably, we cannot track what did not happen, and no mathematical solution will replace the need for people in charge to make judgment calls that may go right or wrong. What we can do is get systematic about trying to evaluate and understand the nature of risks, present sound and meaningful risk-related information to decision-makers at the right level, learn from the limited experience available in the literature, and constantly work to improve.

Risk identification and evaluation

Risks have to be identified and evaluated in order to be managed. The objective of risk identification and evaluation is to identify event sequences with potentially serious negative consequences based on the business model and evaluate the business consequences.

Consequences

Consequences are identified from the business model and rated, typically into low, medium, and high levels, or in a more mature process into more accurate values when warranted within the medium and high value categories.

- Low consequence is typical of business risks like slip and fall accidents and similar readily insurable things.
- Medium risks tend to have serious business impact and include public relations problems, loss of substantial amounts of trust or money, inability to perform on select important contracts, and so forth.
- High consequences tend to involve loss of life, great harm to the environment, collapse of the business, and/or jail time for executives.

Consequences are normally very specific to the organization, however they tend to group into issues of (1) brand or reputation, (2) value, which codifies a variety of financial implications ranging from loss of cash to destruction of stock to loss of information value for periods of time, (3) time which is lost due to people not being as effective at their jobs or the business losing opportunities, and (4) costs which are the direct costs associated with dealing with the incident and its aftermath.

In analysis of risk through this approach, it is prudent to assume that business processes fail and identify the consequences associated with those failures. Any sort of risk mitigation or transfer should be ignored at this point so as to not confuse consequence analysis with other issues in risk analysis and to allow alternative approaches to be tried to mitigate risk without regard to what might or might not be in place at any given time.

Threats and threat assessment

For event sequences involving medium or high consequences, threats should be assessed with increasing attention and detail for higher consequences. As threats are identified, their capabilities and intents are typically taken into consideration in assessing them. As threat assessments are undertaken, the results of a previous effort may apply to the next consequence of import, and when this is the case, the existing assessment can often be applied without additional effort. Because threats may change over time, the process has to be repeated with periodicity appropriate to the circumstance based on cost and consequence.

Threats can be thought of in terms of their capabilities and intents. Capabilities include but are not limited to funding, location, attack mechanisms available, group size, available resources, skill sets, training levels, allies, and access. Intents are harder to identify with specificity, however they are typically assessed in light of group history, motives, group behaviors, group rewards, typical targets, leadership, and declared objectives.

Different threat assessment methodologies are suitable for different circumstances.

Pre-employment and periodic worker checks are part of employee threat assessment. Additional investigation and review is used for positions of higher trust.

Case investigation is used in response to incidents. For example, this is used if an employee gets a threatening letter that rises to the level where the company determines follow-up is prudent and appropriate.

Detailed intelligence is undertaken against specific threats that are known to exist and that are targeting the company for high valued consequence. A good example was the Irish Republican Army (IRA) bombings of financial institutions in the 1980s and 1990s. The specifics of their efforts had to be understood to counter the threat.

Regional intelligence is typically used when moving into a region or when operating in a region under substantial regional threat. For example, building up a business in the Middle East is clearly different from the Pacific Rim in terms of the threats faced.

Local intelligence is used whenever making determinations about placement of facilities, offices, routes, or housing, and when ranking locations for determining where to go and what to do there.

Investigative intelligence is typically used for clearances associated with government jobs, but it can also be used for investigations of employees for high-level-of-trust jobs, and for verification of lifestyle conditions such as rapid changes in wealth.

Table 5-2 shows a methodology for selecting threat assessment methodologies. It includes assessment method, consequence level, time frame, threat level, and cost of assessment. As a general rule, it is better to spend less money and take less time whenever possible, but the problem in threat assessment is that until you look into threats you can't tell whether they are important. As a result, it is common to (1) do an initial threat assessment by type at a generic level and, based on the results of this assessment and consequences, (2) decide which generic threats justify more detailed investigations. Threats change over time, so periodic reassessment is a good idea. Typically, a by-type at the generic level threat assessment is done as part of a protection posture assessment and should be undertaken at least once per year.

Assessment method	*Consequence*	*Time*	*Threat*	*Cost*
By type generic	Medium	Short	Med	Low
By type, classes within groups	Med-high	Med	Med-high	Med
By type, classes, detailed high relevancy	Med-high	Med-long	Med-high	High
Known vulnerability indications and warnings	Med	Short	Low	Low
Detailed intelligence analysis	High	Long	High	High
Investigation-based	Med-high	Med	Med-high	Med-high

Table 5-2 Threat assessment selection methodology

Table 5-3 shows a typical threat roll-up used to do a high-level summary of threats to an enterprise. This table includes assumed values for each of 24 classes of threats including typical funding per job, group size, motivation, skill level, hours of effort per attack process, and initial access. The table is typically augmented by the assessment team to indicate the level of concern and any specific concerns with respect to each threat type.

Threat type	Funding/job	Size	Motive	Skill	Hrs/task	Access
activists	10K	1◁10K	Justice	Med	10K	Insider
club initiates	100	3◁50	Acceptance	Low	48	Internet
competitors	>100K	2◁5	Money	Med	2K	Industry
consultants	0	1	Money	Med	No limit	Insider
crackers	1K-100K	1◁100	Malice	Med	No limit	Internet
crackers for hire	>100K	1◁10	Money	Med	1K	Internet
customers	1K	1◁5	Money	Low	1K	Partner
cyber-gangs	<1K	1◁100	Money	Low	1K	Internet
deranged people	Small	1	Insanity	Any	No limit	Internet
drug cartels	>10M	100◁5K	Money/power	Med	1K	Internet
economic rivals	>1B	10◁1K	Money	High	1K	Industry
extortionists	100-1K	1+10	Money	Low	100	Internet
spies	>1B	>10K	Patriotism	High	No limit	Insider
fraudsters	100-100K	1◁20	Money	Med	100	Internet
global coalitions	>1M	10◁100	Money	Med	10K	Industry
government agencies	>1B	>1K	Patriotism	High	No limit	Internet
hackers	100-10K	1◁10	Exploration	Low	No limit	Internet
hoodlums	100-10K	2◁20	Money	Low	100	Internet
industrial espionage	10K-100K	1◁5	Money	High	1K	Industry
information warriors	>100M	1◁10K	Patriotism	High	10K	Insider
infrastructure warriors	>1B	5◁100	Patriotism	High	10K	Industry
insiders	1K	1◁5	$$/Revenge	Med	1K	Insider
maintenance people	100	1◁5	Money	Low	10	Insider
military organizations	>1B	5◁500	Patriotism	High	10K	Industry
nature	Unlimited	No limit	Randomness	Low	No limit	No limit
organized crime	>10K	1◁5	Money	Med	1K	Internet
paramilitary groups	10K-100K	5◁25	Fun/Beliefs	Low	1K	Internet
police	1K-10K	1◁500	Justice	Med	No limit	Industry
private investigators	100-10K	1◁10	Money	Med	100	Industry
professional thieves	10K-100K	1◁3	Money	Med	1K	Industry
reporters	1000-10K	1	Exploration	Low	100	Internet
terrorists	10K-100K	5◁50	Religion/Power	Med	10K	Internet
tiger teams	15K-150K	3◁5	Money/Pride	Med	100	Industry
vandals	0	1◁10	Randomness	Low	1	Internet
vendors	1K-1000K	1◁20	Money	High	1K	Insider
whistle blowers		1	Justice	Low	100	Insider

Table 5-3 Threats and some related data

This categorization is only one of many that have been used for threat analysis. We have a simulation engine linking the threats to attack mechanisms, but there are many other valid approaches at this level of specificity that can be used in its place.[5.2]

Vulnerabilities

For systems with identified high or medium consequences and whose threats have been assessed as having the capabilities and intents to induce those consequences, vulnerability analysis and mitigation is sometimes undertaken. Vulnerabilities and the paths attackers take to exploit vulnerabilities were described earlier in general terms. They tend to include technical vulnerabilities most commonly associated with computer security, human vulnerabilities that are covered under a variety of topic areas in the psychological literature, structural vulnerabilities that have to do with overall network and infrastructure architecture and dependencies, and organizational vulnerabilities that have to do with weaknesses in the way things are organized and how people interact with each other within the structure.

Vulnerabilities are typically assessed by a testing process of some sort and ranked by criticality and severity in context. The problem with most vulnerability assessments in use today is that they are undertaken as independent efforts and not within the proper enterprise context. They find many vulnerabilities in low-valued systems, fail to properly evaluate their implications, and indicate mitigation that is more expensive and in a worse order than would be found if the task was more properly done. As a rule, for an efficient protection program, vulnerability assessment should only be done selectively and only as directed by results of consequence assessment followed by threat assessment.

Vulnerabilities always exist for risks not avoided, and their total elimination is generally considered impossible, or at least too expensive to warrant serious consideration. It may be wiser to consider vulnerabilities in the context of a threat attempting to induce event sequences that exploit weaknesses and then induce consequences. What we may identify as vulnerabilities can also be thought of as the locations of events along the path of the attack, typically from source to target. The meaning of the term “vulnerability” in this context is that it is one of the steps along that path. Thus it is not necessarily a “flaw” or a “weakness”. It may simply be a choice made of how to implement a system, and all alternatives may also have similar paths. [5.3]

Interdependencies and risk aggregations

Interdependencies are far more complex in information technology than in most other systems because of the high complexity of the individual systems, the high connectivity of information networks over vast distances, and the short time frames associated with the transfer of information over those distances. The manner in which these systems are constructed today is to assemble complex components into composites recursively and use components as part of infrastructures even though those components are not really designed to act as high surety infrastructure components.

Interdependencies

Business utility, or function, typically depends on a large collection of mechanisms that can be couched as a hierarchy. People dependencies start with users, administrators, and support personnel that use systems and keep them functioning. These people depend, in turn, on application programs, data files, and input and output systems for their interactions with the information technology. The applications work through systems infrastructures that include operating systems, libraries, and configurations. The system infrastructures often depend on distant infrastructures such as domain name services, identity management systems, back-end processing facilities, and the protocols used to communicate with these capabilities. The application infrastructure operates over physical infrastructures that include computing platforms, networks, wires, routing protocols, and accessibility to all of those elements. The physical information infrastructure then depends on power, cooling, heat, air, communications infrastructure, governments and political stability, environmental conditions and controls, supplies, people, and the safety and health of those people. In this complex chain of interdependencies, any fault can cause systemic failure unless the fault is masked by some sort of protective mechanism that allows the overall system to continue to function in its presence. Because some of the fault mechanisms, like computer viruses or electrical power supply, are active and potentially systemic in nature, subtle and unanticipated consequences occur when inadequate expertise is applied to this issue. Figure 5-4 depicts the notion of this interdependency structure.

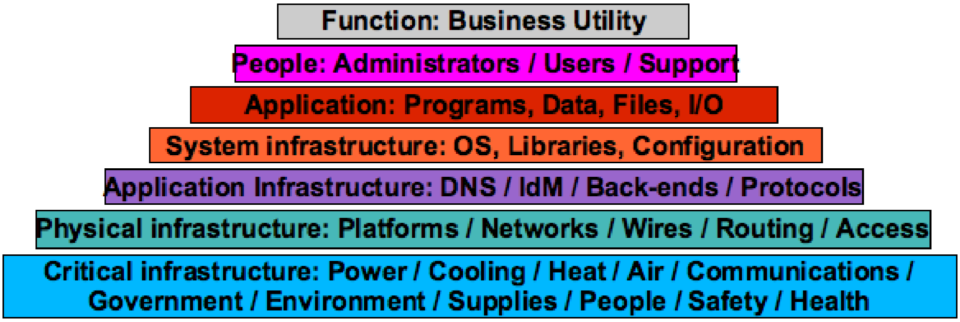

Function: Business Utility
People: Administrators / Users / Support
Application: Programs, Data, Files, I/O
System infrastructure: OS, Libraries, Configuration
Application Infrastructure: DNS / IdM / Back-ends / Protocols
Physical infrastructure: Platforms / Networks / Wires / Routing / Access
Critical infrastructure: Power / Cooling / Heat / Air / Communications / Government / Environment / Supplies / People / Safety / Health

Figure 5-4 – Interdependencies in information systems

Single points of failure as risk aggregations

There are many cases when a single point of failure goes unnoticed in a large infrastructure and its eventual failure leads to infrastructure-wide collapse. Depending on timeliness issues, this can result in consequences ranging from short-term inconvenience to enterprise collapse. To understand single points of failure, the only current approach is to do and interdependency analysis to find the dependency chains underlying business functions identified in the business model. Failure mode analysis associated with all dependencies can then be done to determine business impacts of faults and masking strategies. There are two classes of single points of failure to consider; (1) Any individual system, facility, key individual, or other dependency, within a radius of effect associated with the attack mechanisms within the capabilities of the threats identified in threat assessment, and (2) Common mode failures.

Radius-driven common mode failures

Different threats have different capabilities. Those capabilities lead to radii of effects. For example, nature brings large-scale effects like Earthquakes and hurricanes. To assert single point of failure protection, natural effects within reasonably expected and historically supported radii must be taken into account. Redundant data centers in the same Earthquake zone cannot support the claim to have no single point of failure because a single Earthquake can cause all of them to fail. Redundancy within a single building will not withstand a single explosion at that building, and this is within the threat profile of any substantial enterprise.

Other sorts of common mode failures

Common mode failures are failures that result from commonalities between systems or components. Anything that systems have in common and that can fail will be subject to common-mode failures. The only way to eliminate common mode failures is to use diversity. An example of a widely exploited common-mode failure is the use of large numbers of systems with the same version of the same operating system. A virus that infects those systems tends to do widespread damage. But the operational and cost efficiency of running a single operating system may justify the increased risk of common mode failures. A diversity approach would be to have multiple implementations of each business function. But clearly this at least doubles all development and operational costs, particularly for large-scale systems used for enterprise databases and similar back end processes. Trade offs must be made, and selective diversity is typically practiced only after serious analysis is done.

Key individuals

A particularly good example of risk aggregation is a single individual who controls a substantial portion of information infrastructure and for whom there is no backup. The tradeoffs associated with business failure have to be considered for any small business or small part of a large enterprise. But for any substantial enterprise such a dependency must be eliminated. Consider the implications if the individual has a heart attack or gets hit by a car.

An even more critical element of the key individual problem is the potential that a key individual will turn against the enterprise. For this, checks and balances, separation of duties, and other similar approaches are required. When there are key individuals, they are an insider threat. Examples range from a marketing executive who takes sensitive customer information and goes to a competitor to a network administrator who uses a global network management password to reset all of the passwords on the global routing infrastructure and then reroutes all of the traffic so that the entire enterprise information infrastructure is non-functional. Within a few minutes, such a change can disrupt and enterprise for hours, days, or longer, depending on the controls in place.

Risk treatment options

Risk treatment is the process by which risks that are worthy of attention are managed and risks not worthy of consideration, and other risks as appropriate, are accepted. A risk treatment plan should be identified for all risks identified.

Risk acceptance

Risk acceptance involves a decision by management to accept a risk without further mitigation or transfer, for a period of time. This happens in two classes of circumstances. For risks that are too low to bother protecting against or for which existing insurance and due diligence are adequate, risk is accepted. For risks that are to be mitigated but where mitigation cannot be done instantly, and in cases where rapid mitigation is too expensive to justify, risks are accepted for periods during which mitigation is undertaken. Transfer may also take time and risks not yet transferred may be accepted pending completion of the transfer. Risks that are accepted are, de facto, transferred to the enterprise owners.

Risk avoidance

Risk avoidance is a business strategy in which certain classes of activities or business processes are not undertaken because the risks are too high to justify the return on investment. A typical example is a decision about the maximum value to be placed in a vault, at a site, or on a truck. This strategy avoids the aggregation of risks associated with placing excessive value in one place. Other similar avoidance strategies such as not opening offices in war zones or not doing business in certain localities are commonplace in business.

Less common, but equally important, is the notion that when undertaking or continuing a business activity, the risks should be considered as part of the determination of what to pursue and stop pursuing. Most such risk avoidance due to information-related issues happens when business activities are shut down because the harm becomes too great to continue to sustain. It would clearly be better to identify such circumstances in advance to avoid the losses associated with such situations, but history shows that few decision makers are this anticipatory.

Risk transfer

Risk transfer for low consequences is usually affordable and reasonable if some level of reasonable and prudent controls are in place to meet due diligence standards for low risk systems. There are no definitive due diligence standards, however; there are examples that come pretty close in terms of risk transfer. For example, the payment card industry (PCI) standard published by Visa mandates specific controls for those who process credit card transactions. It is a standard required by credit card processing through the Visa credit card system and is reflected in contract requirements for performing such processing.

Another widespread standard set that is considered representative of due diligence by many is the ISO 2700X control standard series. Some of the standards will be discussed later, however; none of these are formally defined in any way as being diligent from a stand point of risk transfer. Rather, those who follow them tend to be considered diligent enough to meet requirements of others who use contractual and other means to transfer risks.

Risk transfer for medium and high consequences is rare, expensive, and only justified in cases where the worst case loss is not sustainable and an adequate outside insurance capacity is willing to take on the risk. This is a strategy that loses in the long run for medium and high risks because insurance companies have to make a lot of money on each transaction to justify the high consequence of loss and the unknown actuarial nature of the situation.

A good example of this sort of transfer is the insurance of space flight, in which a multi-billion dollar satellite might be insured with a billion dollar deductible and a cost of 33% of the maximum payout. For a $10B space shot, the insurer might get paid $3B and have a $1B deductible, for a maximum $9B loss to the insurance company and a maximum $4B loss for the aerospace company. While this may seem crazy to many observers, if the aerospace company cannot afford to lose $10B and has a $4B profit on the overall effort, it might be worth reducing the profit to $1B in exchange for a guarantee of staying in business regardless of the outcome. For the insurer, it's simply a large bet similar to its smaller bets.

Risk mitigation

Risk mitigation seeks to reduce residual risk by using safeguards to eliminate or reduce the paths to or consequence levels of event sequences that can cause serious negative consequences. This involves reduction of threats, reduction of event sequences that threats can apply, reduction of vulnerabilities (or total relevant event sequences), reduction of the link between event sequences and consequences, and reduction of consequences associated with event sequences. All mitigation leaves residual risk that eventually has to be accepted, transferred, or avoided. The question is how much reduction is desired and how much is afforded by the mitigation strategy employed at what cost.

What to protect how well: risk management

Risk management is the process used by enterprises to turn duty to protect into decisions of what to protect and how well. In other words, it answers the questions:

- What utilities of what content are to be protected to what level of certainty?
- How is this level of certainty to be attained?

These are executive decisions that drive the executive protection management function tasked with carrying out the duty to protect. They generally involve the creation, operation, and adaptation of a business process that facilitates making consistent and rational decisions with respect to meeting protection objectives.

The risk management space

This risk management process involves many intertwined issues and results in controls that are appropriate to the risks. As a rule of thumb, starting in the middle of Figure 5-5 with an information protection posture assessment provides a medium-cost way to get a handle on the overall situation. From there, low, medium, and high risk (in this model a combination of threat and consequence yield the notion of risk) situations are identified and additional work is done to make better decisions for higher risk (closer to the upper right hand corner) issues. Other risk assessment approaches exist and should be considered.[5.4]

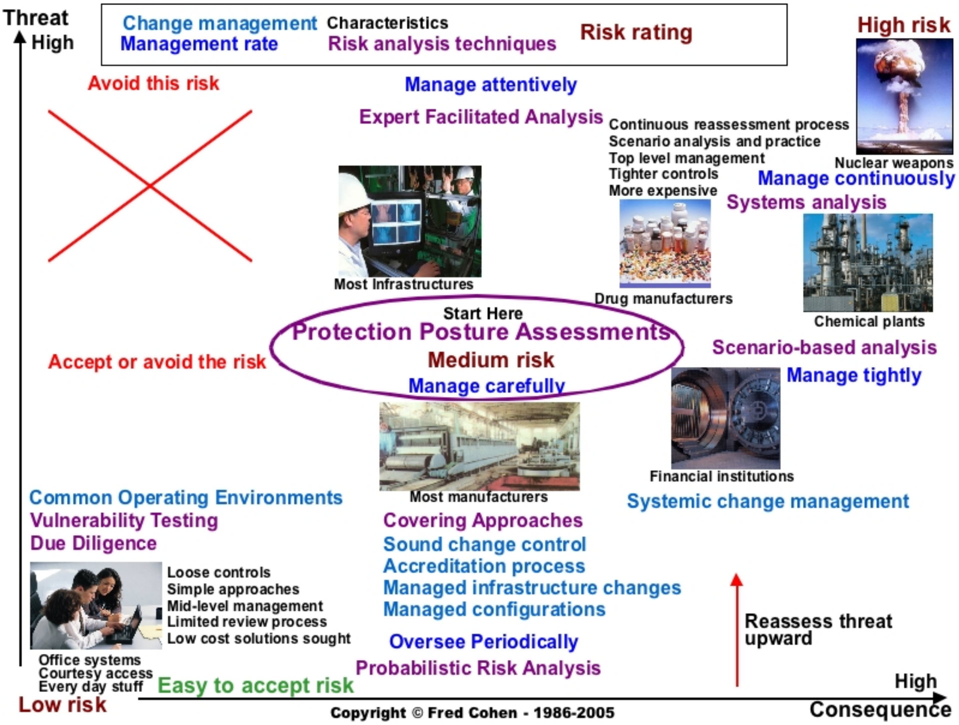


Figure 5-5 – A depiction of the risk management process space

Risk levels lead to management rates and management complexity, change management mechanisms, and different risk assessment techniques.

Low risk options

For the low risk end of the spectrum, where most day-to-day users tend to work, due diligence approaches and vulnerability testing are adequate to the risk assessment process. Diligence with respect to not becoming a hazard is required for any system, and vulnerability testing is a good way to get a handle on easily repaired problems. These are inexpensive and reasonable things to do in most cases. Common operating environments are often used to save on costs of operation and maintenance. At this end of the risk spectrum, it is easy to accept risks. As long as there isn't any really serious consequence associated with failures in these systems, they should be optimized for life cycle cost and business efficiency.

Medium risk options

As risks increase, more demands are made on systems to assure the utility of content. For medium risk situations, many things are different. Sound change control and accreditation processes are necessary, configurations should be closely managed, and infrastructure supporting the application should fall under closer scrutiny and management. Probabilistic risk analysis may be used for natural threats, but covering approaches, protection posture assessments, and expert facilitated analysis are more suitable as the threats increase. While periodic oversight is acceptable at low threat levels, management must keep tighter reins and review at a higher rate for higher consequence systems or systems under more severe threats.

High risk options

When risks reach into the high end, systemic change management comes into play with system-wide testing associated with every significant change. Management rates increase until individual managers are in real-time control over the highest risk systems. Scenario-based analysis becomes increasingly important and, eventually at the highest risk levels, systems analysis becomes necessary. [5.9]

Matching surety to risk

Generally, higher certainty implies greater costs. So the desire to reduce costs has to be balanced with the desire to reduce risks. As a rule of thumb, as risks increase the certainty with which they should be mitigated should also increase. Thus the notion that surety should match risk. Different risk mitigation approaches have different surety levels. For example, separation and limited function can reach high surety, and some transforms can reach medium surety.[O.4] Surety, the certainty with which desired properties of systems can be known to be true, is a continuous range that is most often categorized in sets of levels for convenience. However; protective mechanisms are not continuous over a spectrum of alternatives, so it is necessary to make selections out of a finite set of options and those alternatives must usually be applied to last for a significant period of time to warrant their costs.

Because more certainty is usually desired for systems with higher risks, surety levels should usually be commensurate with risk levels. In other words, as a rule of thumb, everything that high valued operations depend on should be high surety, and as the value of the operations decreases, so can the surety of the technology that supports it. Surety can be continuously considered, however, many people find it easier to consider three surety levels.

Mitigation is non-continuous in that protective mechanisms must be selected and cannot, as a rule, be continuously tuned during operation to meet different surety levels or different cost constraints. There are discontinuous breaks between different mechanisms that force designers to use surety at the highest level anticipated or use lower surety than the worst case requires.

Risk matching would seem to imply that for high risks, only high surety mechanisms can be depended upon. While lower surety methods may also be applied to provide redundancy and additional delays, they will, ultimately, fail if attacked. Similarly, for medium risks, both high and medium surety methods may be applied, and for low risks, any or all of these techniques are applicable. Having said that, it is important to also understand that just because higher surety can be attained, doesn't mean that it always is. Separation can be done with poor quality and be less sure than transformation. Redundancy is also used to increase effective surety.

Low risks and surety

Benign environments that operate at acceptable performance levels but do not handle medium or high utility content are low risk. Typically, low risks are the everyday risks similar to those covered by normal business insurance. They are the normal accidents, errors, and omissions with consequences that are not worthy of additional attention beyond normal and prudent practices and due diligence. If a single low-risk system fails so that it never restarts, all data is lost, corrupted, or made available to the news media, and records of what happened are not available, it should have only negligible effect on the enterprise. Risk aggregation comes into play here because failures of large numbers of low risk systems may, in concert, produce more substantial risk. But as long as these systems are grouped and protected against common mode failures so that computer viruses or similar disruptions cannot aggregate to cause serious business consequences, they are still low risk.

Low surety systems are the typical personal computers and rapidly implemented networks that are often used throughout enterprises to support day-to-day low risk operations like writing letters and emails, preparing studies that aren't particularly sensitive, internal Web sites used for general information, and so forth. These systems can be purchased at retail or wholesale, assembled with minimal effort, and they cannot be trusted to protect integrity, availability, confidentiality, or provide accountability or use control under malicious attack. Low surety systems are the everyday computers available off-the-shelf, run with whatever operating system comes with them in their default configurations, and connect to the rest of the world via the Internet.

These systems vary in quality from a protection standpoint, and certainly this variance can impact operating efficiency, but all such systems are subject to attack and will eventually succumb to attack if the attacker is serious enough about getting to them and they are targeted. In day-to-day operation, they tend to be inundated with malicious code, become part of botnets, become infested with viruses, and so forth, and get cleaned up from time to time.

Typical protective mechanisms fall into the arena of standard features of operating environments in use plus matching known bad to detect patterns indicative of misuse. This pattern matching approach is inherently problematic in that is is easily defeated because detection of known bad is undecidable and even most of the relatively simple problems faced in detection are at least NP-complete.[5.5]

Medium risks and surety

Medium risks typically involve transaction processing, banking, manufacturing, infrastructure operations, and other similarly controlled environments. Many such systems operate in ISO 9000 certified environments and follow other standards for quality and security. Medium risks are substantial enough to cause a need for additional protection. They typically involve consequences that could cause significant implications to shareholder value or public well being, and as such are the sorts of things that trigger the need for mitigation or identification in annual reports under laws like Sarbanes-Oxley section 404. Consequences may include things like halting manufacturing or causing massive recalls, outages that damage the reputation of the enterprise, issues that produce large numbers of disgruntled customers, events that cause closures of facilities or layoffs of employees, and things that cause business plans to change.

Generally if the CFO or CEO has to get notified of significant protection failures, they fall, at least, into the medium risk arena. Medium surety systems are typified by the systems that run most manufacturing plants and many critical infrastructure systems. They tend to use stronger change control, for critical components they use programmable logic controllers (PLCs) rather than general purpose computers, they are not connected directly to larger networks, and they often have regulatory requirements for certification. If these systems go awry, there are generally fail safe mechanisms like lock outs and dead man switches that prevent the physical system from continuing to do potentially hazardous operations.

When low surety systems are used in medium risk environments, they are a serious hazard unless they are protected by additional safeguards, and as a result, they are generally guarded by PLCs or other similar mechanisms.

Medium surety systems are systems designed to do specific tasks well, engineered for the purpose, well tested under a variety of normal operating and exception conditions, kept under change control throughout their life cycle, well audited, and with additional coverage usually supplied by programmable logic controllers (PLCs) or other similar mechanisms that assure that ranges of acceptable values and conditions are met. They tend to have fail safe modes that they enter whenever conditions exceed identified parameters, limit changes to certain rates, are supported by proper administrative and procedural environments that keep them operating properly, and are maintained to assure high levels of availability. They are often in isolated or partially isolated environments, and in some cases regulatory requirements mandate certification and accreditation processes that must be repeated when significant changes take place. Medium surety systems are tested in all known failure modes so that their fail safe responses can be verified, they generally have extensive acceptance tests, and changes are made only based on change orders with the resulting changes similarly tested. Formal change and testing procedures are used, append-only media is used for auditing performance, and records are kept of every important action they take.

In communications and storage, cryptographic systems can reach medium levels of surety if applied between or within systems that are themselves medium surety level. Other related transforms, such as cryptographic checksums, can also be used to provide medium-level surety mechanisms and enhance surety when additional coverage is required or desired.[5.6]

High risks and surety

These are typically associated with high yield weapons systems, some space systems, controls for chemical plants with highly toxic materials in close proximity to people, aircraft control systems, and other life-critical systems. High risk is usually reserved for things that can cause loss of life, business failure, dramatic loss in shareholder value, significant harm to the environment, significant health problems, threats to public safety, and other things that are so important that they justify the extremes in effort associated with high surety systems.

High surety systems have very specific requirements for protection that warrant physical separation, redundancy in protective barriers, and special hardware designs for components. They are very expensive. High surety mechanisms sometimes augment medium surety controls with exotics. Multi-person controls for high risk operations increase human surety, special materials and other defensive measures may be used to create limitations on attack graphs, passive techniques tend to be preferred over active ones when they are as effective, separation mechanisms are effective repeated use protections, while one-time use mechanisms may allow for less separation.[5.7]

Reality check on risk matching

The description provided above discusses how risk management and assurance decisions should be made. The reality of how they are made is, in most cases, quite different. Most real enterprises do not apply these approaches, and many of those who try to do so fail to achieve the desired goals. There are many different reasons for these failures, almost all of them traceable to a lack of diligence, knowledge, and/or process.

- **Diligence failures:** Most people who work in the information protection area tell story after story of how others fail to take even the most simple, obvious, and rudimentary precautions and assume that everything is completely secure unless and until it turns out not to be. Most enterprises fail to even address risk management relating to information protection programs, and many who claim to do so, do only lip service to it. Low surety approaches are used almost everywhere.
- **Knowledge failures:** To quote the author from 1987, "In the information age, ignorance is not bliss, it is suicide."[5.8] The lack of knowledge among those who make risk management decisions for enterprises is stunning. As a field, those who have taught and learned about information technology and business management have failed to seed the knowledge required to make sound decisions. Ignorance of the issues is causing increasing failures with increasing consequences.
- **Process failures:** In cases when there are knowledgeable, diligent, well meaning people in the risk management process, the most common cause of failures seem to stem from processes that are not well enough defined or executed regularly. While these failures are almost certain to occur for some time to come, process failures in well-defined risk management programs tend to be smaller and less severe in consequence than failures from the other causes.

It is commonplace for low surety commercial off the shelf products to be rapidly adopted for use in medium risk environment. The low initial cost is deceptive and executive management seems to miss the point time and again. Unless and until this changes, large scale problems will remain and get worse with time.

Selection of approach

The key decisions in risk management are associated with the mix of risk acceptance, mitigation, transfer, and avoidance.

Risk acceptance is most common today. When risk management is not properly carried out, residual risk is almost always accepted by default. When proper risk management is undertaken, residual risk is quantified and understood by the decision makers.

Risk transfer typically involves insurance of some sort, but risk is indirectly transferred to shareholders when a risk is accepted, and contract terms can and are often used to transfer risks.

Risk mitigation involves the implementation of safeguards intended to reduce risk while leaving an acceptable or transferable residual risk.

Risk avoidance is rarely practiced today, but can be used when other alternatives are unacceptable. It usually means not pursuing risky business opportunities.

The risk mitigation approach be determined by a relatively simple process. Table 5-6 provides guidance by indicating situations under which risk should be accepted without further mitigation, transferred to insurance or some other party, reduced by protective measures, or avoided by not pursuing the business opportunity, depending on whether the risk is acceptable, transferrable, or reducible.

Accept	Transfer	Reduce	Action
No	No	No	Do not engage in this—avoid the risk
No	No	Yes	Propose reduction and re-evaluate
No	Yes	No	Insure or avoid the risk
No	Yes	Yes	Balance reduction with insurance cost
Yes	No	No	Accept or avoid the risk
Yes	No	Yes	Balance reduction x acceptance cost
Yes	Yes	No	Accept or avoid the risk
Yes	Yes	Yes	Balance all three and optimize

Table 5-6 - Risk management action table

A more complex analysis can be done by weighting acceptability, transferability, and reducibility, and applying metrics, but the cases where such analysis is helpful are quite rare.

Risk review rates

Risk management decisions should not be made once and then forgotten. Rather, the decisions regarding risks have to be periodically examined in order to assure that changes in the environment, the enterprise, the systems, the content, or the uses, remain well understood and properly managed. Review rates are generally associated with threat and consequence levels. Table 5-7 gives a typical example of the review rates and responsible parties based on the location in the {threat x consequence} space.

	Low	Medium Consequence	High
High	Should not occur - avoid	3-6-month review cycle, top management update quarterly	Continuous review - top management updates monthly
Med Threat	Mid-management update 9-12 months	3-9-month review cycle, top management update quarterly	Continuous review - top management updates monthly
Low	Mid-management update annually	6-month review cycle, top management update annually	Should not occur – threats are higher

Table 5-7 – How often what management level should review risks

In addition to the mandated periodic reviews, all substantial incidents should trigger reviews to assure that they are within the risk management profiles set for allowable incidents and incident rates. Vulnerabilities may also be uncovered over time or induced by changes, but if the process of risk management is properly done, changes such as these should not require reassessment, but rather should fall within identified tolerances. This implies a change management process that ultimately operates throughout the entire protection program.

Time frames provided here are typical of a well run organization with a well defined and efficiently operating risk management program. Annual reviews of low risk situations, for example, should cover all such systems, but the granularity of the review will not be as detailed as it will be for higher risk systems. Audits are also closely related to risk management schedules since audit results are typically used in verifying risk management profiles and decision criteria. The whole process underlying the risk management effort must also be operable at the same rate as the relevant parts of the program. [5.10]

Risk management questions

1. How are duties to protect turned into decisions about managing risks?
2. If no duty is associated with a particular identified risk, what actions should be taken with respect to that risk?
3. What is a useful and accurate definition of the term “risk”?
4. What are risks comprised of?
5. How does risk gets aggregated in different elements of the enterprise infrastructure?
6. How does risk get aggregated in people?
7. How does risk get aggregated in terms of geographic location?
8. What is a good example of a common mode failure that is prevalent in most enterprises today?
9. Why should vulnerabilities be kept to last when analyzing risks?
10. Why should consequences be considered first when understanding risks?
11. What is the difference between risk acceptance and risk transfer to the shareholders?
12. Why don't probabilistic risk analysis approaches work well for information protection related issues?
13. Who should be able to decide what to purchase and operate in order to mitigate risks?
14. Is someone with the purchasing authority to buy a product or service, also authorized to determine that the purchase is reasonable and appropriate from a risk standpoint?
15. When does top management have to get involved in risk management decisions relating to information technology?
16. How does the COSO process integrate with operational risk management issues?
17. How does risk management use the business model?
18. How does risk management feed the protection governance functions?

6 Information protection governance

The goal of enterprise information protection governance is to control information protection within the enterprise so as to make the overall program effective and efficient at meeting the business needs. Business needs are commonly defined by top management as including the need, without limitation, to:

- Meet due diligence requirements so that adequate care has been taken to protect shareholder value.
- Meet the level of care taken in comparable organizations, and be reasonable and prudent for the situation at hand.
- Meet regulatory requirements associated with laws like the Sarbanes-Oxley Act (SOX), financial and banking regulations, health information regulations, and privacy regulations.
- Meet industry-specific requirements like food and drug production and tracking controls, financial institution controls, and jurisdictional requirements such as privacy mandates and prohibitions against encryption of different countries.
- Meet contractual obligations such as those associated with regulatory mandated contracts for sharing of information, contract language associated with confidentiality and intellectual property, performance goals, and service level agreements.
- Fulfill business needs for integrity, availability, confidentiality, use control, and accountability for actions.
- Assure ongoing utility of content and systems even in the presence of malicious attacks that are widely present in the environment today.
- Address physical factors and disaster scenarios stemming from events such as weather, earth movement, tsunamis, explosions, and other natural and artificial threats to business continuity.
- React effectively to changes in the business environment, competitive threats, and changing worker profiles.
- Meet profit and loss objectives, control costs, and understand and demonstrate what the enterprise is getting for its money.

Governance creates internal business conditions that allow these needs to be met and balanced.

Fulfilling the duties to protect

While it is relatively easy to create a duty to protect, fulfilling the protective duties is often quite a more difficult task. At an enterprise level, without a systematic approach to identifying, codifying, and fulfilling these duties, they are almost certain to go unfulfilled in case after case. Even the simplest duties, like not making illegal copies of copyrighted software, often go unfulfilled because of a lack of a sound approach. [6.1]

One of the key factors in being able to fulfill the duty to protect is the presence of a protection architecture of the sort described in this text. With such an architecture comes a need to inventory and control information assets. With the inventory control system, it becomes feasible to identify and associate duties to protect with content and systems. With the ability to identify and associate duties, it becomes feasible to carry out those duties.

Specific methods used to carry out duties to protect depend on the duties. While relatively static approaches work for longstanding laws and regulations, a more dynamic approach is needed if a company is to deal with many complex and changing requirements.

As a simple technology example, a set of controls associated with access control might be handled by the integration of an identity management approach into the implementation of architectural elements. In such an approach, identities are associated with all protected assets (historically called objects) and rules and roles are associated with human and automated actors (historically called subjects) in terms of the ability of the subjects to act on (sometimes called perform functions with) the objects. The roles allow grouping of subjects based on job function or similar things, while the rules codify the rights that subjects have relative to objects. Those rights can then be translated into specific controls over specific actions within technical systems. [7.1.15c]

In the technology infrastructure, a provisioning system might configure access controls in a computer to allow a user to read a file, elements of public key infrastructure might be provisioned to provide copies of cryptographic keys for objects to specific processes acting on behalf of specific subjects at specific times and from specific places, and entry through doors to areas within facilities might be granted at specific times and places to facilitate the roles associated with current worker tasking. Similarly, audit requirements can be identified in the identity management system and provisioning processes can be used to create audit trails associated with accesses granted over periods of time while end systems and applications can provide audit trails for actions they take on behalf of actors.

But these sorts of technical controls are only the beginning of the issues at hand. Duties range from limits on the behaviors of people with respect to content, to requirements on availability of content for legal purposes, to requirements for privacy, to identifying who is authorized to do what, and on and on.

The vast majority of controls ultimately put in place involve more time, effort, and costs in non-technical arenas than in technical ones. And yet the technical attacks and defenses tend to get most of the focus of attention, perhaps because they seem so exotic and hard for non-technical people to deal with.

Limits on access to facilities, systems, and content can be subverted on a large scale if controls over the people operating identity management systems are inadequate. Hiring someone who, in the end, should not have been trusted, often causes far more harm than a technical attack by a stranger. A lack of documentation often translates in to an inability to continue critical operations when a key individual leaves or becomes destructive.

Whether malicious or accidental, historical losses associated with insiders (those authorized to perform actions) taking actions that they should not take exceed losses from all other causes combined. Accidents have caused spacecraft to fail, lack of defined and fulfilled duties have caused regional power outages, and insiders selling secrets have caused large enterprises to fail.

Governance is about translating all of the duties to protect into a comprehensive protection approach that deals with all of the facets of the information protection program at the enterprise level. Effective governance is required for an effective protection program because a chain is only as strong as its weakest link, and enterprise protection programs are only as strong as their governance and management processes.

What is governance?

Governance is the process by which government operates. This is comprised of structures, rules, power and influence, funding mechanisms, enforcement mechanisms, and appeals processes. Ultimately, for information protection to be effective, it must fit into the governance structure of the enterprise. The goal is to create and operate an effective, efficient, comprehensive information protection program within the context of the enterprise.

Governance structures and fitting in

The structure of the government and its relationship to the governed are fundamental to the way governance works.

- Hierarchical structures are quite common. They are used to leverage increasing numbers of people with decreasing power and influence, lower pay rates, increasing specialization and specialized expertise, fewer privileges, and more differentiated tasks at lower levels of the hierarchy, where they perform more tactical and less strategic tasks. Dictatorships, military groups, and many companies use hierarchy. Knowledge is controlled and propaganda or similar cultural control mechanisms are used to facilitate power and influence over large numbers of people.
- Networked organizations are structured with sets of key participants who take on leadership roles in select areas and many other participants who work independently but form a consensus that moves the group forward. Knowledge is widely available to anyone who wishes to seek it and strategy and tactics are developed by consensus. Pay and responsibility tend to be based on performance levels and infrastructure ownership. These organizations are often called “organic” in the way they operate, but most of them in fact have elite classes that communicate independently in cliques and use the network to their advantage by limiting access to information or selectively feeding information to the group as fits their desires. Sometimes juntas form in these sorts of groups, and these groups sometimes turn into hierarchies as size increases.

- Matrix organizations typically involve sets of leaders associated with different aspects of the functional need. There may be financial leaders, functional leaders, project leaders, line of business leaders, and so forth. Power is distributed and strategy and tactics are shared across groups that form for tasks. Matrix organizations without central leadership or strong management communications tend to produce schizophrenic overall behaviors as individuals are forced to serve multiple masters with differing and often contradictory demands.
- Hub and spoke structures are somewhat more rare and tend to be limited in size because of the critical role of the central leader. The central leader tends to be charismatic in nature for medium-sized organizations and may be a small business owner for smaller organizations. Power and finance are centralized and strategy and tactics are only shared as needed, typically all directed toward fulfilling the vision of the leader.
- Most governments, large organizations, and businesses tend to be composites of these structures if viewed in detail. These are “mixed structures”. For example, the government of the United States is a networked infrastructure at the topmost level with many hierarchies, hub and spoke, and matrix management structures at lower levels. The sharing of power is typically achieved by these mixed structures and each powerful individual at any level of the organization tends to build the structure that they are most familiar with or that they think is most appropriate to their needs and business function.

Each of these structures has particular processes that work more and less efficiently within them, they have advantages and disadvantages, and they are all suitable to different situations. The structure of the enterprise and its components necessarily dictates the structure of the overall information protection program and the manner in which those tasked with governing the process are able to influence the protection posture, measure results, fund the effort, and deal with objections.

Within enterprises, the most common structure is a hierarchy, and this is largely dictated by legal and regulatory mandates, at least as far as the top level of the enterprise goes. Within the hierarchy, there are many different arrangements usually depending on how the business developed over time and executive decisions.

Fitting protection into business structures

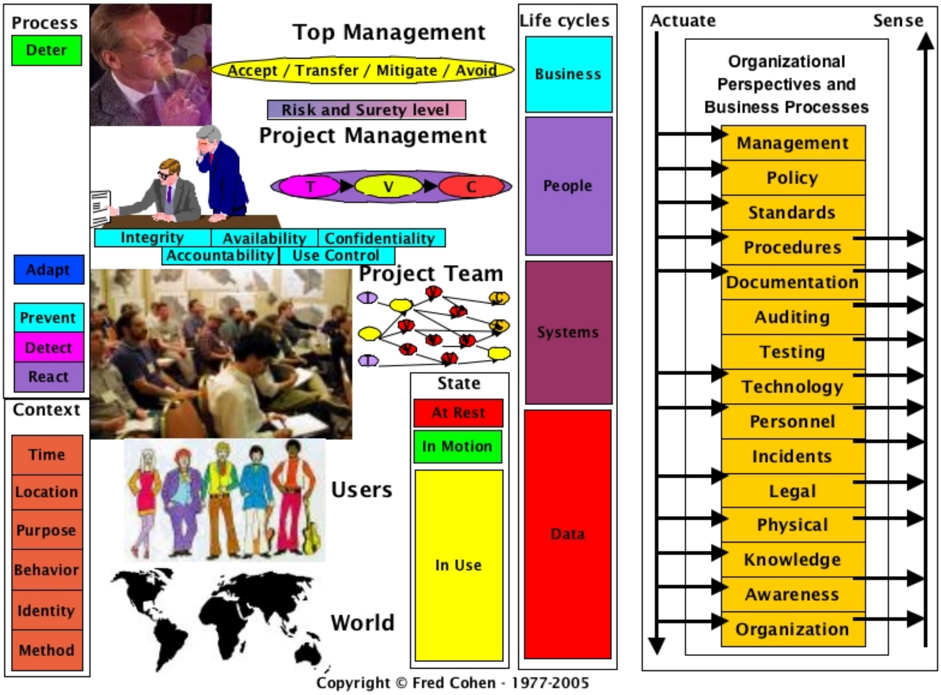
Effective information protection programs are structured for the businesses they serve because this tends to minimize friction and result in more effective control. Most businesses are either oriented toward a market segment with centralized top management and a hierarchy, or are a family of companies with a small central management group and independent internal management in each of the operating entities. Mixes exist where large divisions act more independently than smaller ones with some business functions, like corporate communications, payroll, accounting, human resources (HR), and information services organized as central service organizations. Facilities may have both distributed and centralized elements, and many enterprises divide along national boundaries for legal and regulatory compliance reasons.

Regardless of the overall corporate structure, information protection has to fit into it, and as a rule of thumb, information protection should not be significantly different from other existing corporate-wide functions. If a brand new business model is used for information protection it will likely cause friction because integration with the rest of the enterprise will be unnecessarily complicated. But while integration of information protection into overall business operation is critical to success, the many touch points and need to integrate across a wide breadth of functions, business units, and levels, makes it unique in many ways.

The most common approach is to treat information protection in a manner similar to other crosscutting business functions, like finance and accounting (F&A), HR, or corporate communications. Look at the information protection issues and identify all of the touch points within the enterprise. Consider all of the elements in the picture on the cover and how each of these interacts with the enterprise. Seen through this lens, the enterprise governance picture for information protection may be clarified.

Organizational perspectives and functions

Figure 6-1 shows the linkage between the overall picture of enterprise information protection and a typical enterprise hierarchy. Top management makes different decisions than project and people management and users tend to interact with different mechanisms and aspects of the program.

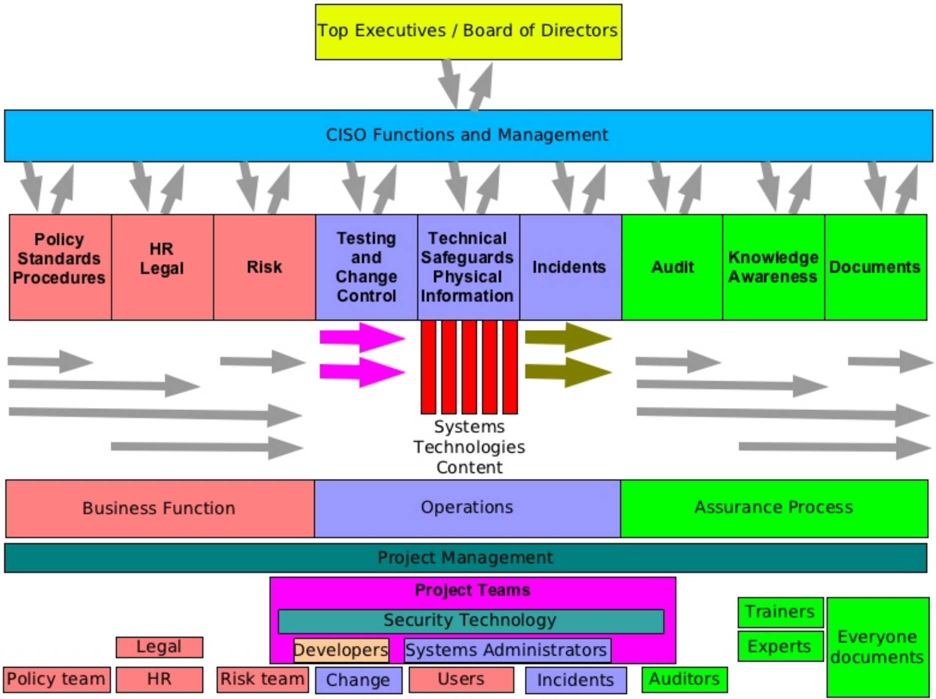


Copyright © Fred Cohen - 1977-2005

Figure 6-1 – The typical enterprise governance structure

Responsibilities for risk management and surety levels lie with top management. Financial risk management is often carried out by a team in the chief financial officer's (CFO's) office, but overall risk management may be undertaken by the audit committee, from within the chief counsel, by a chief risk officer (CRO), by the Chief Executive Officer's (CEO's) office, or by the Board of Directors. Information technology (IT) risk management may be separated from corporate risk management and held within the chief information security officer's (CISO's) office. If there is a separation between corporate and IT risk management, they need to be closely coordinated in order to be effective.

Business life cycles and deterrence are also management responsibilities. For business life cycles, business acquisition teams should include representation from the CISO function. Deterrence depends on positioning of the enterprise, decisions about when to prosecute, policy issues, and so forth. Top management also sets policy, structures protection program management, and defines the placement of information protection by positioning the CISO within the company and defining the linkage between the CISO and HR, legal, the CIO, and others.



Copyright © Fred Cohen - 1986-2005

Figure 6-2 – Enterprise security management architecture

Figure 6-2 shows the overall control system that operates information protection, typically managed by the CISO. It consists of increasingly detail-oriented groups that operate decreasing subsets of the enterprise. The top executives and board of directors control the functions and management associated with the CISO, regardless of the implementation of the function, its organizational location, or the management structures used to implement it. The CISO functional responsibilities include:

- Business functions which include policies, standards, procedures, legal, HR, and risk management activities and involve the policy team, the legal department, the HR department, the risk management team, the users, and some of the project team and developers;
- Operations which includes testing and change control, physical and informational technical safeguards, and incident handling activities and involve the developers, systems administrators, change control team, response team, and project teams; and
- Assurance process includes auditing processes, knowledge, and awareness programs, and documentation functions and involves auditors, trainers, experts, and project teams, and of course everyone has to document what they do.

Project management activities span the entire spectrum of the CISO function, while different groups of people tend to work in different areas associated with the effort. Separation of duties requirements, skill sets, organizational mandates, and other issues result in different groups operating in different parts of the organizational spectrum.

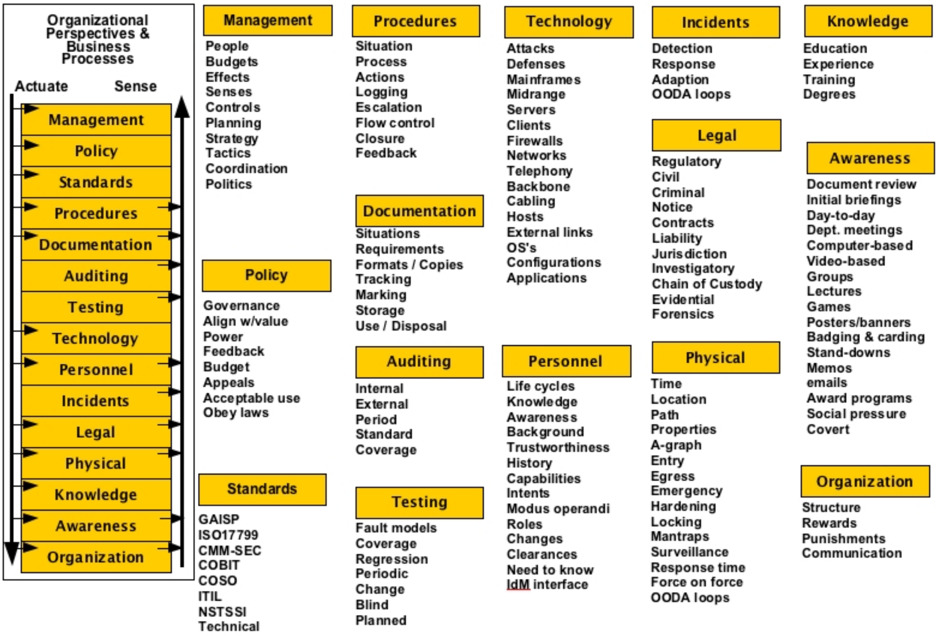
There is also a general flow of information that runs from policy, standards, and procedures through documentation, with functions on the left tending to push more influence toward the functions on the right. Feedback mechanisms lead to adaptations through the control efforts associated with the CISO function.

The most critical function and the purpose for the CISO function is to exert the controls that influence all of the different protection-related functions and to listen to the feedback and make decisions that help to adapt the overall enterprise protection system based on the feedback.

In order to carry out this function, the CISO function has to also be able to communicate effectively with top management or whoever is ultimately responsible on an enterprise level for the proper operation of the business. The CISO function has to fuse together all of the feedback from all of the diverse sources, present the results to top decisions makers, and explain how the controls are working or need to be changed in order for the protection function of the business to operate properly.

From an organizational perspective, the program involves a broad range of activities and contexts. Issues arise in these different contexts. While more in-depth information on each of these areas is necessary in order to carry out the more detailed work associated with them, from a CISO point of view, these details need to be covered at the next level of management. Results are analyzed and rolled-up into reports included as part of the feedback process.

Figure 6-3 shows details of organizational perspectives typically addressed in effective information protection programs. In the following sections, the items listed under each area are explored in a bit more depth. [6.2]



Copyright © Fred Cohen - 1986-2005

Figure 6-3 – Drill-down into organizational perspectives

Management

Management of the protection program is typically handled by the CISO who helps to build management practices and coordinates management of information protection throughout the enterprise. It includes:

People who have to be trained, made aware, tracked, and managed

Budgets that have to be generated, justified, and used wisely.

Effects created by actuators that allow the CISO and others to influence events that take place.

Sensor outputs from sensors including automation, people, and groups are reported to the CISO for situation awareness.

Controls are formed from sensors and actuators. As a feedback system, the CISO uses technologies, procedures, processes, and other things within direct and indirect power and influence to control the process.

Planning is required to cause the complex event sequences involving people and systems to be properly coordinated.

Strategy that translates the long-term vision of the enterprise and the CISO into the plans that result in achieving those long-term goals.

Tactics that provide the short-term event sequences that produce the functional behaviors desired in specific situations.

Coordination that assures that the tactics as implemented remain within the desired set of future event sequences.

Politics that form the basis of the interactions between people throughout the enterprise and, if successfully applied, allow the CISO to control the situation without creating unnecessary friction.

Policy

Policy codifies the intent of the enterprise starting at the top level.

Governance implies the system under which power and influence operate. These are the processes that take place within the enterprise, its institutions, and its structures to allow those in charge to govern.

Alignment with value indicates the need for the policy to tie the value of content to the enterprise to the time, effort, and cost associated with protective functions.

Power issues are codified in policy by granting individuals and groups control over resources, actions, titles, and other influential items.

Feedback provides the means by which policy may be used to close the control loop at the top level of the enterprise.

Budget and the process by which policy dictates that budget is generated and managed to provide the means of controlling critical fungible resources.

Appeals processes define the manner in which policies and decisions made by those granted power and influence under policy may be challenged.

Acceptable use identifies what is and is not acceptable in the use of corporate resources.

Obeying laws is typically codified in policy to assure that employees do not go astray under the corporate banner. It also provides the necessary mandate to provide the necessary resources and knowledge to employees to prevent ignorance as an asserted excuse.

Standards

Standards include specific enterprise control standards that codify policy in more detail and, at the CISO's level, commonly accepted and practiced approaches to information protection that codify due diligence and reasonable and prudent approaches. [6.3]

GAISP are the Generally Accepted Information Security Principles evolved from the Generally Accepted System Security Principles (**GASSP**). These principles are fundamental top-level issues that are key to effective information protection and should be codified as appropriate in every information protection program. They were developed by a committee within the Information Systems Security Association (ISSA), a not-for-profit, international organization of professionals and practitioners. The primary goal of the ISSA is to promote management practices to ensure confidentiality, integrity, and availability of information resources.

ISO 27001 and 27002 are the international standards organization's adaptation of the British Standard BS7799 and as updated from ISO17799. They define issues at two levels deeper than GAISP and codify the most common issues identified by companies in their implementation of information protection. They are designed so that management has the option of determining what to do and to what extent it should be done. Audits against these standards generally adopt the notion that all of the elements must be done to a reasonable and prudent extent based on the situation in the enterprise. These are the most commonly applied information security standards in the world today.

CMM is the capability maturity model. The security interpretation of CMM (CMM-SEC) codifies the maturity level of a security engineering capability. Variations are very useful as management tools because they codifies capabilities from a standpoint of how effectively they are managed. CMM-SEC is not a formal standard. Rather, it is the best codification of these issues available and has utility for the CISO. It differentiates 6 levels of maturity; (0) none, (1) initial, (2) repeatable, (3) defined, (4) managed, and (5) optimizing.

CoBit was developed as a generally applicable and accepted standard for good information technology (IT) security and control practices. It provides a framework for users, management, and IT audit, control and security practitioners. CoBit is sponsored by the Information Systems Audit and Control Association (ISACA) and deals with information criteria, IT processes, and IT resources. Information criteria include quality, security, and fiduciary duties. IT process deals with domains, processes, and activities. IT resources deals with people, applications, technology, facilities, and data. It is used predominantly by auditors and is the global gold standard for audits of information technology security programs.

COSO is the risk management standard created by the Committee Of Sponsoring Organizations of the Treadway Commission. COSO is the regulatory preferred framework for Sarbanes-Oxley section 404 (SOX 404) risk management as specified by the regulators in the United States. As such it applies to all public companies in the US. COSO is also widely embraced around the world. The COSO cube gives an overview of COSO. SOX 404 and COSO say nothing about information technology. COSO has 3 dimensions that are different from those of CoBit. One is organizational, consisting of entity-level, division, business unit, and subsidiary, which really indicates a desire to cover the enterprise at all levels. The second dimension is a collection of perspectives consisting of strategic, operations, reporting, and compliance. The third dimension covers the internal environment, which includes objective setting, event identification, risk assessment, risk response, control activities, information and communication, and monitoring.

The Information Technology Infrastructure Library (**ITIL**) provides a "cohesive set of best practice" to implement British Standard Institute (BSI) standard for service management (BS15000). It includes a guide on implementing security in service level agreements (SLAs).

The National Security Telecommunications System Security Initiative (**NSTSSI**) provides a set of national standards (numbers 4011 and higher) for training personnel who have responsibility for creating, operating, approving, and overseeing secure (classified) systems. This includes information security professionals, designated approving authorities, systems administrators, and system certifiers.

The U.S. National Institute of Standards and Technology (**NIST**) provides a collection of standards in their 800- series of standards that are open for anyone to use and applied within the US government. These standards are quite good in what they cover and are largely equivalent to other similar standards such as those above. They cover everything from risk management to technical safeguards and how to configure specific operating environments.

There are also many related standards such as; the ISO9000 series of standards for quality, which are closely related to the sort of requirements necessary for medium and high surety protection programs; the OECD guidelines for the security of information systems; accounting standards like Generally Accepted Accounting Practices (GAAP); the control requirements of the Health Information Portability and Accountability Act (HIPAA), the standards associated with the Gramm Leech Bliley Act, the payment card industry (PCI) standard for protection of credit and debit card information, and the standards associated with SAS-70 audits used to provide adequate assurance that financial institutions can exchange data with each other.

Technical standards exist for many different aspects of information protection and, as a rule of thumb, it is better to go with standards-based solutions than non-standard solutions because of the complexity of interfacing non-standard systems and capabilities to other systems and capabilities.

Control standards

Enterprises also have and develop their own **control standards** that codify policies to provide detailed guidance for implementation. These control standards act as a bridge between policies and procedures by providing specifics on how to implement policy without delving into the details of systems. They typically codify one or more set of requirements and are increasingly structured to correlate with known standards such as ISO 27002 so that they can be easily mapped into external requirements, applied to demonstrate compliance, meet with audit standards, and otherwise facilitate reduction in effort across multiple competing needs for documentation, verification, and implementation.

Procedures

Procedures implement control standards in specific systems and contexts by creating systematic step-by-step processes that, if properly followed, result in meeting those standards. Procedures codify the processes at the lowest level of implementation and typically generate documentation associated with every step.

Situation leads to process. As a result, most procedures have sets of preconditions for their invocation. They can, in many cases, be codified in a work flow system that identifies the conditions and triggers the actions associated with that condition.

Process is carried out through situation-specific actions that get logged, escalation conditions, control over process flows, feedback loops within and outside of each process to assure quality, and termination conditions that cause processes to formally terminate.

Actions associated with procedures are typically designed to result in some set of specific outcomes.

Logging is used to assure that a record is made of what took place and to allow after-action analysis and reporting for process evaluation and improvement. Logging processes also produce the documentation necessary for legal and other review purposes.

Escalation is typically the result of an exception that is codified as part of overarching procedures in order to assure continued process control even when the process gets out of the predefined control scheme.

Flow control mechanisms typically assure that work is performed in order and that results are checked along the way. Approval processes may also be involved. These are often implemented in ticketing systems or similar control (e.g., work flow) mechanisms. Some flows allow limited parallelism.

Closure is the result of the process reaching a conclusion. Ticketing or other work flow systems indicate that no further work is pending on the process.

Feedback occurs at all levels, from the process components that lead to situational changes dictating further actions, to the overall feedback that improves processes by after-action reports, and other more strategic reviews.

Documentation

Documentation is created throughout the overall information protection process, creating a need to capture and protect that documentation and to use it for investigation, analysis, and other legal and business purposes.

Situations dictate the need for documentation of different sorts. While novel situations may require unique documentation, most situations are recurrent. As they recur, the documentation processes and formats become standardized.

Requirements documents are used to describe what is required for systems when implemented. There are typically specific requirements for different purposes and those requirements are themselves documented to formalize the documentation process.

Formats associated with documents become standardized as the situations leading to them recur. These formats lead to implementation in databases for more systematic analysis and process improvement. Formats also apply to marking and tracking processes.

Copies of documentation are easy to make and should often be tracked. Copies are good for availability but potentially harmful for confidentiality and may be legally restricted.

Tracking takes place at many levels. Documents of certain sorts, such as limited access documents, protected health information, financial records of certain sorts, trade secrets, and classified documents must be tracked throughout their life cycle and the tracking itself is a form of documentation. Large-scale document tracking systems are also vital to retaining and searching large numbers of corporate records, assuring that other processes are carried out at the proper time, and being able to demonstrate that the document control process is operating properly.

Marking is commonly used to allow inspection to identify document types and control information. Markings are required for certain documents, including documents with intellectual property value. Marking is also the basis for much of the automated and manual process associated with document control.

Storage becomes complex and problematic for large numbers of documents, especially mixed combinations of paper, fiche, other physical media, and electronic documents that are interrelated. Tracking systems are helpful in locating and retrieving stored documents as well as determining when storage must be refreshed. Storage also involves environmental controls that are specific to specific media.

Use of documents involves a variety of control issues including access control, application control, protection from corruption, continued availability, how drafts are treated, and so forth.

Retention and Disposition (a.k.a. Disposal) is a key issue that is problematic for many enterprises. The failure to properly dispose of waste is one of the most common faults detected in penetration tests and the results are sometimes dramatic. Documentation of disposal of documents and assurances associated with retention requirements as they relate to disposal are also important to meet legal obligations. [6.4]

Auditing

Auditing provides the means for management to verify the proper operation of the information protection program.

Internal audit processes assure that operations meet internal requirements. This typically involves audit staff and a cyclical process that assures that high valued systems are revisited often while lower valued systems are covered consistent with their value.

External audit processes act as independent verifications that operations are as they are supposed to be and also act to assure that internal audit is effectively doing its job.

Periodicity for audits is a nontrivial matter with audit periods determined by risks, costs, resources, and time and cost to audit. Random audits, surprise audits, regular audits, and other time-related issues all fall under this broad category.

Standards are typically what audits compare realities to. Auditors are generally tasked with relating performance to a standard so that a consistent basis for opinions can be used and comparisons can be done over time and between systems and organizations. It is normal to use the same standards for protection as are used for audit so that the audit provides reconcilable feedback on the adequacy of the program in meeting the standards set for it.

Coverage expresses the extent to which audit processes cover the set of things that could possibly be checked in an audit. It acts as a metric on the audit itself as well as a means to evaluate the value of the audit. An audit that is passed but only covers an unimportant subset of the issues or systems at hand is not a very good reflection of the situation and has little utility.

Testing and change control

Protection testing provides verification that protection does what it is supposed to do.[6.4b] It involves the following issues:

Fault models are basic phenomenological models of the sorts of faults that occur and how they are manifested to the observer.

Coverage is expressed as a numerical value indicating the percentage of the totality of event sequences covered by the testing regimen relative to the fault model.

Regression testing requires that tests against historical weaknesses are used to verify that things that used to be problems do not recur.

Periodic testing requires that the enterprise define periods between tests based on system factors like criticality, rate of change, complexity, and so forth.

Change generates a requirement for regression testing and new test development associated with the changed functionality. Levels of change that require testing must be defined based on the criticality of the systems under test and the nature of the changes. This integrates with the change management system to form a systematic change control process. A sound change control process is preferred for medium risk systems and mandatory for high risk systems. A sound change control approach is outlined under change control architecture in the "Control Architecture" chapter of this book.

Blind testing is testing wherein the individuals operating the system under test are unaware of that a test is underway. A methodology for carrying out such tests is required.

Planned tests also need well defined circumstances and performance requirements.

Technical safeguards - information technology

Risk mitigation often involves technologies.[6.5] Implementing and operating those technologies takes significant effort and expertise.

Attacks are typically codified as sequences of events carried out by a threat. Attack processes are described elsewhere.

Defenses are measures that act to (1) reduce threats, (2) reduce the ability of threats to find and exploit vulnerabilities, (3) reduce the number of vulnerabilities and control their nature, type, and location, (4) reduce the linkage between exploitations of vulnerabilities and consequences, and (5) reduce consequences.

Different technical safeguards apply to different **platforms** and **environments**. The following sets of platforms and environments often have to be considered for technical safeguards:

- **Mainframes** typically have access controls based on user identity involving a subject/object model. They are centrally controlled for large application environment with sound change control over program changes, strong and standardized audit components, and limited user interfaces. They may also house database systems that have query limits. They often have redundant system capabilities, and they tend to use separation of duties.

- **Midrange** computers handle local operations, control production, distribution, and other specialized systems, and have protection similar to mainframes, on a smaller scale.
- **Servers** may be mainframes, midrange computers, or other types of devices, but increasingly they are run on small system platforms with Linux, Unix, or Windows operating systems. They have protection associated with the operating systems. Many have power and disk redundancy, audit controls, and query limits. Some use separation of duties and are change controlled, most have access control in one form or another, and many interact with identity management infrastructure for access control. Applications in those servers may have additional controls.
- **Clients** are typically single user systems, usually run a low surety operating system, and usually have minimal controls. They form the largest set of platforms, tend to be vulnerable, are often poorly managed, and are subject to all sorts of attacks ranging from viruses and worms to Trojan horses. Thin client platforms with strong controls are less expensive alternatives but they tend to be less popular today.
- **Firewalls** are network separation devices and may be implemented as “firewall routers”, as separate components in a perimeter architecture, as software components in end systems, or as separation devices between enclaves or network zones. Firewalls generally have a network interface control capability that differentiates and controls the flow of packets based on source, destination, port, protocol, or content. They may act in concert with or contain proxy devices or similar technologies, may terminate encrypted tunnels, may do load balancing, and may perform a wide range of other audit and control functions.

- **Networks** provide for transport of data. On their own they typically have little in the way of protection other than adequate bandwidth and redundancy to handle expected load levels and survive certain sorts of outages. Networks may have bulk level encryption but this only prevents physical attack on infrastructure from revealing content and provides little or no other protection.
- **Telephony** systems are used to transport data and voice and increasingly are integrated with networks in Voice over Internet Protocol (VoIP) systems. They typically either have large connectivity and can be used to transport arbitrary data or are connected to systems to bridge otherwise secured networks and provide dynamic connectivity. Dedicated systems have very different architectures and require very different protection than other technologies in common use.
- **Backbones** are used to carry large volumes of data between main switching or routing centers. They tend to aggregate a lot of different content and form high valued targets, but they also tend to have little or no protection other than through physical location and cable security.
- **Cabling** provides the media that carries data throughout the enterprise. It typically has physical plant, runs through cable runs of various sorts, and has to be physically secured in order to prevent ready and potentially high-consequence exploitation. External cabling tends to pass out of the facility boundaries and extend to remote sites through paths not physically securable by the enterprise.
- **Hosts** are used for many purposes, like user workstations in development environments, control systems for production facilities, personal data assistants and cellular telephones with computing capabilities, and so forth. Each has unique protection requirements and capabilities.
- **External links** connect the enterprise to the world and present both a path for exploitation and great benefit to the enterprise.

- **Operating systems** have different protective mechanisms designed for sets of intended uses and are commonly extended well beyond the original design purpose.
- **Control systems** are used to control everything from elevators to HVAC to manufacturing systems. They also have indirect effects on other people and systems.
- **Databases** store, retrieve, and search content.[6.11]

Configurations are used to customize systems and platforms. Control over configurations is critical to the protection function.

Applications operate in all of these environments and constitute the purpose for their existence and the basis for their utility. These applications necessarily communicate and these communications must be appropriately protected to protect the utility of the application and related content.

Other technical safeguards include a wide range of technologies that are outlined briefly later under the chapter on Technical Security Architecture.

Personnel

Personnel security issues focus on people involved in protection process and verification that they meet the necessary and appropriate standards and qualifications required for their duties. [6.6]

Life cycles associated with personnel are described in detail under life cycles. They generally involve conception, pregnancy, birth, education, marriage, divorce, training, hiring, promotion, demotion, suspension, vacation, illnesses, leaves, job changes, moves, resignation, termination, retirement, death, and legacy issues. All of these interact with information protection issues in one way or another.

Knowledge associated with personnel helps to determine qualifications and suitability for tasks and jobs. Knowledge tends to be tracked to degrees and related programs, job history, and defined areas of expertise within the enterprise. Advanced degree programs tend to be reimbursed by the company if job-related and these are also tracked in the enterprise.

Awareness levels in defined areas should be tracked to assure that all personnel have appropriate awareness of key issues associated with their job functions and that those who are not properly qualified and aware are not permitted to do things that require that level of awareness. At a minimum, security awareness programs have to touch each individual in an enterprise every 6 months to be effective at keeping levels high enough for effect.

Background checks on all employees should be required in almost all cases. The cost can be as low as $20 for simple criminal record checks and fairly extensive background checks can cost as little as $150. More extensive checks should be made on those with higher levels of responsibility. For some positions, such as those involving classified information or specific interactions with children, detailed background checks are required.

Trustworthiness is hard to assess, but trust is often granted based on limited experience. Many of the least trustworthy people are the most trusted because professional confidence operators are very skilled at displaying the things that generate trust even though it is not deserved. Many companies place excessive trust in insiders and suffer the consequences. A systematic approach to evaluation of trust, including time in position and life-related characteristics is more effective at predicting trust-related behavior than non-measurable qualities associated with personal friendships and liking.

History is often cited as the best predictor of future performance. Background checks and detailed information from personnel records and references tends to produce historical information about personnel that helps make reasonable and prudent decisions in this space. Missing history information on individuals in personnel records is a strong indicator of potential abuses of the system and should lead to detailed investigations.

Capabilities associated with individuals help lead to their assignment to suitable tasks. Specific individuals have special talents or training that produces capabilities that are unusual or hard to train or find. These should be identified for specific information protection tasking.

Intents are more difficult to understand than capabilities. However, indicated intents are often provided in letters, writings, and similar materials and should generally be explored as indicative of likely behaviors. Group memberships and similar factors tend to indicate intent, particularly in groups with widely declared intents such as animal rights groups, ecological groups, and so forth.

Modus operandi is typically associated with criminal behavior, but all people display methods of operation that tend to be reproduced over time. This is useful as an indicator for future tracking and attribution as well as for understanding how likely interactions will take place and be received.

Roles are typically associated with groups of individuals and individuals may be associated with many roles, depending on their tasking within the enterprise. These roles are then translated into authorizations associated with functions on systems. People are moved from role to role as they move from job to job, with the roles refilled for operational continuity.

Changes of employment status, job title, responsibilities, and so forth are all issues that involve information protection functions such as access to systems. Change tracking for personnel and integration into accounts in information systems, access passes, and so forth are critical to effective protection.

Clearances are generally associated with individuals. These are generated through formal processes, screened by authorized screeners, and tracked and maintained by personnel systems. Clearances reflect levels of trust relative to applicable standards.

Need to know information relates to specific work areas and projects. This too is tracked by personnel-related records and must be protected to guard projects against systematic exploitation of associated individuals.

Identity management (IdM) interfaces provide for interactions between the identity management system and personnel, systems, and others tasked with making decisions about individual access. They are typically integrated with personnel systems to assure that records are up to date with authoritative sources.

Incident handling

Incident handling encompasses everything from incident detection and response to disaster recover and business continuity management.[6.7] The goal of incident handling is to detect all event sequences that have potentially serious negative consequences in time to mitigate the consequences to within acceptable loss levels.

Detection is central to any incident handling effort. Incidents that are not detected are not handled. Resulting consequences go unmitigated. While some see this as being relatively unimportant because undetected incidents cannot be all that harmful, history shows that in these cases detection may eventually become clear through extreme consequences that cannot be mitigated by that time. All detection schemes are subject to potentially unlimited numbers of false positives (false alarms) and false negatives (missed alarms). There is a trade off between these and the numbers of alerts can be controlled to meet the available response resources. Usually a thresholding scheme is used to differentiate and rank alerts selected from all detections so that records may show the presence of detections that did not cause alerts but were important to understanding what happened in an investigative process. Most detection systems are not properly designed to meet the enterprise need. They are designed for technical purposes based on available data. A properly designed detection system should detect event sequences that can lead to potentially serious negative consequences and rank those event sequences as they occur by consequences and response timeliness requirements.

Response systems are also very complex. For example, automated responses can be exploited for reflexive control attacks. Human intervention can be easily overwhelmed thus producing a change in thresholds of detection and reaction, leading to serious attacks getting in “under the radar”. Like detection, most response systems are designed for technical response and not oriented toward the needs of the enterprise. An effective system produces responses that mitigate serious negative event sequences by blocking them before the consequences exceed acceptable thresholds. Large-scale responses such as those required to mitigate harm in disasters or when business continuity plans must be invoked are disruptive and costly so they are typically invoked only under well defined circumstances and controlled by a well practiced plan operated by practiced personnel.

Adaptation is the long-term response to incidents that seeks to optimize enterprise performance by strategic changes not related to specific incidents but rather oriented toward changing the way classes of incidents are handled. A good example is the adoption of network zoning. While tactical responses address stopping a current virus and cleanup operations, a strategic response to set up differentiated and separated network zones prevents large-scale worms and viruses from producing the most serious adverse consequences and inherently limits their spread and effect. These changes also reduce other risks without high costs. They are architectural adaptations to the environment in response to incidents in the large.

OODA loops (observe, orient, decide, act), also known as the Boyd cycle, or similar processes exist in all detection and response systems. Observation and orientation are typically associated with the detection problem, which, in classical control theory characterizations may be called detection and differentiation. Decisions and actions form the response processes. A basic idea behind the use of Boyd cycles is that the OODA loop takes time. In conflict situations, a faster OODA loop can make the difference between winning and losing. In incident handling there are many levels of Boyd cycles, from cycles in the time frames of seconds associated with beating the spread rates of network worms, to cycle times in the time frames of years associated with adaptation processes reflected in new network architectures. Issues of timing, sensor placement and design, communications infrastructure, analytical power and technique, and actuator placement and design are all intimately tied into the Boyd cycle.

More details are provided under the protection process area within the Technical Security Architecture chapter.

Legal issues

Legal issues range from the inclusion of proper language in contracts to fulfillment of regulatory compliance requirements for attestation.

Regulatory drivers impact all corporations. Whether your enterprise has EU privacy requirements, US financial reporting requirements, US, Canadian, or Australian health and benefits information requirements, Chinese and French encryption requirements, or other similar requirements, regulatory drivers are increasingly forcing changes in information protection programs.

Civil litigation drives many enterprises in legal areas. A good example of a protection policy that resulted in a lost civil suit comes from a recent case in which a published Web site policy guaranteed privacy of personal information. The policy was not followed and a million dollar law suit was lost as a result. If there were no such policy there would have been no such loss.

Criminal litigation is pending against many executives who failed to report to shareholders on potentially serious negative consequences associated with information technology failures, inadequate assurance associated with financial records, and other similar violations of law. Failures of due diligence are increasingly being treated severely because of prior executive misdeeds.

Notice is required for legal protections to be effective. Good examples are trade secret, telecommunications recording, and worker monitoring notice requirements.

Contracts with inadequate language related to information protection are widespread and result in a wide range of problems, particularly associated with access into enterprise networks used for trading partners. Customer contracts relating to records are similarly problematic. Peering agreements associated with financial and health-related information require a level of due diligence in their perfection. Safe harbor agreements and other similar contracts require that protections be in place and effective. Many existing contracts should be updated to reflect the need to include encryption, access controls, and other protective measures in storage, movement, and use of exchanged information.

Liability issues associated with holding information of certain types, operating systems that interact with third parties, actions of employees with respect to intellectual property, and similar information protection issues are widespread. Even an infection with a computer virus may lead to liability issues associated with the lack of due diligence in protecting peering partners from the infection. Break-ins to unpatched or unnecessarily vulnerable systems at perimeters may lead to liabilities associated with consequential damages to downstream providers and others attacked from your site.

Jurisdiction is a critical issue for large multinationals, however, because of the global reach of the Internet, most businesses are now international. Attacks, scams, and legal processes associated with individuals around the world are commonplace in today's information environment. A business with a Web site has presence everywhere in the world, and sales to foreign nations may result in violations of laws that the seller or buyer are not familiar with. Jurisdictions affect legal issues across the board and mandate a dramatically more complex information protection program than would otherwise be needed.

Investigative processes are linked to legal proceedings including but not limited to legal issues associated with employee sanctions, employee rights in investigative processes, prosecutions associated with criminal acts, civil proceedings related to employee misdeeds, and many other similar types of issues.

Chain of custody issues must be addressed in processes that could ultimately lead to the introduction of evidence in court. While the business record exception in the United States generally provides for these records, other jurisdictions have varying requirements for chain of custody. Records retention processes increasingly require chain of custody to be maintained in order to assure integrity of records and prevent loss of critical information that must be retained in case requested by authorities.

Evidential issues come up whenever information protection issues end up in legal venues. The data presented has to have adequate integrity and accuracy to assure that it can be accepted by the courts and it has to be presented by an expert who is responsible for those records and can attest to how they came to be and what they are supposed to represent. They have to be normal business records to be admissible under the hearsay exception, and as a result, they must be collected in the normal course of business. Preservation orders may require that records be retained beyond their normal life cycles for evidential purposes and these orders must be followed in order to avoid criminal legal sanctions associated with obstruction of justice and disobeying judicial orders.

Forensics efforts associated with identification, collection, preservation, analysis, and presentation of evidence in court require special training and expertise and are involved in almost all investigations associated with information protection issues. [6.8]

Technical safeguards - physical security

Physical security is typically handled by the chief security officer or other individuals responsible for these issues, however, protection of content, technology, and systems at the physical level requires special expertise and is critical to effective protection of the enterprise. Physical security is also critical for health, safety, and protection of the environment. [6.9]

Time has long been a central issue in physical protection and is increasingly becoming a central issue in information protection. Actions take time, whether in attack or defense, and physical security has long recognized this in the design and operation of alarm systems and response regimens. Typically time is measured against attack graphs.

Location is central to physical security issues. Different locations have different situational characteristics, such as proximity to natural hazards like earthquakes, tsunamis, volcanoes, hurricanes, floods, lightning strikes, dust, cold, heat, and so forth. Human hazards are also associated with location, like crime levels in different neighborhoods, cities, states, nations, and continents. Even the location inside office spaces leads to higher or lower profile and susceptibility to attack.

Paths from the initial situation of the attacker to their target and back to safety have various limitations, like topological limits, time to penetrate barriers, equipment and skill requirements, and the number of different ways in and out of areas with and without detection and response. Paths are altered by diversions and other active attacks and defenses.

Properties associated with materials, barriers, and entry and exit processes have substantial effects on available physical attack processes, time to penetrate, noise levels, detectability, and so forth.

Attack graphs express the set of sequences of steps in physical attacks. They are used by attackers and defenders to determine options for entry and egress (exit) on a step by step basis through the successive barriers between attacker and target and target and escape (if planned). Attack graphs are also analyzed for time and equipment requirements in order to properly stage and time processes.

Entry points include normal, emergency, forced, and surreptitious types and are typically identified with different protective measures. The entry concern is typically about who goes in, what they bring with them, if they are allowed, and whether they should be where they are.

Egress (exit) points are similar to entry points except that the actor is going the other direction and different controls are required. On exit the concern is generally about who is leaving, if they should have been there in the first place, what is being removed, and what was left inside.

Emergency situations lead to different entry and exit processes, tend to happen at higher rates with higher volume, and are prime targets for exploitation. This means that the protective process for emergency situations has to be properly adapted for those processes or protection will be ineffective during those times. It is often easy to create an emergency and exploit the altered behaviors.

Hardening of physical structures is widely used to improve protection.

Locking systems of many sorts are used in physical protection. Typically they include keyed, digital, or analog controls of electrical, mechanical, fluid, or gaseous mechanisms that are controlled based on time, location, sequence, and situation. They may have different failsafe features and default settings, may be tamper evident, and may be redundant in different ways.

Mantraps are sets of access points designed to trap individuals within them so that if they fail to properly authenticate through the entire process, they will be unable to leave until forces are able to respond. They are commonly used in physical security systems to deter repetitive entry attempts by unauthorized personnel and to catch those who break part way into or out of a facility.

Surveillance systems include coverage of a range of physical phenomena including but not limited to audio, visual, temperature, humidity, proximity, dew point, pressure, air flow, door and window state, heat, motion, smoke, and chemical presence, absence, and level. These are connected to alarm systems, centralized or distributed data collection, analysis, and response capabilities, may be networked, and operate together with badging and computer-related identification and authorization systems.

Response time is a key issue in physical protection. Typically response times are tuned to mitigation of consequences so that high consequence events that demand rapid response are located close to response forces that are present whenever response may be needed. Response time is degraded by resource consumption and there are almost no systems designed to have adequate immediate responses to handle intentional subversion by multiple diversions.

Force on force issues are inherent in any physical security system. Any defensive force can be overwhelmed by adequate offensive force and firepower.

OODA loops are used to analyze physical security systems and are particularly important in understanding how small properly trained and rapid response forces can defeat larger groups for periods of time.

Like information defenses should be, physical security systems are designed to mitigate potentially serious negative consequences to acceptable levels.

Knowledge

Knowledge is particularly important as it applies to the specialized expertise required for the information protection. Special information protection education, skills, mindset, and experience form critical parts of the knowledge base required to make good decisions about information protection at the design and operational level.

Education in information protection suitable to making high quality technical decisions is highly specialized and typically associated with graduate degrees in specialty fields from accredited universities. Unfortunately there are relatively few such graduate programs and too few graduates to fill the available positions, so highly experienced professionals with proper backgrounds may be used in their place.

Experience is the best teacher in terms of not making the same mistake twice, but experience has its limits. Typical experience levels required for information protection involve 1-2 years per specialty area to become competent to make judgments and have broad understanding of everyday issues. With a proper educational background, the same experience is put in the context of that education, linking theory with reality, and this creates a far more effective individual more capable of understanding the implications of events and more able to think “out of the box”. Given that there are something like 25 major issues in information protection at the enterprise level, at 1-2 years each, the CISO should have from 25 to 50 years of relevant work experience in order to have the knowledge base to understand all of these issues at an operational level. But technologies change over time so while experience of 25 years ago is helpful in understanding the issues from a management perspective, it is not technically relevant at a detailed level today in most cases.

Training is particularly effective for getting an individual prepared for specific tasking. The training will typically be effective at giving them the information they need for a 6-month to 2-year period. Once they start in the task they will adapt to changes if they desire to and be effective for several years. If it is good training it will also provide some of the educational background that will help them understand issues over longer time frames. But training is not a substitute for education and should not be incorrectly treated as if it were.

Degrees are often associated with expertise, but you don't need a degree to be an expert and just because you have a degree doesn't make you an expert. There is of course a strong correlation between degrees and expertise in most fields, but not necessarily in the information protection field at this time.

Awareness

Awareness acts to ensure compliance and create identification with the protection process by providing the necessary information to be able to recognize key situations and respond to them in accordance with the enterprise plan. The total set of awareness programs used throughout the enterprise provides the content used to build an effective operational security process.

Document review is required for all information the employee is required to sign associated with the information protection program. Most people don't read the documents they sign in office settings, so document review is necessary in order to assure that they indeed understand and agree to the terms involved.

Initial briefings are required for all those who access information within an enterprise setting. These briefings lay out the specific things the user has to know in terms that they can act on. Most employees get initial employee briefings through the HR process when they first arrive to start work and this is an ideal place to include the initial information protection briefing.

Day-to-day awareness is fostered by and fosters a properly protective work environment and culture. A goal of the CISO should be to create a culture of appropriate security through their overall program, with a central focus of cultural change and maintenance coming from the awareness program. A culture of security is not a culture of fear.

Department meetings are often used to promote security and bring out protection-related issues. A fairly effective practice is for department meetings to include a review of the security failures of the last month. The CISO's awareness program should provide information for use in these meetings to aid in its effectiveness. This typically includes:

> **A news story** from the media that relates to employees directly, such as a story about someone losing their home after an identity theft cause bad credit,

A current or recent situation within the enterprise involving a security problem found and fixed or a situation that impacted a large number of employees,

Any **changes to the protection program** that have wide-ranging effects in the enterprise,

The introduction of any **new awareness program** or other item of interest, and/or

Any **awards or reward programs** associated with the security awareness program.

Computer-based awareness programs provide a limited way to test for and track awareness of specific issues in specific audiences. As a novelty it may hold interest for a time, but it rapidly becomes drudgery and should only be used as part of a systematic effort associated with specific enterprise needs that cannot be fulfilled otherwise or as a verification of awareness given via other programs.

Video-based awareness programs can be viewed by large audiences or copied for large numbers of smaller audiences. If properly produced with a combination of humor, social references, and examples, it can be effective at conveying important messages in a way that causes high retention of the high-level concepts. It can be repeated periodically but becomes stale over time unless mixed in with other programs. It is expensive to produce on your own but many such programs can be purchased for nominal fees.

Groups are sometimes formed for group processes associated with security issues. These processes can be designed to build up awareness programs, but the most effective and entertaining groups of these sorts for general security awareness tend to be those formed in awareness and training game group settings.

Lectures are often used by large organizations with large technical groups or other widely-attended venues as a means to bring in high-quality experts to enhance internal programs. There are quite a few excellent one-hour lecturers in information protection who charge from $2500 to $8000 plus expenses for a guest lecture and a day of meetings.

Games, typically couched as strategic scenarios and situation analysis, are often used to create policies, work through issues, and understand aspects of a space. But they have also been applied to awareness programs. Typically, a game process is used by top management to develop policies and situations that are then played out for awareness programs by all levels of management and workers with an optional outside facilitator.

Posters and banners are sometimes used to keep up awareness levels. While individual posters typically lose their effect in a few weeks, it is not expensive to put up new posters every month as part of an awareness program. Posters used in one facility can be rotated to the next facility so that a dozen different posters purchased in quantities of a few dozen each can be used to cover dozens of facilities for a year.

Badging & carding systems are often associated with physical access controls but they are also part of awareness programs. The programs should remind people that when they encounter someone without a badge they should take action. The specific actions should be identified and trained. The presence and enforcement of badging and carding systems themselves are also part of keeping people aware of security as an issue.

Stand-downs have been used in extreme circumstances to create awareness at a heightened level. For example, government agencies have used stand-downs that involve decertification of systems until they are repaired. They use the repair period to do in-depth awareness programs for all employees and contractors. In one case tens of thousands of employees were involved in shut-downs during which awareness programs were used all day every day to bring the seriousness of the security issues to light.

Memos, emails, mass voice mails, internal FAXes, and similar corporate communications are often used for awareness issues, particularly when there is a critical time-sensitive issue that requires immediate notice. This may be part of the emergency notification system of the enterprise that is also used in disaster recovery and other large-scale incidents. The use of these means for other aspects of awareness tends to be less effective and has the side effect of reducing the effectiveness of the emergency notification process by making it less unusual.

Award programs provide ways to make information protection activities positive experiences and generate social benefits to those who do these aspects of their job well. Award programs can be run for a few thousand dollars per year and typically include plaques or paper certificates, public notice, notice at department meetings, free dinners for two at local restaurants, or other similar items.

Social pressure is applied by creating a culture that encourages secure behaviors. For example, when someone unrecognized is in a workspace, the employees who normally occupy that space should know to come over and say hello, introduce themselves, and find out if they can help the newcomer. If the newcomer is not forthcoming with useful information about who they are, if they don't have a proper badge, or if they are otherwise suspicious, the social environment should create the response that ultimately leads to the individual being escorted out of the facility, arrested, or otherwise handled. If this is the social environment, security will be effective and people will be friendly, but if it is not, penetration of the facility for long-term access will be easily achieved and sustained. Creating a social awareness program is a good foundation for the material included in the other aspects of the awareness program and leads to both compliance and identification with the desired protective behaviors.

Covert awareness programs have recently been noticed by advertisers and adopted for selling. They involve surreptitiously planting individuals within environments to create social changes. This may take the form of someone who displays protective behaviors in conjunction with a planted intruder, someone who creates a “buzz” around a new idea or program, or someone who uses any of a wide range of other influence tactics to move group behavior toward desired objectives.

Organization

Organizational issues are handled by the power and influence associated with the overall protection program and its leadership.

Structure is used and changed to provide direct or indirect control over behaviors and motivations. It may involve moving a manager out of a position when they fail to cooperate with the program, going to a matrix management system to allow the CISO's organization to more directly control select employees, the creation of new governance bodies to create social pressures at the management level, and so forth.

Rewards at an organizational level are typically out of kilter with protection objectives in enterprises without effective programs. An effective program alters the enterprise reward structure so that managers and workers who show excellence in protection functions are rewarded with raises and promotions. Working on an information security issue should be seen as a path to advancement and might be considered a requirement for promotion into certain positions. Performance reviews should include explicit performance relative to information protection and proper behavior should be rewarded in clear ways.

Punishments associated with poor security performance should include negative management reviews, sanctions of various sorts, and ultimately, termination and prosecution depending on the specifics of the matter at hand. Information protection behaviors should be included as a normal part of worker reviews, and these should be based on performance metrics that are clearly defined, that workers are aware of, and that feed into the overall information protection program's measurement process.

Communication is at the heart of organizational interaction and is a key factor in success for the CISO. Creating and maintaining lines of communication throughout the enterprise and using those lines to control and observe behaviors is fundamental to success.

Top-level governance – the CISO

The CISO is responsible for assuring the ongoing value of all of the non-physical and non-fiscal assets of the company. They manage the enterprise control system associated with information protection through interactions with people, both individually and in groups. The groups are generally of two sorts; (1) functional groups that perform the necessary functions for operating the protection program and (2) review board groups that review and oversee the efforts of the functional groups. The individuals are typically key people within the enterprise at any and every level, and for each specific activity performed, there may be any number of different individuals called upon.

Top-level governance typically happens through interaction with a top-level governance board. This is an outward facing function of the CISO. It is designed to interface between top executives, the board of directors, business unit owners, and other stakeholders who are responsible for overall control of the enterprise and the information protection function. This group ultimately includes the individuals who have legal responsibility for the business and its operations and that determine the placement and reach of the information protection function in the enterprise. This group should meet periodically with the CISO to review overall program performance and inquire about specific issues they deem worthy of their attention. Meetings should be scheduled with this group at least once per quarter and, for select functions of the CISO, like business continuity planning, additional meetings with many of the same people will also be held.

Top level governance process sets the rules, defines the duties to protect, creates and enforces the power and influence, grants information, enables and facilitates risk management, and provides the funding and support that allows the protection function to exist and operate. If top level governance is not done properly, the CISO will have difficulty succeeding and protection will be problematic.

Who should the CISO work for?

In many enterprises, the CISO works for the CIO, and in some cases for the CFO or others, but these lines of authority are problematic. The broad range of issues involved in the CISO's job leads to many high-level interactions with members of the top management team and their staff. CISO decisions are clearly beyond the scope of the typical CIO or CFO function. Placing the CISO at a lower organizational level is problematic for their interaction with the rest of the management team.

It is the role of the top-level governance process and team to determine where the CISO should be placed within the enterprise governance structure and to assure that the function operates effectively wherever it is placed. Regardless of who the CISO actually reports to, the governance group should have unfettered access to the CISO and the CISO should have unfettered access to the members of the top-level governance group. In case after case when the CIOS is placed under the control of an interested party, the CISO is held back from such meetings, told to only work through their management chain to get to the group members, and blocked or threatened if they attempt to "go around" or "over the head" of their manager. This ultimately causes CISOs to be fired and their function in the business to fail. Any CEO that allows this to happen and doesn't require unfettered meetings at least once per quarter is opening the enterprise up to liability, protection failures, executive level abuses, and other similar liabilities.

The most critical reasons that the CISO must be independent stem from (1) the need for the CISO to independently report on security-related matters to the CEO and board of directors and (2) the requirement for adequate influence and access to information to meet the enterprise duty to protect at the enterprise level. Top-level communication cannot be intermediated without putting top management and the enterprise at peril. Fulfilling the duty to protect requires adequate influence and information and the skill to use them effectively. This does not imply an adversarial relationship with other management team members, but the nature of security functions, like audit functions, demands that independence be maintained for objective evaluation to take place. Also:

- The evidence over a long time frame indicates that insiders are involved in 80% of the losses encountered in information system attacks. Many people misstate this statistic as indicating that 80% of attacks involve insiders, but in fact a relatively small number of attacks result in most of the real harm, and many of these involve high-level insiders.
- In case after case, security decisions blocked by executives are used to cover up executive misdeeds. It is critical that a top-level decision-maker act as an independent reviewer of security-related issues, just as it is critical that financial

auditors report directly to top executives and be independent of those who can move funds.

- In case after case, mid-level managers or technical people decided that security enhancements were infeasible because of difficulties in getting them through their management, even though they thought they were the right things to do. In one case a major single point of failure for an entire enterprise was identified in a security review but not passed to top management because a mid-level reviewer determined that it would likely be rejected. This should never happen when a CISO is in charge of these processes but it is common in those without a high-level CISO.
- Regulations like Sarbanes-Oxley require that top management attest to the true state of the enterprise's financial well-being with criminal sanctions for failures to report. A failure like the information system collapse that caused all Comair flights to be canceled over the Christmas weekend in 2004, stranding 30,000 passengers for days, clearly demonstrates the sorts of risks at hand.

An independent top-level position for the CISO is necessary. This may be a side box off of the CEO's office, the board of directors, or the audit committee. It may be placed within the chief counsel's office but this will be problematic for the legal department. Or the CISO can be a member of the management committee.

The reality on the ground today is that most CISOs are placed within the information technology department. That is precisely why we see so many large-scale information protection failures within enterprises in which the CISO knew of the issues, had identified them prior to the serious negative consequences being realized, and was unable to do anything to mitigate the harm in time. In case after case, this key governance decision is poorly made and the result is enterprise-wide, ranging from inconvenience and damage to brand all the way to enterprise-wide collapse.

There are many cases when a CEO has a conversation with the CISO at some point and asks that the CISO directly meet with the CEO and let the CEO know if anything is being inappropriately handled by the CIO or whoever else is in the management chain of the CISO. This is almost guaranteed to fail, and a good CEO should realize this. If the CISO works for the CIO and the CIO doesn't want the CEO to know what's going on, the dependent CISO will be put in a spot where they will either lose their job or tell the CEO what they know. Most CISOs, like most other employees, realize that the decision made about who they work for and who sets their pay and who evaluates their performance is the more meaningful basis for making day-to-day decisions than what the CEO tells them one time in a casual chat. If the CEO meant it, then the CISO would not be working for the CIO.

Almost every CISO who is willing to discuss these issues candidly will tell story after story about how the politics of the enterprise prevented them from doing their job, how top management doesn't have a clue about what's really going on, and about how their boss prevents any of the important things from being known to the top-level decision-makers. Almost every CIO who has a CISO working for them and who is willing to speak candidly will tell you that they don't want their management to know of all of the problems that go on in information technology day after day because it will weaken their position within the enterprise and endanger their job. Most CEOs will tell you that the CIO has it under control.

Should the CISO have other duties?

In many cases, enterprises consider joint positions for the chief information security officer (CISO); the chief of corporate physical security, sometimes called the Chief Security Officer (CSO); and/or the chief privacy officer (CPO). There are situations in which these positions can be joined, but great care should be taken in understanding the implications of this combination. Given the complexity of the CISO job, there is little free time left for this member of the management team. Unless there is a mitigating circumstance, overloading the CISO position will cause degraded job performance. Here are some exceptions:

- If by the nature of the business the physical security function covers mostly information systems and assets, combining the functions may be sensible.
- If there is a large inventory control or production component to the business, or if personnel protection is non-trivial, a

physical security lead is likely needed, and to expect the combined position to do the job as well would be a mistake.

- Placing the CSO function underneath the CISO or the CISO underneath the CSO is a mistake unless there is a large component of one and small component of the other.
- If there is no physical security lead, using the CISO's physical security group as the enterprise lead for physical security associated with information systems may work.
- If physical security is handled by local facilities personnel, a subordinate to the CISO should coordinate facilities security for information technology in a matrix management arrangement with the local facility owner.

The CPO role is usually highly focused in privacy-related issues and deals largely with the enormous global complexity associated with privacy regulations. This is largely a legal-department issue and is most often handled that way. While the CPO is a critical enterprise function, the complexity and technical detail level is high and it is, in practice, infeasible to keep fully up to date in these issues while also operating such a broad management function, except in cases where the enterprise is highly localized.

Who should work for or be matrixed to the CISO?

We will call the individual responsible for operating the enterprise information protection function the chief information security officer (CISO). The CISO has indirect influence over a large number of people and direct control over a smaller team. While enterprises differ significantly on organization of the function, the CISO typically has a staff of about 10 direct reports who control or influence others throughout the enterprise through their leadership in various groups and review boards. In many enterprises these people are matrixed to or only indirectly work with the CISO. The staff members and their qualifications should typically include:

(1) Staff assistant: A well seasoned enterprise employee who knows how enterprise administrative processes and systems work, has strong technical and communications skills, and knows many other workers helps. Strong project management and documentation skills are a must. Project managers do this well.

(2) Policy, standards, and procedures lead: This individual should have some legal training or background and understanding of technical issues in information protection and business systems. They should be detail oriented, have strong writing and language skills, be willing and able to deal with legalese, and have strong library skills. Enterprises often use outside consultants to develop and update policies.

(3) Legal, investigation, and risk management team lead: These functions are sometimes combined because of synergistic skill sets and the need for close coordination of these issues. Typically they have at least 10 years in the enterprise, strong analytical, actuarial, and mathematical skills, and bookkeeping and investigative backgrounds. Retired law enforcement personnel with additional skills and degrees are often used. Many companies split the position using a different individual for risk management and a retired law enforcement person for the legal and investigative lead.

(4) Protection testing and change control lead: Expertise in quality control and quality assurance (QC/QA) and testing is a must. Technical expertise in a wide range of systems, hardware, software, and operating environments helps. A master's degree with a focus on information security or systems testing and evaluation is a help. At least 10 years of experience in a technical setting is required. Special education and training from outside experts is often also required.

(5) Technical computer security lead: At least 10 years of computer security experience is called for with increasing technical responsibility leading to technical team leadership in large-scale complex projects. Strong project management and technical skills across a broad spectrum of system and network types, advanced training or a masters degree in a computer-related area, and the ability to work well with others in groups are also key.

(6) Physical security lead: Typically the physical security program is independent of but linked to the information security program. The lead from the physical security team acts as a liaison to the CISO team for physical security issues. If physical security is handled by facilities managers at each facility, a physical security lead within the CISO office is required to coordinate physical security for IT facilities. This individual usually likes to travel, has 20 years of law enforcement background, perhaps some construction and physical sciences background, and has and continues to go through specialized training in physical, operations, and information security.

(7) Incident handling and business continuity planning lead: This lead is responsible for IT-related business continuity planning and disaster recovery and real-time technical attack and defense - the detect and react processes. They need to understand enterprise business applications and critical interdependencies to design real-time decision mechanisms, but they also need strategic understanding and excellent communication and coordination skills to address business continuity. Typically they are long-term employees with at least 20 years of experience including project management. They need strong vendor management skills and manage external consultants when skills are not internally available.

(8) Audit lead: Typically this individual is a member of IT audit that has become the team leader after more than 10 years of experience because of demonstrated team leadership. The audit lead is typically a representative of the audit team rather than the top technical person.

(9) Awareness and knowledge lead: Experience in corporate training and awareness programs is a must. All the better if the experience is in security-related areas. This position may be filled by a consultant with a history in information security training and awareness, but should be replaced by a staff training and awareness specialist once the program is fully operational, with outside assistance a few weeks per quarter.

(10) Personnel and operations security lead: The ideal candidate has 10 years of counterintelligence experience, work experience leading teams that do background investigations, and sound understanding of technical and human vulnerabilities leading to system and project compromise. This individual is sometimes found in the physical security team.

Other matrixed personnel include almost anybody within the enterprise for short periods of time, depending on the circumstances. For example, during an incident, many individuals may be coordinated by the CISO, external contractors commonly work for or with the CISO to perform their tasking, groups fuse together for mergers and acquisitions, external organizations work with the CISO in detail for activities involving collaborative efforts, and during investigations and incidents, the CISO may get deeply involved with both internal and external individuals and groups, some of whom will end up temporarily supervised by the CISO.

This grouping is an example but is not prescriptive. Background, expertise, and knowledge for these functions tend to be found together in individuals, but other groupings are reasonable depending on the specific individuals involved and the amount of work involved in each area.

Groups the CISO meets with

The CISO, as a high-level executive, tends to spend more time in meetings than anywhere else. Because of all of the control requirements the typical CISO has to meet and the wide range of areas they have to cover, they typically meet with groups of people reflective of the range of duties they have to fulfill. Figure 6-4 shows an extract from Figure 6-2 that shows only the major groups that the typical CISO works with. Through these groups, the CISO influences the enterprise protection program and gains the knowledge needed to do the job.

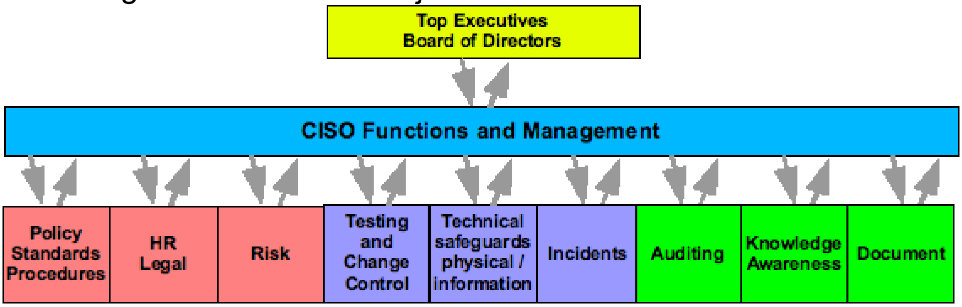

Top Executives
Board of Directors
CISO Functions and Management
Policy Standards Procedures
HR Legal
Risk
Testing and Change Control
Technical safeguards physical / information
Incidents
Auditing
Knowledge Awareness
Document

Figure 6-4 – Groups the CISO meets with regularly

Separation of duties issues

Separation of duties is a key issue in information protection, but at the CISO level, management has to coordinate all aspects of the protection program for it to be effective. What really has to be separated is not the management of the overall information protection program, but the implementation of controls. Different checks and balances exist on management than on implementation. In effect, the audit process covers management issues while management covers the separation of duties for CISO functions.

Figure 6-4 shows a typical overall separation of duties approach. In this approach, specification, execution, and verification are separated at the overall governance level. Policy, standards, procedures, HR, Legal, and risk management operate in the specification arena. Testing and change control, technical safeguards, and incident handling operate in the execution arena. Auditing, knowledge and awareness programs, and documentation functions are in the verification arena.

Within each arena there are also separations. For example, testing and change control are separated from technical safeguards which are separated from incident handling. This separation prevents someone from designing and implementing a malicious information technology element and having adequate control to get it past testing into production and defeat incident detection.

The theory of groups

The depth and diversity of the information protection function and the need for crosscutting implementation and involvement dictates the need for groups of people to get involved in making decisions and carrying them out. The theory of groups says, in simplest form, that effective groups have enough people with top quality skills and knowledge in relevant areas to cover the issues of import to the purpose of the group, and as few excess people as feasible.

- The reason for covering all of the issues is so that important things are not missed.
- The reason for high quality experts is to assure that the best information is brought to the group.
- The reason for limiting redundancy is that it is inefficient. It not only that it wastes time and effort of experts. Where experts disagree, they tend waste the time of the rest of the group over relatively minor differences.

- Adequate redundancy is important to assure that there aren't significant holes, but excessive overlap is not desirable.

Another common understanding about the group process is that it goes through three phases; (1) storming, in which the members have significant friction as they meet each other and adapt to the new context, (2) norming, in which the group members normalize their behaviors to each other by determining what will and will not work in the group context, and (3) performing, in which the group gets work done efficiently with a minimum of friction. The goal is to reach a performing stage quickly and retain it for most of the effort.

A third important group-related issue is that, as groups age, they become stable in their configuration and tend to innovate less. They may form cliques or become a clique as a whole. This tends to lead to similar thought patterns and roles played by individuals that limit their overall utility in the group. The group becomes static and stale and subject to group think, in which even things that would be considered obviously foolish to an outsider are considered reasonable by the group because of their context with each other. To avoid this situation, it is necessary to shake up such groups periodically. This costs some short-term performance but improves long-term performance.

These theories should be applied in governance of the information protection function within the enterprise.

- Optimize group performance by combining the right experts from the relevant areas without excessive redundancy.
- Operate groups on a regular basis and keep them stable over time
- Force a reasonable level of turnover or other dynamics to stay fresh and avoid group think.

What groups are needed

Many different groups are involved in information protection functions within an enterprise, and governance is needed to deal with these groups and keep them running properly. For each IT project, there is an information protection element involved, and the individuals with that function for that group have to have information and understanding necessary to interface between enterprise and project protection functions. Once projects enter into processes that turn them into enterprise business functions, they have to join the fold of enterprise information protection, including all of the facets discussed in the comprehensive information protection approach. That approach involves 15 organizational and business perspectives, 4 types of life cycles, 5 elements of the attack and defense process, 5 different sorts of objectives, 6 different contextual elements, the risk management framework, and how the business works. In large enterprises there are 25 or so enterprise-wide groups involved in this process and a central group that may be mirrored in business units that is responsible for managing this overall process.

Together with leaders of these groups and an enterprise-level leadership role integrated with top management, they form the institutions that form the enterprise protection governance function.

These examples reflect the overall approach of having security functions integrate with business structures. The interconnectivity and interdependencies associated with security-related systems drives the need to coordinate overall security efforts. Integration across the enterprise ends up being a complex myriad of different arrangements that provide necessary function, coordination, and control. Some amount of restructuring may be used to reduce complexity, but the CISO rarely ends up in direct control over much of the security function. Rather, control and influence are indirect and a cooperative environment comes into existence. Failures in control are handled in a political process. Inadequate CISO power and influence leads to control failures and the enterprise suffers. If a necessary function doesn't exist within a portion of the enterprise, the CISO must find a way to get the function in place by directly creating and controlling it, by influencing the business unit to put the function in place, by brokering a deal with another business unit to provide that function, or by other creative means.

Business unit governance group(s)

Business units that are substantial enough to operate more like wholly owned subsidiaries than like departments typically have their own internal information protection functions that fulfill some or most of their needs. Boards exist within individual business units for their internal operations. These are not typically operated by the CISO function. Rather, they interface with the CISO in order to provide enterprise-level information and assure at the enterprise level that information protection is as it is supposed to be. The exchanges also tend to save time and money by reducing unnecessary redundancy and improving process for all.

Policy, standards and procedures groups

The policy, standards, and procedures group is responsible for initial policy development, reconciliation of existing policies, policy rewrites, adaptation of policy to changes in the environment, development and maintenance of control standards from policies in conjunction with the operating environment, and development of procedures associated with meeting control standards.

The review board is responsible for review and approval of policies, which includes top management that makes them official within the enterprise. The review and acceptance of standards by individual groups affected by those standards, approval of those standards by the proper level of management in different enterprise areas, and verification of the consistency of those standards with policies before acceptance is also controlled by this board. Individual managers are responsible for verifying that procedures meet standards and assuring execution of the procedures.

Documentation of all aspects of this process must be generated and kept. This facilitates review for new members of teams, demonstration of regulatory compliance and other legal mandates, and proper execution in context. This should include meeting minutes, periodic plans, deliverables, progress reports, and other related documentation of the process. It should also include original data collected in the process, such as copies of emails associated with policy reviews, schedules for processes in whatever form the projects are tracked, ultimate dispositions of all activities, funding and costs associated with the effort, and of course resulting formal outputs from the process and changes over time.

Project management should be used for this process and should be responsible for collecting, tracking, and reporting on all aspects of project progress, convening and scheduling meetings, and providing the CISO function with ongoing information on the overall effort.

The audit process should verify that these responsibilities are being properly carried out by; (1) selective testing of consistency by examination, (2) verifying that the approval process is generating meaningful review prior to approval, (3) verifying that approval or rejection of changes is done in a timely fashion, and (4) verifying that policies, standards, and procedures are followed. This is done by (1) reviewing the documentation associated with the effort, (2) verifying proper approvals for policies, standards, and procedures in actual use, and (3) verifying the actual operation of the overall system by selective, periodic, random, and blind review of operations against procedures, standards, and policies.

Legal groups

The legal group varies greatly from organization to organization, depending mostly on the size and expertise of internal legal counsel. As a rule of thumb, legal review of all policies is mandatory, standards should be reviewed as well to assure that no laws are being violated, and personnel procedures should be reviewed for issues associated with potential law suits and statutory violations. Privacy laws relating to background investigations, laws related to the specific industry, and the range of related issues associated with legal positions are particularly important in international businesses. The legal group should also be involved in incident response whenever investigatory processes are undertaken.

Legal groups generally control their own documentation and have special privileges for specific situations, so they tend to be more advisory than active participants in the protection program. Often, the CISO function has interactions with a small number of legal staff members and otherwise remains fairly independent of the legal issues except as they are used to review work the CISO is tasked with. In many cases internal legal council are not expert in issues related to intellectual property law or information protection-related issues, and in these cases, outside counsel is advisable.

Personnel security coordination

Personnel security is often coordinated by HR and carried out by a group within physical security that deals with personnel protection, facilities security, and other related issues. There is sometimes also an effort for executive protection that may be yet another group, and background checks may be performed by an outside service. This combination of activities implies that for the information protection issues to be properly covered this activity has to be properly coordinated. HR is usually critical to the proper functioning and coordination of the functions of individuals within an enterprise, and they should almost always be the focus of these efforts. But HR also has to operate in a manner that provides the information required for effective use of protective functions such as identity management, account creation and removal, termination, leaves, and all other aspects of the human life cycle that imply protection changes.

Actions implied by the information protection program as well as issues related to assurance of employee rights and the proper operation of the appeals process for incidents and other matters related to employees is also in the purview of the HR department. Tracking of personnel information is an HR function that has to integrate information protection issues in order for the coordination to take place. Clearance processes and status are HR department functions that integrate with other aspects of security as well.

Documentation requirements are extensive for these processes, legal issues have to be considered, and review boards for processes as well as individual cases are required for personnel actions.

Tracking of training and awareness programs is often handled by either the HR department or a separate training group, however, tracking of educational efforts as it relates to qualifications, benefits, salary, position, and other issues is within the HR function.

The CISO has responsibility to assure that these processes are properly undertaken and that timely and accurate information is used. This means that audit has to be involved to verify the process and that the CISO function has to coordinate the activity and influence changes necessary so that it works effectively.

Risk management groups

The risk management group is responsible for evaluating risks and making determinations about when risk can be accepted, transferred, avoided, or mitigated. This is a core top-level business function that historically falls on top management and rightly belongs there. Unfortunately, many in top management don't clearly understand the issues of risk management when it comes to information and the supporting infrastructure to provide this decision process is not in place in many enterprises. However, there is usually a risk management group that does analysis for other risk management issues, or at least a group that analyzes insurance issues and helps to determine best rates and the need for insurance.

The risk management group for information protection must invoke a process that allows top management to make rational decisions, and this is almost always a core function involving the CISO. In fact, when there is a CISO, they are often responsible for making preliminary evaluations for all risks in this area and sole responsibility for decisions about low and medium risk situations. The decision to delegate risk management for these levels of risk implies a process that identifies risks and differentiates them based on consequence, and this is in the purview of the risk management group.

Risk management has to be a well documented process in order to be meaningfully applied consistently across an enterprise. It should not depend on special expertise for day-to-day operations, however, it requires well qualified individuals who understand how to make good judgments and understand the technology that forms the basis for the evaluations undertaken. The risk management group should tightly integrate with the CISO function.

Protection testing and change control

The protection testing and change control group(s) are responsible for measuring the effectiveness of protection on systems that warrant such controls and assuring to the desired degree of certainty that those systems operate as they are supposed to.

Results of protection testing and change controls are reviewed as a matter of course before results are accepted and systems are transitioned from testing into operational use. This is typically done by a manager responsible for the systems affected and by others who are potentially affected by the changes. For example, a change to enterprise domain name services (DNS) has to be approved by all of those responsible for medium and high risk infrastructure and systems that the DNS server supports. This serves the additional purpose of notifying all affected owners of the pending changes, and of tracking all of the testing that has been undertaken to assure that is meets the requirements of all interdependent systems.

The change control and protection testing group(s) must be independent of other groups because they are tasked with separating research and development from production and assuring that errors, omissions, and acts of malice do not reach the production environment.

Protection testing is different from the sorts of vulnerability scans undertaken by common vulnerability assessment tools designed to operate in low surety environments. These sorts of control are commonly used by systems and network administrators to detect things they should be fixing and by auditors to verify that this maintenance work is being done. Generally speaking, systems under change control are medium of high surety systems in medium or high risk applications and, as such, they tend to be, or should be, isolated from external influences to a large extent.

Technical safeguards group and review board

The technical safeguards group is responsible for the job of technical risk mitigation. They apply technologies to systems in order to reduce the vulnerabilities of those systems and the consequences of failures in those systems.

For low risk systems, as determined by risk management, the technical safeguards group is often left on their own in terms of protection. The CISO function typically oversees the protection of low surety systems and seeks to make certain that they are not able to unduly influence medium or high surety systems through architectural methods, like the network zoning policies, etc.

For medium and high risk systems and content, the technical safeguards team has to gain approval from risk management for mitigation approaches but takes on the primary lead for the design and implementation of technical safeguards.

They are subject to audit as well as oversight, including review by the zoning board for zone-related changes and oversight by the CISO function. Documentation is critical, legal approval has to be gained for certain potentially invasive surveillance technologies, and interface to the HR application environment is central to success of technical safeguards depending on identity management solutions. The technical safeguards team has to implement policy, help develop and follow standards, create procedures and get their approval, send changes through change control for high and medium surety systems, act as experts for some aspects of training and awareness, and receive education in order to continue to be effective in their tasks. They also have a heavy documentation burden and form part of the response capability associated with the incident handling function.

Separation of duties limits the technical team in many ways. Their central role in protection and their access to so much of the potentially harmful systems, places a high burden on them for reporting to others, following the rules set forth by others, and dealing with highly complex situations at high speed with strong oversight. The CISO function is typically a central aspect of this integration across the enterprise. Many enterprises mistake the control of technical safeguards for the CISO function and miss the bigger picture.

Zoning boards and similar governance entities

While network zoning and related matters are typically part of the technical safeguards function, there are often independent groups that review zoning policies, including system owners, network owners, risk management, audit, and incident response. Zoning boards typically include only those impacted by a change in zones or, during the creation of zones, those responsible for working within those zones.

Similar requirements may exist for classified systems and other special purpose environments that have to meet additional regulatory or jurisdictional requirements. For example, manufacturing facilities in certain industries have very specific requirements that have to be met for certain systems, and these are typically reviewed by special groups. Classified computers have special review and approval processes associated with their creation, operation, maintenance, and decommissioning. Special requirements exist for some countries, and the regulatory involvement implies participation of the legal department in review.

Physical security group and review board

Physical security is often handled by an independent business function with special requirements and collaboration associated with data centers, wiring, wire closets, conduits, perimeters for medium and high risk systems, protection of paper and other media in storage, before input, and after output, physical aspects of information and equipment life cycles, and integration of physical and informational access controls.

But in cases when physical security is oriented towards the facilities function rather than overall enterprise protection, or when it fails to cover all aspects of the information protection function, the CISO function has a responsibility to the enterprise to report the problem and, if mandate is given, to manage the protection. Depending on the organization, the CISO function may have only a peripheral role in the physical security review board, may be a member of the group, or may chair the physical security group and convene the review board. In the latter case, the CISO acts more as a CSO and has broader responsibilities.

Incident handling group and review board

The incident handling group is responsible for information technology aspects of business continuity planning, disaster recovery, and day-to-day incident detection and response within the information technology function. It is, necessarily, separate from the technical safeguards team because it is tasked, among other things, with detecting trusted insider abuse. At the same time, the incident handling group is not permitted to control production systems, acting only through the systems administration group for low-risk systems and change control for medium and high risk systems to carry out any changes. This separation of duties is key to proper operation and thus the incident handling team is part of the assurance process, while the systems administrators, developers, and others involved in changes are part of the operations process.

The incident handling team is responsible for (1) identifying event sequences that can cause potentially serious negative consequences, (2) devising the means to detect these sequences in a timely enough fashion to mitigate harm to within enterprise specified tolerances, (3) devising the warnings and response regimen that mitigates these consequences in the required time frames, (4) defining the conditions under which these response processes get invoked, (5) initiating, managing, and carrying out these responses when they are required, (6) devising the process used to determine when response processes can be terminated and normal operations continued, (7) carrying out those termination processes when necessary and appropriate, and (8) after-action reports, documentation, and other related matters that produce an incident handling system that adapts properly over time.

Incident handling is often integrated with the computer security implementation team but in enterprises with medium or high valued systems, such as financial institutions, separation of detection from operation is very common and critically important to preventing high consequences. It is highly advisable to maintain this separation. However, some level of information flow is required in order for intrusion detection mechanisms to be properly tuned to the changing situation. For this reason, incident handling is part of the review process for technology changes. This review element serves as notice and as a means to mitigate and configure detection for problematic technologies before they are deployed.

For low consequence systems, intrusion detection and response processes may be embedded in the systems themselves and run by systems and network administrators, however, it is useful for these systems to provide feeds to the incident handling group so they can remain aware of situations in those environments that may eventually effect them. The seeming inefficiency of separate teams may be outweighed by the larger number of incidents that have to be handled in low surety environments and the need for the higher surety environments to be far more carefully operated and attended to. The additional duties of disaster recovery and business continuity planning also tax the incident response team for medium and high surety systems and may have little impact on low consequence systems. It might be wise to use the low surety environment for experimentation and to help train individuals who eventually move into the medium and high surety incident handling arena.

Incident handling includes a lot of documentation requirements, not the least of which is the collection and retention of forensic evidence associated with legal matters, and the documentation of event sequences that ultimately lead to employee sanctions and other related actions. The business continuity and disaster recovery plans have a lot of documentation as well. The interface to the legal department typically runs through a manager or perhaps the CISO for incidents of significant import. HR records get generated as a result of these actions and the HR information associated with positions, roles, and other elements used in identity management are key to understanding and characterizing event sequences as incidents. Incident handling policies, standards, and procedures are part and parcel of the function, not only because they have to be followed but because they have to be developed and updated. Risk management helps to decide how much incident handling effort is required for which systems, and change control provides information used in incident handling through test results that provide calibration information and configuration management that helps to determine criticality and severity of incidents.

Incident handling feeds data to auditors for evaluation of the incident handling capability and its operation and as information for audit review of the operations area. Incidents often drive awareness programs and the incident response team often acts as a provider of critical information for the awareness and knowledge requirements. Incident handling team members sometimes participate in the awareness process and are key members of the higher level activities associated with business continuity and disaster recovery planning practice sessions.

The incident handling review board is designed to provide management with information about incidents and to get feedback on the process so as to improve it over time. Most enterprises should have quarterly reviews of incident handling and additional reviews when incidents cause substantial harm. Reviews of individual incidents should be created as part of the documentation process complete with after action reports indicative of suggested process improvements. The review board should review after-action reports prior to quarterly meetings and summaries of these reports should be included in the overall review of the program.

Audit group and review board

The audit group is often but not always part of the corporate internal audit function. If no such function is capable of dealing with the rigors of internal information technology audit, a separate group has to be created either within the existing enterprise organizational infrastructure or within the CISO's functional responsibility.

The audit group has a very broad range of responsibilities for reviewing and reporting on CISO functional responsibilities. This generally means that audit reports should go to the top executives or board of directors. The audits of each of the functions of the CISO should also go to the CISO so that the CISO can adapt the operation to meet the need. There is an apparent conflict of interest presented by this need to report to the CISO as part of their feedback and to report on the CISO to top management. One of the best ways to reconcile this is to have an information technology (IT) audit group that reports to the CISO and use internal audit or external audit and review processes to review the performance of the IT audit group.

IT audit has the responsibility to review the performance of every aspect of the information protection program as well as responsibility to verify that no serious undetected incidents take place by acting as an independent incident detection group. This implies a mix of expertise in every technical mechanism used in IT, from telephone systems to identity management infrastructure, as well as understanding of issues related to all of the functions of the CISO.

Awareness and knowledge group and review

The awareness and knowledge group is tasked with providing a comprehensive information protection awareness program to the enterprise. This entails the collection, creation, and dissemination of information appropriate to all of the individuals in the company, translated into proper language and written so as to meet social norms, and presented in a manner that both conveys the important information and provides specific instructions on how to behave with regard to information protection issues relevant to the situations and tasks of the individual.

In order to be effective, critical awareness issues have to be repeated twice a year, and employees who have not received the awareness training and demonstrated their understanding of it have to be decertified from performing tasks until they come into compliance. This implies a system of tracking all users and their currency in security training and awareness for all tasks they are assigned to perform. As changes in responsibility occur, training and awareness have to be updated.

In addition to all of the tracking of program execution and compliance, the awareness program has to be updated on a regular basis so that it does not become stale. A variety of techniques are available and should be rotated and applied over time to keep interest levels high. The program should produce well-documented results that can be readily reviewed on an annual basis to assure that the program is operating properly. This review is typically done by the CISO as part of their normal process. Legal review and long-term documentation should be retained to mitigate any disputes for the duration of the applicability of the training material or its historical value.

Documentation group

The documentation of information protection tends to be extensive and involves many different people. Typically there is a corporate documentation standard, an archival function and document repository, a tracking process that includes aging and life cycle management for destruction processes, and a set of retention policies, standards, and procedures that support this function. A library system is often used to track all of this information, including the requirement to categorize and retrieve data, librarians, and off-site backup storage of important documents. This system should track all of the documentation produced through the CISO function and provide easy retrieval and access for authorized individuals, including the CISO and all of the review boards, relative to the material they review. This group should also provide the means for audit and other related functions to gain access to materials, and provide historical data and research capabilities.

Documentation is often systematically produced through the use of professional project managers as part of the project management process. It is helpful for the CISO to have a project management process to track and provide clear documentation of processes and outcomes. Documentation has to have proper classification and applicability in order to assure that it is properly protected within the enterprise protection architecture.

Special projects and other groups

Project-specific groups are commonly formed for short periods as needed. The CISO is often involved. Some typical groups are:

- Many companies have a centralized firewall implementation and operation group, usually under control of the technical safeguards and coordinated with incident handling.
- If the firewall team is within the networking group at the enterprise level, it may become a CISO function or matrix managed to assure effective operation and coordination.
- A telecommunications group that handles firewalls may exist in a business unit. They may be integrated with the CISO organization or kept within the business unit. Coordination with the CISO team at the management and technical levels may be used. Reporting structures are then designed to provide the CISO information and access as needed.
- A new acquisition might have a firewall team. For transition, a firewall may be put between the acquisition and enterprise to limit damage. The new acquisitions lead then regularly communicates with the CISO team as needed. As the new business unit is integrated, structuring changes to meet enterprise needs.

The CISO's schedule

Table 6-5 gives a sense of a CISO's meeting schedule.

Management Activity	Group size	Frequency	Duration
External top-level meetings	8	2/yr	4 hrs
Internal top-level meetings	10	2/yr	16 hrs
Internal top-level teleconference	10	52/yr	1 hr
Policy group	12	12/yr	16 hrs
Policy review	8	2/yr	4 hrs
Audit review	12	12/yr	4 hrs
Testing group	n/a	Continuous	n/a
Testing review	10	4/yr	4 hrs
Technical safeguards	n/a	Continuous	n/a
Technical review	12	4/yr	8 hrs
Personnel group	n/a	Continuous	n/a
Personnel review	4	2/yr	2 hrs
Incident handling	n/a	Continuous	n/a
Incident review	12	4/yr	4 hrs
Emergency management	12	As needed	As needed
Business continuity planning	n/a	Ongoing	n/a
Disaster recovery	n/a	Ongoing	n/a
Strategic incident team	4	Continuous	n/a
Legal group	6	As needed	As needed
Legal review	12	As needed	As needed
Physical security group	6	12/yr	8 hrs
Physical security review	12	1/yr	8 hours
Facility security groups	6	As needed	As needed
Awareness group	n/a	Continuous	n/a
Awareness review board	4	12/yr	2 hrs
Insurance and risk transfer	6	As needed	As needed
Internal technical review board	12	12/yr	8 hrs
External technical review board	12	4/yr	8 hrs
Internal management advisory	6	2/yr	4 hrs
External management advisory	12	1/yr	16 hrs
Zoning boards	12	12/yr	2 hrs

Table 6-5 – The typical CISO's meeting schedule

Scheduled group meeting times account for more than 600 hours per year of CISO time. Add in preparation and analysis time, the political and budgeting processes, keeping abreast of current events, keeping the CISO’s knowledge level high, strategic planning, emergency involvements, and all of the as needed activities. This constitutes most of the full-time CISO’s schedule.

What are the rules?

The CISO has responsibility for creating and following formal and informal, written and unwritten rules. Formal rules are easier to understand and define. They come from a variety of sources and have varying punishments associated with failures to follow them.

Policies are the codification of internal rules in documented form. But real rules of how companies work are rarely codifiable in those terms. Policies are used to derive control standards that codify more detailed situations, change more often, and have shorter approval processes at more local levels. Control standards constrain procedures that codify sequences of specific actions for specific circumstances. They change even more often, are more locally controlled, and take less time to change. For example, a policy that indicates audit information must be read-only may produce a control standard mandating append-only files for systems with access controls. This leads to a procedure to set protection bits in a mainframe. These include, produce, and require documentation which forms a contemporaneous record of the rules and how well they are followed.

Official rules tend to pass through organizational structures. For example, in a hierarchy, orders come down from above and may not be appealed unless they are thought to be in violation of policies set by top leadership. Even in these cases, the challenger is facing an uphill battle and has the burden of proof. In a matrix environment, different people have responsibilities to fulfill their mandates, and their ability to command effort derives from those mandates. But the individual has to decide how to prioritize the different requests and to understand where their loyalties lie. The rules may be more complex and even contradictory in some circumstances. In hub systems, the center of the hub is simply in charge. The ability to control these systems is limited by the concentration of power and the limits of cognition, focus of attention, and bandwidth of the leader. Networked organizations are driven by effort more than rules. This makes explicit control far harder to accomplish, leaving control to those who have the ability to build consensus by exerting other sorts of influence.

Principles and standards

The Generally Accepted Information Security Principles (GAISP) provide a starting point for understanding principles of effective information security governance. GAISP provides a standard approach to understanding the rules that should be in place and, as a standard, it may be used to assert diligence. It includes pervasive principles, broad functional principles, and detailed principles. The pervasive principles include:

- **Accountability:** Information security accountability and responsibility must be clearly defined and acknowledged.
- **Awareness:** All parties with a need to know should have access to principles, standards, conventions, or mechanisms for the security of information and information systems, and should be informed of applicable threats to the security of information.
- **Ethics:** Information should be used, and the administration of information security should be executed, in an ethical manner.
- **Multidisciplinary:** Security should address the considerations and viewpoints of all interested parties.
- **Proportionality:** Controls should be proportionate to risks.
- **Integration:** Security should be coordinated and integrated.
- **Timeliness:** Accountable parties should act in a timely and coordinated way to prevent or respond to threats and attacks.
- **Assessment:** Risks should be assessed periodically.
- **Equity:** Management must respect the rights and dignity of individuals when setting policy and implementing protection.

Generally accepted system security principles (GASSP) included:

- **Certification and accreditation:** Information systems and information security professionals should be certified to be technically competent and approved by management for operations.
- **Internal control:** Information security forms the core of an organization's information internal control system.
- **Adversary:** Controls, security strategies, architectures, policies, standards, procedures, and guidelines should be developed and implemented in anticipation of attack from intelligent, rational,

and irrational adversaries with harmful intent or harm from negligent or accidental actions.

- **Least privilege:** An individual should be granted enough privilege to accomplish assigned tasks, but no more. This principle should be applied with increased rigor as the potential for damage increases.
- **Separation of duties:** Responsibilities and privileges should be allocated to prevent individuals or small groups of collaborating individuals from causing unacceptable harm or loss.
- **Continuity:** The organization's needs for continuity of operations should be anticipated and adequately protected and planned for.
- **Simplicity:** Information security professionals should favor small and simple safeguards over large and complex safeguards.
- **Policy centered security:** Policies, standards, and procedures should be established to serve as a basis for management planning, control, and evaluation of security activities.

These principles are also codified in the Organization for Economic Cooperation and Development (OECD) principles approved by nations around the globe. Other approaches are also available.

In each of these systems of governance, the information protection program is supposed to use the formal rules and the rule generation and appeals process to create a protection posture suited to the needs of the enterprise. Policies, standards, and processes are put in place to create the environment that fosters appropriate protection. But this is only the start of the overall process.

In addition to written rules of the enterprise, there are many unwritten social rules that play an important part in governance. People have to get along with each other, understand where they fit into the enterprise and where the CISO and those carrying out information protection functions fit. Social processes and influence are involved in clarifying these relationships. This means that, at the enterprise level, the people tasked with information protection must effectively work with both the written and unwritten rules to create the entire information protection program. This is where power and influence come into play and this is one of the many reasons that the CISO position must be properly placed within the enterprise in order to be effective.

Power and influence

Power and influence have been studied for a long time. The basic principles are outlined and depicted briefly in Figure 6-6. [O.6]

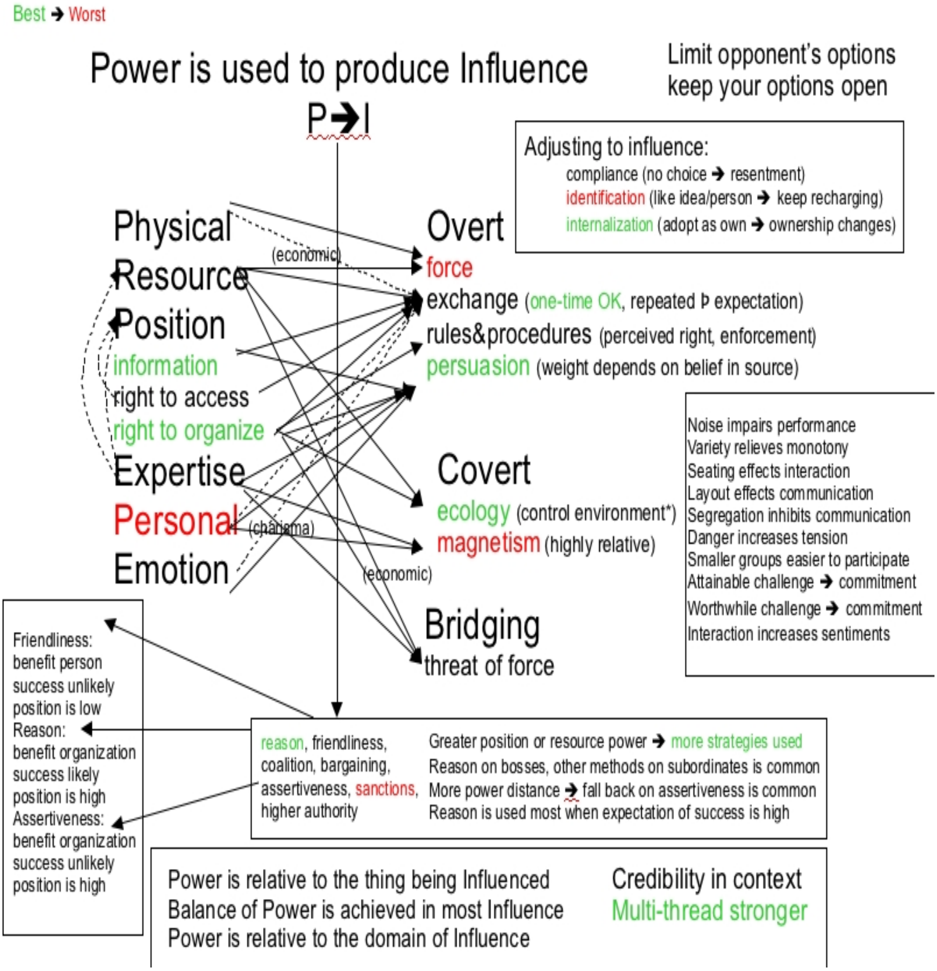

Best → Worst
Power is used to produce Influence
P→I
Limit opponent's options
keep your options open
Adjusting to influence:
compliance (no choice → resentment)
identification (like idea/person → keep recharging)
internalization (adopt as own → ownership changes)
Physical
Resource
Position
information
right to access
right to organize
Expertise
Personal
Emotion
(economic)
(charisma)
(economic)
Overt
force
exchange (one-time OK, repeated Þ expectation)
rules&procedures (perceived right, enforcement)
persuasion (weight depends on belief in source)
Covert
ecology (control environment*)
magnetism (highly relative)
Bridging
threat of force
Noise impairs performance
Variety relieves monotony
Seating effects interaction
Layout effects communication
Segregation inhibits communication
Danger increases tension
Smaller groups easier to participate
Attainable challenge → commitment
Worthwhile challenge → commitment
Interaction increases sentiments
Friendliness:
benefit person
success unlikely
position is low
Reason:
benefit organization
success likely
position is high
Assertiveness:
benefit organization
success unlikely
position is high
reason, friendliness,
coalition, bargaining,
assertiveness, sanctions,
higher authority
Greater position or resource power → more strategies used
Reason on bosses, other methods on subordinates is common
More power distance → fall back on assertiveness is common
Reason is used most when expectation of success is high
Power is relative to the thing being Influenced
Balance of Power is achieved in most Influence
Power is relative to the domain of Influence
Credibility in context
Multi-thread stronger

Figure 6-6 Power and influence

Power comes in many different forms and is directly applied in order to indirectly produce influence. The key to understanding the role of power and influence from a governance standpoint is that the individual responsible for coordinating overall information protection in an enterprise must ultimately have enough power to influence the enterprise to produce a comprehensive program that works. This calls for a combination of skills and mandates, as well as a capacity to deal with people. Power is also associated with physical capacity, resources, position, information, right to access, right to organize, expertise, personal charisma, and emotion. These play on (1) overt use of force, exchanges, rules and procedures, and persuasion; (2) covert uses of control over environment and personal magnetism; and (3) the threat of force.

In different organizations and with different people, different methods work. The skilled handler of power and influence will use these methods while remaining friendly with the vast majority of those they have to work with. The selection of such an individual requires an appropriate management decision and a negotiation process that produces the conditions required for success.

The methodology used to influence different structures is dependent on the specifics of the individual or group trying to create influences and the structure they are trying to influence. For example, a hierarchy is typically moved by moving someone at a high enough level to affect the desired change. If you go too high in the hierarchy you may produce the change by virtue of power, but the change may be considered trivial by the person making it and your credibility may therefore be reduced. If the person is above you in the hierarchy, you cannot change them by force, so you must change them with some other means of influence, perhaps a personal relationship, perhaps the force of logic, perhaps by elements of the environment, and perhaps by influencing others in their peer group to indirectly influence them. The number of situations is large, and the potential number of different paths of influence is enormous, but the interested reader can apply these principles and investigate these issues on their own to figure out how to reach and persuade the audience they need to influence.

Many enterprises make the mistake of taking a highly skilled technical person without adequate tools and mandates and expecting that the protection program will grow and prosper under this tutelage. Others make a similar mistake by putting a manager in charge of information protection who lacks the necessary understanding of the information protection field to make sound judgments about those decisions. Neither is adequate to the task. Some enterprises choose to manage protection by groups, but effective programs have a single individual who is in charge of the overall effort at an enterprise level. That individual creates and operates the groups that produce the results.

The key to success is choosing a champion who understands how to influence the organization, understands the technical issues to the point where they cannot be easily fooled, and understands the business to the point where they can help make reasonable and prudent decisions. For this individual to be successful, they will also need some other things, including but not limited to positional power, adequate resources, and expertise. Specifics should be identified in context and, for those who seek to hire or be posted into such a position, it will be important to identify these things during the interview process. A decision must also be made about whether to build or buy the champion. Hiring an outsider for such a position may be better if internal expertise is not available or if it will create personnel issues. If inadequate internal expertise is available in an individual, it may be prudent to augment expertise while an internal appointee is further developed.

Applying power and influence

Power produces influence and different forms of power are used to influence different people and groups in different ways. Skilled CISOs use power wisely.

Physical power

Because of physical security mechanisms and guard forces, physical security can be a means of exerting power. Having physical access to information systems and infrastructure, being able to lock offices or lock people out of facilities, and the use of guards to escort individuals to termination meetings are all examples of how physical power can be used to influence outcomes. Physical force tends to be overt and direct, although some level of indirectness can be used to imply a physical threat. But this is often considered undesirable in the context of an enterprise and should only be used when necessary for the situation. For example, physical escort is normally used when an employee is terminated, as disputes often arise in this context. Physical power almost always produces compliance in enterprises rather than identification or internalization. [O.5]

Resource power

Resources can come in many forms, typically from things that can be exchanged like money, facilities control (space), people (time), computing resources, network resources, control over the environment (ecology), and the threat of force. Overt resource power tends to produce compliance and, in some cases, identification. If the resource power is directed toward something that is already desirable to the target of the influence, identification and internalization can both be achieved, but usually only as long as the resource continues to be made available.

Positional power

Positional power stems from three aspects; access to information, the ability to grant access, and the right to organize. Information can be used for its exchange value or as a tool of persuasion. For example, it is sometimes used as leverage by allowing its use without disclosure or is sometimes concealed to create an elite class with the power to apply it.

- The ability to grant access can be used for exchanges, however if done repeatedly it creates an expectation of trading value that may be undesirable.
- Information and access rights are effective at producing compliance when that information forces an unpopular move, but more often it leads to identification in small quantities and internalization when it is particularly useful.
- The right to organize is typified by work roles, assignments, titles, pay levels, and so forth. It tends to lead to compliance at best when it is a demotion, but when used for advancement or restructuring with positive attributes, it leads to internalization or identification.

Positional power in information protection often operates through matrix management, project teams, reassignment of people to teams under the CISO, or other similar steps. But this influences other issues like budgets, organization sizes, and so forth, and all of these have impacts on how people think of themselves and the relative importance of managers. As a result, positional power has management interactions that can be touchy from a political standpoint and influence apparent financial results of organizations. As a simple example, if security is a cost center that is handled by the business unit but can be moved into the CISO's office and be handled as a cost by the enterprise, it reduces the apparent costs to profit centers so that individual business units can claim higher margins even though they may not have changed anything but the structure of the enterprise. These results may be tied to performance bonuses or other metrics that benefit individuals.

Expertise, personal, and emotional power

Expertise can be used for persuasion, magnetism, and as a threat of force. For persuasion, the weight of influence depends on the belief of the audience in the source. Magnetism is highly relative and, with some exceptions, is ineffective at high levels in the enterprise environment. The threat of force associated with expertise is based on the notion that the expertise can be used against others, but this is an oppositional perspective only really used in the context of questioning suspects and similar interview processes. Personal persuasion is based on a trust relationship that goes to liking and is commonly used in the enterprise as a way to leverage relationships for benefits. Emotional persuasion and exchange involves elicitation techniques. In the context of information protection, building up personal relationships is always a benefit in working through complex issues, and it provides a great deal of access and information that is of great value.

Persuasion model

For cases where persuasion is key, which comprise most of the situations encountered by CISOs, there is a well known persuasion model that helps understand these issues and address them systematically. This model is due to Chester L. Karrass[9.1.3] who is famous for his graduate work and subsequent courses on negotiations. The basic principle of persuasion is that change comes from the learning and acceptance of the goal viewpoints.

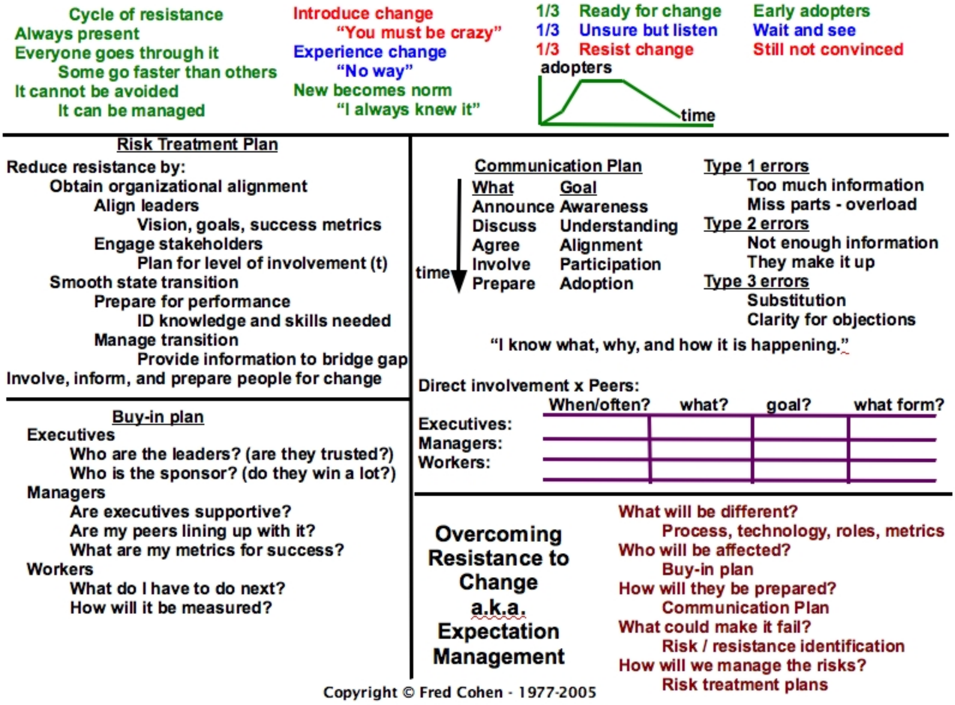
Learning comes from perceiving the message and understanding it. Acceptance comes from comfort with the message; it must be relevant and the person being persuaded must like the idea. This implies a certain understanding of the audience. Specifically, audience motives and value, information and language, perception and role, and attitudes and emotions lead to selection of techniques for persuasion.

- **Message content and appeal:** Studies have shown that persuasion is more effective if both sides are presented with the favored viewpoint presented last. The start and end of a presentation are better remembered with the end remembered best. Conclusions should be clearly stated, and repetition helps, thus the formulaic approach of saying what you are going to say, saying it, and saying what you have said. It helps to arouse a need and then fulfill it. Threats tend to be rejected, and it is better to put the desirable message first. In negotiations, the more you ask for, the more you get. It is better to ask for everything and only back off slowly in exchange for large concessions. It is better to stress similar points of view and reduce disagreements without belittling other views. Tying hard issues to easy ones sometimes helps to solve the hard ones. Being friendly and sympathetic help, and asking advice on how to resolve problems without sacrificing enterprise needs often generates a cooperative environment. Avoid creating defensive situations to prevent hardening views. Appeals to excellence, self worth, and fairness work.
- **Situation setting and rewards:** Setting is important to delivery of the message. Try to make the audience feel worthwhile and to reinforce their opinions. People like balance, but ambiguity upsets them and there is a tendency to resolve ambiguity quickly. Balance should be presented without unnecessary production of lingering ambiguity. If a problem is created it should be readily resolved by agreeing with the presenter's view. Social forces should be considered and the audience point of view must be accounted for. Facts, methods, goals, and values are used

to influence decisions, and power issues are always relevant.

- **Credibility:** If introduced as an expert, the presenter will be seen as one. Media, presentation, clothing, degrees, experience, and references tend to increase credibility. Avoid opinions on issues you don't know much about to retain credibility, particularly among experts.
- **Choice of media:** Letters are good when establishing justification or to get a letter back or when interruption is dangerous. Face to face is better when presence brings regard or respect when visual indicators help guide direction, or when more or less may be desired.

Managing change



Cycle of resistance
Always present
Everyone goes through it
Some go faster than others
It cannot be avoided
It can be managed

Introduce change
"You must be crazy"
Experience change
"No way"
New becomes norm
"I always knew it"

1/3	Ready for change	Early adopters
1/3	Unsure but listen	Wait and see
1/3	Resist change	Still not convinced

Risk Treatment Plan
Reduce resistance by:
Obtain organizational alignment
Align leaders
Vision, goals, success metrics
Engage stakeholders
Plan for level of involvement (t)
Smooth state transition
Prepare for performance
ID knowledge and skills needed
Manage transition
Provide information to bridge gap
Involve, inform, and prepare people for change

Communication Plan

What	Goal
Announce	Awareness
Discuss	Understanding
Agree	Alignment
Involve	Participation
Prepare	Adoption

Type 1 errors
Too much information
Miss parts - overload
Type 2 errors
Not enough information
They make it up
Type 3 errors
Substitution
Clarity for objections

"I know what, why, and how it is happening."

Direct involvement x Peers:

	When/often?	what?	goal?	what form?
Executives:				
Managers:				
Workers:				

Buy-in plan
Executives
Who are the leaders? (are they trusted?)
Who is the sponsor? (do they win a lot?)
Managers
Are executives supportive?
Are my peers lining up with it?
What are my metrics for success?
Workers
What do I have to do next?
How will it be measured?

Overcoming Resistance to Change a.k.a. Expectation Management

What will be different?
Process, technology, roles, metrics
Who will be affected?
Buy-in plan
How will they be prepared?
Communication Plan
What could make it fail?
Risk / resistance identification
How will we manage the risks?
Risk treatment plans

Copyright © Fred Cohen - 1977-2005

Figure 6-7 – Managing Change

Changes are always met with resistance. It is the nature of things. Managing change is fundamental to making the changes associated with information protection programs over time. The greater the change, the more resistance you are likely to encounter. When change is introduced you will hear “You must be crazy” and similar phrases. As people experience change, they will express disbelief. Eventually they will embrace the change if they believe it benefits them, and you will hear things like “No way! Really!!”. As the new becomes the norm, the doubters will brag that they were with you all the way and that they knew it all along. Such is life.

The success of the change plan depends on the way expectations are managed. The basic plan is to understand what will be different, who it will affect, how to prepare those affected, how the change plan could fail, and how to treat the things that could cause it to fail before they do so. The effort in planning change involves the creation of a buy-in plan, a communications plan, and a set of risk treatment plans. And of course it must be understood that no plan survives contact with the enemy.

The buy-in plan

Executives and leaders in non-hierarchical structures need to know who is leading the efforts for change and must build up trust in those leaders in order to buy into the plan. This often means finding a champion at the executive level in the right organization to help sponsor the effort. While some CISO's have direct access to and trust of top management, others are forced to create alliances for change with those above them in the structure. The sponsor of the effort for change is typically a top-level executive who, for one reason or another, decides it's in their or the company's best interest to make the changes indicated. The higher they are, the better the chances for success.

Managers and other facilitators have to gain executive support in order to see benefits in helping changes happen. While many security changes are started at with workers and grown “organically”, managers who become champions ultimately need to find executive support in order to make large scale changes. Managers also have to line up support with their peers in order to make changes and this is often a complex process involving a lot of leadership and time. Finally, managers have to find ways for their efforts for change to be reflected in the metrics used to measure their success in the organization, or they will be punished for their efforts. If executive management cannot be persuaded to create metrics that support managers making these changes, the program cannot be expected to succeed.

Workers predominantly need to know what they have to do next and how their performance in those tasks will be measured in order to buy into the effort. Their cooperation is vital to success, and they need to have sets of rewards and/or punishments in order to motivate them to willfully join in the effort to change.

The communications plan

Over time, the CISO will announce things for awareness to target audiences, discuss things with those audiences to develop mutual understanding, come to agreement so that people are aligned to the change, involve the targets to gain their willing participation, and prepare the targets so that they can successfully adopt the changes. The goal should be for the targets of the efforts to say “I know what is changing, why it is changing, and how it is happening.”

The targets of the change effort depend on the structure of the organization. For hierarchies, they typically include executives, managers, and workers that are directly and indirectly affected by the changes. For network organizations, replace executives with key influential people and work down the influence hierarchy to match up with managers and workers, providing all affected individuals with the information they need to understand, from their point of view, (1) what is changing, (2) why it is changing, and (3) how the change will happen. The communications plan should specifically codify when and how often each target audience should be communicated with and by whom, what is to be communicated with them and toward what objective (the what, why, or how of the change), and the form of the communication should be selected to meet the need per the persuasion model.

The communications plan should seek to avoid errors of omission (type 1), errors of commission (type 2), and errors of substitution (type 3). Errors of omission come when too much information is provided, leading to cognitive overload which causes important things to be missed. Errors of commission occur when not enough information is provided, resulting in people making up their own versions of what, why, and how. Errors of substitution happen when inadequate clarity is present to overcome mental objections and predispositions toward other answers to what, why, and how.

The risk treatment plans

The risk to change stems from the combination of natural resistance as described earlier, vested interests such as ownership of the previous approach, and specific reasons associated with risk management and other performance metrics of the enterprise. If these objections cannot be rationally overcome and influence approaches are ineffective, the road to change will be very hard indeed.

The process by which project-related risk of completion is treated is typically very different from the risk processes associated with protection-related risks, and this difference must be clearly understood. Organizational risks are mitigated by alignment of human forces and creating smooth transitions that don't unduly disrupt the normal course of business or create unnecessary friction.

Typically, organizational alignment starts with aligning the leadership around vision, goals and metrics for success. Once the leaders agree on these factors, stakeholders have to be engaged. It is usually a good idea to start engaging the stakeholders long before leadership makes a commitment, because leadership will ask stakeholders about these issues and, if the stakeholders they ask don't buy into the program, the risk of failure will increase substantially. It is usually a good idea to have a plan for involving stakeholders in various processes associated with the change This means getting their initial and ongoing support and continuing to keep them informed and involved at the appropriate level. In some cases this means finding ways to get stakeholders who disagree with the change to not disrupt the process, and this can be complex as well. For example, in some cases, effective change is attained by getting disruptive stakeholders to engage in other activities that make them too busy to interfere with the change. In other cases marginalizing their views to key group members or finding ways to give them something they want in exchange for cooperation works.

While this may seem complicated, underhanded, and unnecessary to the typical technical expert who has been moved into a management position, it is commonplace among managers and executives who have to find ways to get the system to support their, often competing, ends.

Smooth transition is desirable but not always attainable. The goal is to minimize friction and this is done by preparing people for what they will have to do and managing the transition from the previous operation to the subsequent one. To prepare for performance it is important to identify the specific information, skills, and knowledge needed by each of the different sorts of individuals involved. To manage the transition more smoothly, information has to be provided to bridge the gap between the previous and subsequent states. In other words, the idea is to involve, inform, and prepare people for change

Adaptation to contact

Of course all of the planning for change will not prevent resistance and that resistance may come in any number of forms. To prevent change, many people may push back against the change by using their influence and power, they may refuse to cooperate by withholding information, or they may try to use any of the other aspects of power discussed in the power and influence section of this guidebook. While risk treatment plans can cover many of these issues, there will always be some that are missing. Don't panic! Learn the ways to counter verbal and non-verbal attacks by studying the subject matter in more detail. Practice the methods outlined here, and try to understand what underlies the resistance so you can turn it away or build bridges that help you overcome it.

An example managing security consulting jobs

One of the best examples of the resistance to change comes when information protection posture assessments (IPPAs) are done for the first time in an organization. By the nature of an IPPA there has to be an internal sponsor and an external assessment team. If the internal sponsor doesn't properly prepare the field, the external team will meet with great resistance, but ultimately it is the job of the external team to help the internal sponsor do their job. The assessment team leader has a customer relationship and expectation management problem every time, but in this case it is at an extra level of indirection because the assessment team manager has to get the internal sponsor to manage the internal politics. There are two basic problem cases. They are; (1) someone with inadequate skills, power, and influence as the insider; or (2) an insider who is so powerful that they create powerful resistance.

In this recent assessment, the sponsor of the effort was the CEO of the enterprise. This is a very good situation in the sense that nobody is more likely to be empowered to get the job done than someone who comes at the request of the CEO. But on the other hand, the members of the executive team felt a bit put upon by the mandate from above to review their parts of the enterprise. So the resistance came from the top-level executive team members. In the meanwhile, one of those team members was told by the CEO to identify an appropriate person to coordinate the IPPA, and that team member chose someone who had essentially no power or influence, but a great deal of technical expertise to run the effort. So the full force of the CEO was theoretically available but the CEO was unavailable to use any of that power to actually help get the job done, creating the most powerful possible set of resistance and the least powerful internal lead.

The assessment itself had many problems, including people who refused to attend meetings or missed scheduled meetings, people who told their employees not to be helpful and tried to mislead the assessment team, people who instructed protective forces to act in very abnormal ways so that they could keep the assessment team from finding any flaws, and the list goes on and on. Eventually, a draft assessment report was provided that indirectly indicated the need for a powerful lead for information protection without unduly offending the internal assessment lead who was the first reviewer. Over the assessment team leader's objections, the document was sent to too many people in its draft form, eventually leading to a great number of complaints about factual errors on parts of the draft report that had not been vetted yet.

When a private meeting with the CIO was requested to review the draft report and its findings, the CIO agreed but then set up a meeting with all of the leads responsible within the CIO organization and tried to get the assessment team leads to come to that meeting. It was explained to the CIO that this was the part where we interviewed her and not the part where her team was supposed to be present. We indicated that many of the things we were going to discuss were likely to be things that she might not want her team members to know about, like their performance in various areas from her perspective and staffing level issues. The meeting was held with the CIO alone, and lasted for several hours, much of which was spent discussing the sorts of things she should have been informed about through an effective communications plan, but was not, because the internal lead did not have the necessary access to the CIO to get the job done.

The CIO seemed to start to embrace the process after this meeting, but the resistance was not yet over. A copy of the current draft report was sent to the CIO that evening as she prepared for an executive committee and board meeting the following day. She had decided that she was being treated unfairly by the process and was determined to discredit the report before anyone else had a chance to read it. Of course the report was still a draft because she hadn't had the opportunity to make her comments on it, and she was still unhappy about many of the features of the report because someone else within the organization had prepared her for the report by indicating to her that it would be very negative toward her, even though it has actually said nothing about her because she had not yet been interviewed. The study lead got wind of the situation through other internal sources close to the CEO and the CEO was prepared for any conflict based on the notion that the report was still a draft and that with her cooperation, the CIO could help to form the report into a helpful internal document. She could not really argue the point.

Several weeks later the CIO requested an additional meeting to discuss the draft report in more detail. This meeting started at 4PM at her office and went till about 9PM by the time it was all over. In this meeting, the assessment lead spent time one-on-one with the CIO discussing many of the issues in the report. It was clarified that just because many potential vulnerabilities were found didn't mean that they all needed to be fixed. The risk management process was better explained as a process of executive decision-making and not a process in which the executive is forced to spend budget on things they don't want. Some disputes from a key manager as to findings were reviewed and the assessment lead indicated that the individual had put up substantial resistance forcing the team to check the facts out very carefully. The lead indicated they he personally saw the records, checked the facts, and that the manager had tried to avoid the process and prevent the information from being made available. This was indicated as a management problem that the CIO needed to address. The top priorities for fixing critical issues were discussed at length and the CIO came to agree that the identified items really did need to be fixed and that nothing identified as having to be done was unreasonable or inappropriate. Cost and internal political issues were discussed at length and a lot of clarity was achieved.

About 2 months later, a verbal presentation was made to the executive committee regarding this assessment and a series of related assessments that were underway. When the IPPA was discussed and the results reviewed, the assessment team lead was prepared for further resistance, but none came. The CIO indicated to the executive committee that 5 of the 7 key recommendations were either already completed or were being implemented. The remaining recommendations were long-term and would likely be undertaken in the appropriate time frames.

It should be clear that this was not the ideal situation and that the key elements of the change process could not be handled by the internal assignee. But despite the problems along the way, the change was made and it eventually came to be embraced. Learning to manage change is key to the success of the CISO.

Enforcement, appeals process, and disputes

People don’t always get and stay with the information protection program as well as they should. Attackers almost never do. Insiders who are abusing systems for their own advantage also tend to not obey all the rules. But even the average worker who wants to do the right thing may have some difficulties always getting it right from a protection standpoint. An example of doing the right thing and having the program fail, is when diligent employees place sensitive information in shred bins. In many installations these bins are easily opened. When opened by strangers in the area, no challenges are made. The result is a program that provides attackers with ready access to the most sensitive information.

Enforcement

When the program is not followed, there must be an enforcement mechanism that results in detection and reaction to significant protection irregularities before they result in serious negative consequences. Similarly, when a program is followed and the result is worse than if it were not followed, there must be mechanisms in place to detect and adapt to these irregularities before they cause serious negative consequences. These mechanisms ultimately tie into the overall management feedback system because, in addition to any automated technical responses, the management process must also produce appropriate human responses that are consistent with company policies and needs.

Enforcement is produced through the enterprise governance process. It may involve a direct supervisor, the HR department, computer security personnel, the owners of affected systems, law enforcement, private investigators, the legal department, internal committees, and executive management. It might result in anything from a minor change in a detection threshold to a large-scale adaptation of the enterprise protection posture and civil or criminal litigation. Because of the wide range of possibilities, the governance process should produce protection policies, standards, and procedures that intermingle with the HR policies, standards, and procedures to define the enforcement process.

Significant documentation requirements exist for these processes including documentation that demonstrates that:

- processes are followed,
- forensic data is properly treated,
- legal and regulatory reporting requirements are met,

- sanctions are properly and consistently applied without discrimination based on legally restricted bases, and
- other enforcement documentation as dictated by the legal department.

Despite the need to assure a uniform process, there are different processes for individuals of different status, such as employees, contractors, visitors, suppliers, customers, and others. Each has different background, training, awareness, restrictions on, and punishments for violations of terms of use. For customers this is particularly problematic because a violation of a policy or work rule cannot be summarily dismissed, but it might be a sensitive area that causes loss of business if improperly handled. Misdeeds of contractors are the responsibility of the contractor, however, in practice, it is difficult to do the same level of training for contractor employees as for internal employees who, presumably but not uniformly, have more longevity with and dedication to the company.

Like the legal system in general, administrative punishments may be defined in terms of the specific violations of policy involved and have to meet a standard of consistency in order to prevent wrongful discharge and discrimination-related legal processes from being successfully invoked.

Many enterprises have made the mistake of treating executives differently from lower-level employees when they violate protection policies, and some have found themselves in situations where they end up covering up serious criminal acts as a result. This is always a mistake – executives should be treated like any other employees when it comes to violations of laws or policies. In many cases executives are actually employed by a holding company that may have different work rules, benefits, and so forth. This is problematic as well, and yet it certainly does and must exist in many enterprises.

Enforcement may, of course, have some discretionary aspects to it. For example, a supervisor may determine whether an event warrants action based on the context of the event. An employee who uses a short password on a system in violation of a generic enterprise password policy may be in technical violation of that policy, but if the system only allows short passwords, the employee cannot reasonably be taken to task over it. Many policies allow discretion and the employee history or relationship with management may dictate different responses to similar incidents. Again, these are governance issues that should be addressed and included in training for all those who have a role that allows such discretion to be exercised. Documentation is also critical for such situations.

At higher levels, enforcement becomes problematic because of power issues. Even though a corporate policy is in place, a powerful enough manager or executive may simply disregard the policy for their part of the enterprise. If management doesn't enforce policy, the policy becomes an even bigger problem because the company may be subject to legal actions for failure to follow their policies when they would not be subject to those actions without those policies. Enforcement at these levels sometimes must be dealt with by top management. The power and influence of the individual tasked with information protection governance must be effective enough to protect the enterprise at this level.

Contractual obligations create other security-related enforcement issues. For example, the inability to effectively control consultants and other off-premise contractors leads to their ability to violate policies, standards, and procedures almost without recourse. In the United States, independent contractor work rules cannot be enforced under the IRS code without turning them into employees and thus granting them undesired legal status, access to benefits, and so forth. There are contractual ways to create proper conditions for these consultants and independent contractors, but they must be specifically attended to in information protection governance in order to get them right and adapt them to changing needs over time. This implies that security governance must interact with the legal department to create the proper conditions for contracts.

Contractual issues also apply to duties to protect provided under:

- Intellectual property controls, including trade secret, patent, trademark, and copyright,

- Private information associated with trading partners such as those covered under contracts for services,
- Contract terms that may be required with trading partners for one reason or another,
- Privacy policies that are announced to the public such as those relating to information collected on Web sites and paper contracts,
- Safe harbor agreements such as those associated with European Union (EU) privacy regulation,
- Classification-related contractual requirements typically associated with government contracts in different jurisdictions,
- Health-related information about individuals, like those covered by the Health Information Portability and Accountability Act (HIPAA) agreements required for exchange of data with third parties, and
- Financial information protected under the Gramm-Leech-Bliley (GLB) act, the Sarbanes-Oxley (SOX) act, and other similar and related regulations requiring contracts associated with information protection.

These and other contractual obligations may place nearly arbitrary constraints on select classes of information, and thus it is vital that the enterprise be able to separate information based on applicable protection requirements (typically called a classification scheme) and enforce different rules about information protection with respect to each of these different sorts of information. Typically these approaches include (1) a clearance process so that individuals who have proper characteristics, backgrounds, and training associated with handling of different sorts of information are given clearances to access different categories of information, and (2) need-to-know and need-to-use provisions so that only those individuals working on efforts relevant to the information have access based on their use of that information in their jobs.

Disputes

Ultimately, disputes happen and people don’t like decisions that are made. This can end up in one of three ways. It can lead to a legal process that extends beyond the border of the enterprise, it can lead to an acceptance of the decision, or it can lead to an internal adjudication process. Acceptance is preferred, but internal process is necessary to reduce the number of external processes.

The internal appeals and dispute resolution process is a critical part of protection governance. It should be built into policies, typically through provisions for policy override by someone at an appropriate level of management. Standards and procedures should codify the policies regarding appeals and disputes so that they are handled in a uniform manner that is consistent with HR standards and processes as well as regulatory or contractual requirements. The appeals process and each instance of its use should be documented.

This process is greatly complicated in cases where subcontractors are involved and inadequate contractual provisions are provided. This is one of the reasons that effective governance must integrate with the legal department in creating standard terms and conditions related to all external contracts. Appeals processes for contractors are typically non-existent in non-government contracts. The lack of such processes tend to lead contractors to violate policies that interfere with doing their job. For example, if internal requirements indicate the need to encrypt all email but the contractor has an internal email system in trusted infrastructure, they might ignore enterprise policies for internal communications and risk being dismissed. If they could notify the enterprise of the condition and ask permission, they would not have to ignore policy and might get the exception.

Disputes at higher levels tend to be settled by negotiation processes. Some of these are subtle and indirect, while others are far more formal, in some cases leading to litigation. A typical example is worker monitoring where disagreements between unions and management have led to several legal and arbitration cases. In most such cases, proper governance can dramatically reduce the number, cost, and complexity of these disputes.

In some cases, disputes have to be settled by external bodies. For example, in resolving matters associated with clearances an external body is often involved for higher level clearance issues associated with government contracts. Other legal processes may get involved as well, sometimes leading to arbitration, settlement negotiations, and civil or criminal cases. It is almost always better to settle disputes internally, however; sometimes there is no choice.

The enforcement and appeals process is perhaps the trickiest area of governance. While physical force, like the use of guards for escorting newly terminated employees, may ultimately be involved in security processes, clearly this is to be avoided where possible. This is particularly problematic between executives who may not buy into information protection ideas quickly.

Top management buy-in and support

A goal of any effective information protection program from the standpoint of the CISO is that the top management team embrace and assist in the effort in a highly cooperative process. This is yet another reason that the CISO must be part of that management team or closely linked to its members. Without this sort of cooperative arrangement, the tensions and disputes that ultimately arise will cause the program to fail, power struggles will dominate the best interests of the enterprise, and business will suffer.

It is fundamental and yet often inadequately understood or stated that the goal of information protection should be to optimize business results. On this, all top management should agree. But one of the major challenges of the CISO is providing adequate understanding to other top management team members to gain their support in initiatives. If a policy has an enforcement mechanism that results in an override by a management team member, the CISO should agree on the override and it should be a sound business decision, not a reaction to a newspaper article or a feeling of some sort. This implies a level of communication and maturity that is sometimes hard to find, and a common basis for judgments.

Power and influence and managing change

Dispute resolution often comes down to who has what power and influence and how that power and influence is applied. Different organizations also apply power in different ways. The direct application of power is considered uncouth in many organizations, while other organizations are run by fear with power exercised directly and brutally. The successful CISO must have the necessary elements of power and influence to get the job done, must be able to apply them properly, and must be able to extend them to team members to get their jobs done. Many disputes can be avoided by proper influence and clear communications. Power and influence are wielded in many ways. The effective CISO has to learn to manage change within the organization. This involves understanding how to effectively communicate the right information to the right people at the right time and providing the proper motivations to allow others to embrace or at least allow the changes to take place. The change process and how it is managed is a key governance issue, particularly for organizations in transition and newly anointed CISOs making their mark.

Responses to power and influence

Effective use of power and influence can result in compliance, identification, and/or internalization.

- **Compliance** is associated with forced behavior with no choice. This produces resentment and may also result in malicious compliance. Work-to-rule responses are common in compliance situations where workers are very careful to follow every rule to the letter and with great care, thus dramatically reducing output but still not breaking any rules.
- **Identification** comes from liking the idea or the person who is asking for the behavior. It results in behaviors that, with a little awareness and support, get done reliably and without resentment.
- **Internalization** is associated with the adoption of the influence as their own. The individual takes ownership of the idea and works not only to support it but to improve and personalize it.

The objective in most cases should be to go beyond compliance. Identification results in increased adoption and is far better than compliance. Internalization is sometimes better, but in a security context, internalization is not always most effective or desired, because adopting an idea may lead to personalization and this can lead to non-compliance, particularly when the standards are very specific and inflexible. Regulatory compliance and contract issues are examples where identification may be better.

Measurement of the difference between compliance, identification, and internalization is important to the success of a program but may be problematic given available measurement techniques.

Other power issues

The independence of the information protection function is critical. While friendly persuasion is a valuable approach, force must be available to allow the CISO access to the resources required to do the job, independent of, and without interference from, or knowledge of, others. The power relationship between the CISO and the rest of the company must be designed to allow protection to be effective. It is a fundamental separation of duties issue as identified by GASSP and the OECD guidelines and it is necessary in order to assure independent evaluation for the executives who have to attest to conditions that put the enterprise at risk. Force must be an option to handle situations without other options, but it must be used wisely, be transparent to authorized reviewers, and be properly attributed so that responsibility can be maintained.

Force is necessary when time is of the essence and cooperation is not given or is not feasible because those in control are subjects of the process. Force may have to be covert to be effective, suggesting substantial complications in its use. Both overt and covert forces should be available under proper control. Covert investigations typically involve physical devices and logical mechanisms and thus physical security elements are involved in the process. Covert force generally ends up involving the legal team unless the legal team is the subject of the process, in which case it involves the CEO and external counsel. Investigations of the CISO are typically done by the CEO in conjunction with outside counsel and private investigation firms.

There are typically inflection points in the exercise of power to achieve large-scale changes. Most people try to avoid the use of overt power because of its tendency to produce compliance rather than internalization. But internalization takes time and, in many cases, time is of the essence in information protection. Most CISOs try to work cooperatively except in cases of dire need. They try to bank good will as an approach to achieving identification with the large-scale changes desired for the enterprise and to allow them to continue to be effective when they have to use force. There is a tension between the use of force and good will, and good will wanes with time unless propped up on an ongoing basis. So a key skill for the successful CISO is knowing when the severity of the situation warrants the use of force and when good will is more important than resolving an immediate situation with force.

The enterprise protection control system

Together, the elements of governance described here produce and implement a control system that is at the heart of what information protection governance is about. The control system typically involves a decision-making body that is guided by an individual who is tasked with information protetion governance. The decision-making body depends on inputs it gets from external sources, from the individual tasked with governance, and from enterprise-wide feedback mechanisms. It exerts influence through the full diversity of actions available using the powers vested in it, with the enterprise lead ultimately using those powers to influence the enterprise to be properly protected.

Most complex systems are hard to control. As a result, many different viewpoints exist on how this control works, and it works differently in different enterprises. Some enterprise CISOs assert that they control the protection program by directly controlling budget, having approval processes in place, and otherwise placing observation and actuation points throughout the enterprise. Others assert that control is grown by providing moral guidance and setting minimum standards of behavior that create an environment in which protection-supportive behaviors become the social norm. Still others create and operate collaborations with many others throughout the enterprise and become an advisor with high-level influence, never forcing any issue at all. Most CISOs find mixes of these strategies for different situations.

These approaches and many others are all valid, as long as they are effective at meeting the information protection governance requirements of the enterprise. This brings us to the fundamentals of control. In order for a control system to be effective it must have three things:

- Adequate numbers, placement, and sorts of accurate enough sensors to measure meaningful observables and report them.
- Adequate numbers, placement, and sorts of effective actuators to induce desired system behaviors.
- An analytical system that uses sensor and state information to produce actions that operate the enterprise within the desired control envelope.

Most enterprise information protection control systems have limited sensors and actuators and rely on individuals with special skills to compensate for all of the other failings in the system on a case-by-case basis. This often produces failed governance programs, negative audit reports, dramatic failures and mitigations, unpleasant security-related surprises, costly regulatory-forced changes, and other similar substantial negative consequences.

The enterprise information protection control system is comprised of all of the security-related actions taken by, information gathered by, and decisions made by all of the groups and individuals discussed here. Each security function has responsibilities and tasks to carry out. Meetings are designed to provide control over these functions. The power exerted through the groups and the information they bring back form the actuators and sensors of the control system at the management level. The decisions taken by the groups and individuals in those groups form the analytical part of the control system and should be designed to meet the control objectives of the group and, by implication, of the enterprise.

Within these control systems there are more in-depth and detailed control systems. For example, when a network zoning request shows up at the zoning board, it has to be passed through technical and non-technical experts in other groups for evaluation, and actions on the request have to be documented through their systems. A record is created reflecting the life of the project that involves records and decision processes at every level and in every area appropriate to the need. Each of these groups and the actions of their members are themselves control systems acting within the zoning process and, of course, also within their own functional areas. Ultimately, every individual is a control system that participates with others in other control systems, forming the overall set of controls that operate the enterprise protection program.

Metrics

For any control system to work, it must have ways to measure inputs from sensors and lower level control systems and to be measured by higher-level control systems. Measurements come in many forms. Some measurements may be turned into metrics to allow them to be compared with other measurements. In many cases, metrics are created for comparison but have little to do with the control objectives of the systems that use them. These are a waste of time and effort and tend to obfuscate the control issues. Meaningful metrics have utility in meeting the control objectives of the control system. Two sorts of metrics must be generated by each control system to be effective in context; (1) metrics that allow local control objectives to be met, and (2) metrics that are inputs to higher-level control systems and that are meaningful for achieving the higher level control objectives.

Costs

Generally, costs are critical metrics for business-related control systems. At the top level, costs are typically rolled up for accounting purposes and associated with a classification system to allow management to understand what they are spending money on and what they are getting for that money. While security costs are hard to track because of hidden costs, indirect costs, cost center accounting practices, and so forth, tracking costs is really the easiest part of security metrics. Costs are typically tracked in only one form, monetary units, and they are fully commensurable with each other and fungible, even if, from a budget standpoint, they may not appear to be either commensurable or fungible.

Performance

The hard thing to get a handle on is meaningful metrics for performance. Metrics that are useful at the governance level are those that provide measurements of the effectiveness of the comprehensive approach. The Capability Maturity Model for

Security (CMM-SEC) provides a way to measure progress of an overall program in terms of normalization into enterprise operations. It associates any of levels 0 through 5 (none, initial, repeatable, defined, managed, and optimizing) with each of 11 process areas and 11 organizational issues and is mapped against each of risk management, engineering processes, assurance, and coordination to provide an overall picture of the maturity of the information protection function within an enterprise. The ISO 17799 standard is often used as a basis of comparison by qualitative scores ranging from poor to excellent for each of the areas covered. Similar scoring with GASSP or the elements of the overall program also work. Roll-ups of these scores are often presented as a status and measurement of the overall program. For example, if out of 50 high-valued systems, separation of duties was poor in 35, fair in 10, and good in 5, separation of duties might be ranked as poor for the enterprise.

Time

CISOs normally provide top management with a plan intended to go from where they start to where they want to be in some time frame. They measure progress against those goals over time and sometimes alter the plan to suit changed assumptions or conditions. Once the overall protection program reaches stability, it may be run on a maintenance basis with efficiency measurements dominating progress measurements. Progress is typically measured by the CISO and independently validated by periodic information protection posture assessments (IPPAs). An IPPA is often performed at the start of a CISO's tenure to get initial values for the program and set prioritized objectives over time.

Lower-level metrics

Measurements from lower-level control systems feed into higher-level control systems eventually reaching the CISO's measurement process by providing information that is relevant to governance issues. But most of the groups that provide this information to the CISO operate within more detailed areas. As a result, these groups tend to think in terms of the work they do instead of the issues addressed in top-level governance. It is the task of the CISO to:

- Define metrics that are meaningful to the top-level governance issues
- Identify which metrics are to be delivered in what format by each group
- Provide groups with what they need to generate those metrics
- Turn these lower-level metrics into CISO metrics.

As an example, suppose the CISO wants to measure the maturity of the incident response process according to the CMM-SEC methodology in order to provide a metric on that element of the overall program. This measurement goes across risk management, engineering, assurance, and coordination and involves ratings for each identified area. Answers should be in the range 0-5 for none-optimizing.

Risk	*Risk mgmt*	*Engineering*	*Assurance*	*Coordination*
Administer controls				
Assess impact				
Assess risk				
Assess threat				
Assess vulnerability				
Assurance argument				
Coordinate security				
Monitor posture				
Provide input				
Specify needs				
Verify and validate				

Table 6-8 – A sample roll-up metrics table

A roll-up metric can then be generated by weighting each cell in the matrix and producing a weighted average. Program roll-up can be done by normalizing and weighting program elements to provide an overall weighted average.

Budgets and funding

Without adequate funding, security governance will fail. But the problem of determining how much funding is required is far from simple to solve. There is very little publicly available information on security budgets or numbers of incidents and losses, and even if this information were available, what is covered by the protection program and the specifics of the enterprise are critical to gaining useful understanding of these issues. In addition, security costs are often hidden costs, so it's hard to get a good grip on the cost issue even if someone cares to try.

The hidden costs of security

Those who have been bold enough to publish this sort of information come to several generally agreed values.

- Systems administration and security, which are inexorably intertwined, account for about 5% of all of the time and expenses involved in using computers. This includes 1 out of every 20 people using computers, and 1 out of 20 dollars spent on hardware, software, maintenance, operations, and management for everyone who deals with information technology in their work.
- Costs of regulatory compliance during periods between the introduction of new regulations and their normalization in the enterprise have been reported in the range of 8% of total IT budgets by large enterprises. In normal operation, these same companies reported 4% of total IT budget for meeting compliance requirements.
- Financial institutions evaluated as having a reasonable, prudent, and effective program relative to common standards such as the GAISP, ISO17799, and this book's evaluation scheme typically spend from 10% to 20% of their IT budget on information protection.
- Physical security is also required for information protection to be effective. The cost of physically secured facilities, depending on the specific security needs, ranges from 2 to 10 times the cost of normal unsecured office space.
- Costs tend to be dominated by operations rather than acquisitions. As a rule of thumb, operating costs per quarter equal acquisition costs for security-related hardware and software if operated for effective protection.

Each enterprise is unique and these numbers depend on factors ranging from business location to competitive environment. Metrics widely available today are inadequate for providing better information.

But perhaps more important than how much funding is needed is where the money comes from and goes to. This is largely dictated by organizational structures and responsibilities. Sometimes it is taken “out of hide” from budgets of programs and business units throughout the enterprise, or gets expressed as a tax or charge-back for services rendered. But the overall protection program clearly involves many different hides and trying to assess strict values to protection services is problematic. Companies do not normally account for security costs accurately and tend to keep them as hidden costs. When the accountant who knows something more than their neighbor about a Trojan horse scanner helps others run the scanner, this is not accounted for. The training time associated with information security awareness programs is rarely accounted for, and the increased time and effort associated with security-related inconvenience is generally ignored in bookkeeping systems. As a result, it is difficult to really say where the costs of an information protection program come from or go to.

One of the things that CISO's should try to do is get a handle on these costs and try to track them. A simple approach is to get estimates from individuals throughout the enterprise by making a list of the places they likely spend time and effort by asking them to fill in a chart. The chart is then collected for a statistically significant number of individuals per type and analyzed to get an expected cost across the enterprise. A more accurate and more expensive approach is to observe individuals and try to gather enough statistics to give a good indication of the actual costs. In either case, this is typically done as the maturity of the program increases with the objective of trying to optimize the program at a fairly fine-grained level. Table 6-9 can be used to carry out such a survey. Additional rows may be needed for your enterprise.

There are some costs that are clearly designated as security-related and accounted for as such. These are usually related to the structure of the program.

Area of likely cost	*$$/hrs*
Time to authenticate	
Performance degradation from encryption	
Time spent helping others with security issues	
Time spent reporting or responding to incidents	
Time spent in security awareness training not charged back	
Cost of extra software for security requirements	
Installation, maintenance, and update time for security software	
Time delays in booting up or logging in from security scans	
Delays while running programs for security-related issues	
Costs of multiple authentications after initial sign-on	
Help desk calls related to lost passwords	
Costs of having to shut down and restart for security reasons	
Time wasted during security-induced outages	
Time spent in backups not centrally managed and accounted for	
Time spent in security-related documentation	
Time spent reviewing security-related policies and contracts	
Time spent in gaining additional approvals for exceptions	
Cost of delays from authentications for external access	
Relationship costs of security requirements met and unmet	
Excess costs of fire and media safes over file cabinets	
Time spent in doing security-related paperwork and process	

Table 6-9 - A sample hidden cost data collection sheet

Typical budget numbers

Some industry sources provide gross budget numbers for information protection, but these have little relevance to the actual budgeting process in most enterprises. Different enterprises use different budgeting processes, from central committee-determined project approaches, to hierarchical budgets derived from top management projections, to emergency funding for critical projects in response to incidents. The lack of clear guidance and figures for information protection budgets belies a lack of clear understanding of the protection process and a lack of meaningful ways to codify true costs. A good top-level CISO who is in charge of the process should be able to generate meaningful financial metrics within a year or so of taking over the position if adequate cooperation is provided. But these will almost always be cost metrics with performance associated with fulfillment of organizational goals rather than income metrics or return on investment numbers.

Direct budget for the CISO

While no attempt will be made here to create the overall budget associated with the protection process, some things are clear just from the governance requirements identified. Funding on the order of $2-3M per year for the CISO and top-level team will be required to cover salaries and overhead. These are critical components of overall enterprise protection and trying to cut corners here is a mistake. The CISO team typically also has $2-3M in discretionary funds available for meetings, improvements, travel, initial tests of new security technologies, expert consultant time, keeping up-to-date, and other similar items. This does not include any of the funds required to run security operations. It only covers management of the process at the enterprise level.

Identifiable costs

To effectively measure the information protection program in terms of costs and performance, costs must be identified and performance measures put in place. While hidden costs remain, it is important to track identifiable costs from all sources in order to get a handle on what they are and where they come from.

Each enterprise and many business units within enterprises make their own decisions about how to budget and pay for protection. Many enterprises ignore some of the protection issues or consolidate them within other areas. Ownership of issues varies widely, but as a rule, if they are not otherwise owned, they end up owned by the CISO. Budgets for widely used common items like anti-virus solutions or forensics tools often start in some niche area but ultimately end up borne on an enterprise-wide basis. These budgets and the functions they pay for should end up owned by the CISO so they can be properly accounted for, optimized over time, and eliminated or changed if appropriate.

Some costs are readily identified. For example, a typical security awareness program costs from $10 to $100 per user per year, not including trainee and team leader time. But budget sources for training and awareness vary from company to company. Some companies track training time to an overhead account within each user’s organization, some use charge-back systems to account for the time, and others roll training time into general overhead budgets or project budgets for the project generating the training requirement. Some companies have separate training budgets and treat educational efforts associated with specialized information security experts as part of the benefits package. Regardless of how the budget process works, it takes money to perform these functions, and over the long run that money must come from some identified funding processes in order for the program to become normalized within the enterprise.

A typical centralized cost is a corporate license to a virus or spam defenses and other similar content controls. One example of the advantage of centralization in the CISO' s office was played out in a recent decision by a large enterprise to switch vendors. The savings amounted to $500,000 per year in reduced cost for a solution that the protection testing and change control team found to be otherwise equivalent to its competitor. The process put in place before there was a CISO to make this enterprise-level decision was based on a brand name and involved no significant testing of alternatives. It was operated for several years out of the CIO's office without any attempt to optimize. Several other similar steps by the same CISO have resulted in millions of dollars of cost reductions while improving protection effectiveness and regulatory compliance.

Another critical area that can only be handled centrally is business continuity and disaster recovery planning. Budgets in this area can be very substantial, including the creation of a redundant capability for all critical business functions. Amounts in the range of $20M per year for business continuity planning and related facilities are commonplace. Coordination involves activities ranging from holding scenario development exercises with top management to regular practice of components and the whole plan. These pay off only if disaster strikes, but without them, business collapse is inevitable.

Area	Source	Annual Costs	Hidden costs
Security management	CISO	$5M	n/a
Policy	CISO	$200K	Churn, time, morale
Standards	CISO	$200K	Churn, time
Procedures	Distributed	$200K	Churn, time
Documentation	Distributed	Unbudgeted	All hidden
Security Auditing	Audit or CISO	$150K	Churn, time
Protection Testing	Varies	$450K	Time
Technology	Distributed	$5M	Churn, time, morale
Personnel (training)	HR	$300K	Time
Incident handling	Varies	Unbudgeted	Churn, time
Legal	Legal	$250K	All hidden
Physical	CSO	Provided	Time, morale
Knowledge	Benefits	$2K/course	n/a
Awareness	Varies	$250K	User time
Organizational	CISO	Unbudgeted	Time, stress
Business lifecycles	Business Units	Unbudgeted	Many of them
People lifecycles	HR	Unbudgeted	Time, moral
System lifecycles	System owners	Unbudgeted	Time, churn
Data lifecycles	Data owners	Unbudgeted	Time, churn
Deterrence	CISO	Unbudgeted	Time, churn
Prevention	Varies	$100K (HW)	Time, morale
Detection	Varies	$100K (HW)	Time, morale
Reaction	Varies	$1M(IDS team	Time, morale
Adaptation	Owners	Unbudgeted	Churn, time
Integrity	Data owner	Unbudgeted	Churn, time
Availability	System owners	$20M (BCP)	Churn, time
Confidentiality	Data owner	Unbudgeted	Churn, time
Use control	Business owner	$5M (IdM)	Time, morale
Accountability	Business owner	$5M(retention)	Time, morale
Risk management	Enterprise	$500K	Error costs
Insurance (transfer)	Enterprise	Unbudgeted	Policy limits
Losses	Enterprise	Unbudgeted	Churn, time, morale
Mitigation	CISO / owners	Unbudgeted	Churn, time, morale
Public relations	Communication	Unbudgeted	Stress, morale
Brand	Communication	Unbudgeted	Stress
TOTALS	N/A	$43.2M	N/A

Table 6-10 - Typical budget figures

In making up a list of costs, the structures used for identifying protection processes above are used. For each area, determine what costs are tracked, who pays for what, how much they pay, how they do it, and what costs remain hidden. These can be listed in a table. The example in Table 6-10 is from a newly created business unit in a large financial institution. Some costs are expected to be double in the first year because of the use of consultants for program development.

As this example shows, costs vary across different functions. Costs of $5M per year for retention of records is more than many enterprises spend in information protection budget, and the $20M in disaster recovery costs vary greatly depending on the business and the nature of the operation.

How long will it take?

Hierarchical control systems tend to have more and faster actuators, sensors, and control mechanisms at lower levels of the hierarchy than at higher levels. As a result, time scales in information protection range from microseconds at the lowest levels to years at the strategic management level. One of the goals of an effective control system is that decisions with high consequences requiring in-depth analysis can be made in longer time frames than decisions with low consequence requiring little analysis. This provides the added observation, orientation, decision, and action (OODA) time required to get better answers deployed at larger scales for higher consequences. Information protection sometimes requires large-scale efforts be undertaken at a fast tempo, but a well designed information protection control system does not need to react to events as much as it needs to adapt to changes in the environment and the enterprise direction.

To get from an ad-hoc system of controls in an environment that is not structured for proper governance and control to an efficiently working governance and control structure takes a long time. Even if everything were instantly put in place from a systems standpoint, the people and processes take time to develop and adapt. There are life cycles associated with businesses, people, systems, and data that cannot be pushed too hard without causing them to break and the overall system to fail. One of the useful ways to measure time in terms of governance is to identify how long it typically takes to move from one level of performance in the governance metrics to the next level, assuming that adequate support in terms of power, influence, cooperation, and funding are present. If these are not present, the times are of course longer.

In typical IPPA studies, three time frames are considered. The urgent time frame is typically from immediate to 6 months and involves high consequence situations with glaring vulnerabilities and subject to threats with demonstrated capabilities and intents to attack. In the tactical time frame of 6 to 18 months, the time typically required to complete a substantial infrastructure project, governance issues are typically addressed in terms of moving from the “none” or “initial” level of the CMM-SEC metrics to the “repeatable” or “defined” level. Similarly, ratings in ISO 17799 and GAISP metrics can be improved by one step (e.g., from “poor” to “fair” or “initial” to “repeatable”) in this time frame. In the strategic time frame of 18 months to 3 years from the start of the effort, enterprises can move from the “defined” to “managed” level of the CMM-SEC and up another few levels in other metrics. Once objective levels are achieved, they are typically suitable to ongoing operations, and the protection posture is kept up to date over time.

Changes involving people take from a few months for large-scale awareness programs to 4 or more years for educational processes leading to degrees.

Mergers and acquisitions produce additional processes reflective of the enterprise-wide process and typically take from 6 to 18 months to integrate into the overall information protection program. They operate in an ongoing integration cycle starting when they become part of the enterprise. Breakups do not typically create large-scale governance changes, however, as expertise moves from place to place, lost functions must be replaced and this typically takes some time and effort. It is common for a stable protection program to re-stabilize within 6 months after a large-scale breakup.

Major changes in system security typically correspond to system life cycles if done cost effectively. For systems with long life cycles, these changes usually involve external protective devices. Minor system changes happen all the time with time frames limited only by the research, development, testing, and change control process. Data changes are almost universally made over short time frames depending only on the time required to make and validate the changes.

Startup of large-scale information protection governance and control programs is often problematic because of inadequate expertise or inadequate control over expertise in place. A program can be created from scratch and brought up to a reasonable level of performance in 2-3 years. This start-up challenge can sometimes be reduced by using external experts. Outsourcing of critical high-level functions is feasible for periods of time, particularly when building a program up, however; in the long run the level of expertise required for key positions makes it expensive to outsource and may make governance more difficult. Key areas where outsourcing works well are likely to include:

- The staff assistant can typically be outsourced but is easiest to find.
- An awareness and knowledge lead can be brought in part-time to start the program. Replace this person with an inside staff member in a year.
- Consultants can be used on select high-consequence reviews if desired, but as internal testing leads they are too expensive.
- A security audit lead is often brought in from the internal audit department. External expertise is usually used for protection posture assessments or to augment audit processes on a case-by-case basis.
- A policy, standards, and procedures lead, or an initial development group is probably better outsourced than internally created. Eventually an insider should take over the lead.
- As an interim step outsourced private investigation background check firms can be used for select augmentation purposes.

Whole programs can be outsourced in some cases. Even the CISO can be outsourced for some enterprises for some time. But this is not a long-term solution. It should only be used to bridge between CISOs or for program start-up when no internal CISO is available and a CISO and program must be built.

Summary

The CISO, in whatever form that position takes on, governs enterprise information protection by a combination of power and influence. This involves the creation and operation of institutions that cross all organizational boundaries and involve business functions, operations, and assurance processes at all levels. In each area of involvement, the CISO exerts control, gets feedback, and acts to continue operations within the desired control envelope. At the overall level, business functions push operational needs that push assurance processes, and the CISO must also control the process at this level in order to be effective.

When there is no CISO, governance suffers, and the utility of content to the enterprise suffers along with it. While many have tried committee approaches to management of the function, many have ignored substantial portions of the problem, and many have tried to keep a weak CISO under the control of some other business function, many of the spectacular failures can be directly linked to these decisions. Many CIOs have been fired because of the inability of the CISO who works for them to get the job done under the constraints put in place by the CIO's own actions. Many have ended up trying to cover up things discovered by their CISOs and the enterprises have suffered as a consequence.

Different people in different roles interact with content and the content control system in different ways. Pure business functions like the policy team, HR, Legal, and the risk management team interact with pure operations functions like the testing and change management team and incident response team, and with pure assurance functions like the audit team, education and awareness teams, and documentation specialists. But mixed teams also exist to bridge between these areas, like project teams, technology teams, and project management teams. The CISO crosses all of these boundaries in order to assure that the enterprise prospers and that it can effectively deal with the risks inherent in the efficiencies gained by modern information technology within the enterprise.

Governance questions

1. How does governance turn the determination of what to protect and how well into the reality of protection?
2. What is governance, in your own words, in one sentence?
3. In a networked organization, how does governance operate, and can it operate effectively?
4. For each of the organizational perspectives and functions, what would be the effect of not addressing it?
5. Given the reality that most CISOs work for CIOs, what are the implications for the near-term future of enterprises in terms of information protection?
6. For each of the groups a CISO meets with, assume that there is no such organization in place and describe how you might go about creating such a group by using power and influence granted to a typical CISO?
7. Given the wide range and high number of standards potentially involved in a protection program, what are the advantages of identifying a particular one to rely on, and what are the problems associated with relying on only one such standard?
8. What is the difference between an enterprise control standard and a standard?
9. What is the difference between a policy and a control standard?
10. What is the difference between a procedure and a control standard?
11. How does GAISP relate to ISO 27002?
12. Write a sample appeals process for information protection decisions made at an enterprise and describe the elements that are likely to to involve HR, Legal, management, the CEO, the Board, and other stakeholders.

13. Given the nature of an enterprise control system for information protection, why couldn't the system be operated by the CFO or CIO instead of having a separate individual in charge of it?

14. What is the difference between program metrics and technical performance metrics, which would work better for an enterprise protection program, and why?

15. Create a sample budget for an enterprise protection program, including all hidden costs you can think of, and try to justify the actual budget of 5% of gross sales for a manufacturing business with your list of costs as the basis. Do you get 5%, more, or less, and why is this?

16. Given your answer from question 15, suppose you only get ½ of the funding identified in the budget above, and describe which things you would cut out of the budget and the effect that would have on information protection in the enterprise.

17. Suppose you are a CISO working for a CIO and the CIO tells you to report security incidents and issues only to them, while the CEO told you in a private meeting that you should report important incidents and issues to them. What should you do and how should you go about doing it?

18. Suppose you are a CISO working for a CIO and the CIO is accidentally detected using enterprise computers to do stock trades during the hours they should be working, against enterprise policy and control requirements. How should you handle it?

19. Suppose you are a CISO who works for a CIO and the CIO is worried that the problems you find will cost them their job. Suppose the CIO tries to limit your communication, and then determines to fire you while they look for another job. How would you respond to this situation and based on your responses, what would be the likely outcomes? Are you still certain you want to be a CISO?[6.10]

20. Download the “Influence” software from http://all.net/ and try it out. Does it help gain clarity around power and influence?

7 Control architecture

The control architecture creates the overarching objectives and structural approaches to protection without drilling down into the details of how those objectives are met or those approaches are implemented. It is a theoretical structure that ultimately gets implemented by the technical security architecture.

From the beginning of the computer security field, there have been models of information protection developed for different purposes, many of them with mathematical analysis as underpinnings. As more and more attacks succeeded against systems of all sorts, the mathematics of protection became an increasingly critical element in that the use of mathematics allows some sorts of proofs that demonstrate properties of components and composites to a defined level of certainty. This ultimately led to the development of theoretical models of system security concepts.

The reader is strongly urged to read endnote 7.1 here[7.1].

Although mathematics is quite limited in what it can realistically cover because of the inability to model and analyze reality to the level of granularity necessary for perfection, it is very helpful in finding obvious flaws. Because mathematics ultimately depends on modeling, the question of how we view protection becomes limiting to the things we can do and the way we can think about the issues. These models have produced, over time, a set of structural assumptions and characterizations that underlie most of the thinking and modeling of information protection as we know it today.

As an architectural concept, the elements of control architecture are as vital to information protection as the concepts of floor, ceiling, and walls are to building architecture. They stand as assumptions that, even when we violate them, we should remain aware of. They have been generalized to some extent in the ongoing development of the field and in its search for approaches to meet the ever changing and harrowing nature of the information environment now faced. But they are increasingly vital and being pushed beyond their limits in the wide open and harsh environments in widespread use today.

Protection objectives

While the overall objective of information protection is to assure the utility of content, at a more detailed level, specific objectives have to be identified to help achieve this overall objective. The utility of content is often lost or decreased as specific objectives fail to be accomplished. They are typically codified as shown in Figure 7-1 as integrity, availability, confidentiality, use control, and accountability. These objectives exist in the context of interdependencies, they are often associated with technologies that are used to assist in achieving them, and there are different facets to each of them.

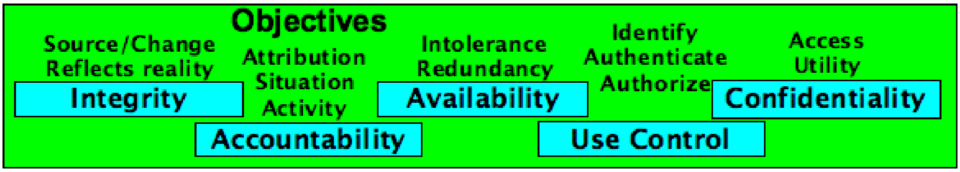


Figure 7-1 - Protection objectives at a high level

Integrity

In most cases, the integrity of content is most important to its utility because, even if it is available and kept confidential, properly audited, and under use control, if it is wrong, its utility is poor. If it is wrong in specific ways, it can be very harmful. Integrity is often broken down into the integrity of the source, protection from inappropriate or unauthorized changes in the content, and assurance that the content represents an accurate reflection of reality suitable for the purpose. Source integrity expresses the association of reliability of content with its source and is an example of an approach to assuring the correspondence of the content with reality. Many cryptographic technologies are associated with integrity in the sense of freedom from unauthorized change and attribution to source, however, cryptography has serious limitations in integrity protection.[5.6] Change control is a vital component of an effective integrity control scheme because it provides redundancy-based controls over changes to verify that they are reasonable, appropriate to the need, and that they operate correctly in the environment before the changes are deployed. Changes also have potentially recursive, complex, and indirect effects that lead to unintended consequences. For example, computer viruses use changes in software to cause transitive spread of the virus from program to program.[5.5] This is an unintended but predictable consequence of combining general purpose function with transitive information flow and sharing. Integrity technologies include, without limitation:

Redundancy allows faults to be detected and, sometimes, corrected. [7.1.10.2-3]

Validation provides the means to increase assurance through independent (redundant) confirmations or refutations of form.

Consistency checks use redundancy to validate data.

Verification provides the means to increase assurance through independent (redundant) confirmations or refutations of content.

Multi-source verification provides specific a sort of independent confirmation or refutation.

Multi-factor approaches use independent sorts of measurements or factors to independently verify content or source.

Trust models are sometimes created and applied to provide metrics on trust. [7.1.19]

Submit/commit cycles provide independent confirmation over an independent channel. [7.1.17]

Watermarking is used to provide a self-validation of the media on which content is sent.[7.1.11]

Cryptographic checksums provide redundancy that allows validation of use of specific keys or confirmation of content against published coded values. These are typically many-to-one functions that are harder to forge than the material they cover.[5.6]

Integrity shells are real-time just before use verifications of content, typically based on cryptographic checksums. [7.1.11]

Digital signatures allow validation of use of private keys. [7.1.12-13]

Certificates provide validation of the authority to sign based on an authoritative third party source.[7.1.12-13]

Trusted Computer System Evaluation Criteria (TCSEC) systems use high surety access controls to assure flow control and limit corruption. [7.1.8]

Trusted Computing Group (TCG) systems use integrity shells, cryptographic checksums and similar methods to assure the integrity of process lineage and content.[7.1.11]

Availability

If information is not available in a timely fashion, its utility decreases, but may not completely disappear. Availability is typically measured in terms of mathematical formulas for availability and reliability of the function when needed. Availability is typically measured as percentage of down time per unit time. For example, hours of system outage per year is used for some systems. Sometimes it is normalized for utility in the enterprise, such as the use of user outage hours per month. It can also be calculated based on mean time to failure (MTTF) and mean time to repair (MTTR) as MTTR/(MTTF+MTTR). Assuming that everything is properly accounted for, these are measurements after the fact, but not as useful for prediction, which is critical for design. [7.2]

Interdependency analysis is used to determine availability of systems based on availability of other systems they depend on. It is central to the analysis of availability, even though it is often ignored.

Redundancy is used to increase availability by making independent resources available in case of failure. Generally, redundancy increases availability but reduces reliability. That is, there will be more failures, but the percentage of time with unmasked failures will be lower. Redundancy must be carefully implemented to avoid brittleness and common mode failures.

Higher quality components are also effective at increasing availability. The approach of using higher quality implies a trade off between the cost of quality and the cost of quantity associated with redundancy. This is also called “fault intolerance”.

Confidentiality

If confidentiality is lost, some content may become useless or even dangerous, but this is rare. In most cases the consequences are limited to potential liability. When classified information, trade secrets, or similar content is involved, consequences are higher.

Confidentiality is usually controlled based on the clearance of the identity, certainty of the authentication of that identity, classification of the content, and need for the authorized purpose. The means of creating and operating this basis is often more easily attacked than the real-time protection in an operating system or application.

Information flow controls are the only really effective way to limit the movement of information from place to place. All other techniques are leaky in one way or another and most can be defeated to great effect by any reasonably astute attacker. These controls are implemented at routers through network separation technologies (e.g., VLANs with quality of service controls to eliminate covert channels), in computer systems through access controls, in physical technologies by separation of systems and networks by distance and with shielding, and in applications through application-level access control. [7.1.2-8]

TCSEC systems are systems implemented under the trusted computer system evaluation criteria and are designed and rated relative to their ability to correctly implement flow controls. High TCSEC ratings imply a high degree of certainty that flows from more sensitive to less sensitive areas only pass through covert channels. If used in limited applications, like as network control devices, they can be highly effective at allowing only specific sorts of controlled flows. However, covert channels are found in all such systems and for general purpose use they are subject to virus attack where viruses then carry covert channel exploitation code. TCSEC systems are often given as examples of how confidentiality control depends on integrity control for its effectiveness. [7.1.8]

Cryptography is often used as a separation mechanism to prevent those who gain access to data from meaningfully using the content it represents. Cryptographic systems are hard to get right, typically have many covert channels, key distribution issues, recovery issues, performance issues, and are hard to manage on a large scale. However; there are significant products in the market that greatly ease this burden at a substantial financial cost.

Abyss processors and similar containment devices are special purpose physically hardened devices that are used for high surety processing. They typically use physical security barriers in small devices like smart cards and other similar platforms to provide special purpose functions, like cryptographic key management and commit components of submit/commit cycles. They are costly to develop but effective in providing secrecy for small amounts of key leveraged to secure larger volumes of content. FIPS certification demonstrates high quality in these systems and critical for medium or high consequence situations. [7.3]

TCG systems use standardized specifications to implement abyss processors within normal computing platforms. They are predominantly used for authentication and encryption and address many of the recovery issues at relatively low cost (on the order of a few dollars) per computer.[7.1.11]

Digital diodes are separation mechanisms that allow one-directional information flow. They are designed so that less secret or higher integrity information can be passed to more secret or lower integrity areas without the potential for the more secret or lower integrity information passing back to or affecting the less secure or higher integrity area. Thus they provide the basic mechanisms for one-directional information flow. They have no covert channels in high surety implementations and known and identifiable covert channels in lower surety implementations.[7.1.10]

Use control

If use control is lost, either content is not usable by those who are supposed to be able to use it, which corresponds to a loss of availability, or content is usable by those would should not be able to use it. This can lead to loss of integrity, availability, or confidentiality, depending on the specifics of the uses permitted. Use control generally associates authentication requirements with identified parties for authorized uses. The basic notion underlying use control is that identified individuals or systems acting on their behalf are granted appropriate use based on their identity and the demonstrated extent of authenticity of that identity. If the current level of authentication is inadequate to the need, additional authentication is required to meet the level required for the use. Use may be more permanently disabled via fail safe if warranted, for example by disabling system use for a period of time.

Biometrics are used to provide authentication based on physical characteristics typically associated with individuals out of a group.

Something you have like a smart card, secure identification card, universal serial bus (USB) authentication device, proximity card, radio frequency identification (RFID) tag, or other device is used for authentication.

Passwords, pass phrases, and other similar mechanisms based on user knowledge, skills, and capabilities are used to indicate something the user knows or can do as authentication.

Separation of duties is fundamental to the administration of use and acts as a control over potential abuses. Separation of duties is typically operated without consideration for time, however, time transitivity of use is critical to proper separation. As a simple example, the requirement to separate purchasing from payments is based on preventing a single individual from placing an order and paying for that order. This can be exploited to pay the person or their relative based only on their own stipulation. Without time transitivity controls, a person who works in purchasing today can get a job in accounts payable at a later date, perhaps at a different facility and under a different name, and carry out the rest of the fraud. To counter such methods, which have been perpetrated in the past, life cycle tracking of individuals and uses associated with those individuals is necessary.

Process controls limit how processes can proceed. An excellent example of a failed approach to process control is the placement of purchasing and payables in the same computer system. Even if separation of duties is applied to the people who work in AP and Purchasing, the systems administrator of the system can typically violate the process controls to grant themselves access to both capabilities. They might directly place purchase orders and payments into the database, thus avoiding the programmed controls that exist to prevent normal users from doing the same thing through normal program interfaces. This is why change control is required for such systems and why separation and process controls must go beyond the boundaries of any individual person or computer system.

Submit/commit systems are use control devices that separate the preparation of a transaction from its approval process. If properly implemented, a device is used for taking the submitted information and committing the transaction, and that device is physically separated, unforgeable, uncircumventable, and independently controlled.

Roles and rules are often used to associate individuals with the roles they play in order to perform their job functions. Rules are applied both to the allocation of people to roles and the actions permitted by people in those roles. This abstraction layer permits organizations to create processes independent of individuals, allows easy changes of people associated with roles, and reduces administrative effort associated with maintaining individual access by replacing it with a two-step process of (1) maintaining roles and rules and (2) associating people with roles. This abstraction reduces management complexity but care must be taken to assure that it does not prevent proper accountability and that time transitivity of role assignment is done at finer granularity. Roles and rules also tend to aggregate risks unless properly controlled.[7.1.14]

Identity management (IdM) infrastructure provides a means by which role, rules, identification, authentication, and authorization processes can be joined together in an administrative mechanism and functional infrastructure elements. By doing this, IdM also tends to aggregate risks so that the IdM infrastructure rapidly rises to a higher risk category as it gains efficiency by centralizing use control and audit functions. [7.1.14]

Accountability

Loss of accountability reduces the certainty with which proper operation can be verified either now or in the future. Accountability is often considered in terms of attribution of actions to actors, the accurate identification and recording of the situation, and the association of the activity with the actor in the situation.

Attribution of actions to actors is particularly problematic, however, we generally use user identity information associated with authentication processes to assert attribution (to a level of certainty associated with the authentication process) of actions associated with the identity (to a level of certainty associated with the surety of the systems and infrastructures involved) to the individual associated with the identity (to a level of surety associated with the initial registration process, background checks, and surety of the systems that maintain and promulgate the identity information). Of course this begs the issue of how certain we are of each of these elements and leads to a level of uncertainty associated with any accountability process, particularly such a process that might be subject to internal malicious attack.

Audit trails are the reflections of the attribution of actions to actors in tangible form. These records are generated by various systems and functional elements of applications and sent to storage and recording media for retention, transfer, evaluation, and other use. The storage and recording media is subject to attack and may reside in the same system as the audits are generated, or elsewhere. It may be append and read-only, may pass through protective barriers, and may be well controlled – or it may not. Surety of the audit process limits surety of accountability. Audit trails tend to record specific information thought to be of value to the need, but they may not be useful for situations in which systems are not used as designed. For example, in a direct attack against the operating system leading to a modification of the underlying database files, the accountability mechanisms of the database engine are often circumvented, leaving no audit of changes.

Granularity issues drive to limits on accountability. A seemingly simple transaction, such as the movement of money from account to account might seem simple and readily accounted for. But if this transaction involves remote Internet access to a Web server that uses a back-end application server to modify data in a mainframe database that stores results in a storage area network, the number of auditable events could easily run into the tens or even hundreds. Every system involved and every aspect of their operation may result in audit records.

Analysis issues associated with audit trails include timing, correlation of the many audit trails associated with any transaction, identification and explanation of excess or missing audit items, accounting for failure modes and the results of all such modes when they occur, reconstruction of events from partial audit records, and protection of confidentiality associated with the data the audit trails reflect. The analysis process must not be capable of corrupting or deleting the audit records or many other problems occur. Audit trails may be correlated for data aggregation and this too must be understood in context to determine what separation of duties and use controls over audit information is required.

Preservation issues are driven by a combination of organizational and legal requirements. Legal requirements call for retention of accountability information of many sorts, most particularly as it is associated with business records. Retention periods vary, but 4-7 years is a minimum period to expect, and the ability to identify and retain specific records associated with incidents or legal matters for longer periods is often required. Some accounting records must be destroyed in a timely fashion in order to follow EU regulations and other privacy requirements, so a method for separating different audit records based on these requirements becomes a critical part of the accountability process.

Access control architecture

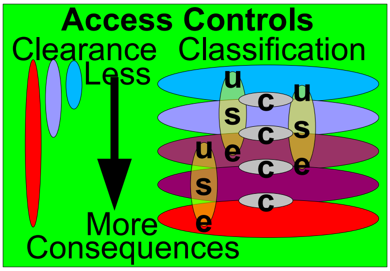


Figure 7-2 Access Controls

Access control is used to implement the basic separations that assure integrity, availability, confidentiality, use control, and accountability. Based on individuals with clearances and classifications of content and systems in terms of consequence, use is limited. Use control dictates that use should only be granted to content and systems based on need, thus the principle of least privilege applies and use-based compartments extend between different consequence levels. Compartments may also be made based on the desire to limit risk aggregation. Controls (c) are placed between levels of consequence to assure to a desired degree of certainty that consequence levels do not interact except in well defined and properly controlled ways. These controls are implemented through combinations of physical and informational functional units. Figure 7-2 presents the typical structure of a Bell-LaPadula or similar control system. [7.1.2-7.1.8]

Components and composites: functional units

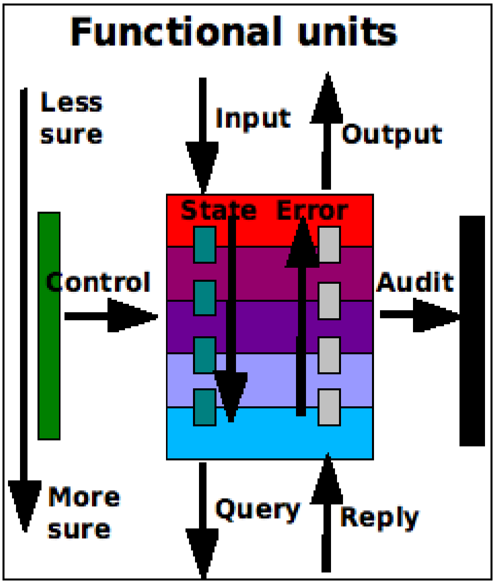


Figure 7-3 Functional Units

The technical architecture is made up of composites formed from functional units that are usually themselves composites. The functional units are made up of layered sets of protective barriers and functional applications that take input and initial state and produce output and next state. Units are controlled through a control plane, audited to an audit plane, and may send queries and get replies associated with external data sources. Surety increases as more layers are passed, because and only if layers are independent in that they can have failures and yet other layers continue to operate properly. They should pass state and error information back and forth and the audit process can detect behaviors that are out of normal functional unit behavior. Typical layers include firewalls to eliminate clearly invalid inputs and sources, decryption and authentication, data input validation, state machine modeling for proper context, identity-based access controls, validation of queries and replies against known valid classes, redundant sourcing of data based on requirements, back-end process selection for the query and reply process, encryption and authentication for back end processing, and verification of results from queries. Figure 7-3 presents a pictorial of a functional unit from a conceptual standpoint. Unfortunately, This is greatly impacted by composition issues and surety can really only be evaluated in light of these issues. [7.1.20]

Perimeter architectures

Figure 7-4 Perimeters

Perimeters are implemented in both physical and logical senses, with logical perimeters often placed at physical perimeters for the added surety associated with their co-location. For example, a cryptographic mechanism might be placed at the physical barrier between an enclave and the outside world so that from inside the enclave only plaintext is visible and from outside the enclave only ciphertext is visible. By locating the mechanism at the physical barrier there is no chance that a cross connection between the two sides will occur because the physical protection prohibits it. Similarly, the encryption makes the physical barrier more effective at separation because there is a reduction in the logical mechanisms that can be applied to bypass the encryption mechanism. This is shown at a simplistic level in Figure 7-4. Perimeters are often judged by the set of barriers present against illegitimate passage, the quality of implementation of those barriers, and the ease of passage for legitimate purposes.

Physical perimeter architecture

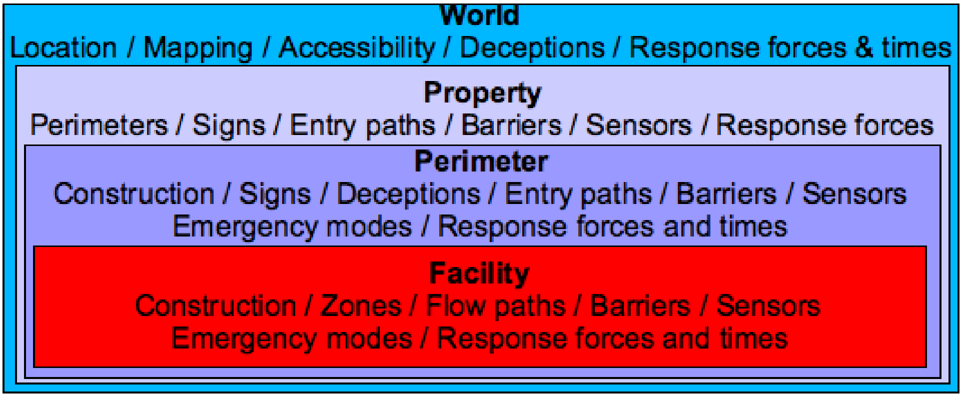

World
Location / Mapping / Accessibility / Deceptions / Response forces & times
Property
Perimeters / Signs / Entry paths / Barriers / Sensors / Response forces
Perimeter
Construction / Signs / Deceptions / Entry paths / Barriers / Sensors
Emergency modes / Response forces and times
Facility
Construction / Zones / Flow paths / Barriers / Sensors
Emergency modes / Response forces and times

Figure 7-5 Physical perimeters

Physical controls are integrated into informational controls. Figure 7-5 presents a general notion of how this might work. For deterrence there are signs, terrain, location, and deceptions. For prevention, perimeters use a wide range of barricades including but not limited to steps, fences, cement separators, moats, mounds, walls, and even mine fields. Detection involves a wide range of sensor technologies including visual, infrared, ultrasonic, sonic, chemical, pressure, motion, and even animal mechanisms. Reaction involves the movement of forces or use of fires of various sorts. Adaptation is undertaken by structural redesigns, movement of facilities, increased or enhanced perimeters, and so forth.[5.3.1]

World

Different technologies are typically placed at different places. For example, concealment of location by not advertising it, putting signs on doors, or putting an address in the corporate directory are effective at limiting the number of people who know where a facility is, but this really only works for those who do not have legitimate access. Some locations are in remote areas making them inaccessible for most people who don't have a good reason to be there, and this forms an extensive distance barrier to approach without detection. Modern mapping capabilities provide global positioning system-based maps and overhead satellite photography so that preventing the mapping of an area is far harder than it was many years ago when simply not putting streets within a land area on the map would prevent it from being mapped by hostiles. Deceptions ranging from false locations in directories, to addresses that don't seem to exist, to concealment of a facility within another business, have all been successful at limiting the knowledge of attackers of a target. Response forces and times associated with their responses are also key to the analysis of location. For example, being located near emergency services provides increased security through decreased response times.

Property

Property location and characteristics such as grades, soil makeup, weather, and surrounding topology are important factors in the protective function played by the property on which a facility is placed. Properties in flood zones, at the end of airport runways, on known fault lines, next to active volcanoes, in tsunami areas, below large bodies of water, near hazardous chemical plants or explosives factories, and in other paths of natural or unnatural disasters are subject to the outrageous fortunes associated with those locations and require additional protective measures in order to achieve the same level of protection that would commonly be afforded by a different location. Perimeters surrounding properties and property lines with natural barriers, and barriers within properties such as rivers, lakes, arroyos, cliffs, and similar natural and unnatural barriers, are important to characterizing attack paths into and out of properties. Accessibility from the air, ground, water, and underground are all important to understanding attack paths as well.

Perimeter

Perimeters surrounding and within properties provide distance and distance, has advantages. Distance implies time in physical movement while also reducing electromagnetic, sonic, and other emanation levels. It increases power levels required for exfiltration of data, makes running wires take longer and cost more, makes it more obvious when someone tries to go from one side of the perimeter to the other, and makes it harder to tunnel under or fly over without being detected.

Barriers are typified by moats and walls. They provide added reduction in emanations of various sorts, perhaps blocking visual inspection from easy to enter proximate locations. They prevent penetration by different sorts of mechanisms, ranging from a simple fence that prevents walk-ins, to a barrier capable of deflecting a high explosive blast. They also provide cover for attackers who may be able to hide behind or between barriers to defeat detection.

For the vast majority of cases, barriers have to be permeable to be useful because some amount of legitimate use has to pass into and out of the protected area. Entry paths are provided to allow barriers to be bypassed in controlled ways and under proper identification and authentication processes that grant authorization to pass. Mantraps and similar technologies may be employed to trap individuals who try to pass a barrier without authorization to do so, but there are liability issues and potential criminal liabilities associated with this sort of restraint in some situations. For volume entry and exit facilities, entry paths have to be fairly direct, proximate to parking or entrances, and able to handle the volumes required. Construction of barriers and emergency modes for bypassing barriers are critical to understanding behaviors under unusual circumstances as opposed to normal operational modes.

Signs are commonly required to provide legal notice as to trespass, proper entry, authorized access and use, and safety and health hazards associated with the property. Sensors around and within properties can be very helpful in allowing smaller numbers of people to more rapidly detect and triage attempted entries and passage. A wide range of sensor technologies are available, ranging from unified heat, sound, light, motion, shape, humidity, temperature, and dew point sensor arrays to simple trip wires and touch sensitive devices that sound alarms. Response forces are required in order for these methods to be effective with the time required for response at different force levels. They act as a critical factor in the effectiveness against specific threats.

Facility

Facilities have topologies that dictate how things and people go from place to place. Internal barriers, sensors, zones, and similar protective mechanisms are analogous to those on properties, but typically with better controls. For example, buildings often have sound, temperature, and humidity controls, motor generators, doors of different quality with locks of different quality and hinges on one side or the other. Construction materials and processes dictate the classes of threat capable of bypassing barriers such as walls and doors as a function of time with or without detection. Passage under floors, over ceilings, through air ducts, by picking or tricking locks, electrically or mechanically fooling sensors or tripping opening mechanisms, and removing or cutting hinges from doors, all grant human access. Tailgating, introduction of noxious gases to invoke emergency modes, fires, floods, and any number of other reflexive control attacks may be induced or occur by accident. Response forces and times also limit the time for an attack.

Logical perimeter architecture

Logical perimeters act in much the same way as physical perimeters, providing a series of barriers that slow or stop attackers. They include transforms and separation mechanisms at the outer perimeters, access controls, transforms, enclaves, and filters at facilities perimeters, and a range of other technologies closer into the higher valued content. Figure 7-6 shows the logical barrier architecture commonly considered for enterprises.

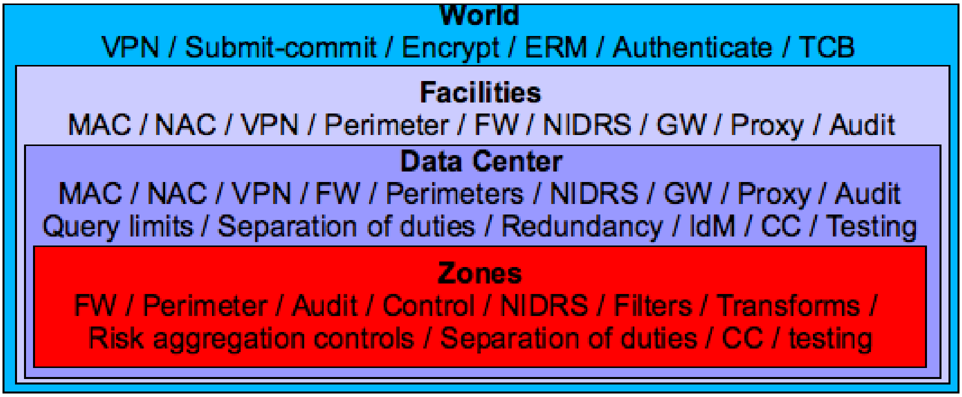

World
VPN / Submit-commit / Encrypt / ERM / Authenticate / TCB
Facilities
MAC / NAC / VPN / Perimeter / FW / NIDRS / GW / Proxy / Audit
Data Center
MAC / NAC / VPN / FW / Perimeters / NIDRS / GW / Proxy / Audit
Query limits / Separation of duties / Redundancy / IdM / CC / Testing
Zones
FW / Perimeter / Audit / Control / NIDRS / Filters / Transforms /
Risk aggregation controls / Separation of duties / CC / testing

Figure 7-6 – A typical logical barrier architecture

World

From the outside world, perimeter mechanisms are generally oriented toward things that permit the perimeters to be permeated with relative safety. Virtual private networks (VPNs) are used to provide encrypted tunnels between areas while authentication technologies are designed to allow the identity to be authenticated to the degree appropriate for the use. Submit-commit mechanisms provide physically secured devices to the user (to the desired level of surety) so that any mechanism desired can be used to submit a request but an adequately secured method is used to commit to that use. Enterprise rights management is used to pack protective mechanisms with content for low surety levels and can be used at a distance. Trusted computing bases (TCBs) can be used to provide higher assurance at remote locations.

Facility

Facility-level protection typically includes mandatory access controls at the network level, low-level communications card or processor identification and authentication mechanisms for devices attaching to internal networks and systems, VPN termination or internal VPN capabilities, physically secured logical network separation perimeters such as virtual local area networks (VLANs), firewalls, network intrusion and anomaly detection and response systems, gateway systems, proxy servers, and audit mechanisms.

Data center

Data centers typically have additional protections both at the physical level in terms of internal areas within facilities, and at the network and logical level in terms of similar protections to those for the facility but with tighter settings and more restrictions. Additional protective measures include query limits that limit the syntax and semantics of database queries, separation of duties protections to assure that risk aggregation is limited from a logical perspective within the data centers, redundancy for increased assurance levels against denial of services and single points of failure, identity management systems and interfaces to increase the surety of and specificity of access control decisions, change control mechanisms to increase the surety of software and configurations for systems with higher valued content, for utilities, or for aggregations of lower valued content that form medium or high risks, and more extensive testing processes.

Zones

Zones are used to further (1) separate portions of networks at a logical level both from a standpoint of classification and need to know as implied by the access control architecture, and from a standpoint of disaggregation of risks, (2) separate control from data, and (3) meet other protective requirements or constraints associated with functional unit design and risk management mandates. Zones are implemented with firewalls and other perimeter mechanisms, audit mechanisms, control mechanisms, and separation of audit from control from content. Network anomaly and intrusion detection and response systems may be used along with filtering technologies such as virus detection and transform technologies such as those identified for content control, to augment zoning solutions in some areas. Separation of duties tends to be implemented so that different individuals have responsibilities in different zones, and this helps with risk aggregation controls as well. Change control and testing processes are also varied depending on the specific needs of the zones as defined.

Perimeter summary

While perimeter technologies vary widely they have some commonalities that define their utility at an architectural level. While a zone may have substantial size, perimeter mechanisms tend to operate at a boundary and not within that boundary. As a result, perimeter architecture is oriented toward the fundamental notion of what will pass the barrier in what direction at what rate and how long the barrier will withstand what sorts of forces. Put in other terms, the barriers act to either sever attack graphs or increase the time to traverse links of the attack graph depending on the capabilities being used in order to defeat it. At the same time, it is desirable that perimeters provide as little friction to normal operation as possible, and for high volume perimeters such as airport entrances or network perimeters, their design should facilitate low delay times under high load.

Access process and trust architecture

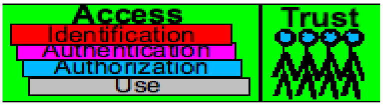


Figure 7-7 Access and Trust

The utility of the overall information capability of the enterprise depends on the ability to legitimately access information resources with minimal friction while still assuring the continuing value of the information in light of the hostilities of the environment in which it works. The access process architecture defines how identified subjects demonstrate their identities through authentication, and how the properly authenticated identified subjects can then use the content through an authorization mechanism. Access implies trust, which is transitive and risky. This is shown conceptually in Figure 7-7 [7.1.15, 7.5]

Identification

Identity of people and things, including programs and processes, is in itself a purely informational item. It is a, hopefully unique, tag that allows an individual to be associated with other properties. An identification system is a system used to track identities and associate them with these other properties. Initialization of identification processes is fundamental to their success as this is the process by which those things are associated with their identifying set of properties. For low surety situations, anybody will do, but clearance processes with background checks and detailed life reviews are invoked for situations in which people have to be identified with higher surety upon entry to a system of identification, while pedigree for hardware and software is commonly considered in determining its suitability for trust in high risk situations. Once an identity and some of its properties have been established, the identification system can provide a wide range of additional utility.

Authentication

Authentication is a process by which an identity can be verified as authentic by a process of testing that identity against the properties known for it in the identification system. For example, the identification system may have a user name and password associated with a human individual, in which case the presentation of those authentication factors are usable for authentication of the identity. The surety of the authenticity of an identification depends on the available properties in the identification system and the ability to present and verify those factors as present or absent in the entity in question. For higher risk, higher surety is typically desired, and sequential authentications may be used to increase the certainty with which authenticity of an identity is trusted. Different properties have different surety levels and withstand different threats more or less successfully. For example, biometrics are measures of what something is based on their physical properties such as finger prints, hand shape, eye print, DNA patterns, footfall pattern, and so forth. Fingerprints, for example, don't normally change and are unique to the individual, which makes them good for authentication, however; they can also be easily forged by skilled attackers.

Authorization

Once a subject's identity has been authenticated to an adequate level for a decision process to be completed, that subject may be authorized to a certain use. The authorization process typically involves matching a requested use with the identity and surety of authentication to determine how that attempt at use should be treated. Many treatments are possible depending on circumstances and capabilities of the protective system in use. Some options include; permit the use, refuse the use with a reason given, require additional authentication by the user for that use, require additional authorization for that use through an approval process, refuse that use and act to eliminate further requests, redirect the request to a deception system, and audit the request and its resolution. The decision on what to do is typically driven by some sort of table that associates authorities with subjects. This may involve a system of roles and rules used to determine what functions are allowed to what roles, with the roles associated with the individual identities as a second step in the process. Such a system facilitates changes more efficiently by allowing roles to be changed for identities at a high rate and roles to change at a lower rate, allows checking of roles against separation of duties requirements and similar overarching needs, and provides a reduction in errors and delays associated with changing permissions for large groups or for individuals across many systems.

Use

Use is the business utility that access process is designed to facilitate. As a result, the process should be relatively transparent and automatic to the user relative to the utility associated with that use. So the amount of effort and surety required for doing something simple like checking the time of day should, in most circumstances, be minimal. Otherwise the effort required to perform the process exceeds the business value granted. In most cases, authentication allows use of a set of capabilities for a period of time so that a single authenticated identity is authorized for sets of activities which are performed without additional authentication at every step. For high valued transactions, like large financial transfers or setting off explosive devices, additional authentication is warranted and applied, however; the additional authentication associated with that high valued transaction may also be leveraged to allow uninhibited subsequent use for a period of time and to a set of functions. Use over time from locations, and within other contexts, may be highly limited, but in most cases large ranges of usage in excess of least privilege requirements are granted because of the complexity of limiting use at a fine granularity. In these cases the audit mechanisms associated with use are often used to provide additional checks on that use and to limit the effects of illicit use. In all cases, use should be audited if the value of the operation exceeds the threshold of risk requiring audit or if there are regulatory or other drivers that mandate auditing of use.

Change control architecture

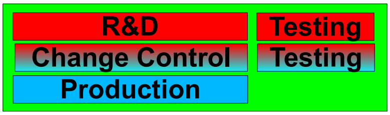


Figure 7-8 – Change control

Change control processes designed to assure that, with increasing surety for increasing consequences, changes to production systems are thoroughly tested, verified to meet need. contain only inappropriate elements, work properly on test data, can be reverted to previous states, and operate properly under emergency conditions. These verification and testing processes involve administrative and technical actions, usually involve a tracking process and ticketing system, require special expertise and technology, and are supposed to precisely reflect the production environment. In some disaster recover programs, the testing environment is kept at a separate site and used as the emergency recovery environment.

The historical approach to sound change control requires:[7.1.17]

- Changes are based on requirements specified and approved by people not involved in making those changes.
- Changes are provided to change control in meaningful form.
- The change control process does not alter the mechanism.
- The change control process examines the mechanism.
- The mechanism must be straight forward, easily understood, and clearly address the specific change requirement.
- There must be no unnecessary or unrelated changes.
- Change control must go through a well-defined testing process after the changed mechanism is analyzed.
- Independent verification is used to assure that the mechanism matches its specifications.
- After suitable management and technical approvals, only unaltered original mechanisms provided to change control are put into production.

Research and development

Research and development includes design and implementation activities and is typically carried out by people who have specific design and implementation goals. For change controlled environments, the changes made are limited to those associated with the desired change in function or behavior of the mechanisms being changed. Implementers are normally responsible for testing their changes against functional requirements and security requirements as well as against historical fault types. These are called regression tests.

Change control

Once research and development completes its changes, the change control process evaluates those changes to assure that they are (1) necessary to meet the defined change in functional requirements, (2) appropriate to the changes functional requirements (3) of no material affect on any other functional properties not identified as part of the change, (4) obvious and well understood in their operation, (5) consisting of only necessary and meaningful parts, and (6) consistent with proper operation of the changed system. The changes that pass these tests are then tested against operational requirements, security requirements, and with regression tests. If they fail these tests, they are returned to research and development and the failures identified with the individuals responsible. Failures of this sort should be used as a negative performance indicator on personnel reviews and managers who fail to react to such failures seriously should also be subject to negative performance reviews. Repeated failures of this sort should result in termination. Reversion copies should be kept and reversion tested.

Production

In production use, mechanisms do not change except in well defined ways, and their status should be regularly verified. Unauthorized changes should be investigated and causes eliminated. Failures in production should lead to reversion if they are high consequence.

Emerging control architecture elements

In addition to the well known control architectures elements, there are an emerging set of control architecture elements that are increasingly being applied in addition to or in place of the elements described above. The reader is warned that any or all of these may or may not stand the test of time or be subsumed in other control architecture elements or elsewhere.

Security policy languages and execution

Notionally, a policy language [7.1.19] that is specified by management and executed by mechanisms could enforce policy automatically without many of the other elements of the protection program in place. In practice this does not work. Rather, policy languages of today are typically implemented by technical specialists to cover technical controls through provisioning. In large enterprises, hundreds of workers might be part of the provisioning process and, while the language helps to automate many technical protection elements, it does not represent a substantial change in the way business is done. Rather, it implements rules associated with roles that ultimately lead to a subject/object system [7.1.15c]

The Web services world

In the Internet's use of the World Wide Web (Web), services are provided to users increasingly through the use of what has been called the “Web services bus”. This is a virtual communications layer that rides on top of the TCP/IP infrastructure used to move sequences of bytes between systems and within the hyper text transport protocol (http) suite. This “bus” provides services like encryption via secure socket layer (SSL) and other mechanisms, authentication processes via signed certificates generated and verifiable through trusted (though not necessarily trustworthy) certificate authorities, interfaces to identity management and other infrastructure to allow authorization processes, and other services added as desired or needed. In this environment, programmers and application designers assume trust levels associated with these services and apply the security services to try to generate more secure application environments that compose large numbers of different information sources for greater functionality.

Wrappers and related approaches

Wrappers started as a security response to incidents when the responders didn't have access to operating system source codes to be able to immediately repair security-related faults.[7.4] The technique was rapidly expanded into all manner of other protection mechanisms and efforts and wrappers or similar mechanisms have acted as proxy mechanisms for security ever since, especially in cases where access to underlying mechanisms are unavailable and time is of the essence. Unfortunately, this composition-based approach[7.1.20] has not been given proper attention in terms of assuring that desired security properties are well defined, and the mathematics of composition have not been well applied to wrapper technologies. While there is a potential for these approaches to yield a new control architecture model, unless and until adequate attention is paid to the underlying issues, it will not be well enough understood to use as a basis for structural decisions.

Cryptographic seals and self-protection

The use of cryptographic systems to “self-seal” content which is then accessed by authorized users, is part of the approach used in many copy protection “schemes”. They are commonly called schemes because that is considered descriptive of their lack of clarity in design and implementation. Most such systems are very weak and rapidly broken, making things a bit more difficult for most of the people most of the time, but not stopping the serious threats from defeating the copy protection approaches. At first blush this seems suitable as a tradeoff until one individual cracks the scheme and publishes a software program that automatically breaks the protection on any copy. At that point, the cracking routine is spread around the World using the Web, individuals unseal their copies and make them available over the Web, and the cat is out of the bag, so to speak. While there is some hope for a future in this arena, particularly through the use of the Trusted Platform Module (TPM) of the Trusted Computing Group (TCG) and other similar mechanisms to provide enforcement of the presence of seals in the use of material, this really only works when the user community is under very strict control or highly cooperative.[7.1.11]

Models of trust

Fig. 7-9 Trust

Many authors have worked on issues associated with modeling trust,[7.1.18-20] and there are many trust models in use today. From a control architecture perspective, models of trust are present in every case, embedded in process elements, the risk management process, the access control notions of clearances, and otherwise spread across the space. They have a place in control architecture, but that place is not yet well enough defined to become systematic in its use or to use it as a meaningful transition into technical security architecture.

Control architecture questions

1. What are the advantages and disadvantages of explicitly addressing control architecture in terms of understanding and managing information protection for an enterprise?
2. Given the myriad of protection objectives, how might these be prioritized and associated with different content, and how can protection be properly associated with content? Address at least three approaches; (1) an approach in which every element of content at finest granularity is associated with protection objectives and controlled individually, (2) a scheme in which content is treated in larger groups such as databases, and protection objectives are associated with those larger bodies of content; and (3) an approach in which content is differentiated at the enterprise level into only a few broad categories and each category is treated without further differentiation. Address the full spectrum of issues identified in previous chapters in your response.
3. Given that access control typically depends on the combination of a classification scheme and a clearance process, describe a completely different approach in which people are not cleared, content is not classified, and yet protection objectives are met. How does your new approach alter the rest of the control architecture and can you still implement the other aspects of control architecture given your approach?

4. Assuming that you have perfect functional units, provide a drawing of how an enterprise might use these units to implement a zoning strategy in which perimeters are used to implement the standard access control model.
5. For a perimeter architecture in which increasingly high consequence content is controlled behind an increasing number of increasing quality perimeters, how can control and audit be implemented without creating risk aggregation in the person of individuals who operate the routing infrastructure?
6. Since the requirement for access demands that there be some way to penetrate perimeters, all perimeters are inherently imperfect. How can the intentional imperfections of perimeters be mitigated using other aspects of control architecture?
7. Change control seems fundamental to any success in protection since, without control over changes in controls, they cannot be kept properly functional. Given that the vast majority of control mechanisms are implemented in commercial off the shelf (COTS) hardware and software purchased at the lowest available cost from vendors from all over the world, how can an enterprise use these low surety systems and mechanisms to implement higher surety in their protective environment?
8. Given the history of secrecy as a focus of information protection and the high requirements for integrity and availability in most critical applications today, what control architecture elements should be applied to increase the emphasis on these objectives other than simply setting objectives and their priorities?
9. Given that cryptographic systems are not high surety mechanisms in most cases, how can they be used to support high risk operations and what provisions must be made in the control architecture to compensate for their weaknesses?

8 Technical security architecture

From the front cover of the book, technical security architecture can generally be broken down into (1) “protective mechanisms” that touch content, users, and systems (2) other technical mechanisms that support the protective mechanisms, like work flows and inventory, called “protection process” mechanisms, and (3) portions of life cycles, data state, defense process, and context controls that are used to make decisions about and keep track of decisions. Most of the study of information protection from a technical standpoint tends to focus on the protective mechanisms, and these are well covered in standard texts and publications.[8.1]

Using the term architecture, there is an implication of an underlying model or structure that meets some need through its use. As used here, we are really discussing the elements of technical security, with the models used as a basis for piecing those components together, described under the heading of “control architecture” in the previous chapter. As such, the manner in which technical protective elements are fitted together is largely ignored here, and the mechanisms themselves are examined at a rudimentary level. The astute reader will find that this lacks in what might be called sound advice as to how to apply the mechanisms to the models to generate decisions. This goes under the heading of teaching the reader to fish rather than handing them fish. Before making decisions about how to apply models to enterprises and then selecting technologies, it is necessary to first understand the available technologies. That is the purpose of this exposition. In application, mechanisms are selected and then implementations are chosen to meet the specific criteria of the situation. As in any design, there are loops in which a technology selection fails to meet the need and another view of technical alternatives must be undertaken to seek a different solution. Think of this chapter as the tool chest used to start the next level selection process.

Protection process

“Protection is something you do, not something you buy” [2.6.1] The things that you do and that directly affect content are the focus here.

Issues of context

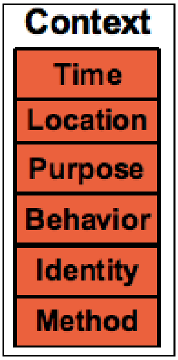


Figure 8-1 Context

Context is important to almost all decisions in information protection. As Figure 8-1 shows, context generally includes the basic questions of who, what, where, how, when, and why as they relate to content, systems, infrastructures, people, and functions. These are the same questions that are often asked in an investigation or a mystery novel, except that the answer to the questions and the methods available for attaining those answers tend to be very different when high surety is desired or when computers are the context of the process.

Time (when)

Time is important in tracking behavior, associating events across infrastructure, and making determinations about what is authorized and over what period something happened. It is also a central issue in all processes, particularly in attack and defense processes, where time is of the essence and OODA loop timing issues[8.5] may determine outcomes.

Time zone: The time zone associated with the action under consideration is the common time people think of and deal with on a day-to-day basis.

Time: The time within a context, or more commonly, the universal coordinated time (UTC) associated with the item of interest. UTC is the time typically used internally in system clocks and many applications and audit systems. It is useful for getting a common context to compare systems in diverse locations. In outer space and under-sea systems, time may be kept in some other frame of reference, but UTC is most commonly used.

Context: In some situations, time is relative to context, and this must be expressed in those situations where there might be a difference. For example, mission-oriented systems tend to keep time in the relative context of the mission.

Accuracy: Time bases have different errors type and magnitudes. For example, average error, skew, and skew rate are common issues in times based on line frequency, which is commonly used in electric clocks.

Differential: Differential time is a common issue in synchronization and differential limits are critical to many timed operations.

Location (where)

Location historically dominated access control, but mobility has made location harder to determine, the mobile and remote work force has caused location to become harder to limit, and location-independent approaches to computing are increasingly deployed. The utility of location has changed but not disappeared.

Network location determines large-scale controls to a large extent. Zoning policies are generally effective over locations, with different locations in the topology granted different sorts of access. While some networks in the general sense allow broad locational deviation, others are still located within physical enclosures or in limited areas.

Address, whether related to a map location, an Internet Protocol, or another similar sort of address, is a common method for differentiating systems and uses. Within banks, for example, teller functions can only be undertaken from select access points and this can still often be determined based on addresses.

Lines associated with telephone systems, terminal connectors, and direct or switched communications systems, are very widely used to indicate location, and this location is then used to determine controls.

Numbers, often special phone numbers, are used for maintenance access. They are typically restricted to select remote telephone numbers.

Global Positioning System (GPS) locations are used in increasing numbers of systems to provide location information that can be correlated with other factors to provide information ranging from routing to assistance calls. GPS has been used to limit access and to provide location-based authentication. Location can also be correlated with time for travel rates and to associate physical and logical access.

Physical locations are associated with devices and protective barriers and are often used as a basis for allowing or denying access. Known physical locations may have known protective conditions that allow extraordinary access based on facilities protection, personnel characteristics, and so forth. Local access to consoles is commonly used to grant maintenance access.

Logical location codifies conditions associated with a device or operating environment used to associate a level of trust. Typically, proxy servers or similar mechanisms provide a local presence that is used to gain access from another location.

Delta expresses the accounting for location changes that is sometimes used to determine physical possibility and other related conditions. For example, a credit card presented on the West Coast of the United States and then presented again on the East Coast of Africa an hour later cannot be the same card under current transportation systems – someone is lying.

Purpose (why)

Controls based on reasons are fairly rare in the sense that the controls are rarely tied directly to purpose. However; this is implemented indirectly in many systems by associating a purported purpose to specific usage patterns. As a rule of thumb, and as a prudent practice, unless there is an affirmative reason to grant access, access should be denied.

Authority is usually allowed as a basis for authorization. The purpose is to fulfill the mandates of positional power.

Context leads to use. For example, access to a database with financial records is granted to processes within applications acting for users who normally would not have database access, in order to provide them with prices for goods being sold.

Applicability of an action to a purpose is the basis for allowing use, while risk associated with access is a reason for denying use. Since there is always a level of risk involved in any action, a level of applicability required for access must be defined in order to grant access.

Basis expresses the underlying rationale that justifies use. It is typically expressed in terms of a rationale that makes sense to the owner of the thing being used. Typically the basis dictates the decision process over use, however; since this is not readily codifiable in computer terms, human judgments over classes of uses and applications authorized for those uses are substituted, at the cost of accuracy, but to the advantage of easier decision-making.

Rationale typically consists of a logical argument of some sort that balances risks against utility.

Explanation is used to provide additional details to the decision-maker. This is used after the fact to validate the decisions when independent reviews are undertaken and in periodic and situation-specific reviews.

Validity of explanations, rationale, and basis are subject to external inspection. For example, if the basis is a rationale using an explanation that doesn't make logical sense, or the rationale depends on a fact that is not accurate, the basis is not valid and use should not be granted.

Behaviors (what)

Behaviors are particularly useful and increasingly important to making protection-related decisions, whether it be the behaviors of individuals or the behaviors of systems. While the science of understanding behaviors is old, it is not precise and there is a lot left to do in this area.

Actions are tracked in behavioral modeling and analysis systems and the actions taken are used to inform protection decisions.

Sequences of actions are particularly informative because, in many cases, attacks are composed of sequences of actions that individually seem benign.

Situations dictate actions. The combination of system and world situation and behavioral sequences leads to the action that should take place. Without understanding the situation it is impossible to make determinations about the sensibility of many actions.

Inputs to systems include all things that can cause system effects. Inputs include behaviors that are not available to the machine even though they exist. For example, low-level signals are digitized, which covers up low-level details. Original inputs must be used in some cases to understand what is going on.

Outputs from systems can indicate problems. In many cases, outputs are directly detected as unacceptable. In other cases, different outputs are acceptable in different contexts.

State information is rarely available to analytical systems, however; the state of the machine dictates input processing and the resulting state changes and outputs.

Changes to states are the result of inputs interacting with previous state. Internal behaviors are almost always reflected in state changes. Typically, attacks on systems generate undesirable state changes that produce undesired side effects.

Identity (who)

Identity and the management of identity in the world and within systems is commonly considered a fundamental aspect of protection. A basic principle is to take identity, authenticate it to the desired level of surety for the need, and use the identity to authorize access and actions. This allows the protection architecture to work.

Names are typically associated with all of the identified items of interest, whether they be individuals or things. The design of name spaces is important, among other things, because many things may be identified by systems and name conflicts can cause incorrect system behaviors.

Types are usually associated with identity information. For example, there may be people, things, and subtypes associated with them.

Properties are typically associated with named identities. This includes linkage to roles and rules, properties associated with identity for controls of various sorts, information about locations, times, their capabilities to authenticate, biometric properties, and any number of other things.

Basis for identity implies that different reasons for associating identity may exist. The differentiation of basis has utility. For example, the assertion of an identity may be associated by way of a federation with a trading partner who provided the information, or it may be associated with a DNA examination tracked to parents. The former is clearly adequate for different uses than the latter.

Certainty is reflected in the application of the basis to validate identity and the need for certainty in the application of that identity. For users from the Internet accessing library records, no certainty is typically required or desired. They may even make up their own identities for that purpose. But for access to financial systems of an enterprise to perform large dollar value electronic funds transfers, far more certainty is desired. Multiple authentications, proper locations at times, proper basis, and other similar factors may be considered in the decision to grant authorization.

Method (how)

How things happen is often used as a means of control, and of course surety levels and similar facets of protection are inherently tied to the methods used to accomplish them.

Hardware tends to provide more certainty of function because it is less flexible and less subject to the sorts of design flaws commonly found in software. It is also less complex than software from a standpoint of the sizes of the state machines, it takes longer to create and modify, it can be more thoroughly tested, and it is more expensive to reproduce in low volume.

Software is the opposite of hardware in the respects described above. It is more flexible, more subject to design flaws of certain sorts, less sure, more complex, easier to create and modify, less thoroughly tested, and less expensive to reproduce than hardware.

Route controls are designed to use the path from place to place to increase the level of certainty that content is what it is purported to be. In practice this may be a network path, a physical path, or whatever part of these paths is identifiable by the recipient.

Means is generally associated with the way something was done and is used in legal parlance associated with patents, which are means and methods for accomplishing some task.

Transforms can seal information and be used to prove to those that can verify the seal or unseal the package that the creator had or had use of the transform and key.

Protocols are used to differentiate request types. Typically, any protocol can be "tunneled" through any data path, and steganographic encoding can be used to supply arbitrary content over any data path while meeting syntactic requirements. Protocols are, nonetheless, used to verify certain properties of communications.

Packet or line are often used to differentiate how content arrives or is sent and these are often controlled to limit paths.

Physicality is used in certain interfaces, such as the console interface to most systems, to differentiate actions that are allowed. Most systems have limitations on non-console access. For example, changing "Basic Input Output System" (BIOS) settings can sometimes only be done during the bootstrap process through the console interface.

Voice, data, and video paths are often differentiated so that certain functions can only be performed over certain types of interfaces or with certain types of content. For example, some systems use audio validation processes, or perform a challenge with audio information that can only be responded to with data, forcing the user to have proper capabilities and configurations on their systems in order to gain access.

Life cycles

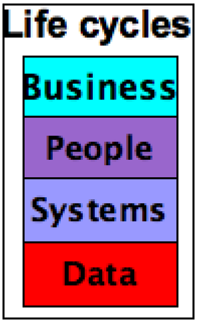


Figure 8-2
Lifecycles

As identified in Figure 8-2, life cycles are associated with businesses, systems, people, and data. These life cycles include a wide range of different processes over which protection must be considered in order to have an effective overall protection program. As a rule of thumb, very few circumstances call for all aspects of all lifecycles to be addressed; however, any of these aspects might have to be addressed for any given situation.

Business

Business life cycles have many interactions with information protection programs that are ignored in the literature to a large extent, even though their effects can be dramatic. Business changes often have significant impacts on employee behaviors and there are many cases in which these produce disgruntled employees, layoffs, firings, and organizational changes. These imply significant information protection issues beyond what is listed here.

Formation of businesses and the processes involved increasingly expose a lot of information to public view. For example, in order to form a corporation and get a bank account in the California today, you may have to provide a fingerprint to the bank and personal information to the state to allow them to track you down if they should want to. When businesses are formed, there are automatic processes that notify vendors, who pay the state for information about the formation, and use the data provided under force by the state to contact the owners to sell things to them. When new businesses are formed by existing businesses, there may be effects on credit and other similar interactions.

Funding processes involve a lot of detailed financial information, often including credit checks on individuals associated with the business and containing a wide array of private information. Funding processes are also used to feed data into large databases that are widely accessible for a fee or, in some cases, for free. The funding processes often involve information that can be readily used in identity theft, or in rare cases, business identity theft in which the identity of the business is stolen and used to perpetrate frauds. Funding profiles for businesses often ignore information protection issues and, as a result, protection is often lax in start-up processes to the detriment of the shareholders.

Operation of businesses include the sorts of information protection requirements described throughout the widely published literature.

Initial public offerings (IPOs) lead to the need to run companies as public rather than private entities and this has dramatic effects on the legal and regulatory requirements in terms of information protection. The basic issue with an IPO is that the value of their investment depends on the integrity, availability, confidentiality, accountability, and use control of the enterprise's information and infrastructure. As a result, the enterprise must meet due diligence requirements, be reasonable and prudent, and produce results that the CEO and CFO can attest to.

Joint ventures and similar business arrangements require special protective measures, particularly when companies compete in other markets. This is necessary in order to prevent (1) collusions or revelation of pricing information, which might violate restraint of trade requirements, (2) competitive information from being leaked, (3) corruption of one enterprise by the other through the joint venture, and (4) other similar negative consequences. However, the participants in the venture must still effectively work together and reach back into their respective infrastructures for day-to-day operations and provide content relevant to the joint venture.

Mergers and acquisitions lead to the combination of information technology components, capabilities, and systems, mixing of staff, and exchanges of content that are typically controlled by completely different information protection programs. There is a very significant cost associated with the transition of an entity into a new security operations process. Someone ultimately has to end up in charge, firewalls between entities have to be created so they can interoperate while the protection infrastructures are reconciled, information classifications have to be reconciled in order to gain proper controls, clearances and need-to-know designations have to be reconciled, interdependencies change, risk aggregations shift, and so forth. This is an effort comparable to the start up of a new protection program in one of the entities and major infrastructure changes in the other. These changes tend to produce disgruntled and laid off employees and this must also be considered.

Divestiture typically involves the splitting of content and systems between the two resulting entities. There are many implications for information protection. For example, for every role in each resulting entity, the split has to result in appropriate membership levels. Since those in roles tend to be organizationally bound, critical roles may be moved wholesale into one entity resulting in critical unfulfilled operational roles in the other. There are many solutions to this. Some of these situations involve large business units with their own mirror of the CISO organization, which makes it a lot easier. One of the entities may have to add positions to mirror what the enterprise did for them before divestiture. In a sale to another entity, that entity may have necessary functions already. In other cases, large parts of the IT organization are retained in one entity and its services leased to the other entity for a pre-arranged period of time for the transition. Typically these arrangement are for 3 years or more. These changes also tend to produce disgruntled employees and this must be considered.

Bankruptcy can either be for reorganization or for termination of the existence of the entity. Reorganization is not very significant from a protection standpoint other than the effect of creating disgruntled and frightened employees. Termination of a business leads to termination of all employees and sale of assets. This implies a variety of information protection functions that are usually poorly fulfilled and brings possible liability to the officers. Private information protected by law includes, but it not limited to, protected health information, individual financial information, human resources information like employee records, and business financial records. All of these must be properly stored or disposed of according to the legal requirements for that sort of data. Proprietary materials from third parties, like trade secrets, must be protected. Items covered by intellectual property rights, like copyrighted materials, may have to be protected. Classified or similarly controlled information has to be properly handled regardless of the business status of the entity. In short, end-of-life processes must be properly managed during a bankruptcy process.

Dissolution for any other reason than bankruptcy, or at the end of the bankruptcy process. also has to deal with the life cycle issues associated with systems, data, and people.

People

People have life cycles, and every facet of their life has implications for the enterprise and its information protection program. From before conception to long after burial, there are life cycle issues in the enterprise.

Conception is typically a private matter, however, prior to conception, health care programs at the enterprise have to reflect proper status of the mother in order to assure that medical care and job assignments are proper for the status of the individual. Women of child-bearing age are restricted from certain roles for liability and health reasons. These issues are handled by information systems and must be properly protected from disclosure or corruption while still being reflected in roles available to the individuals in use control processes.

Pregnancy usually brings more use restrictions and changes behavioral patterns of individuals. This leads to differences in behavioral detection models and responses to different sorts of behavioral detection results. Work hours may change, location may change, and in the latter stages, leaves may start, with the corresponding changes in use.

Birth creates new identities within enterprise systems, for example, associated with health care programs and in similar areas. These identities have different status than others within the enterprise records and require different protections.

Education impacts qualifications of employees for different positions and benefits are often associated with education. For children of employees, school and day care records may be available at the company for emergency contact or other purposes. These have special protection requirements as well because they may involve protection of minors.

Marriage often brings about name changes that need to be reflected in identity records, but these changes require historic association in order for time to be properly accounted for. Most current identity management systems handle such changes poorly. Marriage also has impacts on benefits and other similar issues that lead to the need to protect different information in different ways. Marriage changes behaviors, and the protection system must compensate for these changes as well.

Divorce, like marriage, often brings about name changes, requires tracking processes, changes of status, benefits, and other information, and has implications for privacy of records. Divorce is also a life change that may produce erratic behaviors. It tends to remove stabilizing factors that effect suitability for certain tasks, however; these effects are not universal. As a result, divorce should trigger an evaluation relative to life stability for people in sensitive positions. Divorce may also change identity-related information, contact information, and so forth, and this leads to potential tracking issues associated with granting access, just as marriage does.

Training and the tracking of training are important to the protection program because training affects qualifications, and because training requirements associated with certain job functions must be fulfilled in a timely fashion or the individual has to be decertified for those tasks.

Hiring processes involve background checks, verification of resume facts, and checking of references. These are important to initial establishment of clearances at hiring. For sensitive positions, more in-depth checks are required. In the information protection program, such checks are typically made part of the personnel reliability program. Hiring processes also involve requirements for initial awareness and training that must be fulfilled and documented, creation of new enterprise identity information, association of roles with individuals, and other similar processes associated with granting access to enterprise systems and the initiation of behavior and life cycle tracking processes.

Promotion typically comes with new responsibilities associated with information protection. The training and awareness program needs to include new security-related duties in the promotion process, including issues associated with the evaluation of security performance in subordinates, where appropriate. Promotion may result in changes in authorized access and this has to be reflected in role changes and access to systems, facilities, and information. Behavioral changes associated with the new position have to be reflected in detection profiles. Promotion also requires a process for hand-off of content and capabilities to replacements as appropriate.

Demotion is usually not a happy moment in a career and it is a time of change that can often generate a disgruntled employee. Behavioral changes must be watched as well as recalibrated for the new roles and responsibilities. Demotion typically results in role and access changes and these are typically supposed to happen during the meeting when the employee is notified of the change. Demotion also requires a process for hand-off of content and capabilities to replacements as appropriate.

Suspension of people mandates suspension of many but not all information technology privileges for the period of the suspension, tends to generate disgruntled employees, and results in behavioral changes that need to be reflected in behavior tracking systems. This also requires a process for hand-off of content and capabilities to replacements.

Vacation should lead to temporary suspension of many, but not all employee access rights for the period of the vacation. Vacations tend to lead to short-term changes in employee behavior upon return, but these end in a day or two in most cases. Training and awareness levels should be checked on return as well. A process for hand-off of content and capabilities to replacements may be needed.

Illness severe enough to produce days away should generate changes in access for the period of the illness.

Leaves typically run for periods of days, weeks, months, or more, and should be associated with temporary suspension of many, but not all, access rights. Upon return from a leave, training and awareness typically has to be undertaken to catch the individual up to the current situation. This includes updated security awareness and recertification on systems where the training requirements may have lapsed. Extended leaves also require a process for hand-off of content and capabilities to replacements, as appropriate, and return of the hand-offs upon return.

Job changes produce changed roles in most cases, resulting in the need to terminate previous accounts, create new ones, and so forth. This also requires a process for hand-off of content and capabilities to replacements as appropriate.

Moves involving home address changes or changes in workplace or office number lead to changes in access controls associated with network connections, and other similar changes within systems and tracking. Updates to historic records to reflect these changes are needed in order to assure that mail gets redirected, and movement of content and systems from place to place requires physical protection during the move. Inventory processes should be undertaken before and after such moves to assure that lost items of value are identified and that loss is prevented where possible. Moves often result in end of life processes for stored data, and this has to be properly handled as well.

Resignation typically involves a planned departure. The circumstances may dictate special precautions, and because resignations, unlike terminations, are not surprises, there are typically concerns about theft of proprietary information between the notice and the termination of duties. As soon as resignation is notified, information protection actions need to be taken to protect against actions of the terminal employee, and sensitive access should be removed or closely surveilled for the duration of employment. Most resignations are given on a few weeks notice, which provides time for transfer of content and knowledge, however; content should be immediately secured to the extent it is in tangible form to assure against any actions by a disgruntled employee who may be resigning. A standard resignation process should be in place to manage this process properly. Many resignations correspond to competitive moves and these should be examined if potential harm could result.

Termination typically involves a formal meeting in which the employee is notified of the termination. During this meeting, access should be suspended or terminated, all equipment and access devices should be gathered, and proper forms should be signed to acknowledge termination requirements and reaffirm employee agreement issues. Information technology should preserve data associated with the individual at this time and provide means for administrative access. The employee should be escorted from the start of the termination meeting until they leave the premises. If they need to clean out their desk, this should be supervised by an adequately knowledgeable person to assure that only authorized material is removed. This process should be well defined and consistently applied at all levels. Home access should also be terminated and any equipment or other materials in the worker's home should be gathered as part of the termination process. Remote control mechanisms may be used to disable access to content on uncontrolled worker systems, and keys to buildings and systems should be disables or otherwise rekeyed to prevent exploitation. A common practice is to withhold the last paycheck until extant material, like badges and equipment, is returned in good condition.

Retirement is usually a ceremonial time with a party and memories of various sorts displayed for fellow employees. From an information protection standpoint it is very much like any other termination. The process should be similar, well defined, and strictly followed.

Death of a worker may seem like the end of the life cycle tracking but it is not. It is processed similar to a termination except that the employee is unavailable for participation in the process. If there is a death in the worker's family rather than the worker, the life change will result in some behavioral changes as well as th need to invoke processes associated with insurance and so forth.

Legacy of employees, even after termination or death, continues for a substantial period. Records have to be retained for different time periods depending on specifics, but normally 7 years of history are retained for business records unless other requirements apply. Accounts and data may be used over a long time frame and these should be reassigned to those who take over the workload. The identity information associated with an employee may remain associated with their identity and data life cycle processes must be careful not to mis-associate identity with legacy information. Retirement funds and other similar financial or health-related information may continue to be handled for a long period of time, and benefits may accrue to dependents and descendants indefinitely.

Disgruntled employees and ex-employees

There are really only three choices here; (1) terminate their employment, (2) make them happier with work, or (3) let them fester and eventually cause harm. Making them happy is preferred if they are highly productive. If this fails or if they are marginal in terms of performance, termination generally is preferred. Festering is undesirable but often done. Ex-employees without access predominantly threaten leaks and harassment and must be met with court orders and similar mechanisms when they get hostile.

Systems

Systems have life cycles that can be as short as a few weeks to as long as decades. Hardware replacement cycles typically dictate that components are replaced within 10 years of installation for most computer systems, but some supporting infrastructure equipment like telephony systems and cabling, air conditioning, and heating units last for 30 or 40 years. And many systems have all of their hardware and software replaced over time in an evolutionary process. As a rule of thumb, changes in systems have costs that increase by a factor of 3 to 10 for each step in the life cycle up to maintenance. So every poor protection-related decision made early that could have been changed for a dollar in the conceptual phase of the system results in repair costs in the range of hundreds of thousands to millions of dollars in operation.

Conception of systems typically comes from a few people who think up the idea of what the system will do. This is the point where considerations about information protection should start to enter the picture. The protection concept should be an inherent component of the idea underlying the effort. This is more important for bigger ideas that will have longer life cycles because the errors made early will turn into larger and larger costs over the life cycle.

Design of systems must consider information protection issues in order to make choices that lead down more fruitful, more securable, and less costly paths in the long run. Designers should consider all of the life cycle areas as well as the need for integrity, availability, confidentiality, use control, and accountability. They also need to have adequate expertise to make reasonably good design decisions with regard to these issues, and this requires adequate background and education in these specialty areas that is largely lacking in most engineering and computer backgrounds today.

Engineering systems to work within an environment often involves a lot of systems integration. In this effort there are many sources of incompatibilities between systems that have to be resolved in order to allow interoperability. These interface issues are also security issues in most modern systems. In many cases the engineering design has faults that are carried into implementation because the problems were not thought through as deeply as they should have been. Since there is no systematic approach to engineering solutions, it is the creativity of the engineers that has to be counted on. A large part of the engineering experience is related to what the engineers have seen before, so it is important that they be exposed to many of the more common security-related design faults in order to avoid them in future designs. There are also some limited tools that help check designs for known fault types. Design processes associated with high quality are typically applicable and the CMM-SEC and NSTSSI processes are good first steps to doing reasonably secure design.

Implementation involves security issues associated with procurement of components, design and code review processes, protection testing, audits, change control processes for the larger environment, and so forth. Implementation has to integrate system audit with enterprise audit and enterprise control into system control. Integration of intrusion detection and response systems, identity management, zoning policies, and other similar protection measures into systems happens at this time and, of course, it had better have been considered in the earlier phases.

Operation of systems involves all of the enterprise protection processes and has to produce metrics, generate audit trails, take control signals, fail in a safe mode for the rest of its environment, remain within control requirements, and perform useful tasks efficiently.

Maintenance processes introduce many opportunities for attack, often including remote maintenance or similar capabilities that bypass other protective barriers and controls. These require special maintenance modes and controls, including separation from other systems while in maintenance, sound change control processes for making changes, and verification and reintegration after maintenance. Maintenance periods typically involve different people than normal operation periods. Proper control over their presence and access has to be maintained. Storage media used in maintenance has to be protected as does data associated with testing processes, special access, and passwords associated with maintenance processes. Maintenance access should be disabled during normal operating periods.

Disasters occur from a wide range of causes and with enough frequency and range of effect that they destroy or disable components of systems within significant radii. Overall business function for substantial businesses has to survive disasters that leave most of its potential business operating, but not global catastrophes that would put it out of business regardless of information technology. This can only be done by redundancy in capabilities and people, and diversity of locations. During disasters, normal physical protections in place will almost certainly fail, but the overall protection, in terms of risk management requirements, must not fail, even at this time. Planning must include the potential for disasters.

Recovery processes involve the ability to restore business operations in a timely fashion after a disaster or other less harmful event. This requires people, systems, data, and business change-overs and a well-tested and practiced plan. Recovery should have well-defined starting and ending conditions and process checks along the way. During recovery, normal protective measures are often bypassed. Risk management should either dictate that the change in risk profiles be acceptable or otherwise mitigate these increased risks as part of the recovery process.

Upgrades to systems are commonly done without significant concern about protection, however, for medium and high valued systems, change control processes should be required. These processes assure that upgrades are thoroughly tested before being put into use. Testing normally covers operation over a period of time under benign circumstances. Protection testing for malicious attacks is a far different challenge. Malicious upgrades have been used by attackers, so verifying the source and integrity of the upgrade is vital to effective change control. Control over systems changes is often not feasible at the level desired, so at some point risk has to be accepted in most cases. As the value of the system increases, acceptance of risk should be made harder and harder.

Transformations of systems from function to function tend to happen over time. Transformations are typically evolutionary and, when not properly planned, they often result in protection issues. As a general rule, planning these changes to start at the conceptual level and work through all of the other early systems phases is an effective way to deal with transformations.

Consolidation of systems to join functions is a common cost saving activity, but as systems are consolidated, the risks associated with the preconsolidation systems are aggregated into the consolidated result. The resulting risk aggregation has to be analyzed and proper safeguards taken to compensate for the change in risk and resulting change in requirement for certainty associated with the result.

Obsolescence happens as systems near the end of their useful life cycle. As systems enter this phase of operation they are generally replaced or a decision is made to terminate the functions they provide. Over time the maintenance costs go up until it is more cost effective to recreate the system than to run it any longer. During this phase of operation there is a tendency to reduce the utility of the system and its criticality, thus reducing it protection requirements. The key thing to assure here is that protection is reduced only as the risk is reduced.

End-of-life happens for all systems eventually. As systems become decommissioned, care must be taken to assure that they are no longer needed. This typically involves running at least one full business cycle of every still desired function of the system before shutting the old system down. After the system is shut down, residual data remains an issue from a confidentiality standpoint and accountability remains an issue until all value is certified as gone. Formal policy, procedures, standards, and documentation are required for system end-of-life.

Reconstitution of systems after the end of their life cycle is rare but it can and sometimes does happen. In this case, all of the protective functions associated with its creation must be followed and reviewed for changes in situation between the time the system was decommissioned and when it will be reconstituted. After reconstitution, normal processes associated with end-of-life must be redone when the system is again decommissioned.

Resale of systems after decommissioning should be straight forward. The only real requirements are verification of the decommissioning process, its resulting elimination of residual data and value, and documentation associated with the accountability aspects of the sale and retention and disposition of content.

Destruction of systems, once data has been removed is used for cases where the junk value of the components resulting from destruction exceeds the resale value of the system or where disposal is less expensive than alternatives. Destruction can also happen as a result of events. If destruction is for resale value or disposal, end of life processes should assure that residual value is appropriate and destruction may proceed following all applicable laws and regulations associated with environmental and health standards. Many computer systems include parts with hazardous chemicals, such as PCBs, and care must be taken in disposal to avoid downstream liability. For systems destroyed as a result of events, additional end-of-life processes may be required to assure that residual value such as confidential data is not present in the “destroyed” form.

Recycling of components and materials is fairly common in the computer industry and it should be considered as an alternative to destruction and disposal. One of the best programs is the use of old computer equipment in schools, where 3-5 year old personal computers may be well-used for many years. Recycling of materials within systems, such as gold, silver, and other metals can often pay for the destruction and disposal process associated with the remaining components. Many companies now put used computers up for sale on e-bay or other auction sites. They may only get 10 cents on the dollar, but this is 10 cents they didn't have before, and they avoid the expense of proper disposal. If fully depreciated, income may need to be balanced against disposal costs. Finally, computer museums are starting to arise, so old high-valued systems may be turned into museum pieces at the end of their life.

Data

Life cycles for data are often ignored because data is thought of as passive, however, data is the representation of the content that is vital to business operations. Throughout its life cycle, data must be properly cared for to assure that the business operates as it should. The terms data, information, knowledge, and wisdom are often intertwined and misused. Generally, data as presented here is the representation (i.e., a realization in tangible form) of content (the stuff that has utility). Information is defined as symbolic representations in any form. Knowledge is something that computers don't really have, but if they were to be considered in this light, knowledge would likely be considered the combination of information and processing suited to applying it to useful purposes. Wisdom is rarely found in people and certainly never found in computers except as data representing human wisdom if properly interpreted.

Inception of data occurs when real world events take place outside of the realm of the computer system or when the computer generates some internal signals at an electromagnetic, optical, mechanical, or other physical level. All sorts of data exists that cannot be sensed by computers and this is ignored by the computers leading to limitations on their cognitive input capacity.

Observation depends on the sensor capabilities and limits of the device doing the sensing and the ability of the system reading that sensory data to interpret it and transform it into a form that it can use. For example, many programs read inputs and ignore certain characters, and systems typically strip off protocol elements in the receipt of data. The limits of observation are also limits on the ability of the system to differentiate inputs of different sorts and a resulting loss of capacity to detect many deviations that could yield useful information about source and integrity.

Entry is generally considered the time at which the data becomes something that can be stored, used, processed, output, or otherwise comes into the control and possession of the computer system at the logical level of programs being able to act on it.

Validation processes are often used to check for proper syntax, limits, and internal consistency. Syntax checks are fundamental to effective security and failure to do proper syntax checks at input is responsible for the vast majority of current technical computer attacks. Generally, no input sequence that is not legitimate and valid for the application in context should be accepted. This includes limits on length, value, symbols and symbol sequences, and all of these in the context of program state. Limits are used to prevent excesses based on policies or design. For example, input length limits should correspond to designed storage for inputs and dollar value limits on transaction amounts should be determined by user, context, and company policies. Many inputs contain redundancy, such as the entry of a postal code and state in a form. Since many postal codes map to one state, any sort of inconsistency between an entered postal code and the entered state can lead to a detection of invalid input. Addresses can often be tracked to zip codes today because of the increasing accuracy of geographic data, so these checks can be very effective at correcting input errors as soon as possible.

Verification is the use of redundancy to confirm or refute assumptions. For most cases, verification implies a separate and different method of confirmation than the original source. For example, if the weather report indicates high humidity, it can be readily verified by a sensor. The level of verification typically depends on costs associated with verification and risks associated with the use of unverified data.

Attribution associates data to its source. Generally, there are 4 levels of attribution discussed in the literature. Level 1 attribution is associated with the physical input channel, such as the remote IP address, the wire that the signal arrived on, the telephone number of the remote data entry terminal, or the terminal connector that was used for the entry. Level 2 attribution seeks the indirect version of level 1 attribution, attempting to track data to the system or hardware device that first transmitted it. Level 3 attribution, also known as source attribution, associates data with its human or other real-world source, the party or condition responsible for its entry. Level 4 attribution associates data with the organization behind its source. Level 1 is usually relatively easy. Level 2 is very complicated unless a great deal of surveillance is in place. Level 3 depends on psychological characteristics and may be easier than level 2 if differentiation of source rather then specific identity is desired. Level 4 attribution requires an intelligence operation to be effective in a malicious environment. Attribution and the ability to verify attribution leads to assessment of trust. For example, when a well known expert says a product is good it may be taken far more seriously than when an anonymous reviewer on e-bay says it is good.

Fusion of data takes place in systems that typically do normalization and correlation of some sort. The result is typified by proximity to known situations in a state space. This produces secondary, tertiary, and n-ary derivative information that is applied or stored as data for other processes. Fusion is fraught with errors and assumptions and is thus a far more complex issue from a protection standpoint than data. Fused data also has mixes of the properties associated with the sources and processing mechanisms used to derive it. For example, if highly sensitive data like the schedule of a military operation is fused with common data, like weather information, the result may be highly sensitive (i.e., the change in schedule due to a storm) or far less sensitive (i.e., the total fuel consumption estimates for the operation which may vary because of weather, time, target location or other factors). Fusion leads to data aggregation as well, and this can cause two otherwise non-sensitive pieces of information to be sensitive when combined. For example, departmental total salary may not be sensitive while individual salary might be. But if you can get departmental totals before and after each new employee is hired into the department, you can readily derive the starting salary of each individual. Similarly, because of the nature of pricing of medical procedures and tests, knowing the medical fees paid leads to the procedures and tests performed, which in turn leads to the medical conditions of the patients. Thus medical bills become sensitive protected health information because of the ability to fuse them into protected health information.

Separation requirements associated with data are generated because only separation technologies are sure to limit the flow of information. Data separation is typically at the heart of zoning policies and other related issues. Generally, data is associated with classifications and users are associated with clearances. Data is only accessible to users when the user clearance is commensurate with the data classification. Functions performed are then limited based on the needs of the user with respect to the data.

Analysis of data involves the processing of the data through state machines so that the output of the state machine has utility in a different context. This is typically the sort of thing done when so-called raw data is mixed with other data, transforms, and process models to produce meaningful content for the user that is only indirectly related to the data itself. For example, temperature gradients on a wing may be mixed with simulation models and analyzed to determine aircraft stability. Errors in analysis may produce dramatic side effects, so the integrity of the analysis process is often critical to the business function. For example, analysis of data associated with a bridge may reveal or fail to reveal structural limitations that could cause the bridge to fail under load conditions.

Transforms are commonly used to change data media, representation, form, format, or utility. For example, data associated with a simulation may be transformed into graphical format and mapped into a picture to produce a graphic depiction of an event. Transforms are commonly used to extract subsets of data, for example to differentiate intrusion-related audit data from unrelated data. Transforms are used to change data into formats used in different applications or systems, like a transform from EBCDIC to ASCII for moving content from mainframes to personal computers. Transforms are used to reformat data, like putting a presentation into columns. Transforms are used to change media, for example to place the data on a backup tape. All of these transforms are critical to the function they support and thus transforms must be protected for business function to be protected.

Transmission is generally associated with the data in motion as described elsewhere. In transmission, data becomes susceptible to a larger set of attacks associated with the larger physical space and increased number of media and systems it passes through, or comes into contact with.

Storage is generally associated with the data at rest state which is described elsewhere. In storage, data tends to be localized and concentrated in a small physical space, and thus has the advantage of being physically securable and the disadvantage of aggregating risk in space and time.

Use of data is generally associated with the data in use state described elsewhere. When in use, data must be in usable form. There are few options for protection of the data without protection of the mechanism that uses it. Thus protection of data in use typically involves protection of the operating environment.

Presentation of data typically involves transformation into a presentation format and display on an output device. This may be presentation for human consumption or for automation such as process control systems. It is critical that the presentation accurately represent the intent of the application. For example, many presentations are intentionally deceptive, or at least misleading in that they emphasize things the presenter wants to put forth and minimize issues the presenter wants to be ignored. The presentation of statistical information is notorious enough to have its own saying: “lies, damned lies, and statistics”. From an information protection standpoint, this has a range of implications.

Modification of data can be accidental, intentional and appropriate, or malicious. Accidental modification is generally undesirable and can be covered by relatively simple statistically verifiable controls such as redundancy and fault tolerance. Intentional and appropriate modification is desirable from the standpoint of being able to enter and alter values associated with the business utility of the system. For example, changing your address so you can continue to get your mail when you change offices is a business function that involves legitimate alteration of address data. Malicious modification of data is highly undesirable and protection typically involves the use of cryptographic checksums for detection and access controls for prevention. Someone else changing your address as part of an identity theft is an example of the same change used for a malicious purpose. Integrity is a function of intent, and computers are notoriously bad at dealing with issues of intent.

Loss of data can cause business consequences associated with the value of the data unless appropriate protections are in place. Value comes in the form of business utility associated with its use. That utility may be lost from the loss of data. Redundancy protects against loss of utility unless all redundant copies are also lost or unavailable in a suitable time frame for use. Preventing release depends on confidentiality protections, typically mandating the use of encryption or prevention from physical access even when in possession of the data's container.

Recovery of lost data comes in one of several forms. The data may be backed up or otherwise kept, sent, or created redundantly, it may be regeneratable at a cost, it may be recoverable from partially broken or deleted media, and it may be located and recovered by physical or electronic means. Insurance may cover the value and the legal process may aid in recovery of the value through civil and/or criminal sanctions. With the exception of risk transfer techniques, these typically involve outside specialized expertise associated with data recovery, computer forensics, private investigation, or law enforcement processes.

Reconstruction of data is sometimes a choice if the data is derived from other data that is available, if fragments exist at different places, or if the original values can be derived from other data values associated with or derived from it. A really good example was a data set that was destroyed in a fire but was reconstructed from portions of it that were previously emailed to other parties. Those parties sent back copies of partial subsets and they were combined together to reconstruct enough of the original data to meet the need.

Backup of data is a fundamental process designed to assure availability over time. Different sorts of backup are required for different circumstances. The decision about which types to apply stem from timeliness, redundancy, transportation, quantity, and duration issues. For data that has to be restored from backups in near real time, duplicate (hot standby) systems are typically used. For data that has to be very redundant, the redundancy requirement leads to the number of copies and their diversity in space and media. For data in large quantity or that has to be at distant locations in some time frame, different media and bandwidth are acceptable. For backups required to last different amounts of time, different storage media and processes are used. All of these vary with the specifics of the application, almost all combinations of these are attainable, and the costs vary with the need. More and harsher requirements increase costs. For typical data, typical backup regimens include daily incremental backups of changed data kept for one week, weekly incremental or full backups of all data kept for a month, monthly full backups kept for a year, and annual full backups kept indefinitely or retained for the legally mandated duration for business records. Backups have to be tested by restoration on a regular basis in order to assure that they work, tracking backups and selectively restoring from them is problematic for sequential media such as tapes, and large-scale backup facilities on-site and off-site are commonly used for data centers. Data retention and disposition issues also drive back processes to an increasing extent.

Restoration from backups is a process that its typically tuned to the backup process. Restoration processes depend to a large extent on timeliness requirements and media. Restoration in real-time usually requires backups on media similar to the original, and in many cases is implemented by transaction replay processes at secondary sites. Less real-time restoration can involve wider ranges of processes.

Destruction of data is problematic. Generally there are several types of destruction processes associated with digital data and different methods associated with paper, CD-ROM, DVD, and fiche data that are most commonly used.

- For digital data stored on disk or tape, deletion of files is most common and least effective. It is trivial to restore this data and it should never be used to destroy data of substantial value. Secure deletion based on multiple pattern-based overwrites is used against medium-grade threats. For higher grade threats electromagnetic erasure with high Oersted field generators can be used but is limited because generators may inadequately penetrate the media. Physical mangling of disks is ineffective against high-grade threats because remaining fragments store large quantities of data per unit area. Destruction of media and contents by burning at high temperatures or boiling in acid for long enough time is most effective.
- For paper media, strip shredders are the most common method of destruction. They are ineffective and easily defeated, leaving only a false sense of security. These

shredders are consistently and easily defeated. Cross-cut shredders are more secure but to be reasonably safe, shreds should be sized on the order of a few square millimeters for typical printouts. Sensitive and non-sensitive data should be joined in the shred bins to increase volumes. Shredding should be done by the individual doing the disposal, not through a service that shreds elsewhere. The best common process cross-cut shreds, then burns or pulps in a recycling process physically controlled by cleared personnel.

- For CD-ROMs and fiche, data density is far higher than for paper. Shredders of the sort described above are effective but leave shards large enough to extract useful content. Burning or emulsification with acid is preferred.

Data states

Data is generally treated differently when at rest, in motion, and in use, and the data state has a great deal to do with the protection mechanisms and need for protection. There is an age-old characterization of data at rest, in motion, and in use, but today, with mobile computing, data is often at rest in a disk and in motion because the disk is in a personal data assistant or laptop computer. Or it may be in motion because the laptop is on an airplane and in use because the user is using it. So in a sense we have a new state of data. But tapes were always moved from site to site as part of backups and airplanes have had computers in them every since computers come into substantial use. So when these terms are used, they refer to properties of the data. More than one set of properties may be in effect at any given time.

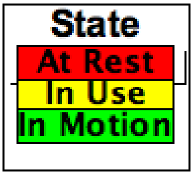


Figure 8-3
Data states

Data at rest

Data at rest is, in essence, data stored at a physical location in a physical device, typically a disk, CD-ROM, USB storage device, etc. In most cases, computers with high valued information in large quantity remain in one physical location. This means that the physical security measures associated with that location act as a significant part of the protection afforded to that data.

Storage of data is in physical devices like disks, tapes, CD-ROMs, DVDs, paper, fiche, and more recently electronic storage devices like PCMCIA cards, USB drives, and the like.

- **Tapes** are usually disconnected from any computing device and only come in contact with those devices when passing by the tape head that reads or writes them. They are often manipulated using robotic devices to move them between large storage areas and tape readers and writers. Those readers and writers are most often disconnected from the computers that use them and they are accessed at a distance over internal cabling. Tapes are large enough that they have to be concealed with something else that is noticeable in order to be removed, often have bar codes or other similar markings to allow them to be identified and tracked, and are usually stored within hardened data centers and other similar areas. It takes a lot of time to go through a tape and it can really only be accessed sequentially, so while it can have a high burst rate, it is slow to get to any particular place in a tape and it takes as much time to delete a whole tape as it does to write over it all. Because of the large numbers of tapes compared to drives in a typical data center it would take weeks or months to manually erase a large percentage of the data. Tapes are rarely missed over periods of hours to days so they can often be removed, read or written, replaced, and not missed. Tapes need to be read every few years in order to be refreshed, as they age beyond 10 years they start to become unusable, and they tend to lose data when placed into environments in excess of 100 degrees F. RFID tags are usable on tapes and may be applied in some circumstances to track movement into and out of facilities and areas.
- **Disks** are high speed for input and output, typically sized either for laptops (3 inches by 2 inches by ¼ inch give or take) or for internal use (2 inches by 3.5 inches by 6 inches give or take). They store up to about 1 Terabyte each, can read, write, or delete information at about 10 gigabytes per minute, are randomly accessible for rapid access to files, include the electronics needed to read and write them, and

are usually stored in systems that are using them all the time. Because they are generally in use, it is hard to remove them for duplication or destruction without being noticed Because they are usually within cases inside systems, it is often hard to gain physical access without being quite obvious. Disks are usually replaced every 3-5 years because storage is increasing so quickly that 100 5-year-old disks can be replaced by one new disk. They also tend to fail after 2-3 years of use, so replacement is mandatory to reliability. Old disks lose value quickly, so they are often discarded instead of being resold. Proper destruction is critical to the protection process and these disks must be properly handled.

- **Paper** storage is one of the most overlooked protection issues in many modern enterprises. In almost every protection posture assessment paper-based data that is readily accessible contains enormously damaging information that nobody notices moving from place to place, can be easily copied, can be burned in a fire, can be used for illegal purposes, and can be altered or replaced unnoticed. In one recent assessment, paper found at one location included (1) hundreds of completed immigration and foreign worker forms; (2) name, address, medical, pay rate, bank account, and social security data for almost all employees; (3) a complete printout of not-yet-released corporate books for a fiscal year including details of customers, suppliers, prices, expenses, locations, and operations; and (4) medical records for thousands of employees. Protection of paper records is clearly vital.
- **Fiche** and similar records are kept in most cases to allow smaller space to be used to store more historical data that must be retained but is rarely accessed. Fiche is far more easily taken without being noticed, can be taken in larger volume because of the smaller size per datum than paper, will almost never be noticed for a long time, and is rarely inventoried even in the disposal process to assure that everything that should be disposed of is disposed of.

- **Electronic storage devices** today tend to be relatively small, certainly fitting into a shirt pocket, and in some cases embedded in other small devices. They are readily disguised, operate at high speed, and can tolerate substantial harsh handling without losing data. They attach to a system in a second, are recognized and mounted, can be loaded with data within a few minutes or less, and removed immediately. As a result, they are ideal for moving data in and out of environments surreptitiously. They store gigabytes of data, so for most applications, especially espionage, they are very handy. But for corporate storage of high-valued data they are too easily removed, copied, and replaced, too easily stolen, not as reliable as they might be, and hard to control as inventory items or as authoritative sources of data and value.

Retaining stored data requires media-specific processes to assure operation over long time frames. This is typically handled by the hardware and automation in systems. But there are other retention requirements associated with stored data that are far harder to properly carry out. These are the legal requirements for data retention associated with business records and the requirements associated with data retention policy that have to be implemented in information systems. Generally, laws require retention of normal business records for 3, 4, or 7 years in the United States depending on the specifics, and for material records associated with a business, 7 years retention are required in the United States according to Sarbanes-Oxley regulations. Businesses tend to want to eliminate records as soon as possible in order to limit liability, so many have very short retention times for things like email. But this is potentially problematic and may result in fines or criminal sanctions against individuals and companies. EU regulations further complicate issues by requiring that certain privacy-related data not be retained past the amount of time required for its use. For most cases this is something like 7 days to one month for things like passport numbers, telephone numbers, addresses, and so forth. This potentially interferes with customer service requirements, warranty information, shipping and receiving records, and so forth. This also has interactions with backup policies and practices since retention on backups and other media have to be handled. Tracking all of the data at rest also becomes problematic, particularly when it is in motion between being at rest. Even eliminating all records of a particular transaction that is not kept in many records systems becomes difficult. For example, in a recent criminal case all but one copy of a document was removed from records and backups, but one backup copy of a file server copy of the record stored while in transit in an email server ended up being found and the case was dramatically impacted as a result.

Protection of data at rest is often facilitated by **operating system access controls,** which can be highly effective. They are often more effective than alternatives. They are faster, more reliable, and better for survivability and recovery processes. They are easier to use than alternatives like disk, file, or record encryption and cryptographic checksums, respectively, for achieving confidentiality, use control, and integrity. Availability is generally assured with redundant disk storage as a local solution and distributed backups, checkpoints, and transaction records as a solution for transaction systems, databases, and file systems that support this sort of change mechanism. Accountability is typically retained by ownership records and accounting data sent to external audit collection and retention systems, retained locally if adequate system protection is available, or sent to write once read many (WORM) disks if they are available for this purpose.

Backup is described elsewhere as associated with data life cycles. Backup for data at rest typically comes in the form of redundant arrays of independent disks (RAID), removable backup media, file server backup areas, transaction-based remote system backups in hot, warm, or cold standby modes, or long-term storage in other forms at remote backup facilities and recovery sites. For most enterprise data centers, backups go to tapes and copies of those backups are sent to a remote site for disaster recovery. Depending on timeliness requirements, backups may be made continuously, periodically, or on special occasions. Backup scheduling is covered under life cycle issues associated with data under backups.

- **RAID** backups come in the form of multiple disks containing portions of the data in an arrangement that assures that as long as m-out-of-n disks are working, the data will continue to be available in real-time. However, most RAID implementations are designed so that once the (n-m)th disk fails, the data is unavailable and very hard to recover. This makes RAID resilient up to a point and then brittle. Worse yet, because most RAID arrays are implemented with identical disks, there is a tendency for them to be installed together and fail at very nearly the same time. RAID failures out of line with the expectation based on statistics occur because these models ignore the change in failure rates over the life of a device. The so-called bathtub curve indicates that at the start and end of life cycles failure rates are far higher than the steady state rates during normal operation.

- **Removable media** backups typically include CD-ROMs, DVDs, WORM drives, and tapes, however, increasingly disks are being used in this capacity as well, through removable drive bays and firewire or similar interfaces. Of these choices, for enterprises, only tapes are realistic today in large data centers. Disks are expensive and no automation exists for storing large numbers of them and automatically mounting and unmounting them. CD-ROMs and DVDs store too little for effective backups of today's large storage media and inadequate automation for them is also an issue. WORM drives are really only used for specific applications where each operation is backed up for safety or liability reasons, such as in manufacturing facilities. Tapes, with all of their limitations, remain the only real viable removable media for large data center backups.
- **File server backups** are particularly useful for network-based backup approaches. Terabyte (1000 Gigabyte) file servers are now available for under $1000 each. They can be placed on Gigabit or slower Ethernets and used to store backup data from large numbers of systems remotely. Because they are on live systems, restoration can be immediately done by the user who owns the files by copying those files back. Automation is used to backup changed files to these file servers at any desired period, and typically

backups happen at the early hours of the morning or at randomly chosen times after network access is available. The scheduling issues are complex here. Backups done when computers are turned on result in large numbers of backups the first thing in the morning, which collapses network availability. Backups scheduled at a time of day fail because the computers are not always on at that time. Backups done by user fail because users fail to remember or do them. Some companies try policies of keeping computers turned on at night, others automate some sort of overnight startup and shut down process, and others try other methods, but all of them have problems. File servers are also useful for backing up larger permanent systems. Storage area networks are the evolution of these file server approaches. They use a name space to provide very large amounts of storage for backup purposes, sometimes ranging into the 1,000 terabyte scale. But even then, care in what is backed up is required to prevent overrunning available capacity. For example, a company with 100,000 computers each with a 40 gigabyte disk drive that is half used generate 2,000 terabytes in a single full backup. The vast majority of this data comes from almost identical contents such as operating systems, standard applications, and files that are not very important. The enterprise data classification scheme should include information on what needs to be backed up with what level of redundancy and how often so that this classification can be used to make automated backup determinations. These backups have another problem in that they typically generate more files and space over time because they intentionally do not delete things that fail to appear on the new backups. If files are moved or copied, the copies are generated and the subsequent deleted copies do not get removed. Tracking the precise file system state for this sort of backup mechanism is not widely implemented yet today.

- **Transaction-based remote backups** depend on having a transaction system, a transaction-based file system, or another way of turning changes in data at rest into transactions. The initial state is synchronized and then transactions are sent to the remote system and replayed there for updates. As a result, the backup has an identical state to the original. Of course this has its problems as well. One of the problems is that an attacker who does a massive deletion of files will generate transactions that delete those files on the backup system. For this reason it is important that instead of replaying all transactions as they happen, checkpoints are taken and transactions are recorded. In this way any previous state can be restored. But this process involves even more storage because duplicative efforts generate excess transactions. Nevertheless, this capability is very handy and highly desirable if a proper file system state is to be restored and counters to common attacks are to be successful.

- **Long-term backups** in remote backup facilities and recovery sites are also very common practice. Typically, for disaster recovery purposes, off-site backup copies of tapes or other media are made and physically transported to a backup site for use in emergency. This is rarely done more than daily because of the limits of the transportation system as opposed to the communications system that is more real-time but lower volume per time (1,000 terabytes can be shipped anywhere in the world in less than a day, but via communications this is problematic). Increasingly remote backup sites allow transaction-based updates or backups, so disaster recovery processes are implemented by doing electronic off-site backups, but the cost of maintaining high-speed lines for this purpose is substantial and the potential for a remote attack on the off-site backups is also worth considering in an evaluation of the tradeoffs.

Restoration process depends on timeliness requirements and backup approach. For real-time restoration, hot standby systems are the only realistic solution. Warm standby systems work for near-real-time restoration assuming that some amount of state can be lost without consequences outweighing cost savings. Cold standby equipment is commonplace, typically in the form of computers of similar configuration at another location where the off-site backups are kept, tested, and restored in case of emergency. Increasingly, enterprises are recognizing the need for geographic diversity of personnel and systems and moving toward an approach where research and development systems feed change control systems that are identical to the systems in the field. The change control testing area can be turned into a live site as part of a recovery operation at any time, and since it is an exact copy it should work identically. When not used for restoration and recovery purposes, this site is used instead of sitting idle, so it is an even better value for the money than a standby site that is not used on a daily basis. Another alternative is a shared recovery site in which several companies share a computing facility used by any of them when their primary site is out of service. This is fine for extremely local disasters but becomes a contention issue in larger scale failures if not carefully planned.

Data in motion

Data in motion may operate through physically secured wiring and infrastructure. If the physical security is adequate to the need, no additional measures are required. However, the vast majority of information in motion today travels over long distances through insecure infrastructure. In these cases additional protection is required as the consequences increase.

Extracting the data from its at rest state can be on a push or pull basis. Push systems, like email, have transmissions generated by the sender. In these systems, the sender is typically responsible for providing appropriate protection. Pull systems, like Web services, have user requests for transmission that are serviced by servers. These servers may take into account the user request and authorization based on identification and authentication to determine the proper protection associated with the transmission and then use protection as appropriate to the situation. However, most servers are not very good at it.

Encryption is one of the main technologies used to protect content in transit. Because secure socket layer (SSL) encryption is so inexpensive and universally available, it has become a de-facto standard for encryption of data in transmission. Encryption, if properly done (which it rarely is), allows communication to be kept confidential. But, on its own, it provides no protection other than confidentiality. Encryption gets its utility from a combination of the cryptographic algorithms used, the cryptographic protocols used to control the transmission sequences, and the implementation of those algorithms and protocols. While cryptographic algorithms are typically very hard to defeat if well chosen, cryptographic protocols often leave major vulnerabilities in systems, and implementations almost always fail to meet the need if attacked with a reasonable level of effort.

Authentication is used to verify the validity of an assertion of identity. Surety varies with method and implementation. Authentication is usually done by verifying combinations of things that you are, have, and know or can do. Biometrics associate physical properties such as iris patterns, finger prints, DNA structures, facial patterns, keystroke patterns, speech patterns, hand size and shape, footfalls, and other similar recognizable and differentiable characteristics of individuals. Most of these systems are useful for differentiating any of 1000 or so individuals from each other with reasonable numbers of false positives (acceptances) and false negatives (rejections), but they are poor at real-time identification of individuals. Therefore their prime use in information protection is in verifying an identity and not in identifying an individual. Many biometrics can be easily spoofed, are not scalable, and use insecure infrastructure. Things that are possessed typically include badges, software, digital certificates, time or use variant tokens, specialized hardware devices, cryptographic keys, and other physical keys or devices. They can all be stolen and many of them can be duplicated or spoofed. Things that you know are limited because of human memory limitations which makes them potentially guessable. They typically have to be revealed in some form in order to be used, thus leading to their duplication and unauthorized use. Things that you can do are rarely used but can be effective. Multi-factor authentication is used to increase the difficulty of attack at the cost of increased difficulty of use and reduced convenience. Such systems also have to have bypass capabilities for practical use in most enterprises. These bypass capabilities may be less well protected than the rest of the system and they tend to aggregate risk, making the bypass mechanism a prime target for exploitation.

Transmission of the data, possibly in cryptographic or other form, involves the translation of the data into a format and signal form suitable to the transport media. For example, for optical media bits have to be turned into modulations of optical signals. Transmission can be made over multiple channels and paths for diversity. In some cases spread spectrum techniques that change signal channels over time and introduce false signals into other channels to obfuscate messages are used to protect from surveillance. Redundancy with spectrum spreading increases signal effectiveness over noisy channels and resists thin spectrum jamming, forcing jammers to increase their power over a wider spectrum to be effective and making them readily targetable as a result of their increased power footprint. Frequency and path hopping can be as effective as encryption at concealing content but they are less common than other techniques. Path diversity is harder to implement because of the increased cost associated with multiple paths and because the total number of paths available from any given physical location over wired infrastructure are often limited to a very small number. Transmission often uses compression to increase effective bandwidth and may use cryptographic checksums on transmitted data to allow receivers to detect intermediate changes.

Transport media has effects on accessibility. Radio is a broadcast medium allowing anyone within a signal to noise dictated distance to receive signals. In the case of satellite communication, this typically extends to at least one continent at a time, making the signal very widely available. For mechanisms like wireless WiFi and Bluetooth systems, the radius is on the order of a kilometer if the listener is skilled and signal focusing (devices that focus signals directionally) or reduction (like special wall paint or building design) methods are not used. Wired media like cables or hubbed Ethernet systems are also broadcast media over the locations the wire extends to. Switched infrastructure uses point to switch signaling and switch to switch consolidation of signals and allows SPAN ports to access all traffic, but under better control. Routed networks limit paths to relevant paths for the specific bit stream, but can be redirected and used for broadcast and SPAN eavesdropping as well. Telephone transmission systems are line switched point to point communications with consolidation switch to switch, necessitating either central office or wire access for attack. Access is often available at the interface points and outside of structures. Wires transmitting electromagnetic signals also generate induced signals that are readable without physical penetration of the cabling and even optical fibers can be read with laser interference methods, but fiber has far less cross-talk requiring less separation of cabling for effective protection against high grade threats.

Reception of signals depends on environmental conditions that differ with the transport media. For many sorts of optical, infrared, microwave, and similar radio techniques, atmospheric conditions have a substantial impact on reception. WiFi, Bluetooth, and mobile telephone technology have similar limitations. Signal strength for non-fixed systems varies substantially with location. Wired signals have reception problems in some environmental and atmospheric conditions, and power failures, and are subject to damage from Earth movement, electrical shock, and other similar causes.

Verification of transmitted information is typically done at a hardware level, after translation into digital form, through the use of checksums and cyclic redundancy check (CRC) codes. Under malicious attack, cryptographic checksums are necessary in order to verify that received data is identical to transmitted content and these systems are subject to many of the same limitations as cryptographic systems. Verification of syntax, form, and values in context of the receiving system can also be used. Decompression is used to undo the compression associated with the transmission process. Unless verification is properly done, vulnerabilities in subsequent phases of transmission may be exploited.

Decryption is used to undo the encryption process that may be used prior to transmission so that received data is in usable form. Decryption keys must be protected in order to meaningfully decrypt content and prevent others from decrypting it.

Delivery of data to either storage or processing involves operating system operations that may include protection-related meta-data. Generally, access controls or similar protective measures are implemented in this process to assure that information is properly protected on delivery and stored with proper markings and protection settings to allow classification and access controls to operate properly. Translating meta-data is often problematic.

Data in use

Protection of data in use is problematic because it must be in a form that is useful for processing. There are some cases, like comparison to specific known values in password verification, where data can be left encrypted and have utility. But the vast majority of uses require that the data be readable. Data in use is rarely protected against modification beyond process separation mechanisms, because this is not supported by current processors.

Validation of data before use is critical to its proper use. Programs often make assumptions about inputs and those assumptions are commonly exploited by attackers. Input should always be validated for syntax and value ranges based on program state. This is also used to detect inconsistencies and react to them.

Verification is used to increase the surety level associated with data. It can take the form of redundant calculation, redundant data sourcing, or in some cases, a submit-commit cycle. Submit-commit cycles are typically used in conjunction with transaction systems. Submitted data is independently verified before a transaction is committed.

Transforms on data are the dominant functions used in conjunction with the use of data. Inputs are mixed with state data to produce outputs and next states. The outputs represent transforms of the input sequence to an output sequence. Redundant processing is used in some cases to increase surety of results. In some cases processing uses checksums or state verification mechanisms to assure that transformations produce appropriate output. In use, data has to be protected from other data. This typically happens through operating system support of hardware mechanisms for process and memory separation.

Reconciliation is used to verify the consistency of results. This is particularly important in financial transaction processing systems and other high-valued applications.

Instantiation of data involves making copies of instances of data for different purposes. Multiple instances implies a need to mirror protective mechanisms and classifications across all instances.

Attack and defense processes

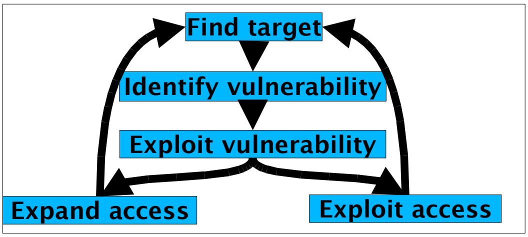


Figure 8-4 – The generic attack graph

Intentional attacks against information systems and technology generally follow a pattern.[8.2] The attacker seeks a target. Once found, the attacker seeks target vulnerabilities and exploits them to gain privileges. Privileges are used either to exploit the access now available or to attempt to further expand access. Exploiting or expanding access can both be used to find additional targets and expand the scope of the attack or they can be used more directly to induce consequences. There is another attack process that has been identified in which the attacker randomly tries an exploit against any target and proceeds based on the assumption that it worked. This attack process is very easy to detect and has a very low probability of success. Figure 8-4 shows this as a generic attack graph.

The attack process is carried out by threats using their capabilities and intents to attack system after system, ultimately leading to consequences. Defenses can sever the attack graph, or at least act to reduce the likelihood and increase the time associated with traversing the attack graph. The empty circles in Figure 8-5 indicate defenses that cover (eliminate) threats, vulnerabilities, or consequences from the example attack graph. In the example, even though there are vulnerabilities, threats, and consequences remaining uncovered, there is no path from threats through vulnerabilities to consequences. As a result, this set of protections will be effective even though it may cost less than eliminating all of the threats, vulnerabilities, and/or consequences (which is likely impossible anyway).

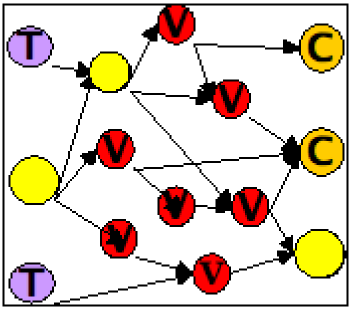


Figure 8-5 – An attack graph

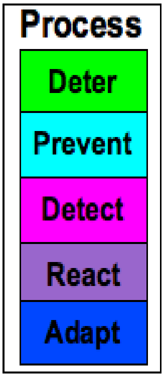


Figure 8-6 – Defense processes

The defense process outlined in Figure 8-6 consists of deterrence, prevention, detection, reaction, and adaptation. Deterrence includes any efforts to reduce the interest of attackers in specific targets. This involves psychological processes directed at the attacker and often uses deceptions of one sort or another. Prevention is typically attained by technical safeguards that limit access or function in some way. Detection should be designed to provide timely notice of event sequences that have potentially serious negative consequences. Reaction is the tactical response to attacks that are detected, which typically involves immediate actions to mitigate harm. Adaptation is a strategic response, typically involving architectural, process, or procedural changes. [8.9]

Deter

Deterrence typically happens at the management level through decisions associated with public relations, business ventures, responses to other attacks, and corporate stances on issues that effect the decision process of the attacker. It includes preventing attacker awareness of targets, reducing interest in them as targets, putting up barriers that make the attackers think targets are not worth the difficulty of attacking them, causing attackers to believe they will be caught and prosecuted if they try to attack, or to believe it is immoral or unethical to attack.

Perception in the mind of threats is the target of any deterrence process. The goal is to influence attackers to prevent them from attacking. There are several common methods used to achieve this.

Deception involves causing attackers to misperceive the object of their attacks. As a deterrent, deception can increase the workload and decrease the certainty of attack. Deception has proven effective in reducing the desire to attack, disrupting group processes among attack groups, and increasing the cost to attackers. [8.3]

The path of most resistance deters most non-directed attackers who seek the path of least resistance. Even those who are determined, tend to try simpler things first and move on when the going gets tough. Only skilled, directed professionals attack hard targets with determination.

Arrest of perpetrators and widespread publication of arrests are the most effective deterrents against the commonly estimated 1/3 of people who will break the rules if they think there is little likelihood of punishment.

Prosecution increases the perception that attack will be harshly met. Unfortunately, too many executives who get involved in abuse of systems end up unpunished even though those who work for them are more likely to get prosecuted. Prosecution of executives and the threat of criminal prosecution under Sarbanes-Oxley have provided a dramatic change in deterrence of corporate crimes and have dramatically increased the adoption of regulatory mandated protective measures.

Policy should provide for **sanctions** that are clear and uniform and identify those sanctions with specific acts so as to deter those acts. Policy should also require that these sanctions be read, understood, and agreed to by those who agree to work for the enterprise so that it becomes a term of employment and is included in contracts.

Awareness of sanction policies and consequences of actions are effective and should be included in awareness programs to help deter criminal acts by employees and other authorized workers.

Deterrence is usually not explicitly considered by enterprises even though they do a considerable amount of deterrence in their processes without thinking of it in these terms.

Prevent

Prevention of attacks is done by stopping the attacker from finding a target, identifying and exploiting vulnerabilities, and expanding or exploiting privilege. Techniques used to prevent attacks are identified in the security database at http://all.net/ and these techniques can be used to sever attack graphs.

Deception techniques prevent attackers from finding targets, identifying vulnerabilities, and expanding or exploiting privileges. They are predominantly used today in preventing target detection by making it harder for an attacker to differentiate a legitimate target from a false target. Once an attacker is detected attacking a false target, real targets are made unavailable to the attacker thus forming a detection/response loop that acts to prevent further attack attempts from finding any real targets. Against random attacks, deception can be used to cause any random attack to have a low probability of success followed by acting to cut off further attack graphs.[8.3]

Firewalls are designed to cut off attack graphs that start on one side of the firewall and go to another side of it. They act as prevention mechanisms at the perimeters of enclaves, or in different terms, the firewall defines the perimeter of an enclave. They prevent the identification of targets and exploitation of target vulnerabilities by preventing information flow between the attacker on one side of the firewall and their target on another side of the firewall. They limit the expansion and exploitation of network access by limiting the range of other network locations that can be reached and the manner in which they can be reached. [8.4]

Authentication is used to prevent an attacker from doing what an authorized user can do. More and more sure authentication techniques are used to increase the level of certainty that the user is who they claim to be. This extends to software acting on behalf of users as well as in roles that are only associated with automated systems.

Authorization associates authorities with authenticated identities. Authorization mechanisms include both the technical mechanisms that allow an identified and authenticated user to perform functions with data and the mechanisms used to grant, revoke, and alter those authorities. The administrative control over authorities is often the weak link in a system of controls. Authorizations are often set incorrectly and fail to properly associate the clearances of the user, the certainty of the authentication of that user, and the classification of the data. As a result, users have too little authority to get their job done or more authority than they require for the tasks they need to perform. The principle of least privilege asserts that the latter is inappropriate and the former implies a lack of proper business function. But getting precision in authorization is difficult because of the complexity of systems and the mismatch between technical protections and management intent. A typical system has millions of protection bits and inadequate technical tools to manage them. And those protection bits usually control technical mechanisms that are too large grained and incommensurate with policies. Finer granularity is achievable at the cost of more time and space. Commensurability is not obtainable because computers do not have mechanisms to operate on intent. Arbitrary rules can be reasonable for people making decisions (i.e., in an emergency where granting this person access to that data might save many lives, grant the access) but trying to codify all such rules in terms of things that computers process is infeasible for the foreseeable future.

Access control mechanisms are typically based on a subject/object model.[7.1.15c] Subjects are typically user identities associated with processes. Objects are usually devices, files, records, or fields. The access control might be based on any number of implementations ranging from capabilities lists to access control lists to group and user identities associated with processes and files, but they can all be modeled as subject/object matrices in which each subject is granted a set of rights with respect to each object. This is a stateless model in which the sequence of events that got the subject to where they are has no effect on the rights they have with respect to objects. Models with state dependencies are far more complex and have not been implemented to date. [7.1.2-8,]

Use controls, in the sense of technical prevention mechanisms, are programmed mechanisms that associate functions with situations. For example, when a Web server intermediates between a browser and a database, the Web server is often granted access adequate to perform any of the functions it can perform for any use it is designed to facilitate. This lack of access control is (usually poorly) compensated for by the Web server using applications that only use appropriate functions for the situation. These controls tend to be far weaker than operating system controls because they depend on situation dependent code that is less tested, in larger quantity, less controlled in its development, and more easily exploited than typical access control mechanisms. However, these use controls can make more situation-dependent decisions.

High-speed **intrusion prevention systems** (IPSs) are really just systems that detect intrusions and respond to them before they are exploitable. IPS depends on adequate and appropriate detection of intrusions, an OODA loop that is fast enough to respond before serious negative consequences arise, and a response that is not exploitable to the advantage of the attacker and that is effective at preventing the serious negative consequences. Today there are systems that can detect certain classes of known intrusions and react in time to sever the attack graphs associated with them, however; they produce false positives, false negatives, and can be exploited for reflexive control, typically resulting in denial of services to legitimate uses.

Architecture acts as a preventive measure if properly implemented. **Separation** as an architectural principle is one of the keys to success. For example, by separating networks into areas based on the need to communicate, attacks that otherwise deny services on a large scale are contained to the areas in which they first appear. This limits damage and provides time and controls to help mitigate the problem. The separation of audit from control from data is also central to proper network operations and to meeting regulatory compliance requirements. Separation tends to be far less expensive to implement and operate than more active alternatives, it is more certain to work, and it tends to solve problems over longer periods of time, reducing churn associated with many other more active controls. [8.6]

Surety is a measurable basis for asserting the certainty with which the protective measure will successfully perform the function it is intended to perform and no other. Higher surety indicates more certainty. Different protective measures have different surety levels and costs. Generally defenders should favor higher surety at lower cost, like choosing separation mechanisms in architecture over IPS approaches when they cover the same attack graphs. But very often there are tradeoffs between cost and surety and in these cases the decision is not so clear.

Detect

The fundamental purpose of detection is to identify event sequences with potentially serious negative consequences in time to mitigate those consequences to within acceptable levels. Unfortunately, this is not how most current detection systems work. Rather, they detect what they are able to detect regardless of the utility of those detections. [8.4.3]

Detection is an enormously problematic area. Ideally, detection would never be used because prevention would be perfect. However, we do not live in an ideal world As a result, detection is necessary, at a minimum, to provide redundancy for preventive techniques so that when they fail there is a chance that their failure can be detected and mitigated. But detection technologies have increasingly become preferred over prevention technologies by many decision makers for a variety of good and not-so-good reasons.

Fundamental issues associated with detection make it problematic. For any of the more interesting things that current detection systems try to detect (e.g., viruses, spam, intrusions, anomalies, spyware, etc.) there are a potentially infinite number of false positives (false alarms) and false negatives (missed alarms) for any finite detection time. This means that detection always has to deal with these issues in order to be effective and that response has to take this possibility into account.

Detection of attacks of this sort is and will always be unreliable, it takes time, causes delays, and costs a lot to operate. Detection of attacks generally has to be updated because attackers adapt to detection mechanisms. This means that they must be actively attended to in order to remain effective. Detection implies response and response implies investigation, so the indirect costs of detection are high when many detections take place. Ideally, detection operates in a relatively quiet environment with little noise and few attacks to detect. But when used as a substitute for an effective prevention mechanism, detection is very expensive to operate, uncertain, and complex to get right.

There is, however, a chicken and egg problem associated with detection. Without detection, many attacks go unnoticed in the tactical time frame and may never be associated with their consequences. For example, an attack that reveals pricing information and causes no other harm will be reflected in a more competitive sales environment. It may seem like the competition is heating up and eventually you may even go out of business, which has happened from this very cause. The problem is that, without detection, justifying costs of prevention is difficult, and effective prevention means that little will be detected and most of it will be unimportant. So in order to justify more budget for protection, more detection is used, the number of apparent attack attempts increases, and the case for more defenses grows. As a manager, the hard thing to do is to be effective at preventing and detecting attacks with potentially serious negative consequences while getting adequate funding to meet the need without resorting to scare tactics or creating false impressions in order to get proper budget.

Host-based detection resides at endpoints. It has the advantage of having host state information available for its analysis but the disadvantage of not having access to related information from other hosts. Thus it lacks the context to understand the larger picture. Hosts tend to have excess performance available and thus host-based detection can use more computer time per host. In the aggregate far more unused computer time is available in hosts than in most other places. Host-based detection can look at stored state information over long time frames, giving it more potential for deeper inspection.

Network-based detection operates based on network traffic. It has the advantage of being able to cover many hosts with one mechanism in one location, but the disadvantage of not having host state information available for situation analysis. The performance of a network-based detection system is limited because the bandwidth must be significant in order to gain the advantages of centralized detection. The ability to retain historical information, relate information over long time frames, and correlate information from many different hosts is limited by memory and performance, which means that as bandwidth goes up, the analytical capacity per packet goes down. The leverage gained by centralizing the function is paid for by a reduction in available analytical power.

Intrusion detection is a term associated with known techniques that can be codified in specific terms, but of course many sorts of intrusions only become known to defenders after they are detected by other means, and many are never known. In terms of detection, current systems typically only detect known intrusions, and specifically detect only a very small class of intrusion types that are detectable by observing specific event sequences of short length from the same user. While research has produced far more advanced intrusion detection techniques, they have not been substantially implemented in commercial products. Most known intrusion detection techniques are easily bypassed. For example, in an article written several years ago, 50 ways to defeat intrusion detection systems were identified, almost all of which work against almost all current detection systems. [8.4.4]

Anomaly detection seeks to detect events and sequences of events that are different by some measure of significance from “normal” events and event sequences. In law enforcement there is a saying “JDLR” which stands for “just doesn't look right”. When something just doesn't look right, investigation is necessary in order to figure out what's going on. The same is true in detection for information protection. Anomaly detection leads to investigation. The false positive and false negative problem is reflected in a quantity of detections over time. Too many detections overruns the available investigative response capacity of the defenders, while too few detections reduces the justification for that capacity. Thus the detection thresholds are often set to match the investigative capacity of the organization rather than to reflect the value of the detections. A proper feedback system should use the results of investigations to determine what thresholds on which sorts of anomalies justify alarms and the system and staff should be adapted to those needs rather than using staff levels as a basis for choosing what events and event sequences to alarm on.

Automated response based on anomaly detection is also problematic. Without investigation, anomalies are not to be trusted as a basis for action other than investigation. In situations where anomalies are very serious and known to cause serious negative consequences in time frames that are short, automated response must be carefully predetermined to assure that it will always result in a fail safe condition.

Behavior produces externally observable events. These are the events that detection systems try to observe. Limits on observation are associated with the limits of sensors, the limits of translation of sensor data into representations, and the limits of detection system capabilities for analysis of sensor data. Behaviors associated with systems and people in situations are typically predictable to within some limits and this predictability leads to detection of deviations.

Situation provides context that is used to determine the acceptability and normalcy of behaviors. There are sets of situations in which certain behaviors are acceptable, but codifying all of the situations associated with each behavior at fine granularity is infeasible. As a result, situations are generally split into large-grained classes.

Patterns are matched with event sequences in context to determine if the events are to trigger a detection.

Heuristics are sequential machines used as a more general form of pattern matching mechanism. They are typically coded as sequences of triggering conditions and actions, but may be arbitrary state machines.

History is often used to calibrate anomaly detection systems and historical data is sometimes recorded and replayed for calibration purposes.

Authority of users to perform tasks is sometimes used to differentiate between legitimate and illegitimate uses. By using detection to identify cases when authority is apparently exceeded or does not match actions, attacks that bypass protections can often be detected.

Identity is sometimes mapped into event sequences so that the identity of an individual can be used to differentiate legitimate from illegitimate event sequences.

Collection of data for detection and collection of forensic data related to the detection process is necessary in order to perform analysis and to assure that adequate record-keeping is done for legal and regulatory purposes.

Preservation of data is typically required for its use in any subsequent legal action. This should be done as part of normal business record recording processes and should be well structured and documented to assure that it is not easily challenged in court.

Fusing data together is required for detection of all but the simplest of known attack patterns. This typically starts with session-level fusion so that parts of the same session are translated into input and output sequences as associated with each finite state machine (hardware device, protocol element, software program, or application) for analysis of its impact on that machine. At the next level of abstraction, changes in these machines should be fused against other changes in the total environment to identify implications of those changes relative to expectations. This level of fusion is almost never used in current systems, however, experimental systems at this level have been implemented.

Analysis processes are used to match fused data against criteria to determine what constitutes a detection and what properties to associate with those detections. Analysis for a wide range of classes of attacks have been determined to be undecidable so there are and will always be infinite numbers of false positives and negatives for general purpose computing environments. However, most high valued systems in use in enterprises are really not used for unlimited purposes and adequate characterization of their operation can be used to dramatically reduce false positives and negatives. While the integrity of data is a function of intent, in many systems, intent can be well and clearly defined for all but the most unusual situations.

Attribution of actions to actors is critical for the association of detections to those responsible and for the resulting consequences. [8.7]

React

Reaction is dependent on detection. Without some sort of detection there is nothing to react to, and if detection is not accurate or timely, the reaction will also have problems with being appropriate and timely. Particular problems arise with automated response systems because they form reflexes of the enterprise information infrastructure. If these reflexes can be triggered by attackers they can be used to induce undesired responses that damage the enterprise. A classic and one of the most easily exploited examples, is the introduction of false packets into a network so as to cause the detection system to assess that an attack is underway by one critical system against another critical system it is linked to. The detection system identifies the packets as an attack. This in turn induces an automatic response of cutting off the attacking system from the victim of its attacks. The result is the severing of communications between two critical interdependent systems. This problem stems from a combination of factors, often including poor design, inaccurate detection and attribution, the need to react quickly and automatically to certain classes of attack in order to limit damage, and a lack of proper architectural planning and response analysis.

Investigation of detected event sequences is necessary in order to determine an appropriate reaction. For certain classes of sequences, automated responses are developed, but this allows reflexive control attacks. Unfortunately these same sequences in different contexts may require immediate response in order to limit harm, so investigative processes have to be balanced with immediacy. Investigative processes also have a tendency to produce far more information than a simple explanation of the event sequence of interest. In case after case, seemingly trivial detections have led to investigations that led to larger and larger scope. In many cases, these lead to large-scale criminal or civil prosecutions. Investigations typically start with a triage effort by internal incident response team members and follow through until there is reason to believe that something involving inappropriate behavior has taken place. At that point the investigation has to be handed over to investigative professionals in order to result in a positive outcome. Many amateur investigations end up producing serious problems. These include legal liabilities, harassment suits, inadequate evidence, loss of critical forensic data, and inability to prosecute. When insiders are involved in cases requiring investigations, the potential for investigative leaks and cover-ups increases dramatically. Unless there are professional internal investigators on staff, outside private investigative teams that specialize in computer-related investigations are usually used. Investigations are usually carried out by, through, or in conjunction with corporate legal counsel. [8.8]

Assessments are undertaken in response to high-consequence detected incidents at two levels. An initial assessment and any number of small follow-on assessment are often undertaken as part of triage efforts to determine who to contact, how far to escalate responses, who to get involved, and so forth. In addition, incidents sometimes generate an awareness that results in more strategic assessments such as IPPAs.[2.6] While this is not the optimal approach to resolving enterprise issues, many CISOs and others use the response process to justify such assessments because these are the only times that management shows a willingness to spend enough money on such an issue to get it done.

Refocus of attention and resources often occur in response. The details of each event in context drive processes in different directions, cause sensors to be adapted, thresholds to be changes, forensic data to be generated and analyzed, etc.

Coordination is required for complex investigations that spread to involve large numbers of systems or systems not controlled by the enterprise. Coordination of response process for timing of technical steps is required. Responses may have to be coordinated at a management level. Legal coordination is required at any point where humans get involved or when it is determined that the event sequence is not just a technical flaw of some sort. Investigative coordination is required with law enforcement and the legal system in many cases. Physical security and HR coordination get involved when employees or contractors are involved. Line management gets involved coordinates administrative actions, and executives get involved when consequences are high enough.

Opinions are generated during response processes and these opinions are used to make decisions about how to proceed with the process.

Advice is often given to managers and executives at all levels in order to help them make decisions about actions to be taken, both in tactical incident response and as a result of investigative processes.

Reporting and presentation of detected information and related materials is critical to the response process. Data presentation plays a particularly important part of the reporting associated with response. When a user reports a problem, this is a response to a sequence of events. User reporting is responsible for a significant portion of all detected incidents today. Tracking of reported incidents is used to detect coordinated attacks, and incident reporting data is often used to justify further efforts in information protection. For some events, like the discovery of contraband or the possession of material that is illegal to possess, immediate reporting to legal authorities may be required as well. Legal counsel should be involved in this process.

Covering of vulnerabilities is commonly used in incident response. A typical example is the creation of a firewall rule to limit the use of a port associated with an attack while repairs are done to mitigate the attack.

Disabling of features, capabilities, or select systems is sometimes used to mitigate the short-term effects of an attack. This is typically used when the value of the service is outweighed by the damage of the attack. It is also used during some repair processes to prevent further exploitation until the repair is completed.

Push back is used to try to force the action closer to the attacker. Typically, an attack is detected near its target and as the path toward the target is identified, protective measures are moved closer and closer to the attacker until the attacker is cut off rather than the target being cut off. This strategy reduces wasted bandwidth at the target and in intervening infrastructure but it is problematic against most distributed coordinated attacks in use today. They are so distributed that cutting them off beyond border routers disrupts normal operations.

Deception is a viable response strategy against many attackers, and those who have used deception under the title of honeypots or similar appealing names have been successful at convincing management of its utility and appeal. Deception done properly can be used in a tactical as well as strategic manner and can lead attackers far astray. Depending on situational specifics, deception can be a very useful counterintelligence tool, however; the cost goes up as the fidelity of the deception increases, and substantial expertise is required in order to be effective against high quality attackers with deception. [8.3]

Mitigation is typically associated with repairs of weaknesses in systems that allow them to be attacked. Of course all of the responses described here are part of the overall mitigation effort, but repairs are notionally the path taken to mitigate most harm as a semi-permanent fix. Mitigation of faults in an operational system is several orders of magnitude more expensive than proper design. The deeper problem is that many attacks do not involve weaknesses, but rather exploit normal operations. Unless systems are designed so as to avoid unlimited flexibility and control changes, many sorts of attacks will continue to be mitigated on a one-by-one basis.

Administrative changes to systems are also typical of response processes. Typically this involves cleaning up a lot of side effects of the attack, often in a highly manual and time consuming process.

Prosecution of attackers is last step in response and typically takes years if it is pursued at all. It is rarely pursued because the benefit to the enterprise is only indirect and the cost in time and inconvenience is substantial. When prosecution is eschewed the result is a criminal that continues to commit crimes.

Adapt

Adaptations typically happen at an architectural level and operate as a long-term strategic response to enterprise needs. While rapid adaptations are used in some cases, these usually result in poor solutions that are ineffective and expensive even though they may fulfill an administrative need.

Management of overall information architecture is critical to any adaptation. A management and technical team typically oversee architecture and zoning to make strategic adaptations. This team is often augmented by specialized security architects.

Process requirements for any architectural change require approvals of all sorts, from zoning board functions to design approvals and so on. An excellent source for design process criteria is the NSTISSI series of standards on security design processes for classified systems. While most enterprise systems don't have a need to meet this level of rigor, reduction in rigor is easier than trying to develop a new process from scratch. NSTSSI standards include 4 different sorts of individuals with different responsibilities: [8.10]

- **“***Designated Approving Authorities***”** describes the purpose, applicability, responsibilities, and minimum performance standards for approving authorities. It covers legal liabilities, policies, threats, incidents, access, administrative responsibility, communications security, tempest protection, life cycle management, continuity of operations, and risk management.
- *“Systems Administrators*“ covers purpose, applicability, responsibilities, and minimum standards for administrators. These include access controls, administrative requirements, audits, operations, contingencies, and platform-specific security features and procedures.
- “*Information System Security Officers*” covers purpose, scope, applicability, responsibilities, and performance standards. They include (1) maintaining a plan for site security improvements and progress toward meeting the accreditation, (2) ensuring that systems are operated, used, maintained, and disposed of in accordance with security policies and practices, (3) ensuring that the system is accredited and certified if it processes sensitive information, (4) ensuring that users and system support personnel have the required security clearances, authorization, and need-to-know, are indoctrinated, and are familiar with internal security practices before access is granted, (5) enforcing

security policies and safeguards on all personnel having access, (6) ensuring audit trails are reviewed periodically (e.g., weekly, daily), and audit records are archived for future reference, (7) if required, initiating protective or corrective measures, reporting security incidents in accordance with policy, (8) reporting the security status of the system, and (9) evaluating known vulnerabilities to ascertain if additional safeguards are needed.

- "*System Certifiers*" responsibilities cover purpose, applicability, responsibilities, and minimum performance standards. These include documenting mission needs, registering the new application for tracking purposes, negotiation of security requirements, preparing a plan for accreditation, supporting system development, performing certification analysis, recommending certification, evaluating compliance, and maintaining the certification over time.

Engineering approaches to architecture adaptation include the need for compatibility with legacy systems, meeting cost constraints, integration with enterprise operational capabilities and systems, and understanding how to analyze architectural measures against protection needs. The CMM-SEC approach is also a design and engineering methodology that can be applied to track and measure adaptation.

Architecture of the enterprise network and the application under adaptation require a detailed understanding of existing enterprise information technology architecture, a clear understanding of protection requirements that are driving adaptation, and the history that led to the situation on the ground. Security expertise relevant to the architectural issues is a must, but in many cases architectures go from bad to worse when security engineers ignore the context of the systems. The goal of adaptation should be to provide a reasonably smooth and low cost transition from one architectural state to the next. The overall path may take a long time to follow. Along the way, it is critical to take steps that move toward the objective without disrupting the organization beyond its ability to adapt to the changes.

Organizational adaptation is sometimes called for. Many failures in protection are the result of inadequate separation of duties or similar failures to follow basic principles. These are often driven by organizational issues such as power struggles between managers and executives. When organizations have to be adapted, skill at exercising influence comes to the fore.

Detect/react loop

The detect/react loop is particularly critical to the effectiveness of response process because, if fast enough, it provides prevention, if too slow it is ineffective, and if the attacker can tune the attack to it, they can create positive feedback to amplify the attack.

OODA loops, otherwise known as the Boyd cycle, dominate much of the discussion surrounding this issue because they are convenient if imprecise way to discuss the issue in understandable terms. Observation, orientation, decision, and action (OODA) expresses the process by which events outside a system interact with a system. But of course there are many systems involved in typical attack and defense processes, so there are many OODA loops underway at any given time. Systems exist in layers of context and response processes happen at all layers. Attackers and defenders have limits associated with times between different events. Performance limitations of computing systems and algorithms play into the issue. Human performance is also an issue in many situations.[8.5]

Autonomics are used in cases when human reaction time is too slow or unreliable for effective reaction and the situation can be accurately enough characterized to allow for effective and non-harmful response that fails in a safe mode. An example of an autonomic systems used in a high risk situation is the computers that control the space shuttle during its hypersonic S-curves as it enters the atmosphere. Because reaction times have to be so fast (milliseconds) that humans cannot maintain control on their own, the dynamically unstable system has to be managed by computers. But computers fail, and a single computer failure in this critical operation could cause the shuttle to disintegrate. So redundant computers that detect other computer failures and react to them very quickly are used. Unfortunately, most enterprises do not think though autonomic responses this well, but fortunately, there tend to be clear fail safe conditions for most enterprises and few of their systems cause such large negative consequences so quickly.

Operations have slower Boyd cycles in response to event sequences than autonomics, but many operational mistakes have led to dire consequences. For example, operational errors have brought down large computer networks for hours to days. The brittleness of operational decisions with respect to information infrastructure implies the desire for a way to check things out before instantiating changes. Thus change control is commonly used in operational responses to assure that changes can be undone and that changes don't cause more harm than good. Ultimately simulation systems would be most helpful in this arena. Some networks have such systems in place for verifying changes before testing and doing larger numbers of checks than would be possible to cover by testing alone. Most large enterprises that handle high valued information in automatic systems for a long time have solid testbeds for testing operational changes and have strong change control to assure that this testing is completed before changes are made to the operational network. However; in emergencies, certain classes of changes are sometimes made anyway, and some of these emergency changes cause more harm than the problem they were intended to solve. Risk management is necessary as part of the decision to forgo change management.

Organizations have far longer OODA loops. The need for committee decisions, meetings, verification with legal processes and policies, and the rest of the organizational process that supports information protection requires patience that is sometimes quite taxing. In many cases the detect react cycle for organizations is so broken that individuals responsible for criminal acts continue those acts and are given time to get away. In some cases, they are even notified of the pending actions by the process or team members with mixed loyalties. In these cases speeding up the process may not be feasible, and slow processes lead to extended problem periods. While it is advisable to act quickly in information protection responses at an organizational level, this is only really feasible in hierarchies when the top decision-maker makes a rapid decision. This also tends to have high risks for all involved, tends to reduce the amount of legitimate consideration required for high-valued decisions, and creates more tension than is generally desired in a large organization.

Work flows

Protection process is typically implemented in terms of a set of work flows; standardized event sequences with inputs, state, outputs, and systems that take state and input to produce output and next state; with the explicit purpose of carrying out the processes identified for protection. There are many work flow systems available. They typically handle help desk operations or other similar ticketing systems, and similar mechanisms have been around for many years in the legal profession, medical systems, in aerospace, and in other fields. Manual work flow systems were commonplace up until the last several years and many continue to persist and will for a long time to come.

The advantages of automated work flow systems for security come in several areas. They (1) assure that work gets done in the proper sequence, (2) can act to assure that approvals are properly undertaken prior to actions, (3) can provide automated provisioning integration for automatable work flows like adding user identities based on roles and similar functions, (4) can document the entire process, (5) allow verification, (6) help to reduce the work load for audit, and (7) provide support for process improvement. However, because of their central role in operational aspects of protection they also form risk aggregation points that pose significant risk. For example, identity management solutions that automate some limited components of security work flow associated with access controls, can be attacked to cause all access to cease, to grant access to unauthorized individuals, to destroy the information functions of an organization, or to disrupt operations in automated manufacturing or processing facilities. Providing adequate surety for these systems and disaggregating risks by creating sets of these systems with zones of control and potentially overlapping authorities is complex and problematic, but necessary for the enterprise that wishes to succeed in light of the realities of threats in the information world.

Work to be done

Many facets of information protection exist and the work that has to be done for all of these facets comprises a very significant portion of the total effort in information protection. Work has to be described and standardized in order to fit into work flow systems and this itself is a very substantial effort. There are some partial work flow systems that exist for security but they are nowhere near the level of completeness required for an enterprise and they cover only a small subset of the overall work flow of the enterprise security operation. This book includes high-level overviews of hundreds of processes that all have to be codified into work flows in order for them to be properly handled in a systematic manner for an enterprise. For the small or medium sized business this book can be combined with metrics to form a set of checklists for many of the common functions, and they have been used for that purpose.

Process for completion and options

For each item of work to be done a process for completion should be defined including the conditions for its invocation, times associated with different actions to be undertaken, primary and auxiliary contacts for performing the identified tasks, optional processes for emergency, standard, and exceptional conditions including appeals processes and overrides, and enough details to allow any authorized and properly trained and competent person to carry out the work. The processes should identify points for workers to certify that work has been done, and for those who certify work to do so and notify the system of the verification.

Control points and approval requirements

Most processes have control points of one sort or another. For example, a worker may prepare all of the elements for a building to be wired for electrical systems, but until the building inspector comes and approves of the plan of the building ready to be wired, the wiring waits. In information protection there are similar control points defined, typically when risks beyond thresholds of the level of the current worker are reached. The approval process should identify someone with adequate authority and knowledge to make a reasonable and prudent decision about the risk, identify the risk and the options to the authorized person or people, and seek their approval or rejection or optional paths. In some cases multiple approvals or more complex voting systems may be used and timeliness issues may require actions be taken urgently. Presumably the overall system has to be able to handle this in order to be effective in these cases.

Appeals processes and escalations

Work flows have to have suitable provisions for appeals and escalations when something that one person wants to have done is at odds with someone in the approval path. While most processes don't get appealed in hierarchical systems because of the nature of the structure, in matrix organizations there may be many paths to getting work done. In networked organizations the organic nature of the process often allows many paths to getting something done. But even in a hierarchical process there will be times when escalation is used, for example, when timeliness is an issue and normal approval paths are not available in a timely enough fashion.

Authentication requirements & mechanisms

The quality and quantity of authentication associated with different functions typically varies across a wide spectrum. For example, a simple lookup of the work to be done might require only a user identity and password, while the ability to change a work order may require an additional authentication such as the presentation of a time variant password from a secure token. For some actions physical presence may be required and this may mandate a third party authentication to certify presence along with biometric data and other similar methods. The work flow system has to support the use of different authentication mechanisms to support the different levels of surety required to perform different operations.

Authorization and context limitations

Authorizations associated with identified subjects under different levels of authentication may change with context (see details of context elsewhere) and different situations within work flows. The work flow system has to be capable of handling complexities associated with the specific identified needs of data owners for access to the resources necessary to do work, and in some cases, alternative sources with different authentication requirements may be sought because of circumstance. For example, if time is of import and any two of eight approvers are adequate to the need for a process to continue, the work flow might request responses from all eight authorizers and notify the authorizer's that the work has been approved once two have approved, so that they don't have to look at the issue if it is already settled. Similarly, context may change during the process, thus changing approval requirements. Appropriate methods must be used to properly deal with these situations. The work flow system should also help to prioritize work so that more important or time critical work is given proper priority.

Work flow documentation and audit

The work flow system should provide documentation of what was done and what is to be done and allow this information of be read for audit purposes as appropriate. Detailing should be available to the specific actions taken by specific individuals at specific times, the approvals required and obtained. The work flow requirements of the situation at the time should be documented so that all of the information needed to validate an action after the fact can be made available to the reviewer or auditor. Thus everything needed to determine what was done, why, when, how, where, and under what situational circumstances should be available to check on any specific process undertaken or all of the processes of the system.

Control and validation of the engine(s)

Whether work flow the mechanisms are manual or automated, the mechanisms that control the processes have to be controlled, verified, validated, tested, reviewed, and tracked to assure that they do what they are supposed to do. This includes both the normal operation of these mechanisms and all of the exception conditions and malicious sequences that might circumvent the system at every level of its operation. For example, if work flows are implemented using a paper system to cover regular backups of systems, the process will typically involve the use of a piece of paper that indicates what to do on a given shift. The shift workers then use the checklist, perhaps doing a backup and reflecting that on the checklist with date, time, tape number, and initials. The verification may be done by going to the proper tape number and restoring its contents to a test system to verify that it has the data it should have from that time and date and that it properly restores. Verification of this activity by random sample will validate that the mechanism is being used and operating properly. Additional malicious abuse testing might include seeing whether making a false entry causes a backup to not be done (for example a worker could claim to have done the work on a prior shift even though they did not do it and cause backups to go undone) or by taking away the sheets of paper and determining whether a work around is used to still do the backups and how the escalation process works in that circumstance.

Risk aggregation in the engine(s)

Automated work flow systems tend to aggregate risk by centralizing and unifying the processes that the system supports, by combining the information and capabilities of the work flows into a single computer or at a single location, by unifying the administrative aspects of managing those systems, by using common operating environments with common mode failure mechanisms, by combining previously separate mechanisms, and by creating dependencies on the work flow system for proper execution of work. At the same time these systems reduce costs, increase efficiency, improve auditability and accountability, reduce time to get many tasks done by using computer communications to replace paper processes, provide for more efficient and effective backups of the work flows, and so forth. The question for executive and risk management to answer is how much risk can be aggregated before additional protective measures are required. As a rule of thumb, and based on the notion that the surety should match the risk, as risk gets to the medium level, medium surety techniques should be used. As the work flow system reaches to risk levels where single individuals can no longer be permitted to make decisions, multi-person control must be added, and risk disaggregation by multiple work flow systems or the use of other compensating controls must be used.

Inventory

Inventory is perhaps the least understood area of protection architecture and yet one of the most developed areas of information technology. Inventory started being used in businesses before the start of written history and is considered one of the key reasons that numbers were first developed by humans.

In the case of information protection, inventory plays a supporting role by providing a collection of content, protective mechanisms, people, things, systems, software, hardware, network elements, controls, process elements, and potentially everything else discussed in this text. The overall security inventory is the list of things and properties associated with those things that allows systematic tracking of what is to be done with what, where, when, why, how, by whom, and the basis for a repository for the history of everything that was and everything that was done with it. Taken in this light, the information protection inventory seems like it is the key component to the protection program. And yet you can almost never find such an inventory within an enterprise. [O.7]

In reality, most enterprises have many different inventories and those inventories are used for all manner of different things. The information protection function today usually leverages existing inventories, creates selective inventories of its own, and misses a lot of things that might otherwise be better protected. Processes, audits, automated processes, and many other such things help to form the inventory elements involved in different aspects of the program.

For example, a requirement for a network audit might drive the networking support group to generate an inventory of networked devices. This might be gathered automatically by a network discovery tool and fed into a risk management system by a network security team to form the rudiments of a top-level device inventory. At the same time, somebody in risk management might be generating a different inventory of important content based on discussions with top management. An inventory of software on personal computers (PCs) might be maintained by the desktop management team as part of their systems management efforts, and an inventory of servers might be maintained by another team.

Occasionally an ambitious CISO might generate a far more comprehensive inventory by compiling content from many different sources. Increasingly, there are tools that provide the means to consolidate these inventories into a common repository and recollect the content over time from the systems that originate it through network-based connectors and automated processes like hiring and termination processes. In some cases, these mechanisms leverage directory infrastructure as part of the identity management process to hold and consolidate these records because directory integration has produced a variety of methods including meta-directories that do very much the same sort of activity and often have connectors for common databases.

In order to be really useful, the inventory should include a lot of information about the inventoried items. For example, it is important to link content to the aspects of that content that make it subject to the protection program. If content is subject to particular laws or regulations, it should be linked to those laws and regulations, and the database should permit queries about all content related to a particular law. The data retention and disposition process for the enterprise should be closely linked to this mechanism so that, for example, all files to which a given individual had access over time can be identified and searched for activities related to a pending legal matter. This has to happen quickly in order to meet legal mandates, but also to prevent the eternal retention of all content because it might be related to a pending legal matter.

The granularity issue quickly comes to the fore when issues like the time transitivity of access are analyzed for histories of user access. Since tracking of this is known to be high storage and computation time,[7.1.16] it becomes necessary to find alternative approaches. For example, tracking individuals to roles and tracking roles accessing content over time will cause enormous expansion of possible access, but can be used as a first cut to limit searches. It is likely that programmed searches will be required for some time, especially when the size and complexity of meta-directory-based inventories is as high as it is today.

In most cases, inventories will be kept to the level of the database or file rather than to the level of the record. When more detailed information is required, audit trails or record-level information will be queried from relevant databases. This corresponds to the likely granularity of requirements for control as well as management awareness and association with applications. Shared databases become a more complicated problem because of the transitive effects of content flows over time, but this is almost never managed today or even tracked.

Inventories are used in risk management, among other reasons to determine when the process is done. As such, inventory is vital to not missing big risks. Interdependency analysis also links into inventory issues and, for a large enterprise, becomes extremely hard to track without dependencies in an inventory and automated analysis. Change management process also depends on or creates its own form of inventory embedded into schedules. Things like periodic reexamination of risks requires that inventoried items be linked to review schedules. Access controls require inventories of personnel and their clearances and content and its classification in order to enforce protection requirements. Access methods require the same inventory along with identification and authentication methods and associated surety levels to grant authorizations, which in turn operate through identity management infrastructure that is a database which acts as an inventory for the purposes of tracking access and provisioning processes. Control objectives have to be mapped into the inventory in order to properly associate controls with content. The organizational and functional processes have to be tracked and controlled and this means an inventory of some sort, if only embedded in the regularized process and work flow systems.

Whether realized and explicitly controlled or not, there is an enterprise information protection inventory, it is widespread, likely highly distributed, very large, non-uniform, diverse, and vital to the operation of the protection program. At a minimum, it should be recognized and controlled. Risk aggregation issues associated with inventory failures and access should be analyzed and considered, and inventory process should be included in the protection management process.

Protective mechanisms

Protective mechanisms are the technical mechanisms that are directly in contact with or control of the content, threats, vulnerabilities, or consequences, and assure the security of the content and its utility while supporting its business utility with minimal friction. [8.11]

Perception

Perception-related defenses are typically used to influence the attacker and as such they are directed at the attacker. While outsiders are subject to many defenses, insiders can also be affected by perception. Perception is a rather substantial field and a lot of research has been done in this area, however; the basics from an information protection standpoint can be characterized in terms of:

Obscurity: By making facilities, people, systems, and other elements of the information infrastructure and capabilities of the enterprise difficult to understand and find, it becomes harder for the attackers to locate and attack or exploit them. While obscurity as a defense is often viewed negatively by some members of the information protection community, obscurity plays a critical role in almost all protection schemes. If everything is known about a system and its defenses it is indeed far easier to defeat than if it is less well known to the attacker.[8.12]

Profile: Keeping a low profile is very helpful in defeating many attackers. While many business leaders and experts make their living largely by being public personalities, for most people, keeping a low profile is relatively easy to do. Buildings that have data centers, for example, should not be marked as such, because by marking the building or making its purpose easy to understand, attackers are given an easy method for target identification. In addition, random attacks and attacks by many group threats tend to be oriented toward high profile targets, and those attackers can be avoided by this approach. Even against insiders, computer centers with large glass walls in imposing spaces may end up being targets of opportunity or foci of resentment. Keeping locations with high value obscure can be highly effective at reducing insider threats to physical infrastructure just as keeping the names and locations of financial and critical systems obscure can reduce the number of insiders likely to try to attack them. [8.13]

Appearance: Many people have misimpressions of enterprises that stem from historical information, rumors, competitive advertisements, industry norms, or small numbers of negative contacts. These people may become threats if they believe that the enterprise has done something wrong or illegal. The appearance of an enterprise, a system, a facility or a business venture has a direct effect on the set of threats that are likely to be faced by it. While there will always be some threats that will get past appearances, many will not. [8.14]

Deception: The general field of deception for information protection is substantial and growing. It is based on the notion that there are error mechanisms in automated systems and people that can be exploited to induce faults in their processes. These error mechanisms can be effectively exploited to defend systems from attack. A common deception is the use of a firewall to suppress the information about internal systems, and another one is using the same reply for a failed password as a failed user identification in a login. Deceptions can also be use to induce false signals, for example to disrupt threat-induced network intelligence efforts or to present easily exploited systems so that attackers will exploit them instead of higher valued systems. [8.3]

Structure

The structure of networks, systems, applications, facilities, and businesses can effectively limit risks. These structural mechanisms provide layers of defense against attacks from different sources. Of course a fundamental thing to understand about structures is that those inside any given area of a structure are essentially past the barriers and those barriers are of no effect except in their role as separators or delays. Separation mechanisms lead to zones that are differentiated from each other by those barrier mechanisms and form a set of logical spaces.

Mandatory or discretionary access controls are mechanisms that enforce separation controls based on subject/object models that control access of subjects to objects. Discretionary controls are controlled by the user while mandatory ones are controlled by the system itself. Mandatory controls generally follow a control scheme like the Bell Lapadula model[7.1.4] used in the Trusted System Evaluation Criteria [7.1.8] that is designed to protect users from getting access to higher classification levels than their clearances allow. Discretionary access controls like those in Unix and similar systems allow users to set protections of files to determine whether they are private to the user, the user's group, or not private.

Information flow limitations are used to form barriers between regions[7.1.7,10] as opposed to the lower-level subject/object controls of typical access control mechanisms.[7.1.15c] This is usually used for network separation, like in the use of virtual local area network (VLAN) technologies with rate shaping to separate areas of networks, or the use of router-based controls to limit network addresses, physical interfaces, and network ports across routers or switches. Rate limits on networks are used to limit denial of services attacks and routing can be used to force specific traffic to travel along specific routes.

Digital diodes and similar mechanisms provide high assurance that information can only go where it is supposed to go.[7.1.10] They do this by physically limiting the flow of information. A digital diode physically assures that information can only flow in one direction with the side effect that reliable transmission requires redundancy as it does in a broadcast media, because not even protocol confirmations can be allowed to pass. Lower surety diodes exist as well as similar mechanisms with small covert channels. The ACAT guard and similar technologies are used to allow outward flow of information from more classified to less classified areas by passing it through a human and automated guard station that is certified for the purpose. This allows many covert channels as well as leaking classified information in steganographic or other coded forms.

Firewalls and similar permeable barriers are used to limit the effects of issues on one side of the barrier from impacting other sides of the barrier while still allowing select information to pass.[8.4] Modern firewall appliances also include content control mechanisms, however those are covered separately. Firewalls tend to have demilitarized zones (DMZs) and/or proxy servers designed so that packet-level and transport-level attack mechanisms cannot pass through the firewall but are stopped by it. They tend to control allowed protocols, ports, addresses, and to a lesser extent, sub-protocol elements. They tend to save state information and often perform network address translation so that direct access between separated network segments cannot work without the firewall present, thus reducing the problem of cross connects and eliminating addressing of systems on other sides of the firewall except those that are supposed to be accessed.

Content controls

Content controls consist of three classes of controls; separation mechanisms such as those identified above, transforms, and filters. These controls operate by examining the syntax and markings associated with content and the situation or context that the content is being applied in.

Transforms are used to encrypt, encode, or authenticate the representation of content so that it is meaningless if illicitly examined, of utility for use in legitimate applications, or detectable if modified. Digital rights management[8.15] software, encryption hardware and software[8.16] , virtual private networks (VPNs),[8.16] and digital signatures[7.1.12] are the most common transforms. Transforms are also used on markings associated with content to reflect changes associated with functions performed on the content. Transforms are medium surety level mechanisms if properly used and can reach high surety for very specific circumstances. [7.9.11,]

Filters limit what is allowed to pass, and include such mechanisms as virus detectors, spam detectors, spyware detectors, Trojan horse detectors and similar known bad content detection mechanisms. They can also be used to prevent unauthorized syntax and data sequences from passing outward. Known content filters are subject to large numbers of false positives and negatives, leading to their low level of surety as protective mechanisms. [8.17]

Markings are used as part of content control to allow known and readily readable small amounts of content to be associated with larger non-structured or more easily modified content so that rapid decisions can be made about access and treatment based on the markings without looking at the syntax of the content itself. For example, classification markings are used in trusted systems to track and separate data and determine accessibility. Markings can also be used in many other similar applications.[8.18]

Syntax checking is used in most low surety mechanisms to examine the content as a way to determine whether it is known to be of some sort or another. The sort detected by the syntax checking is then used to determine whether to pass or hold the information or treat it in some other manner.

Situation checking is used in conjunction with syntax or markings to determine, based on the state of the application, machine, network, or other situational elements, what to do with the content. Typically this involves passing the content, altering it, deleting it, auditing the process, or other similar options.

For high surety content control, it is appropriate to use separation mechanisms such as those identified under access control above.

Behavior

Behavioral mechanisms are used to deal with situations that can be detected by external observation, situations in which behavioral limits can be set regardless of the content or its use, or situations in which controlling behaviors facilitates protection. They include detection of change, times of events, rates of events, fail-safe mechanisms, fault tolerant computing techniques, intrusion or anomaly detection and response, human behavior characterization, detection, and analysis, separation of duties, and least privilege.

Change detection and prevention: Read-only media limits change behavior effectively. As an example, bootable CD-ROMs are used to provide high assurance against changes in the operating environment. When combined with read-only floppy disks for configuration, these systems form medium surety firewalls and similar mechanisms with high surety of regaining original state at reboot. Change detection is useful for many purposes, typically for detecting attempts to alter information as is done in the use of cryptographic checksums in transmission or in storage. Program change detection is critical for assuring integrity of software and is used for virus and malware detection in systems where programs don't change except under proper controls.[7.1.11]

Time and rate controls: Control over times and rates are typical of behavioral detection mechanisms. For example, detection of a worker accessing systems at unusual hours or prevention of tellers from making bank transactions after the bank is closed are common mechanisms used in limiting or detecting behaviors based on time. Rate controls have to do with how much happens in a period of time. Examples include methods to limit the rate of transfer of data between locations, rate limiting by traffic type or application, limitations on the rate of responses or queries to databases, and so forth. [8.19]

Fail-safe systems: Failure modes that can be identified in advance and safe modes for operation during those failures that can be created to work in the presence of those failures are used to create fail safe systems. For example firewalls that disallow all traffic when they fail are common and act as fail safes (for cases where that mode is safe) while detection and response systems that fail by allowing all traffic (which is common for active defenses) are safe for different situations. Most high surety fail safe mechanisms are based on physics. For example, a fail safe water release valve may have a maximum flow rate so that even if a computer tries to release more volume than is allowed, the fail safe mechanism prevents it by physical limitation. Programmable logic controllers are commonly used in manufacturing processes to provide fail safes and similar methods are used in many critical infrastructures.[8.20]

Fault tolerant computing: Fault tolerant computing is designed to be able to detect faults and tolerate them by responding in ways that cover the fault. For example, triple modular redundancy is used in cases where any single failure must be tolerated. Similarly, coding schemes provide for single error correction and double error detection in memory for higher integrity computers, in transmission schemes to recover from noise, and in the detection of behavioral patterns that are out of normal behaviors for applications. [8.20]

Intrusion or anomaly detection and response systems: Detection of most classes of attacks is undecidable while automated response is problematic for all but the simplest situations. Intrusion detection systems detect known intrusion sequences while anomaly detection systems detect changes in behavior that are outside of the normal changes associated with the operation of the system under examination. Response systems can involve anything from logging and notification to automated severing of access. The problem with automated response is that it can be used for reflexive control or similar disruptive effects unless detection is very precise and accurate. The goal should be to detect event sequences with potentially serious negative consequences in time to respond so a to mitigate consequences to within tolerable limits. [8.4.3-4]

Human behavior detection and analysis: Detection and analysis of human behaviors and behavioral changes when using computers is also within the realm of modern technology. Various characteristics of individuals can be identified, characterized, mapped to individuals, and used to detect various conditions. Examples include detection of different people by keystroke and error patterns, detection of command selection and usage in command-based computer use, and detection of normal patterns such as reading email periodically at intervals during the day versus reading it as it arrives. As in the case for intrusion detection, the goal is to detect event sequences with potentially serious negative consequences in time to respond so as to mitigate consequences to within tolerable limits.

Separation of duties: Many behaviors are undesired to the point where they have to be independently verified. A good example is the separation of submission of a transaction from commitment to that transaction, the so-called submit-commit cycle. Another good example is the need for multiple approvals before performing a dangerous operation, such as the requirement that two independent operators turn keys simultaneously in order to launch a missile. Separation of duties is a behavioral constraint mechanism that prevents illicit behaviors from happening even when an authorized insider decides to undertake an inappropriate action of significant magnitude.

Least privilege: The principle of least privilege is based on the notion that users, processes, and other subjects should not have privileges they don't need. As a behavioral control mechanism, least privilege limits behaviors to those that are required in order to carry out the functions that provide the desired utility. For example, process lineage has been used to limit which programs can be run from where so that users can carry out all of their normal activities but any attempt to run a program different from the normal user process behavior is blocked or changed to an appropriate alternative. Similarly, many server programs give up privileges after startup to force any attacker attempting to exploit a flaw to be limited in the privileges they can gain as a result.

Behavioral mechanisms come in different surety levels depending on the specifics of the mechanisms used. There are many behavioral detection and response technologies to choose from, but care must be taken because this particular class of technologies provides widely varying surety depending on the specifics of the implementation and the behaviors being allowed or blocked.

Technical security architecture questions

1. Why can't computers handle “why” very well?
2. Given the many aspects of the human life cycle discussed, what are examples of how each of these elements might become issues to enterprise information protection? Give a specific example of content that could be at issue, the protective requirements, and the consequences of loss of protective function for each aspect of the human life cycle using only examples not in the book.
3. Why can't intrusion detection and response completely replace separation mechanisms in current systems?
4. What are examples of cases where deterrence will almost certainly not work for enterprises?
5. Work flows and inventory seem to be so distant from the notions of technical security, how can they really be part of this issue?

6. Given that people move around so much and computers are no longer static devices in glass houses, what is a better approach than the division of data states into “at rest”, “in motion”, and “in use”?
7. How can we protect data in use through the use of cryptographic methods?
8. Given the seemingly extreme importance of the protection inventory and the claimed lack of any such inventory within real enterprises, what can be done to manage the overall protection program of an enterprise and make sure we don't miss anything?
9. How can perception controls be managed and what are the right set of qualifications for an individual who is to manage those issues?
10. If we cannot detect all known viruses and have periodic misses that result in massive global infections, how can we reasonably expect to detect intentional human attackers trying to circumvent content-based detection and filtering mechanisms?
11. If an insider uses an unauthorized steganography program to encode content that they have access to and send it to an outsider who pays them for the information, is there a technical control scheme that can detect or counter this attack?
12. Given the limits and costs of structural defenses and the limits of behavioral defenses, how should they be traded off against each other to balance protection for different circumstances? Is there a general rule of thumb that would apply to where each should be preferred?
13. Technical protective mechanisms actually touch the content and control it. Isn't there a way to create a completely technical solution that eliminates the need for most or all of the other protection things discussed in this book? If not why not? If so, what is it?

9 Making better protection decisions

Given the complexity of the information protection field and the scope and magnitude of the challenges facing the enterprise of today, decisions about protection seem far too complicated for any human to make correctly, or even to characterize as correct or not.

But management decisions don't have to be perfect in order to be effective, and in most cases there are only limited numbers of reasonable alternatives to choose from at the level of decisions that have to be made for an enterprise. As such, the big decisions that are most important tend to relate to long-term architectural choices and directions, urgent decisions related to emergencies underway, and mid-range decisions that affect operational matters over periods of months to a year or so.

- **Urgent decisions** tend to present few choices and offer high gains or high losses. If high perceived or actual losses are avoided, the decision-maker is often viewed as a savior. If high losses occur, the decision-maker may be vilified and terminated, or it may be seen as inevitable, depending on how the individual manages their situation within the enterprise.
- **Tactical decisions** are the most fraught with danger because they are not so urgent that they can be costly without doing harm and they don't take so long to get through that people fail to remember the origination and process undertaken.
- **Strategic decisions**, if made foolishly and highly disruptive or if they cause major losses, produce terminations. But they can be taken slowly and outside advice may be sought and used to reduce risks of the decisions.

Making better decisions typically involves doing a better job of understanding the issues, identifying viable options, and selecting among the options. Doing better at each of these elements generally depends on avoiding mental (i.e., cognitive) errors, being thorough and thoughtful, and taking the whole situation into consideration.

Common decision processes

Certain processes are commonly applied to decision-making in this context, and they apply to protection related decisions as to others. They are largely codified in the literature on judgment and decision making that started in the shadow of the development of operations research as a field.

As operations research proved its worth starting in World War 2, purely analytical method emerged for optimization of different criteria for select classes of problems when adequate data was available. This led to a wide variety of automated decision-making and optimization systems such as systems for placing loads in lanes and sending multiple shipments to sets of destinations with minimum fuel usage. But despite scores of years of effort in the risk management arena, these methods are rarely of much use in making high-level decisions. Information protection related metrics are only in their infancy, leaving little common basis for statistical analysis. This leaves human judgment and reasoning as the real source of these decisions. And of course these methods are fraught with errors, biased by individual experience, subject to all manner of influence, and generally problematic, especially when viewed in hindsight.[9.1] We will start by focussing on avoiding errors.

Much work has been done on identifying error mechanisms in decision-making, and most of the published results appear to be directed toward understanding the limits and improving the quality of high valued decisions in which substantial resources are expended in the development of options, weighting of factors associated with those options, and producing results. However, the vast majority of common decisions made in businesses, in personal lives, and in other activities that people undertake are not high valued, do not consume much time, and do not justify the high cost of a thorough decision process. Many of the high valued decisions in enterprise information protection also suffer from a lack of time and thus limited process is available. This is where experience and expertise come together, group process becomes vital to success, and where the technical methods identified in the scientific literature are of limited utility.

The basic steps of the decision process include (1) identification of options, (2) identification of factors used to weigh those options, (3) comparison of options in terms of those factors for making the decisions, and (4) presentation of decisions to the decision-maker (DM) and others, also known as justifying decisions. In a business context it may also be helpful to (5) record the process for later review and analysis.

For the DM within a business, it may also be important to note that the total aggregated value of these lower valued decisions may far outweigh the total value of the higher valued decisions to which so much attention has been paid. Nevertheless, executives are most often fired because of failures relating to high-valued decisions.

Identifying options

Identifying options is greatly facilitated by using the content of this book, in the sense that it provides a large volume of approaches and aspects to be considered. Depending on the amount of time available and the value of the decisions, it is sometimes worth going through essentially all of the elements contained herein to identify the different things that can be done to meet a particular protection-related challenge. When there is too little time or when the decision is less important, fewer options are considered. Group processes are also used to generate more options and to engage others in the development of options so as to both improve the alternatives and engage the stakeholders in the decision. This can also reduce individual risk associated with the decision.

Identifying factors

Factors that are used to weigh most protection-related decisions are complex and very often personal to the DM.

A personal interest in a particular approach, a liking relationship with a vendor, the chances of getting fired for taking a non-standard approach, the ability to spread the risk of the decision to others, and the desire to limit liability, are all examples of factors in a decision that are commonly used, highly individualized in application, and not in a standard list of published factors.

The factors associated with analytical approaches identified in the risk management section of the book are certainly important considerations as well. Cost and loss, for example, elements of the duty to protect as defined for the specific enterprise, and ability to operate with existing personnel and methods are important factors that should be considered for any enterprise scale decision. Many more of these can be gleaned by going through the text as well, picking out the factors that seem important to the present decision.

Other factors are highly dependent on the specific kind of decision, the enterprise social and cultural environment, and the technical situation.

The enterprise may mandate the use of specific factors in presentations of decisions to management. For example, a return on investment (ROI) analysis may be required for capital expenditures, even though there is no real ROI justification for a security device in the sense of other investments. Specific factors may be required or specific presentation mandates may be associated with assurance processes. A legal checklist may have to be included and an HR review may be required. All of these and many other factors may be required or desired in order to get a decision supported by management and through the proper channels in order to make an enterprise-wide commitment to a course of action over time.

As a more technical example, the presence of a mainframe as the database platform for an application may make it infeasible to apply internal controls on the mainframe that are available only for personal computers. This may then necessitate the use of mainframe controls that are available, optionally augmenting those controls with off-platform controls such as network-based security devices, controls at an application gateway, or combinations of these controls to provide the desired coverage. Each selection may depend on other aspects of enterprise capabilities, incur added costs. have management and maintenance requirements, etc.

All of the relevant factors should be identified as time permits, and listed or otherwise codified to allow them to be systematically examined. Over time, most people build increasingly large libraries of these factors and tools to apply them more easily to situations as they arise.

Comparing options based on weighing factors

Weighing factors in simple decisions can perhaps be done in your head. But for most enterprise-level protection decisions, there are simply too many factors involved to weigh them mentally and too much potential for making big mistakes or forgetting critical items. People are reasonably good at comparing two factors against each other, but they tend to be far worse as the number of factors increases. Somewhere around 5 to 7 factors, people usually become unable to keep track of things at all.[9.2] They may end up in a frenzy over the decision, incur a lot of stress, and may decide to oversimplify or leave it to someone else. This results in a poor decision unless the enterprise takes a more systematic approach.

For complex comparisons of options, people perform best when they use tools to facilitate their activities. These tools also provide the means to present results of analysis, which plays into the presentation and justification of decisions.

The use of comparisons implies the presence of metrics, and in order to make sense of decisions, these metrics have to be meaningful to the issue at hand. This is where many protection analysts fail miserably. There is a great lack of effective metrics for information protection, and even when metrics are defined, they rarely track well to the factors involved in these decisions. For example, the number of detected viruses blocked by a virus defense mechanism may be readily available, but how does that value relate to the complexity of implementing the new firewall strategy in terms of the reorganization required for implementation across the enterprise? This problem of apples and oranges is not solved by putting everything in terms of monetized units because of the difficulty in getting sound numbers, the high degree of variation in numbers when they are available, and the multi-dimensional nature of decisions and single dimension of monetary units.

It ends up that almost all such comparisons involve a combination of quantitative and qualitative factors that have to be compared to each other and weighed against each other. This is where all of the mathematical models in the world cannot and do not replace a decision maker putting values on factors and varying those values to suit the decision that makes the most sense to them.

Presentation and justification of decisions

In enterprises as elsewhere, decisions must be presented to others and justified. Depending on the nature of the enterprise, social norms may dictate factors that must be presented and the structure of the presentation. But these issues don't alter the fact that compelling presentation and proper justification make the difference between getting better decisions for the enterprise and getting worse ones and keeping the presenter's job or losing it.

The literature on judgment and decision making asserts that decisions are often made well before the formal decision process ends. This is true in my experience. Once a decision is made, it may still be subject to change, but the effort tends to move from making the decision to justifying it at that point.

There are psychological factors involved in making effective presentations of decisions and making those justifications work, and there is a lot of related literature on this subject.[9.1] The key elements that are best studied and well known are those presented in the area of negotiations, where the world has learned from the early research.[9.1.3] Presentation can typically be made more effective by the ordering of factors and related techniques, while avoiding undue influence involves eliminating the effects of those same influence factors.

As the number of factors increase and the complexity of the decision grows, graphical aids can be useful in presentations, and the appearance of a mathematical model or other similar approach often helps to create the necessary flourish to make the listener believe in the result and reduce questioning.

A thorough job for a complex strategic decision should involve the inventory analysis indicating the implications of the decision across the enterprise and including the magnitude of the changes to be made. The business model should be used to test out the proposed alternatives that seem most viable and to show that interdependencies will not be create risk aggregations in excess of mandates. Work flow process should be used to characterize the time and effort involved in the implementation. And generally, all of the aspects of the protection program should be considered. This may result in hundreds of pages of documentation.

Recording decision processes and results

In the enterprise environment, decisions of import generally have to be documented, and in the process, protection-related decisions should be integrated into the enterprise documentation and change processes. For example, for a strategic decision, there is likely a study performed to look at the requirements, current state, gaps, and likely future state. This study would become part of the permanent documentation of the effort and provide the background for later analysis of failures and successes as well as providing content used in the presentation and justification process.

After the decision is made, the process of carrying out the decision is typically tracked through a project management system with project managers involved in keeping things on track and within schedule and budget requirements. The project management approach typically provides a lot of added documentation and addresses decisions at finer levels of detail along the path to implementing the strategic vision associated with a high-valued decision.

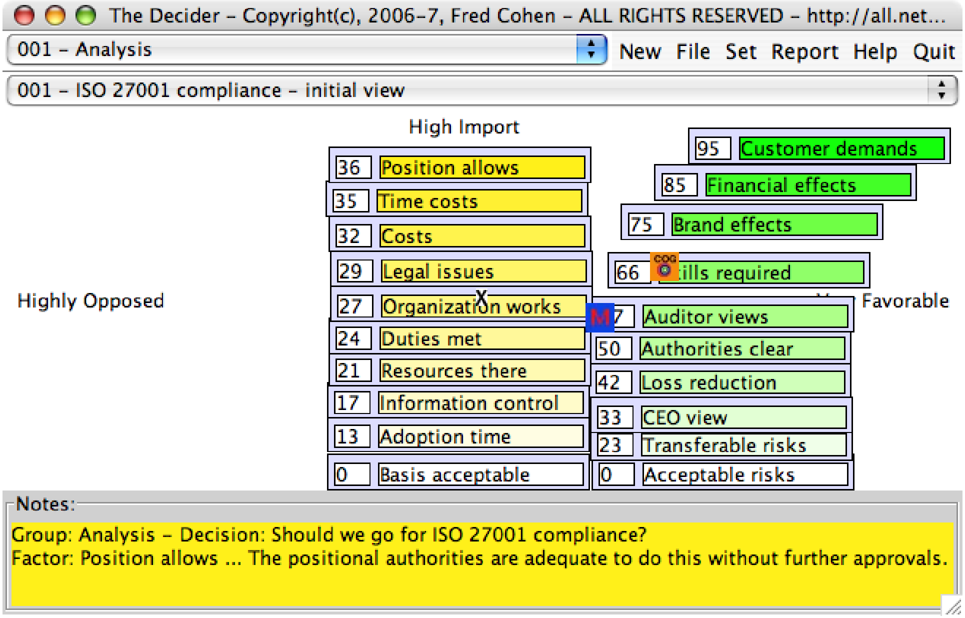
Tools to support decision-making

Because decision-making can become complex to the point where people perform poorly on their own, tools to support decision-making become necessary. In the security space, there are specific sets of tools that are used for real-time analysis and decisions that are pre-made in the sense of being programmed. These are typically embedded as expert system mechanisms (usually some variation on production systems[9.3]) within products. But those do not support nor are they helpful in the sorts of management decisions discussed here. Human decision support typically comes in the form of processes and mechanisms.

Processes for decision support usually help to identify more alternatives and factors and weigh factors. These include various human meeting methods in which people get together to brainstorm in a structured process, strategic scenario simulations in which specific aspects of the space are explored in more depth to both educate the audience and gather information from them, and other similar approaches. These approaches can be facilitated with technologies, typically including pieces of paper posted on walls throughout meeting rooms, and informational tools that provide inputs such as votes, displays of lists, and depictions of the decision space.

The most common tools in use today involve spreadsheet programs that encode factors and options in rows and columns, and perhaps take votes from individuals to produce weights for each pair of factor and option. The weights are summed and alternatives sorted based on the summed (or averaged) weights.

An advancement on that approach uses a two-dimensional space to characterize factors relative to each other in importance and favorability of the option. Figure 9-1 gives and example of a decision support tool called Decider applying this approach to a decision by an enterprise to move toward ISO27001 compliance.



The Decider – Copyright(c), 2006-7, Fred Cohen – ALL RIGHTS RESERVED – http://all.net...

001 – Analysis

001 – ISO 27001 compliance – initial view

Notes:

Group: Analysis – Decision: Should we go for ISO 27001 compliance?
Factor: Position allows ... The positional authorities are adequate to do this without further approvals.

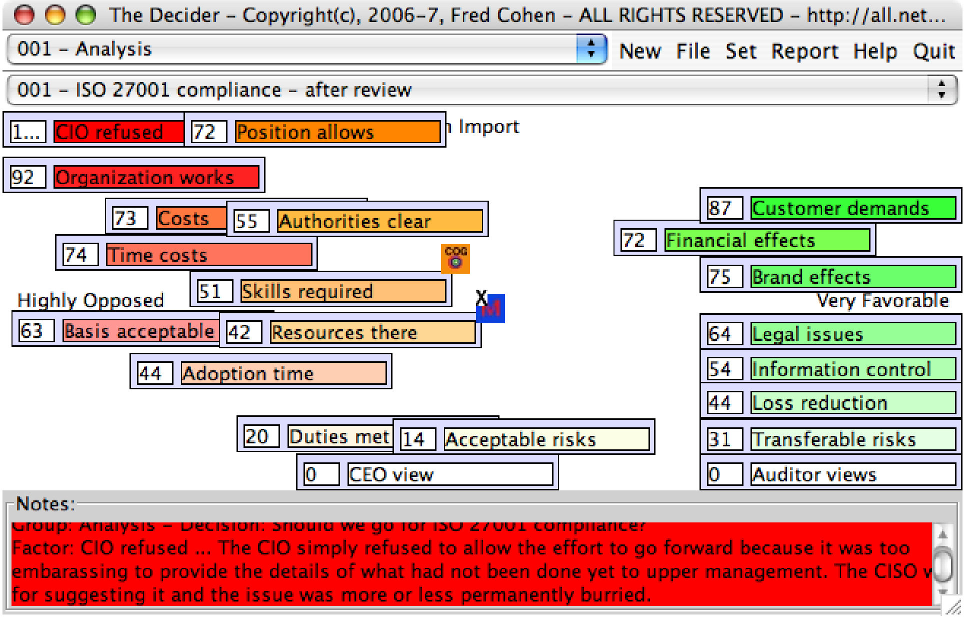
Figure 9-1: An initial review of a decision to support ISO 27001 compliance

This tool in this case indicates strong support for this decision and brings 20 different factors to bear, all presented on one screen and movable with respect to each other. It also allows details of the basis for each of these factors to be encoded so that a report can be generated that supports the decision.

The reporting process in this tool orders outputs for favorable or unfavorable presentation by applying the psychological principles discussed earlier, and thus provides an initial outline that can be used to provide a more formal report.

This tool is also useful for eliminating many of the cognitive errors associated with decision making such as ordering of presentation given as input to the use of the tool and reduction of the impact of loaded terms.

But the use of such tools and reduction in cognitive errors does not necessarily lead to good decisions. People still anticipate things differently than they see them after experience. Figure 9-2 shows the same analysis after internal review of the initial decision, some months later in the process.



Group: Analysis - Decision: Should we go for ISO 27001 compliance?
Factor: CIO refused ... The CIO simply refused to allow the effort to go forward because it was too embarassing to provide the details of what had not been done yet to upper management. The CISO v
for suggesting it and the issue was more or less permanently burried.

Figure 9-2: An review of the same decision after more experience

The same factors are viewed very differently once they are put to the test of time, and this example shows pretty clearly that things can change. This is a good reason for a process that takes time and does some test runs before going into full scale production.[9.4]

So-called best practices – don't buy it

The term “best practice” is often bandied about in the information protection arena as if there were some set of standard practices that could solve the problems faced in the field without bothering to think through the issues. This is foolishness, as I hope the rest of this book has helped to point out. You cannot codify the rich set of issues of information protection into a simple list of things to do or a set of practices that are “best” at a generic level.

> ***WARNING: USE OF THE TERM “BEST PRACTICES” WITH REGARD TO INFORMATION PROTECTION SHOULD BE VIEWED WITH EXTREME SKEPTICISM!***

When someone comes out with an assertion about using best practices, it may be best to ask them to identify the documentation that indicates this practice as “best”. In situation after situation, the expression is used without any basis whatsoever. To drive the point home, and assuming aggressive interaction is desirable, you might ask them if this is in ISO27001 or 27002, if it is from CoBit, and list other standards asking if it is from one of those standards. If it isn't from a standard, perhaps it is from a publication of some sort. If so, you should be anxious to find it so you can review it for details and make sure it applies to the present situation. Chances are very good that there is no basis whatsoever for the claim of best practice, or if there is, it is from a vendor brochure or some other similar publication.

One of the best ways to tell the difference between someone who knows almost nothing about information protection and someone who knows something may be from their response to queries in this area. Anyone who asserts that they use best practices is almost certainly not doing so and doesn't know what they are.

I have been quoted as saying something to the effect that there is only one “best practice” in information protection, and that is to take a systematic comprehensive approach. That's another way of saying that the best practice is to do everything discussed in this book diligently and professionally and spend your time and effort understanding these issues.

Some sample sound practices with a basis

What can be done, with time and effort, and in the context of the situations of the day, is to create a set of reasoned security decisions and decision criteria that are sound. Soundness, in this context, means that the decisions are sensible, have a solid basis in what we currently know, and can be reasonably applied with proper thought and consideration. These practices tend to change over time because of the change in the environment. They can be thought of as sound only in the context of the basis upon which the decisions are made.

Because times change, the sound security decisions of today differ from those of a few years ago and will vary from those of a few years from now. As a result, there is a time dependency involved in protection-related decisions. Nevertheless, in an effort to be helpful for today and to provide some sense of how to develop decision criteria for the future, select security decisions are presented here as examples of what can reasonably be codified.[1.1.4]

The decisions presented here are in a standardized format and approach. They start with a title, followed by a question, provide a set of options, identify a decision algorithm, and indicate a basis for decision-making. The goal of applying them is that the decision-maker can read and understand the issue, make a reasonable decision quickly, identify the basis for the decision, and move on. In their use within an automated tool, this allows decisions to be made very quickly and supported with documentation. Reviews can be undertaken and changes made as appropriate.[9.5]

Avoiding radius driven common mode failures

Title: Backup facility distance: How far should I go?

Question: How far away do I have to put redundant data centers and people to assure reasonable business continuity?

Options: (1) No distance requirement. (2) At least 5 miles away. (3) In another city. (4) At least 250 miles away. (5) On another continent.

Decision: For enterprises with no requirement for a separate backup facility, Option 1 is fine. Option 2 is used for highly localized enterprises wishing to only protect against facility failures, riots, explosions, calamities on local highways, and similar conditions. Option 3 is used for regional enterprises not trying to continue operations in the face of large-scale events like hurricanes. Option 4 is appropriate for any large enterprise not limited to a region and is suitable for most natural disasters and most single event human activities. Option 5 is for multinationals that are global in nature and need to survive national level governmental changes, severed international communications, and wars and insurrections. Options 4 and 5 are also helpful in assuring high performance because global communications takes time and tends to be less reliable than local communications.

Basis: Distance requirements for backup facilities are driven entirely by scenarios that can lead to simultaneous loss. These in turn are threat driven. If an enterprise has to be able to survive limited nuclear attack, government insurrections, regional and world wars, and comet strikes that don't end all human life, transcontinental diversity is a necessity. There are many global multinationals that have this requirement and the efficiency gained by reducing this sort of redundancy does not justify the collapse of the business under conditions that, in the case of wars and governmental changes, happen quite frequently.

Many enterprises operate within only one continent or country, or substantially do so in terms of their need for information technologies associated with business continuity. In these cases, distance becomes an issue when regional events take place. In the US, for example, there are hurricanes that effect one region, earthquakes in another region, floods and storms in another region, winter related infrastructure failures in another region, and so forth. If the enterprise exists only within a certain region, then it doesn't particularly help to have redundant data centers elsewhere because they will be of little use when the rest of the business does not operate. The exception is the large enterprise that is dominant in a region but has enough diversity in operations to continue even when that region is hit by a regional event. Enterprises that cross regions usually have substantial facilities in more than one region, and these are often, but not always, ideal locations for redundant data centers.

In some cases these data centers may not be completely redundant but rather may have a primary function and a backup and recovery capability. Another common approach is to use one location for research and development and the other for operations. Since the same sorts of systems are used in each and separation of duties are required for medium and high risk elements of information technology operations, it is relatively low cost to physically separate the functions, use redundant hardware in the facilities, and use the redundant operational expertise of research, development, and testing in one location for operations if the primary operational center fails. If the secondary fails, research and development may collapse but operations continue.

For strictly regional companies, city to city redundancy is called for unless the enterprise is single facility or very limited in geographical scope. By spreading across cities, local government failures, telecommunications outages, common energy and infrastructure failures, most riots and fires, many tornadoes, hurricanes, earth movements, most floods, and many other similar events can be survived. Care must be taken to assure that commonalities are avoided in each of the interdependent areas so that business continuity is assured. For example, placing all data centers along coast lines could be subject to common mode failures associated with increases in sea level, for example, from tsunamis. Cost differences are typically very small for assuring proper infrastructure redundancy, but expertise is needed to verify this redundancy.

For very small localized businesses, if the owner survives, a set of backups that can be recovered within days to weeks and the ability to purchase new hardware is adequate. Redundant data centers are not a requirement at all. To the extent that there is redundancy, most protection is against single facility failures such as the main place of business burning down or some service outage. Redundancy may be limited to a backup tape taken home with the owner at night or every week.

How many redundant data centers do I need?

Title: Data center redundancy: How many do I need?

Question: I understand that I need redundancy to protect business continuity and provide for disaster recovery, but how much redundancy do I really need?

Options: (1) A single data center well protected from all identified threats. (2) Two data centers, a primary and a backup. (3) More than two data centers distributed across the regions where the business functions.

Decision: For most small businesses, a single data center, if there is a data center at all, is reasonable and prudent, so option 1 is a good choice. For a medium sized business with only one facility where all value of the enterprise resides, option 1 is also viable because any event large enough to cause serious damage to the data center if it is well protected within the facility is likely to also impact all of the other elements of the business. Option 2 is good for distributed medium sized businesses or large enterprises that don't have a high threat profile, are not highly dependent on information technology, and are not highly geographically diverse. Option 3 is appropriate for any large enterprise that is geographically distributed or a medium sized business with high dependence on information technology and geographic diversity.

Basis: For small businesses, the cost of redundant data centers is fairly high and there is rarely data so critical to operations that multiple data centers are justified. A better strategy is often to have backups retained off-site, perhaps in a bank safe deposit box or fireproof media rated safe. If and when disaster strikes, the backups can be used for recovery without high losses to the business. The cost is low, the consequences are relatively low, and the resources are spent if and when recovery is needed. Of course the backup and recovery process must be tested periodically to assure that it will in fact function. Similarly, a well protected data center inside a single-facility medium-sized business is adequate because there is no reason to protect the data center more than the rest of the business it supports.

For medium scale businesses that are geographically diverse or highly dependent on information technology and large enterprises that are not geographically diverse, a primary and secondary data center are appropriate in order to assure continuity of operations across facility-related failures without long delays in recovery. Backups are mandatory, but these backups should be reflected in a timely fashion in the backup facility so that recovery and continuity of operations is assured at all times and within time frames that prevent serious negative consequences to the information utility of the company. The backup site should also be populated with adequate personnel to continue operations if the primary fails catastrophically and people cannot be transported to it. Putting all of the enterprise eggs in one basked or allowing single points of failure is irresponsible.

For any large enterprise that is geographically distributed or medium scale enterprise with high dependency on information technology and geographic distribution, geographically distributed data centers in major regions of operation should be in place to support critical business functions while also affording higher performance for the local area and retaining appropriate expertise in multiple facilities to continue business operations even if regional disasters or government failures take place. The larger and more distributed the company, the more opportunities there are for geographic distribution and redundancy. Not all data centers must have copies of all content. Rather, distribution of content over data centers and levels of redundancy should be determined by utility of local versions of information combined with business impacts of failures. As in the two data center case, recovery times are important to understanding the design of the redundancy. Infrastructure, and other dependencies should be considered, and personnel redundancy is critical.

In all cases, backups facilities, backups, and backup and recovery processes should be tested and verified periodically. This is typically done at least once per year as part of business continuity planning efforts. Backups should be verified as they are taken, so that loss of backup data because of media failures should never be at issue for the short run. Redundancy is a complex subject and the exact number of redundant data centers is highly dependent on the criticality of information. Many financial institutions have five or more data centers each capable of running all financial transactions. Many companies are sufficiently diverse that they have limited redundancy for individual systems, but many facilities with data centers holding different capabilities, so that only a small fraction of the business fails if a data center is lost. Less diverse businesses and businesses undergoing data center consolidation for cost savings sometimes end up with inadequate redundancy. Some large enterprises have had complete business failure because of a single point of failure in a business-critical system.

Decision-making questions

1. Why are protection-related decisions any different than other enterprise decisions and in what ways are they the same and different?
2. How do you identify most relevant options?
3. How do you identify most relevant factors?
4. How do you solve the “apples and oranges” problem in comparing different factors and options in these decisions?
5. Is it a bad idea or a good idea to use psychological principles to try to present protection decisions you favor in the most favorable light and those that others favor over yours in a less favorable light?
6. Why do protection-related decisions need to be recorded?
7. Make a sample spreadsheet to support a decision about a firewall selection. Include all of the relevant elements from this book.
8. Why is the term “best practices” a poor choice for use in the context of information protection and what are ten good reasons to avoid the use of the term?
9. What are the problems not addressed by the sample decisions? What factors are missing? What options are missing? How could these decisions be made better?
10. Given that there may be very little time to make a critical decision about real-time response to an incident, what decision-making processes should be used for this purpose and what should be planned out in advance?

10 Summary and conclusions

Content and its business utility drive the need for protection, but protection can also drive out some of the business utility. The goal of effective protection is to assure business utility by efficiently and effectively protecting that utility in order to facilitate proper business operations. This is done by the protective mechanisms put in place and the processes that control them.

Protective mechanisms contact the content and the mechanisms that use that content as well as the threats to that content and its utility. These mechanisms include perception-based, structure-based, content-based, and behavior-based mechanisms that are controlled by protection processes.

Protection processes include the inventory, work flow mechanisms, attack and defense processes which deter, prevent, detect, react, and adapt to threats, data state controls that deal with information at rest, in motion, and in use differently, and interact with elements of the technical security architecture.

The technical security architecture deals with issues of life cycles, particularly in the area of systems and data life cycles, and context in terms of time, location, purpose, behavior, identity, and method of use within the context of the overall control architecture.

The control architecture typically consists of requirements for integrity, availability, confidentiality, use control, and accountability that are fulfilled by sets of access controls, functional units, perimeters, access methods, trusts, and change controls. Change controls limit what changes can be made based on the surety requirements of the content and its utility, which in turn is driven by the risks associated with that content and utility. Access methods provide the means by which identified subjects are authenticated and authorized to use content and gain its utility. Perimeters form the separation mechanisms between zones that allow grouping of surety requirements and risks into pools of manageable size and similar control requirements. Functional units are the mechanisms that implement the architectural concepts and the business utility of the content. Access control schema associate clearance requirements of subjects with classifications of content and its utility in conjunction with risk aggregation requirements, perimeters, and access methods to model the control of access. Trust is the amount of harm allowed from an identified source. Protection objectives stem from business needs and risk-related desired surety levels.

In execution, organizations have to do everything that is done to operate their systems and protect their content and its utility. This involves management, policies, standards, procedures, documentation, auditing, testing, technologies, personnel, incident handling, legal, physical security, knowledge, awareness, and organizational aspects and processes. These processes are influenced by the CISO and the results of these influences are measured and sensed by the CISO using a combination of technical and non-technical means within the overall governance architecture of the enterprise in order for the CISO to carry out their duties. These processes are also critical parts of the life cycles of people and businesses within the enterprise and these aspects must be managed as well.

Decisions must be made in order for enterprises to carry out their security responsibility. These decisions are often too complex and important to be done by an individual or within a purely mental framework. Groups processes and tools are needed in order to make and codify decisions. Notions like “best practices” are fantasies that must not be embraced if the enterprise is to prosper. Rather, thoughtful people must diligently peruse in-depth understanding of the complex issues in information protection in order for the enterprise to assure the utility of content in a reasonable manner and consistent with its specific needs.

The CISO is guided by and contributes to the risk management process that turns duty to protect into what to protect and how well through the evaluation of risks and matching of surety to risks. Oversight defines these duties to protect by combining legal and regulatory requirements with contractual requirements and self-imposed requirements. This is ultimately driven by business requirements that support the people and things and the business processes that cause the business to operate and prosper.

11 Detailed table of contents / outline

12 Endnotes

Chapter 1

Background and Introduction

1.1 These include:

1.1.1 F. Cohen, “The CISO ToolKit: Governance Guidebook”, ASP Press, 2005, ISBN# 1-878109-34-0

1.1.2 F. Cohen, “The CISO ToolKit: Security Metrics”, ASP Press, 2005, ISBN# 1-878109-35-9

1.1.3 F. Cohen, “The CISO ToolKit: Governance Checklist”, ASP Press, 2005, ISBN# 1-878109-37-5

1.1.4 F. Cohen, “The CISO ToolKit: Security Decisions”, ASP Press, 2005, ISBN# 1-878109-38-3

1.1.5 F. Cohen, “The CISO ToolKit: Information Security Awareness Basics”, ASP Press, 2007, ISBN# 1-878109-39-1

1.2 Sir Isaac Newton is the one responsible for the original thought regarding giants. I believe the actual quote is: “*If I have seen farther, it is by standing on the shoulders of giants.*” Unfortunately, throughout the text, the reader is likely to encounter many appropriations of others' thoughts and ideas that are not cited. While an effort has been made to cite relevant papers and other works where feasible, there are indeed many who have contributed greatly to the field that are not cited in this book and many original ideas that are only cited through other authors whose papers have been included. While eternal improvement is the goal, books eventually have to get published and the author's capacity to review everything ever written is increasingly taxed by the quantity of the task and the available time.

1.3 For details of the author's writings, see http://all.net/. This site is used as a resource throughout this book, particularly for papers that are old and perhaps hard to find, for dynamic content, and so that another 500 pages aren't required to cover the subject matter.

Chapter 2

Enterprise Information Protection

2.1. The term “utility” and the phrase “assure the utility of content” are, perhaps the most complicated to understand in the field of information protection and remain, even today, a part of an ongoing debate within and without the field. In my view, assuring the utility of content is the purpose of information protection. Some may say that the effort should be limited to confidentiality, integrity, and availability (CIA), while others identify only privacy (or confidentiality) as the key issues. Many argue that availability and integrity are properties of quality of service and don't belong in the “security” arena. I, respectfully when I can muster it, disagree. From a standpoint of a business, there is a dire need for assuring (making certain to within a desired degree of certainty) that the content (the meaningful – useful thing that we associate with information) provides the desired utility (useful purposes) for the business. If utility is reduced by leaks of the information to unauthorized people, then confidentiality is part of the task. If utility is decreased by the represented content not being available for use, then that falls under the information protection arena. You get the idea.

2.2. This particular characterization is not the most commonly used one. In fact, I think it is, for the moment, unique to my approach. This “control architecture” in most cases does not exist in any codified form outside of the notions within peoples' minds. But it really should be formalized. The IACUA elements (an expansion on confidentiality, integrity, availability – CIA) – are ordered by typical import and include elements of use control and accountability that are relatively new concepts to most of the practitioners in the field. Without getting into a debate over it, it is my view that including these additional elements is helpful from a practical standpoint in dealing with modern issues, regardless of how other models might well handle them. The notion that access control models are necessary is subtle in that most people in the field implicitly assume a well known model based on the Bell-Lapadula model [D.E. Bell and L. J. LaPadula Secure Computer Systems: Mathematical Foundations and Model. The Mitre Corporation, 1973 *This was the classic paper in which Bell and LaPadula described a model for maintaining secrecy in timesharing computer systems*]. But this is not the only model available, and indeed, it is not really the model in widespread use today. The models in widespread use today have not really been codified or analyzed to date, except that they typically fit into a transitive information flow model similar to the Poset [F. Cohen, ``Protection and Administration of Information Networks with Partial Orderings'', IFIP-TC11, ``Computers and Security'', V6#2 (April 1987) pp 118-128.] or Lattice [Denning, D. E. Secure Information Flow in Computer Systems, Ph.D. dissertation, Purdue Univ., West Lafayette, Ind., May 1975.] models. These models are not very helpful in the current situation because they indicate, perhaps rightly, that current protection is largely ineffective and will remain so. It would be useful to have other models, perhaps with statistical or other bases, that are not as definitive in their partitioning of the space but still useful for description and analysis. The functional units come in many forms, from firewalls to software elements within applications, to cryptographic transforms in which the content is enveloped for transit and storage. Perimeters, while touted by many as disappearing, remain present in different forms and at finer and finer granularity, leading to greater complexity and management difficulty. Access mechanisms have been around since before the common era, and change management has been a critical element of information protection for a long time. Again, these are typically not conceptualized as part of the protection model but they are critical to protection and I think they are vital to understanding what is going on and seeking new solutions and viable alternatives when faced with real problems encountered every day in enterprises.

2.3. In summary, content and business utility are protected by mechanisms, processes, and architectures that are structured through the control architecture and managed via influence on organizational elements by the CISO. The CISO acts to meet the duties to protect by determining how to protect the things that need to be protected and controlling the organization so as to affect those protections. The risk management process and feedback mechanisms guide the CISO and act as the means by which oversight is accomplished with the ultimate objective of assuring that business processes are not interfered with in ways that cause serious negative consequences. The CISO reports to top management and oversight individuals and groups that have ultimate authority over and responsibility for the business. The effective CISO provides the oversight function with adequate accurate information to make reasonable and prudent decisions and carries out those decisions once they are made.

2.4. This is really a key point that many in top management miss and many in information protection fail to point out to them. As an example, a CIO at a company I was working for at one point got fired because she took it upon herself to accept a risk that could have put the entire enterprise out of business. She did so without consulting the CEO or anyone else on the top management team that she was a part of. Whether she knew this and did it anyway or didn't know it was irrelevant. She should have known it. I have seen this again and again, especially in data center consolidations and other similar cost cutting efforts. When too much redundancy is squeezed out of a system, the result is inadequate assurance and the consequence is disaster. But when a decision-maker takes on risks that are above the level of risks they are authorized to take on, even if there is no disaster, when found out they are invariably punished.

2.5. A governance structure is typically a set of definitions, rules, practices, and processes that facilitate governing. In corporations, the government is typically, but not always, designed to allow those in charge to have their way while providing those who are governed (and paid to work there) with the flexibility and benefits that motivate them to help the enterprise succeed. This may be very different for different sorts of workers in different companies. For example, when labor is non-creative and in high supply relative to the demand, management may have a governance process that is strictly top-down and authoritarian. But in a high technology high growth business where workers are highly educated and the environment is very competitive, it may seem more like a cooperative.

2.6. The information protection posture assessment process was first given in:

> 2.6.1 F. Cohen, “Protection and Security on the Information Superhighway”, Wiley and Sons, 1995, with examplkes from previous assessments at: http://all.net/books/superhighway/index.html

2.7 James A. Schweitzer, “*Protecting Business Information – A Manager's Guide*”, Butterworth, 1996, ISBN#0-7506-9658-3.

2.8 Trust models are the subject of a lot of study, but no definitive understanding has yet been gained. See also endnote O.3.

Chapter 3

How the business works and is modeled

3.1. A presentation on business modeling related to the content here is available at http://all.net/RM/BusModels.pdf It includes more detailed slides than are presented here. It was originally presented in June of 2006.

3.2. The story of the blind men and the elephant is one I commonly use to discuss how information protection gets viewed within enterprises. It is from a Buddhist cannon (Udana 68-69): A number of disciples went to the Buddha and said, "Sir, there are living here in Savatthi many wandering hermits and scholars who indulge in constant dispute, some saying that the world is infinite and eternal and others that it is finite and not eternal, some saying that the soul dies with the body and others that it lives on forever, and so forth. What, Sir, would you say concerning them?" The Buddha answered, "Once upon a time there was a certain raja who called to his servant and said, 'Come, good fellow, go and gather together in one place all the men of Savatthi who were born blind... and show them an elephant.' 'Very good, sire,' replied the servant, and he did as he was told. He said to the blind men assembled there, 'Here is an elephant,' and to one man he presented the head of the elephant, to another its ears, to another a tusk, to another the trunk, the foot, back, tail, and tuft of the tail, saying to each one that that was the elephant. "When the blind men had felt the elephant, the raja went to each of them and said to each, 'Well, blind man, have you seen the elephant? Tell me, what sort of thing is an elephant?' "Thereupon the men who were presented with the head answered, 'Sire, an elephant is like a pot.' And the men who had observed the ear replied, 'An elephant is like a winnowing basket.' Those who had been presented with a tusk said it was a ploughshare. Those who knew only the trunk said it was a plough; others said the body was a grainery; the foot, a pillar; the back, a mortar; the tail, a pestle, the tuft of the tail, a brush. "Then they began to quarrel, shouting, 'Yes it is!' 'No, it is not!' 'An elephant is not that!' 'Yes, it's like that!' and so on, till they came to blows over the matter. "Brethren, the raja was delighted with the scene. "Just so are these preachers and scholars holding various views blind and unseeing.... In their ignorance they are by nature quarrelsome, wrangling, and disputatious, each maintaining reality is thus and thus." Then the Exalted One rendered this meaning by uttering this verse of uplift

> O how they cling and wrangle, some who claim
> For preacher and monk the honored name!
> For, quarreling, each to his view they cling.
> Such folk see only one side of a thing.

Chapter 4

Oversight

4.1. A detailed description of these issues is contained in *The Sedona Guidelines: Best Practice Guidelines & Commentary for Managing Information &*

Records in the Electronic Age, A Project of The Sedona Conference Working Group on Best Practices for Electronic Document Retention & Production, September 2004 Public Comment Draft.

4.2 Many standards have sections related to disaster recovery and business continuity planning for information technology. For example, ISO 27001 and 27002 both have business continuity planning and disaster recovery planning elements. There are professional societies that specialize in these areas and many small businesses that support process development. There is also a large industry providing backup and recovery of information and information technology including a wide range of facilities with different protection profiles.

4.3 Intellectual property law is another substantial specialty area with patents being the largest area of legal specialty. Patents are universally understood but recognized country by country (or group by group for some groups of countries), so in practice patents are very expensive and problematic for global enterprises. Copyrights are sometimes hard to enforce across borders and international treaties apply to them more than other areas. However, there is an enormous trade in illegal copies of copyrighted material worldwide. Trade secrets are far more complex but are increasingly being unified both within countries and across the globe. However; many countries do not respect intellectual property laws associated with other countries, particularly in the trade secret arena, and this creates conditions for large-scale industrial espionage, which is increasing worldwide and has been for some time.

4.4 Customer content, in general, has all of the potential issues that any other sort of content has, plus all of the contractual obligations associated with the responsibilities taken on contractually when taking on the content. In practice, this is an area that is very complex and many companies fail to properly deal with it. It is the reason for things like SAS-70 audits and peering agreements as well as the “safe harbor” provisions in laws and treaties.

4.5 Limited coverage of this topic is commonly available and it is widely ignored in the information protection arena compared to other professional arenas. Recently several of the major professional societies in the security arena have gotten together to try to work out a common code of ethics that follows along with the codes of the IEEE, ACM, and other professional societies. For more details on this see: http://all.net/Analyst/2007-04.pdf and a subsequent article in the same forum at http://all.net/Analyst/2007-08.pdf followed by an interesting event discussed in http://all.net/Analyst/2007-12.pdf to see how things change over time and how effective social action can be in this arena.

4.6 In one consulting job I worked on the status of production was considered to be very confidential and critical to the competitive edge of the business. And in my own business experience, I have encountered occasions when production problems limited output at critical times when release of information about these problems could have destroyed substantial portions of the business. In the power industry, for example, where there is a spot market that operates in near real-time, companies have used published information on the status of the power grid to intentionally (and illegally) take assets offline, thus increasing the use of other assets that were sold at higher margins. In the power crises near the beginning of the 21st century, this sort of tactic was used by many companies, as it is in all markets. So called “gaming the system” is often exploitive of such operational status information.

Chapter 5

Risk Management

5.1. See F. Cohen, "Managing Network Security" (series of articles in Network Security Magazine) - Risk Management or Risk Analysis?", Network Security, Mar., 1997 for details on the limitations of probabilistic risk analysis, and other articles in that series for some alternative approaches to managing risks. There are many risk assessment methodologies. The most popular among them is probabilistic risk assessment (PRA). Unfortunately, this is almost always a poor choice in the information protection field because PRA assumes independence of events, that all events are characterizable, and that they are characterizable in terms of random stochastic processes that do not change substantially over time. When malicious human attackers attack enterprise information systems, they don't follow these rules, so these sorts of assessments tend to be substantially wrong. Furthermore, when planning defenses, due diligence, regulatory and contractual requirements, and many other factors come into play that are beyond the capacity for PRA to handle in helping to assess defenses. PRA assumes a prevention model as opposed to fusing together deterrence, prevention, detection, reaction, and adaptation. PRA also has no mechanism for dealing with issues of time or high levels of uncertainty. The result is an approach that is useful for certain simplistic situations but not very useful for governance.

5.2. See F. Cohen, "Simulating Cyber Attacks, Defenses, and Consequences", IFIP-TC11, `Computers and Security', 1999, vol. 18, no. 6, pp. 479-518(40). An alternative model is “FAIR”. It uses motive, primary intent, sponsorship, preferred general target characteristics, preferred specific target characteristics, preferred targets, capability, personal risk tolerance, and concern for collateral damage as evaluation criteria for threats, and arranges them into categories: {Internal (employees, contractors, partners), External (criminals, spies, non-professional hackers, activists, nation-state intelligence services, and malware authors)}.

5.3. See F. Cohen, "Simulating Cyber Attacks, Defenses, and Consequences", IFIP-TC11, `Computers and Security', 1999, vol. 18, no. 6, pp. 479-518(40) as above. An outstanding reference in this area for physical security is

> 5.3.1 Mary Lynn Garcia, “The Design and Evaluation of Physical Protection Systems”, Elsevier, 2001, ISBN 0-7506-7367-2

5.4. Other risk assessment approaches exist and should be considered, however, the most useful approach in most enterprises does not produce the sorts of results that most risk assessments in other fields are likely to produce.

Rather, the end result is a series of ranked consequences with associated and validated threats. After consequences are put in a ranked list, threats are associated with those consequences, and vulnerabilities validated to determine if identified threats could exploit identified vulnerabilities to induce those consequences. Once this is done, a set of validated risks are available. That is as far as risk assessment can really go.

5.5. Undecidability issues are rampant in information protection. Starting with the problems associated with computer virus detection, proofs and similar demonstrations have been made for the undecidability of all sorts of detection and the high complexity of detection methods for finite environments. This drives up the cost and ultimately makes detection of known bad problematic. Similarly, anomaly detection, which is detecting deviation from known good suffers from the same sorts of challenges. For more details see:

> F. Cohen, ``Computer Viruses - Theory and Experiments'', DOD/NBS 7th Conference on Computer Security, originally appearing in IFIP-sec 84, also appearing as invited paper in IFIP-TC11, ``Computers and Security'', V6#1 (Jan. 1987), pp 22-35 and other publications in several languages.
>
> F. Cohen, ``Computational Aspects of Computer Viruses'', IFIP-TC11, ``Computers and Security'', V8 pp325-344, 1989.
>
> F. Cohen, “Computer Viruses”, Dissertation at University of Southern California, 1986.
>
> B. W. Lampson, A Note on the Confinement Problem, CACM 16(1), October, 1973:613-615

5.6. Cryptography has been used for a long time for information protection and there are many journals covering this area. Cryptographic checksums and related areas are typically applied. For further details see:

> F. Cohen, ``A Complexity Based Integrity Maintenance Mechanism'', Conference on Information Sciences and Systems, Princeton University, March 1986.
>
> F. Cohen, ``A Cryptographic Checksum for Integrity Protection'', IFIP-TC11 ``Computers and Security'', V6#6 (Dec. 1987), pp 505-810.
>
> F. Cohen, ``Models of Practical Defenses Against Computer Viruses'', IFIP-TC11, ``Computers and Security'', V7#6, December, 1988.
>
> Look at the more recent efforts of the “Trusted Computing Group” which has started to implement these and related mechanisms in hardware.

5.7. Finite state machines are ultimately the goal of high surety designers with the machines based on well understood physical mechanisms and provable analysis methods that demonstrate that the totality of states and transitions under identified environmental, state, and input conditions all meet the design criteria. There is extensive literature on this associated with the design of digital circuits and systems over the last 50 years or more. Flow controls for separation have also been widely explored ranging from the Bell LaPadula model to partialy ordered sets. For more details see:

> D.. E. Bell and L. J. LaPadula Secure Computer Systems: Mathematical Foundations and Model. The Mitre Corporation, 1973
>
> Denning, D. E. Secure Information Flow in Computer Systems, Ph.D. dissertation, Purdue Univ., West Lafayette, Ind., May 1975.
>
> F. Cohen, ``Protection and Administration of Information Networks with Partial Orderings'', IFIP-TC11, ``Computers and Security'', V6#2 (April 1987) pp 118-128.

5.8. Originally stated in F. Cohen, Introductory Information Protection, 1987-9, available at http://all.net/books/IP/index.html (Introduction)

5.9. Typical approaches include (1) strategic games used to (a) develop long-term plans, make complex decisions, and enlighten mixes of participants, (2) automatic games used to (a) analyze specific situations and (b) optimize specific goal-sets, and (3) simulations used to test out possibilities under controlled conditions. Game theory applies here and there are many excellent books on game theory, particularly as used to examine strategic options. These books, however; tend to be theoretic while reality never quite meets the models portrayed by the mathematical game mechanisms. Common strategic games include, without limit, day after games, prosperity games, and opposition games. These are often used by high-level decision-makers to examine long-term issues and to explore a space of options in the short run. A slide presentation on this subject area can be seen at http://all.net/Talks/Game/sld001.htm.

5.10 The interested reader is encouraged to look at "Security Decisions", which is part of the CISO ToolKit cited above, but which is also available as a software program at http://all.net/ under "Management Analytics". The book and program contain more detailed decision criteria associated with different circumstances for different enterprises including the specific basis for the decisions in different contexts.

Chapter 6

Governance

6.1 Largely as a result of inadequate attention to detail, groups like the Software Publishers Association (SPA) find companies in violation of licensing contracts and commonly gain treble damages (three times the manufacturer suggested retail value of every copy in place) and a right to inspect (typically the ability to look at what software is running in every computer in the enterprise using whatever means they see fit) for years to come. A common outcome is that the SPA finds further unlicensed software and gains more damaged and rights to inspect. This is not typically a matter of a company trying to cheat the rightful owner of the copyright, although there are certainly companies and individuals that do that. It is, more often than not, a company that just doesn't have very good control over content and technology. If this happens by accident in case after case, imagine what happens when someone tries to intentionally defeat the duties to protect.

6.2 The approach shown here was originated in the late 1980s and early 1990s and was first described in about 1997 (F. Cohen, “Protection and Security on the Information Superhighway”, Wiley and Sons publishers). It is now available online at http://all.net/books/superhighway/index.html. The basis of this approach is the somewhat philosophical perspective that if you do things right and don't forget anything the protection program will function, and if properly tracked and enhanced with time, it will improve with time. It has since been widely adopted in different forms and is increasingly embraced as an approach that is workable, even though nobody can really be definitive as to why it works or how well. In particular, the lists in each area are somewhat arbitrary and based solely on experience and rationalization. Many others use different lists, use different levels of abstraction in their lists, seek to find commonalities to avoid extensive lists, and so forth. Yet in the author's experience the lists as presented, if interpreted broadly, are reasonably comprehensive and tend to reveal what is important to reveal. Put another way, if you check everything listed, you won't miss any of it, even though you might miss something else. If you eventually find that you have missed something, you should add it to the list you use unless you can place it within the existing list somewhere reasonable.

6.3 Standards are in a state of flux in information protection and likely will be for some time. The list included here has changed by about 20% in the last 2 years, including renumbering of the 17799 to 27002 and the addition of 27001 which is more or less an outline of 17799 using only the control objectives. More standards are being built every day, and at all levels. NIST has a substantial standards effort that does an excellent job at making these standards open to all, while other organizations have chosen to charge for standards, which creates inhibitions to their widespread adoption. Technical standards arrive at a rate of well more than one a month now and range from changes in protection in wireless networking to intrusion communications language specifications for allowing ISPs to collaborate in fighting botnets. As in any broadly scoped field that matures, internal groups are needed to keep up with the relevant standards.

6.4 Data retention and disposition is increasingly critical because, among other things, of recent legal cases that have had large judgments when companies attempted to cover up wrong-doing by destruction of records. The Sarbanes-Oxley Act and similar laws around the world have been largely a response to abuses, and those abuses included destruction of records. Excellent coverage of the legal aspects of this issue is provided in *The Sedona Guidelines...* cited in endnote 4.1 above. Technical aspects of disposal are also complex and poorly understood and in case after case poorly thought out attempts to dispose of data have resulted in massive reconstitution at a later time. Excellent coverage of these issues is provided in Simson L. Garfinkel, “*Design Principles and Patterns* *for Computer Systems That Are Simultaneously Secure and Usable*”, dissertation from Department of Electrical Engineering and Computer Science, MIT, 2005

6.4b The current situation in protection testing is poor at best. For example, to do a test, the function has to be specified, and in most cases, units under test are not specified, incorrectly specified, or under-specified. However, many companies are starting to emerge in the area of protection testing and increasing rigor is being applied to testing in this area. The more well established fields of digital circuit testing and program analysis are increasingly being applied in this area and, as a result, the testing regimens are getting far better. Nonetheless, testing today is very limited and most claims made in this arena must be viewed with extreme skepticism. In almost all cases extremely large gapes are easily found in the testing approaches. Metrics of coverage are undefined, poorly defined, misleading in presentation in most cases, with their primary use being for advertising rather than giving accurate information on what to expect as a result of the tests. Also, as tests are applied, even given these limitations, large numbers of faults are found in almost all software and systems. These typically include large numbers of false positives (often 90% or more) and after removal of false positives, large numbers of remaining faults in the unit under test. Repairs often result in the introduction of more faults, and it seems clear that humanity just does not yet know how to build secure systems or test them definitively. Change control is usually not done at a level of surety appropriate to the consequences, and this is an area where surety fails to meet risks in case after case.

6.5 Technical controls for information typically involve a combination of many different elements that often get divided into “virtual” vs. “physical”. Since virtual is a complex issue to address, we take the approach of dealing with information technology, which we take to mean finite state automata executing with electrical, optical, or other similar signaling techniques, as opposed to physical, which we take to be the control of movement of people and things, typically through the use of barriers of some sort and sensors related to the movement, location, or state of the physical things. Areas in which these things intersect to make the issues particularly unclear are things like optical and electrical signal leakage and introduction of wave forms. But regardless of how the problem is ultimately divided, these issues may all have to be addressed.

6.6 Personnel issues generally involve human resources, which is almost never within the purview of information technology. The challenges of interfaces between personnel and other aspects of content control have long been problematic and have been largely solved in the information technology arena by the increasing integration of automated HR systems with identity management (IdM) systems and the integration of IdM into a multitude of automated systems through provisioning of access controls and other related authorization mechanisms and audit mechanisms. While this set of mechanisms is complex and problematic in many ways, it dramatically improves the timeliness and accuracy of controls, even if it aggregates risk in these control mechanisms. It is important in considering this to also recognize that the majority of issues addressed ultimately involve human judgment, human processes, and other things that are not highly automated or integrated. It is also important to understand that the HR mechanisms have to be integrated into all manner of management processes that support all aspects of the protection program because of the need for HR involvement whenever enterprise management or personnel are involved, which is the vast majority of the time.

6.7 Incident handling as described here is broader than common usage. In most cases, this is separated from disaster recovery and business continuity and is treated only in terms of technical incidents. In this broader approach, any event sequence that is detected and responded to is part of incident handling.

6.8 Forensics is an emerging field that is getting increasing attention in the enterprise. Networked forensic capabilities are increasingly being used and tied into the data retention and disposition process. The enormous volume of content in the typical enterprise is driving an increasing need to have centralized data storage facilities with search capabilities in order to meet search requirements of litigation as well as to provide unified archiving and disposition processes that are very hard to track. Because forensics is often subject to legal process the requirements for integrity of records and care in process drives up the perception of the need for a higher surety and quality level.

6.9 As described in endnote 6.5, physical security in the context of this discussion is related to the movement of people and things. Endnote 5.3 cites “The Design and Evaluation of Physical Protection Systems” which is an outstanding book for understanding the issues of physical security in this context. However; there are many physical security systems and requirements outside the scope of facility protection described in that excellent work. Even for relatively minimal surety requirements, physical security is very important and it is often poorly done, even when a lot of effort is put into information protection at the “logical” level. For example, it is common to be able to walk into a facility largely unaccosted, place a wireless access device on a network by plugging it into an RJ45 jack, walk out, and gain access to the enterprise network for an indefinite period of time. Facilities protection is often handled by a local facility manager and not coordinated across the enterprise. In this case, the CISO ends up helping manage facility protection for facilities containing systems handling content.

6.10 I don't make these things up. This is something I have seen time after time, although in slightly different forms each time. CIOs in large enterprises tend to be highly sensitive to problems that could become visible to other top managers and tend to cover them up rather than deal directly with them. This perhaps reflects the idea that a CIO does not belong at that high a level in an enterprise, or perhaps the idea that the CISO must be at the same level as the CIO. Perhaps it implies inadequate tolerance for the limitations of information technology, or perhaps over-promising by those in information technology. But I have seen time after time where CISOs got fired for doing their jobs to delay the firing of the CIO till they could find a new job elsewhere and move on. This is the nature of politics within an enterprise.

6.11 Database security has a long and storied history, and database vulnerabilities continue today as they have for more than a half a century. If anything, they have become more populated with content and less secure in their operation. An excellent resource to start reading about this issue is:

> 6.11.1 Dorothy Elizabeth Robling Denning, "*Cryptography and Data Security*", Addison Wesley, Reading, MA, 1982.

Denning's coverage of this area is important background for understanding the protection issues in computer systems because databases are, in many ways, a reflection of computer systems in general.

For example, data aggregation in databases is a form of the covert channel problem (see endnote 7.1.6) that is still problematic today after more than 35 years, and in far larger measure than before because of the enormous number of databases with open access. One of the best known examples is "Google", which increasingly provides information used to break into systems, hunt down spies, and do intelligence operations against opponents. Covert channels have become so widespread that they create a flood of security problems, and search engines augmenting the Web have produced such a change that they seem to defeat the vast majority of controls put in place in the first golden era of databases.

Databases today still often contain more of the most valuable content of an enterprise than any other content arrangement, are required in near-real-time for operational continuity, and their failure can cause even some of the largest enterprises to fail, in some cases, literally over night. The databases increasingly include more computational capability and for more and more varied algorithms then they did in the past, and many databases today operate in distributed computing environments, link with other internal and external databases, and are highly interdependent on each other and intervening infrastructure. The failure of a database at once company can ripple through to an inability of many other companies to price and take orders, process purchases, deliver goods, and collect payments. Consider that a simple book purchase today over the Internet likely involves (1) a database of books, their availability, stocking information, and so forth, (2) a pricing model associated with a customer database and historical information on purchase prices that are themselves gained from external brokers and distributors, (3) a processing network of databases for processing credit cards that involves a processing center, a credit card company, a credit card issuer, and details of available account balances, (4) a database used for shipping and inventory control that may link to an external shipper and their databases that control vehicles, warehousing, inventory, and so forth, (5) an accounts receivable and payable database that tracks the transactions and processes the financial information associated with the purchase for clearing monies through the financial system and yielding financial records and reports. and (6) all of the things that all of these databases depend on that also have databases, such as the domain name system, which is a large distributed database. Database protection is clearly vital and a microcosm of the larger enterprise information protection challenge.

Chapter 7

Control Architecture

7.1 The development of these concepts come through many long term efforts of outstanding researchers and scientists who cannot all be adequately credited here. But for a brief historical perspective, and a list of papers and writings that anyone who wants to be top flight in this field should read, here is a start at a reading list:

- 7.1.1 C. Shannon, “*Communications Theory of Secrecy Systems*”, Bell Systems Technical Journal (1949):pp656—715. In this paper, Shannon applied his information theory to breaking all of the known ciphers up till that date and provided a proof that the only theoretically unbreakable cryptosystem was the so-called perfect cipher. This paper also introduced the concepts of diffusion and confusion, and introduced the concept of work load which is the basis for using imperfect cryptosystems today.
- 7.1.2 W. Ware, “*Security Controls for Computer Systems (U)*”, Report of the Defense Science Board Task Force on Computer Security, Feb. 11, 1970, available online at: http://all.net/refs/ware70.pdf. This report covers the issues associated with creating a time-sharing multi-access computer system serving geographically distributed users, and processing the most sensitive information. Among the things it points out clearly is that solving this problem includes a combination of “hardware, software, communications, physical, personnel, and administrative-procedural safeguards...” and identifies the fact that, at that time, attaining a system meeting all of the identified security requirements was not feasible (as it appears to still not be today).
- 7.1.3 James P. Anderson, “*Computer Security Technology Planning Study*”, ESD-TR-73-51 Vol 2, October, 1972 Prepared under Contract # F19628-72-C-0198. This study overviews the requirements for research and development required for the development of multi-level trusted computer systems. It is available online at http://all.net/refs/ande72.pdf
- 7.1.4 D. E. Bell and L. J. LaPadula, “Secure Computer System: Unified Exposition and Multics Interpretation”, ESD-TR-75-306, March, 1976 This report presents the Bell LaPadula model of computer security in which information flows from less trusted to more trusted users only as well as its implementation in a concentric ring structure within Multics. The concepts consolidated in ths report include many of the key concepts used since that time and the models they use were the foundation of models for secure computing through the 1980s and continue in use

today to a large extent. The references from this paper are particularly seminal in the field as well. This is available at http://all.net/refs/bell76.pdf

- 7.1.5 P. G. Neumann, L. Robinson, K. N. Levitt, R. S. Boyer, and A. R. Saxena, "A Provably Secure Operating System", Final Report, Contract DAAB03-73-C-1454, June 13, 1975 (see http://all.net/refs/neum75.pdf) is a competitor to the Multics approach and discusses the issues of developing a secure operating system including all of the mathematical analysis involved in completing the effort. Again, it lays out the foundations of defining the control objectives and attaining them in a provable manner.

- 7.1.6 B. W. Lampson, "*A Note on the Confinement Problem*". CACM 16(1), October, 1973:613-615 This paper described the covert channel problem for the first time. The implication is that no system that shares resources in a non-fixed fashion can ever provide perfect separation of the sharing parties. It is an example of how models of security and control architectural concepts can collapse based on a single missing concept in the analysis. The impact of this cannot be underestimated, even though it has been by many.

- 7.1.7 Denning, D. E. "*Secure Information Flow in Computer Systems"*, Ph.D. dissertation, Purdue Univ., West Lafayette, Ind., May 1975. In this thesis, Denning describes the use of a Lattice structure to model secure information flow within a computer system. It is the logical progression of the previous information flow models and consolidates them into a unified mathematical framework.

- 7.1.8 The Trusted Computer System Evaluation Criteria (TCSEC) moved this concept into a formalized approach to creating secure operating systems to meet high surety requirements. However, just as the first systems were ready to be accepted under the TCSEC...

- 7.1.9 F. Cohen, "*Computer Viruses - Theory and Experiments*", IFIP TC-11 Conference, Toronto, 1984 (also appearing in Also appearing in DOD/NBS, 7th Security Conference, Sept, 1984. Also appearing in IFIP-TC11 Computers and Security, V6(1987), pp22-35 and many other subsequent publications). This was the first and is the most widely cited paper about computer viruses. It includes most of the scientific results about virus detectability, the infectious nature of viruses, and the vulnerability of systems found early in anti-virus research. Among the things it identified were that trusted systems including all of those listed above could be subverted by the use of computer viruses to spread infections from the least trusted user to the most trusted user and, because of Lampson's paper, that this could be used to leak secrets, thus defeating all systems of the day and under the models of the day and putting the fundamental models in peril as an approach to building secure systems to meet high surety needs. Only truly separated systems with provably secure one-way communications could meet the requirements, and that required...

- 7.1.10 F. Cohen, "*Designing Provably Correct Information Networks with Digital Diodes*". This paper relied on from

    7.1.10.1 C. Shannon, "A Mathematical Theory of Communications", Bell Systems Technical Journal, V27 #3,4 July+Oct, 1948 pp. 379-423, 623-656;

    7.1.10.2 E. Moore, C. Shannon, "Reliable Circuits Using Less Reliable Relays", J. Franklin Inst #262 pp 191-208, Sept 1956; and

    7.1.10.3 J. von Neumann, "Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components", Automata Studies, Princeton University, University Press, pp 43-98, 1956

  to show that a physical device could provide one way communications with adequate separation to allow a Partially Ordered Set (POset) to be implemented (which is a generalization of Denning's Lattice structure cited earlier). This made perfectly secure systems, in the sense of information flow control, functional, but it eliminated the ability to have multi-level secure operating environments (as opposed to multiple security levels which it enables ideally).

- 7.1.11 F. Cohen, "*Models of Practical Defenses Against Computer Viruses*", IFIP-TC11, ``Computers and Security'', V7#6, December, 1988. In this and earlier related papers, the integrity problems identified by the computer virus problem were largely addressed based on the notion of providing computationally strong integrity maintenance mechanisms and examination of the interdependencies of all executed content on all other content in the transitive closure of dependency. This was closely related to the longstanding practice of watermarking and the use of watermarks for document authentication but substituting cryptographic methods for printing methods. This in turn led to the creation of integrity shells and similar mechanisms that ultimately led to the current developments in "Trusted Platform Module" (TPM) mechanisms under the Trusted Computing Group (TCG) consortium. The watermarking line was also followed and increasingly steganography-based watermarking is used to provide attribution at select points in the lifecycle of data, or more commonly of files. In order to implement all of these things, it was also important that developments in cryptography take place to allow cryptographic integrity mechanism to increase the cost (per Shannon above) of defeating integrity to be very high without excessive costs for protection. This in turn depended on public key cryptography.

- 7.1.12 In the late 1970s, Diffie and Hellman created the notions underlying public key cryptography and, while their original systems were defeated over time, they eventually led to the creation of the RSA cryptosystem. R. L. Rivest, A. Shamir, and L. Adleman, "*A Method for Obtaining Digital Signatures and Public-Key Cryptosystems*", Comm. of the ACM (February 1978) 21, no. 2:120--126. This famous paper introduced the best known and oldest largely unbroken class of public key cryptosystems. These are systems that allow digital signatures or encryption both from and to a trusted or untrusted source. There were also other models of security introduced through these approaches including, without limit, a model wherein anybody could store their data on public forums with assurance of the integrity and secrecy of that data, the notion of publication of keys to cryptographic systems thus reducing the n^2 key problem to 2n keys, and turning what required exponential space into a solution requiring only linear space while reducing secrecy requirements and distributing trust.
- 7.1.13 With the introduction of public key systems and certificates based on applying those systems to decentralized certification of authority by trusted parties in chains of trust, came Kerberos; a system for managing and distributing session keys covered by public keys and trusting a central repository. This is the basis of the current Microsoft key management infrastructure and is the most successful of the current systems in surviving attacks. It was based on earlier work used in IBM networks for allowing trusted third parties to authorize actions through similar trust mechanisms, but Kerberos extended these concepts to operate under public key systems, thus solving a variety of technical problems that limited scalability.
- 7.1.14 Identity management (IdM) started to emerge in the early 2000s with the advent of automated provisioning systems (systems that automatically set protection values, user identities, passwords, and other configuration parameters on remote systems). These systems used cryptographic channels, and databases (typically directory systems initially based on LDAP) to automatically configure large numbers of systems, thus reducing systems administration overhead. A single administrator could set a new permission configuration for a group of users and the provisioning system could automatically propagate this setting to thousands of computer systems. This then led to integration of HR systems into the provisioning system and allowed work flow mechanisms to automate new employee additions, termination processes, and other similar activities that were previously human intensive, highly subject to errors, and untimely. As these systems deployed to larger populations, more platforms were integrated and it became increasingly difficult to specify what to do across these systems. Roles and rules were implemented to help to automate this process, and then...

- 7.1.15 In the middle 2000s, policy description languages started to come into increasing use to allow the rules for provisioning to be codified linguistically and integrated into workflows for complete automation of the operational mandates of an enterprise based on higher level enterprise policy. These systems deal primarily with what we will call technical security policy, and allow arbitrary logic expressions, and ultimately, arbitrary programs, to be executed to determine what who can do at what time, from where, and using what programs (who, what, where, how, and when). However; the "why" question cannot be answered by these systems. That ultimately limits their ability to make determinations or mistakes in the manner that people would.

- 7.1.15b These systems also have other technical limitations, but they dramatically changed the model of protection into one of highly differentiated access based on arbitrary rules, which in many senses is a return to the notions underlying the original work on subjects and objects stemming from

  > 7.1.15c M. Harrison, W. Ruzzo, and J. Ullman, "*Protection in Operating Systems*". CACM 19 no. 8 (August 1976):461—471.

  This paper introduces an early formal model of protection in computer systems and forms the basis for the subject/object model of computer security. It also proves that determining the protection effects of a given configuration is, in general, undecidable.

- 7.1.16 This last result is particularly important because it then becomes clear that all such languages are fundamentally limited in what they can do and shows that the computational complexity of the identity management system must be limited by limiting what it can actually do. That is, in order for it to be reasonably usable, it must not be truly general purpose. Issues like revocation or rights create performance problems if attempted, and problems like the time transitivity of information flow (see F. Cohen, "Protection and Administration of Information Networks under Partial Orderings", Computers and Security, 1986.) lead to extreme complexity when dealing with separation of duties and similar issues over time and across personnel changes and relationships.

- 7.1.17 Meanwhile, change control has been a vital issue that was discovered early in engineering and has been the subject of engineering controls ever since. Change control generally provides increased assurance around changes, ranging from limiters on slew rates and values for automated control systems to submit/commit cycles that force independent verification through independent channels before allowing high-valued transactions to take place. Cryptographic checksums and other mechanisms described earlier for integrity protection are also variations on change control as enforced by different technical mechanisms. But in the computing arena, the fast and loose development cycles of the 1990s changed all of this for quite a while. Time to market started to dominate quality in the era of the personal computer as the “glass house” mentality that had people with a high degree of expertise taking care of and managing all computation and computing facilities. The facilities were expensive and thus there was a need to make sure they were well used. As the price went down, the need to reduce waste was seen as less important, and budgets allowed almost anybody to purchase computing power. As more and more people gained computing power and computers, there was no simultaneous development of expertise, so expertise grew organically and most of those in charge of computers had little or none except in their own area of specialty. They had enough knowledge of computers to be dangerous – and that word is well used in this context. The carefully controlled software environments that assured that money was properly counted in banks and books were accurate yielded to the era of spreadsheets with wrong answers, formulas based on questionable data producing profit and loss statements that were fantasies, and eventually to down right frauds. The older models of change control still exist and are coming back and increasingly becoming a vital part of building medium and high surety systems. Unfortunately, in the commercial software arena, change control has not been put in place in most projects and, as a result, there is less of a basis available to depend on the programs prior to the point at which change control can be put in place by the enterprise.

- 7.1.18 As the Internet emerged, the problem of transitive trust has emerged as increasingly problematic. The demonstration of inadequate trust models given by present computer viruses is only the tip of the proverbial iceberg when it comes to the difficulty associated with making determinations about what can be reasonably mixed with what else while maintaining assurance to desired levels. And yet today most uses of the Internet and many functions within enterprises operate based on content that cannot be determined to be accurate or even reasonable. Caveat emptor seems to be the way of the Internet, and yet enterprises must find some model of what to deal with in what manner to allow a reasonable amount of commerce while still inhibiting criminal and accidental losses to reasonable amounts.
- 7.1.19 The notions of trust is another area where security has attempted to build models. Starting with the notions of trust underlying background checks and separation of duties, the field has always been related to trust. In the time frame just after computer viruses were first being published, Peter Denning gave his Turing Award lecture titled “*Reflections on Trusting Trust*”. This was a reflection on the limits of the ability to detect Trojan horses in software, a problem long believed to be extremely hard for programs of any substantial size. At the time, programs of 100 lines of code were the focus of attention because the security kernels of the early trusted systems were on the order of 100

lines of code. It was identified that unless the programmers were trustworthy, it could not be assured that they had not planted a Trojan horse in their code. Thus the requirements for cleared programmers were put into evaluation criteria at that time. Moving forward in time, more complex trust and reputation models were associated with public key cryptography, certificates, trusted certificate providers, and ultimately complex systems such as those supported by eXtended Access Control Markup Language (XACML) and similar languages that allow the extension of trust across domains through federation and other loose affiliations. These models of trust do not have a sufficient basis today to demonstrate their trustworthiness or an analytical framework that can be applied to reason efficiently about trust issues widely faced by enterprises today. See also:

7.1.19.1 "P. G. Neumann, "*Reflections on System Trustworthiness*", A chapter in "Advances in Computers" by Marvin Aelkowitz, pp 269-310

- 7.1.20 Finally, it is important to touch on the subject of functional elements and understanding their surety and interoperation. The secure development process identified in the earlier papers is almost non-existent today and, as a result, other approaches may be needed. At present, there are researchers who have worked on the notions underlying the creation of composites from components so as to derive properties of composites based on properties of components and the way they are connected and used. For details see:

7.1.20.1 P. G. Neumann, R. Feiertag, "*PSOS Revisited*", available at: http://www.csl.sri.com/users/neumann/psos03.pdf

Reference 7.1.19.1 above

While we await better models we do continue to connect components to form composites, but we have little in the way of effective models for how to analyze them or determine their properties. Still, this element of the control architecture is fundamental to decision-making and structures are used to limit zones and control content within systems. As such, the functional elements are necessarily modeled.

This very brief outline of the issues that are central to the issues underlying the concepts of control architecture is intended to allow the reader to start their study of this area, but hardly dents the surface. We hope they start to bring light to the challenges underlying this field, which is not adequately studied as a field and yet so vitally critical to the design of protection.

7.2 For good coverage of the issues associated with reliability look at the field of Fault Tolerant Computing which has several fine conferences and journals, particularly in the IEEE where much of the development of this field for integrated circuits and computer systems is documented starting largely in the 1970s. See also references 7.1.10.2-3 for very early coverage of concepts.

7.3 See: White, S.R, Comerford, L. "*ABYSS: an architecture for software protection*", IEEE Transactions on Software Engineering, 1990 V16#6 pp 619-629. FIPS certifications exist for different levels of surety. ABYSS (a basic Yorktown security system) is an architecture for protecting the execution of application software. It supports a uniform security service across the range of computing systems. A novel use-once authorization mechanism, called a token, is introduced as a solution to the problem of providing authorizations without direct communication. Software may be transferred between systems, and backed up to guard against loss in case of failure. The problem of protecting software on these systems is discussed, and guidelines to its solution are offered. Certified systems under the FIPS certification process are listed online at http://csrc.nist.gov/groups/STM/cmvp/documents/140-1/1401val-all.htm.

7.4 Wietse Venema, "*TCP WRAPPER Network monitoring, access control, and booby traps*". 3rd UNIX Security Symposium, Baltimore, September 1992.

7.5 For more on trust see "Other" at the end of the end notes

Chapter 8

Technical Security Architecture

8.1 The history of technical security and the concepts in use in that arena are largely from time frames beyond the history of professional societies and publications. They are embedded in culture and historical documents ranging from religious texts to military histories. As such, finding citations is often problematic. And yet things like passwords, which certainly existed long before the most ancient texts were written, have survived and remain in use today. While the first use within a computer may be only 70 years old, physical security concepts were largely carried over into the computing arena, and to cite the first person to reuse an ancient concept is hardly to give proper credit to the ancients who invented the concept or the natural phenomena that many common security concepts depend upon.

8.2 This characterization of the generic attack graph and the use of attack graphs for modeling information protection is due to:

> 8.2.1 F. Cohen, "Red Teaming Experiments with Deception Technologies", 2001 and subsequently in IFIP TC-11 Computers and Security. http://all.net/journal/deception/experiments/experiments.html .
>
> 8.2.2 F. Cohen and D. Koike, "*Leading Attackers Through Attack Graphs with Deceptions*". May 2002 subsequently in IEEE conference proceedings: http://all.net/journal/deception/Agraph/Agraph.html

8.3 Deception articles can be found at http://all.net/journal/deception/index.html. The most important ones to read for historical coverage is:

8.3.1 Fred Cohen, Dave Lambert, Charles Preston, Nina Berry, Corbin Stewart, and Eric Thomas, "A Framework for Deception" which was published in IFIP TC-11, Computers and Security, 2001.

8.3.2 F. Cohen, "*The Use of Deception Techniques: Honeypots and Decoys*" (the encyclopedia article at the previous URL), 2003.

8.3.3 F. Cohen, "*A Mathematical Structure of Simple Defensive Network Deceptions*", 1999

Many more recent attempts at the use of deceptions have been tried and the area is increasingly becoming popular, but it is still shied away from by many and remains an emerging topic in many areas. Historical deceptions, such as refusing to indicate whether a user identity is valid without a valid password and concealment of files not authorized for read access are in common use and not viewed as deceptions even though they fit into this arena.

8.4 Firewalls started out as an extension of the principles of trusted systems to networking environments. They were implemented first as access control mechanisms such as trusted network guards and later as routers or similar infrastructure mechanisms using IP addresses, port numbers, or similar addressing mechanisms to differentiate computers and grant them selective access to other addresses. Over time they have evolved into complex combinations of components that provide far deeper inspection of content as it moves form place to place. Nevertheless, the complexity of detecting malicious content makes it impossible for a practical system of this sort to be precise. For details of why, see:

8.4.1 A. Turing, "*On Computable Numbers, with an Application to the Entscheidungsproblem*", London Math Soc. Ser 2. Vol 42,Nov 12,1936,230-265. This is the famous paper that shows that any problem that can be solved by any general purpose computer can also be solved by any other general purpose computer, given enough time and space. It also shows that a large class of problems can never be solved by a computer. The so-called Halting problem, in particular, is proven unsolvable.

8.4.2 F. Cohen, "*Computer Viruses*", ASP Press, 1986. This Ph.D. dissertation provides the mathematical basis for most of the mathematical work on computer viruses to date, including the formal definition, proof of undecidability, properties of viruses and viral sets, and proof that transitivity, Turing capability, and sharing lead to viral spread.

8.4.3 F. Cohen, ""National Technical Baseline Study: *Intrusion Detection and Response"*, Lawrence Livermore National Laboratory and Sandia National Laboratories December, 1996. http://all.net/journal/ntb/ids.html

8.4.4 F. Cohen, "*50 Ways to Defeat Your Intrusion Detection System*", December, 1972, http://all.net/journal/netsec/1997-12.html. This particular paper was a method to poke fun at the research community that was

> continuing to develop poor intrusion detection systems, because of their failure to listen to the more academic version of the situation as presented in the national technical baseline study cited in 8.4.3. It resulted in many of the members of that community becoming upset at the author but also helped to move forward the more serious aspects of considering these challenges in a serious way.

These papers show some of the computational basics of why detection and limitation based on content is so difficult.

8.5 The Boyd Cycle (http://www.boydcycle.com/) is described by the observe, orient, decide, act (OODA) loop as a method by which people act on stimulus and that limits their ability to act. To a large extent, tactical warfare in many circumstances can be dominated by the individual that gets inside the OODA loop of their opponent and thus makes in impossible for their opponent to make and cary out sound decisions in a timely fashion. The concept has been applied in many other areas.

8.6 Limited and well controlled separation is the basis for the Bell LaPadula model cited above and much of the history of access controls and trusted systems. Over time, as described in 7.1, separation was found to be limited by computational complexity and covert channels, leading to the use of digital diodes as the only effective means of providing for one way information flows in computer systems. Architectural separation is often attained in limited ways through the use of firewalls and network segmentation with trusted systems as intermediaries between zones.

8.7 Attribution is enormously problematic in today's networked environments. The challenges range from being able to identify the direct contact point of events under scrutiny to attribution to original actors and groups that they are part of or that supported them. In general, an intelligence effort is required for attribution and even the best intelligence forums in the world are not completely successful at defeating deceptions intended to obfuscate attribution. In most systems today, even with identification and authentication properly working, tracing events to a human source is often problematic because the operating environments on which the authentication was done is untrustworthy. Proving who was at what keyboard is a serious problem in digital forensics that almost always depends on external events not associated with the computer system and the correlation of those events to computer-related events.

8.8 The interested reader might want to read F. Cohen, “*Challenges to Digital Forensic Evidence*”, ASP Press, 2008, to gain additional understanding of the difficulty with proving who did what to a legal standard.

8.9 For an alternative run through of defenses and mapping them into different aspects of this characterization, see the database at all.net, which includes definitions and linkages between threat types, attack mechanisms, and defense mechanisms as well as their properties. http://all.net/ select “Database” and press “Go”.

8.10 NSTSSI standards (http://all.net/books/standards/NSP/index.html) provide the means by which the US government produces and executes on secure systems. Note that the many spelling errors and lack of complete sentences underlie that nature of the beast supplied to the educational community.

8.11 Protective mechanisms are the heart of what most people think of when they discuss technical aspects of information protection. They are presented here as broken down into perception, structure, content, and behavioral controls, and this is a very unusual breakdown of the field. Others have chosen other structures and there is no particular reason to prefer one over another except as it brings utility to the reader. This particular approach was thought up when trying to review the previous literature and consolidate it under the notion that technology should not guide understanding, but rather understanding should deal with human views and technology should fall where it may. That does not mean that others are less humanist, only that this is the way the thinking went. This was originated in discussions with Eric Maywald of Burton Group's Security and Risk Management Strategies team during a dinner we were attending at the RSA conference and was fused soon thereafter once it came clear that this approach might yield fruit. It was closely related to the notion that there are maximum expected surety levels associated with separation (structure), transformation (content-based), and pattern matching (behavior, content based, and perception).

8.12 Security through obscurity is widely touted as problematic, and yet at the end of the day, only physical separation and obscurity really underlie all protection, since a person or group that knows enough and has the proper physical characteristics and things ultimately cannot be defeated by protective measures that allow legitimate use. To complete the argument in favor of this position, it might be considered that any legitimate use must differentiate between at least two different inputs or outputs that have consequences and, at the end of the day, some set of circumstances must be able to produce each of the outputs and their consequences or the system is not useful. Any attacker that can produce the “wrong” output defeats the system, and a collusion of all of the sources of input can certainly accomplish this. Since physical separation is not likely to be reasonable in most circumstances as the sole protective mechanism, it then becomes clear that knowledge must be limited, and hence obscurity is fundamental to security. Consider the password as an example for clarity.

Now having said this, the argument against the use of obscurity generally admits these possibilities and restricts the problems of security through obscurity to the use of easily defeated systems which leave their assurance to a lack of even simply attainable knowledge by the attacker. This is hard to argue with. In the cryptography arena, the notion is that something that must be obscured should be limited to cryptographic keys, and thus cryptography must normally obscure the keys but nothing else. Another way to say this is that anything that has to be obscured is a key. Cryptographic systems that are useful normally have to be distributed for use. If the system itself must be kept obscure, all uses would be defeated if the system itself was ever examined. This severely limits its utility.

8.13 Low profile buildings as a security measure became popular in the 1960s when high profile computer facilities at universities and elsewhere were subject to attack as part of anti-war protests. This continued well into the 1980s and 1990s as the Irish Republican Army targeted high profile buildings associated with the financial industries for bombings. And of course the World Trade Center bombings are another in the long line of examples of the problem with placing high valued targets in high profile buildings. The same has been true in general for ever since wars involved attacks on facilities, but in the information arena, the issues became profound only as information technology was associated with facilities.

8.14 Appearance is closely related to deception and is often grouped under “perception management” as part of the same general area. There is a large body of psychological research behind the use of appearance and other aspects of perception that can be very helpful in understanding and anticipating what will happen as a result of decisions about what appears how, where, and when. From being ecologically sound as a company stance, to concealing where animal research is done, appearances can have substantial effects on protection of information content by its ability to reduce threats. The interested reader should review the citations from 8.3.1

8.15 Enterprise rights management (ERM) and digital rights management (DRM) are methods of applying cryptographic seals and signatures to authorize or track uses of content. In these systems, it is most common for a display or use mechanism to take content distributed without constraint and allow access to it only by someone authorized by a key or through a rights management infrastructure, such as a TPM accessing a certificate authority and key management system. All such systems are problematic in that the secret keys used to unlock content can always be gleaned or the content gathered from legitimate users or the mechanisms that display the content. The classic depiction is a video display being duplicated by a copier, however, screen capture and optical character recognition or other similar methods are commonly used today to defeat such systems when they cannot be readily defeated cryptographically.

8.16 Encryption systems are systems used to transform content into a form that can only be understood through the use of a key. This allows encrypted content to be moved about with reduced risk to exposure of the meaning of the content. Virtual Private Networks (VPNs) create encrypted tunnels through which other protocols allow arbitrary exchanges of information flows that are concealed from external observation. This tends to defeat intrusion detection and other similar sensor mechanisms as well as defeating illicit users trying to view content. This them drives enterprises to decrypt all content in security devices, leading to increased risk aggregation in those devices, reduced performance, and the false appearance of end-to-end coverage.

8.17 The limits on known content filters are immediately apparent in the context of the problem of virus detection identified in 7.1.9, 7.1.10.1, and as discussed at length in 8.4.3.

8.18 The use of markings stems from early trusted systems which used marking mechanisms to associate security information with content. For example, a file that is classified as "secret" might have a marking associated with it in the file system to allow the security kernel to make its determinations regarding accessibility based on the clearance level of a process, which also has a security marking. These marking-based approaches in turn stem from physical security marking, such as a stamp on a document that indicates its classification level and the use of a guard to determine whether the individual requesting access has the clearance by looking them up on a list of authorized users.

8.19 Time and rate controls are very old, however, an excellent coverage of the topic for the lay reader is available from fairly recent times. See: W. Schwartau, "*Time-based Security*", Interpact Press, ISBN# 0-962870-04-8

8.20 Failsafes are generally included in coverage of fault tolerant computing. An excellent early overall book on the subject is D. P. Siewiorek and R. S. Swartz, "*The Theory and Practice of Reliable System Design*", Digital Press, Bedford, Mass., 1982, This book describes many of the underlying techniques and much of the well understood theory of high reliability computer system design.

Chapter 9

Making Better Security Decisions

9.1 The interested reader may want to look at:

9.1.1 Bob Fellows, "Easily Fooled", Mind Matters, PO Box 16557, Minneapolis, MN 55416, 2000

9.1.2 Thomas Gilovich, "How We Know What Isn't So: The fallibility of human reason in everyday life", Free Press, NY, 1991

9.1.3 Chester R. Karrass, "The Negotiating Game", Thomas A. Crowell, New York, 1970.

9.1.4 Robert B. Cialdini, "Influence: Science and Practice", Allyn and Bacon, Boston, 2001.

9.1.5 Charles Handy, "Understanding Organizations", Oxford University Press, NY, 1993.

9.1.6 Peer Soelberg, "Unprogrammed Decision Making", Academy of Management Proceedings, 1966, p3-16.

9.1.7 Gilles Coppin, Frederic Cadier, and Philippe Lenea, "Some considerations of cognitive modeling for collective decision support", Proc 40th Hawaii Int. Conf. on Systems Sciences, 2007.

9.2 Miller, G. A. (1956). "*The magical number seven, plus or minus two: Some limits on our capacity for processing information.*" Psychological Review, 63, 81-97. Available online at: http://www.musanim.com/miller1956/

9.3 Waterman, D.A., Hayes-Roth, F., "*Pattern-Directed Inference Systems*" New York: Academic Press, 1978.

9.4 F. Cohen, "*Making compliance simple – not*", analyst report for 2007-10, available at http://all.net/Analyst/2007-10.pdf

9.5 The “Security Decisions” software tool can be downloaded for free trial use from http://manalytic.com/ to get a sense of how it works and to try out decisions.

Other

In reviewing the book for indexing purposes, I came across things I expected to see but didn't or end notes that I added late in the process. These are they:

O.1 Periods processing is a process by which systems are used for different periods for different purposes. They go through Color Changes, a cleansing process which eliminates residual content, between periods to allow them to be used for different purposes without cross-contamination or covert channels.
O.2 Color changes – see periods processing.

O.3 Trust, trustworthiness, trusted, and similar terms relate to the extent to which one can be harmed. More trust means that more harm can result. Trust in excess of trustworthiness implies excessive risk, while trustworthiness in excess of trust implies less utility than is achievable. There are many models of trust and they are all problematic in their way. The most critical problem with trust is that is is usually transitive in nature, in that when you trust someone or some thing, and they trust someone else or something else, then you indirectly trust the thing they trust. This chain can go on and on, possibly with a reduction in effective trust over distance, but not always. Because digital systems are so good at transitive extension, for example they make perfect copies of the bit values in most cases, computer viruses use this transitive extension of trust to reach the transitive closure of information flow in survivable environments.

O.4 Generally, high surety mechanisms, such as separation, physics-based, and limited function mechanisms can be implemented with lower surety and often are, but they can reach extremes of surety. Medium surety mechanisms, typically transforms such as encryption, cryptographic checksums, and so forth, can also be implemented poorly and lead to lower surety. Low surety mechanisms, typically pattern matching mechanisms such as malicious code detectors and vulnerability scanners, can never achieve medium or higher surety unless they are used in specific situations wherein their properties can be shown to be definitive with respect to criteria. Redundancy can compensate to some extent and drive up the effective surety within a category, and proper implementation and operation also increase surety, but these do not allow low surety mechanisms to reach medium surety or medium surety mechanisms to reach high surety.

O.5 The issues of compliance, identification, and internalization, and is detailed under “Responses to Power and Influence” on or about page 154 and in 9.1.3-5.

O.6 Issues of power and influence are more broadly explored in:

F. Cohen, “*Frauds, Spies, and Lies and How to Defeat Them*”, ASP Press, 2005 ISBN# 1-878109-36-7.

O.7 According to a recent study, 70% of investigated successful network attacks targeted content or systems that the enterprise protection team didn't know existed. No inventory – no protection.

13 Index

For a far more detailed index to the book, look at:
http://asp-press.com
under the detailed listing for this book.