CISO Essentials Guide

A three-pronged approach to leading successful security programs

Sriram Lakshmanan

bpb

CISO Essentials Guide

A three-pronged approach to leading successful security programs

Sriram Lakshmanan

CISO Essentials Guide

A three pronged approach to leading successful security programs

Sriram Lakshmanan

www.bpbonline.com

First Edition 2026

ISBN: 978-93-65891-201

LIMITS OF LIABILITY AND DISCLAIMER OF WARRANTY

The information contained in this book is true and correct to the best of author's and publisher's knowledge. The author has made every effort to ensure the accuracy of these publications, but the publisher cannot be held responsible for any loss or damage arising from any information in this book.

All trademarks referred to in the book are acknowledged as properties of their respective owners but BPB Publications cannot guarantee the accuracy of this information.

www.bpbonline.com

Dedicated to

My Amma and my (late) Appa for shaping my ideals, my value system and for being a constant source of support, sense and food.

About the Author

Sriram Lakshmanan (Sri) has over 25 years of experience in wide-ranging global security and risk leadership roles at Healthcare, Financial Institutions, Security Consulting, and Professional Services organizations.

He has experience in driving techno-human capabilities around defining, implementing, and running measurable security programs globally. He has built insights and gathered experience in the domains including infrastructure and application security, threat management, incident handling and reporting, open-source intelligence, and GRC (including third party risk management). He has built teams and processes from the ground up in not just information security but also in infrastructure delivery. The core of his approach is to drive the culture to equip the workforce to adapt to evolving threats and to enable business objectives securely.

Sri often speaks and participates in public forums such as industry events and security conferences like the RSAC.

Sri currently works as a Deputy CISO at a large technology-driven professional services organization. He has a Bachelor's Degree in computer science from Delhi University and a Post Graduate Diploma in Business Management. Over the years, he has acquired several certifications from leading institutes like the ISC2, ISACA, and SANS.

About the Reviewers

- **Rahul Anand**

  Rahul Anand is a passionate cyber resiliency and risk management leader with over 15 years of experience in cybersecurity, digital forensics, disaster recovery, IT governance, and offshore delivery strategy. He has successfully led global programs across industries, delivering measurable improvements in cyber posture, compliance, and operational resilience.

  Rahul holds multiple industry certifications including CISM (ISACA), Certified in Cyber Security (ISC2), and Certified Disaster Recovery Professional. He earned his MBA from XLRI, Jamshedpur, and a bachelor's in engineering from Shivaji University. He specializes in building and scaling cyber resiliency frameworks aligned with NIST, PCI DSS, and ISO 27001. He is currently working as CISO at Trilegal, where he leads global cybersecurity initiatives and is part of the leadership team driving innovation, compliance, and resilience across critical systems and services.

- **Senthil Subramaniam**

  With over 21 years of experience in cybersecurity, ESR Senthil Subramaniam currently serves as the Global **Chief Information Security Officer** (**CISO**) at Infinite Computer Solutions. In this role, he leads enterprise-wide initiatives to safeguard digital assets, manage cybersecurity risk, and ensure compliance with global regulatory standards.

  Through his diverse experience, Senthil has developed deep expertise in **governance, risk and compliance** (**GRC**), security operations, incident management, cloud security, Zero Trust Architecture, and cyber resilience. As CISO, he works closely with executive leadership and cross-functional teams to align cybersecurity strategies with business objectives, ensuring agility and responsiveness in an ever-evolving threat landscape.

Senthil has received several prestigious accolades, including the Security Accelerator 2025 award from the CIO500 Forum and the Leadership Excellence award from CXO Junction, recognitions that reflect his significant contributions to the cybersecurity domain.

A passionate advocate for building a strong security culture, Senthil actively mentors aspiring cybersecurity professionals and regularly participates as a panelist in industry conferences. He also engages with academic institutions as a speaker and thought leader. His certifications include CISA, CCSK, CPISI, ECSA, and ISO 27001 Lead Auditor. Additionally, he has authored articles for leading regional magazines, sharing insights on emerging trends and best practices in cybersecurity.

Acknowledgement

It takes a village to move a mountain. This book, for me, was the mountain. I am eternally grateful to the villagers: my wife Uma, our daughter Ankita, and my niece Dyuthi, who have stood by me patiently as I went through this journey over 7 months.

A large part of my success is attributed to family, especially my sisters - Shanthi, Sareswati, and Jaya, who have always given me the encouragement to experiment and to have my back. My brother-in-law Kannan, who gave me the laptop on which this book came about. And my father-in-law, who has more confidence in my book than anyone else.

There are several people who have shaped my journey in this exciting field of work. To all those mentors, supervisors, and colleagues, I am indebted to you.

A bunch of friends, who remain anonymous here, pushed me hard to complete the chapters and are more excited for the book than I may have been. Thank you guys.

I am grateful to the team at BPB publications, who had the trust in me for this project. I am thankful to the team of editors and technical reviewers whose insights and suggestions to the manuscript will enrich the reading for the audience.

I thank you, the reader, for making this purchase. I am hopeful you will like the way I have tried to bring the concepts to bare with examples in simple terms.

Preface

The profile of the CISO has evolved over the years. It is not just about protecting information systems but about protecting the future of the organization. In today's world, a CISO has to wear multiple hats - that of a strategist, business enabler, communicator, crisis leader, advisor to the management executives, and more. Boards of organizations are increasingly paying attention and even demanding more from the role of a CISO.

This book outlines this complex role in a practical handbook with clear frameworks, actionable strategies, and lessons from real-world scenarios. These approaches are designed to help a CISO lead with confidence, align the security program to business objectives, and help overcome future threats while being razor sharp on the business acumen and effective communication.

From seasoned CISOs to first timers, this book promises to be a quick reference guide to navigate challenges, make informed decisions, and make your organization's security program an advantage to the business.

The book is divided broadly into three aspects:

a) Functional and domain-related skills

b) Acquiring and applying business acumen to security programs and

c) Communicating them effectively to the various layers in the organization, including the board

Chapter 1: Explains the commonly used terms and concepts in information security, i.e, the triad of security – CIA. These concepts form the building blocks for many of the topics covered in subsequent chapters, and thus, a common understanding will enable you to relate to the topics better.

Chapter 2: Explores cyber / information security risk, including its identification, classification, articulation, and mitigation. Using a lifecycle approach to risk management, the chapter provides you with guidance on how to think about risk in security decisions and, more importantly, who should own them.

Chapter 3: Is focused on the role of standards and controls. As a CISO, you will be required to define and implement organizational policies for the chosen sets of controls and often demonstrate conformance to the applicable/chosen standard(s).

Chapter 4: Covers the principles of privacy and explains their linkages to security controls. The cost of non-conformance is often steep and irreparably damages the reputation of the organization.

Chapter 5: Explores the principles of security and privacy to be applied when designing an application or a process. We examine the similarities between the two and the ways these requirements can be commonly understood.

Chapter 6: Discusses how the information security team understands the technology, its architecture, and thinks through the controls. The CISO and his/her team would frequently encounter some of the common security terminologies and technologies.

Chapter 7: Focuses on the user identities, principles of need-to-know and need-to-have at depth, the common protocols in use, understanding the risks to identities, and how to manage them. These concepts will help the reader protect their information assets against attack techniques.

Chapter 8: Covers the foundations of the cloud, its types, and how to secure the cloud environment. We explore how some of the traditional technologies have evolved in the cloud and how organizations have adapted to the cloud.

Chapter 9: Focuses on the foundations of Zero Trust, its characteristics, and the differences from traditional security models or layered defense. It offers some practical aspects of how the challenges of Zero Trust can be dealt with and the concepts implemented.

Chapter 10: Covers the advancements in technology that have made it easier for cybercriminals to cause harm to organizations and even individuals, especially high-net-worth celebrities. Driven by motives such as financial gains, espionage, mischief, or even simply proving a point, cybercriminals continue to innovate. In recent times, cybercriminals have even operated as organized groups with a specific modus operandi. The CISO and his/her team are required to be aware of these evolving developments and adequately protect their organization from such exposures.

Chapter 11: Is focused on incident management. Several organizations continue to be targets of successful attacks by the adversary. It becomes pertinent for organizations to gain visibility around access and use of their information assets and have mechanisms to detect, contain, and respond to any anomaly as fast as possible. In this chapter, we will explore the concepts of logging and monitoring and their relevance to incident response.

Chapter 12: Focuses on cyber resilience and the importance of having means and measures to bounce back quickly from adverse situations. The aspects of resilience go beyond business continuity and disaster recovery.

Chapter 13: Covers a human centric approach to driving a security-focused culture using some awareness of how the human mind thinks and reacts to an external stimulus. It is an interesting dimension in protecting an information asset.

Chapter 14: Focuses on managing security talent, using a competency framework for upskilling, and some other related aspects. Talent shortage and its management are key dimensions of a CISO's role because not only does the team have to manage current operational program outcomes, but also continually stay abreast and adapt to the changing attack scenarios.

Chapter 15: Brings together the concepts such as the technical controls, principles of security and privacy, Zero Trust, cyber resilience, Threat Intelligence, and security culture, and weaves them into a measurable and the impact of the budget on the program.

Chapter 16: Covers the meaning, purpose, and relevance of business strategy and uses it to derive a possible security strategy for the organization. This approach helps the security programs to be relevant and effective, enabling the security of the organization.

Chapter 17: Focuses on the importance of communication, the mindset of various stakeholders, and the methods of communicating with them. The success of a security program at various stages would require the CISO and the team to use these methods to be effective. The primary stakeholders of the CISO are the CxOs in the organization. We will examine how a CISO can engage with those stakeholders and be successful. Security leaders are often required to present to the board.

Chapter 18: Focuses on providing insights for a CISO to create a compelling narrative of the security program from a board's point of view. It will help to prepare from a board's perspective and bring to bear the things that matter at that level of strategic focus.

Coloured Images

Please follow the link to download the
Coloured Images of the book:

https://rebrand.ly/5bbfdf

We have code bundles from our rich catalogue of books and videos available at **https://github.com/bpbpublications**. Check them out!

Errata

We take immense pride in our work at BPB Publications and follow best practices to ensure the accuracy of our content to provide with an indulging reading experience to our subscribers. Our readers are our mirrors, and we use their inputs to reflect and improve upon human errors, if any, that may have occurred during the publishing processes involved. To let us maintain the quality and help us reach out to any readers who might be having difficulties due to any unforeseen errors, please write to us at :

errata@bpbonline.com

Your support, suggestions and feedbacks are highly appreciated by the BPB Publications' Family.

Instagram *Facebook* *Linkedin* *YouTube*

Get in touch with us at: **business@bpbonline.com** for more details.

Piracy

If you come across any illegal copies of our works in any form on the internet, we would be grateful if you would provide us with the location address or website name. Please contact us at **business@bpbonline.com** with a link to the material.

If you are interested in becoming an author

If there is a topic that you have expertise in, and you are interested in either writing or contributing to a book, please visit **www.bpbonline.com**. We have worked with thousands of developers and tech professionals, just like you, to help them share their insights with the global tech community. You can make a general application, apply for a specific hot topic that we are recruiting an author for, or submit your own idea.

Reviews

Please leave a review. Once you have read and used this book, why not leave a review on the site that you purchased it from? Potential readers can then see and use your unbiased opinion to make purchase decisions. We at BPB can understand what you think about our products, and our authors can see your feedback on their book. Thank you!

For more information about BPB, please visit **www.bpbonline.com**.

Join our Discord space

Join our Discord workspace for latest updates, offers, tech happenings around the world, new releases, and sessions with the authors:

https://discord.bpbonline.com

Table of Contents

CHAPTER 1

The Triad of Security

Introduction

This chapter of the book will begin by explaining the commonly used terms and concepts in information security. These concepts form the building blocks for many of the topics covered in subsequent chapters, and thus, a common understanding will enable you to relate to the topics better. In many places, we will leverage etymology, the study of the origins of words, as a great way to connect the meanings and the interpretations of words better. It helps in remembering the words and their meaning. We may be using the terms data and information interchangeably.

Structure

The chapter covers the following topics:

- Introduction to triad
- Confidentiality
- Integrity
- Availability
- Using CIA to think about security program

Objectives

By the end of this chapter, you will be able to understand and appreciate the differences and the relationships between the three most important concepts called the triad of information security (confidentiality, integrity, and availability). You would be able to align your security strategy (covered in a later chapter) by adjusting the priorities across these three facets.

Introduction to triad

The three (triad) primary facets of information security are **confidentiality, integrity, and availability**, (**CIA**). You will see this acronym referred to many times in this book. Organizations often appoint a **Chief Information Security Officer** (**CISO**) or the head of information security to define and run the security strategy. We will explore the role of CISO a bit later in the book. The core of what CISOs and his/her security team worry about revolves around the CIA. In this book, we will use the terms information security and security to mean the same thing. If we imagine information security to be a stool, then the CIA are its legs. Before we look further into the CIA, we will take a look at some other phrases and terms.

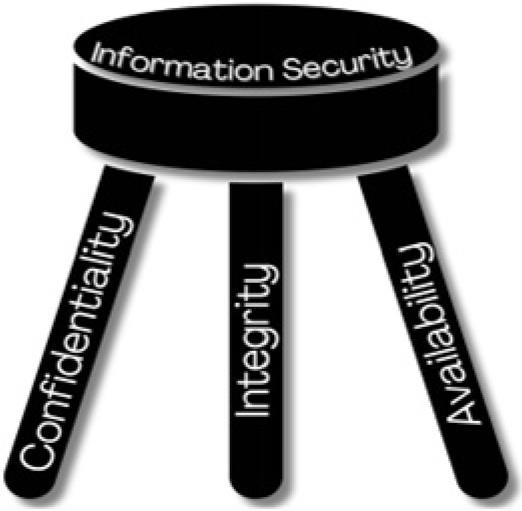


Figure 1.1: *The triad of information security*

Need to know

As simple as it reads, the user/consumer should get access to information/system if they have a solid reason for it. We can call this reason a business justification or simply a justification. The *need-to-know* is determined by the role the user plays, and not by the hierarchy or the position of the user in an organization.

Let us look at some examples:

- As an IT admin, you would need to know all about the servers you manage. The **Chief Executive Officer** (**CEO**) of your company may need to know how much risk is in server environments, but does not need to know the password to that server.
- The **Chief Information Officer** (**CIO**)/**Chief Technology Officer** (**CTO**) may need to know the running status of the critical business servers, but would not necessarily need to know which PowerShell is stored where. The security incident manager, by virtue of the role, would need to know all of the security events, but a senior manager or a vice president in the business teams should not know unless it concerns them or their team directly.

When you share a secret with a friend or colleague in-person, you are in visual contact and you know you are talking to the right person, but in the digital world, where information sharing may be happening via emails and/or chat forums, etc, you would need to establish that the person you share the information is who you believe them to be. In other words, you will rely on the authentication of the identity of the person. We will cover this concept shortly.

Need to have

This principle expands the requirement of needing to know to what you can do with it. It guides the setup of the type of permissions the user must be accorded to achieve his/her business requirement.

For instance, as an IT admin, you may have access to check the configuration of the domain server but not have access to change any configuration. Additionally, based on your role, you may have rights to run PowerShell scripts on several servers, but you may not have rights to upload or download anything from the internet from those servers.

It is important to note that most mature organizations focus on the quality or detail in justification as an auditable record; therefore, reconsider saying *need access* by adding the why to it. *I need access to run PowerShell scripts as part of my role*. It also gives the approver an opportunity to make more informed choices.

Minimum necessary

The concept of *minimum necessary* requires thinking about access to data with as minimal possible access as required. The more access that is irrelevant, the higher the chances of it being compromised.

Let us say you and your family consume 4.5 kg of rice every month. That is your minimum necessary. However, if the grocery store sells only packs of 5 kg, you end up having half a kg extra. That extra grain is likely to also be susceptible to natural rotting, and or ants or food moths attacking them.

In the digital world, as an IT admin, you have a role and need to manage a web server, but you are given permission to manage all web servers. In this case, the minimum necessary principle has not been applied.

Let us take an example from everyday life: how many times have you approved all permissions that an app you downloaded from the Play Store, App Store, or the Microsoft Store? Imagine a calculator app asking you to give permission to all your contacts and SMS on your phone. There is no valid reason for you to allow such permission for an application like a calculator. Think about gaming apps requiring permission to read files from your hard drives.

Tip: Good apps should not require excessive permissions. If they do, you may be able to turn off some permissions even after installing. If not, maybe that is not an app you should use.

Authentication

Authentication is the process of establishing that the user trying to access data/information is the authentic user (from an identity perspective). It is designed to prove that s/he is who s/he is claiming to be, i.e., *who you are*. The most common implementation of authentication is the use of the factors of username and password. The core assumption is that the factor—password is unique, not easy to guess, is set by you, and is not shared with anyone. Only you have your own password.

While native authentication attempts to prove *who you are* (user-id) by using *what you know* (password), multi-factor authentication mechanisms may be needed for important systems to include checks on *what you have*. You are expected to have a device (such as shown in *Figure 1.2*) on your person. This device would have a randomly generated number that you will need to use to gain access. Sometimes this additional authentication can be achieved using many of the popular apps like *Google Authenticator* (*Figure 1.3*), which also shows a unique time-bound and one-time password on your phone:

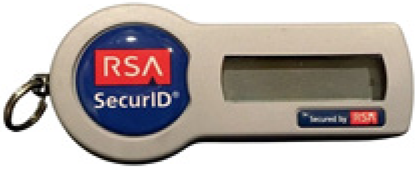
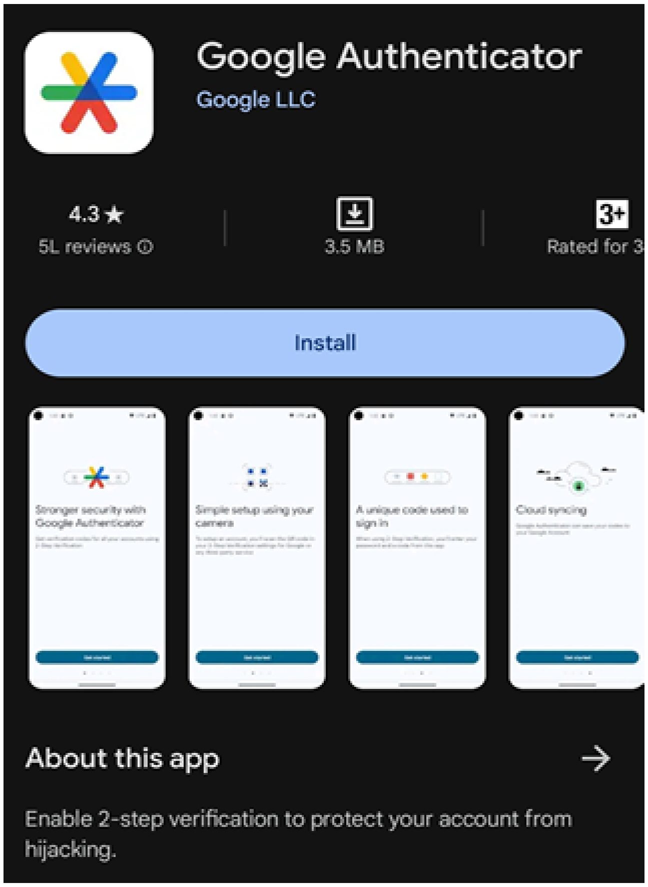


***Figure 1.2**: RSA token*



Google Authenticator

Google LLC

4.3★
5L reviews ⓘ

3.5 MB

3+
Rated for 3

Install

Stronger security with Google Authenticator

Simple setup using your camera

A unique code used to sign in

Cloud syncing

About this app →

Enable 2-step verification to protect your account from hijacking.

***Figure 1.3**: Google Authenticator app*

RSA/Google own the respective trademarks for their brand and products; the image is shown here for illustration purposes only.

Tip: It is estimated that at least 40% of breaches are because of the use of weak authentication methods.

Additional authentication mechanisms may include authenticating with *Something you are*. This factor typically involves using one of the biometric features, like your fingerprint or your iris scan (face ID).

Tip: With advancements in technology, we are slowly moving away from passwords for logging into laptops and desktops using technologies like Windows Hello.

Authorization

Authorization implies the rights/permissions that the users must have on the data to appropriately execute their job responsibilities. This is the *need to have* an aspect. The authorization process would ideally require a formal request from the user with a clear justification. Another approved user (read admin or information owner) may need to take a call to reject or approve the request and then take the required action to provision the access if needed. You may think of authorization as something you have authority over, to see/read, or to even create or modify.

Typically, organizations will have different authorization levels to ensure the information systems/data are duly protected from any unauthorized modification. Some of them are:

- **Read-only**: You can read most of the information, but cannot change anything.
- **Read-write**: You can read, create, modify, or even delete any part of data (or configuration).

Confidentiality

Confidentiality takes its origins from confiding, i.e., sharing and maintaining secrets, especially in a security program. The purpose of the confidentiality leg of the triad is to ensure that corporate secrets, data/information are duly protected from unauthorized access or transmission by anyone. Confidentiality in the digital world is established by requiring a user to authenticate prior to gaining access and then ensuring the access meets the need to have criteria (authorized user). In common parlance, it is said that if a secret exists between more than two people, it may not remain a secret anymore.

Why should one protect the confidentiality of data? Well, some of it may be because you want to protect the business secrets—like the formula of aerated drink, the spice and condiment mix in your hot-selling jar of pickle, or the secret algorithm used in predicting market trends for the stock market.

You may also need to protect some documents/information by law, such as unpublished and unaudited financial reports of a listed company, because a leak of such documents may be used to manipulate the stock market, bringing huge financial implications, reputational damage, and even regulatory issues.

It may also be for more humane reasons—to protect the identity and details of your customers so that you do not become the cause of their agony. Such customer identity information, when in the wrong hands, may be misused and cause real damage to your customers with problems like credit rating, illegally created IDs, and used in nefarious criminal transactions, or even using such IDs to defame someone.

Several laws across the globe talk about protecting the confidentiality of data. Any unauthorized access or disclosure of such information is required to be reported to regulators (like the CERT-IN in India, and the **National Health Services** (**NHS**) for data about nationals of the United Kingdom, or to the **Health and Human Services** (**HHS**) for any healthcare data of American nationals. We will explore the concept of privacy of data in subsequent chapters.

Cybercriminals increasingly attack organizations and break the confidentiality of data/information, they collect it and even sell it.

Integrity

Using etymology, the Integrity of data and information can be defined as absolute truth or completeness, much like a human's integrity is being true and authentic.

Therefore, to preserve the integrity of data, there should be no means by which it can be tampered with. There must be means, also called controls (we will cover controls in subsequent chapters), put in place to inhibit any unauthorized changes to the data.

Interestingly, the data itself may be incorrect. In the context of integrity, we may still treat that data as the absolute truth to protect it from tampering, intentionally or unintentionally. Such incorrect data may be rectified by appropriately authorized personnel/system through an approved process.

Let us say in the retail store you run, the price of the pack of cartons of milk is incorrectly entered as Rs90 vs. Rs72 by the data entry operator at the time of inventory creation. While it looks factually incorrect for the purpose of integrity, this is the data to protect from tampering. Why? Since the authorized personnel entered the data. There must be other means by which identification of this error and its correction by authorized personnel may need to occur; we will discuss them later.

The **point of sale** (**POS**) operator is tasked with just billing the product, applying available discounts, and collecting payment. S/he is not authorized personnel to change the price of a product. If this person manages to gain access and change the price, it would be unauthorized access, and the data would no longer maintain its integrity.

Tip: Need to know/need to have are the principle aspects in CIA.

Again, the *need to have* comes into relevance. The POS operator, in the scheme of things, does not need to have access to change the product pricing. The fact that s/he may be able to have it corrected by the authorized personnel is still possible.

Availability

This principle revolves around making untampered information available to authorized users when they need it, where they need it, as permitted by business. Aka information required for business needs to be *available*.

So why might information become unavailable to an authorized user in the first place? Some examples are as follows:

- The application is not accessible on the Internet, and the user is not in the office.
- The data on the application is corrupted.
- There is a natural calamity inhibiting access.

In all such scenarios, it is pertinent that the company plans its business objectives, and it clearly defines its information needs and creates measures accordingly. For instance, taking regular backups of data can help you restore to a particular time in the past. Or ensuring the right levels of placement of offices in zones that are generally not subject to seismic activity or prone to fires.

Tip: Tape-based backups are now increasingly losing flavor, and cloud-based backups are fast emerging as the mechanism of choice.

Using CIA to think about security program

As a CISO, your role is to define, establish, execute, and continually evolve the security program. The CIA triad gives you a framework to think through the most foundational elements of the building blocks for a security program. The security program is based on controls, a topic we will be going into detail in *Chapter 3, Role of Standards and Controls*. For now, consider controls to be key to the means to achieve CIA. Several controls will remain foundational in all scenarios. Organizations will need to ensure appropriate authentication methods are implemented. However, at times, one leg of the triad may take more precedence over the other. For instance, at a hospital, the availability of the operating digital equipment is more important for saving a patient than having read-only access to the data on that equipment. In the same light, for a beverage manufacturing company, protecting the confidentiality of the recipe is likely to be more important than the need to produce large quantities of beverage, aka availability. Let us take an example where Integrity would take more precedence. For instance, in a bank, the core functioning is based on the exact money in which account and its transfer. The financial institutions would give utmost importance to ensuring the right amounts are entered in all transactions for all accounts.

It is important to appreciate that each organization will have a different focus on one of the legs of the triad, but that does not mean the other two are not relevant or will be ignored. Much like in everyday life, we adapt to scenarios, we flex between our choices of controls. This choice of controls is determined by the strategic objectives of the company and the management direction with respect to risk. We shall cover the aspects of risk in *Chapter 2, About Managing Risks*.

Conclusion

In this chapter, we covered some key terms and explored how they are connected in the security program with examples.

Key takeaways

- The primary role of the CISO and the security team is to protect the **confidentiality, integrity, and availability** (**CIA**) of the information systems while ensuring business objectives are met.
- The following principles are foundational to a security program:
    - Need to know
    - Need to have
    - Minimum necessary
- Authentication and authorization are important factors to achieve the CIA.
- Traditionally, the security team focuses largely on CI of information, while the availability aspects, such as backups, uptime of servers, networks, and redundant telecom connections, are usually the highest focus of **information technology** (**IT**) teams.

Join our Discord space

CHAPTER 2
About Managing Risks

Introduction

In this chapter, we will explore aspects around cyber/information security risk, such as, its identification, classification, articulation, and mitigation. We will walk through a lifecycle of risk management and provide you with guidance on how to think about risk in security decisions, and more importantly, who should own them. We will be using the term cyber risk or information security risks interchangeably.

Structure

The chapter covers the following topics:

- Defining risk
- Types of risks
- Lifecycle of risks
- Roles in risk management
- Using risk management in business practices

Objectives

By the end of this chapter, you will be able to define and implement a risk management process and use it in making sound business decisions. As a CISO, you will be balancing business growth and its requirements, and will need to ensure the right authority in the company is accountable for the risks to those requirements and solutions. Management of risk will involve cross-functional engagement and a method to record and report them. You would also learn how to consider the requirements of a GRC tool, while the product selection itself is out of the scope of this book.

Defining risk

Risk, in our day-to-day life, is things that may bring about harm or danger. For instance, an open electric socket that a toddler may poke his/her finger into and experience an electric shock. Or a speeding truck hitting your vehicle on a highway. Or being in the path of lightning strikes and succumbing to a fatal consequence. Or someone using your credit card for an online purchase without your knowledge.

All of these are possible events, but think about:

- How often are those likely to happen and materialize the danger of electric shock to the toddler, or injury to you, while travelling on the highway, or being struck by lightning, etc.? The answer lies in probability.
- What is the extent of the adverse impact?
- Is there anything you could do to minimize that possibility or remove that chance altogether? For instance, it is easy to make electric sockets child-proof, consider doing so even if your household currently does not have children in that age group. If you are a keen traveler by road, unfortunately, accidents can happen. Your focus should be to drive carefully and ensure you are covered by insurance for the vehicle, bodily damage, and life itself. Similarly, while it is possible to never use a credit or debit card at all, it may not be practical in today’s age, and thus, setting limitations on online purchases is a better alternative.

From the examples above, it would have become apparent that:

- Adverse scenarios exist all around us, and we will continue to live our lives around them. For instance, traveling by road or highway is mostly unavoidable for most of us.
- The chances of a risk happening are a function of probability.

- Some weakness in the environment causes the risk to materialize and results in an adverse/negative impact.
- There are ways to eliminate or minimize the impact of those adverse scenarios.

Risk, at an organization, could be several; let us see some examples:

- Seismic activity is forcing the manufacturing process in the plant to be halted.
- The building was broken into by miscreants, causing damage to the property.
- A short circuit triggering a fire and destroying the finished products kept for shipment to distributors.
- A doctor unable to use the robotic equipment in the middle of a surgery due to a power failure.
- The company's e-commerce website suffers a cyber-attack, resulting in the service not being available for customers.
- Failure to hire or hire the right type of talent, resulting in issues with client deliverables, causing contractual penalties.
- The **Chief Operating Officer** (**CEO**) is unable to access his/her corporate folders containing crucial materials.
- The sales team is unable to log sales numbers via the app at the end of the month.

Much like the day-to-day world, all these risks are possible and may cause loss or damage or both. The nature of loss or damage may be financial, reputational, operational, contractual, and even that of litigations. There is usually something the organization can do to eliminate or reduce the exposure to the weakness and thus the risk.

We will now explore some common definitions of risk from renowned sources to use and apply in our corporate scenarios:

- The **National Institute of Standards and Technology** (**NIST**), a US government body, defines risk as *a function of the likelihood of a given threat-source's exercising a particular potential vulnerability, and the resulting impact of that adverse event on the organization.* They have also articulated that *Information system-related security risks are those risks that arise from the loss of confidentiality, integrity, or availability of information or information systems and reflect the potential adverse* *impacts to organizational operations (including mission, functions, image, or reputation), organizational assets, individuals, other organizations, and the Nation.*

- The **International Organization for Standardization (ISO)** is considered a leading authority and source of standards. We will cover standards in the next chapter. ISO 31000, the standard for risk management, defines risk as the impact of uncertainty on achieving objectives. This standard further clarifies that an impact is any deviation from what is expected. *It can be positive, negative or both, and can address, create or result in opportunities and threats*. And that, *Risk is usually expressed in terms of risk sources (3.4), potential events (3.5), their consequences (3.6) and their likelihood (3.7)*.
- The **International Information Systems Security Certification Consortium (ISC^2)**, a leading and one of the oldest respected cyber security content and certification organization defines *risk as the possibility of damage or harm and the likelihood that damage or harm will be realized.*
- As per the **International Information Systems Security Certification Consortium (ISACA)**, *risk is the combination of the likelihood of an event and its impact.*

Using the learnings above and to ensure we remember CIA well, let us try to adapt a slightly modified definition of risk for an organization as *events that may lead to the loss of confidentiality, integrity, or availability of information or information systems and resulting in adverse impact to its financial position, reputation, contractual obligations, regulatory commitments, and stakeholder interests.* Stakeholders are investors, promoters, employees, contractors, business associates, customers, and suppliers.

Tip: Our digital ecosystems are so connected today that a risk in one organization may lead to risk for others. For instance, in July 2024, a software component update glitch made several thousand computers dysfunctional. An organization that saw the impact ran a risk of loss of revenue and other operational issues.

Types of risks

By now, it would have been clear that risks are all around us. All those risks can be classified into one of the following types to help us think in a more structured manner. Let us also examine the cause-effect to elaborate the types. See *Table 2.1*:

Tip: For a more formal categorization of risk, consider using the ISO:31000 standard.

S. no	Type of risks	Caused by	CIA triad	Effect it may have (impact)
1	Regulatory and compliance risks	Changes in laws and/or their rules. Lack of oversight in meeting legal requirements.	A	Loss of reputation. Monetary setback due to penalties or legal costs. Disruption to operations.
2	Financial risks	Inaccurate accounting processes. Unexpected expenses from an adverse event. For instance, Covid19 required unplanned expenditure on many companies to stay relevant. Disproportionate contractual liability vs revenue.	I, A	Loss of revenue or impact to profits. Increased regulatory scrutiny.
3	Technology or cyber risks	Malfunctioning of information systems. Cyber-attack on information systems. Failure to adapt with changing technology or keeping the environment updated.	C, I, A	Loss of data/information. Inability to run operations. Loss of customer trust.
4	Operational risk	Inaccurate or outdated business processes or systems. Lack of oversight to key business processes. Skill deficit in manpower deployed.	C, I, A	Incorrect work product (or service) delivered. Penalties due to **service level agreement** (**SLA**) breaches. Damage to functioning of technology systems. Reputational damage. Loss of revenue.
5	Environmental risks	Inadequate management of **heat, ventilation and air conditioning** (**HVAC**).	A,I	Disruptions to operations. Damage to technology system. Impact on human lives and/or property. For instance, lack of fire extinguishers.

S. no	Type of risks	Caused by	CIA triad	Effect it may have (impact)
6	Other risks	Intentional human activity such as fraudulent activity. Strikes, vandalism etc. Lack of strategic direction. Geopolitical events such as armed conflict between two nations.	C, I, A	Loss of revenue. Reputational damage. Damage to lives and/or property. Loss of market share. Limitations on access to certain markets.

Table 2.1*: Mapping of risk types to CIA*

In a customer-centric market, a brand built on trust and years of reputation is of paramount importance. Any negative impact on reputation is mostly irreparable and may lead to other types of losses.

Tip: In January 2025, the Government of India released the draft version of the rules for the Digital Personal Data Protection Act (DPDP Act) 2023. According to this version, e-commerce and online gaming companies with significant users would need to maintain data/logs for at least 3 years. This change requires the companies to evaluate and prepare for compliance, which might require cost and time.

Tip: Geopolitical changes may drive unexpected and unplanned market changes. In early 2020, when the COVID-19 virus started raging around the world, a lot of new demand for semiconductor chipsets was generated to cater to the needs of hybrid workforces and devices like oxygen concentrators. This situation led to additional stress on the manufacturing capacity in Taiwan, which was already in distress due to resources like water. Equipment manufacturers had to quickly find alternate capabilities in other geographies to reduce their operational risk.

Note: In *Table 2.1*, the causes listed in the column are events, circumstances, or situations that cause the harm (i.e., impact). These causes are also referred to as threats. We will cover threats in detail in subsequent chapters. For now, remember that the threat is the agent that causes the risk to materialize.

Before we explore this risk management lifecycle, we will need to be clear on the scope of our **risk management** (**RM**). You would want to know the most important things, such as key information assets of your business, that you should worry about and focus on applying the RM lifecycle approach to. Like they say, if many things are a priority, none is. For instance, the inventory of all your information assets is, say, 100 assets. You would club/categorize these into groups of similar functionalities, like all customer application servers into one group, all employee-facing services like payroll application into another, and so on. Then you can apply the risk management lifecycle to those groups of information assets.

Tip: Risk assessment (RA) is the process of identifying and assessing risk, and risk management

encompasses the process of RA and furthers the efforts to reduce the risks.

Lifecycle of risks

As risks exist all around us, we must find ways to eliminate or reduce risks to meet our objectives. Risk management is defined as a process to identify, analyze, prioritize, treat/mitigate, and continually govern them to make informed decisions. We can represent the **risk management** (**RM**) lifecycle as shown in the following figure:

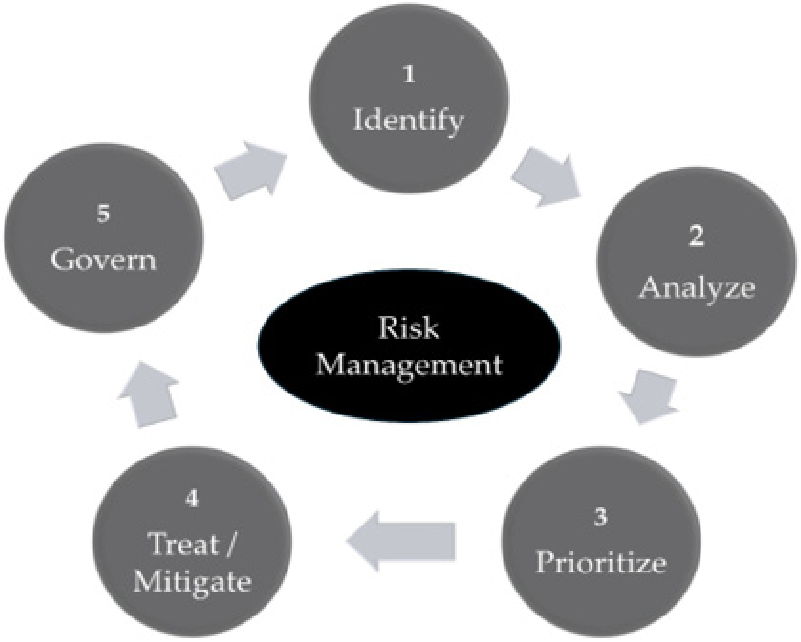


Figure 2.1: *Lifecycle of risk management*

The explanation is as follows:

- **Identify risks**: As we begin the process of identifying risks to the groups of assets, we should reflect on the business objectives and determine which group of information asset(s) are key to meeting those objectives. All the information assets would seem important, but each of them will not be as critical. Therefore, a conscious effort must be made to identify those critical assets and identify risks to those assets. For instance, as an e-commerce website company, you would want to make sure the infrastructure delivering those services is most critical to your organization in comparison with the information asset group of user laptops or desktops.

Each information system will always have some risks, irrespective of how good it is. This is called an **inherent risk**.

Identified risks are formally documented in the **risk register.**

What could be some of the factors that can compromise or cause harm to the CIA of our critical asset, the e-commerce website:

- **Denial of service** (**DoS**) attack on the web server, causing it to crash and become unavailable.
- The e-commerce website is defaced with meaningless content, and the links on the page are disabled.
- The backend infrastructure that processes the payment by credit cards or innovative technologies like **Unified Payments Interface** (**UPI**) is unable to take any requests.
- Malware or viruses infect the servers hosting the e-commerce website from an internal network.
- An IT admin incorrectly changes the configuration of the web server, causing it to stop working.
- The telecom links that bring the web server to the internet are dysfunctional.
- A disgruntled employee physically switches off the server in the data center or powers down the server hosted in the cloud.

Tip: Organizations generally require the owner of the information asset to start identifying r
to his/her information asset. In our example, the owner of the e-commerce website service le
shall be responsible for identifying initial sets of risks. This is because they are the most awar
the relevance of each part of the process and the systems that support it. Thus, the informa
asset owner can articulate what could go wrong. These efforts are supported by risk experts
risk assessors.

Generally, a common template is used across the company to assess risks. This may be a separate file or part of the risk register. The risk assessor will augment the initial risks shared with more insightful aspects from various occurrences in the industry. For instance, the risk assessor may add the unpatched software vulnerabilities as a risk to the e-commerce website going down, as s/he is aware of the increasing trend of attackers capitalizing on vulnerabilities in web server software.

- **Analyze**: As risk assessors, our next steps are to determine the impact of that risk and how probable it is to occur. Our analysis may be backed with historical data and thus allow us to quantify the risk in monetary terms, or it may be a subjective judgement of the risk assessor. We may also choose to do a qualitative assessment based on professional judgment. In qualitative risk analysis, you would rate the risk in terms of scales like very high, high, medium, low, and very low. It is important to apply a consistent approach to qualitative analysis. For instance, for the same risk for an asset of similar criticality, the determination of impact and probability should not be dissimilar. If such a dissimilarity arises because of some unique conditions, then the risk assessor must document that too.

Tip: It is difficult and impractical to quantify each risk, and thus it is recommended to incorpo both qualitative and quantitative views into the risk analysis.

- **Prioritize**: During the process of risk analysis, a host of risks would have emerged. To handle each risk, it would take time, effort, and cost. It is thus important to define and use a mechanism to prioritize those risks. There may be some risk that remains below a threshold defined by the organization; such an outstanding risk is called **residual risk**. And this threshold is called **risk appetite**. For instance, the e-commerce company may decide that any risk that is very low and/or has an impact of less than $10,000 would not need any additional efforts to manage them. We would start by considering all risks above the acceptable limits. While it may be smart to handle high risks first, it may also be efficient to examine and see if there are any risks, especially low ones, that can also be treated/mitigated as quick wins.
- **Treat/Mitigate**: The actions we take to reduce risk are called **risk treatment** or risk mitigation. The word mitigate means to soften or reduce in degree or measure. The plan to mitigate risk is called a **risk treatment plan** (**RTP**). The plan should be documented as it serves as the guiding light for the actions to take and the record of decisions taken to reduce the risk. Elimination of risk is nearly impossible; we can only reduce it to acceptable levels. The decisions taken to treat/mitigate risks are called **risk decisions**. The following are broad means by which risks are treated.
    - **Avoid**: Risks can be avoided by eliminating the cause of the hazard itself. For instance, if an e-commerce company determines a business risk in accepting payments from a particular issuer of credit cards, it can easily block accepting cards on the website. Similarly, if the company feels it cannot meet the stringent requirements of a particular

region, it can choose not to expand its business into that region. It must be noted that you are not avoiding the risk by ignoring it; you are avoiding the possibility of a cause materializing.

Tip: Countries operate their critical infrastructure, such as nuclear power plants, space programs, and utility systems like power generation and distribution. In order to avoid the risk of cyber threats from the internet, these critical environments are generally not connected to the internet, thereby avoiding the impact.

- **Remediate**: Remediation is the steps or efforts taken to reduce the possibility and/or impact of a cause (weakness). Remedial actions may imply deployment of some protection technologies, and/or changes to processes, and/or ensuring the right skilled manpower is available to execute and maintain the actions. This costs money and time. Mitigating risk generally implies taking remediation steps. For instance, for the risk of a web server being susceptible to compromise due to a software vulnerability, you can remediate this risk by having a strong patch management process.
- **Transfer:** An important strategy for risk mitigation is risk transfer. When you transfer the risk, you are not transferring your accountability; you are just transferring the risk to another organization. For instance, taking an insurance policy for decisions of key personnel can help limit the cost of key decisions that might run into trouble with regulators. Some of the risks may be transferred using indemnification clauses in contracts. For instance, e-commerce companies may sign a contract to indemnify themselves in the event of a pandemic breaking out or indemnify against cyber-attacks originating from specific countries.
- **Risk accept**: Risks that are open would need to be accepted by the risk owner. Generally, a risk owner is the owner of the information asset and is responsible for managing the CIA of that asset. A record of the risk acceptance must be kept. Risks would be open in the following scenarios:
  - Risk remediation will take time to implement fully.
  - Even after implementing all the risk remediation actions and/or having transferred via insurance or indemnification.
  - When the open risk is below acceptable levels.
  - When a requirement is technically not feasible to implement, for

instance, a minimum of 9-character alphanumeric password may not be possible on network devices, only a numeric pin is possible. Such deviations from requirements are also called exceptions.

Refer to *Table 2.2* for examples of risk and their RTPs:

S. no	Identified risks	Some suggested RTP
1	DoS attack on the webserver causing it to crash and becoming unavailable.	**Remediate**: By configuring protection for DoS attacks. **Transfer**: Buy Insurance to cover disruption due to cyber-attacks such as DoS.
2	The e-commerce website is defaced with meaningless content and the links on the page disabled.	**Remediate**: By implementing mechanisms to disallow any changes to the webserver content. Have relevant periodic backups to use for restoring to last published status.
3	An IT admin incorrectly changes configuration of the webserver causing it to stop working.	**Remediate**: By implementing a strong change management process.
4	A disgruntled employee physically switches off the server in the data center or powers down the server hosted in the cloud.	**Remediate**: By implementing a strong security process for permitting access to physical data center or cloud console. Ensure personnel given such access is monitored.

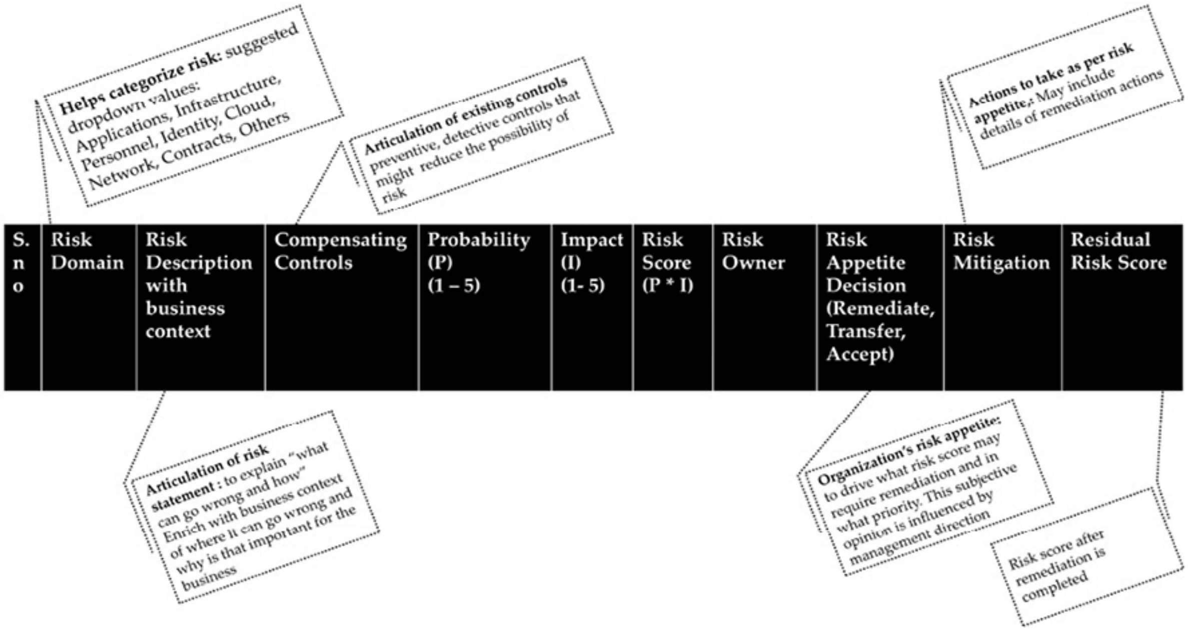
***Table 2.2**: Risks and RTPs*

The execution of the RTPs will need subject matter experts in teams such as technology, legal teams, human resources, and so on. Just that the risk owner is accountable for having the issues remediated. Organizations may choose to have the risk assessor empowered to drive the execution of RTP as his/her responsibility.

- **Govern:** Risks continue to be influenced by changes to strategy, business plans, technology, regulations, or even personnel. Therefore, we should periodically revisit the risk and reassess its relevance, probability, and impact. For instance, let us assume that the risk assessment at the start of the year concluded that email-based threats are less risky for you. At the present juncture, you observe that with the growth of business, more and more of your employees have started using email systems for a variety of business communications. You may now need to reassess and might conclude that, because more users are now a potential target for an attack, the risk to the information system is now higher. Organizations define the levels of risk appetite and their approving authorities. For instance, all very high risks would need to be approved by the CEO, but all medium risks may be approved by the information asset owner. A Risk Officer is sometimes appointed to ensure the overall process is consistent and has oversight of all risks irrespective of the owner. The CISO is responsible for all cybersecurity risks.

The documentation of risk at various stages is the key to success. The documentation helps in making sure appropriate decisions are being reviewed, are consistent, relevant, and reasonably comprehensive, and are in line with managing business objectives appropriately.

Figure 2.2 illustrates an example of a risk register:



S. no	Risk Domain	Risk Description with business context	Compensating Controls	Probability (P) (1 – 5)	Impact (I) (1- 5)	Risk Score (P * I)	Risk Owner	Risk Appetite Decision (Remediate, Transfer, Accept)	Risk Mitigation	Residual Risk Score

Helps categorize risk: suggested dropdown values: Applications, Infrastructure, Personnel, Identity, Cloud, Network, Contracts, Others

Articulation of existing controls preventive, detective controls that might reduce the possibility of risk

Actions to take as per risk appetite,: May include details of remediation actions

Articulation of risk statement : to explain "what can go wrong and how" Enrich with business context of where it can go wrong and why is that important for the business

Organization's risk appetite: to drive what risk score may require remediation and in what priority. This subjective opinion is influenced by management direction

Risk score after remediation is completed

***Figure 2.2**: Example of a risk register format*

During the process of an assessment, issues/findings may be identified. For instance, the e-commerce application uses weak methods of authentication, making the application unsafe. Such issues should be captured in an appropriate tracker.

Roles in risk management

Risk management is a team sport; many people must come together to collectively work on it. Each person involved in the process brings a perspective and capability to the process. In the previous topic, we have defined several such roles. See *Figure 2.3* to see those roles and their accountability in the lifecycle:

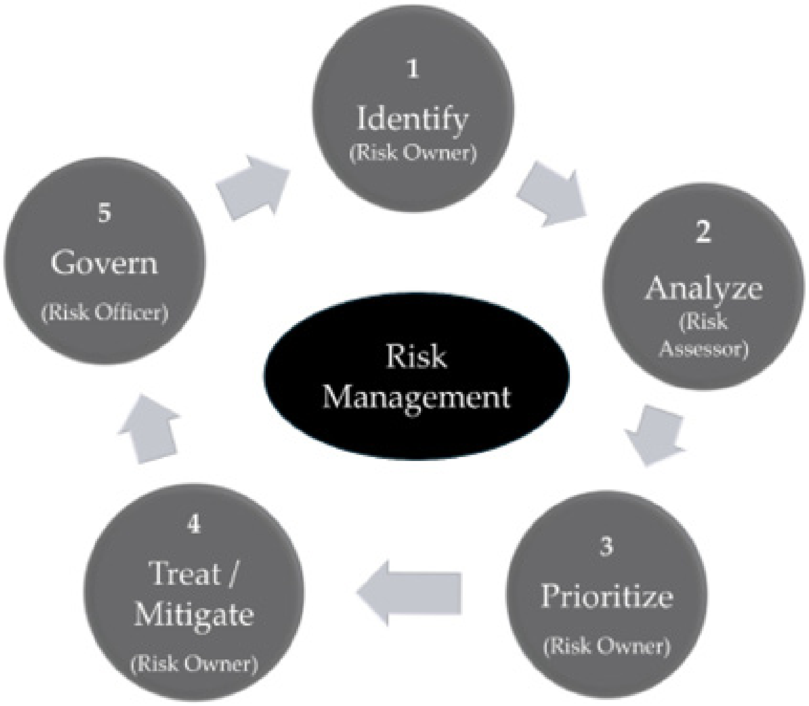


***Figure 2.3**: Roles in risk management lifecycle*

Using risk management in business practices

While we have examined several business examples throughout the chapter, it is essential to remember:

- The risk management process helps organizations prevent untoward events and prepare for adversities.
- RA serves as a key tool to make even proper and informed business decisions. The goal of risk management is to have the business make informed decisions. If the risk outweighs the benefits, the business must ideally refrain from overriding the risk.
- Mature organizations use the process to ensure their business objectives are met and there are no unknown unknowns. RA should have already determined any unknowns, and RTPs would exist for those.
- An organization must choose the RA methodology, define its risk appetite, and identify the risk approving authorities.

- Top management engagement is critical to the success of RA and must diligently support the RA efforts, even if the CISO is held accountable for the cybersecurity risks.

Using governance risk and compliance tools

Risk management is quite an intensive effort involving several people in the organization at various stages. Documentation at each stage is important. To assist and aid in making the process of the RA itself simple and easy to do, there are several tools. These tools are called **governance risk and compliance** (**GRC**) tools because they allow you to govern the risks and compliance requirements. A GRC tool may be used to aid in:

- Library of common risks and their impact.
- Management of risk registers with appropriate access controls.
- Leveraging out-of-the-box risk scenarios, for instance, if there is an internet-facing information asset, the risks from the external world can be automatically populated saving time.
- Common repository of risk artifacts to collaborate during various phases of RM lifecycle.
- Automatic workflows and email notification capabilities for risk lifecycle phases.
- Tracking some compliance requirements and reporting on them, such as the ISO27001 standard. We cover some of the standards in the subsequent chapters.
- Aggregation and analytics for top management, for instance:
    - Which business unit has the most risks.
    - What the top 3 risks.
    - During (aging) a risk is open.
    - Risks due for review based on time, renewal condition, etc.

Each GRC tool will have some features better than the other; it is upon the organization to choose the one that fits best in terms of organizational purpose, flexibility of use, and cost. Some of the popular GRC tools are:

- Archer
- ServiceNow
- MetricStream

- Jira
- Microsoft Excel and similar spreadsheet tools are commonly used as well.

Tip: It is strongly recommended to use the out-of-the-box features of the GRC tools as much as possible and not do unnecessary customizations. You should consider tweaking your manual process instead if needed to align with the tool.

Conclusion

In this chapter, we explored several concepts of risk using examples. A core part of a CISO's job involves managing cyber risk. You would now be able to think across various types of risks, prepare ways to mitigate them, and govern this process of assessment periodically. *Figure 2.4* summarizes some of the key concepts you have learned:

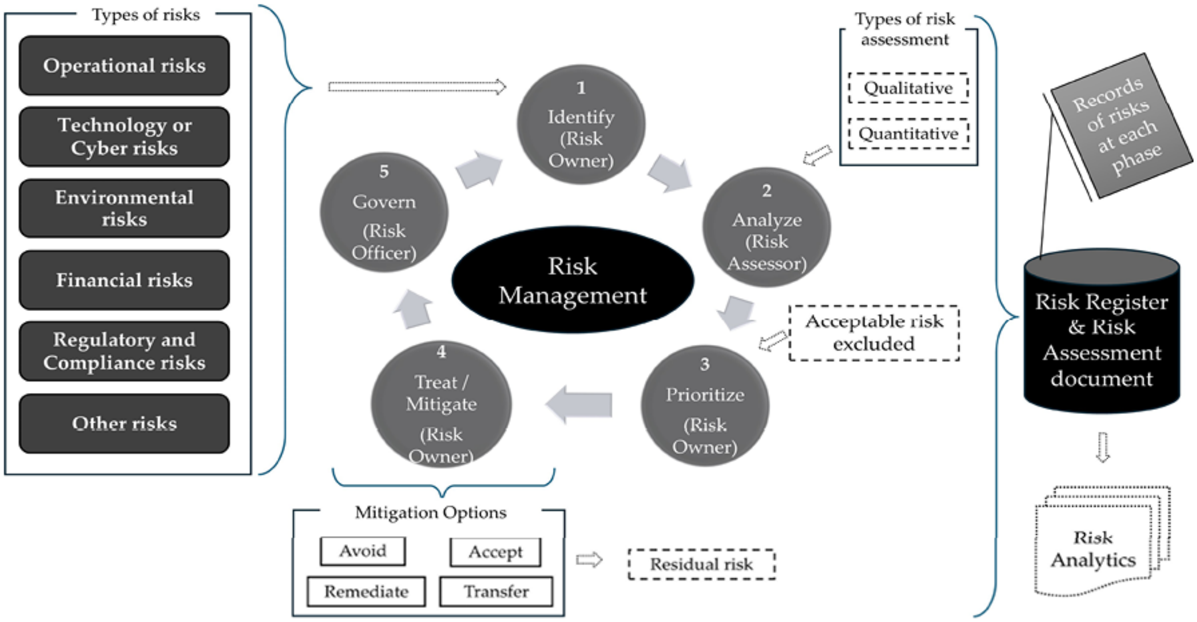


***Figure 2.4**: Composite view of the RM*

We will be using the risk concepts covered in this chapter throughout the book. In the next chapter, we cover standards and controls. It will help you in thinking about the journey from your current state of the security program to its future. We will explore the controls from the people, process, and technology angles.

Key takeaways

Some of the key takeaways from this chapter are:

- Risk can be considered as a loss of CIA of information systems and a potential impact on an organization's mission, objectives, and operations, etc.
- The information security team oversees cyber risk and enables the appropriate stakeholders to make informed decisions in managing those risks by periodically performing RA.
- Issues determined in risk assessments are aggregated and analyzed, where patterns of risks may emerge.
- Such issues and risks must be mitigated holistically.
- Risks may remain even after several measures have been taken to mitigate them.
- Organizations should consider formally defining their risk appetite to use in their risk decisions.
- Records of risk should be maintained in a risk register using any suitable commercial and free tools.
- An organization may have specific processes and assigned executives for managing risks, such as a CRO.

References

- **https://csrc.nist.gov/glossary/term/risk**
- **https://www.iso.org/obp/ui/en/#iso:std:iso:31000:ed-2:v1:en**
- **https://www.isc2.org/certifications/cissp/cissp-student-glossary#r**
- **https://www.isaca.org/resources/glossary#glosss**
- **https://www.crowdstrike.com/en-us/blog/falcon-sensor-issue-use-to-target-crowdstrike-customers/**
- **https://static.mygov.in/innovateindia/2025/01/03/mygov-999999999568142946.pdf**

Join our Discord space

CHAPTER 3
Role of Standards and Controls

Introduction

In this chapter, we will explore using information security standards to define the controls we need to manage the risks. We will walk through a lifecycle of controls and provide you with guidance on how to think about maturing the security program periodically. We will often be referring to concepts of risk management.

Structure

The chapter covers the following topics:

- Exploring information security standards
- About controls and their lifecycle
- About the NIST Cybersecurity Framework
- About policy and its governance

Objectives

By the end of this chapter, you will be able to understand some of the most commonly used information security standards and their applications. You will be able to use them to define, implement, or augment controls to reduce risks to your business. As a CISO, you will be required to define and implement organizational policies for the chosen sets of controls and often demonstrate conformance to the applicable/chosen standard(s).

Exploring information security standards

In the metric system, a kilometer is exactly 1000 meters, and 1 liter is 1000 milliliters anywhere in the world. That is a standard. We can think of a standard as something that is expected to meet certain norms, predictable dimensions/reliable quality specifications. A standard contains minimum requirements and their specifications. An information security standard specifies the minimum and desired requirements for CIA. It may cover aspects of privacy, a topic we shall explore later in this book. Generally, standards are defined and designed by **subject matter experts** (**SME**) and thought leaders to ensure consistency of performance or reliability of execution that can be expected from a product or service. The reliability of the conformance to a standard comes from an independent external assessment to certify the implementation of such requirements. We introduced a few standards in *Chapter 2, About Managing Risks.*

The genesis of a standard may arise from:

- **Regulations** or legislations (laws) passed by various national or even state(provincial) governments.
    - The **Health Insurance Portability and Accountability Act** (**HIPAA**) of the **United States of America** (**US/USA**) requires healthcare organizations to handle patient healthcare data and breach notification requirements. Industry had to define mechanisms/means to meet compliance with this law. The US healthcare system adopted a voluntary framework called the **Health Information Trust Alliance** (**HITRUST**) for securely handling patient care-related data, called **protected health information** (**PHI**). Over some years, HITRUST has positioned itself as a framework for other industries too to protect confidential information about individuals called **personally identifiable information** (**PII**).
    - The **European Union** (**EU**) came out with a legislation called the **General Data Protection Regulation** (**GDPR**) in April 2016 to protect the privacy of citizens of the European Union. It is widely seen as a primary legislative direction that spells the requirements for

protecting the PII, but it has very specific requirements for data breach notification.

  - The **California Consumer Privacy Act** (**CCPA**) is a state law requiring the protection and privacy of consumer PII. Much of the law's requirements are focused on privacy principles. We shall explore these principles later in the book, but please consider that these principles also translate into some CIA requirements

- The **needs of an industry** to have common minimum requirements for its own use cases, such as handling a specific type of data to protect against financial fraud. For instance, financial institutions require the credit or debit cards they issue to be used safely without compromising the integrity of transactions. Thus, the consortium of financial institutions, aided by experts, created the ***Payment Card Industry Data Security Standard (PCI DSS)***. This is a voluntary standard, but interestingly, it is mandated by the industry for any organization handling card data. There are other PCI standards, see *Table 3.1*:

S. no	Some types of PCI Standards	Scope	Applicability
1	**Personal identification number (PIN) security**	Secure management and use of PIN data (cryptographic data) during any payment card use by online or offline means.	Whenever any organization is processing payment cards.
2	**Mobile Payments on COTS (MPoC)**	Securing any contactless payment made. For instance, when you tap the card at a **point of sale** (**POS**) your card data travels without you having to enter a PIN in contactless fashion.	Contactless acceptance of payments.
3	**PCI DSS**	It provides guidelines around protecting **card holder's data** (**CHD**), also called PCI data. Name, card number, its expiry and the **card verification value** (**CVV**) are all part of CHD. Every part of the environment that handles PCI data constitutes the **Card Data Environment** (**CDE**).	This standard applies to all organizations involved in a financial transaction involving cards.

***Table 3.1**: Types of PCI standards*

- Guidance from Government bodies is issued to all government institutions/private institutions under its jurisdiction to protect their citizens' PII or PHI.

  - In the US, the **National Institute of Standards and Technology**

    (**NIST**), through its **Special Publications** (**SP**), issues guidance to protect its citizens' PII/PHI, etc. While this standard is a mandatory requirement for US Federal government organizations, the industry has adopted it as an important guiding light. One benefit is that organizations working with US Federal organizations can demonstrate compliance with the same standard. We shall explore another important NIST framework called the **Cyber Security Framework** (**CSF**) later in this chapter.

  - The **Reserve Bank of India** (**RBI**), the central bank with oversight of banking institutions in India, mandated in 2016 that all commercial banks operating in India should define and implement a robust **cybersecurity and resilience framework** for adequate cybersecurity preparedness (against cyber-attacks).
  - The **Indian Computer Emergency Response Team** (**CERT-in**) issued *Information Security Practices for Government Entities* in 2023. The guidelines are applicable to all government organizations and departments in India. It is likely to be adopted voluntarily (at least in parts) by private organizations in India too because the requirements are general best practices and even specify expectations from a CISO (of government bodies).

- **Independent organizations** bring in best practices/benchmarks that are globally accepted.

  - The **International Organization for Standardization** (**ISO**) and the **International Electrotechnical Commission** (**IEC**) are non-government organizations that assimilate experts from the world over to bring out commonly accepted standards on a variety of topics. The IEC focuses on all electro-technical topics that include information security and works alongside ISO to jointly bring out some of the standards. We learnt about the ISO 31000 standard on risk management in *Chapter 2*, *About Managing Risks*. ISO/IEC has a standard on information security/cybersecurity and privacy protection referred to as **ISO/IEC 27001**. The latest publication of this standard came out in 2022 and hence is notated as ISO/IEC 27001:2022. The complementary standard that has the details of the controls to be implemented is notated as ISO/IEC 27002:2022.

Tip: ISO/IEC 27001, commonly called 27001 or ISO 27001, is one of the most widely accep

standards on information security management systems (ISMS) in about 150 countries and ac all sectors such as manufacturing, agriculture, professional services, financial institutions, an on.

Tip: Artificial intelligence (AI) started gaining immense popularity in late 2022, with an estima 250 million users in January 2025. The ISO/IEC 42001 specifies the requirements and prov guidance for establishing, implementing, maintaining, and continually improving an management system.

Please note that ISO 27001 is the certifiable standard, meaning the assessment and its independent third-party evaluation can happen against this standard. ISO 27002 defines how to implement the requirements.

- The **Association of International Certified Professional Accountants** (**AICPA**) has published a few standards.
  - **System and Organization Controls** (**SOC**) uses 5 **Trust Services Criteria** (**TSC**) to evaluate a service organization's approach to internal security, confidentiality, and privacy. The 5 TSCs include CIA, security, and privacy. We shall cover security and privacy principles later in the book. The **Statements on Standards for Attestation Engagements** (**SSAE**) details the rules for the review of internal setup (also called controls) and the issuance of attestation in the form of the SOC 2 reports.

Tip: The reporting framework now known as System and Organization Controls (SOC) formerly referred to as Service Organization Control.

There are three varieties of SOC reports, see *Table 3.2*:

S. No	Name of the report	Scope	Used for	Standard used in evaluation
1	SOC 1	**Internal control for financial reporting** (**ICFR**).	Independent assurance of responsible and accurate financial operations and its reporting.	ISAE 18
2	SOC 2	Information security controls employed to protect organizational and customer data.	Certification by an independent assessor.	ISAE 18
3	SOC 3	Information security controls employed to protect organizational and customer data.	Largely marketing purpose as it barely has any details and thus is good for communicating to public.	Generally audited for only the ISAE 18 - Security Trust Principle.

		Type 2 only		

***Table 3.2**: Varieties of SOC reports*

SOC reports are of two types.

- **Type 1**: A test of design at a particular point in time. It only covers what methods are used in protecting the organizational and customer data.
- **Type 2**: A test of the effectiveness of controls over a period of time, typically 6 months.

Tip: A SOC 2 Type 2 has confidential information about the organization's implementation of security controls and how effectively they work. It is, therefore, shared only with stakeholders and prospective and/or current customers where a non-disclosure agreement (NDA) exists. In the wrong hands, this report can be misused by attackers.

Sometimes, such control requirements are issued/published as a guideline; for the purpose of this book, we will treat them as a standard. A standard required to be met because of legislation or a government mandate is called a **compulsory standard**; others are **voluntary standards**.

Let us examine some of the differences between the two types of standards in *Table 3.3*:

S. no	Type of standard	Examples	Implications of non-compliance	Methods to prove compliance/conformance
1	Compulsory	HIPAA GDPR CCPA RBI CSF	Penalties imposed by government. Possibility of further legal action. Loss of reputation and/or business.	Documentation of risk decisions. Regulatory filings at a set frequency/on-demand.
2	Voluntary	ISO 27001 HITRUST PCI DSS SOC	Loss of business.	Certification by an independent assessor.

***Table 3.3**: Differentiating the types of standards*

For the remainder of the book, we will refer to information security/cybersecurity standards and/or guidelines as **standard(s)** interchangeably.

Tip: Some voluntary standards may be treated as mandatory in business operations. For instance, US-listed companies often require their service providers, i.e., organizations to which they outsource any part(s) of their business operations or functions, to furnish a periodic SOC 2 Type 2.

Structure of standards

Every standard is built differently. However, the means to the end of protecting the CIA remain the same. We will look at the standard's general composition. A standard focuses on its security requirements and spells out the objectives for it, called the **control objective**. For ease of alignment and logical grouping, each control objective is further clustered under **domains** and/or **subdomains**. Each objective in a standard has one or more requirements, called **control**. A control is specified using a **control statement**. Each control statement may also specify the required configuration, called the **control configuration element**. See *Figure 3.1*:

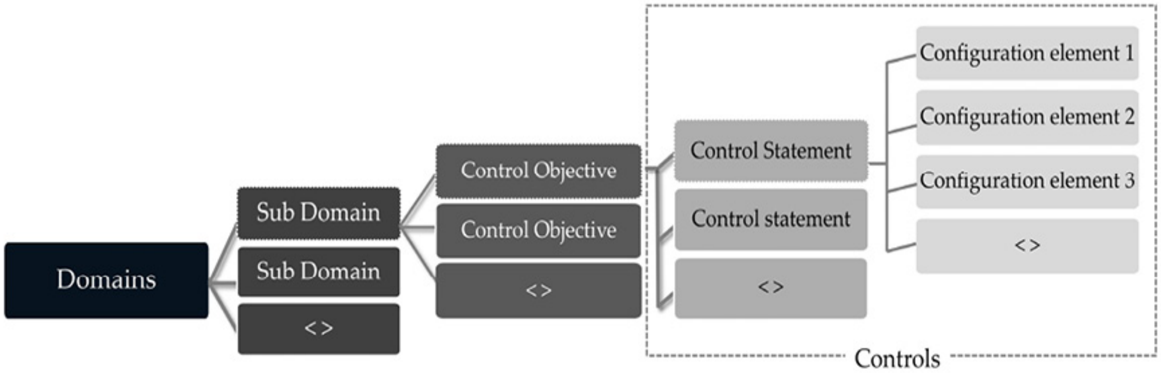


***Figure 3.1**: Structure of a standard*

About controls and their lifecycle

A control can be classified as:

- **Administrative controls**: Pertains to process, policy, and any other means to reduce risk, generally based on user behavior. For instance, providing training to upgrade a server to avoid any human errors.
- **Technical controls**: Pertains to the use of technology to meet the security objective. For instance, implementing authorization restrictions on only specific people to upgrade servers.
- **Physical controls**: Pertains to protecting the physical environment itself, such as restricting buildings from unauthorized entry, protecting against fire by using fire retarding construction materials, etc. For instance, security personnel manning the door to the data center and disallowing any unapproved entry.

  Further, based on the nature of controls or their purpose, the controls can be of the following types:

- **Preventative**: A control that prevents the risk from happening or causing damage. For instance, using **multi-factor authentication** (**MFA**) to protect from unauthorized authentication. We covered MFA briefly in *Chapter 1*, *The Triad of Security*.
- **Detective**: A control that detects any deviation from expected behavior/routine. For instance, a motion detection alarm is installed in sensitive areas of a bank's locker room.
- **Corrective**: A control that triggers a change to remediate an observed gap. For instance, patching a server when a new applicable vulnerability is announced.

See *Table 3.4* for examples and to learn more about controls:

Tip: Some controls only deter the risk from happening but do not eliminate the possibility of risk. For instance, the camera monitoring entry/exit to a building serves as a deterrent for unauthorized people to attempt an entry, but does not prevent them from making the attempt. Such deterrent controls are generally detective in nature.

Preventative		Nature of controls		
		Detective	Corrective	
Categories of controls	Administrative	SOP for server patching. Policy on password length, complexity and age. Performing pre-employment **background checks** (**BGC**).	Process to send patch update logs for periodic review. Check on password policy configuration.	SOP to identify missed patches and prioritize them. Coaching users with weak passwords to set stronger passwords. Disciplinary process to handle unfavorable **BGC** results.
	Technical	Limiting access for upgrading a server.	Logging access made to a server and monitoring logs for unauthorized accesses.	Deploying patches for the identified vulnerabilities. Disabling access for personnel with unfavorable BGC.
	Physical	Restricting physical access to the server room/data center.	Review access logs to server room/data center and determine any unauthorized access.	Deploying motion detection cameras to alert security guards.

***Table 3.4**: Examples of controls*

From *Table 3.4*, you can infer that more than one control may be applied to meet a control objective, and those may be in different categories of controls, but shall be complementary to each other. For instance, to meet the objective of limiting access to personnel with favorable background the following may have to be considered:

- People (such as BGC personnel).
- Process (such as a disciplinary process for unfavorable BGC—a corrective administrative control).
- Technology (such as disabling access for personnel with unfavorable BGC).

Further, across aspects of people, process, and technology, the control itself may be preventative, detective, or corrective in nature.

Tip: Security events related logs, such as authentication attempts to servers, are usually sent to a common log collation system. A team called the security operations center (SOC) team monitors such logs and uses technology to determine any anomalies that might indicate any risk to the CIA. Such teams may be called the cyber defense center (CDC).

Tip: The SOC team, pronounced as sock, is focused on the prevention and detection of security weaknesses. This SOC team must not be confused with the AICPA-prescribed report—SOC, which is also pronounced as sock. The report has a trailing number 1,2, or 3 and is used in the pronunciation: SOC 1 (sock 1), SOC2 (sock 2), and SOC 3 (sock 3). To avoid confusion, SOC reports are sometimes pronounced as ess-oh-see, such as SOC 2 as ess-oh-see two.

We will use *Table 3.5* to understand the typical control domains and their focus areas, irrespective of how each standard may have defined them. We will explore some of these domains in subsequent chapters of the book.

S. No	Control domains	Focus area	Control examples
1	Application security	Design, develop, deploy and/or maintain security of applications.	Security testing of source code.
2	Cloud security	Secure the cloud infrastructure, applications and services.	Auditing cloud configuration.
3	Data security and data privacy	Protect the CIA of data stored in any form and ensure privacy requirements are met.	Applying need-to-know, need-to-have (covered in *Chapter 1, The Triad of Security*).

S. No	Control domains	Focus area	Control examples
4	**Identity and access management (IAM)**	Manage identities and their access permissions including authentication and authorization requirements.	Complex password requirements.
5	Incident management and response	Program to detect, prevent and respond to security events.	Create and test an **incident response plan** (**IRP**).
6	**Information technology** (**IT**) security	Security of IT infrastructure with appropriate configurations.	Deploying anti-malware and disabling access to USB storage.
7	Legal and compliance	Review and inclusion of applicable regulatory requirements to ensure compliance.	Perform internal assessment to meet ISO 27001 requirements.
8	Personnel security	Define and review of controls applicable to personnel.	Pre-Employment background checks.
9	Physical security	Deploy physical access and building safety related controls.	Deploying and monitoring camera at building perimeter.
10	**Risk management** (**RM**)	Define and implement risk management lifecycle and related controls.	Risk Management process and templates.
11	Resilience, **business continuity planning** (**BCP**) and **disaster recovery** (**DR**)	Preparing to bounce back quickly from any untoward security event and enable continuity of business.	Resilience program for critical business functions and IT infrastructure needed. DR Plan.
12	Security architecture and engineering	Oversight on changes being planned due to new technology, and/or modified process.	**Change Advisory Board** (**CAB**) to examine and review changes.
13	Security operations/Defensive security	Ensure security logs are assimilated, collated and analyzed for any anomaly and corrective/preventive action is taken.	Isolating systems impacted with malware.
14	**Threat and vulnerability management** (**TVM**)	Identify threats and vulnerabilities in the business environment that bring risks to the CIA of information.	Implement a **vulnerability management** (**VM**) program. Review **threat intelligence** (**TI**) feeds and monitor for any anomalies or **indicators of compromise** (**IOC**).

Table 3.5: Typical domains of controls

Review the *Figure 3.2* to understand the control requirements with an example. To expand on the control statements and their configuration:

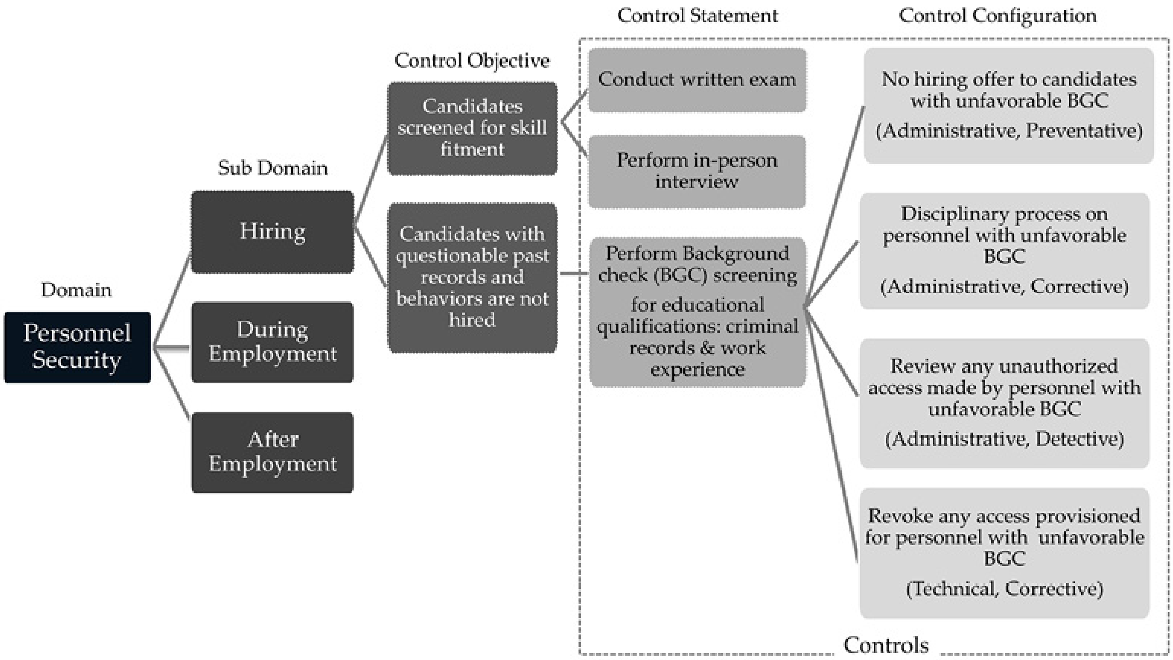

Domain
Personnel Security
Sub Domain
Hiring
During Employment
After Employment
Control Objective
Candidates screened for skill fitment
Candidates with questionable past records and behaviors are not hired
Control Statement
Conduct written exam
Perform in-person interview
Perform Background check (BGC) screening for educational qualifications: criminal records & work experience
Control Configuration
No hiring offer to candidates with unfavorable BGC (Administrative, Preventative)
Disciplinary process on personnel with unfavorable BGC (Administrative, Corrective)
Review any unauthorized access made by personnel with unfavorable BGC (Administrative, Detective)
Revoke any access provisioned for personnel with unfavorable BGC (Technical, Corrective)
Controls

***Figure 3.2**: Example of control structure*

A **compensating control** is an alternate control configuration that meets the requirements of reducing the risk but may do so by a more manual, less costly, or easier to deploy. For instance, instead of installing cameras at all of the entry/exit points to a building, placing either a lock or having the doors manned by a guard can fairly compensate for the risk of unauthorized entry.

Lifecycle of controls

We learned in *Chapter 2, About Managing Risks,* that we take actions to reduce risk. These actions are nothing but the implementation of controls. As the organizational dynamics/environment changes, risk assessment changes, and therefore, the redial actions/controls may change too. This lifecycle of continual change to the controls is referred to as the lifecycle of controls, as depicted in *Figure 3.3*:

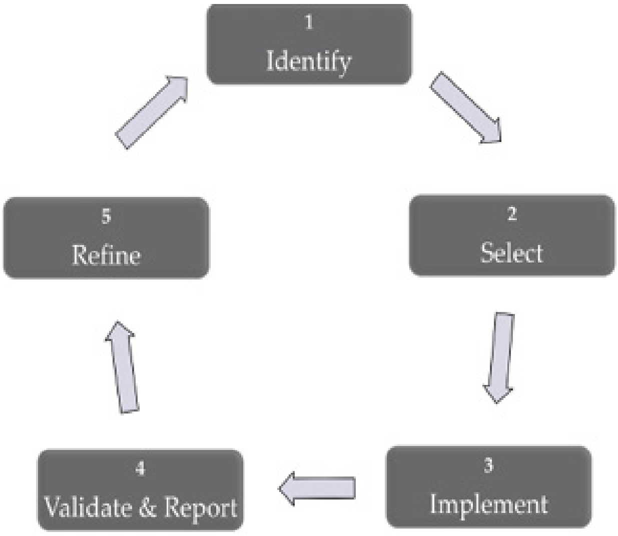


Figure 3.3*: Lifecycle of controls*

The explanation is as follows:

- **Identify**: An organization needs to identify the probable sources for the list of controls it needs to meet the security objective of its CIA. Some of these sources/factors are depicted in *Figure 3.4*. The frameworks from **MITRE*** (an independent advisor on the topics of information security), **Cloud Security Alliance (CSA)**, and **Centre for Internet Security (CIS)** benchmarks are widely accepted for their thought leadership on various aspects of security. Several of the standards assimilate these concepts into their requirements and vice versa.

  The following aspects are important to know:

  - Each of the factors depicted has several control objectives and configuration recommendations.
  - There will be an overlap of controls across all factors, even if the control statements may differ. In *Figure 3.4*, the intersection of various factors is shown; the size of the intersection is just representational. Many controls across factors will overlap because they are all primarily focused on CIA and/or Privacy.
  - At least two of the factors (regulations and management direction) are always applicable to an organization.

  Tip: The list of controls in a standard can be referred to as a control list, and the list of cont from all such applicable factors can be called a control library. There are several control libra available along with mapping of controls across all major standards and regulatory requiremer

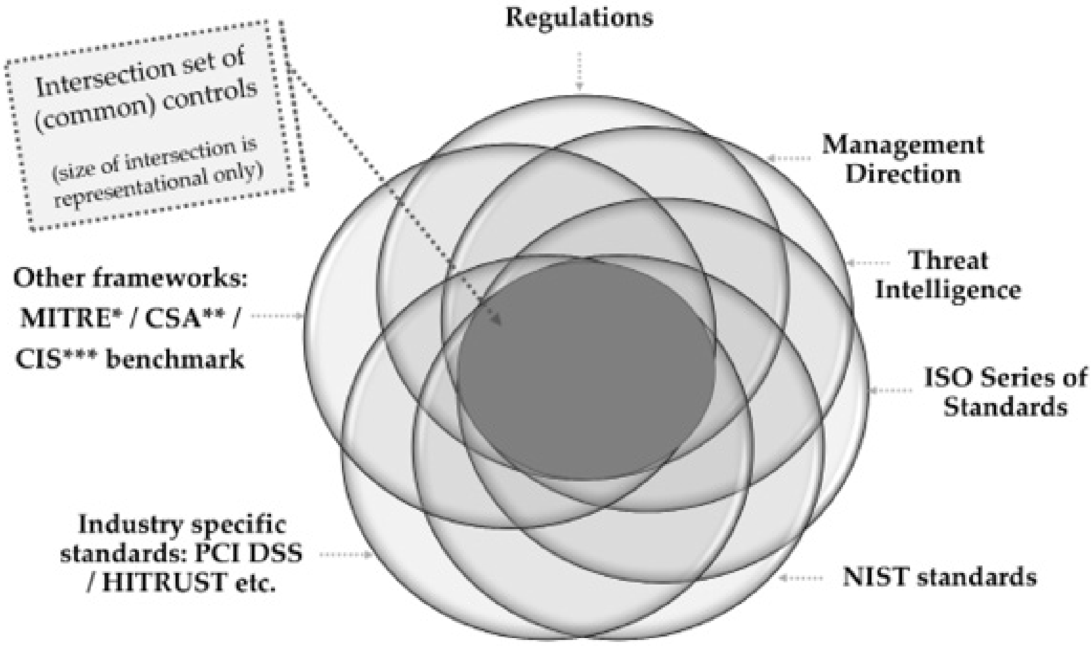

Intersection set of (common) controls
(size of intersection is representational only)
Regulations
Management Direction
Threat Intelligence
ISO Series of Standards
NIST standards
Other frameworks: MITRE* / CSA** / CIS*** benchmark
Industry specific standards: PCI DSS / HITRUST etc.

***Figure 3.4**: Factors for consideration in identifying controls*

- **Select**: The organization's management may select the controls by one or all of the following:
  - Using the **risk treatment plan** (**RTP**).
  - Based on advice from internal or external SME/advisors.
  - Based on management's previous experience.

Generally, the management is aided by the CISO and his/her team in this selection. For instance, as depicted in *Figure 3.5*, the organization may choose only some of the factors indicated with a tick mark.

Tip: It is possible for an organization to choose only a subset of controls from the list. For instance, the organization may choose to implement administrative control of performing the BGC but not implement the control for review or revocation of access. This may be because they are very deliberate about not hiring and onboarding a candidate unless the BGC is favorable.

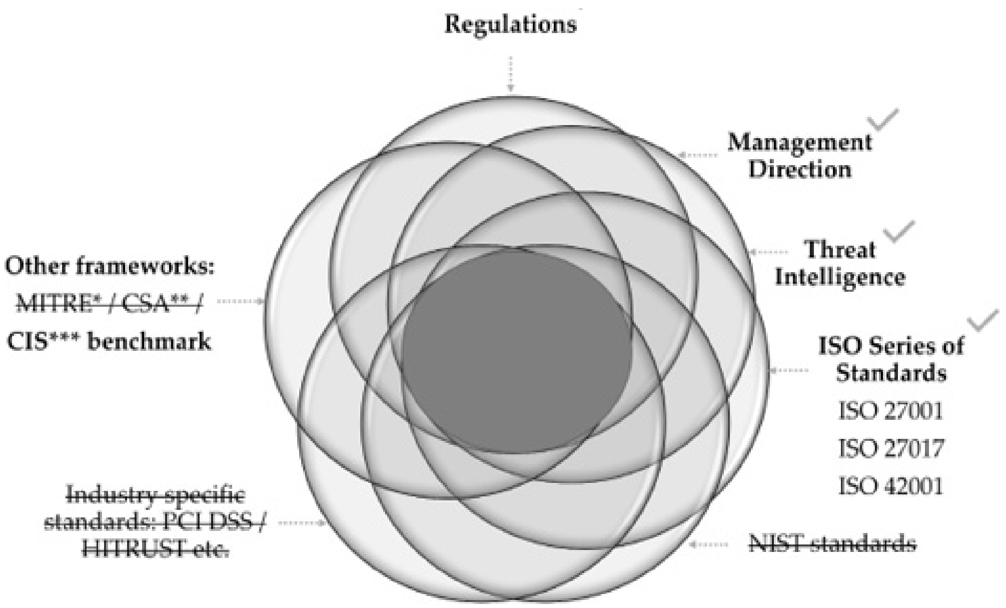
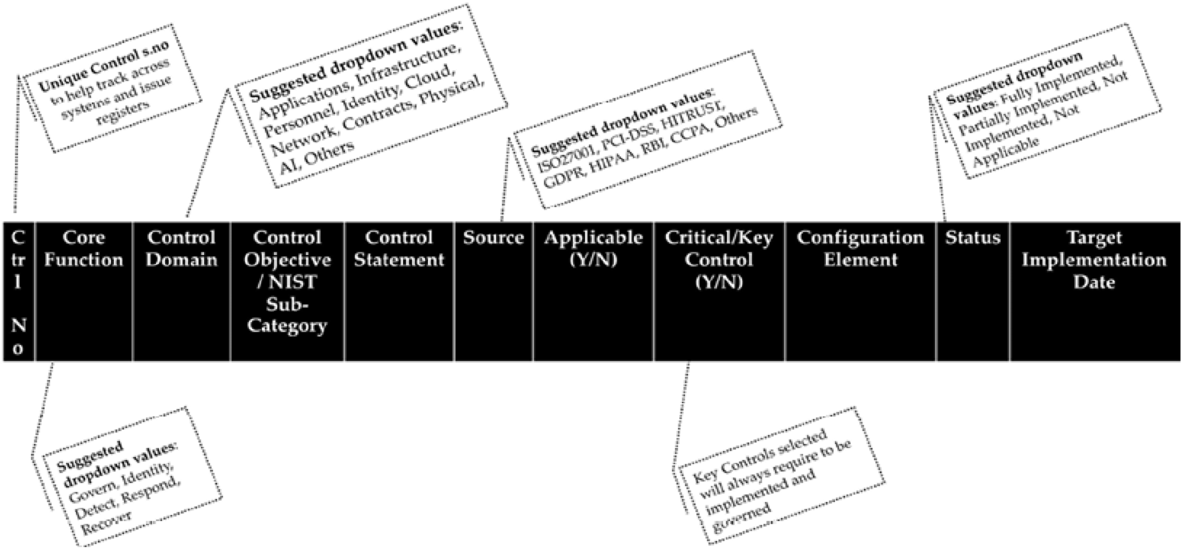


***Figure 3.5**: Selection of factors*

Figure 3.6 illustrates a control library format that may be used:



Ctrl No	Core Function	Control Domain	Control Objective / NIST Sub-Category	Control Statement	Source	Applicable (Y/N)	Critical/Key Control (Y/N)	Configuration Element	Status	Target Implementation Date

Unique Control s.no to help track across systems and issue registers

Suggested dropdown values: Applications, Infrastructure, Personnel, Identity, Cloud, Network, Contracts, Physical, AI, Others

Suggested dropdown values: ISO27001, PCI-DSS, HITRUST, GDPR, HIPAA, RBI, CCPA, Others

Suggested dropdown values: Fully Implemented, Partially Implemented, Not Implemented, Not Applicable

Suggested dropdown values: Govern, Identity, Detect, Respond, Recover

Key Controls selected will always require to be implemented and governed

***Figure 3.6**: Example of a control library*

- **Implement**: The selected controls must be implemented to reduce risk. We will prioritize the controls to implement based on constraints such as:
  - **Process readiness**: For instance, if the control chosen is to perform background checks, it might take time to set up the vendor(s) for it, train the hiring team, draw new/edited employment contracts, and so on. We may have to take one step at a time.

- **Technical readiness**: For instance, the ability to automatically process to disable access for personnel with unfavorable BGC may not be ready yet.
- **Cost**: For instance, conducting BGC checks in the USA or EU may be expensive, and enough budget may not be available for it.
- **Legal considerations**: For instance, performing drug testing during BGC may not be permissible in certain countries (where the organization may hire personnel).
- **Organization culture**: For instance, while the BGC may be a risk, the management feels the culture of the organization is not yet ready for the BGC to be done.
- **Existing alternate or compensating controls**: For instance, BGC may not be required if the candidate furnishes government-issued proof of past employment, or the candidate was hired right out of college.

See *Table 3.6* for a few examples of selecting the controls and documenting the justification of the choice made:

Domain	Control requirement	Applicable (Y/N)	Implement (Y/N)	Justification for exclusion
Personnel security	Perform **background check** (**BGC**) screening.	Y	Y	Based on cost or legal constraints some regions may be excluded from this requirement.
Personnel security	Review any unauthorized access made by personnel with unfavorable BGC.	Y	N	The organization does not hire unless the background check is favorable.
Application security	Implement a secure software development program and govern it.	N	N	The organization only uses commercial or freeware software and does not develop software.
Incident management and response	Manage information security incidents.	Y	Y	
Physical security	Secure of buildings from unauthorized entry.	N	N	The organization does not have any dedicated office buildings and its personnel work from home.

Domain	Control requirement	Applicable (Y/N)	Implement (Y/N)	Justification for exclusion
IAM	Access to information assets shall be based on *need-to-know*.	Y	Y	
IAM	Allow access to information assets only upon successful authentication using complex password.	Y	Y	Password complexity requirements shall not apply where it is technically not feasible, for instance, on voice systems-**call data records** (**CDRs**).
TVM	Deploying patches for the identified vulnerabilities.	Y	Y	

Table 3.6*: Selection of applicable controls*

Tip: It is a recommended practice to document justification even for the inclusion of a contro that the rationale is clear and connects to the risk management documentation.

Tip: ISO 27001 has a list of controls in its Annex A. Organizations may choose to include exclude those controls based on their applicability by documenting the justification in a Staten of Applicability (SoA). For instance, an organization can exclude control 5.11: Return of assets does not issue any computing devices and relies on its personnel to use their personal machi The certifying body's auditor will closely examine such exclusions in the SoA for their validity.

In *Table 3.6*, we included control on BGC. The justification for excluding the control to review access of personnel with unfavorable BGC is not valid because there are personnel in regions where BGC will not be conducted. We do not know the BGC of personnel in those regions, and thus it is important to appropriately monitor their access usage to protect the company.

An organization will continue to operate with risks until the selected controls or compensating controls are implemented. It is also possible that some of the controls originally selected in the previous phase may not be implemented at all due to changes in the business or regulatory environment.

Tip: It is important to make incremental progress instead of waiting for everything to be per for implementing controls. As good risk management practice, we will reduce risk as mucl possible and leverage risk treatment plans to document it.

- **Validate and report**: The success of a security program depends on continually validating the implementation of the control. There are two aspects to validate:

  - **Coverage**: This test and its measurement focus on whether all the control is implemented or not. For instance, when validating the implementation of password control, we will test the implementation of password requirements on all IT setups. One of the metrics to consider is the percentage of environments in a control that are not implemented as a function of the total environment. For instance, 6 out of 120 servers that do not have the required password control can be reported as 5% of servers do not meet password requirements.
  - **Effectiveness**: This aspect validates whether the implemented control is working as designed. For instance, on the servers, the password requirements are enforced, are they working correctly? Any deviation in implementation may also be noted as a %. For instance, 11% (i.e., 13/114) of servers do not have effective password control. Notice here that we removed the six servers where the control was not implemented from the denominator. You may, however, report effectiveness over the total population as well. The idea is to stay consistent in whichever way you choose.

Tip: Coverage tests are also called test-of-design (TOD), and Effectiveness tests are also called t of-effectiveness (TOE).

Validation of applicable controls may be done (internally or externally) through:

  - A formal process of assessment, wherein manual effort is done to check, and thus, it is slow and prone to mistakes.
  - Using tools or scripts in addition to the manual checks. These tests are scalable, fast, and more reliable.

Validation of controls brings out the gaps in implementation to remediate and/or enhancements to be made. The focus should be on remediating the risk.

- **Refine:** The environment around us is continually changing with evolving technology advancements, such as AI, regulatory requirements, such as India's **Digital Personal Data Protection Act** (**DPDP Act**) 2023, or business environments, such as an organization's growth plans. The risk assessment process continually evaluates the implications of such changes and may trigger modifications to control or even the selection of new ones. This concept is in line with the continual improvement that all standards require from a security program. Failure to refine controls may put the organization at risk of non-compliance or even a cyber-attack. For instance, a new requirement in regulation for data localization in a country, the architecture of the IT may need to be reviewed, and due adjustments made to not just meet the regulatory needs. Similarly, by watching out for increasing trends in cyber-attacks using phishing or impersonation, tweaking the controls to protect inbound emails, and making the users aware would be paramount.

The cycle of identifying relevant controls and their configuration will get triggered, and this process continues. When done well, the security program not only stays abreast of important developments but also continually improves.

About the NIST Cybersecurity Framework

NIST, a widely accepted authority on cybersecurity, published the **cybersecurity framework** (**CSF**) in February 2023 to guide government institutions and private organizations on managing cyber risk. The CSF:

- Is simple to understand and be used by all stakeholders in an organization.
- Helps organizations with *what* security outcomes they wish to focus on.
- Is increasingly becoming a framework for organizations to specify their target security outcomes and thus map the journey to them.
- Does not prescribe *how* those outcomes shall be achieved. It is meant to be used in conjunction with other frameworks or standards, such as ISO 27001, that define the how part.

The following are the components of the CSF:

- **Core functions**: Govern, Identify, Protect, Detect, Respond, and Recover. These are the outcomes an organization wants. For instance, an organization would want to ensure it has governance of its security program and can respond to adverse events to ensure the business can continue to run. The structure of CSF, see *Figure 3.7*, is similar in design to the structure of standards we reviewed earlier. Please note that the figure is representational and does not reflect all the functions, categories, and subcategories.

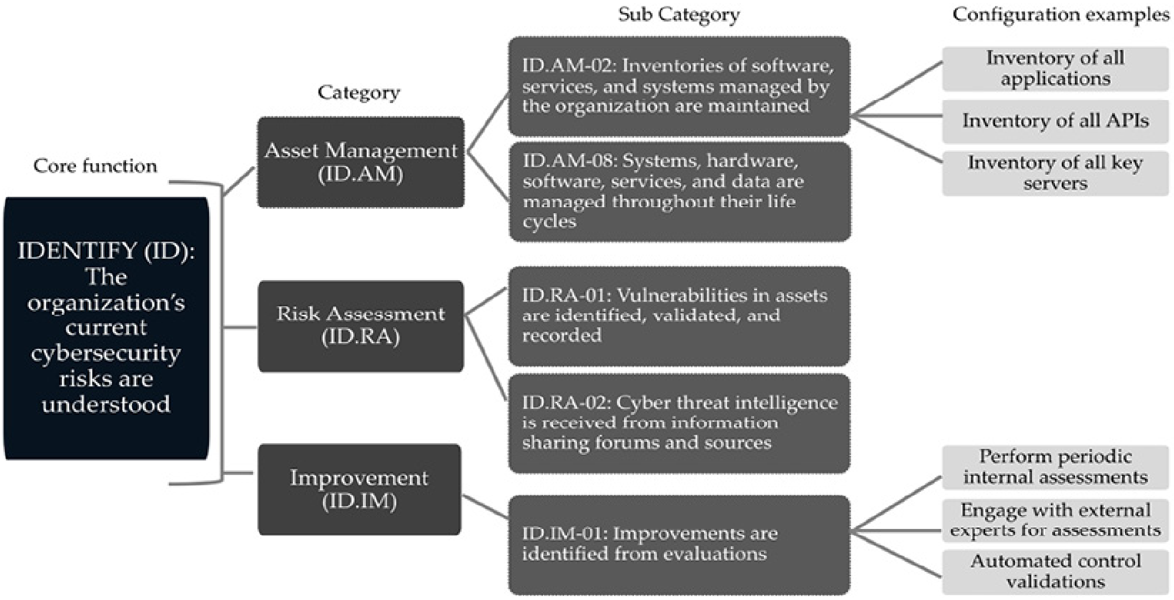

Core function
Category
Sub Category
Configuration examples
IDENTIFY (ID): The organization's current cybersecurity risks are understood
Asset Management (ID.AM)
Risk Assessment (ID.RA)
Improvement (ID.IM)
ID.AM-02: Inventories of software, services, and systems managed by the organization are maintained
ID.AM-08: Systems, hardware, software, services, and data are managed throughout their life cycles
ID.RA-01: Vulnerabilities in assets are identified, validated, and recorded
ID.RA-02: Cyber threat intelligence is received from information sharing forums and sources
ID.IM-01: Improvements are identified from evaluations
Inventory of all applications
Inventory of all APIs
Inventory of all key servers
Perform periodic internal assessments
Engage with external experts for assessments
Automated control validations

Figure 3.7*: NIST CSF*

- **Organizational profiles**: Helps the organization define the current state and choose the desired state of security posture. This profiling helps in drawing the roadmap to the desired state.
- **Organizational tiers**: Provide a means to grade the control's level. NIST suggests the tiers for the security program/security posture to be:
  - **Tier 1: Partial**—when cybersecurity efforts are ad hoc and reactive. For instance, an organization does not have the means to know and make its users aware of email-based phishing attacks. It may just be blocking phishing emails in an ad hoc fashion.
  - **Tier 2: Risk-informed**—when the cybersecurity efforts are consciously but reactively done based on risk in a piecemeal fashion. For instance, the awareness of phishing attacks and their remedial measures is known in parts of the organization, and controls are implemented on a case-by-case basis but not organization-wide.
  - **Tier 3: Repeatable**—when the cybersecurity program elements are documented and are applied consistently in a repeatable fashion. For instance, tools are used to continually block phishing attacks, and users are periodically made aware of the risks.
  - **Tier 4: Adaptive**—when the cybersecurity program continually adapts to changing environments and is prepared to quickly deal with any exigency. For instance, apart from blocking the rising phishing attacks,

the organization keeps a close eye on other types of email attacks and continually tweaks its preventive and detective controls.

The best use of the CSF is to examine the current state of the security program without any bias and set the bar for your desired state. This gap will help you plot the next steps and create a plan to work on. See *Figure 3.8*, showing an example of an organization's current state and desired state of security posture:

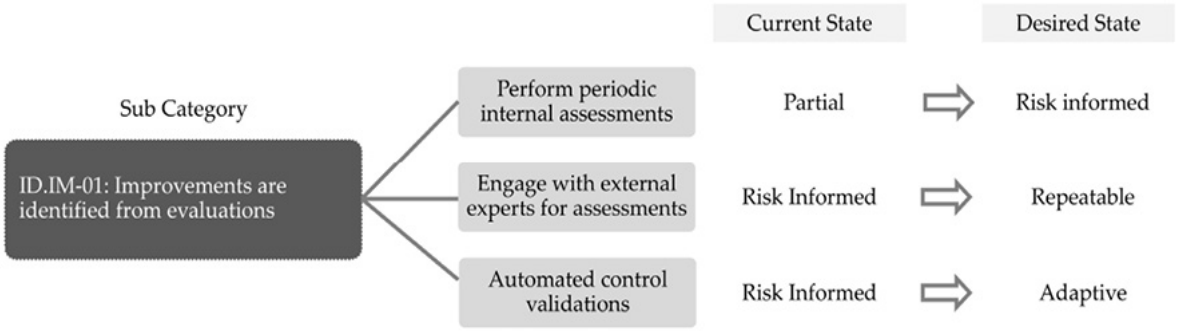

Current State
Desired State
Sub Category
ID.IM-01: Improvements are identified from evaluations
Perform periodic internal assessments
Partial
Risk informed
Engage with external experts for assessments
Risk Informed
Repeatable
Automated control validations
Risk Informed
Adaptive

***Figure 3.8**: CSF: Organization profiles and tiers*

We can also pictorially demonstrate the journey to cover against the desired state, and as a comparison to where the industry currently is. It is possible that an organization may be performing better than the industry average. See *Figure 3.9*, the desired state is depicted in a darker shade to optically convey the gap.

Tip: The journey from the current state to the desired state may need several months or years; the focus should be to make continual progress in line with management direction and the budget available.

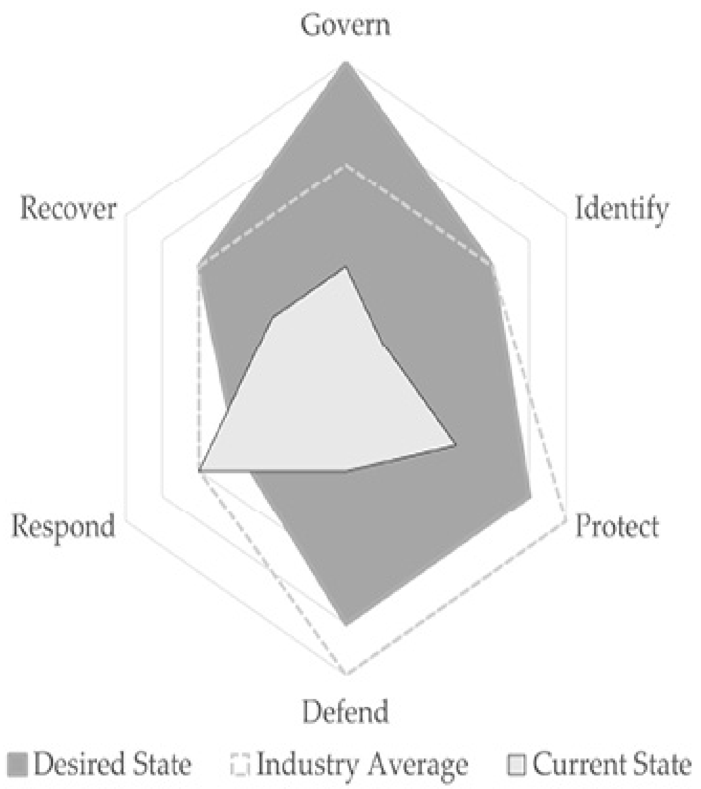


Figure 3.9: *CSF: Current and desired state benchmarked against industry average*

About policy and its governance

A policy is a written statement of management intent for any control. The management may authorize the CISO to draft and implement policies in line with the organizational objectives. All standards require some form of policy as part of an organization's security program. For instance, ISO 27001 has a mandatory requirement (that cannot be excluded even through SoA) for organizations to define and implement policies. The standard, however, does not mandate what the policy statement should be. Similarly, the CSF also has the function of govern (GV), which requires the organization to establish, monitor, and communicate policies to stakeholders.

For instance, if an organization requires its users to have unique passwords for accessing information assets, the policy may be drafted to read something like this: *Users shall be required to use their individual passwords to gain access to their authorized access information assets.*

In the example, we have added the constraint for them to access only the assets authorized for them. It is also normal to have these two separate policy statements, such as:

- Users shall only access information assets authorized for their use.
- Users shall be required to use individual passwords to gain access to the information assets.

Generally, each of such policy statements is clubbed under one or more policy documents based on the control domain. Additionally, organizations collate all policy statements they expect all their users to follow into a single document called the **Acceptable Use Policy** (**AUP**) for easy communication and implementation.

The organization must adopt a policy to define, document, implement, monitor, review, refine, and communicate requirements to relevant stakeholders. A policy framework must consider the following for the governance of the policy:

- Formally documented, preferably in a standard template.
- Approved by CISO or top management.
- Is reviewed at least annually, but is updated as needed as well.
- Must be relevant to the current organizational environment and the controls chosen for implementation/already implemented.
- Should be tool-agnostic.
- Stakeholder feedback must be solicited and considered at the time of revisions.
- Relevant stakeholders must be part of the review of policies in their area of work. For instance, running a BGC process is in the domain of the **human resources** (**HR**) function. They must be part of how the statement of intent is drafted and control is implemented. Similarly, policy statements on passwords or vulnerability patching must be consulted with the IT department.
- A record of revisions is maintained and has version numbers.
- Deviation from policy shall be maintained and handled as a process within the risk framework. For instance, the organization may choose to disallow USB ports on laptops, but may formally approve them for sales executives and the top management
- Nonconformance to policy must be handled, such as via administrative control or a disciplinary process.
- The policy may be tested from time to time by internal or external sources, and relevant learnings from this assessment will be incorporated into the

next versions of the policy.

Tip: Organizations commonly define a policy for making and releasing policies as well. This drives accountability and demonstrates the maturity of processes.

Tip: AI, especially generative AI, has been taking the world by storm since late 2022. There are risks for an organization with such AI models fed with the organization's confidential data. It is recommended to draft, document, and communicate to all users the policy on the use of AI tools.

Conclusion

In this chapter, we learnt about standards, types of controls, and policies. We covered the NIST CSF, which is becoming a well-respected framework.

In the next chapter, we will explore principles of security and their relevance in securing an organization's data.

Key takeaways

Some of the key takeaways learnt include:

- The operating environment of any organization is influenced by changes in the business environment, technology, and regulations on data security and privacy.
- Organizations may be subjected to industry or geography-specific requirements, too, such as HIPAA and GDPR.
- There are several globally accepted standards, such as NIST CSF, ISO 27001, and PCI-DSS, that help structure the requirements of protecting the CIA of information systems.
- Carefully selected technical, administrative, and physical controls to prevent, detect, and correct the gaps, perceived or real, help in this endeavor.

References

- **https://www.iso.org/home.html**
- **https://www.iso.org/standard/81230.html**
- **https://www.nist.gov/cyberframework**

- **https://www.mitre.org/focus-areas/cybersecurity**
- **https://rb.gy/zbhtyw (YouTube: look for "The one about Controls RSA")**
- **https://www.rbi.org.in/commonman/english/Scripts/Notification.aspx?Id=1721**

Join our Discord space

Join our Discord workspace for latest updates, offers, tech happenings around the world, new releases, and sessions with the authors:

https://discord.bpbonline.com

CHAPTER 4

Role of Privacy Principles

Introduction

In this chapter, we will explore the principles of privacy and understand their linkages to security controls. The relevance of privacy has especially grown since 2016 as regulations, such as the **General Data Protection Requirements** (**GDPR**) and the Privacy Act, began enforcing the requirements. The cost of non-conformance is often steep and irreparably damages the reputation of the organization.

Structure

The chapter covers the following topics:

- Privacy principles
- Important global regulations on privacy
- Relevance to security controls
- Data breach notifications

Objectives

By the end of this chapter, you will be able to understand the origins of privacy requirements, the developments in this aspect since the 1980s, and their implications for the CIA. As a CISO, you will be required to partner with various organizational teams, such as the legal department, technology, business operations, and human resources, to enable conformance to privacy requirements/regulations.

Privacy principles

The Organization for Economic Co-operation and Development (**OECD**) was established after *World War 2*, to advise global Governments on evidence-based policy on a variety of topics such as environmental pollution, taxation, and education. The OECD's privacy principles, first published in 1980 and then revised in 2013, are considered a gold standard in understanding privacy and its principles. We will first explore some key definitions:

- A **data subject** is a person whose personal data is being collected, stored, or processed. For instance, each of us individually are data subjects for our respective national or state/provincial Governments. Our full name, **date-of-birth** (**DOB**), postal address, contact details, gender, and similar fields are called **personally identifiable information** (**PII**) and constitute personal data. There are typically 18 such PII fields, and at least two fields of personal data are required to positively identify a person. Only the full name or the date of birth is not enough to zero in on a person.
- A **data controller** is an authorized body that is accountable for the privacy of the PII of data subjects. It collects and stores PII and decides how and what can be done. For instance, the Income Tax department in India, or the **Internal Revenue Service** (**IRS**) in the USA, may decide to use the taxpayer's information to auto-enroll them into the election voter list in an area. An employer will also be a data controller as they have accountability for candidate and employee/contractor PII.

  Tip: A data controller may outsource the process of collecting a data subject's PII to a third party, but it would still be considered a data controller.

- A **data processor** is an organization that handles the PII on behalf of the data controller and may store, transmit, or process the PII. For instance, the payroll process is generally outsourced by companies. This third party is a data processor and acts on the PII, name, bank account, DOB, and similar fields to process salaries and credit them into bank accounts. They are also required to provide any supporting evidence of controls for protecting such PII when asked by data controllers, local governments, or data subjects themselves.

Tip: A data processor cannot use the PII unless instructed by the data controller.

OECD privacy principles

Let us explore the privacy principles defined by the OECD and widely accepted globally. Some regulations in other countries may have slightly abridged principles and/or names for them, but the intent remains the same. They may differ from each other on the extent of rights of the data subject. See *Table 4.1*, in the last column, we have tried to indicate the most likely/ most significant leg of the triad that the principle will impact:

S. No	Principle	Purpose	Example(s) of violation	Triad to focus on
1	**Collection limitation principle**	Collecting only relevant and required PII by lawful and fair means and where possible with due consent of the data subject.	Taking pictures of data subjects using hidden camera. An e-commerce store collects excessive and unnecessary customer details like number of dependents in the family.	C, I
2	**Data quality principle**	Collected PII should be relevant, accurate and kept up-to-date as much as possible.	The phone number of the data subject is taken from the internet without validating the ownership. The motor vehicle seller mis-spells the name of the vehicle buyer in their records.	I

S. No	Principle	Purpose	Example(s) of violation	Triad to focus on
3	**Purpose specification principle**	At the time of collection, the data controller must specify the purpose of collection to the data subject.	A data controller collects phone numbers from data subjects but does not specify the reason. For instance, an employer collects phone numbers at the time of joining for purpose of communicating with the employee but does not specify so. A phone's fitness application does not make the data subject aware that it collects the number of steps using the inbuilt pedometer for recommending healthier lifestyle choices.	C
4	**Use limitation principle**	Using the PII collected must only be used for purpose(s) it was taken for and not disclosed without consent of the data subject, except if required by law enforcement or such authorities.	An insurance company collects the phone number from its candidates for employment purposes but then starts using it to attempt to sell insurance policies to them. An e-commerce collects data subject's (customer) date-of-birth for determining age but uses the information to send birthday wishes, coupons and vouchers during that month.	C
5	**Security safeguards principle**	Deploying reasonable security controls to protect the CIA of the data subject's PII.	Allowing data subject's PII access to everyone. A health diagnostic center provides access to the report just by entering a phone number without any additional checks, such as an OTP. Thus, anyone can see anyone's diagnostic report.	C, I

S. No	Principle	Purpose	Example(s) of violation	Triad to focus on
6	**Openness principle**	Transparency in demonstrating the developments, practices and policies about PII.	An airline's website allows online bookings but does not have a privacy policy to specify how it collects and uses PII. The power utility company of a city does not provide its consumers any information at their offices or website on how it uses the PII details, like consumer contact details.	A
7	**Individual participation principle**	A data subject has the right to know what all PII a data controller has and must be able to update or delete such records.	An employee does not get to see his/her PII that the employer has maintained. A consumer bank does not provide any means for the data subject (its customer) to view and edit/request correction to his/her PII such as contact number.	C, I, A
8	**Accountability principle**	A data controller is accountable for meeting all applicable privacy principles.	The data controller outsources the collection of PII to a third-party and wishes to transfer even its accountability to them. A data controller at a fitness equipment company allows the data subject's information to be used for emailing tailor made fitness plans without consumer's consent.	C

***Table 4.1**: OECD privacy principles*

A few concepts started to emerge as governments started to find ways to implement these principles. Notable among them are:

- **Data anonymization** is a process under which PII is fully or partially altered in a copy of the data it has in such a way that the PII can no longer identify a specific individual. Using etymology, the word anonymization is derived from anonymous, which implies without a name. Anonymized data can be shared easily for use in any analysis. For instance, if the insurance organization wants to analyze the nature of healthcare claims pertaining to waterborne diseases from a particular region, it can use anonymized data to evaluate aspects like the number of such claims, the amounts of such claims, and the timing of such claims. This can be used for better underwriting/pricing premiums for the future and/or work with governments/social workers to reduce the probable causes for such claims. Some of the techniques by which anonymization can happen are:

- **Data masking**: Under which the real PII is replaced/obscured with random letters and/or numbers so as to protect its confidentiality. For instance, the phone number of an individual may be shown as +91-123 xxx4 987. This technique is commonly used in handling PCI data, where all digits of the card number except the last four digits are masked. Just the last four digits do not bring any risk to the financial transaction or the PII of the individual to whom the card is issued.
- **Data generalization:** Involves removing parts of the data in a field such that it becomes generic and less identifiable. For instance, in the insurance company example, instead of the full phone number or even a masked phone number, just having the country and/or location code would generalize the data. Such a technique is most commonly used for area codes, year of birth (by omitting date and month), or using age in blocks of range such as 20-30 years, 31-40 years, and so on.
- **Data swapping:** Involves randomly replacing characters and numbers from the same data set. For instance, swapping the data in row 1 of the data sent with the DOB of the 4th person, the address of the 7th person, and so on. These numbers, 4th and 7th, are also randomly chosen and are stated here only for example. Again, there should not be a pattern in the randomization. In this example, the data seems near real but is no longer PII.
- **Data perturbation:** Involves modifying actual data using techniques like rounding off or adding random numbers, called noise, to phone numbers or salaries. This technique helps maintain confidentiality, but if the factor of perturbation is not chosen wisely, it can expose the original data. For instance, by rounding off employee age to the nearest 10 multiple may protect the employee’s

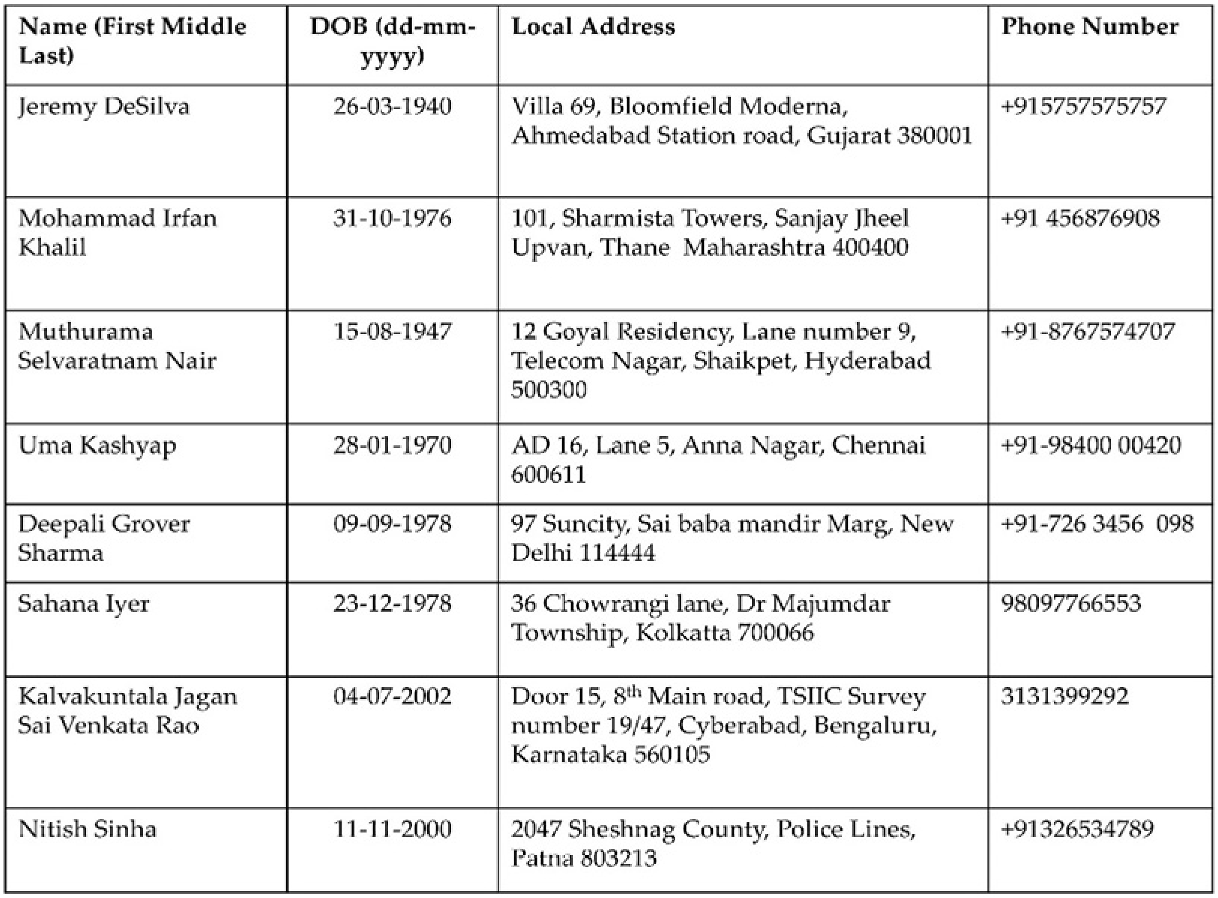
PII if the year of birth is like 1976 or 2013, but will be meaningless if the year of birth is something like 2010.

- ○ **Pseudonymization**: It involves replacing original values with pseudonyms (false) values. For instance, replacing the date and month with some fake date and month.
- ○ **K-anonymity**: It involves using techniques like data generalization and data masking in tandem to make the data indistinguishable. Data generalization is applied to a dataset containing individuals with the same set of a few fields PII together, and then masking portions of some of the fields. A group must contain more than two people. For instance, an organization can apply to insurance companies for group medical insurance of its staff by generalizing gender, age, and medical condition in some meaningful groups and then masking their postal code. This way, the insurance companies can make a reasonable judgment for their underwriting and compute premiums, while the organization would have maintained the confidentiality of its employees' data.
- ○ **L-diversity**: Extends the K-anonymity by grouping three or more individuals with atleast one sensitive field as different. For instance, in the example above, the medical condition (the sensitive information should be different) while the age, gender, and zipcode (quasi-identifiers) are common. It will not be possible to identify individuals.
- o **T-closeness**: Involves the distribution of sensitive information, which is a quite close representation of the generalized information of the overall population. For instance, if the K-anonymity group for age 30-40 is shown as prediabetic as the medical condition, the same should be true for the entire population/cohort as well.

- **Synthetic data:** Involves artificially generating data sets and using them for testing. For instance, if the insurance company wants to test its sales campaign for a new policy, it can randomly generate a name, DOB, contact email ID, and location, and use that to test its application's functionality. Remember, in this case, there is no real PII.
- **Data encryption:** Involves using technology to make part or all of PII undecipherable or useful only to the authorized person who has the

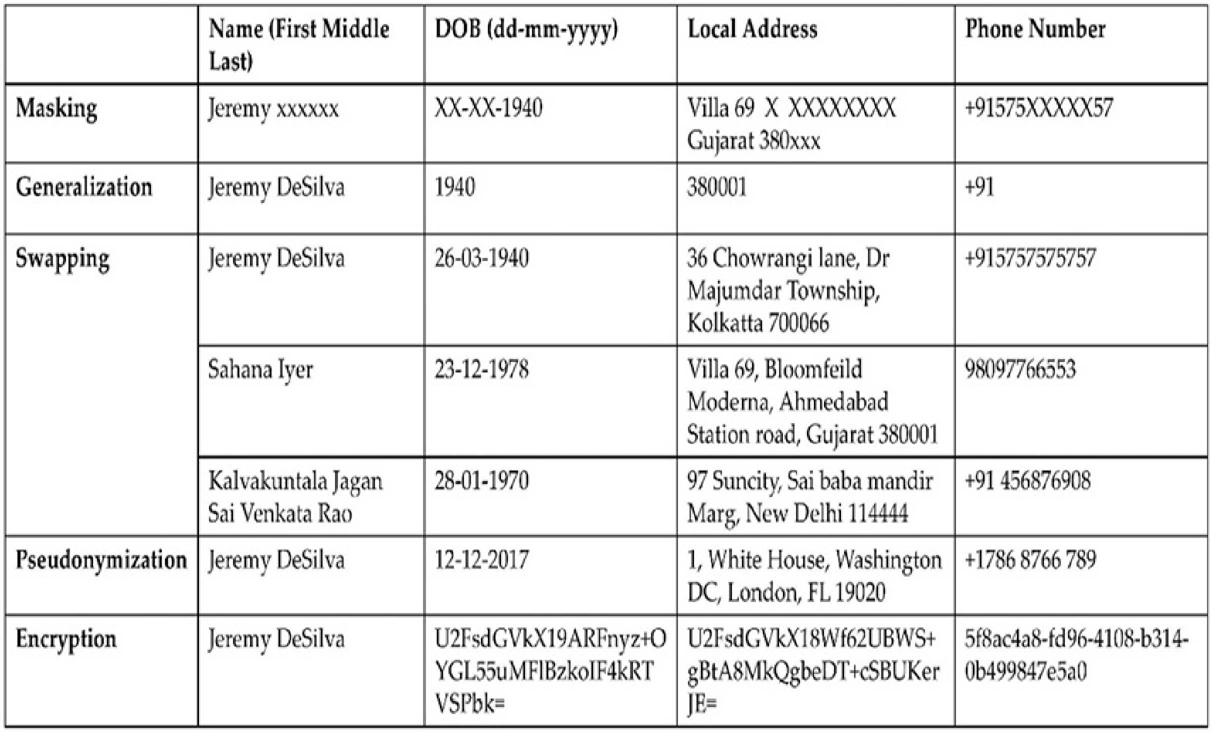
access and means to decrypt. This technique is useful for allowing the continual use of the original PII for the purposes it was collected, but protecting it from unauthorized access.

See *Figure 4.1* and *Figure 4.2* for examples of various data protection techniques discussed above:



Name (First Middle Last)	DOB (dd-mm-yyyy)	Local Address	Phone Number
Jeremy DeSilva	26-03-1940	Villa 69, Bloomfield Moderna, Ahmedabad Station road, Gujarat 380001	+915757575757
Mohammad Irfan Khalil	31-10-1976	101, Sharmista Towers, Sanjay Jheel Upvan, Thane Maharashtra 400400	+91 456876908
Muthurama Selvaratnam Nair	15-08-1947	12 Goyal Residency, Lane number 9, Telecom Nagar, Shaikpet, Hyderabad 500300	+91-8767574707
Uma Kashyap	28-01-1970	AD 16, Lane 5, Anna Nagar, Chennai 600611	+91-98400 00420
Deepali Grover Sharma	09-09-1978	97 Suncity, Sai baba mandir Marg, New Delhi 114444	+91-726 3456 098
Sahana Iyer	23-12-1978	36 Chowrangi lane, Dr Majumdar Township, Kolkatta 700066	98097766553
Kalvakuntala Jagan Sai Venkata Rao	04-07-2002	Door 15, 8th Main road, TSIIC Survey number 19/47, Cyberabad, Bengaluru, Karnataka 560105	3131399292
Nitish Sinha	11-11-2000	2047 Sheshnag County, Police Lines, Patna 803213	+91326534789

***Figure 4.1**: Original data set (example)*



	Name (First Middle Last)	DOB (dd-mm-yyyy)	Local Address	Phone Number
Masking	Jeremy xxxxxx	XX-XX-1940	Villa 69 X XXXXXXXX Gujarat 380xxx	+91575XXXXX57
Generalization	Jeremy DeSilva	1940	380001	+91
Swapping	Jeremy DeSilva	26-03-1940	36 Chowrangi lane, Dr Majumdar Township, Kolkatta 700066	+915757575757
	Sahana Iyer	23-12-1978	Villa 69, Bloomfeild Moderna, Ahmedabad Station road, Gujarat 380001	98097766553
	Kalvakuntala Jagan Sai Venkata Rao	28-01-1970	97 Suncity, Sai baba mandir Marg, New Delhi 114444	+91 456876908
Pseudonymization	Jeremy DeSilva	12-12-2017	1, White House, Washington DC, London, FL 19020	+1786 8766 789
Encryption	Jeremy DeSilva	U2FsdGVkX19ARFnyz+OYGL55uMFlBzkoIF4kRTVSPbk=	U2FsdGVkX18Wf62UBWS+gBtA8MkQgbeDT+cSBUKerJE=	5f8ac4a8-fd96-4108-b314-0b499847e5a0

***Figure 4.2**: Data set after data protection techniques were applied*

Tip: To meet the Individual Participation Principle, organizations implement a process called Data Subject Request (DSR). This allows individuals to know the current details of their PII and to request an update or even deletion/destruction if they so desire. The deletion of their records would need to be done as far as possible and confirmed back unless there is a regulatory reason to retain.

Important global regulations on privacy

In the decade of the 1970s, a digital revolution, the era of personal computers had started gaining rapid pace. This meant more and more PII was being stored on computers and was accessible faster. Computerization had started picking up pace and national governments started developing laws for implementing privacy principles, especially in the USA, Australia, and the **European Union** (**EU**).

Tip: The United Kingdom (UK) was part of the EU until 31st January 2020.

In *Table 4.2*, we examine some of the global privacy regulations and one or more of their striking features/differentiators:

S. No	Country	Some of the latest global regulations on privacy	Key features
1	India	**Digital Personal Data Protection Act (DPDP Act)** 2023	Only digital personal data is in scope. Specifies blacklist regions (where data of Indian nationals cannot be sent, processed or stored). Data Fiduciary defined which is loosely a data controller. Provision for a consent manager to manage individual consents across organizations. Requirement for organization with significant data on individuals to appoint a **Data Protection Officer** (**DPO**). Consent is valid only if the purpose of the organization's service is directly relevant for it. For instance, if an insurance company's app sought and got permission from an individual for all the contacts on his/her phone but because the service of insurance has no requirement to access the individual's phone book that consent is automatically invalid. Insurance company can use the consent to only contact the individual. Verifiable parent (or guardian) consent is required for any data pertaining to a child (under age 18 years). Even then behavioral monitoring, targeted ads are not permitted to be sent to a child. Information to data subjects should atleast be his/her native language.

S. No	Country	Some of the latest global regulations on privacy	Key features
2	UK/EU	**General Data Protection Regulation (GDPR)** 2016.	Applies to all members of the EU/UK and on any organization processing data of EU/UK citizens irrespective of their location. For instance, a US based e-commerce portal would need to comply to GDPR requirements for orders from EU/UK customers. **Data subject rights (DSR)** that enables the principles of transparency and individual participation. Data subject's consent is an explicit requirement. Data Minimization—need-to-know, need-to-have at the data collection stage itself. International data transfer is permitted but under strict conditions. A child is generally defined to be 16 or above. Non conformity can lead to steep fines as high as 4% of global revenue or £20million whichever is higher.
3	USA	**Health Insurance Portability and Accountability Act (HIPAA)** 1996	Patient's rights to their healthcare data. Responsible use of patient healthcare data by the organizations involved in patient care.
4	Canada	**Privacy Act** **Personal Information Protection and Electronic Documents Act (PIPEDA)**	Privacy Act covering PII data of data subjects handled by Canadian Federal government. PIPEDA handles PII of Canadian data subjects by private organizations and excludes its applicability to 3 Canadian regions. It has 10 principles for privacy, though similar to the 8 OECD principles.
5	Australia	**Privacy Act 1988**	Has 13 privacy principles but are like the 8 OECD principles. Permits individuals to use pseudonym. Scope of jurisdiction includes organization providing services in Australia irrespective of their location. Cookie consent management is part of the law.

S. No	Country	Some of the latest global regulations on privacy	Key features
6	China	**Personal Information Protection Law (PIPL)**	Data of Chinese nationals collected, stored and processed must be within territorial boundaries of **Peoples Republic of China** (**PRC**). Cross border data transfer is permitted but strictly though framework and government consent. Provides more powers to Chinese government access to individual data and permitting some access to those data subjects. Non compliance can lead to fines upto 5% of revenue or cancellation of permissions to run the organization.

Table 4.2*: Various global privacy regulations*

Tip: At the end of 2024, a report by the United Nations reported that around 137 out of 194 countries had some legislation in place for the protection of data and privacy in their respective countries.

The state of California in the **United States of America** (**USA**) enacted the **California Consumer Privacy Act** (**CCPA**) in 2018. The Act had provisioned several privacy-related rights similar to the privacy principles covered above, such as the *right to know* the information collected. Organizations providing services to citizens of the state are required to comply with the state law. The CCPA was further strengthened by another state law, the **California Privacy Rights Act** (**CPRA**), 2020, which focused on several landmark requirements, such as organizations being required to honor consumers' data not to be sold. Several organizations used a concept of *opt-out,* where a data subject could choose not to participate in email-newsletters outreach, mailing list campaigns, have their data sold/exchanged with other organizations without consent, or even cookies that are used by the organization's website.

Tip: Regulations such as the GDPR and CCPA have forced organizations to grant their website visitors greater control over the privacy of data collected by cookies. Websites now implement a largely opt-in method for cookies that can be collected and used by the website. Generally, only cookies necessary for their direct operations are enabled by default, and all other third-party ones, like online behavior analytics or advertising, are blocked by default.

Relevance to security controls

Privacy and security have a symbiotic relationship. While the privacy of individual data stems from regulation(s) and specifies the rights of individuals for their own data, security, on the other hand, protects all data, including personal data, and is generally an outcome of several requirements we discussed in *Chapter 3, Role of Standards and Controls*. We may also look at security controls as to how the privacy principles may be met, whereas the privacy principles focus on what to protect. For instance, an organization collects location information from a user to deliver goods from the online shopping experience. With regards to privacy, the user's location, name, and phone number are required to deliver the goods, and security controls can be designed and implemented to protect even this data. The app can be built so that the number is invisible to the delivery agent, and yet s/he can call the data subject to coordinate. In *Table 4.1,* we examined the triad of CIA impacted by each OECD principle. For instance, to meet the *Individual Participation Principle*, the organization must ensure that appropriate authentication and authorization are applied to limit access to any authorized individual. The OECD principles and features of several of the privacy laws require organizations to have reasonable security.

Privacy principles focus on the rights of individuals to their PII that an organization collects and uses. Individuals are entitled to know what is collected, why, and how it will be used. Additionally, individuals have the right to update their information and even request that it be deleted if they desire. In order to enable these rights, the CIA triad of security principles is used. For instance, an individual provides name, postal address, email ID, and phone number to the e-commerce portal and wishes to:

- View the current PII information that the portal has maintained. The e-commerce would implement some means for the individual to authenticate and see their profile. This authenticated access upholds the principle of confidentiality. Additionally, the ability to see the information as and when reasonably possible also supports the concepts of availability.
- Update their current PII in a secure fashion. The e-commerce portal would implement a means for an authenticated individual to directly update their information on the portal, or require the e-commerce company to be contacted via registered phone number or email ID. This

verified update would uphold the integrity of the information. Similarly, the individual can request his/her PII to be deleted.

We examined the relationship of the OECD privacy principles to the CIA earlier in *Table 4.1*. It may be inferred that several security controls may come into relevance. For instance, an e-commerce organization:

- Spells out a policy on what their practice is/shall be for handling the DSRs, such as *individuals shall require authentication to the portal to access their PII information on their profile and to request any modification requests*. The organization must also create a **standard operating procedure** (**SOP**) for handling DSRs.
- Decides to require individuals to call their helpdesk for any PII modifications. While the privacy individual participation principle will still be met, the procedure may run a risk of unauthorized modification, as an unverified caller may get anyone's information changed. This risk will need to be handled via some remediation and documented accordingly.
- To manage any untoward changes to PII, the security controls of monitoring, incident management, and response shall come into force.

Data breach notifications

A security incident is a security event where authorized access and/or leakage may have happened. If such data pertains to individuals, that security incident turns into a data breach. Each of the privacy regulations may give a general limit to the penalty that shall be imposed in the untoward event of a breach. A great way for organizations to mitigate this risk of penalty is to work with an external counsel/organizations can also purchase cyber insurance to cover the money payable because of such untoward events. Broadly put, insurance is a strategy for risk transfer. In this light, cyber insurance is similar to property or car insurance. Here is a list of some of the penalty ranges in some of the regulations:

- The DPDPA requires prompt notification and may impose a penalty of ₹ 250 crores, i.e. $ 30 million or higher.

- The GDPR may impose a penalty on an organization for failure to report within 72 hours, a sum of €10 million, i.e. $10.36 million.
- The *Australian Privacy Act* may impose a penalty of AU$30 million. i.e., $ 31 million.

Some of the recent fines awarded to organizations for not meeting privacy regulations include:

- **Jan 2025**: $60K on a healthcare organization for not honoring a patient's right to know.
- **Nov 2024:** €500 at a café in Spain (GDPR).
- **Nov 2024:** €100K on call center for a power utility for unconsented telemarketing (GDPR).
- **July 2022**: Oklahoma State University—Center for Health Services pays $875,000 (HIPAA).
- **Jan 2019**: €50 million by French authorities on Google for requiring the creation of a Google account on Android phones (GDPR).

Conclusion

In this chapter, we learnt the privacy principles and examined several global privacy regulations. In a nutshell, privacy specifies the aspect of personal data that needs some control and protection, while security controls protect the personal data. Organizations include technical requirements of privacy in their security control selection and implementation, and measure them.

In the next chapter, we will explore how security and privacy requirements can be built into the design of an organization and its ways of working.

Key takeaways

Over the years, several regulations have emerged and evolved globally to require organizations to protect the personal information of individuals they serve. Some key takeaways are:

- The eight globally accepted principles of privacy have also become the cornerstone of the information system architecture, design, development, and deployment.

- These principles and regulations define what to protect, while the security controls are the how, i.e., the implementation element.
- DSR, one of the most visible implementations of the privacy principles, must be carefully planned and executed.
- It is possible that an organization will have lapses in its security controls, leading to a security incident. Care, however, must be exercised to differentiate those from a breach that has more stringent legal implications, with reporting on time and on penalties and other procedural customer notifications that the regulator or courts may levy.
- Privacy breaches can have significant reputation and financial implications for an organization.

References

- **https://www.oecd.org/en/topics/sub-issues/privacy-principles.html**
- **https://www.bbc.com/news/uk-politics-32810887**
- **https://www.meity.gov.in/writereaddata/files/Digital%20Personal%20Data%20Protection%20Act%202023.pdf**
- **https://www.enforcementtracker.com/**

CHAPTER 5
Security and Privacy by Design

Introduction

In this chapter, we will explore the principles of security and privacy to be applied when designing an application or a process. We will also explore the similarities between the two and the ways these requirements can be commonly understood.

Structure

The chapter covers the following topics:

- Designing with proper security and privacy controls
- Defining security by design principles
- Defining privacy by design principles
- Challenges in implementing these design principles

Objectives

By the end of this chapter, you will be able to understand the various design principles, such as those proposed by OWASP, for the security and privacy of applications/products. The same concept can also be applied when designing new processes/services.

Designing with proper security and privacy controls

Security and privacy requirements are required to ensure that the CIA and the privacy of the information are maintained throughout. The application/product/service that has built such requirements from the start is likely to be more secure and compliant. Some of the advantages of focusing on security and privacy at the design stage itself are:

- **Reduced rework on coding**: If an application is designed to consider the requirements for authentication and is prepared to code accordingly, the chances of having to retrofit the requirement later are reduced. For instance, an e-commerce portal accepting credit cards for payments must ideally conform to the PCI-DSS standard from the start. The application and its interfaces need to be coded and tested accordingly. Similarly, a healthcare application must be designed and coded in such a way that the **protected health information** (**PHI**) is safe from unauthorized access/modification.

  This would not only make the applications meet compliance requirements but also enhance customer trust. To retrofit such a key business requirement later and be compliant will not be optimal. **First-time-right** is a mindset where a future-relevant design is so well thought through that it is likely to be the right way to accomplish. Rework may cause software developers to feel demotivated, and therefore, it may impact the quality of coding.
- **Better alignment to business objectives**: The purpose of a for-profit organization is to continually look for growth and create a relevant impact in the market it operates in. The business objectives are defined by top management to drive the strategy for the path forward. The role of the CISO and the function of privacy is to enable the business on that growth path in a risk-based approach and enhance customer trust. Any product or service that the consumer is not able to place trust in is unlikely to grow. The security and privacy practices can help augment

that trust for the consumer.

- **Meeting compliance needs**: Organizations have to constantly conform to regulatory, contractual, and management demands. Not meeting regulatory requirements may have punitive impacts, while others may have organizational repercussions, such as reputational and/or financial impact. All such constraints and requirements form a part of the compliance needs. We explored several ways to define, select, and implement controls. Compliance should not be thought of as a check-in-the-box activity; instead, the requirements should become part of the culture of practice and a business requirement.
- **Speed to outcome:** It is likely to be faster. Given that the code already should be meeting security/privacy requirements, the tests prior to release are likely to have minimal gaps, and therefore, the chances of delivering new functionality/features are faster. For instance, a technology product company uses secure coding practices as prescribed by standards such as the **Open Worldwide Application Security Project** (**OWASP**). Such applications are likely to be free of vulnerabilities and meet customers' third-party risk management requirements easily.
- **Reduce potential risks:** It reduces risks later on and helps the organization make more informed choices for controls and user experiences.

Overall, factoring security and privacy requirements at the design stage itself is less expensive, directly helping with the finances of the organization. The secure coding practices are likely to help organizations be compliant and reduce the chances of security incidents and breaches. As the old adage goes, *a stitch in time saves nine*, and a timely fix is most cost-efficient.

Much like the standards we explored in *Chapter* 3, *Role of Standards and Controls,* there are several frameworks that help us to think about security by design and privacy by design in a structured fashion. Most of them have similar concepts. We will leverage frameworks such as the OWASP, which is a not-for-profit, community-driven organization that works on the security of software. The digital revolution runs on software, and thus, the OWASP recommendations are ideal to consider for applicability. We will discuss the OWASP Top 10, a widely consensus-based list of critical security risks, later in the chapter.

Defining security by design principles

OWASP, secure by design, applies whether the development process used is the age-old waterfall model or the new-age agile-development methodology. The agile methodology is nimble to recognize the needs of the customer and incrementally develop, test, and deploy products (applications) to market. Irrespective of the software development methodology used, the most appropriate place for thinking about security is in the design stage itself.

- **Principle of least privilege (PoLP)**: Access to data/information with as minimal access as possible, as required. For instance, the logistics team of an e-commerce retailer only sees the customer's name, shipping address, and details of goods on the software, but does not see the credit card used for payment. Similarly, in a healthcare application that handles patient care to insurance-related functions, the medical care provider must be able to see all medical history and treatment-related information, but may not be able to see other PII, like phone numbers and email IDs. On the other hand, the insurance company agent should only be able to see aspects like a blood test was conducted, but all the results of the blood test may not be relevant for the claim.
- **Segregation of duties (SoD)**: Implementing **maker** (the one who raises the request) and **checker** (the one who verifies/fulfills the request) to be at least two independent people. For instance, the e-commerce retailer's procurement team can submit the invoice of its suppliers, but the payment is initiated and verified by a separate team.
- **Defense in depth**: Deploy multiple layers of controls to prevent and detect any malicious or cyber-attack activities. For instance, use email security tools to protect inbound email attacks and also administratively train users to handle suspicious emails, thus ensuring that a process to detect and respond to such suspicious emails exists.

- **Secure by design**: The default configuration of the IT assets and/or the applications should be secure. For instance, a new application being developed for handling credit card data must ensure that data is encrypted, and the default server password must be changed.
- **Fail safe**: The systems must be able to handle error conditions such that information is not exposed even for unknown errors. For instance, a webpage's search capability should be able to handle symbols and special characters without crashing. Similarly, an incorrect command line option for a directory search must not return confidential information results.
- **Economy of mechanism**: Keep it simple and straightforward—choose the simplest implementation for coding and deployment. For instance, an application can be deployed with a database on the same server and yet securely segregated for access.
- **Openness principle**: The implementation of the design should be independent of the design itself. For instance, in an application development program, a formal code repository should be part of the design and can be openly represented. This does not expose any risk to the application. However, where the code repository is, how it is accessed, and how it is kept updated are implementation aspects and should be secured.
- **Least common mechanism**: An application should use the least common shared function/compiled code to ensure undue authorization does not materialize. For instance, applications using the same code library for provisioning access must ensure that the access is not granted for all applications automatically. It should only be for the access minimally required.
- **Psychological acceptability**: Balance user experience and security controls using a risk-based approach. For instance, in an application showing just the status of fulfillment of an IT request, there is no need to encrypt the data and/or deploy **multi-factor authentication** (**MFA**), versus an application that has PII, such as an employee payroll application.
- **Reuse code appropriately**: No new vulnerabilities or configuration weaknesses should be introduced by the use of common library code.

For instance, an application reusing an existing code for formatting the table output must not additionally pull information that is not authorized for the report. Similarly, end-of-life or vulnerable open-source libraries must not be used.

- **Security logging**: Anomalous, unexpected conditions and authentication attempts to an application must be logged and kept secure. For instance, monitoring and analyzing login attempts to an application from unexpected locations, such as outside the country, should be done.
- **Secure coding**: Proper secure coding should be done with practices using OWASP, such as:
    - Input validation
    - Strong authentication
    - Controlling access to information and code itself
    - Proper session management
    - Data security—anonymization, encryption, etc.
    - Encoding output
    - Code review and testing

For instance, improper session management can cause information disclosure and data leakage, defeating the confidentiality requirements.

In *Table 5.1*, we explore the principles and some key security domains (covered in *Chapter 3, Role of Standards and Controls*) that it addresses:

S. No	Principle	CIA Impacted	Security domain addressed
1a	**PoLP**	C	**Identity and access management (IAM)**—authorization
1b	**SoD**	I, C	IAM—authorization
2	**Defense in depth**	C, I, A	All security domains
3	**Secure by default**	C, I	IT security
4	**Fail Safe**	C, I, A	Application security, IT security
5	**Economy of**	C, I	Application security

	mechanism		
6	**Openness principle**	C, I	Application security, security architecture and engineering
7	**Least common mechanism**	C, I	IAM, application security
8	**Psychological acceptability**	C	Risk management
9	**Reuse code appropriately**	C	Application security
10	**Security logging**	C, I, A	Defensive security, incident management and response
11	**Secure coding**	C, I	Application security, offensive security

Table 5.1*: Security by design principles and most relevant security control*

Defining privacy by design principles

Embedding privacy requirements into the design and development of applications, systems, and processes not only helps meet the regulatory requirements but also helps in ensuring improvements in customer trust in the growing digital revolution. Much like security, there are benefits when such requirements are already built in at the design stage itself.

In the 1990s, *Information and Privacy Commissioner of Ontario, Canada*, *Ann Cavoukian,* proposed the most widely accepted **privacy by design** (**PbD**) principles, detailed as follows:

- **Proactive vs. reactive**: Incorporates measures to prevent data breaches proactively. For instance, using security principles like authentication and session management helps maintain confidentiality.
- **Privacy by default**: Automatically protect PII by default, which requires no user intervention. For instance, a user profile on social media automatically restricts publishing of all PII fields and lets the user choose if s/he prefer to display their phone number in full or masked, etc.
- **Privacy embedded into design**: Privacy requirements are foundational requirements when developing any application or business process. For

instance, a new payroll application ensures data encryption, segregation of duties, and the principles of least privileges are applied in line with risk management practices.

- **Positive sum**: Incorporate all business requirements, which include privacy ones, without making unreasonable tradeoffs, and document decisions based on risk. For instance, a web application that is able to anonymize user preferences and yet provide meaningful contextual content.
- **End to end security**: Protecting PII throughout the lifecycle from collection, storage, processing, and destruction. For instance, a web application that collects fit-for-purpose user PII and ensures it is protected for CIA at all times.
- **Visibility and transparency principle**: The implementation of the design should be independent of the design itself. And the privacy requirements are met. For instance, the application should deploy all measures to ensure that the PII collected is only used for the purpose for which consent was given, and that is independently verifiable, irrespective of the process used.
- **Respect for privacy**: The focus should be in line with the data subject's privacy considerations and protection. For instance, the application or service must by default focus on protecting the PII and respect the user's choices. For instance, if the user disables performance cookies, the application shall not be able to optimize some of the front-end renditions.

OWASP, ISO 31700, and many other standards were adapted from that guidance. See *Table 5.2* for the explanation of PbD, some examples, the impact on CIA, and some security controls that directly apply:

S. No	Principle	CIA impacted	Security domain addressed
1	**Proactive vs. reactive**	C, I	Application security, risk management, incident management and planning.
2	**Privacy by default**	C	Application security, operations security.
3	**Privacy embedded into design**	C	Security architecture and engineering, application security, risk management.

S. No	Principle	CIA impacted	Security domain addressed
4	**Positive sum**	C	Security architecture and engineering, risk management.
5	**End to end security**	C, I, A	All security domains.
6	**Visibility and transparency principle**	C	Application security, security architecture and engineering.
7	**Respect for user privacy**	C	Operations security, defensive Security.

Table 5.2*: Privacy by design principles mapped to most common security controls*

Challenges in implementing these design principles

The principles for security by design and privacy by design are quite self-explanatory, and yet application coding continues to have the same avoidable errors. Generally speaking, any risk to confidentiality is a likely loss of privacy as well because exposure of PII to unauthorized information implies that at least one of the privacy principles was not met. See *Table 5.*3 for **OWASP Top 10** security issues that all security coders and testers must be familiar with:

S. No	OWASP Top 10	Brief explanation	Impacted CIA	Example of fix
1	Broken access control	Users must be able to see/access only what is authorized to them.	C, I	Implement role-based access control.
2	Cryptographic failures	Information systems must protect the confidentiality of data at rest and in transit.	C	Use strong password hashing algorithms.
3	Injection	Applications must handle the user supplied information/queries appropriately to prevent unauthorized disclosure.	C, I,	Using safe techniques like **application program interface** (**API**).

S. No	OWASP Top 10	Brief explanation	Impacted CIA	Example of fix
4	Insecure design	Security requirements were incorporated into coding. For instance, a self service password reset process does not require any validation of identity.	C, I, A	Using a library of common security controls.
5	Security misconfiguration	Application, database use default passwords or error handling gives away too much information, for instance, a login attempt failure informs if the username was incorrect or the password was.	C, I, A	Hardening guidelines. Build proper error handling flows.
6	Vulnerable or outdated components	**End of life** (**EOL**) software/software components.	C, A	Governance of EOL.
7	Identification and authentication failures	Inappropriate handling of authentication such as not implementing MFA.	C	Proper management of sessions, or MFA.
8	Software and data integrity failures	Failure to verify the authenticity of patch/update of software.	I	Strong release management and patch management governance.
9	Security logging and monitoring failures	Information systems must be able to log relevant security events and those must be triaged/screened and acted upon as needed.	C, I, A	Implement verification of sufficiency and relevance of logging.
10	**Server-side request forgery** (**SSRF**)	Applications not validating the source of the request and allowing a malicious link to be clicked and opened.	C, A	Implement deny-by-default.

***Table 5.3**: Privacy by design principles*

Tip: In November 2021, a critical vulnerability (referred to as CVE-2021-44228) was announced. Log4j was a widely used component and had several million users globally. This event was almost a game-changer that called for organizations worldwide to update the library code immediately or risk a serious cyber attack.

Tip: OWASP Top 10 was last formally revised in 2021, and the next update is scheduled to be published around November 2025.

The OWASP Top 10 application security issues have largely remained the same for decades. The following factors may be considered to improve the quality of secure software (that also factors in privacy requirements):

- Enhanced stakeholder engagement, such as the product management team, in specifying the requirements, or when the development team does not feel the relevance of such requirements.
- Ensure formal and proper training for developers on secure coding practices and the implications of insecure code.
- Incorporating security and privacy requirements as use cases in agile development and business requirement specification in the traditional waterfall development methods to ensure they become part of the build.
- Balancing security requirements without compromising on user experience. For instance, not managing the user session appropriately may lead to someone else being able to see information s/he are not authorized to.
- Often, in the garb of user experience, privacy principles such as consent and purpose may be overlooked. Even if it seems a hindrance, consent must be formally and clearly sought.
- Ensure strong governance to ensure secure code is built, tested, and only then released.

Organizations may consider the following best practices:

- Adapt a secure software development lifecycle, where the principles of security by design and privacy by design are an integral part.
    - Define and implement coding guidelines.
    - Govern the creation, appropriate use, and periodic updates of reusable libraries.
    - Use threat modeling and risk assessments to identify potential security issues early in the design phase. We will cover some of these topics later in the book.
    - **Continuous integration and continuous deployment (CI/CD)**: Implement CI/CD pipelines with integrated security checks.
    - Implement governance on integration with other systems and the

    APIs used.
    - Use automated tools to perform static and dynamic code analysis, vulnerability scanning, and security testing.
    - Consider using peer code review, where feasible, to ensure the quality and security of code.

- Define and implement clear roles and responsibilities for the development and testing team, and foster the principle that security is everyone's responsibility.
- Use security frameworks and standards such as ISO27001 and OWASP Top 10 to ensure coding efforts can effectively deal with emerging threats.
- **Impart periodic and relevant training**: Organizations must consider providing training on security by design, privacy by design, and secure coding practices to their application architects, developers, and testers. Some of the popular and effective training on the topic include:
    - **OWASP secure coding practices**: Utilize the *OWASP Secure Coding Practices Quick Reference Guide*. This guide provides a comprehensive checklist of secure coding practices that can be integrated into the software development lifecycle. It covers essential topics such as input validation, output encoding, authentication, session management, and more.
    - **SAFECode training**: SAFECode offers a detailed guide on fundamental practices for secure software *development*. This guide includes best practices for identifying and managing application security controls, which are essential for an effective secure software development program.
    - **SANS secure coding training**: The SANS Institute provides various secure coding training courses, such as *SEC540: Cloud Security and DevSecOps Automation* and *SEC542: Web App Penetration Testing and Ethical Hacking*. These courses are designed to equip developers with the skills needed to identify and mitigate security vulnerabilities.

Conclusion

In this chapter, we learnt the value of deploying the principles of security and privacy at the design stage of applications or processes. The overall benefits of doing so are in the cost of rework and also in the cost of compliance.

In the next chapter, we will explore some security technologies that help protect the data/information of organizations.

Key takeaways

Some of the key learnings include:

- The principles of security-by-design and privacy-by-design correlate to the CIA requirements of information assets.
- Management engagement and support for these principles must be extended and should be duly considered in the design and implementation of any product, application, processes, or services.
- The choice of some of the controls by the organization's team may be driven by convenience or user experience, but ultimately, regulatory requirements and business objectives should drive the final choice.
- Applications should be designed, developed, and tested for weaknesses enumerated in lists such as the OWASP Top 10.
- Building privacy controls at a later stage may be feasible, but they are likely to be costly and cumbersome.
- The teams working on such applications, products, and services should be duly trained and capable.

References

- **https://github.com/OWASP/CheatSheetSeries/blob/master/cheatsheets/Secure_Product_Design_Cheat_Sheet.md**
- **https://www.merriam-webster.com/dictionary/a%20stitch%20in%20time%20%28saves**

%20nine%29#:~:text=used%20to%20say%20that%20it,it%20become%20a%20bigger%20problem

- https://owasp.org/www-project-developer-guide/
- https://www.dsci.in/files/content/documents/2024/Privacy-by-Design-DPLF-SIG-paper.pdf
- https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf
- https://owasp.org/www-project-top-ten

CHAPTER 6

Key Security Technologies

Introduction

In this chapter, we will examine some of the common security terminologies and technologies that a CISO and his/her team would frequently encounter. These concepts will help the information security team understand the technology, its architecture, and think through the controls.

Structure

The chapter covers the following topics:

- Genesis of computers and networks
- Key network security technologies

Objectives

By the end of this chapter, you will be able to understand the various security-related aspects of various technologies, networking principles, and the importance of protocols and ports, algorithms, and offensive and defensive technical.

Genesis of computers and networks

The origins of devices that could calculate/compute, i.e., computers, can be traced back to almost ~3000 BC when Abacus was invented. It was a simple wooden frame with several horizontal rows of steel wires on which the beads were mounted. These beads were moved around to perform basic mathematical functions. In fact, even today, in several countries, such as India, it is probably one of the first toys a kid gets when starting primary education. By the mid-17th century, mechanical devices that could add using gears, such as the Pascaline, were invented. Science grew leaps and bounds in various aspects from the late 17th century onwards, when steam engines were invented. Around 1820, *Charles Babbage*, a British mathematician, invented a steam-powered device called the **Differential Engine** to solve logarithmic problems. About a decade later, he invented the **Analytical Engine** to solve a variety of mathematical problems using punch cards and cards with holes. He is considered the **father of computers**. *Ada Lovelace* is considered the **mother of computers** for designing and writing the algorithm for this *Analytical Engine*. The Analytical Engine never got made. By the mid-20th century, several technologies such as vacuum tubes, transistors, magnetic devices, and printed circuit boards were being used for complex computing to solve for speed to outcome in science and military. Using computers commercially started around the 1960s. A key aspect to note is that these computers were mostly standalone, they did a very specific task or tasks and were not connected to other devices usually except punch card readers and magnetic tapes.

In the 1970s, the first personal computer based on microchips started being commercially built and used. By then, computer programming also had grown leaps and bounds with languages such as **Common Business Oriented Language** (**COBOL**), C, Pascal, FORTRAN, and so on. These were used to program and put the personal computer to a variety of tasks for an individual's work.

In the late 1960s, the realization of interconnecting computers to share/exchange information digitally gained prominence. The space war between the USSR (now Russia) and the USA triggered the **Advanced Research Project Agency** (**ARPA**) to build **Advanced Research Projects**

Agency Network (**ARPANET**) to connect two computers and send information from one to the other using a concept called packets. About a decade later, **Transmission Control Protocol/Internet Protocol** (**TCP/IP**) was developed to specify how computers can communicate with each other. For instance, if any two individuals have to speak and communicate effectively, they use a common language such as English with its vocabulary where each word means at least one thing. Similarly, TCP/IP was built to make dissimilar computer technologies communicate with each other.

Tip: The first ever email was sent by Ray Tomlinson to himself between two computers on ARPANET, and apparently, all he sent were the keystrokes on the top row of a standard keyboard layout "QWERTYUIOP".

Coaxial cables were used for short distances, much like a dish antenna was connected to a television set. In the 1970s *Twisted-pair* cables, also called *Ethernet cables*, were the most common medium to interconnect computers using TCP/IP. Technological enhancements, such as **fiber optic cables**, continued to evolve making these connections more reliable, efficient, and fast even over long distances.

By the late 1960s, more than one computer could be interconnected using special devices such as *switches* or *routers*. These devices had slots for the twisted-pair wires from computers to be inserted. These slots are called **network ports**. Switches have several physical ports to which hosts are connected to form a network. Routers connect servers, switches, and other parts of the network to other networks, typically branch networks and/or the Internet.

The **Institute of Electrical and Electronics Engineers** (**IEEE**) provides thought leadership to modern-day networking and to define and develop technical standards on interoperability. The IEEE 802 series of standards is one of the most important technical standards that define how networks, wired or wireless, should be set up for seamless intercommunication. The network speed was measured in bits per second. In the digital world, everything is a sequence of 0 (zero) or 1 (one), and each of these 0/1 is called **bits**. 8 bits make up a byte. Look at *Table 6.1* to understand network speed and what can be typically done at that speed:

S. no	Speed	Type of network	Typical use
1	10Mbps	Ethernet	Internet browsing, emails

	(Megabits per second)		
2	100Mbps	FastEthernet	Fast file transfers, video conferencing
3	1000Mpbs/1Gbps (Gigabits per second)	Gigabit Ethernet	Gaming, high-definition video streaming, connecting servers
4	100Gbps	Gigabit Ethernet	Used within data centers

***Table 6.1**: Network speed and their typical uses*

Network types

Networks are categorized based on aspects like scope, and geographical spread, some of the types of networks are:

- **Local area network (LAN)** computers are interconnected in a physical space over twisted (or fiber optic) cables using routers or switches. The LAN would typically connect computers in a computer lab, a floor or two in a building, the entire building, or even a home. LAN is typically a wired network and uses ethernet-related standards such as IEEE 802.3.
- **Wireless local access network (WLAN)** interconnects computers and devices, such as phones or tablets, using Wi-Fi (or Wi-Fi) technologies based on the *IEEE 802.11 standard*. This revolutionary standard is commonly used in our day-to-day lives, at home, at hotels, at airports, and even in offices.
- **Campus area network (CAN)** interconnects LANs within a campus to form one network. It may use cables or other IEEE standards on electromechanical interconnectivity.
- **Wide area network (WAN)** interconnects networks over large geographical areas such as between offices, data centers, buildings, campuses, cities, or even countries. The Internet we use today is an implementation of WAN technologies.

Network topology

Network topology is a term used to describe how commuters (nodes) may be interconnected within a network. The term node may be used to represent computers/devices. The following types of topologies are possible:

- **Point to point (P2P)** where two nodes are connected together over a single network link/cable.

- A **bus** where each node is connected to a single cable.

  See *Figure 6.1* for a visual representation of point-to-point and bus topology:

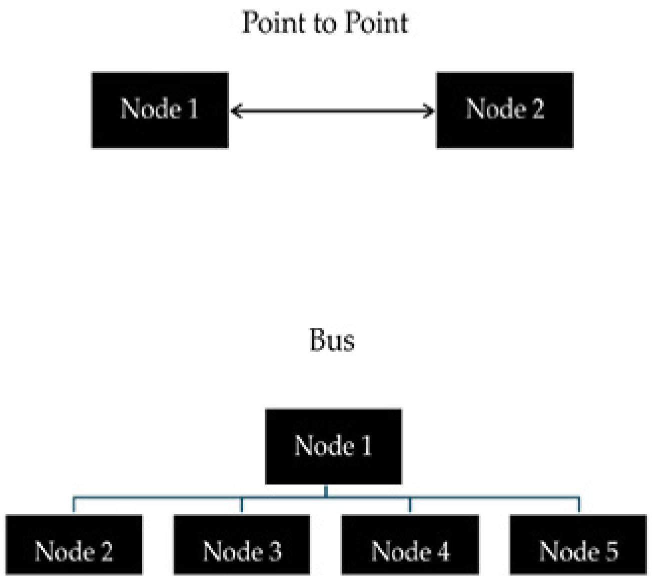


Figure 6.1: *Network topology—point-to-point and bus*

- In a **ring network,** each computer is connected to its left and right with a single connecting medium, like a coaxial cable, such that it forms a ring.
- **Tree** is a hierarchical interconnection between nodes.

  See *Figure 6.2* for a representation of a ring and a tree network topology:

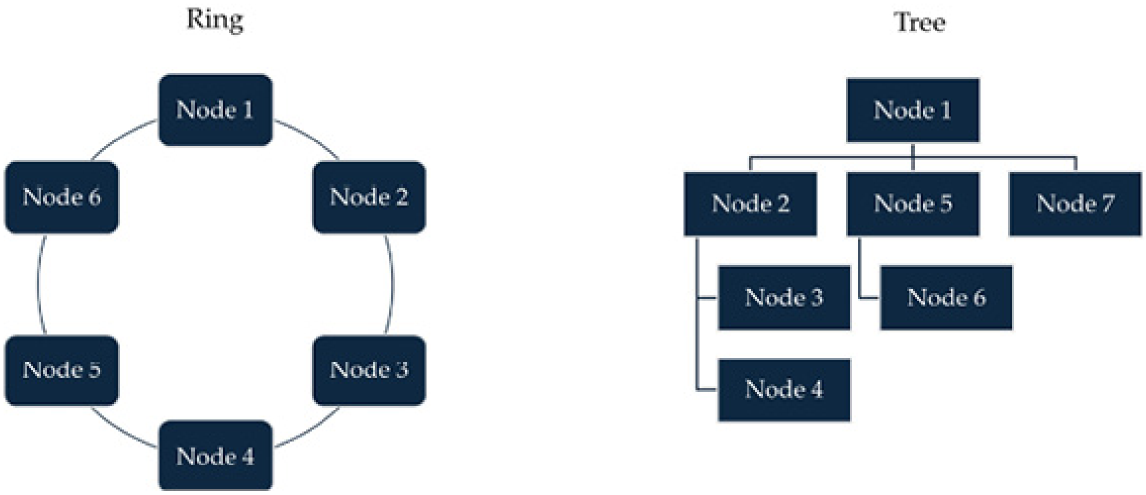


Figure 6.2: *Network topology—Ring and Tree*

- **Star** where each node is connected to a central hub (router or a switch) and can communicate with each other.
- **Mesh** where each node is interconnected to the other.

  See *Figure 6.3* for a representation of a star and mesh network topology:

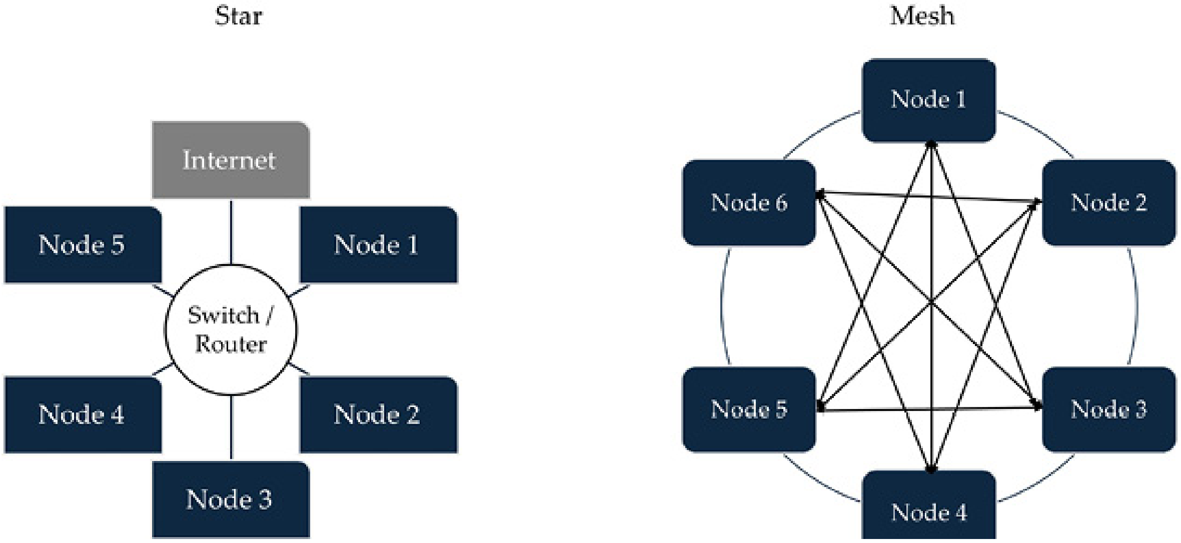


Figure 6.3: *Network topology—Star and Mesh*

- A **hybrid** where the topologies defined above are used in conjunction

IP address

The **Internet Protocol** (**IP**) address is the mechanism by which devices are identifiable on a network and are able to send and receive information. If a device on the network is your digital home, then its IP address is your home address. ARPANET introduced the concept of IP address.

IPv4 address

In the 1980s the first widely used version of IP addressing called **IPv4**, i.e. IP version 4, gained prominence. Even today, IPv4 is widely used across the globe and hence IPv4 and IP address are used interchangeably.

An IPv4 address is a 32-bit notation with two parts: a network portion (or ID) and the host address itself. The network portion of the address identifies the network to which the hosts belong, and the host address identifies the specific host itself. An IP address has four 8-bit portions, and each 8-bit portion is also called an octet. An IP address is represented in a format called dotted decimal notation. Each of the four octets is notated in decimal numbers separated by a period (or dot). For instance, 172.16.9.123 is an IP address of a host.

IPs on a device can be manually configured (called static IP) or be allotted through a process on a server (called dynamic IP). The server that has the functionality to allocate the IP and some other network configuration is called **Dynamic Host Configuration Protocol** (**DHCP**). Static IPs are typically used where the host is required to be reached by the same IP, and hence, it is used for servers, network devices, etc. End-user computers usually use a DHCP service to get an IP. These dynamic IPs are usually segregated into **virtual LAN** (**VLAN**) to further segment the network and limit excessive traffic. On a network, only one host can have the exact same IP address at any given point in time. It is, however, possible to reassign the same IP to another host after a certain duration lapse or a certain purpose is completed. The duration for which an IP is associated with a particular host is called the IP lease.

There are four classes of IP addresses that are in use, as shown in *Table 6.2*:

Class name	Range	Higher order bits	Network octets	Example (with network octet highlighted)	Number of possible networks	Number of unique IPs in network
Class A	1.0.0.0 to 126.255.255.255	1	First (using 7 bits)	10.10.10.10	128	~16.7 million
Class B	128.0.0.0 to 191.255.255.255	2	First two (using 14 bits)	172.16.9.123	16,384	65,536
Class C	192.0.0.0 to 223.255.255.255	3	First three (using 21 bits)	192.168.10.100	2,097,152	256
Class D	224.0.0.0 to 239.255.255.255 (used for multicast)	-	Not defined	224.0.1.253	Not defined	

***Table 6.2**: Classes of IP address*

The minimum value for an octet is 0, and the maximum is 255 because, in binary form, 255 is 11111111. Class E IP addresses in the range 240.0.0.0 to 255.255.255.255 are reserved for research and development and are not used in any commercial network configuration.

0.0.0.0 is not a routable address, implying a host can neither send nor receive any information if it has this IP. Similarly, you may notice a network with an octet starting at 127.0.0.0 is not listed as it is a special Class A network used only for loopback traffic, i.e. traffic to itself. TCP/IP implementation specifically uses 127.0.0.1 for the host to refer to itself. A webserver running on a host can be accessed on the web browser using **http://127.0.0.1**. This IP is also called localhost. This special network is very commonly used in the development testing of applications without having to access the Internet.

In each network, a host cannot be assigned 0 in the final octet, nor can 255 be assigned. 0 represents the network itself and 255 is the broadcast IP for all hosts to receive messages. For instance, 192.168.1.0 is not a host but a network, and similarly, 192.168.1.255 is the broadcast address of the same network.

Routable and non-routable IPs

Internet Corporation for Assigned Names and Numbers (**ICANN**), a not-for-profit organization, has a worldwide mandate to define and oversee Internet resources such as IP address allocation, the root Domain Name Server (DNS), and so on. The **Internet Assigned Numbers Authority** (**IANA**) is the global authority that runs the Internet smoothly. It implements the directions of ICANN. **Public Technical Identifiers** (**PTI**) is the specific body within IANA that now has the worldwide mandate to define and control the key elements of the Internet, such as IP addressing, domain names (including TLDs), and Internet protocols.

IANA reserved the following IP ranges, called **private IP ranges**, for use within an organization:

- 10.0.0.0 to 10.255.255.255
- 172.16.0.0 to 172.31.255.255
- 192.168.0.0 to 192.168.255.255

These IPs are not directly available on any Internet device and are also called non-routable, implying these devices cannot directly access or be accessed from the Internet. Most organizations use one or more of these ranges for their devices on their LAN. Except for these private IPs, the reserved IPs like the loopback IP 127.0.0.1, the Class E IP range, and all other IP ranges are **public IPs** or **routable IPs**.

Hosts with private IP addresses cannot access the Internet or any services on it because their IP addresses are not routable. Public IPs are quite expensive and are in short supply therefore organizations would lease out only the required number of public IPs for their use. Here are a set of possible reasons for organizations to need a public IP:

- To host and maintain its web servers, such as its corporate website, or even the e-commerce portal of the retail company.
- To permit access to corporate infrastructure like corporate email, messaging systems, file servers, and application servers from the Internet.
- To permit remote users to connect to its corporate network via technologies like **virtual private network** (**VPN**).
- Provide its users with access to the Internet.

Netmask and subnets

Network Mask, or netmask, is used to divide larger networks into smaller networks called subnets. It is also a 32-bit / 4-octet number represented in dotted decimal notation. It also helps separate network ID and host ID by using a series of 1s and 0s. For instance, a netmask of 255.255.255.0 means that the first three octets are the network ID. It is also denoted by /24 to signify three octets (8 x 3 = 24). In this case, there are only 256 hosts that are possible. This slash '/' notation is called **Classless Inter-Domain Routing** (**CIDR**). The use of CIDR makes the IP classless, thus adding to the flexibility of IP allocation and the more optimal allocation of available IPs.

The advantages of subnetting/CIDR are as follows:

- **Create optimal networks**: For instance, if an organization decides to use an IP range from Class B for their 2300 hosts, the rest of almost 63000 (of 65536) IPs would lie unused. Using netmask, we can create a smaller subnet say of 3000 hosts and free up the range.
- Enables **efficient routing** of traffic by routers.

- Reduces the traffic load as the larger network will not see traffic meant for hosts within a subnet.
- It allows **granular control of permissions for hosts** in that subnet. This concept lines up well logically with the principle of need-to-have that we discussed in *Chapter 1, The Triad of Security.*

A subnet is created by using some of the host ID bits as part of network bits. This implies that subnetting a Class A network gives you more possible subnets versus Class B or Class C. It is also apparent that more subnets may mean that there are less number of hosts per subnet. *Table 6.3* shows some of the subnets and the formula to calculate the number of subnets and hosts:

Class	Network bits (a)	Host bits borrowed (b)	CIDR /(a+b) (n)	Subnet mask	Effective subnets obtained (2^^{b})	Number of hosts per subnet (2^$^{32-n}$-2)
A	8	1	/9	255.128.0.0	2	8388606
A	8	10	/18	255.255.192.0	1024	16382
A	8	22	/30	255.255.255.252	4194304	2
B	16	1	/17	255.255.128.0	2	32766
B	16	5	/21	255.255.248.0	32	2046
B	16	13	/29	255.255.255.248	8192	6
C	24	1	/25	255.255.255.128	2	126
C	24	3	/27	255.255.255.224	8	30

***Table 6.3**: Examples of subnets in each class of network*

Tip: In the context of the OSI model, the subnets are used in data traffic routing and hence, are a part of layer 3—Network.

Creating a network topology, subnet planning, and routing is both a science and an art. The engineers must ensure the following considerations are made before deciding on a subnet:

- Number of approximate hosts: the expected number of IPs needed in a subnet with the future in mind.
- The number of subnets to manage should be as optimal as possible,

neither too many nor too few, as it also has an impact on routing.

- Class of IP available for use.

IPv6 address

IPv4 has around 4.2 billion IPs, but these are not sufficient. Since the 1990s, there has been a massive growth in the use of information technology and a growth in organizations' digitization efforts. Additionally, around the year 2000, there was a rapid increase in personal devices like smartphones and handheld devices. The **Internet of Things** (**IoT**) further proliferated this demand for IP with smart TVs, CCTVs, robotic floor cleaners, and several industrial devices. On average, a four-member household today is likely to have at least six devices that use an IP address. The **Internet Engineering Task Force** (**IETF**) had the foresight to recognize this eventuality and came up with IPv6 (IP version 6) in the late 1990s.

IPv6 is a 128-bit addressing system with several trillions of IPs. It is notated using a colon-separated *hexadecimal* format. There are eight parts to the address with four characters, each separated by a colon. For instance, 2405:201:c009:305d:1a0e:c005:bed:741b is an IPv6 for a host. If there are zeros in the portion, such as in 2405:0000:c009:305d:1a0e:c005:0000:741b, the four zeroes can be excluded and be represented as 2405::c009:305d:1a0e:c005::741b. Preceding zeros need not be written; for instance, notice the second portion from the right has only three characters, i.e., 201.

A modern device can auto-configure an IPv6 for itself, and therefore manual configuration or a DHCP server is not needed. However, the devices still need other network configurations, like a gateway IP, and a default DNS server to be provided. This configuration is provided by DHCPv6 servers. The IPv6 IP leased or self-assigned also has a lease time, but it can be significantly longer.

IPSec is built into IPv6, and thus connections between hosts can maintain confidentiality and integrity.

OSI model

Open Systems Interconnection (**OSI**) is a widely accepted conceptual 7-layered architecture framework that structures the flow of data between two devices on a network. This **International Organization for Standardization** (**ISO**) framework allows two disparate technology devices to communicate using common specifications. For instance, an organization has Windows-based servers and Apple iOS-based end-user devices. Both these technologies are different and yet would be required to be interconnected for the business to be able to leverage the benefits. The standards, such as IEEE 802, would be implemented by Windows and Apple and would use the OSI model as a reference architecture for their hardware.

Figure 6.4 illustrates the OSI reference model and how the data flows from the sending host to the receiving host:

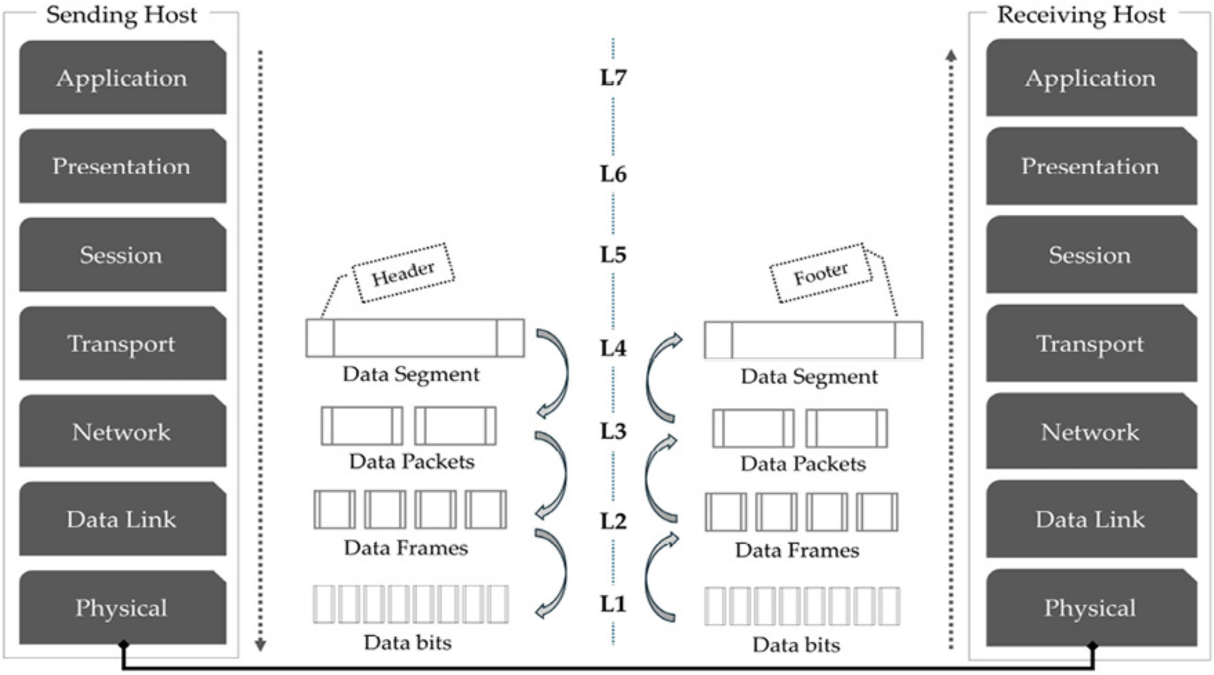


***Figure 6.4**: OSI reference model*

The explanation of the figure is as follows:

- **Layer 7**: The **application layer** is the frontend layer that directly interacts with the user. The user input on the sending node is usually through keystrokes, mouse clicks, or touch on the touchscreen. This is the layer where the data/information is visible to the human eye on the computer screen. The application layer is responsible for rendering the webpage/**command-line interface** (**CLI**) in human-readable form, while it would have received the information in machine-understandable form. The most used protocols are **Hyper Text Transfer Protocol** (**HTTP**), **File Transfer Protocol** (**FTP**), **Simple Mail Transport Protocol** (**SMTP**), and **Domain Name Service** (**DNS**), and secure versions of those like HTTP and SFTP. We cover these protocols later in this chapter.

- **Layer 6**: The **presentation layer** interfaces with the layers below and specifies the formats, encoding methods, encryption requirements, and even data compression methods to use. For instance, you may have noticed that the same site in different countries might open in different languages. In France, you may see the same news website showing the same news content in the French language, while it opens up in American English in the USA. The presentation layer provides such a framework.

Tip: A common encoding standard used for the English language is the American Standard Code for Information Interchange (ASCII). It uses numbers 0 to 127 to represent symbols like @ and the English alphabet in upper and lower case.

Tip: Unicode can encode over 1.1 million characters and is now one of the most common encoding standards in the world, directly contributing to the globalization of content in a reliable, secure, and stable way.

- **Layer 5**: The **session layer** is responsible for initiating, coordinating, and terminating connections that an application or service desires. The session layer will attempt to ensure that the requestor node and the recipient node are able to synchronize periodically to ensure continuity of service. For instance, when you log in to webmail, a session from your computer to the server of the email service provider is established. A session remains active with periodic synchronization between your computer and the server for the duration you use the machine. Sessions usually terminate after a period of inactivity, but some require manual termination, such as logging off.

Tip: The periodic synchronization in the session layer makes the continual availability of applications and services possible. Proper implementation of sessions also causes confidentiality and integrity to be maintained.

Tip: It is highly recommended to explicitly log off from your banking websites, email systems, and even computers after you are done or when stepping away from your machine. This terminates the session and protects your digital access from being misused.

In this section MQTT Broker or Server is used interchangeably though the term Server has been standardized.

- **Layer 4**: The **transport layer** uses TCP and the **User Datagram Protocol** (**UDP**) to efficiently and correctly transfer data across the network. Some of the key functions the transport layer handles are:
    - Facilitate applications to communicate between two hosts/nodes on dissimilar networks using service ports.
    - Provide for *multiplexing,* a process of bringing all the data/signals from a sender and packaging them together for a recipient, and *demultiplexing*, the process of splitting the received signals/data to the correct recipient. To draw an analogy, think about sending multiple papers from all your employees to the tax department after packaging them in an envelope and addressing them to the officer, which can be a form of multiplexing. The taxman will open the envelope and send them to the relevant teams within the department; this is loosely like *demultiplexing.*
    - Combines packets into *segments* to send to network layers and reassembles them when it receives.
    - Manage flow control and error handling.
- **Layer 3**: The **network layer** organizes and selects the best error-free logical path to transmit data packets. This concept of selecting and using the path to transmit is called routing. The other functions of this layer are:
    - Encapsulating data into smaller **packets** to transmit efficiently and to reassemble them at the receiving end. Each packet carries the information on the source and destination node (read IP address). The data that the host wishes to send is also called a **payload**.
    - Within the device, the Network layer has an **interface** for receiving input from other nodes and an interface for output, i.e., for transferring data to another node.
    - **Delivering a packet to the intended recipient** across the network.
    - To assign a **unique logical address**, called an IP address, to devices within a network.

- To transmit the data to an intended recipient, special bytes called *headers* are added as a prefix to the data itself. Headers contain source and destination IPs.
- A *footer*, which is suffixed to data packets, may be used. Footers usually contain data about data and transmission, i.e., *metadata*, such as error-checking numbers called **Cyclic Redundancy Check** (**CRC**) to determine transmission errors or a *checksum* to validate integrity. A checksum is a unique value for any text that can be generated using some algorithms. A checksum will always be the same for the same text when subject to the same algorithm. If there is any change to the data in transmission, the recipient end checksum for the data would not match, hence indicating loss of integrity.

Tip: Networking devices that deliver the OSI Layer 3 functionality are commonly called L3 devices. Routers are the most common of them.

- **Layer 2**: The **data link** layer provides mechanisms to use the physical medium (physical layer), attaches/reads the source and destination address, and ensures the correctness of data (bits and bytes) in transmission and for receipt. For instance, when the bits sent from node one node to another are not received in order, the data link layer will sequence them correctly. It would also be able to detect and discard any data bits that may have been damaged during transmission. VLANs are logically configured on network switches and operate at the OSI Layer 2. The data link layer has two sub-layers:
    - A **media access control** (**MAC**) address is the unique identifier of the physical layer. For a device, consider the MAC address of its **network interface connector** (**NIC**) to be like a US **Social Security Number** (**SSN**) or the *Aadhar number of India*. MAC identifies a unique address for connecting to a physical device. The MAC converts the data packets into frames for a bit-by-bit transmission over the physical layer.
    - **Logical link layer** (**LLC**) manages the flow of the data streams for applications and the error handling of those frames.
- **Layer 1**: The **physical layer** utilizes the hardware components in a network to transport data through electrical or magnetic interfaces.

Consider this to be the layer where the interconnect happens, and the bits and bytes are transported as frames over a physical medium such as a *twisted* or *fiberoptic* cable, *WiFi*, or *Bluetooth*. A computer needs to transmit data over electrical or radio signals and vice versa; this conversion happens at the *physical layer*.

Tip: The OSI Layer also helps IT personnel think and triage network problems, starting first with physical connections/cables and moving upwards to ensure the IP, ports, protocols, sessions, and keys are correctly configured. Several network troubleshooting tools/built-in code are available, such as Wireshark, tcpdump, and nmap.

Evolution of cyber-attacks

The evolving technology with computers, networks, and the Internet has also made disruptions to computers and their networks possible. Some of the infamous disruptive events, being referred to here as cyber-attacks, are illustrated in *Table 6.4*:

Cybersecurity attack	Explanation/Modus Operandi	Damage caused	Impact on remedial action and future of cybersecurity tools
Creeper Virus (1970)	The code self-replicate and move from computer to other on ARPANET and flash *I'm the Creeper: Catch me if you can* the teletype screen (early avatar of modern day computer screen).	None	This was just a proof of concept and triggered the development of an alternate code, called Reaper, that removed the Creeper code. The genesis of one of the first anti-virus.
Bran Virus (1986)	Credited to Pakistani brothers, who created a code in their software to prevent the piracy of their medical. The virus would infect the boot sector of a machine that used pirated software installed from a floppy disk. The virus would flash the home address and phone numbers of the brothers.	None	Prompted the need for organizations to think about anti-virus spreading through floppy drives and planning for backups and recovery.

Cybersecurity attack	Explanation/Modus Operandi	Damage caused	Impact on remedial action and future of cybersecurity tools
Morris Work (1988)	Cornell student *Robert Morris*, created a code (first computer worm) that could self replicate independently. It had no malicious intent and was meant to only prove his point on security flaws.	Within 24 hours, about 10% of world's computers on the Internet were rendered unusable.	Morris was convicted under the then US law, triggering crimes against computers and networks to be taken seriously.
ILOVEYOU worm (2000)	*Guzman*, a Philippines national, created a worm to seek/collate the dialup-username and passwords from computers via an email with subject ILOVEYOU. The unsuspecting user would open the email and trigger an executable that took the credentials and transmit them for Guzman to use. The intent was to defraud unsuspecting people using email handling flaws in Windows 95.	Apart from disrupting scores of computers and their internet access, the worm is infamous for disrupting UK Parliament's computers.	Social Engineering and efforts to protect users from being tricked started evolving. This also brought to light email as a means for cyber-attack and therefore thinking about its protection.
Stuxnet (2010)	It is the world's earliest known nation-state attack using cyber on a country's critical infrastructure. The malware impacted the safe and efficient function of Iran's centrifuges.	Capitalized a software vulnerability/weakness on a **programmable logic monitor** (**PLM**) equipment to damage the connected centrifuges and also propagate through network.	Focus came to attackers could be anyone anywhere and thus an approach to attacks from anywhere on anything became prominent.
NonPetya (2017)	Self-reproducing Russian worm propagating by capitalizing the vulnerability on a Ukrainian accounting software. The malware would encrypt the machine and demand payment to a randomly generated fictious number.	Thousands of computers worldwide made dysfunctional across 60 countries.	Earliest massive disruption brought about user awareness efforts. The most infamous ransomware involving nation-state actors. Cyber insurance started insisting stronger governance on anti-virus.

Cybersecurity attack	Explanation/Modus Operandi	Damage caused	Impact on remedial action and future of cybersecurity tools
Wannacry (2017)	Using a vulnerability on a Windows OS protocol (called SMB).	Made about 200,000 computers dysfunctional across various industries in 150 countries. One of the most effected was the UK's healthcare system.	Vulnerability identification and patch management efforts gained importance even though such capabilities were available.

Table 6.4: *Cyber-attacks and their impact on cybersecurity*

We will cover some details of attackers and their tactics later in the book.

The motives of such attacks were one or more of the following:

- The spirit of experimentation, i.e., to prove a technical point, sometimes inadvertently caused a disruption.
- *Script Kiddies*, young kids with access to coding, could create programs and cause nuisance or even actual disruption to organizations.
- Nation-states began using cyber as a weapon to bring harm to their enemy.
- Cyber criminals are motivated by financial incentives.

Algorithms used for data protection

In *Chapter 1, The Triad of Security,* we learnt that the focus of information security is on the CIA. The need to protect data stored on computers/servers and when in transit within a network is protected through technologies that rely on one or the other algorithms. An algorithm is typically a mathematical function that takes an input and transforms it.

One-way functions, hashing

A key concern in data transmission is its integrity. For instance, an e-commerce retailer would not like any unauthorized changes to its transactions, such as an increase or decrease in the price of a commodity it sells. If the payment system used by the e-commerce retailer organization is not appropriately protected, the shopper's cart can be modified by cyber criminals or even pranksters for quantity and/or price. This would erode customer trust and may even cause financial implications. Similarly, when a sender wishes to transmit a message, s/he would like to ensure it reaches the recipient without any alteration in transmission.

Special one-way algorithmic functions can be used to generate unique fixed-length bytes/values, called a **hash**, for any text/message. This message is also called a **message digest.** The hash of the same text, when using the same algorithm, will always be the same; any modification of the original text will change the hash. Therefore, this process is deterministic and returns the same-sized hash irrespective of the size of the input. The process of using one-way functions to compute the hash is called **hashing** algorithms/functions. This process is one-way because even if you pass the generated hash to the function, it will create a new hash but not the original text.

The recipient generates the hash of the message it receives using the same algorithm and compares it to the hash value provided by the sender. If those match, then the message can be guaranteed not to have been tampered with, i.e., its integrity was not compromised. It can be inferred that the message's confidentiality may still have been at risk/compromised. Hashing is also commonly used to determine the integrity of system files, installable programs, and key documents.

Figure 6.5 shows how hashing algorithms can be used to compare the hash value and determine if the message was intercepted and changed in transit:

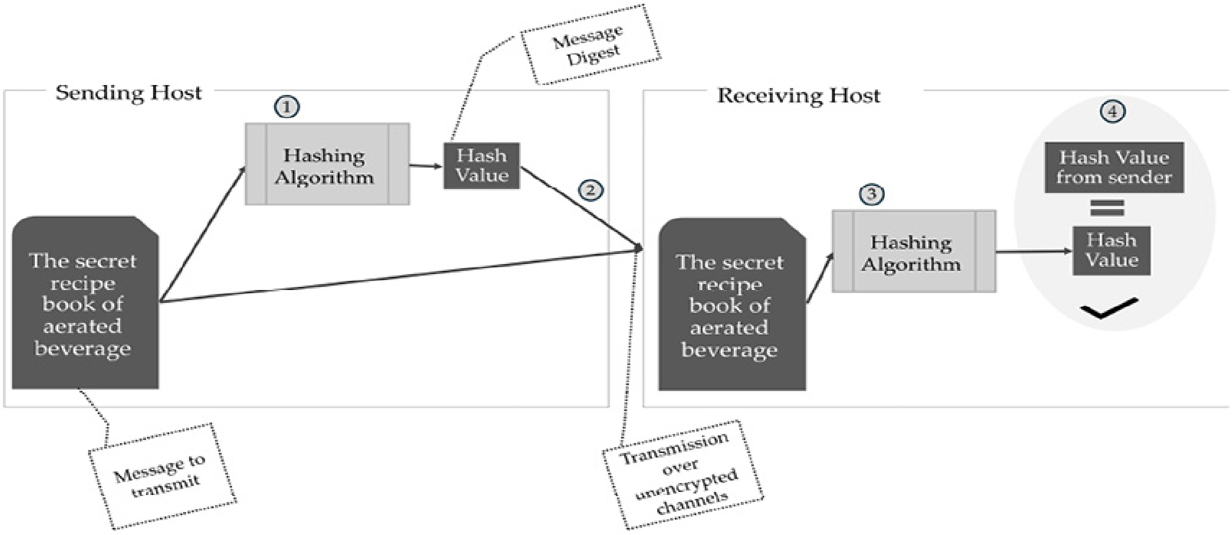


Figure 6.5: *Using hashing to determine message integrity*

Some of the popular hashing functions in use today are:

- The message/text is used to create a 16-bit hash of the message by using algorithms called **message digest** (**MD**). The 5th version, MD5, is still partially in use. It generates a 128-bit hash value for every message/plain text. Research has proven that it has weaknesses and is no longer fit for use.
- **Secret hashing algorithm** (**SHA**), the first version of SHA-1, generates a 160-bit hash value and was considered an able replacement for MD5 for encryption uses. However, the algorithm is already deprecated. SHA-2 supports variable bit lengths, namely 224, 256, 384, and 512, for the hash values. The hash key length is notated, such as SHA-224 or SHA-512.

Tip: Hashing is often used by software companies on source code to generate a hash and publish it. The user/buyer of software can compare the original hash value to ensure the software has not been tampered with during distribution/on the Internet.

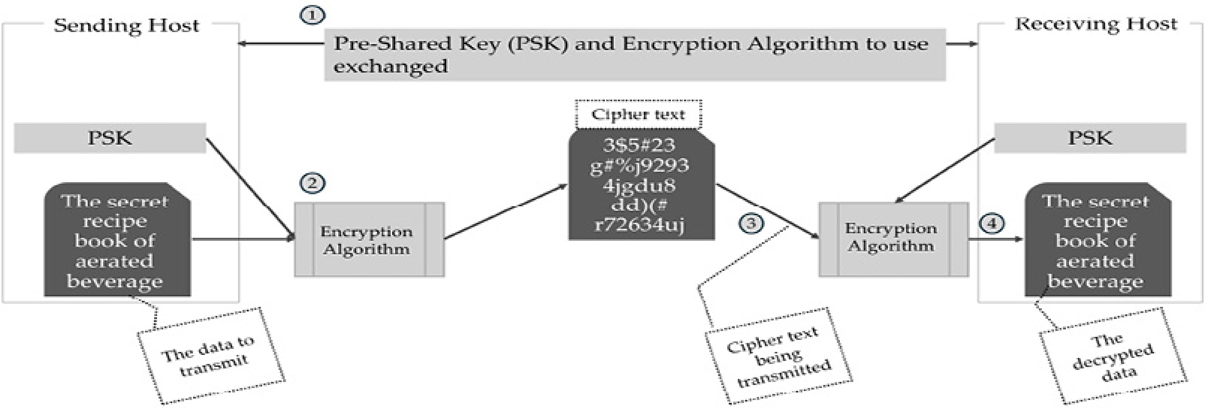
Encryption

Protecting data from unauthorized access or use is paramount for businesses, especially over networks. One of the most important ways is to encrypt the data. Etymologically, encrypt is derived from a Greek word and means hidden or concealed. Encryption is based on principles of cryptography, the science of using codes to obscure or hide information. The use of encryption is a centuries-old practice where kings and their courtiers would be able to send and receive messages secretly. Even if the message was seized by the enemy, the inscription was difficult to decipher. Thus, the message was protected for its confidentiality. In the infamous *World War 2*, the German army used encryptions for all its military messages; They used a device aptly named **Enigma**. *Alan Turing*, a British mathematician, led his team and is credited with breaking the Enigma code during World War. This event is considered one of the major reasons for Allied forces to have defeated the German-led enemy forces. Encryption algorithms use secrets to convert plain text (i.e., human or machine-readable) to create *ciphertext* in a way that only the intended destination can decrypt (reverse the encryption) to plain text. While on the computer, the text will still be binary, but the ASCII equivalent would not be decipherable/meaningful. This means that confidentiality is protected. Needless to say, the secret key used must be only known to the sender and receiver, else an intruder/eavesdropper shall be able to get unauthorized access and may potentially change the content and transmit with the same secret key and encryption algorithm, as a new ciphertext to the original intended recipient. Encryption algorithms operate in the *OSI Layer 6: Presentation*.

There are two types of encryption:

- **Symmetric encryption**: The sender and receiver use a pre-decided secret key and the encryption algorithm to exchange data. This type of encryption is useful when the focus is on efficiency, the volume of data to encrypt, and the exchange of the secret key itself is not a concern. The secret key is of varying length between 128 bits and 256 bits; these are also called *encryption bits*. Common uses of symmetric encryption are for VPNs, databases, and local file encryption. The steps involved in the symmetric encryption are as follows (shown in *Figure 6.6*):
    1. The **pre-shared key (PSK)** and the encryption algorithm to use are exchanged and shared between the hosts.
    2. The encryption algorithm then uses the PSK and the sender's text to create the encrypted text, called cipher text.
    3. Cipher text is transmitted using TCP/IP protocols to the recipient's IP address.
    4. The encryption algorithm uses the PSK and the cipher text to decrypt it and get the original text

  See *Figure 6.6,* which shows the steps in symmetric encryption:



Sending Host
1
Pre-Shared Key (PSK) and Encryption Algorithm to use exchanged
Receiving Host
PSK
The secret recipe book of aerated beverage
2
Encryption Algorithm
Cipher text
3$5#23 g#%j9293 4jgdu8 dd)(# r72634uj
3
Encryption Algorithm
4
PSK
The secret recipe book of aerated beverage
The data to transmit
Cipher text being transmitted
The decrypted data

***Figure 6.6**: Symmetric encryption*

- **Asymmetric encryption**: The sender and the receiver use a pair of keys to encrypt the data. The sender encrypts using the intended recipient's *public key*, and the intended recipient uses their secret *private key*. In asymmetric encryption, the public key is known to all, and each user's private key must only be known to the user. The key length is usually around 2048 bits, largely attributed to the mathematical formula used in algorithms. Common uses include TLS protocols, when we use a HTTPS website, encrypting emails, and or in digital signatures. The steps involved in the symmetric encryption are as follows (as shown in *Figure 6.7*):

  1. The sender uses an encryption algorithm on the data it wishes to encrypt and sends it using the receiver's public key to create the ciphertext.
  2. A packet header is added to the cipher text with information on key length, encryption algorithm used.
  3. The receiver uses their own private key to decrypt the message using the algorithm and key details from the header.

  Figure 6.7 shows an implementation of asymmetric encryption using the public-private key pair:

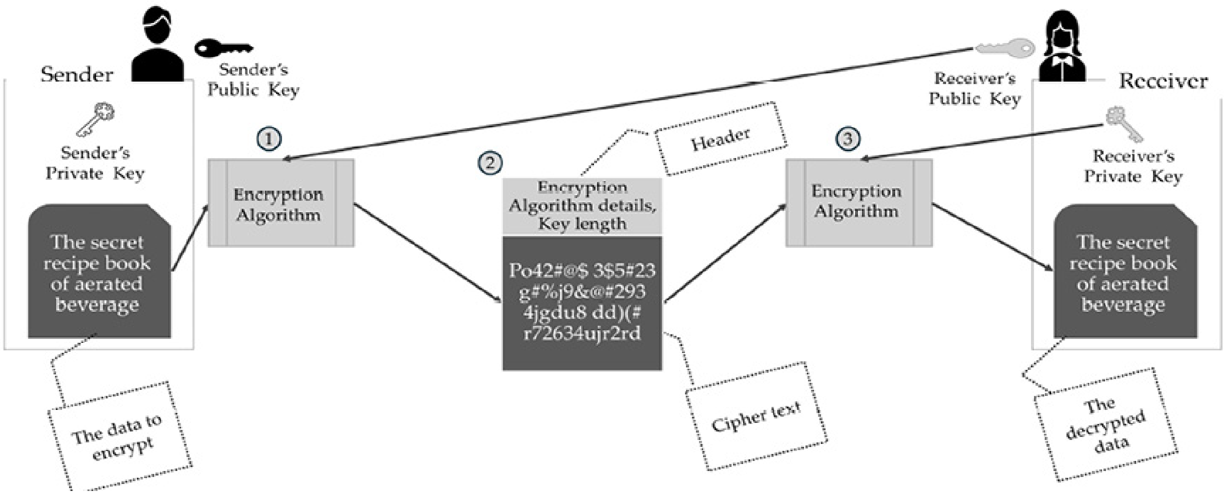


***Figure 6.7**: Asymmetric encryption*

Tip: To break even 256bit encryption requires immense computational resources and many years. With the advancement of technology and processing power, it is entirely possible for such algorithms and secrets to be compromised. However, for now, these are considered safe enough to use.

Asymmetric encryption is also called **public key cryptography** or **public key infrastructure** (**PKI**). PKI is an integral part of trust certificates on websites. These certificates are also called *SSL certificates*.

Some of the common encryption algorithms used today are tabulated in *Table 6.5*:

Algorithm	Encryption type	Key properties	Important use cases
Data Encryption Standard (DES)	Symmetric	Has short key length of max 56 bits. Already been cracked and is no longer considered safe for confidentiality uses.	Now defunct.
Triple Data Encryption Algorithm (3DES)	Symmetric	Encrypts the data 3 times using DES.	Bank ATMs. Unix operating systems internally use 3DES for its passwords.
Advanced Encryption Standard (AES)/Rijndael Encryption	Symmetric	Has 128bit encryption but can also support 192 and 256bit encryption.	Several protocols we use on the internet such as HTTPS, IPSec.
Blowfish	Symmetric	Breaks the message block into 64 bits and encrypts with individually using variable length encryption bits from 32 to 448 bits.	Encrypting local files.
Twofish	Symmetric	Breaks the data block into 128 bits and uses a variable length of encryption bits ranging from 128 to 256 bits. Implements a concept of randomizing and obscuring data, called *data whitening*, in the 128 blocks using part of a key even before actually encrypting the data. Can run on 32-bit processors.	Disk encryption.
Rivest-Shamir Adleman (RSA)	Asymmetric	Uses prime numbers in its algorithm to encrypt, decrypt, and exchange key information and digital certificates. Is widely used inspite of being resource intensive and having limitations of data	Digital Security certificates used in websites. Encrypting data in transmission.

		packet size (typically 234 bits).	
Diffie-Hellman (DH)	Asymmetric	Sender and recipient agree on a shared public key based on mathematical formulas involving prime numbers, root and modulo and then generate their respective private keys for their own use.	Used for key exchange, several network protocols for secure connection.

Table 6.5*: Encryption algorithms*

Digital signatures

Digital signatures are an accepted practice in businesses for signing agreements, contracts, and emails, filing and signing income-tax returns, etc. The digital signature uses principles of hashing and encryption together to authenticate and validate the sender. It also establishes the creed that the sender did send the message. This is called non-repudiation, i.e., taking/establishing accountability of the sender.

The **Digital Signature Standard** (**DSS**) specifies the requirements of algorithms that can generate a digital signature and how to use them. It ensures control over integrity but not the confidentiality of the message itself. The process of attaching a cryptographically generated value to a message is called a *digital signature.* The digital signing involves the following steps (as shown in *Figure 6.8*):

1. A one-way hashing algorithm is used to generate the message digest for the message.
2. Sender uses his/her private key to encrypt the message digest itself.
3. The message, along with this signed message digest, is sent to the recipient.
4. The receiver uses the public key of the sender, decrypts the packet to get the original hash. This signature can only be decrypted by the same public-private key pair, and thus a successful decryption implies the sender was authenticated. This concept of decrypting a digital signature using the sender's key pair is also called the **Authentication Header** (**AH**) protocol.
5. The receiver regenerates the hash of the message received and matches it with the sender’s decrypted hash. If they match, then it can be trusted that the message was not altered, and the sender is validated.

Figure 6.8 depicts the use of a digital signature. Notice that the actual message is not encrypted here, but it also can be.

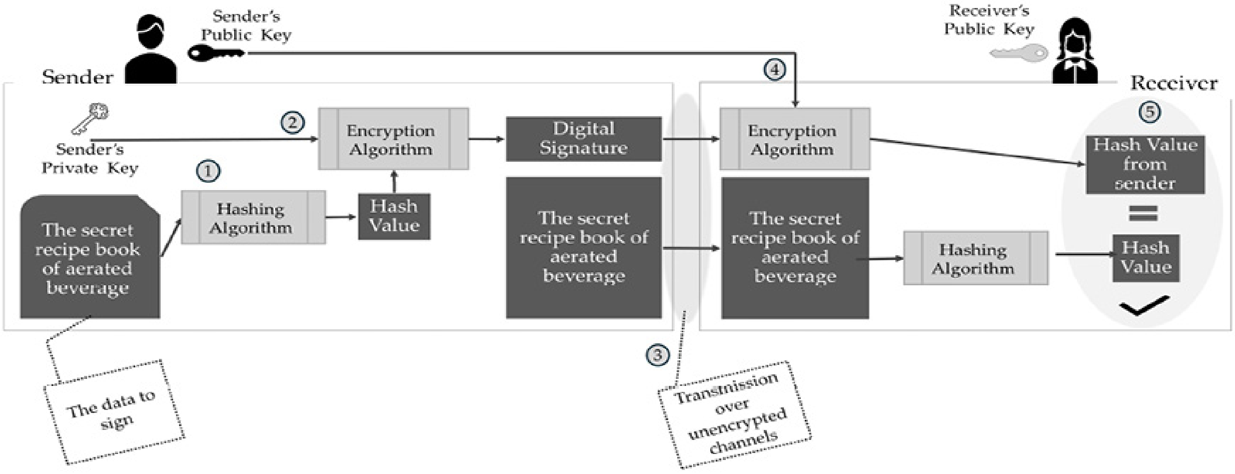


***Figure 6.8**: Digital signature*

PKI is a key technology control in use for several of the purposes defined above. The key components of PKI include:

- **Certificate authority (CA)**: It issues and manages the digital certificates and the authenticity of the certificate holder. For instance, all major e-commerce websites use the SSL issued by trusted CAs such as Verizon, Digicert, Amazon, Entrust, etc.
- **Digital certificates**: Issued to subscribing users. This could be all users of an organization.
- **Certification revocation list** (**CRL**): A very important list that tells a requestor if the certificate is still valid or not. Modern-day browsers automatically warn users on their page of an expired or revoked certificate.
- **Certificate practice statement** (**CPS**): The policy and procedure a CA will use to issue, manage, and revoke certificates. It also carries the documentation on algorithms, key length to use, and so on.

Organizations may choose to use external CAs for external websites to enhance customer trust versus managing a PKI setup internally. Organizations may choose to manage and maintain an internal CA/PKI for internal consumption, such as issuing its email users digital certificates for internal use or for its internal websites and digital certificates for internal devices.

Tip: NIST has a couple of standards, namely FIPS 186-5 and NIST SP 800-57, for PKI.

Protocols and ports

Protocols can be thought of as a way of doing things in a specific way based on agreed-upon rules. In the context of technology, a protocol defines the way data can be exchanged between hosts for a defined purpose(s). We will cover the family of TCP/IP, **User Datagram Protocol** (**UDP**), and some of the most common TCP and UDP protocols.

Ports, on the other hand, can be imagined in the digital world, akin to waterways or airports, where vessels dock/land. Each port in the physical world has a code or a number, such as London Heathrow's airport code, which is LHR, and Hyderabad, India's code, which is HYD. Similarly, in the world of technology, connections between devices happen over/via a port and on layer 4. For instance, the **Simple Mail Transfer Protocol** (**SMTP**) is a TCP protocol that lays down the specifications and rules for a computer to send and receive emails over port number 25. In 1981, the US **Department of Defense** (**DoD**) published the standard for TCP for ensuring reliability in *interprocess communication* amongst hosts. The port numbers range from 0 to 65536 and are primarily the reason that different applications and services on the same computer can share the computer's infrastructure. You might imagine these ports to be different airplane parking bays in the same airport.

The layered protocol architecture, as shown in *Figure 6.9,* was introduced in the DoD standard referred to as **RFC 793**. The figure also shows the purpose of each OSI layer:

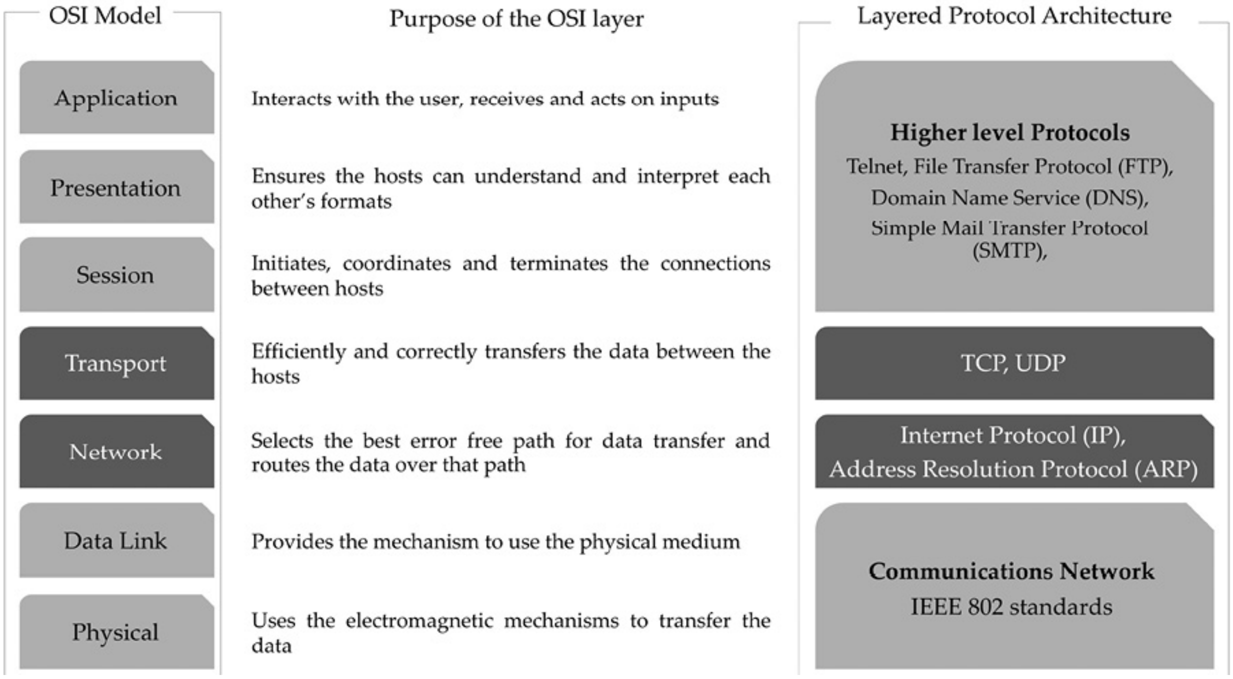

OSI Model
Purpose of the OSI layer
Layered Protocol Architecture
Application
Interacts with the user, receives and acts on inputs
Presentation
Ensures the hosts can understand and interpret each other's formats
Session
Initiates, coordinates and terminates the connections between hosts
Transport
Efficiently and correctly transfers the data between the hosts
Network
Selects the best error free path for data transfer and routes the data over that path
Data Link
Provides the mechanism to use the physical medium
Physical
Uses the electromagnetic mechanisms to transfer the data
Higher level Protocols
Telnet, File Transfer Protocol (FTP),
Domain Name Service (DNS),
Simple Mail Transfer Protocol (SMTP),
TCP, UDP
Internet Protocol (IP),
Address Resolution Protocol (ARP)
Communications Network
IEEE 802 standards

***Figure 6.9**: Layered protocol architecture*

UDP is a connectionless protocol and is very useful for the host to be able to transmit data without requiring confirmation of receipt. For instance, a server can be set to periodically broadcast the time of the day for any host desirous of using it. This protocol is called **Network Time Protocol** (**NTP**). In the physical world, this is akin to the clock towers many cities have. TCP, on the other hand, is a connection-oriented protocol. The core purpose of TCP is to reliably transit the packet to the intended destination.

TCP handshake

The TCP uses a mechanism called a 3-way handshake to reliably establish a connection between the source and destination/receiving hosts before it starts to send data. The TCP 3-way handshake process is as follows:

1. The source host sends a packet called *SYN* to the destination to ensure that both hosts can synchronize.
2. The destination host responds back with a packet called SYN-ACK to acknowledge the synchronization request and that it is ready to receive data packets.
3. The source host then sends a packet called ACK, letting the destination host know that it received the SYN-ACK and that the reliable

connection is now established.

See *Figure 6.10* for a pictorial view of the handshake over port 4321:

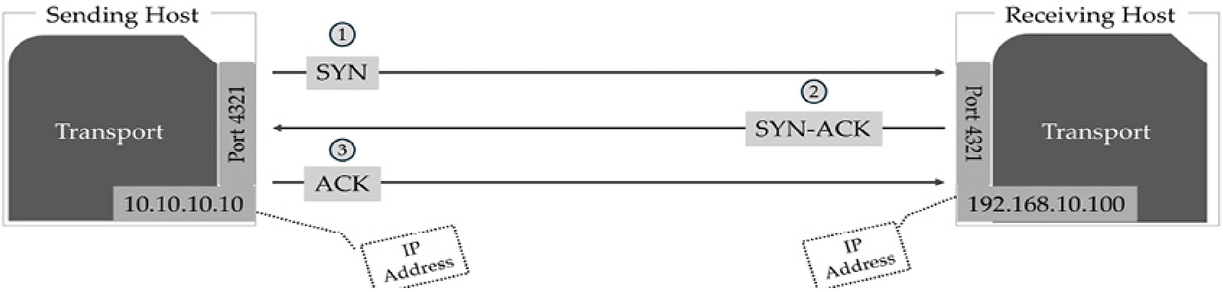


***Figure 6.10**: 3-way TCP handshake*

Common ports and protocols

Table 6.6 lists the common protocols and their standard purposes:

Protocol	Purpose	Port number	OSI Layer
Neither TCP nor UDP			
Address Resolution Protocol (ARP)	Communication between hosts using MAC address.	-	2
Internet Protocol (IP)	Transmission of data over network.	-	3
Internet Control Message Protocol (ICMP)	Used by network layer to check connectivity.	-	3
TCP			
Secure Shell Protocol (SSH)	To remotely access hosts over the network in a secure fashion.	22	7
Secure File Transfer Protocol (SFTP)	Securely transfer files between hosts (uses SSH).	22	7
Domain Name Service (DNS)	Resolves names of websites to IP addresses.	53	7
Hyper Text Transfer Protocol (HTTP)	Allows host browsers to access webpages/websites on webservers.	80	7
Hyper Text Transfer Protocol (HTTPs)	Allows host browsers to securely access webpages/websites on webservers. Uses SSL or TLS.	443	7

Protocol	Purpose	Port number	OSI Layer
Secure Socket Layer (SSL)	To authenticate communications on the network and encrypt the data between server and hosts.	443	5
Telnet	Remote access to a computer, the protocol is insecure and not recommended to be used.	23	7
Border Gateway Protocol (BGP)	Allows networking equipment to share network information and routes with each other.	179	3
Lightweight Directory Access Protocol over SSL (LDAPS)	Secure version of LDAP that authenticates and enumerates users and devices.	636	7
Transport Layer Security (TLS)	Security of communications between networks, similar to SSL but is faster, more robust and better at handling security keys.	443/465	4 to 7
UDP			
Dynamic Host Configuration Protocol (DHCP)	Assigns IP address and other network configurations to hosts.	67/68	3
Domain Name Service (DNS)	Reliable information sharing between DNS servers.	53	7
Network Time Protocol (NTP)	To advertise a common time for hosts to synchronize their clocks.	123	7
Internet Protocol Security (IPSec)	IPSec is used for establishing secure connectivity such as VPN.	50 / 4500	3

***Table 6.6**: Common ports and protocols in OSI layer*

A protocol itself may be part of the OSI architecture at higher layers, but shall use the ports and transmission mechanisms in the lower layer. For instance, the DNS protocol operates on the application layer. It uses port 53, which is established on the session layer, and the transmission (TCP/UDP) and routing (IP) are managed by the transport and network layer, respectively. Similarly, a DHCP operates at the network layer (3) when assigning IPs to hosts, but may leverage the capabilities of the data link layer (2) when discovering servers on the network.

Tip: SSL is responsible for encrypting the data/emails and operates on the OSI Network layer,

while HTTP uses SSL/TLS to encrypt web traffic and operate on the OSI Application layer.

Address translation protocols

To be able to route the traffic on the internet, the source IP and the destination IP must both be public IPs. To route a private IP to a public IP, **Network Address Translation** (**NAT**) is used. NAT takes a private IP and masks it with a public IP for outbound (to the internet) traffic. External IPs will not see which specific private IP made this request. A device on the network, such as a router, is designated as the network gateway. The interface on the LAN side has an IP from the same subnet, while the other interface on the device has a public IP. For any outbound traffic, also called egress traffic or north-south traffic, the gateway device translates the IP to a public IP and maintains a record of which internal host attempted to make what external service request (such as access to a website). When the external server responds, it retranslates the public IP to the right private IP and sends the traffic to the right host. Each of the hosts in a subnet/network can use the same public IP or a pool of the organization's public IPs. The routers need information about public IP and their hosts, and the available paths to them. The routers maintain this information in the routing table. The shorter the routing table, the faster the traffic flow. Core Internet routers use a protocol called **Border Gateway Protocol** (**BGP**) to publish optimal routes that other routers can use.

A server on the organization's network may be hosting its corporate website, and the same server may also be the mailing server. We learned earlier that the https works on port 443, while mails (SMTP) works on port 25. The perimeter devices can receive requests from different clients for different services on the same server by using **Port Address Translation** (**PAT**). It uses the socket (such as **`https:200.10.200.10:8080`**) to send it to port 8080 on the internal host that hosts the website.

Network Time Protocol

NTP is a connectionless protocol that publishes a standard time that every device on a network can synchronize to so that the time on each device will be exactly the same. The requirement to use NTP is fundamental in several regulations worldwide to ensure devices can keep to a standard time. This comes in handy in several business scenarios, but most importantly for security incident investigations. The time at Earth's longitude 0, which is at the English town of Greenwich, is called the Coordinated Universal Time/**Universal Time Coordinated** (**UTC**). The time configured on the server running this protocol can be set:

- Manually, but it is tedious and may be impacted by conditions such as the battery quality on the motherboard.
- Synchronize to publicly available NTP services from large organizations, such as Google, Microsoft, or NIST.
- Synchronize with atomic clocks.

Tip: India's time is UTC + 5:30. It means that when it is midnight at Greenwich (longitude 0), the time in India is 5:30 am.

Domain Name Service

Instead of the dotted decimal notation for websites, we use a name, called a **domain name**. The domain name makes it easier for organizations to associate their digital presence brand with a name for their consumers/stakeholders to remember and use. Even Government bodies have domain names as they also wish to communicate their work to citizens and other services people may need, such as passport issuance or filing income tax.

The domains that end with `.com`, `.info`, `.gov`, `.edu`, etc., are called **top-level domains** (**TLD**). The DNS servers maintain the list of domain names and the actual servers and their public IPs associated with them. For instance, Google.com is one of the TLDs and a very popular domain name of a very popular search engine in the world. The DNS servers will maintain that the `google.com` domain is hosted on a public server with a known IP address. *Table 6.7* lists the various domain-related information that DNS servers have:

Name of the record field	Purpose
A Record	Contains the IP address of the webserver that hosts the website.
Canonical Name (**CNAME**)	The alias name of the host.
Mail Exchanger (**MX**)	The hostname of the SMTP server for the domain if any.

Nameserver (NS)	The server that can provide the details of the DNS information for a domain.
Service of Authority (SoA)	Indicates whether it is an authoritative source for a domain and periodicity of updates to the domain information.

***Table 6.7**: DNS record*

A user types the website they wish to reach on their browser. The browser (working at the *OSI Layer 7: Application*) sends a UDP request over port 53 to a DNS resolver, which in turn contacts servers called Authoritative Name Servers to get the actual IP address the domain is hosted on. This IP address is relayed back to the user's machine. Organizations often store the already resolved DNS information to be able to skip the DNS resolver query on the Internet and thus make browsing a tad faster. This does run the risk of DNS not being updated if the actual website changes its hosting location.

Tip: To protect the confidentiality and the privacy of users' access to websites, DNS queries and responses can also be encrypted using DNS over HTTPS (DoH).

When a new website/domain name is registered, the nameservers they will use may also be provided. The authoritative nameservers, in turn, use TCP port 53 to periodically send such updates to nameservers. This ensures that all websites, new or old, are always associated with an IP that requesting hosts can reach. It is possible to register a domain name without providing any nameservers or the public IP of the server it will be hosted on. Such domain names will not be reachable. Organizations often choose to register several similar-looking/spelled domains for their brand to protect it. **Cyber-squatting** is a tactic used by ill-meaning individuals to profit from registering and/or using a domain that resembles a brand, trademark, or even a famous individual. It is also a common tactic used by cybercriminals to commit fraud or even host inappropriate content. While not directly linked, this tactic can cause the availability of the brand on the Internet and may even lead to registered user credentials to popular websites being stolen, causing a loss of confidentiality and integrity. The domain names we use on the browser are also called **Uniform Resource Locators** (**URLs**). We will explore some of these tactics in later chapters.

Tip: Organizations register misspelt names of their brand to redirect traffic to their main website. For instance, aple.com will automatically redirect to apple.com.

Tip: Yahoo.com successfully won the case against an individual who registered and had a

cyber-squat on YahooIndia.com. The court ruled in favor of the company as the domain name was very close to the more popular brand and amounted to the possibility of it being misassociated with that brand.

Internet Protocol Security

On the Internet, millions of computers connect. It is possible to eavesdrop and know what is being exchanged between different hosts. For instance, an eavesdropper can see all requests and responses between the host and web server in clear text for a website using only HTTP. This poses a serious threat to confidentiality, especially when dealing with PII. A news website or an online library of books/music is meant to be accessible by everyone for use; confidentiality is not a constraint. Confidentiality and integrity are key to banking sites, and for web-based emails, they should be. All sites that require authentication and/or have PII or PHI must use HTTPS. Similarly, **Secure FTP** (**SFTP**) must be implemented vs. FTP.

In case an organization wishes to use a secure method to transport information between two hosts, irrespective of protocols, it can set up an IPSec connection.

IPSec is a family of protocols that allows a secure, encrypted connection between two hosts and even authenticates the source of the data, providing for CI goals. While IPSec typically uses UDP port 500, it can switch to TCP if packets face problems with UDP; this allows for availability goals to be considered. The protocol is commonly used in setting up and operating a VPN.

An IPSec connection can operate in two modes:

- **Tunnel mode**, also called IPSec tunnel, wherein everything between the two hosts is authenticated for source, and everything that is exchanged between the two hosts is encrypted.
- **Transport mode** wherein only the payload is encrypted. This attempts to maintain the confidentiality and integrity of the data.

IPSec Tunnels can be created between various office locations, such as the main office and branches, and are called site-to-site tunnels or site-to-site VPNs. Similarly, remote users can also make VPN connections to the organization, which are called client-to-site VPNs. In both these scenarios, organizations can capitalize on the IT infrastructure and provide users with the IT services they need. IPSec uses encryption, hashing, and digital signature technologies to provide confidentiality and integrity. In today's time, VPNs between various offices of the organization are established over the Internet but can also be established as a WAN technology. *Figure 6.11* shows a representation of the client-to-site (or remote VPN) over the Internet and site-to-site VPN using extranet (WAN). The pipe in the diagram represents the tunnel to symbolize that no one can eavesdrop on the tunnel to compromise confidentiality.

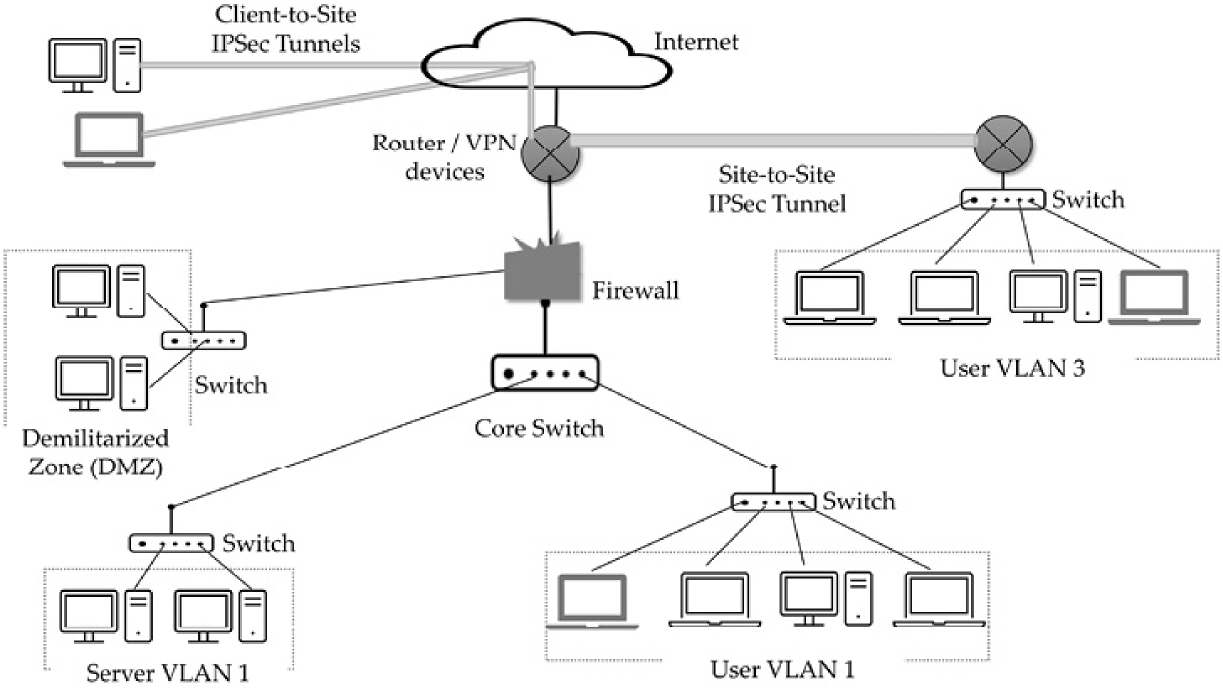


***Figure 6.11**: IPSec for VPN connections*

Key network security technologies

A key role of network security technologies and tools is to:

- Enable the required access to business digital services. For instance, the e-commerce retailer would want its portal/website to be available to its customers.
- Block any cyber-attacks that can impact the CIA.
- Implement the principles, such as need-to-have, need-to-know, that we covered in *Chapter 1, The Triad of Security*.

A network for a typical organization is shown in *Figure 6.11*. The key components are listed as follows:

- **Router (R)** handles the routing functions described in the OSI model. It is placed as the first line of defense for inbound traffic from the Internet to the corporate network. It also serves as an *egress* point, i.e., an exit path for outbound traffic from the corporate network. Routers use **access control lists** (**ACLs**) to define/specify the rules for traffic permitted or denied. The modern-day routers can be configured to automatically prevent cyber-attacks and any anomalous traffic. Routers are often denoted by the letter *R* in network diagrams.
- **VPN** devices, sometimes called VPN *concentrators* for their ability to congregate all VPN connections, allow remote users to connect to the network. The VPN users will generally have the same access as if they were in the office building. VPN devices use IPSec tunnels/SSL technologies-based tunnels to protect communication and data over insecure channels such as the Internet. Users will use the VPN client on their host machine to connect to the pre-configured VPN server on the corporate network and authenticate to get valid access. Organizations use authentication methods, including multi-factor authentication, to allow only legitimate users to connect. It is important to ensure that proper ACLs based on *need-to-have* and *need-to-know* are implemented on VPNs so as to protect the organization's network from unwanted remote user risks.
- **Firewalls** (**FW**) are also an egress device that uses rules to protect the IT infrastructure for inbound and outbound traffic at ports and protocol levels. Firewall rules are represented with reference to the source, destination, port/protocol, and permission. For instance, Any-*UserVLAN-All-block* implies the firewall is configured to block all inbound traffic from the Internet that is meant for internal user VLANs. This is done because the user VLAN accesses required services on the Internet through specific server(s). Firewall rules are processed sequentially until the condition is met; the most restrictive rule should be in the correct sequence, allowing required traffic to succeed but block unwanted ones. Similarly, the firewall can block any network traffic from user VLANs to the **demilitarized zone** (**DMZ**). The DMZ is a

virtual network for services that the organization wishes to accept from the Internet. For instance, the e-commerce portal and its payment gateway will be placed in the DMZ. Firewalls have evolved over the years to automatically even prevent intrusion attempts. Such new-age firewalls are called **next-generation firewalls** (**NGFW**). NGFWs use a concept called **stateful inspection**, i.e., it is able to maintain the context of the packets arriving in the stream of transmission and thus recognize any anomaly. FW, on the other hand, performed stateless inspection, i.e., inspection of each packet as an individual packet. It may be recalled that in data transmission, data packets are broken into smaller portions to allow the network protocols to transmit them efficiently. NGFW operates at *OSI Layer 7: Application*, while the FW generally operates only at the *OSI Layer 4: Transport*.

Tip: One of the most common pitfalls in firewall rules implementation is having a last row rule as any-any-any rule allows. This implies that any traffic towards or from the Internet is allowed. Using the principles of need-to-have, this should be a block.

- **Virtual local area networks** (**VLANs**) configured on the switch use ACLs to control allowed and disallowed traffic and add the required granularity of control to protect confidentiality and integrity.

  Some of the key protocols already discussed are not represented in *Figure 6.11*, but their role is essential. For instance, the DHCP, DNS, PKI, Internet proxy, and email gateways, etc.

- **Web proxy**: It is used to NAT outbound Internet traffic for port 80/443 from user or server VLANs to the internet and provides filtering of URLs using categories. Organizations would generally restrict websites that are not fit for use in their organizational culture and/or not in line with their business use. They may also do so to protect bandwidth consumption on the Internet and to ensure user productivity. For instance, organizations typically would not allow a gambling website or one showing pornographic content to be accessed from their machines. An organization in the business of online gaming will typically allow all its users to access other gaming websites on the internet. Internet proxies also have a feature of storing previously and commonly visited pages on their server, thus making the subsequent webpage requests for those sites load faster.

- **Email gateway**: Life without email seems near impossible; at the same time, the non-repudiation of emails and disallowing email spoofing are key aspects of email security. Email gateway solutions provide such features. The SMTP server of the organization is protected by an email gateway, which can filter malicious emails, spam, and spoofed emails. Modern-day email gateway solutions can also rewrite a URL in the body or the subject of the incoming email to protect it from harm. Such technologies are becoming very relevant and table stakes given the increasing rise in email-based attacks (like phishing) on organizations. We cover several of these attacks later in the book.
- **Intrusion detection systems** (**IDS**) are technologies that protect the devices on a network from any inbound Internet attacks. Generally, these technologies use the knowledge of cyber-attacks in the global network to see if a similar pattern is visible in the organization's network. In case it does detect such an anomaly, the system is configured to alert the IT administrator/security manager for immediate action. **Network IDS** (**NIDS**) is deployed for a network or a LAN and protects it from network attacks. **Host IDS** (**HIDS**) is typically deployed for servers, like databases or applications, to protect them from intrusion attempts.
- An **intrusion prevention system** (**IPS**) can detect the anomalies, an IDS can alert the right authority, and can be configured to automatically take evasive or corrective action. For instance, if the IPS senses a lot of unexplained and/or malicious traffic coming to the organization's internally hosted web server and deems that it will overwhelm the server, causing availability issues, the IPS can simply block the source of the traffic. **Denial of service** (**DoS**) is a common attack where so much traffic is sent to a server that it will not be able to handle it and/or crash, i.e., stop functioning. This implies that the users requiring those services will be denied service. We will cover DoS and other attacks later in the book.
- **Data loss prevention (DLP)**: These technologies aim to prevent any data from leaking/being sent unless authorized. They can be implemented on email gateway, web gateway, and endpoints to protect egress from machines using email, web, and data transfer. These tools

have the context of data based on their labeling and/or pattern. For instance, DLP tools can detect credit card numbers and US social security numbers and prevent them from being sent. There may be limitations of DLP's capability, such as PII in compressed files, encrypted files, and content in languages other than common ones like English. However, they are a key element of modern-day data security.

Devices such as firewalls and routers at the digital boundary of the organization are called **perimeter devices**. Other external-facing websites, email gateways, and web proxies are typically protected behind a firewall, even though they may have an external IP and ports open on the Internet.

Another key element of a secure environment is to ensure the configuration of devices meets or exceeds the required policy and standards. In *Chapter 3, Role of Standards and Controls,* we covered controls and control configurations. Organizations must ensure the device configurations are consistent and are periodically updated in line with business requirements and environmental changes. These device configurations are documented and published as internal standards and are called **hardening documents**. Some of the important hardening elements to consider for appropriate and safe configuration are:

- Changing the default manufacturer administrator ID and password. Even today, original equipment manufacturers often build and ship their products with administrative usernames and passwords as admin/password. These documents are also easily available on the internet. Cybercriminals can use such information and gain unauthorized access.
- Local users and/or guest users must not be created. Or must be kept disabled.
- Unwanted ports and protocols should be turned off. For instance, telnet is a protocol that is not secure and must not be used, even though the OEM may not have deprecated the protocol on their product. Similarly, where possible, the TLS protocol must be used instead of SSL.
- Excessive and unwanted permissions must not be set up. The principle of need-to-have must always be applied.
- Administrators of key technologies must use strong authentication such as MFA.

- OEM-specific hardening guidelines must be considered and implemented where possible.

Endpoint security

Servers and end-user machines are the core of how today's businesses operate. Several of these devices use the Internet and are thus directly at risk of having their data compromised. Technologies are required to protect these endpoints and data within them. Some of such technologies include:

- **Disk encryption**: Using encryption technologies, the entire computer disk is encrypted, called **full disk encryption** (**FDE**), or to make it more efficient and practical, only portions of the disk/storage are encrypted, which is called **used disk space encryption** (**UDE**). It is also possible to encrypt only certain files on a disk and volumes of server disks. Such a practice has management overheads but is cost-effective and operationally efficient, as encryption-decryption time is avoided. Databases can also be encrypted to protect the PII/PHI they contain. Any of the common encryption algorithms, such as AES, may be used.
- **Antivirus**: To protect host machines against computer viruses. These are typically signature-based and are dependent on the virus definitions to be continually updated to be effective. Viruses have the ability to self-replicate and spread over the network once they infect a host in a network.
- **Antimalware**: A new age replacement of antivirus that can detect and act against a variety of malware (malicious software), including a computer virus. Malicious software is often available for free download for uses like playing media files and creating slideshows. Often, behind these seemingly harmless functions, such software affects the host machine and causes disruption. Antimalware and antivirus tools do have the ability to take some evasive action, such as quarantining the infected file or disconnecting the infected machine from the network.
- **Endpoint detection and response** (**EDR**): It is deployed on endpoints to detect patterns of malicious activity and immediately execute a corrective/preventive action. For instance, an EDR on a host computer

can notice the unusual behavior of any of its files attempting to change the computer registry (in a Windows machine) or attempting to create a root user on a Linux OS machine. It can then prevent such action from being executed by the operating system of the machine. It can also stop any newly downloaded and unauthorized tool from being executed. EDRs analyze patterns of malicious behavior and thus are more likely to protect the hosts. Thus, they are fast becoming the standard in large organizations to protect computers, including servers. EDRs often take in intelligence feeds about growing attacks on the Internet and accordingly protect the host.

Tip: Network detection and response (NDR) is also an evolving technology that can determine patterns of bad/malicious network traffic and take corrective/preventive action at the OSI Layer 3/2.

- **Backup:** They are one of the most important technologies that should be the focus of the organization to ensure it can bounce back from disruptions. Proper strategy and implementation of a backup plan will help the organization ensure better availability and also help in responding to adverse environmental, technological, and/or user actions. Backup can also be encrypted using the same algorithms to protect the confidentiality and integrity of the data on the media. Popular backup media includes magnetic tapes, storage systems available to a LAN called **Network Attached Storage** (**NAS**), or a large network of storage used by servers to backup large volumes of data using a technology called **Storage Attached Networks** (**SAN**). Periodicity of backup, such as incremental, daily, weekly, and monthly, applies risk principles and available use of resources as optimally as feasible.

Security controls are a bouquet of a variety of tools, technologies, processes, and people that need to work together to maintain the CIA objectives of the organization. This is a concept called **defense in depth** (**DiD**). In fact, several layers of security controls work in tandem to protect against a variety of risks. We explored risk management in *Chapter 2, About Managing Risks*, where compensating controls play a role in adding to the layer of defense. While the term DiD seems to indicate that they are focused on protection, it may be noted that the controls cover detective, preventive, and corrective controls as well. *Figure 6.12* shows a representation of some of the controls in a DiD approach. In the later chapters, we will cover some more topics, like vulnerability management and incident response, that will add to this DiD philosophy.

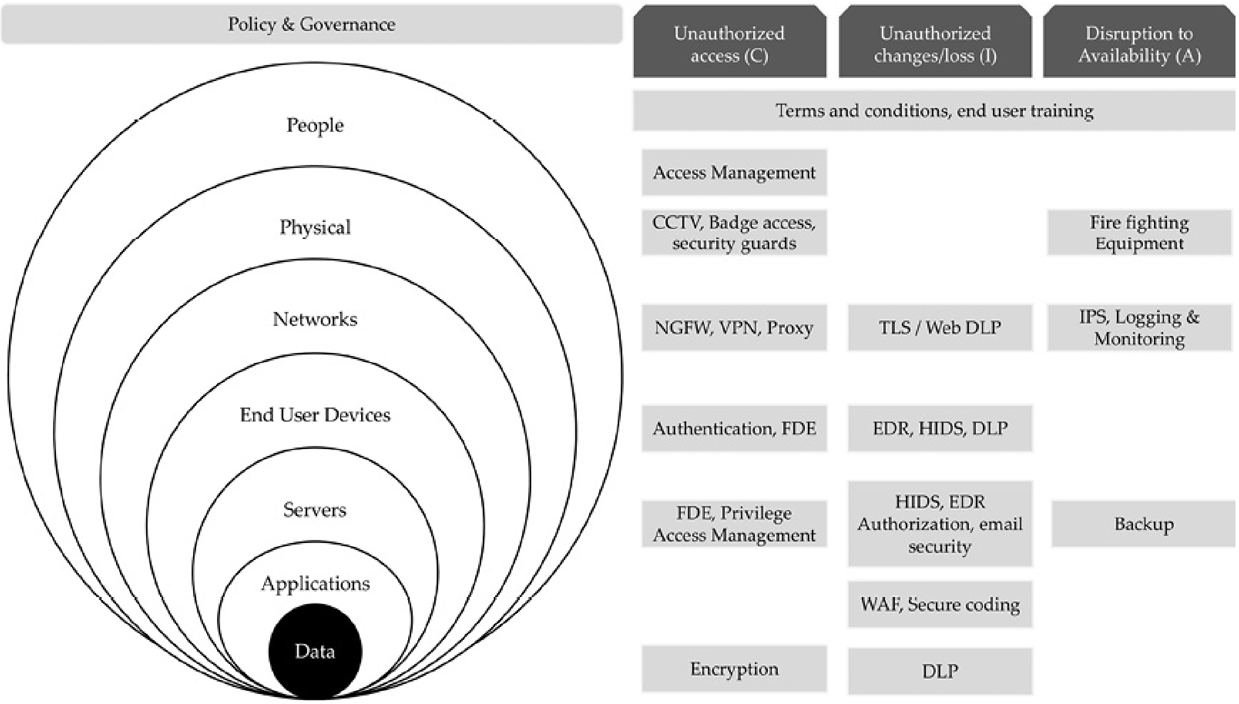


***Figure 6.12**: Defense in depth*

Earlier in the chapter, we covered IoT and the proliferation of IP-based devices, including handheld. A similar approach to protecting such devices becomes key. The use of IP-based devices in manufacturing setups and inpatient care at hospitals is not new. Those devices are referred to as **operations technology** (**OT**). Many times, the network of IoT, OT, and IT systems is kept separate with no interconnection to each other; this is called an **air-gapped environment**. However, it is generally the same IT team that manages the setup. Over the coming decades, IT-OT convergence is likely to play an increasingly major role, where the CISO would have to reimagine the network of today.

Advancements in AI and quantum computing

Self-service chatbots and **interactive voice response** (**IVR**) systems have greatly eased the burden on consumer service agents while positively impacting the customer experience in most cases. Chatbots got a further boost with the adoption of AI-based content. **Generative pre-trained transformer** (**GPT**) and related technologies started to be widely adopted around Nov'22, when ChatGPT was launched to help users generate their own content, including software code. **Artificial intelligence (AI)** is no longer a buzzword. Within two years, there is barely any technology available that does not claim to use AI and GPT in its product. Even service companies are leveraging the power of AI using the data at their disposal. The power of AI in executing tasks is immense. As algorithms and **large language models** (**LLMs**) learn how to use data and make decisions, much of the repetitive, analytical, and content-generative work is likely to move to AI-based systems. In fact, it is expected that the businesses will adopt an ecosystem of agents, each of which is capable of independently executing a task. For instance, an agentic AI will be able to use your prompt (instruction/request) to not just plan out your trip's itinerary, book hotels and flights, organize your packing list, and, if needed, even order the travel accessory/apparel you need, based on the destination's weather. The use of AI does come with risks, especially when it comes to PII and PHI, and must be factored into the AI strategy the organization wishes to adopt:

- Maintaining the guardrails of data that the AI LLMs need to train on. For instance, using aggregate PHI might be ok (with consent), but allowing AI agents to determine patterns of health and disclose them to unauthorized people could be dangerous.
- Decision bias is pertinent, as AI systems learn from experience and feedback. The system can therefore be made to believe a bias is a correct decision and thus create inappropriate outcomes.
- The language models require exposing data, and organizations would need to implement their AI in a way that their data is not exposed to other organizations. The LLMs they use must remain within their control.
- **Adversarial attacks on AI models**: Cyber criminals can compromise the integrity of the models and create conditions for incorrect and biased judgments by AI. In a healthcare AI application, this could even mean a life-and-death impact.

Like any other technology, AI will have pros and cons. It is important to find the right balance between using risk management principles and using them responsibly to allow businesses to innovate and grow ethically.

Tip: In the early part of 2025, the popular AI models, such as Copilot from Microsoft, Gemini from Google, and Q from Amazon, were challenged by a Chinese version called Deepseek, which proved to be not just efficient but way less costly. Several commercial AI content providers switched to this AI model in February '25.

Exciting developments in the field of AI, called agentic AI, are now encouraging organizations to have autonomous agents that can independently make decisions and take actions with or without human intervention (human-in-the-loop). This approach enables extreme speed to the outcome for routine tasks. For instance, when prompted to create an itinerary for a vacation to a beach, an agentic AI solution can not just suggest the itinerary; it can automatically make airline and hotel reservations on your behalf and suggest your packing list based on the temperature in that period.

Quantum computing is likely to challenge the way computers work and, more importantly, the foundations of cryptography and encryption in a big way. The power of quantum computing, which uses qubits, is the ability for a qubit to have multiple states/values at the same time. The 256-bit encryption, which needed several years and astonishing computing power to be cracked, would easily be compromised by quantum computers. The researchers are working on a quantum-safe encryption, also called **post-quantum cryptography** (**PQC**), that is likely to be strong enough to ensure confidentiality even when a quantum computer tries to solve the algorithm.

Tip: NIST has a standard for PQC, and several efforts are underway to counter the challenge using newer tools and capabilities.

Conclusion

In this chapter, we learnt that technology is ever-changing, and so are the cyber risks. In learning the principles of network security and several technologies, we explored the concepts, strengthening the defense's in-depth philosophy and approach. Knowledge of these protocols and technologies will help in thinking about the security program with technical acumen.

In the next chapter, we will explore identity and its related technologies, risks, and handling.

Key takeaways

As we explored the origins of computers and those of cyber attacks, the key takeaways are:

- Innovation in security technologies is also keeping pace to protect against risks.
- Gaining foundational knowledge of reference frameworks such as **Open Systems Interconnection** (**OSI**) that help the security team understand networks, routing, and the flow of information.
- In the digital world, information flows through IPs, ports, and protocols, by using key services such as **Domain Name Service (DNS**), **Lightweight Directory Access Protocol** (**LDAP**), **Simple Mail Transport Protocol** (**SMTP**), **Transport Security Layer** (**TLS**), **Hyper Text Transfer Protocol** (**HTTP**) and **Extensible Markup Language** (**XML**) using web services and/or **application program interface** (**API**).
- Protecting information as it flows through, using concepts like encryption, supports the foundational CIA principles, but is also required by regulations and/or customer contracts.
- Principles such as *least privilege* in provisioning access, ensuring secure configuration baselines, such as strong authentication mechanisms, and using current standards of communication, such as TLS1.2, are important.
- The security incidents and breaches around the world indicate that several organizations may not have implemented baselines and best practices.
- Advancements in quantum computing have now necessitated a relook at encryption algorithms and preparation of a world that uses **post-quantum computing** (**PQC**) approaches. This applies not just to data at rest or in transit, but even to the security of the **virtual private network** (**VPN**) and digital certificates used for email security.

References

- https://www.iso.org/ics/35.100/x/
- https://www.geeksforgeeks.org/application-layer-in-osi-model/
- https://www.eso.org/~ndelmott/ascii.html
- https://www.codecademy.com/article/osi-model
- https://www.ietf.org/rfc/rfc793.txt
- https://www.cisco.com/c/en/us/support/docs/ip/routing-information-protocol-rip/13790-8.html

Join our Discord space

Join our Discord workspace for latest updates, offers, tech happenings around the world, new releases, and sessions with the authors:

https://discord.bpbonline.com

CHAPTER 7
Identity and Access Management

Introduction

In this chapter, we will explore digital identities that power the access to the information we collect, work on, and consume. We will explore the principles of need-to-know and need-to-have in depth, the common protocols in use, and understand the risks to identities and how to manage them.

Structure

The chapter covers the following topics:

- Authentication, authorization, and accounting
- Privilege access management
- Identity-based attack tactics

Objectives

By the end of this chapter, you will be able to understand the key role digital identities/user-ids play in the digital world. As a CISO, you will be required to define and implement organizational policies and procedures to manage users and the access they have in various scenarios. We will also be able to realize several of the attacks and ways to tackle them.

Authentication, authorization, and accounting

Access to an organization can be thought of in the following two facets:

- **Physical access**: This pertains to all in-person access, such as access to an office campus, an office building, specific rooms of the building, and so on. There are multiple physical access controls that are implemented, such as:

    - **Photo-id** based entry/exit.
    - **Radio frequency ID** (**RFID**) based **proximity card**.
    - **Biometric**-based access, where the face recognition/Iris scan or fingerprint access to specific areas is configured.

    Organizations may implement additional controls like manned security guards, vehicle body checks using cameras, motion detection camera-based alarms, hand-held metal detectors, and baggage scans based on their security objectives

- **Logical access**: This pertains to access to information systems using software rules and configurations. While some of the concepts covered in this chapter apply to physical access as well, for the remainder of this chapter, we will focus on logical access.

In *Chapter 1*, *The Triad of Security*, we explored concepts of authentication and authorization as a key means for meeting CIA objectives. Accounting implies monitoring and tracking users' key activities, such as login/logout, duration connected on a website, VPN, and their key decisions on the business applications, such as approving an invoice for payment. **Authentication, authorization, and accounting** (**AAA**) together form the core of what is called an **identity and access management** (**IAM**) system. *Figure 7.1* represents the three primary questions AAA focuses on:

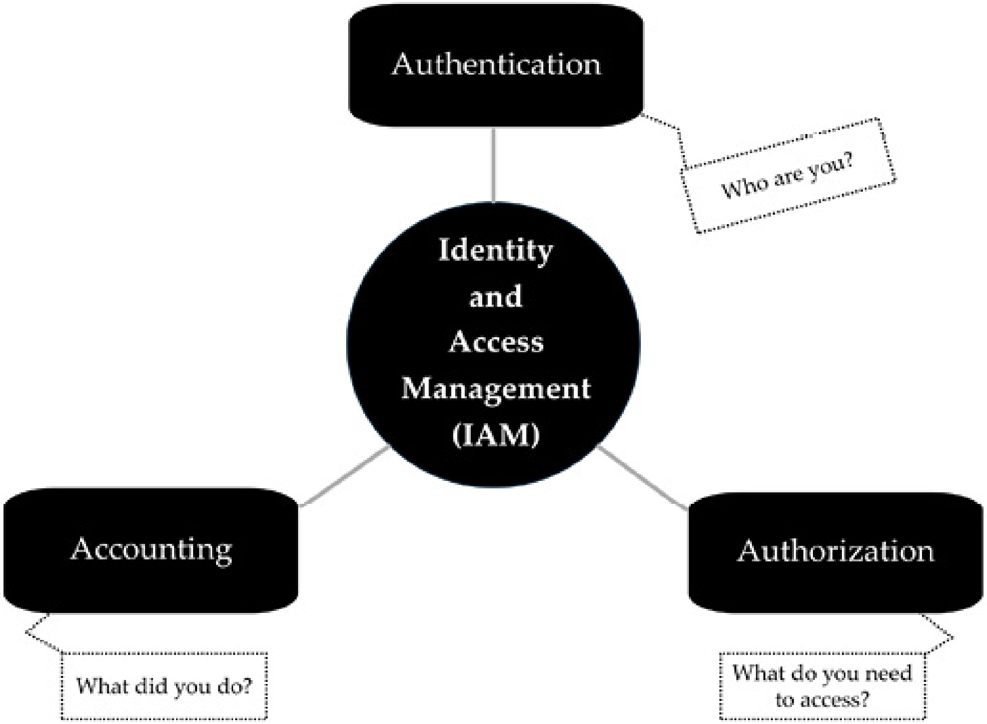


***Figure 7.1**: AAA and the key questions*

We will explore some of these concepts at length.

User ID management

Logical access to computers, networks, corporate applications such as emails, file storage, and business applications like insurance claims processing, inventory management of retail stores, employee-related services like leave management or pay-sip generation, and even risk management using **governance risk and compliance** (**GRC**) tools. **User identities** (**IDs**) can be created locally on each individual computer/information system, and hence called *local user IDs*, but it is impractical and does not meet the management requirements of centrally administered policies for user management. Several standards, covered in *Chapter 3, Role of Standards and Controls*, suggest organizations have a centralized mechanism to manage users. The standards also require that user IDs are not shared and are configured securely. In this chapter, we will focus on centrally managed user IDs.

The lifecycle of a user ID management is broadly considered in 3 phases for an employee or contractor, namely:

- Onboarding, where the user ID of the employee/contractor is created. This process is called **user ID provisioning** or just **provisioning**. Each user must have a unique ID.

- During, where access is granted based on the need-to-have principle upon request/review. The assigned access must be reviewed for validity at periodic intervals and at the change of role.
- Offboarding, where, after the employment/end-of-contract, the user ID is deprovisioned/made inactive, or sometimes deleted. This process is called **user ID deprovisioning** or just **deprovisioning**.

Tip: The terms user ID, ID, account, user account, username, and identity are used interchangeably in the context of IAM.

The lifecycle of the user management is shown in *Figure 7.2*:

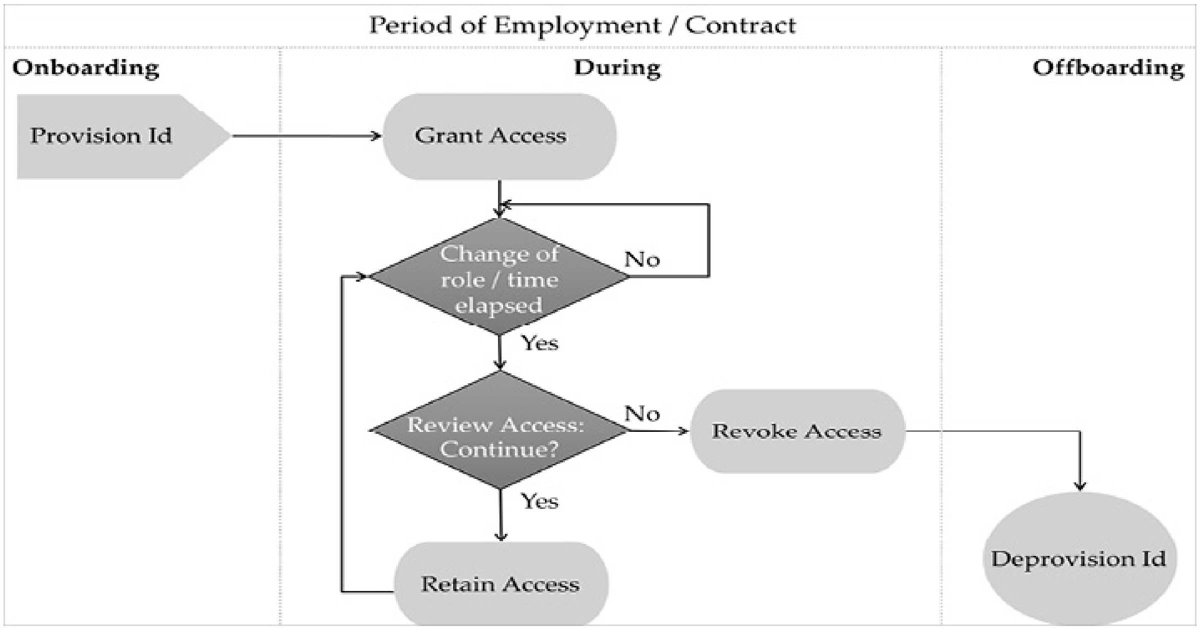


***Figure 7.2**: Lifecycle of user ID management*

Tip: During the phase of employment, additional rights may be assigned or modified. Those may be considered to follow the grant access process.

Tip: Like user-id management, physical access to a user also goes through the same lifecycle, but may use different software, generally called a building management system (BMS).

Apart from creating user IDs on domains, organizations may need to create and manage access to applications within the application or devices, such as firewalls. For instance, the organization may decide to maintain different user ids for the payroll application because even ex-employees may also need access to it for their full and final settlement or taxation-related documents. In such cases, it is essential for organizations to ensure there is an efficient and well-managed process to provision and de-provision access. User ID that remain enabled even after the person no longer serves the organization and/or where the original purpose is completed, are called **stale IDs**. These IDs can be misused by attackers or disgruntled internal personnel. The problem compounds if these stale IDs have privileges or access rights that can bring harm to the computer network or its resources.

Tip: In standards such as SSAE/ISO27001 and for SOX compliance, user ID management and physical access review are key controls, and the need-to-have principle is often a check that organizations fall short on. It is essential for organizations to ensure relevant access is configured only to an authorized and active username.

User ID types are explained in *Table 7.1*:

Types of user accounts/IDs	Purpose	Common pitfalls to watch out for
Normal users	For authentication and authorization requests by all employees and contractors individually.	Some organizations allow contractor IDs to be shared amongst individuals for convenience. This risks non-repudiation and even malicious use.
Administrators or privilege accounts	IDs created/provisioned for performing administrative tasks on computers, servers and other domain resources. Generally, there may be a default administrator id on devices. Such as root on Linux/Unix systems and administrator on Windows OS systems, or on home Wi-Fi routers. The *su* id is a powerful super user with complete administrative rights on Linux/Unix systems.	Organizations do not rename them or fail to ensure stronger usage restrictions leading to misuse.
System accounts	Identities created and managed by the **operating system** (**OS**) itself to perform a specific task during its own functioning. These are also administrative accounts. For instance, the *Apache Webserver* on a Linux server runs as a built-in system account. Similarly on Windows computers, there is a built-in SYSTEM account. These accounts can only be used by the system and are not available for a user for logging-in.	Other than the OS, no one else should be able to use it to login.

Types of user accounts/IDs	Purpose	Common pitfalls to watch out for
Service accounts	These are accounts created/provisioned on domains for an application or an external webservice to access resources on the servers without needing human intervention. For instance, an application will authenticate with the DC using a service account or launch a reporting service on another server at a scheduled time.	Service accounts are often a target of the cyber criminals and must be governed properly for usage and by implementing stronger password requirements for them.
Domain accounts	These accounts are created with privileges on the entire domain to perform any administrative function such as run scripts or program to determine system weaknesses called **vulnerability management** (**VM**) program. We will cover VM later in the book.	These accounts have far reaching permissions on the entire network and their use must be restricted and governed properly.
Guest users	User ID called **Guest** were often built in computers that were meant to be used by guests such that they can use the computer minimally, much like a guest at home may be permitted to use the couch.	Keeping guests users enabled on computers are susceptible to attacks by cyber attackers.
Break glass account	This is a special administrative account that is meant to be used only in emergencies when the primary administrator(s) are inaccessible, or the domain system is not responding to any user or application.	This is an account to be used at last resort, the ID and the password must be kept protected and not used. The purpose of using a *break glass* account is to quickly restore the *availability.*

Table 7.1: *Type of user accounts*

Tip: In an Active Directory (AD), groups of the account types may be created or built-in ones used to centrally manage permissions to the group, thereby each member would get the permissions.

With the increasing adoption of **robotic process automation (RPA)**, or bots for short, a lot of human step-based IT activity is automated. These bots typically need administrative permission. AI is further augmenting automation with machine learning and the deployment of agents that serve specific purposes. For instance, a customer service chat agent today can quite easily and accurately answer standard user questions. Organizations have started to focus on creating a family of such made-for-purpose AI agents that together can handle several user decisions. This set of agents that use AI to make decisions on behalf of humans is called **agentic AI**. These agents will also need system/application privileges based on the use case.

Some examples of how an organization's naming convention for a username are:

- Using the user's first name and last name or any part of it, for example, if the username is Uma Iyer, the username may be UmaI, or Uma_Iyer, or Uma10. The number here is the sequence number of users named Uma in the organization; in our example, it is 10.
- User's email such as **Uma.iyer10@company.com**.
- User's HR record number such as 1023409, where the number corresponds to the employment number on the HR file.

Tip: Generally, contractor IDs, service accounts, and privileged IDs follow a distinct naming convention to make it easy for user governance.

Domains

In an organization, various IT devices, such as servers, user computers, network printers, etc., are required to be managed centrally. Such a grouping is called a **Domain**. An organization's domain will have:

- A domain name. Generally, the internal network domain name will be part of the organization's website domain and hence a subdomain. For instance, Microsoft may run its internal network domain as corp.microsoft.com, while its **top-level domain** (**TLD**) website is Microsoft.com. The TLD domain name must be available in the public DNS records. We covered TLD/DNS in *Chapter 6, Key Security Technologies.*
- At least an administrator is needed to manage the server, called the **domain controller** (**DC**).
- At least one server that serves as the **primary domain controller** (**PDC**). Depending upon the network architecture, the spread of an organization's offices or network segmentation, **backup DCs** may need to be additionally configured to synchronize with the PDC. The DC is used for:
  - Set and assign the organization's IT policies and controls. We explored some of the control and control configuration elements in *Chapter 3, Role of Standards and Controls*. For instance, we can use the DC to disable the ability to copy/transfer data to a USB storage device.
  - Creating and deleting user IDs (**provisioning** and **deprovisioning**) and managing permissions assigned. Organizations may use applications called **identity governance and administration** (**IGA**) to manage permissions for network, business applications, and other resources that

users need.

- Manage the directory of devices and users (directory services). A directory is a grouping of similar attributes. The repository of the devices and users is called **Active Directory** (**AD**) on Windows OS and **Red Hat Directory Services** on Red Hat OS.
- Supporting **Local Directory Access Protocol** (**LDAP**) to handle requests from domain computers and users for the AD, including creating, accessing details about members of the directory, and deleting directory entries.
- Authenticate and validate authorization requests.

Tip: On the DC, the Local Directory Access Protocol (LDAP) runs on TCP port 389, and the secure version of LDAP over SSL (SLDAP) uses TCP port 636.

Note: The acronym DC is also used for data centre; however, in the context of this book, we shall imply domain controllers.

Authentication

In *Chapter 1, The Triad of Security*, we defined **authentication** as a process of establishing that the user trying to access data/information is the authentic user (from an identity perspective). It is designed to prove that s/he is who s/he claims to be, i.e., *who you are.* The user provides the key pair of username/password to the login screen of the network/application to complete authentication in the simplest of forms.

Firewalls and network security devices often use an authentication protocol called **Remote Authentication Dial-In User Service** (**RADIUS**), which encrypts the password in transmission between the user's machine and the network device.

The most common and one of the oldest authentication protocols used in Windows-based environments was **New Technology Lan Manager** (**NTLM**), which used a challenge-response method to authenticate a user. It is an insecure form of authentication that is no longer widely used.

In the modern day, a network authentication protocol called **Kerberos** is used. The protocol is supported by Windows and Linux/Unix operating system environments, and several of the network devices, too. Kerberos is preferred over LDAP for its security. This protocol handles authentication in the following way:

1. The requester’s username and the password hash are validated against the hash stored in the DC using its service called **Key Distribution Centre** (**KDC**).
2. The KDC sends back a **Ticket Granting Ticket** (**TGT**) that gets stored on the requesting device.
3. A network device/user requesting a service from the domain (requester) sends its service request with TGT to the DC.
4. The DC validates the TGT and the permissions the identity has and accordingly sends back a **service ticket**.
5. The requester then sends this validated service request to the domain resources to get the service it needs.
6. The domain resource internally validates whether the service ticket is valid and is permitted to be serviced.

A pictorial view of the Kerberos-based authentication is shown in *Figure 7.3*:

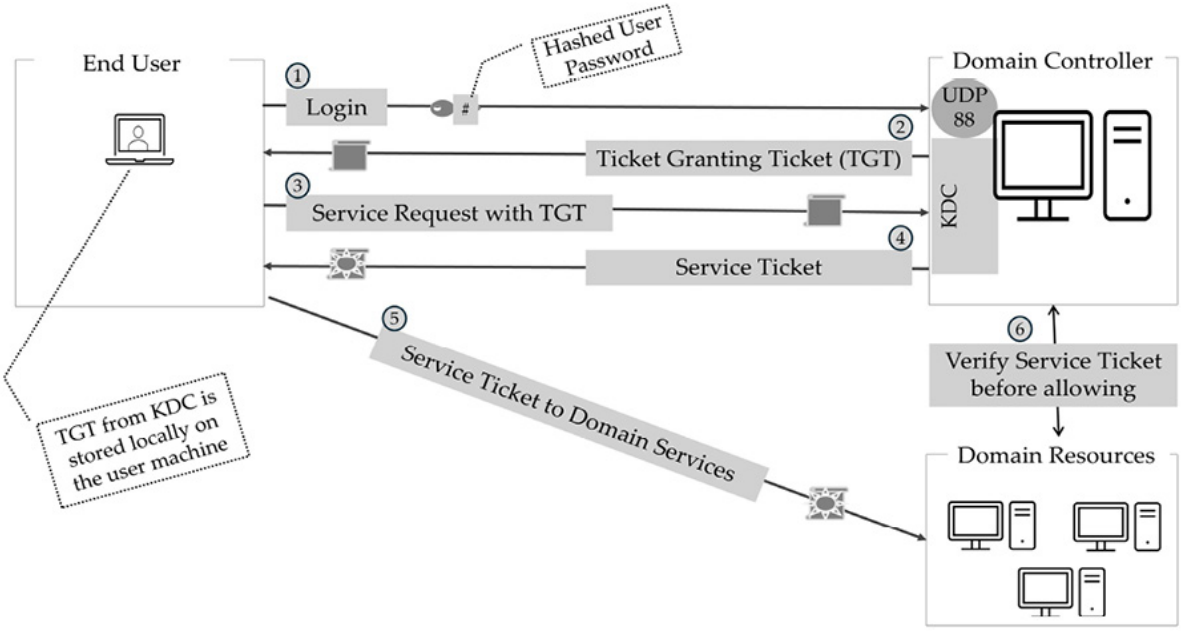

End User
Hashed User Password
Domain Controller
1
Login
#
UDP 88
2
Ticket Granting Ticket (TGT)
3
Service Request with TGT
KDC
4
Service Ticket
5
Service Ticket to Domain Services
6
Verify Service Ticket before allowing
TGT from KDC is stored locally on the user machine
Domain Resources

***Figure 7.3**: Kerberos authentication*

When a human ID provides the combination of username and password for login, it is called an **interactive login** because there is a human-to-machine interaction to prove *who you are*. These logins are also called **type-3** logins. Instead of passwords, it is possible to log in interactively using a PIN, a biometric feature such as a fingerprint or facial recognition.

Tip: Interactive login for system and service accounts should ideally be disabled because these identities have excessive permissions on the domain and can be misused by attackers.

Organizations must choose the right level of authentication to implement based on principles of risk management. Some of the criteria to consider are user experience, technical feasibility, and regulatory requirements. For instance, banks in India require credit card transactions beyond a certain value to be authorized by the cardholder using a bank-supplied **one-time password** (**OTP**). The authentication can be based on one or more factors, namely:

- **something that-you-know**: Such as a password/or **personal identification number** (**PIN**) or an OTP.
- **something that-you have**: Such as a hardware/software token that prompts a number to enter.
- **something that-you-are:** Such as facial recognition, iris scan, or fingerprints (using such human features is also called **biometric authentication**).

Using any of the factors above, additional authentication, called **adaptive authentication**, may be triggered based on some conditions being met, such as:

- **Time-based,** such as login time to the network. For instance, if a user logs in beyond his/her regular working hours, s/he shall need to answer security questions with pre-registered answers.
- **Location-based factors**, for instance, when the user is on the corporate network within the office premises, a single authentication factor is applied, but the same user may need to use the OTP sent to the phone for an additional step of authentication. Such a method protects from attacks called **account takeover** (**ATO**). We shall cover these attack types later in the book.

Tip: The pair of usernames and passwords is also called user credentials.

Types of authentications for a pre-registered user are:

- **Single-factor authentication (1FA)**: Where a user must use a password or a PIN to authenticate. This concept uses one secret factor, i.e., *something that-you-know*. The assumption is that the password/pin is not shared and only one user knows it. This method is commonly used to log in to websites, user machines, and network devices, and is risky for important processes like contract signing or accounts payable/remittance to suppliers. In some cases, a pre-registered response to questions might also be needed to be provided (challenge response) to authenticate. This would still be *something that-you-know* and hence a single-factor authentication.

- **Multi-factor authentication (MFA)**: Where a user needs to provide two or more factors. While this method may make the user experience not so desirable, it is certainly relevant for protecting identity. For instance, modern user applications such as e-commerce websites, webmail, and professional social media enhance users' identity security by requiring a password/pin and an OTP/verification code sent to an alternate email or phone number.
    - **Two-factor authentication (2FA)**: Where a user must use any two of *something that-you have*, *something that-you-know* and to authenticate. For instance, we use the bank's ATM card and its PIN for banking transactions. The banks require that ATM cards be used only by the person to whom it is issued. Organizations may use hardware tokens (covered in *Chapter 1, The Triad of Security*) or application-based methods for the second factor. This method is called **token-based authentication**.
    - **Certificate-based authentication**: Where one of the factors is a PKI certificate installed by the organization on the device on the network. Devices that do not have such a certificate will not be able to connect to the organization's network. It is widely used in a concept called **device trust**. The user may be required to authenticate to some network resources based on adaptive authentication rules. Increasingly, PKI-based authentication, coupled with biometric authentication, is being implemented to make user login **passwordless**.

With increasing digitalization, almost every application and network requires some user credentials. This is onerous on the user as s/he must remember several usernames and password combinations, each with different names. The following concepts have emerged over time and are in use:

- Organizations may establish a relationship called **domain trust**, allowing the full or part of Active Directory resource information to be shared, and the user does not need to authenticate to the second domain. This domain trust is commonly used when organizations are in the process of **mergers and acquisitions** (**M&A**). The domain trust is configured on the DCs.
- Two organizations may choose not to establish this formal AD-level trust and yet allow each other's authenticated domain user to be permitted

access to the resources on their domain. It may also be done to leverage user management of larger web platforms. This form of trust between is called **trust federation**, and the user-ids are called **federated ids**. For instance, several e-commerce websites or social media websites allow the use of the login of services like *Gmail* and *Outlook*. You would not need to create new user credentials on these websites and instead use your *Google* or *Microsoft webmail ID* itself.

Single sign-on

Single sign-on (**SSO**) allows all applications, network services, and websites to be configured to leverage an authentication service that validates user identity and its authentication from a service/server called an **identity provider** (**IdP**). Active Directory is an IdP. The access authorization information may be maintained centrally or may continue to be maintained at the application or network resources. For instance, organizations may allow their payroll applications to divert the user to their corporate **directory services** (**DS**) and authenticate to the domain instead of maintaining the user credentials at the payroll application itself. Similarly, once a user has successfully logged in to its network, it will allow access to some birthright applications like leave management, corporate emailing applications, or file servers.

SSO uses an **Extensible Markup Language** (**XML**) called **Security Assertion Markup Language** (**SAML**) to exchange authentication and authorization data using TCP port 80 or 443. The service provider receives the SAML assertion that the user has been authenticated and that the request is valid.

SSO services are increasingly available as an external web service and are leveraged by organizations. We will explore some details of the SSO service using the **cloud** later in the book.

SSO might leverage an open standard-based protocol called **OpenID Connect** (**OIDC**) that enables websites to redirect authentication requests to IdP using a secure method called **JSON Web Token (JWT)**. **JavaScript Object Notation** (**JSON**) is a commonly used text-based exchange format used between clients and servers. JSON is easily human-readable and is independent of programming languages.

Password management

Organizations and publicly available websites require the password to meet or exceed certain requirements. These requirements are based on several trends of common hacking activity and best practices that standards such as ISO 27002 have defined. Some of the commonly implemented requirements are:

- **Password complexity**: The composition of the password must meet at least three of the following:
  - **Password length**: Should be 12 characters or more.
  - At least one uppercase and a lowercase.
  - At least one special character out of !@#$%&.
  - No repeating numbers such as 111.
  - Avoids common and cracked passwords such as *password* or *P@ssw0rd.*
- **Password age**: The maximum days a password is valid. It may also signify the days since the last change to the password was made. Organizations typically require their employees and contractors to change their passwords every 90 days. Once the password's maximum age is reached, a user will not be able to log in without changing the password. The password would need to be **reset**.
- **Password reuse**: The number of times a new password can be used before the same one can be used. Organizations typically require users not to set the last 24 passwords.
- Passwords must neither be shared nor written down.
- ID/Account must be temporarily disabled after a certain number of consecutive unsuccessful attempts. This configuration is called **account lockout**. Usually, organizations set this number at 5 for remote connections like VPN and 10 for internal applications.

As of early 2025, NIST and other bodies have started strongly making some of the following recommendations to organizations to:

- Focus on periodic password reset, say every 90 or 180 days, and only do so sooner if there is evidence of compromise.
- Using passphrases, i.e., using a phrase that a user likes/dislikes to create a string of alphanumeric letters as the password. It is easier to remember such a passphrase instead of so many passwords.
- Determine simpler and secure ways for using tools called **password managers** that can generate and store strong random

passwords/passphrases for applications.

- Use encoding methods like Unicode as well as ASCII.

NIST has based such recommendations on extensive research and on patterns of past password-based attacks. While such recommendations are not yet widely adopted, some organizations have relaxed password age requirements and have allowed password managers.

Fast Identity Online (**FIDO**) is an open and more secure authentication standard based on PKI that aims to:

- Augment MFA with a secure fingerprint-based hardware token.
- Replace the need for users to enter passwords.

FIDO is widely being implemented across organizations to negate password-based threats to organizations' data. This protocol also vastly improves the user experience and helps counter some of the phishing-based attacks. We will cover cyber-attacks later in the book.

The latest version of the standard is FIDO2 and leverages *something-you-are* and/or *something-you-have*. FIDO implementation uses aspects such as:

- **Passwordless authentication**: Most personal computers and handheld devices have facial recognition or fingerprint to access the device; the same concept is used for an organization's machine, where the device itself may also use the organization's digital device certificate. The user will just need to use his/her fingerprint to log in to the network.
- **MFA with FIDO2-compatible hardware tokens** that can read fingerprints is also used. Without the token inserted on the local computer, the website will not be accessible. For instance, in a FIDO2 compliant implementation, the hedge fund manager of a financial organization would need to have a hardware token physically inserted in his/her computer to authorize the token and then authenticate with his/her fingerprint. Hence, it reduces the chance of misuse, fraud, and other identity-based attacks.

The administrator can apply risk-based decisions to apply the FIDO-based authentication only for certain websites. The modern-day **IdP/SSO** platforms support multiple FIDO options.

Authorization

In *Chapter 1, The Triad of Security,* we learnt that authorization implies the rights/permissions the users must have on the data to appropriately execute their job responsibilities. This is the *need to have* an aspect. The authorization process would ideally require a formal request from the user with a clear justification. Another approved user (read admin or information owner) may need to take a call to reject or approve the request and then take the required action to provision the access if needed. You may think of authorization as something you have authority over, to see/read, or even to create or modify. Typically, organizations will have different authorization levels to ensure the information systems/data are duly protected from any unauthorized modification. Some of them are:

- **Read-only**: You can read most of the information, but cannot change anything.
- **Read-write**: You can read, create, modify, and/or delete any data (or configuration).

In the user management lifecycle explained in *Figure 7.1*, the process of granting, reviewing, and revoking access is part of authorization. The types of users are grouped together for easier administration. For instance, an organization will want all its normal users to have the same password policy. The domain administrator will maintain a DC group, also called a **security group**. The collection of permissions and configurations that are applied to these security groups on the DC is called **Group Policy Objects** (**GPOs**). For instance, the organization may use GPO to allow or disallow users to edit the Windows registry, USB storage, and define password complexity rules. Earlier in the chapter, we learnt the types of users. On the DC, security groups for each of those applicable user types may be created for easier administration.

Tip: Every employee and contractor will need some default permissions, such as the ability to log in to a network, the ability to send and receive emails, and access to the attendance marking system and e-learning platforms. Such rights are called birthrights and should be configured at the time of onboarding without needing a specific request.

Tip: Access to common areas of the building, like the reception area, cafeteria, mailroom, etc., is usually birthrights for physical access setup.

Some of the key strategies for implementing *need-to-have/need-to-know* in defining and specifying access controls:

- **Role-based access control (RBAC)**: Administrators of the IAM assign permissions to users **based on the role** of the user/group; every member of the group gets the same permission. Other aspects are:
    - Aids in security by design principle by ensuring only relevant people

  - inherit/get required access.
  - Aids in user management (provision and deprovisioning). Users can simply be added or removed from such role-based groups.
  - Changes to permissions can be made at the group level instead of user by user.
  - This approach is an impediment when exceptions have to be made, i.e., a set of permissions is preferred not to be extended to a partial set of users. For instance, the organization has a read-only view role for all security devices that it assigns to relevant internal team members, but it does not want the firewall rules to be available for members of the *internal audit*. In this scenario, unwanted permissions will get assigned to all members, or the organization will have to consider more granular role-based groups, making it complex to administer.

- **Mandatory access control (MAC)**: MAC is based on allowing/disallowing based on a user's level of authority or clearance. The level may not be hierarchical.

  - It can be considered a stronger implementation of need-to-have/need-to-know.
  - Generally used in government organizations or in research organizations where some information is required to be guarded with utmost care.
  - Users are assigned security clearances and labels/attributes that are used in defining access security policies. For instance, in a clinical research organization, only a qualified user with top-secret access is allowed to view or modify the formulations of the medicine. Similarly, only very specific personnel with government clearance may be allowed to access the security detail on the movement of heads of state, such as the prime minister or president.
  - This approach is not convenient for business organizations in most scenarios, as the benefits are fewer than the complexity of implementation.

- **Discretionary access control (DAC)**: Where the owner of the information asset, such as the application server or the network administrator, uses his/her discretion/judgement:

  - To allow or disallow access requests from users.

    - To decide on the access rights to provision.

  The asset owner, in turn, may use RBAC within the information system to make administration easier.

  Tip: It is important for the requester to be clear and specific about the access being reques with clear and valid justification. This not only aids the information owner but also serves record of audit/access assessments.

- **Attribute-based access control (ABAC)**: Where the access permissions are based on the attributes of the data. In an ABAC, rules and policies are defined and implemented to determine the access permission by using the attributes of a user, system/device, and/or environmental factors. Generally, ABAC is used in conjunction with other access models. For instance, the HR recruitment personnel will have access to PII but may not have access to compensation-related fields. Similarly, a department's finance executive may not have access to see the budget amounts of another department, though they are on the same application and the same database.

  ABAC can also be conditional, i.e., access may be allowed based on certain conditions being met. For instance, an e-commerce portal may allow its administrator to update the unit price of high-value goods only from the corporate network. An administrator working remotely will not be able to make such modifications. Similarly, the building access to the data center of an organization may only be permitted during certain hours of the month. Such measures are used for additional protection of the CIA.

Some of the protocols used in authorization are:

- **Terminal Access Controller Access-Control System+** (**TACACS+**) for securely authenticating a user of a network device, such as a router, to a TACACs server and to know the authorization for an ID. In an organization with several networking equipment such as routers and switches, a TACACs server is configured with user management also done on it. Some network devices do support the **Active Directory** (**AD**).
- **Open authorization** (**OAuth**) is an authorization standard that uses JWTs to exchange authorization securely without sharing the user credentials. This is also called **delegated authorization**. For instance, a user can permit a third-party app to post their preferred updates to another website. For instance, several fitness applications can publish the latest activity, such as running, to social media. In this case, the user has to authorize the

fitness app to use their identity using OAuth, but without actually sharing the social media app’s user credentials.

Periodic **access reviews**, also called **entitlement reviews**, are a key detective security control that can help reduce the possibility of an attack. The focus on access review, as also shown in *Figure 7.2,* is to ensure only the appropriate user IDs have the necessary access. All unwanted access must be revoked.

Accounting

To establish who did what and when, the activity logs are generated and stored on servers. These logs establish the trail of events in the chronology of events and serve as **detective control**, which we covered in *Chapter 3, Role of Standards and Controls*. Such logs are extremely useful in the following scenarios:

- Determine ownership of activity, i.e., who did what.
- Probable cause of failure, such as an application not being available.
- Trail of activities that may have resulted in a positive or negative outcome.
- What privileges were used by the administrator, and when.

There are IT operational logs, such as the memory utilization of the server every hour, the throughput delivered by the router on any egress line, or the uptime of the website. However, for the purpose of security, such logs may be of secondary interest because they may impact the availability of information assets, but not due to a security attack. The operational logs will be of prime interest in a security investigation if a cyber-attack may have caused the deviation from normal. For the remainder of the book, we will consider logs as security-related logs.

The organization will use several criteria, such as its risk appetite and organizational security objectives, including regulatory or customer requirements, in its log management strategy. Some of the other aspects are:

- What activity to log, i.e., scenarios or activities, it must capture.
- How to secure the log, i.e., to send the logs directly from the source server/device to the log storage server without it being tampered with.
- The period of log retention is, such as one year.
- The storage medium for these logs is low-cost, high-capacity storage servers and/or the local device itself.
- The process and the analysis to derive from those logs.

- The organization may need to use a risk-based approach to choose the scenarios or business rules it wishes to log.

The following are some examples of security logs (also called **audit logs**) that must be enabled:

- **Login-related**: Logs of successful and unsuccessful logins to any application, VPNs, Domain, etc. We may use a risk-based approach to not log in to user activity for some internal devices, like Wi-Fi access points, when connecting from a corporate device.
- **Key activity related**: Logs of action on business apps. For instance:
    - Filing of an invoice, its approval, and its remittance.
    - The manager of an e-commerce portal is updating the unit price of a product.
    - The HR manager updates the employee information, such as promotion or compensation.
    - The employee is updating their own bank account details on the HR portal.
    - Details of which administrator provisioned what access to whom.

The logs should capture information such as:

- User ID that made the change.
- What change was made.
- The time when the change was made is captured in a standard timezone such as **Universal Coordinated Time** (**UTC**).

Tip: One of the most common pitfalls in IT and digital environments is their servers not being in sync with standard or consistent time servers running Network Time Protocol (NTP).

Most security standards and regulations require logs of key events to be generated and kept tamper-free and safe. Logs that are stored such that they cannot be tampered with are called **immutable logs**.

They may be sent to a specialized server for log management. We will explore logs and their role in incidents later in the book.

Privilege access management

Access is considered **privileged** when the identity (Id) has the permissions to make modifications to important security and/or business configurations. Administrator/privileged access is the ability to perform certain task(s) that a normal user would not have and impacts the CIA of the information asset(s) in a significant way.

Some of the examples of privileged access in an organization are:

- Ability of the firewall technologist to add/remove rules of traffic from and to the organization's network.
- Ability of the website administrator to change the content of the organization's webpage(s).
- The ability of the network administrator to change the Wi-Fi password and/or change the allowed websites on the web proxy.
- Ability of the domain administrator to set GPOs.
- Service accounts to launch a specific scheduled task, such as backup, on every server at a certain time.
- The ability of a local administrator of the computer:
    - To install and remove software.
    - Modify configurations, such as turning off anti-malware or permitting USB storage functionality.
- Ability of an application and/or database administrator to enable/change the features or values. For instance, the e-commerce platform administrator's ability to change the unit price of commodities.
- Ability of the financial controller to set and modify budgets of the department in the finance tool.
- The ability of the HR manager to change the compensation structure or export all of the HR records.

In an organization, several activities may require read-write permissions and approvals, such as an employee requesting leave and the supervisor reviewing and approving it. This access would not be considered privileged. Similarly, approving an invoice and making payments has financial ramifications to the organization, but having to instill strong governance for all such aspects is neither practical nor warranted.

Good governance of the privileges assigned and used by the user is key to a secure environment. Attackers have often used such privileges under types of user-id cover to cause breaches in organizations. PAM governance may include:

- Prior to granting privileges and at a periodicity such as per annum, the human user ID with privileges must undergo some security awareness training on the importance of the privilege and the consequences of non-adherence to secure practices of the assigned role.
- **Background checks** (**BGC**) are required prior to being assigned privileges.
- Any additional MFA/FIDO requirements for authentication/authorization to enforce.
- The information asset must apply or define proper discretion before granting access, and the granularity of such permissions.
- Review of roles and their current access.
- The password complexity of the privileged Id should ideally be stronger.
- The passwords of service accounts/system accounts must not be hardcoded in the application.
- Organizations may choose to provision and assign such administrator privileges to a secondary user ID so that the normal user ID continues to have only regular privileges.
- The privileged ID and the privilege itself are used only by their intended user.
- Each privileged activity is logged and reviewed. For instance, when access to the domain controller is required by a domain administrator, a service account is used only for its known purpose.

Note: PAM is also called privileged identity management (PIM).

Tip: In the DAC model, an administrator makes the decisions for permission to the informa assets.

Segregation of duties

User permissions are defined and implemented in such a way that no single ID can cause a disruption. For instance, the person who checks and authorizes a supplier invoice for payment must be different from the person who approves the actual remittance. **Segregation of duties** (**SoD**) is very commonly defined for several such business processes and is an important consideration for **Sarbanes-Oxley** (**SOX**) related audits (covered in *Chapter 3, Role of*

Standards and Controls) and for overall IT controls. For instance, the e-commerce portal website administrator, ideally, must not be the same person who decides on inventory to stack, unit price, etc. The core reason for the concept of SoD is to have a **maker-checker** approach. There will be a maker (of the request) and an independent checker/approver (of the request). This concept is borrowed from banking operations, where financial instruments like cheques go through such validation.

The RBAC access model is often used to achieve the business objective and the security objective, where the information security policy manager would not have access to manage firewall rules.

Tip: Organizations may choose to have all security-related administration rights with IT teams or even specific individuals. In such a scenario, it would need to take a risk-based decision and implement stringent independent monitoring to protect itself.

Note: Segregation of duties is also called SoD.

Integration with other systems

Identity serves as the fulcrum of digital environments and is the basis of how information access is provisioned and delivered. Service and system IDs are also used for information exchange within a business environment. *Figure 7.4* represents how several systems interact with each other using identity:

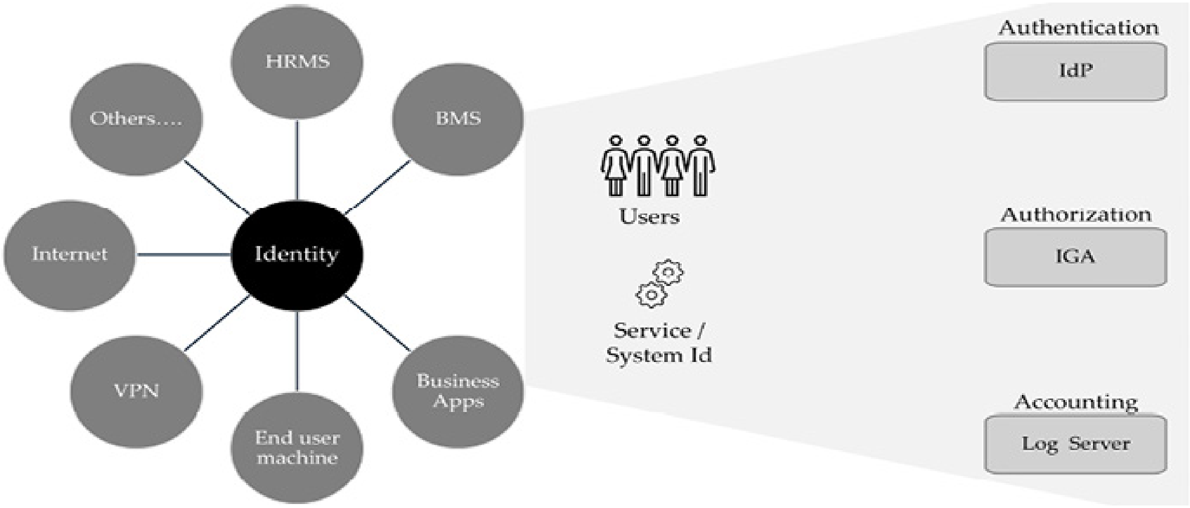


***Figure 7.4**: Integration with other apps using identity*

Identity-based attack tactics

Access to information is based on identities and, therefore, a key aspect for attackers to use. In the **defense-in-depth,** covered in *Chapter 6, Key Security Technologies*, humans are likely to be the weakest link. For instance, attackers compromise weak passwords to gain access or steal identities. *Table 7.2* explains some of the common credential-stealing attacks focused on compromising identities:

Attack tactic	Explanation	Prevention controls to mitigate the attack
Password guessing	Attacker uses an iterative mechanism to guess the password of a user(s).	Disallow passwords that are easy, and/or have common dictionary words. For instance, passwords must not be the word password.
Password cracking	Attackers use an iterative mechanism to generate hashes of passwords in their list and comparing it against the hash of the password obtained from AD or network until a match is found.	itself or common dictionary words. Set account lockout for consecutive unsuccessful login attempts. Use Passphrase where possible.
Password spraying	Attacker uses some specific and generally used password(s) against several ids with the hope atleast one of the account may have the same password.	Harden system for default accounts like Guests/Administrator.
Brute force	Attackers try and break the password by continually and systematically guess/crack the password.	Location based limitations for login, for instance disallow login attempts from geo locations that organization has no business interests in. Disable/delete ids that are no longer required, especially when they have administrator privileges. Use MFA where possible.
Pass the hash (PtH)	Attackers obtain the hash of the password and use it to send its request to the application/network resource. The attacker does not get to know the password itself but still gets the desired outcome and breach confidentiality or integrity of information.	Where possible restrict use of same passwords for multiple accounts (**credential overlap**) and impart user training. Harden applications and OS for not accepting unvalidated hashes. Strong user ID governance, especially PAM IDs.

Attack tactic	Explanation	Prevention controls to mitigate the attack
Man-in-the-middle (**MITM**)	Attackers use the network to sniff and determine key information aspects such as passwords and then subsequently use that for compromising confidentiality and/or integrity. For instance, in an MITM attack, the attacker can change the unit price of an e-commerce portal thereby causing financial loss to the organization. Similarly, using network protocols like **Address Resolution Protocols** (**ARP**) or DNS, covered in *Chapter 6, Key Security Technologies*, the attacker can force future traffic to flow through IT infrastructure controlled by them MITM is a type of **session hijack** attack.	Use encryption in transmission, such as SSL connection. Deploy strong **access control lists** (**ACLs**).
IP / Domain spoofing	The attacker tricks the user to authenticate to a page that looks like the application/website they use. This spoofing may be also IP based. Attackers essentially masquerade their IT infrastructure as the original one. Such spoofing is commonly used in banking websites, with the link generally sent through emails. A similar technique is also used to spoof MFA.	Implement TLS 1.2 or higher for key websites and train users to recognize the correct URLs. Disallow newly registered domains and implement a strong proxy filtering control where possible.
Stealing and forging Kerberos tickets (TGT)	Similar to session hijack, the Kerberos TGTs (called **golden ticket**) is captured and used to generate malicious tokens which are subsequently used.	Strong governance of privilege identities. Consider reset of the **golden ticket** (**GT**) periodically, though it is cumbersome and time consuming. Hardening of domain controller.
MFA attacks	Attackers may compromise passwords and then use several ways to compromise MFA enabled identities such as, by generating MFA requests, sending push notifications or bombarding users with several MFA requests where user may keep accepting the push notification (MFA fatigue).	Use FIDO2 hardware tokens. Deploy device trust. Use MFA with rotating code, such that the user has to observe the prompt on screen and then chose the MFA code on the second device (such as a phone).

Table 7.2: *Common credential stealing attacks*

Passwords are also increasingly being retrieved from web cookies, password managers, and cached domain credentials.

Note: **Organizations considering an enterprise-wide password reset would need to have a planned way to reset the golden ticket first and allow for some time for all DCs to sync before triggering user password resets. While the control is relevant and important, it may impact the availability of the organization's services.**

Tip: **Credential Stuffing attack techniques use breach data that contains user credentials and**

attempt to use the same password against the user's other accounts. It is common for users to set up and use the same account in more than one application.

Figure 7.5 illustrates the MITM, also called **adversary in the middle** (**AITM**):

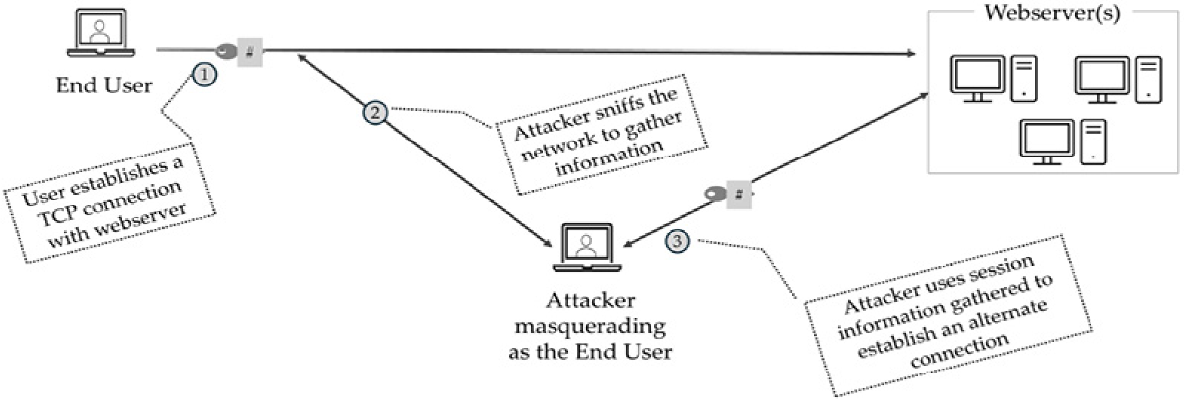

End User
1
User establishes a TCP connection with webserver
2
Attacker sniffs the network to gather information
Webserver(s)
Attacker masquerading as the End User
3
Attacker uses session information gathered to establish an alternate connection

***Figure 7.5**: AITM attack*

In December 2016, Ukraine's electric power infrastructure suffered a cyber attack that crippled its ability to supply electric power to around 220,000 consumers for about 6 hours. The attackers used brute force to gain access to the critical infrastructure, i.e., the power supply network.

Attacks such as those listed above can be detected by using appropriate logging mechanisms and monitoring them. We covered some of those logging requirements earlier in this chapter.

Later in the book, we will cover more details on cyber attacks.

Conclusion

In this chapter, we covered several aspects of IAM and learnt the importance of ID, PAM, and the governance over them. We also covered several identity-based attacks and their mitigation methods.

In the next chapter, we will explore the core principles of the cloud, its uses, and how to secure it.

Key takeaways

Today, access to information systems is through identities: human or otherwise. Here are some important takeaways of topics we examined:

- Authentication, authorization, and accountability are key aspects of identity security.
    - Authentication options such as passwords/passphrases, PIN, passwordless features such as Windows Hello or Apple Face ID, and SSO must be deployed, balancing feasibility, user experience, and security. Technologies such as MFA should be used to prevent account takeovers.
    - RBAC or others, using modern protocols such as OAuth, it is important to design controls based on principles such as *least-privilege* and *minimum necessary*.
    - Security-related actions taken by an identity, such as a password reset, and any action on an information asset, such as changing the data purge settings on an application, should be recorded and traceable for the purpose of accountability.
- Attackers try to elevate privileges, and thus, a governance program for PAM is essential to monitor and control privileged access. Similarly, governance on non-human (non-interactive IDs) is important.
- Identities integrate with other systems in a digital world, and due care must be taken to define and design controls around them.

References

- **https://pages.nist.gov/800-63-3/sp800-63b.html**
- **https://www.w3.org/TR/webauthn-1/**
- **https://attack.mitre.org/tactics/TA0006/**
- **https://attack.mitre.org/campaigns/C0025/**

CHAPTER 8

Cloud Security

Introduction

Cloud technologies have been creating a massive impact on how technology runs today; barely any aspect of today's digital world has not been touched by the cloud. In this chapter, we will understand the foundations of the cloud, its types, and how to secure the cloud environments. We will also explore how some of the traditional technologies have evolved in the cloud and how organizations have adapted to the cloud.

Structure

The chapter covers the following topics:

- Foundations of cloud computing
- Using cloud technologies

Objectives

By the end of this chapter, you will be able to understand the nuances of the cloud, its relevance, and how organizations are leveraging it. As a CISO, you will be required to define and implement organizational policies and procedures for the cloud and its security.

Foundations of cloud computing

In the traditional IT data center setups, organizations often had to plan, strategize, and budget capital expenditure to enable their business applications to run uninterrupted. These setups were costlier, cumbersome to continually and efficiently manage, and often lacked right-sizing due to business demand fluctuations. The IT team also had to be skilled in several aspects of technology. With the rapid advancements in technology, the organization's environment could not justify the return on investment. The gap between new technology and the old environment, called **technical debt**, continued to grow. For instance, the e-commerce portal would need more capacity when the organization runs a sales campaign or during the festive season, compared to an average day.

The idea of internet-based computing resources that could be configured and provided on a pay-per-use basis, much like water and electricity, was floated in 1961 by *John MacCharty*. The idea caught on and was commercially implemented by Salesforce.com in 1999. The company provided internet-based **customer relationship management** (**CRM**) software to enable its customer organizations to service their end-consumers better. Several companies like *Amazon, Google*, and *Microsoft* also started offering cloud-based resources in the early part of the millennium (the decade of 2000).

Note: Organizations that directly launch their service offerings using cloud resources are called born-in-the-cloud. Salesforce.com is one of the first such organizations. Today, there are many organizations bringing unique capabilities, including artificial intelligence (AI), to the cloud.

Cloud computing

According to a definition by NIST, *Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.*

The terms cloud and cloud computing are used interchangeably. The core reasons for cloud computing to be successful were:

- **Reduction of IT capital expenditure:** Reduction of the costs for servers, storage, and network connectivity. For instance, organizations need to set up branch networks and VLANs for their operations. Instead, they could host their application(s) on the cloud and let branches/remote users access those over the internet. This saved telecom costs and other costs in managing IT networks.
- **Easier to test, scale, and expand**: Unlike traditional IT setups, test environments and prototypes could be created and made available much more easily and quickly.
- **Easier to use modern versions**: Software updates happen at a rapid clip, updating the operating systems of servers. Network devices are complex and require skill and planning. However, with the cloud, the switch to the latest version is easier as those technologies can be provisioned and tested faster.
- **Easier reachability/availability**: Unlike traditional IT setups, where complex setups of VPN and MFA, etc, would be required, in the cloud, a remote user could just use the internet to connect to the application from anywhere.
- Advancements in technologies such as **virtualization**, where a single physical server/computing resource could be configured to host multiple versions/instances of the application.

The success of using cloud computing is probably well epitomized by Netflix. Netflix is a popular digital content streaming service that transformed itself from a DVD rental service based in the US in 2008 to a cloud-enabled service with a worldwide reach by 2015/16. They leveraged the capabilities of Amazon’s cloud to not just increase their global reach to about 130 countries but also deftly manage the scalability of infrastructure, relevant regional content to be delivered to users’ devices when they want. Needless to say, their revenues also grew manifold within that period.

Cloud architecture

The cloud architecture has several components, namely:

- **Cloud service provider (CSP)**: the organization offering its services using cloud computing. There will be several CSPs on the cloud. Each CSP:

  - Has a large pool of physical enterprise-grade servers.
  - Leverages some virtualization technology, such as *VMware* V*sphere*, *Microsoft Hyper-V,* that runs on top of these servers.
  - Offer service(s) using these virtualized servers. For instance, a CSP may offer services to view content such as movies, tele-serials, live performances, and even video conferencing.
  - Provides an administration panel, called the **control panel**, to manage entitlements, such as servers needed, storage required, and user management.
  - Mechanism to define and implement security and privacy policies such as ACLs and access entitlement, encryption, etc.
- **Cloud consumer**: The user/consumer of the service offered by the CSP. While generally the consumer is an organization, individuals can also subscribe to cloud services. For instance, we may be using cloud-based storage from *Apple*, *Google*, and *Microsoft* on our personal devices.
- **Cloud service broker**: They are intermediaries that may be used by the cloud consumer to broker the contract with CSPs and/or act as an interface for taking requirements and lining up various cloud services behind the scenes.
- **Cloud carrier**: Provides interconnections between the user and the CSP, for instance, the ISP used by the consumer.

Figure 8.1 illustrates the architecture with one CSP shown:

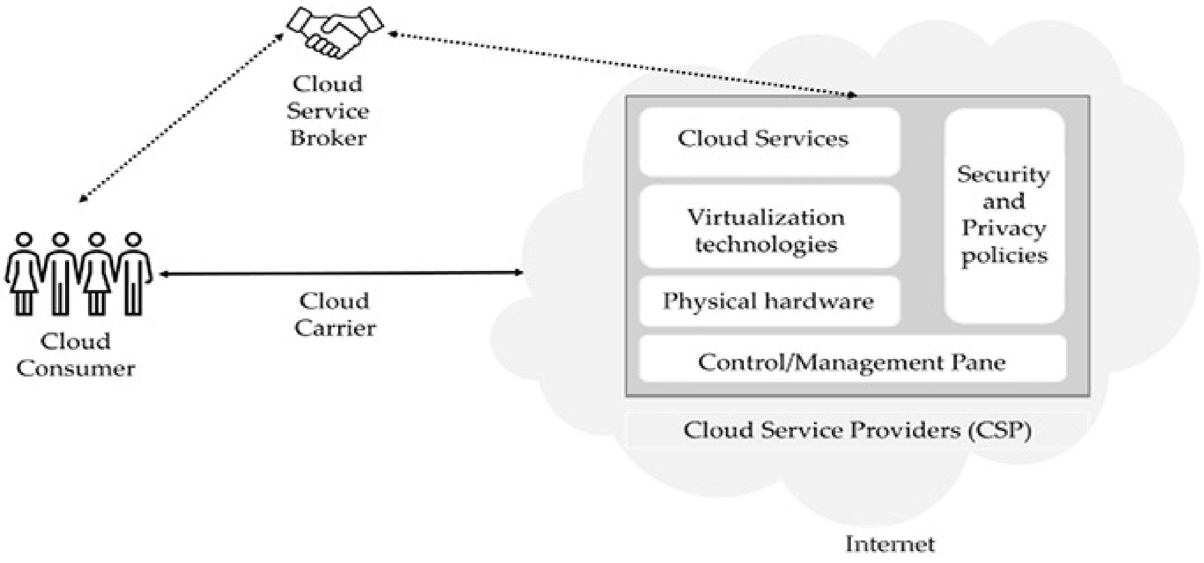


***Figure 8.1**: High-level cloud architecture*

Amazon Web Services (**AWS**) is one of the leading CSPs that offers a wide range of cloud services, including computing power (called **EC2**), storage (called **S3**), databases (called **RDS**), machine learning (called **SageMaker**), and more.

Airbnb is a popular portal and app that allows people to list their homes/properties for other guests to rent and use. Airbnb uses AWS to manage its global infrastructure, handle peak loads, and ensure high availability of its platform. AWS's scalability and reliability have enabled Airbnb to grow rapidly and provide a seamless experience to its users.

Cloud characteristics

As per NIST, a CSP must meet five characteristics for it to be considered, as defined in *Table 8.1*:

Characteristics	Explanation	Security considerations
On-demand self service	The consumer should be able to request and provision cloud services by themselves whenever they want and without having to wait for long time gaps. For instance, in traditional IT, servers would need to be procured, installed, hardened and then made available. These would take several days. On the cloud, it would be almost instantaneous.	Managing cloud resources and their configuration as per security objective using automation.

Characteristics	Explanation	Security considerations
Broad network access	Availability is a big focus in cloud computing to ensure access to cloud resources is available anywhere anytime over the Internet. The proliferation of personal devices like smartphone and digital tablets make it easier for users to access such services.	Information assets on the cloud must be protected for authorized access using principles covered IAM, such as MFA.
Resource pooling	The cloud service should be able to prioritize and leverage a pool of unutilized resources and provide a seamless service to the consumers. For instance, several consumers of the cloud-based application can use the same physical server and such IT environment.	Proper segmentation and zoning should be implemented such that while resource pooling can be automatic, the organizations data does not get exposed to unauthorized user.
Rapid elasticity	The ability of the CSP to adjust the storage space, number of virtual servers needed and similar IT requirements on the go without needing manual intervention or downtime. For instance, the cloud consumer of an e-commerce service temporarily requiring additional server compute time would automatically get those seamlessly.	The CISO should be aware of the environment to protect, and therefore implement required controls such as monitoring and governance on the new **virtual machines** (**VMs**).
Metered service	Monitor and track usage and share those transparently with the consumer to see and adjust limits as needed. For instance, an organization leveraging cloud based storage for its employees can pay for actual storage used per month/per period and/or can also set limits for its user.	Cloud accounts are like privileged Id and must be governed appropriately. If cloud accounts provision resources without checks and appropriate timely revocation, the organization may suffer financial losses. Attackers use unprotected cloud accounts to launch their own malicious services.

Table 8.1 *: Characteristics of cloud computing*

Note: Dropbox leverages broad network access to allow users to store, share, and access their files from any device, anywhere in the world. This ensures that users can collaborate and share files seamlessly, enhancing productivity and convenience.

The primary purpose of cloud computing is to provide better **availability** (refer to the CIA triad). That does not mean that cloud computing does not focus on confidentiality or integrity. The concepts and technologies covered earlier for protecting the CIA are also applied in cloud computing.

See *Figure 8.2*, which illustrates the five cloud characteristics, their key advantage, and one of the security considerations:

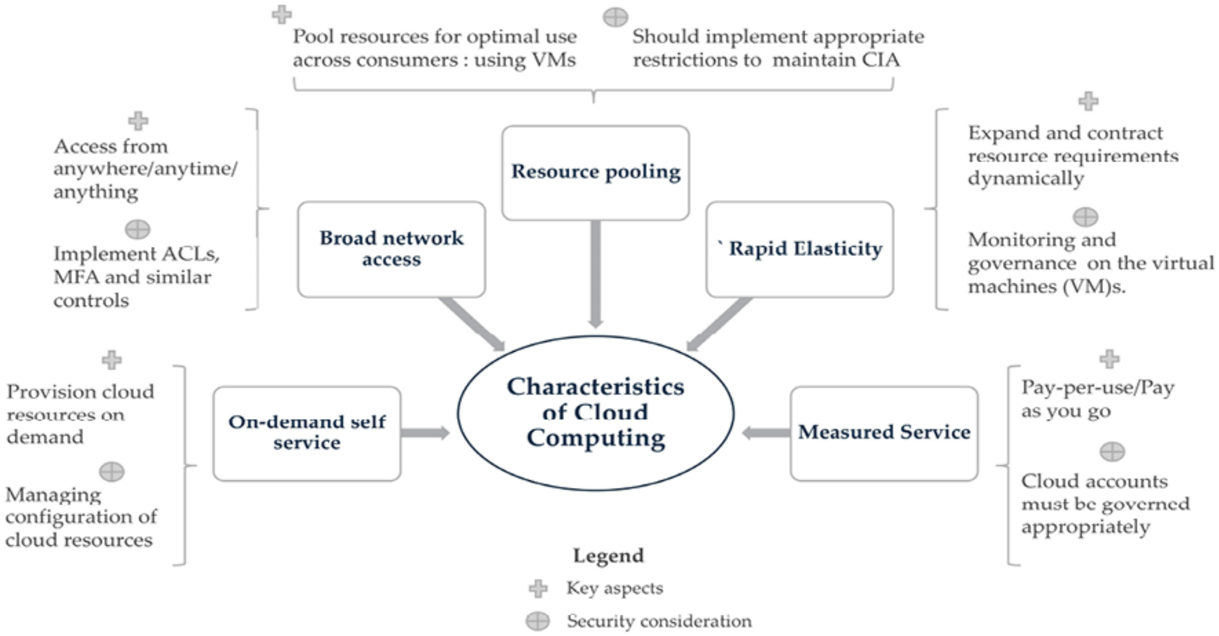

Pool resources for optimal use across consumers : using VMs
Should implement appropriate restrictions to maintain CIA
Resource pooling
Access from anywhere/anytime/ anything
Implement ACLs, MFA and similar controls
Broad network access
` Rapid Elasticity
Expand and contract resource requirements dynamically
Monitoring and governance on the virtual machines (VM)s.
Characteristics of Cloud Computing
Provision cloud resources on demand
Managing configuration of cloud resources
On-demand self service
Measured Service
Pay-per-use/Pay as you go
Cloud accounts must be governed appropriately
Legend
Key aspects
Security consideration

***Figure 8.2**: Cloud characteristics*

Cloud computing largely leverages virtualization concepts, which have three main components:

- **Physical server**, which has the memory, storage, computing power (processor), and networking, and runs the operating system called **host OS**. The host OS could be any of the OS, like *Windows* or *Linux.*
- **Virtual machines** (**VMs**) run the cloud service. The OS on the VM is called the guest OS, and it could be different from the host OS.
- **Hypervisor** manages the resource pooling and provides elasticity to virtual machines. It enables the VMs to function as independent machines and allows them to share the hardware resources of a single host. It can be **bare metal/embedded,** i.e., implemented at the hardware level of the server. It may be a software installed on the host OS, and is called Type2.

Figure 8.3 illustrates the typical hypervisor architecture:

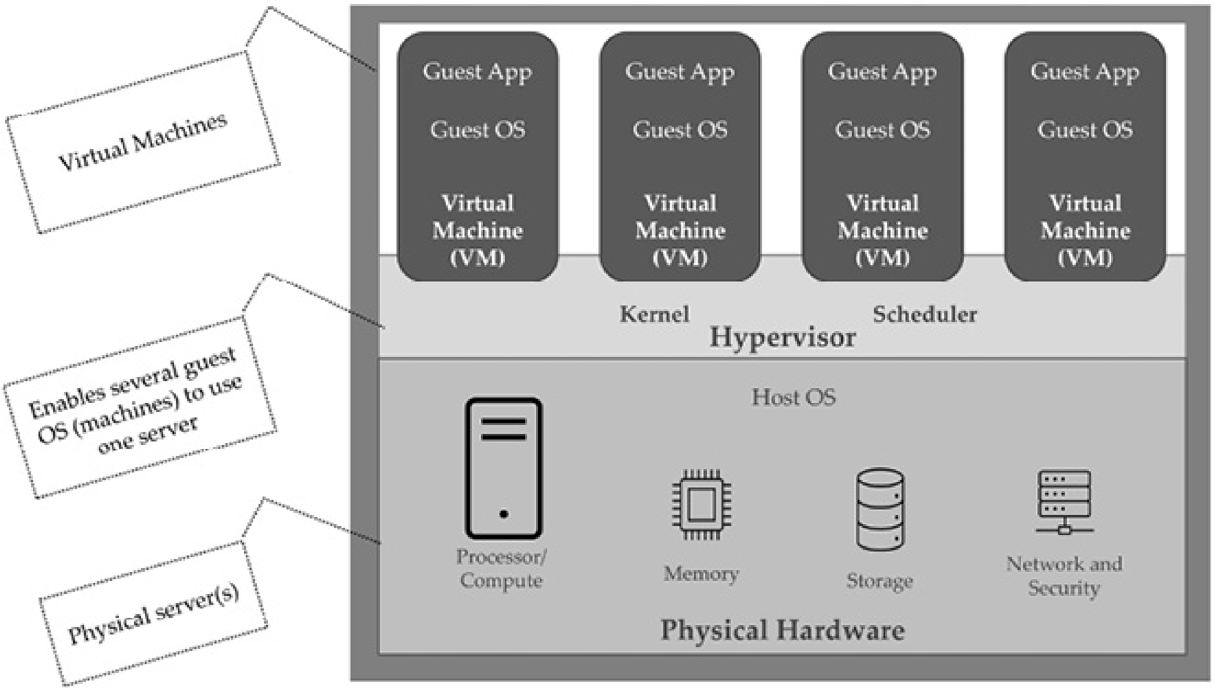


***Figure 8.3**: Hypervisor architecture*

As part of resource pooling, it is possible for two organizations to use the same server/application as if it were their dedicated application. Each logical partition in the cloud is called a **tenant**. Such a copy of the application/service may be referred to as an **instance**. The features licensed (authorized) to be used within the instance are often referred to as the **subscription**.

Note: Cloud-based apps have the ability to provide each customer an independent tenant, whereas an internet-based application allows multiple users on the same application, using the same database and the same storage, etc. For instance, Gmail is a popular webmail service, and every individual who uses a mailbox primarily uses the same application; however, each organization may have a different instance of salesforce.com configured just for them. It is now common for the internet-based application to also be on the cloud infrastructure.

Some of the common security considerations with respect to cloud computing include:

- It is important that hypervisors are configured to ensure independent VMs/tenants are isolated and do not allow inter-VM traffic by default. Otherwise, inter-tenant access may lead to loss of confidentiality of information. This is called **VM hopping**. For instance, when two organizations leverage their own instance of **salesforce.com**, neither of them would want their customers' information to be exposed to the other. The ability of the VM to access other VMs is called **guest OS breakout**.

- Preventive, detective, and administrative controls must be selected and implemented in line with the organization's security objectives. The benefits of the cloud should be leveraged, and organizational policies must be adjusted accordingly without causing undue risk to the confidentiality and integrity of the organization's information assets. Some examples include:
    - The host OS and guest OS must be hardened as per organizational policies. Using automation, the VM should be configured in a way that it meets the organizational policies, even though it may remain on-demand for the user.
    - Protecting and governing cloud accounts, i.e., the administrator of the CSP itself.
    - The control panel (see *Figure 8.1*) should also be protected via a relevant RBAC policy.
    - Network segregation should be implemented such that traffic between the host OS and services is on a need-to-have basis.
    - Appropriate logs must be configured.

Cloud service models

Cloud computing is structured in three main service models. Each of these models provides the cloud consumer options that suit their business purpose the most. Irrespective of the model, the five characteristics hold true. According to NIST, these models are defined as:

- **Infrastructure-as-a-service (IaaS)**: *The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include OSs and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over OSs, storage, and deployed applications and possibly limited control of select networking components (e.g., host firewalls).*

    - Used when an organization wishes to build and manage its own configuration of an operating system and the application/cloud service on top of it. For instance, an organization may use IaaS to build, host, and manage its website.

      A CSP may also use the IaaS services of other CSPs.
    - CSPs offering IaaS include **Amazon Web Services** (**AWS**), **Google Cloud Platform** (**GCP**), **Microsoft Azure**, **IBM Cloud**, **Oracle Cloud Infrastructure** (**OCI**), **Rackspace**, and **Alibaba.** Each of these IaaS providers has unique characteristics, even though their fundamentals remain the same. For instance, AWS is a large global player with extensive IaaS capabilities, GCP is commonly used for data analytics and **machine learning** (**ML**), and OIC is used for leveraging Oracle databases and products on its cloud.
- **Platform-as-a-service (PaaS)**: *The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure, including network, servers, OSs, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.*
    - Used when an organization wishes to leverage a CSP's cloud infrastructure to develop and/or deploy its custom or procured application/service. For instance, a healthcare insurance company may choose to deploy its claims management software on the cloud using a CSP's PaaS model. Similarly, app-based cab rental/cab-hiring services such as *Uber*, *Lyft*, or *Hertz* may use PaaS to host their application on the cloud.
    - Common PaaS CSPs include AWS Elastic Beanstalk, Google App Engine, Oracle Cloud Platform, Heroku, and Microsoft Azure Pipelines. Each of these PaaS brings different approaches and native support to different programming languages, databases, and integration with code repositories. For instance, AWS Elastic Beanstalk supports languages such as *Java*, *.NET*, *Python*, *Node.js*, and *Ruby* on Rails. Google App Engine allows applications to run

seamlessly across several Google-managed servers and supports the default management of SSL certificates.

Tip: Heroku, a CSP, provides a PaaS environment for developers to build, run, and scale applications. Companies like Toyota use Heroku to develop and deploy applications quickly, without worrying about the underlying infrastructure. This allows them to focus on innovation and delivering value to their customers at speed.

- **Software-as-a-service (SaaS)**: *The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure, including networks, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.*
  - Used when a cloud consumer wishes to leverage the capabilities of software already on the cloud for its own use. For instance, an organization providing business services to other organizations may leverage Salesforce.com, a popular CRM, to acquire and manage its customers throughout the deal cycle, i.e., customer hunting, acquisition, contracting, renewals, and so on.
  - Organizations may develop and offer their services by leveraging existing cloud capabilities. Some of the common SaaS offerings available are:
    - **Communication and collaboration platforms**: For instance, Slack, Microsoft Teams, or Google Business.
    - **Project management software**: For instance, Atlassian.
    - **Video conferencing services:** For instance, Zoom.
    - **File sharing**: For instance, Dropbox.
    - **Code repositories**: For instance, GitHub.
    - **Document management**: For instance, Adobe.
    - **Entertainment content**: For instance, Netflix or Amazon Prime.
    - **Secure code testing**: For instance, Fortify on Demand.

        - **Network security:** For instance, Cloudflare.
        - **Identity management and SSO**: For instance, Okta and SailPoint.
        - **E-learning platforms**: For instance, Cornerstone *OnDemand*, Adobe Captivate.
        - **Internet proxy**: For instance, Zscaler and Netskope.
        - **Endpoint detection and response (EDR):** For instance, CrowdStrike, Sentinel One, and Microsoft Defender.

    - The CSP has the onus of managing secure coding, privacy by design, and managing the configuration weaknesses of its software. Cloud consumers would need to focus on security measures for their data. For instance, it would need to define and implement RBAC, MFA, Encryption, and other such controls. Depending upon the agreement with the CSP, a cloud consumer may also be responsible for ensuring their instance is appropriately secure and continually tested for weakness, and a risk-based approach to fixing such gaps is used.

Tip: Almost every software capability is available as a service and is denoted as XaaS.

While the physical security of cloud infrastructure is the responsibility of the CSP, the security and privacy of the data and implementing risk management practices will always be that of cloud consumers. As per NIST definitions of IaaS, PaaS, and SaaS, it is clear what a cloud consumer controls and what they do not. This aspect is a key determination of what is called the **shared responsibility model**. The cloud consumer and the CSP both have some shared responsibilities to manage the CIA. The CSP will remain focused primarily on availability. The **cloud consumer** must additionally focus on the following to meet the business and security objectives for confidentiality and integrity of data:

- **Service level agreements (SLA)**
- Roles and responsibilities in the shared responsibility model, for instance, the CSP may own the responsibility for maintaining the security of the platform and tenant, but the responsibility and security of the access to the tenant/instance are that of the customer.

- Security controls—design and effectiveness.

Tip: Even cloud environments are susceptible to cyber attacks. Between June and September 2025, several organizations that use Salesforce for their CRM were part of a targeted attack across several business verticals. The largest known organizations include Salesforce instances of Google, Workday, Allianz Life, Adidas, and KLM-Air France. These cases highlight the importance of the shared responsibility model.

Figure 8.4 illustrates the various cloud service models and the ownership and management:

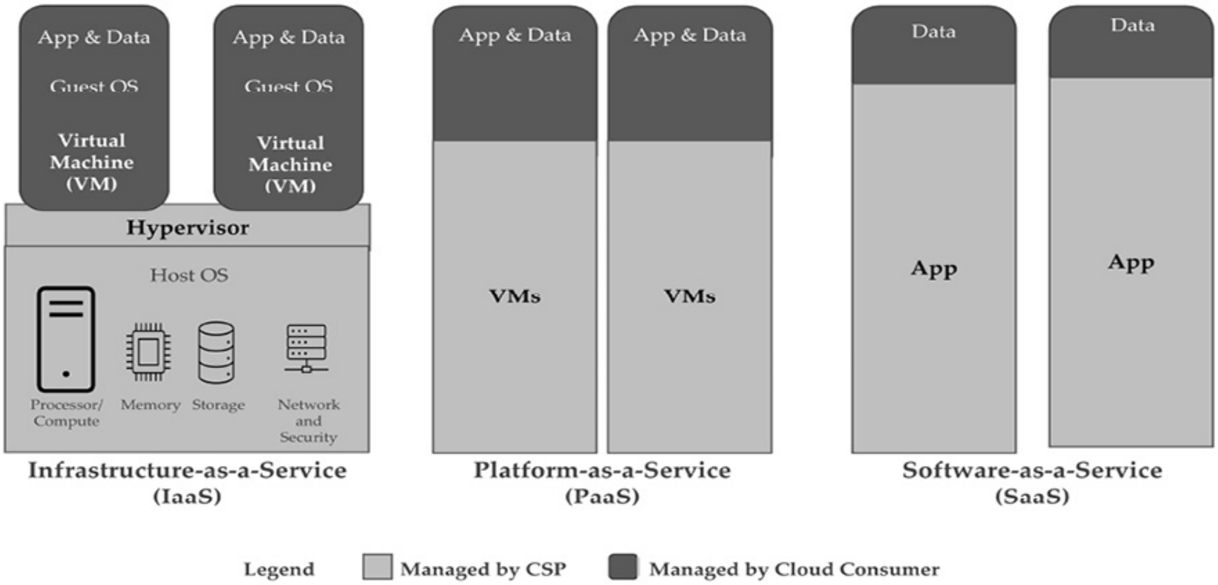


***Figure 8.4**: Cloud service models*

Applications and services running on the cloud are also called **cloud workloads/workloads**. Workloads can be set up to be always available or stood up using automation just-in-time when needed. For instance, the e-commerce website may be hosted on the cloud. Based on its user, it may additionally make features like purchase analytics and suggestions for user purchases available, running on an on-demand workload. The inter and intra operability of workloads is managed with a capability called **orchestrator**.

VMs need an operating system to function and thus may require more memory, storage, and computing power. All of these cost money. In lieu of VMs, a lightweight alternative called **containers** can be used. The container encapsulates the functioning of the application and its dependencies (such as environment variables) in such a way that it can run independently, irrespective of the underlying host OS. This makes containers highly efficient in portability across CSPs, too. Technologies such as **Docker** for creating containers and **Kubernetes** for managing information exchange between containers and/or host OS are used. In each of such technologies, security aspects like hardening, RBAC, and monitoring are required to be implemented.

Large/complex Applications may be structured as small independent portions/components (called **microservices**) with independent code, databases, and protection mechanisms such as access management. They are independently developed, tested, and deployed, usually in containers. Microservices interact well through APIs.

Cloud deployment models

According to NIST, the four possible deployment models of cloud, their definitions, and characteristics are:

- **Private**: *The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.*
    - Generally, organizations in highly regulated markets or industries, or organizations whose risk appetite is quite low, would choose to be on a private cloud. For instance, a healthcare service provider would prefer a private cloud as it would be the only tenant, and the confidentiality and integrity of PHI would be easier to manage.
    - Private cloud deployment may not be an optimal representation of the five cloud characteristics. For instance, resource pooling and elasticity, though technically possible, are less relevant.
    - Popular private cloud providers include **Hewlett Packard Enterprise** (**HPE**), Google Cloud, and VMware. Generally, organizations would leverage IaaS/PaaS services from CSPs.
- **Community**: *The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated*

by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off-premises.

  - Organizations or bodies with common/shared objectives of security, privacy, and operational constraints may use this model as a unit. For instance, resident welfare associations or various administrative bodies of a city may come together to offer their citizens services on the cloud. Similarly, various clinics of primary healthcare services may leverage a **community cloud** to keep the patient data in a controlled way while allowing for easier exchange amongst themselves. The data protection requirements of laws such as the US **Health Insurance Portability and Accountability Act** (**HIPAA**) may be easier to demonstrate to the regulator. Similarly, several regional banks in the **European Union** (**EU**) may use this model for their intra-bank financial settlement processes and branch-level customer services while meeting the requirements of the **General Data Protection Regulation** (**GDPR**).
  - Community cloud deployments are useful for government bodies that may be starting on their cloud journey. This also enables the government to keep the data of its citizens within its territorial boundaries (called **data localization**).
  - All major CSPs offer the ability for a community cloud.

- **Public**: *The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the CSP.*

  - One of the most popular deployments that best embodies the five characteristics of the cloud. It is used in most of the cloud offerings across IaaS, PaaS, and SaaS. All major CSPs offer a public cloud; the organization may select the CSP using criteria such as reputation, reliability, global reach, built-in security and privacy features, cost, ability for easier and cheaper data transfers, etc.
  - Security controls, as per the Shared Responsibility model.

- **Hybrid**: *The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that* *remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).*

- Organizations may opt for a hybrid approach to protect intellectual property, such as in a pharmaceutical manufacturing setup or aerated beverages, by keeping such business-critical aspects on-premises while leveraging other cloud deployment models from CSPs offering services like CRM.
- Microsoft Azure, Google Cloud, and AWS offer strong capabilities to run cloud services within an organization's on-premises data center. It is possible to leverage cloud resources when the capacity of on-premises infrastructure is at its peak. This is called **cloud bursting**. Thus, a private cloud can also leverage public cloud resources. For instance, an e-commerce portal deployed on the organization's data center can expand its computing, storage, and processing capacity during its sales campaign and/or during the weekend or festival season.
- The deployment model provides organizations with an extra layer of security for information assets, but it may be complex to manage and maintain.

Organizations may use a cloud strategy leveraging offerings of various CSPs to minimize their cost, maximize efficiency, and increase global reach, etc. This mix-and-match approach is called the **multi-cloud** strategy.

Figure 8.5 illustrates the cloud deployment and its security considerations:

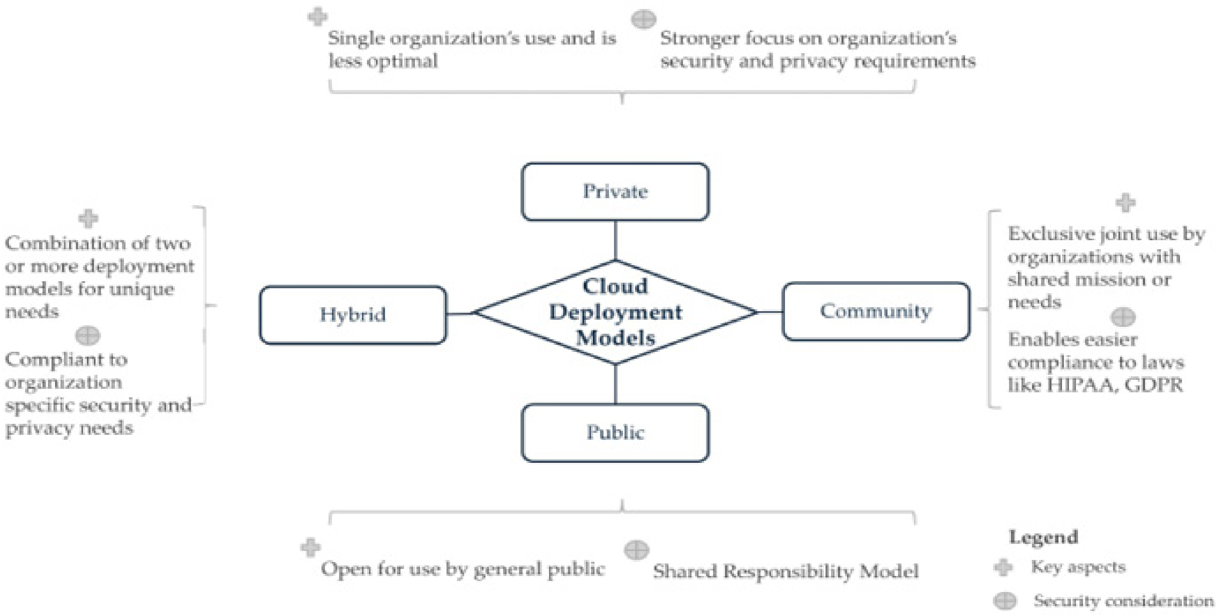

Single organization's use and is less optimal
Stronger focus on organization's security and privacy requirements
Private
Combination of two or more deployment models for unique needs
Compliant to organization specific security and privacy needs
Hybrid
Cloud Deployment Models
Community
Exclusive joint use by organizations with shared mission or needs
Enables easier compliance to laws like HIPAA, GDPR
Public
Open for use by general public
Shared Responsibility Model
Legend
Key aspects
Security consideration

***Figure 8.5**: Cloud deployment models*

Using cloud technologies

Organizations with traditional on-premises infrastructure are adopting cloud services and their preferred deployment model. The organization's business strategy is a primary tether for any decision on adopting the cloud. The business strategy is expected to include the following:

- The growth plans, service offerings, and the geographical scope of offerings. For instance, the body and mind fitness application may choose to deliver all its training programs worldwide, but may choose to have the in-person consult over video call only in India. This strategy is a key determinant when looking for the right CSP.
- Regulatory constraints, if any. For instance, the fitness application may need to offer a local version in China to keep the data within China.
- Cost/Budget available for cloud spend.
- Risk appetite to determine what can be hosted on the cloud or not.

The cloud technologies have enabled even application development methodologies to be faster and scalable. The ability to develop and deploy applications at scale and at a faster pace by having all required teams work together seamlessly. This concept is called **DevOps**. For instance, the application team can start coding by using on-demand provisioning of a test environment, and the IT can use that awareness to build the required capacity and monitoring mechanisms. **Security by design** and **privacy by design** can be built into this development process that has specific use cases to solve for in an agile way (called **Dev pipelines**). We covered these topics in *Chapter 5, Security and Privacy by Design*. The expanded model, which also focuses on security requirements as part of DevOps, is called **DevSecOps.** The code can be automatically set to scan for code vulnerabilities and logic gaps.

Note: Etsy is a cloud-based marketplace for the sale and purchase of goods. They adopted DevOps practices to streamline their development and deployment processes and integrated security into their DevOps pipeline (DevSecOps). This approach enabled Etsy to make security a part of every stage of its software development lifecycle and thus improved its ability to deliver secure, high-quality software quickly.

Choosing the right CSP

The strategy will be useful in choosing the right CSP based on factors such as:

- The deployment model that the organization decides to adopt. Each organization will need to choose largely based on its risk appetite, the regulatory environment, and its data security strategy. The Organization not in the business of cloud computing infrastructure itself would choose IaaS, PaaS, or SaaS from one or more of the CSPs.
- Geographical spread of the CSP. CSPs have massive data centers clustered as **availability zones/zones** in various countries on all continents. Some of the common ones include US East, US West, Germany, the **United Kingdom** (**UK**), Brazil, Japan, India, and China. The groups of these availability zones are called a **region**. For instance, Delhi and Mumbai in India are two regions for Google Cloud. The CSP designs its cloud architecture with high availability; in the event of any zone or region becoming unavailable, the alternate zone/region shall automatically be available. The organization may choose the region to host its offerings in.
- The transparency of controls at the CSP. The organization may choose to use the CSP's SOC2 attestations and/or the CSP's cloud controls assessments, such as **Security, Trust, Assurance, and Risk Register**

(**STAR**), on industry-neutral platforms like **Cloud Security Alliance** (**CSA**). The important security practices include key management for encryption.

- **Cost and metering rates** of the CSP. For instance, what does the CSP charge for porting data from on-premises to their cloud or between CSPs or even between regions of the same CSP?
- Interoperability and data exchange mechanisms between CSPs and organizations' on-prem environment, or between CSPs using APIs.
- The exit clause in the CSP contract should not prohibit the organization from terminating the contract if it so chooses.
- **Possibilities of automation**: For instance, an organization can use coding to provision its cloud infrastructure with the right type of hardening of servers, access control of network and security devices, and monitoring rules. This kind of automation, called **infrastructure as a code** (**IaaC**), is transformative for cloud computing as it makes speed-to-outcomes for customers faster and more predictable. IaaC is also useful for on-premises environments. Some of the popular IaaC tools include Terraform, Ansible, Azure Resource Manager, and Pulomi. Each tool has its own special capabilities, and describing them is outside the scope of this chapter. However, as a security officer, you should be looking to ensure IaaC also uses secure coding practices, is not vulnerable, is managed with appropriate rights, and is part of overall governance, etc.

Note: Spotify, a popular music/podcast streaming service, chose GCP for its data analytics capabilities and machine learning tools. GCP's global infrastructure and advanced analytics services have enabled Spotify to analyze user data, deliver personalized recommendations, and enhance user experience.

Cloud migration

Similarly, migrating to the cloud requires careful planning and execution. For the most part, cloud environments are more modern and not necessarily compatible with technologies used by the organization in its data center. Some of the considerations for **migrating to the cloud** include:

- **Strategy for the cloud**: The objectives for migrating to the cloud, its success factors, and top management support are key and must be documented and appropriately socialized within the company. For instance, the deployment model to use and the security and privacy controls to be implemented. The CISO and his/her team would need to be mindful of the security configuration of the on-demand server and its accessibility, encryption, and vulnerability management. Once the strategy is defined, a cloud service broker and cloud architects can help define the next steps.

- A defined **scope** of infrastructure and/or applications to migrate, and the order of their **priority**.
- **Timeframe for migration**: Traditional organizations have complex application connections, networks, and integrations. Often, those dependencies are not well documented. Any approach to migrating the current tech stack must be done with caution. Older technologies may not work well on the cloud. It may be fruitful to leverage modern-day technologies to revamp the organization's applications on the cloud.
- **Cloud infrastructure, application, and security skills**: The concept of cloud is different, it operates differently, and needs to be managed differently. The cloud migration team would need to be equipped with the skills accordingly. This is also relevant to ensure that concepts like **secure by design** and **privacy by design**, covered in *Chapter 5, Security and Privacy by Design,* can be met.
- **Regulatory, contractual, or industry requirements**: For instance, the on-premises environment may be set up to meet the requirements of GDPR, HIPAA, or the **Payment Card Industry Data Security Standard** (**PCI DSS**), but those may need to be appropriately set up on the cloud.
- Clarity on the **shared responsibility model** and the implications for the organization's efforts.
- Cost/Budget constraints in building or moving to the cloud, cost of data portability and transfer between on-premises and the cloud, etc.
- **Data portability** between CSPs may also be explored, and thus, the contracts with CSP must be carefully drawn to avoid unreasonable data transfer costs. For instance, the organization must be clear on the cost of migrating its existing data to a CSP's application. Similarly, any charges for API data movement costs must be transparent.

- The **ability to terminate the CSPs'** leverage without unreasonable roadblocks. CSPs may try to prevent the loss of the cloud customer's business by forcing a multi-year contract and limiting the possibility of the cloud consumer exiting. This is called **vendor lockout** and should be avoided.

Note: Capital One, a leading American bank, closed its eight data centers and migrated to the AWS cloud. They selected AWS for reasons that included AWS's perceived alignment with Capital One's vision of being a technology company that did banking, and their focus on customer service.

Note: Netflix's cloud migration journey took years of meticulous planning and execution. As part of their cloud migration, they also focused on ensuring their on-premises/unused data centers are decommissioned.

Managing cloud resources

Earlier in the chapter, we determined that the cost of scalable infrastructure is a key reason for the success of cloud computing. CSP uses metering to allow customers to pay by use. The customer must apply proper due diligence and governance to prevent unwarranted costs. Some examples are:

- A developer provisions a test server to demonstrate the prototype of the new version of the software. However, they do not de-provision the server.
- The storage used by the organization continues to grow without real business need, and periodic data purging or data storage limits are not implemented. For instance, all popular cloud-based email providers allow only some **gigabytes** (**GBs**) of storage.

Provisioning cloud resources is a privileged activity and must be governed under a **privileged access management** (**PAM**) program. The cloud administrator uses the control panel/management panel on the CSP's portal to administer the cloud. An organization may have multiple cloud identities (cloud accounts) to manage the organization's requirements.

Tip: Adobe uses AWS to manage its cloud resources efficiently. By implementing a robust governance framework and using AWS's cost management tools, Adobe ensures that its cloud infrastructure is optimized for performance and cost-effectiveness. This approach has enabled Adobe to scale its services, manage costs, and deliver high-quality products to its customers.

Important cloud technologies

There are several technologies used on the cloud that are a significant reason for the success of cloud computing and are relevant from a security standpoint. Some of them are:

- **Software-defined networks (SDN):** Software is used to efficiently control and manage aspects such as the network topology, its routing, updates, ACLs, and performance parameters. Unlike traditional architecture, the SDN decouples the control from routers and switches and centralizes it for easier management. SDN may be considered to be operating at the OSI model *Layer 2* (data link) and *Layer 3* (network) primarily for core networking and data-forwarding-related aspects; it also operates at higher layers such as *Layer 7* (application). SDN uses an **application programming interface** (**API**) to manage the network. SDN is structured into three layers, namely:

    - **Application layer**: Where security technologies like **intrusion detection systems** (**IDS**), Firewalls, and availability aspects like Load balancers operate.
    - **Control layer**: Abstracts the actual underlying physical hardware and allows the application layer to use a component called **SDN controller** to route traffic to the desired destination. This is akin to the L3 capabilities performed in the OSI model.
    - **Infrastructure layer**: The physical infrastructure, such as the switches that the control layer oversees.

  Using SDNs, network segmentation can be defined and implemented in a more efficient and dynamic way. This helps in ensuring the traffic within the network can be configured to adhere to principles such as *need-to-have/minimum necessary*.

  The SDN concepts can be applied to **wide area networks** (**WAN**), i.e., have software that shall define, manage, and dynamically optimize network paths and routing. This is called **SD-WAN**. SD-WAN can automatically optimize the network path to take, reducing costs on dedicated telecom circuits between branches/partner offices. SD-WAN shall also be able to allow prioritization of access to specific applications. A view of branches and the main data center connecting to each other and over broadband to an SD-WAN on the internet is represented in *Figure 8.6*:

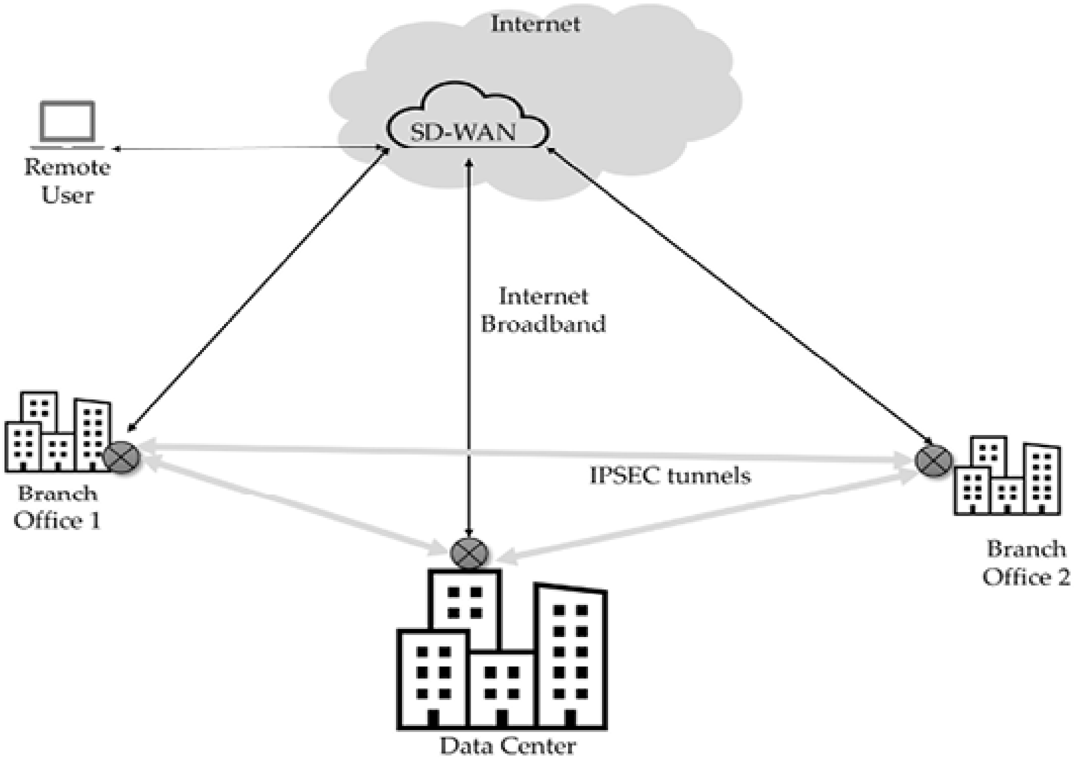


***Figure 8.6**: Representational view of SD-WAN*

SD-WAN may prove to be a bane as it establishes an inherent trust between the sites (say branches), thereby expanding the possible attack surface via the internet. Due care must be taken with appropriate rules, blocking unwanted ports and services, and ensuring the ACLs are structured well.

- **Cloud access security broker (CASB)**: As applications and databases continue to make cloud their home, the protection of data and its access becomes more relevant compared to an on-prem solution. A CASB solution, which is generally a SaaS offering itself, brokers/intermediates the user access to cloud and/or on-prem applications. This ensures the organization’s security requirements for data access and provisioning and that other principles of IAM are met. CASB plays an important role in security with capabilities such as:

- **Data security**: Ability to encrypt data, provide **data loss prevention** (**DLP**) capabilities
- **Visibility**: To know and monitor all corporate applications and other shadow IT infrastructure.
- **Compliance:** Provide the ability for authorization of access at a granular level, such as those mandated by regulations like GDPR, HIPAA, or standards like PCI-DSS.
- **Threat prevention**: Prevent malicious traffic by using anomaly detection techniques and/or integration with the audit and logging capabilities of the organization. We cover threat protection, intelligence, and incident response later in the book.

  Figure 8.7 illustrates a CASB that may be on-prem or on the cloud, but can broker user requests for access and its authorization:

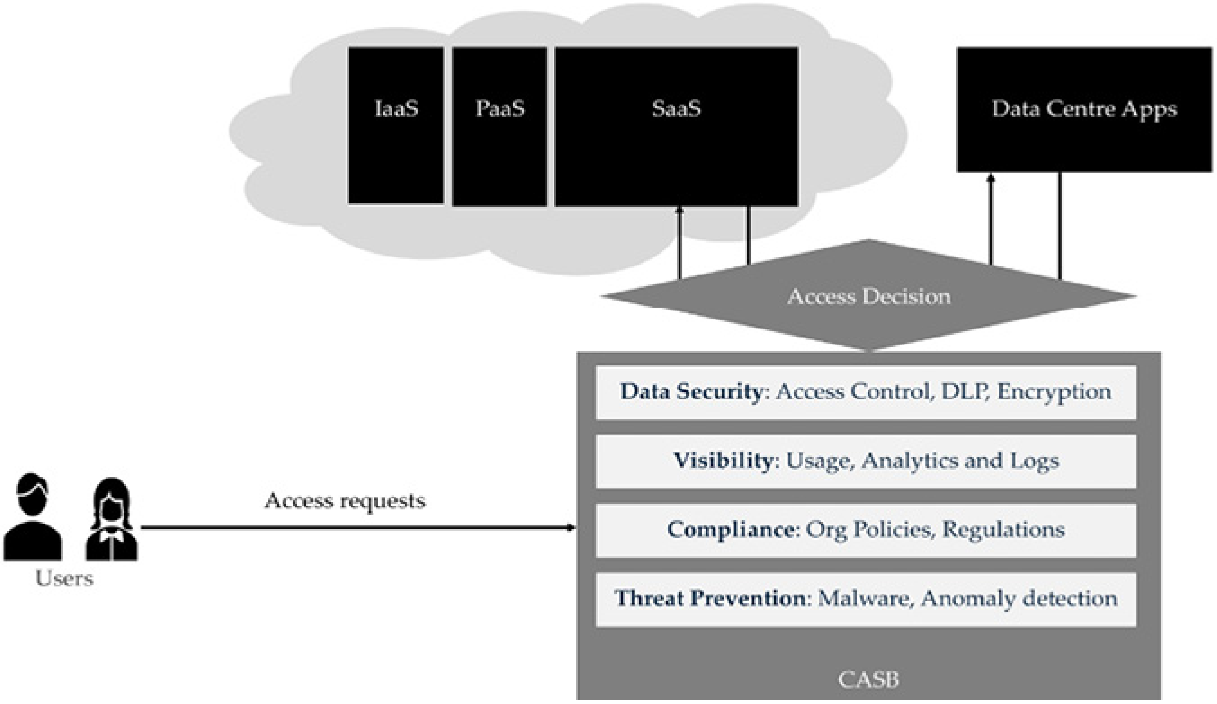

IaaS
PaaS
SaaS
Data Centre Apps
Access Decision
Data Security: Access Control, DLP, Encryption
Visibility: Usage, Analytics and Logs
Compliance: Org Policies, Regulations
Threat Prevention: Malware, Anomaly detection
CASB
Access requests
Users

***Figure 8.7**: Representative view of CASB*

Some of the popular CASBs used are:

- **Microsoft Cloud App Security** (**MCAS**) is commonly used to administer cloud security controls for Microsoft's cloud-based email service called M365. It can govern user and their access to the

  - approved M365 services based on the identity and/or device used.
  - **Next-Gen CASB**: Palo Alto Networks' CASB solution couples its strong capabilities in firewall security and its cloud security offerings.
  - **Symantec CloudSOC CASB**: Integrates well with their DLP product offering.
  - Other popular products include those from organizations such as *Netskope*, *Zscaler*, *Proofpoint*, and *Forcepoint*.

- **Cloud security posture management (CSPM)**: One of the key characteristics of cloud computing is **on-demand self-service**. From a security standpoint, this meant that anyone could configure and set up an environment of their choice, but it may not meet the organization's security objectives. A user in the organization may be able to set up a cloud account, provision a server using that, and host PII data with limited or no access controls. Such a setup defies the organization's security objectives of protecting the CIA of data. Sometimes, such an absence of access controls may even be inadvertent. For instance, the cloud user may not have turned off the *everyone access* to cloud storage. To overcome these challenges and to continually monitor configuration lapses to either prevent or remediate them, organizations leverage a SaaS offering called **CSPM**. Using a CSPM also enables an organization to consistently apply its security baselines, especially for its production environments, and prevent untoward security issues. An organization may allow its development team to host test websites/application prototypes on demand with fewer constraints. CSPM also has the ability to alert the appropriate teams for issues it detects that need attention. Some of the popular CSPMs include Wiz.io, Microsoft's Defender for Cloud, and Palo Alto's **Prima Cloud**.
- **Secure access service edge (SASE)**: This brings about an integration of security controls of access, data/application, network security, user devices, and even identities. The key features a SASE offering delivers are:

  - Efficient management and administration of cloud IT operations, including WAN networking.

    - Stable and secure remote access, for end users and for branch office setups. Fundamentally, irrespective of the user’s location, access to the organization’s cloud and on-premises infrastructure can be managed with security considerations built in, for instance, disallowing rogue devices to connect to the organization or disallowing users from accessing the corporate payroll website from handheld devices.
    - Integrates services such as internet access (proxy/secure web gateway), CASB, and device firewalls.

  Many organizations capitalized on the capability of cloud-based SASE to enable work from anywhere during the infamous COVID-19 pandemic of 2020. This technology allowed remote users to connect to corporate environments securely and also be governed for the websites they can access. Administrators could also use the internet to connect to cloud-based SASE services and then manage remote servers. Some of those benefits continue to be used even today.

- **Cloud workload protection platform (CWPP)**: Workloads are a key component within the cloud. In order to monitor and manage the appropriate workload security configurations and their runtime security issues, for instance, the elevation of privilege of a user to admin when the application crashes.
- **Cloud-native access protocol provider (CNAPP)**: Protects various aspects of cloud infrastructure throughout the lifecycle, including code, runtime, and configuration of infrastructure, application, and databases. It covers microservices, containers, and even cloud orchestration. CNAPP also delivers the functionality of CWPP.

Note: As covered in earlier chapters, we use the term security to also imply privacy aspects of security.

Conclusion

In this chapter, we covered the cloud, its characteristics, service models, migrating to the cloud, and security considerations.

In the next chapter, we will explore core concepts of **Zero Trust** (**ZT**).

Key takeaways

Cloud technologies have been creating a massive impact on how technology runs today; barely any aspect of today's digital world has not been touched by the cloud. Some of the learnings include:

- Organizations continue to leverage the benefits of cloud computing to scale their operations, make their applications modern, reduce the cost of running the infrastructure, and enhance their reach.
- The most important security aspect for the organization leveraging cloud technology is the **shared responsibility model**. Protecting the cloud computing environment may be that of the **cloud service provider** (**CSP**), but more often than not, protecting the data on the cloud application will be that of the organization.
- An organization's business regulatory environment, capability needs, risk appetite, and current state of feasibility of application migration may compel an organization to choose a **hybrid** or a **private cloud** model, though arguably, **public cloud** may be the most effective from a functionality and cost perspective.
- Cloud resources may be provisioned in an automated fashion using IaaC, and aspects of cloud configuration may further be protected using CSPM tools.
- Cloud computing has unleashed the power of several interesting applications, such as cyber ranges to learn, phishing campaigns to use for training workforce, managing prospective customer leads and their conversion as customers, and many processes such as investments, audio-video content, and gaming. Securing those and from those are relevant aspects for the security team.

References

- **https://www.salesforce.com/products/what-is-salesforce/**
- **https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf**
- **https://cloudsecurityalliance.org/star/registry**

- **https://learn.microsoft.com/en-us/defender-cloud-apps/what-is-defender-for-cloud-apps**
- **https://www.netflix.com/in/**
- **https://www.airbnb.co.in/**
- **https://www.heroku.com/**
- **www.etsy.com**

Join our Discord space

Join our Discord workspace for latest updates, offers, tech happenings around the world, new releases, and sessions with the authors:

https://discord.bpbonline.com

CHAPTER 9
Zero Trust

Introduction

Data is said to be the new oil. The value of data and information is immense. They are also one of the most sought-after valuables. Organizations and institutions, including those that are not-for-profit, continue to be the focus of cybercriminals. In this chapter, we will explore the concept of Zero Trust and apply it to protect the data.

Structure

The chapter covers the following topics:

- Foundations of Zero Trust
- Applying Zero Trust

Objectives

By the end of this chapter, you will be able to understand the definition of **Zero Trust** (**ZT**), its characteristics, and the differences from traditional security models or layered defense. We will also understand the challenges ZT aims to solve and some practical aspects of how it can be implemented.

Foundations of Zero Trust

John Kindervag, a researcher with Forrester, coined the term ZT and outlined the philosophy in 2009/10.

The traditional security model has layers of defense to protect access into the perimeter/network using firewalls and VPNs, but once inside the network, an authenticated user/device is largely treated as trusted. The traditional model is often referred to as the castle-and-the-moat model or **trust-but-verify** model. The ZT model professes **never-trust–always-verify**.

As per the *National Institute of Standards and Technology (NIST), Special Publication SP 800-207, Zero trust (ZT) is the term for an evolving set of cybersecurity paradigms that move defenses from static, network-based perimeters to focus on users, assets, and resources.*

In a ZT model, trust is not implicit for either the user or for his/her authorization, even in scenarios such as when a user:

- Has already authenticated.
- Is in the known physical or logical network.
- Is the owner of the asset and/or a known user of the asset.

The five key reasons for the increasing popularity of ZT's adaptation are:

- Increasing cyber-attacks, especially **advanced persistent threats** (**APTs**), where the attacker launches the cyber-attack much later than the time of actual infiltration into a corporate network. For instance, in 2010, a malicious code called Stuxnet was found in scores of Iran's nuclear facilities. The Stuxnet worm crippled the centrifuges and thus Iran's ability to enrich uranium. The Stuxnet was seemingly active on the system for many weeks before it orchestrated the attack on 1000s of centrifuges. This example is relevant because there was malicious code inside a trusted environment that turned hostile.
- Anywhere-work/hybrid work models, i.e., providing users access to the corporate environments from anywhere, anytime.
- Increasing leverage of the cloud lent itself to the relevance and popularity of the ZT philosophy/framework.
- Security vendors and consultants are pushing the ZT terminology into their products and services.

- Government agencies provide a fillip to the ZT's popularity. For instance, the US's **Cybersecurity and Infrastructure Security Agency** (**CISA**) produced a maturity model for the ZT framework.

The ZT framework does not imply that firewalls and such perimeter tools are irrelevant. It just changes the way identity and data access may be approached. To implement the ZT philosophy, an organization needs to define the relevant architecture encompassing people, assets, and relevant resources. Organizations can leverage NIST's **Zero Trust Architecture** (**ZTA**) to progress on the journey of ZT without having to undo/dismantle their current defense-in-depth security models. They can strategize, plan, and incrementally progress on the ZTA.

TIP: BorgWarner, a global manufacturing enterprise that built over the years in the traditional castle-and-moat approach, embarked on a transformation journey using principles of Zero Trust. Their focus on optimizing security architecture, reducing cost, migrating to the cloud, enhancing user experience, and faster integration of their mergers and acquisitions (M&A) led them to choose a leading ZT technology vendor. In a span of six months, BorgWarner moved significantly ahead on its ZT journey, during which it eliminated about 90 firewalls and several complex network connections. They leveraged a cloud-based proxy solution to enable a café-like experience for its workforce.

Key tenets of Zero Trust Architecture

The seven key tenets for the ZTA, based on the NIST definition, are as follows:

- **Everything in the organization's information system eco-system is a resource**: The applications, servers, databases, APIs, cloud containers, network components, including Wi-Fi access points, the users of the information systems, and the data itself are all resources.
- **All communication must be secured irrespective of the network**: Unlike the traditional network, where the internal LAN/VLANs and thus devices on them are inherently trusted, ZTA professes that the location of the device, internal or external (internet-based), does not determine the trust. They must be treated as untrusted devices, and their access and communication constraints should be much like those of devices not owned/managed by the enterprise. For instance, in the traditional network approach, only a remote user connecting via a secure VPN is required to use **multi-factor authentication** (**MFA**). In the

ZTA, even a user in the physical premises of the organization and connecting to authorized network ports would also require MFA.

- **Session-based access to each resource**: Every user should authenticate every time they access a resource within a set limit of time, and such access would not automatically extend to another similar resource by default. In the traditional model, once logged into a suite of products such as M365, once you authenticate to Outlook, you may not need to authenticate again to MS Teams or SharePoint. Similarly, once authenticated to Gmail, all other subscribed services like *YouTube, Google Meet*, etc., are automatically authenticated. In the ZTA, each such resource would, ideally, require authentication by the user. This may be challenging for the user experience and the culture of the organization. The organization may need to take a conscious risk-based decision to allow collaboration apps in a suite (M365/Google) to not require re-authentication, but at the risk of compromising on true ZT principles.
- **Access to a resource is granted using a dynamic policy**: The policy will evaluate the whole context, i.e., identity context, user logins/programmatic service account logins, device contexts, such as software versions, network traffic path and anomaly, and current attack intelligence. In the traditional model, the authorization policy is usually static and generally based on the source of the network and may require a step-up authentication like an MFA or a phishing-resistant MFA form factor (covered in *Chapter 7, Identity and Access Management*). In the ZT approach, the policy will dynamically determine and trigger an MFA for access to a resource based on the requester, its device, network path, etc.
- **Security posture of the asset is validated prior to granting access**: In the traditional model, device posture is validated, using concepts like **host-checker**, for connections to VPNs/VDIs. Once the asset establishes a VPN connection, the security state is not rechecked. In the ZT approach, validation may be done using a risk-based approach for each resource access. The ZTA professes **continuous diagnostics and monitoring** (**CDM**) for validating the safety of access to the asset. Using the principles of *least privilege*, any asset that does not meet the

security expectation/baseline, such as the latest updated patch, may not be allowed to access some resources. For instance, the organization may allow limited internet access to get the patch updated, but deny access to corporate employee leave management applications or other business applications until compliance is met. The CDM comes in handy as the state of security is dynamic with the changing security landscape. Records of this access are kept as part of the *Accounting* requirement we learned in *Chapter 7, Identity and Access Management,* and are used for determining behavioral anomalies. Therefore, if an asset may have been secure when it was first connected, the security state may have changed with time, and thus, the access must be re-evaluated when it requests access to a resource.

- **Authentication and authorization to all resources are dynamic, continual, and risk-based**: This tenet implies that for each resource access, there shall be a lifecycle of granting and reviewing access, re-evaluation of trust even in an ongoing communication channel, its current risk and threat context. It may mean reauthentication and reauthorization. IAM and a strong asset management program work in tandem to ensure that user and device context are up-to-date and can be validated for continual trust even during an established connection (such as an SSL connection).
- **Use threat awareness, user, and device context to automatically refine policy enforcement**: This tenet proposes to continually adapt the access and authorization policy and its enforcement using the awareness of factors such as network traffic, security posture of the device, and access requests. For instance, when the threat intelligence suggests traffic originating from a certain geography may likely be malicious. The policy should dynamically correlate such aspects and access requests from such regions for any resource, which must automatically require more stringent authentication, such as step-up MFA and updated device security posture. The policy may automatically provision only a read-only view and elevate that to the administrator's access only when certain/additional checks have been successfully completed.

Figure 9.1 represents the tenets of ZTA with key tenets:

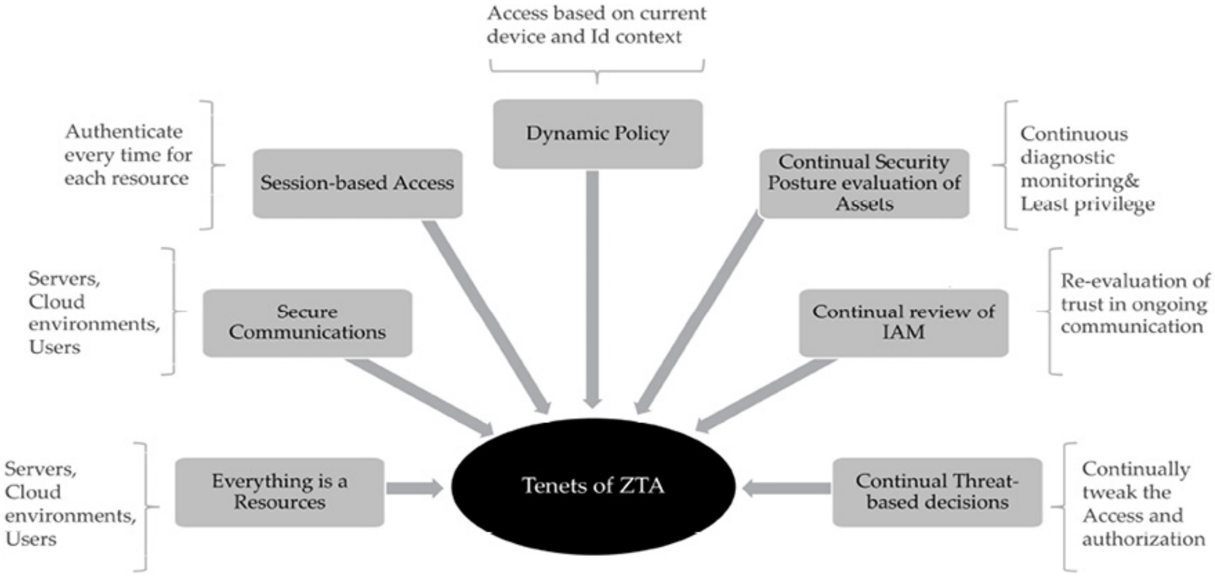

Access based on current device and Id context
Dynamic Policy
Authenticate every time for each resource
Session-based Access
Continual Security Posture evaluation of Assets
Continuous diagnostic monitoring& Least privilege
Servers, Cloud environments, Users
Secure Communications
Continual review of IAM
Re-evaluation of trust in ongoing communication
Servers, Cloud environments, Users
Everything is a Resources
Tenets of ZTA
Continual Threat-based decisions
Continually tweak the Access and authorization

Figure 9.1*: Zero Trust Architecture and its tenets*

ZTA is also defined using simpler principles, namely:

- **Never trust, always verify**: Much like the tenets defined above, no access to any resource is trusted by default (and hence Zero Trust). The validation (authentication) must occur every time for every request, and specific/role-appropriate authorization may be provided based on the context.
- **Least Privilege/need-to-have**: We have previously covered these principles, in the context of ZTA, the authorization must be as granular as possible so as to disallow any permission creep and risk to CIA of any information asset.
- **Continual verification**: ZT professes never to trust. Thus, even authenticated sessions/users/devices must continually be evaluated for being relevant, current, and not bringing any risk to any resource. For instance, the user's device connected to the VPN for several days on the trot may not have gotten the regular **endpoint detection and response** (**EDR**) upgrades or maybe even patch updates, thus the device posture may pose additional risk to the information asset. Such a connection must be re-evaluated and terminated based on the organization's *risk appetite* and policy. We covered risk management in *Chapter 2, About Managing Risks*.
- **Assume breach**: This principle is probably the most diametric change

from traditional security models. Cybercrime is on the rise. APT attacks continue to happen, and thus, it is possible that the organization's environment is already compromised/breached, even if the data or information may not yet have been taken from the environment. The ZTA encourages the security team to assume a breach has already occurred and then define/augment mitigating controls. This approach also reinforces the core ZT philosophy that the perimeter-based castle-and-moat approach is not useful when the Trojan horse (malicious code/attacker) is already inside the castle (i.e., organization's network).

Key pillars of Zero Trust

ZT framework applies its principles/tenets to the following pillars to validate their security requirements and conformance to organizational policies:

- **Devices**: Focuses on ensuring that only a device meets or exceeds the security requirements of the organization and is permitted to authenticate and be authorized to the resource. For instance, a personal device with a missing anti-malware update will be denied access.
- **Identities**: Focuses on verifying the identity for interactive logins, i.e., a human user, and for programmatic ones, such as a service account.
- **Network**: Focusses on ensuring up-to-date ACLs to restrict traffic within network zones (covered in *Chapter 6, Key Security Technologies*).
- **Application and cloud environments**: Focusses on security applications, their infrastructure, and, where applicable, cloud environments such as containers, workloads, microservices, and APIs.
- **Data**: Focusses on protecting the CIA of data in transit, and at rest.

Figure 9.2 illustrates the key aspects of ZTA:

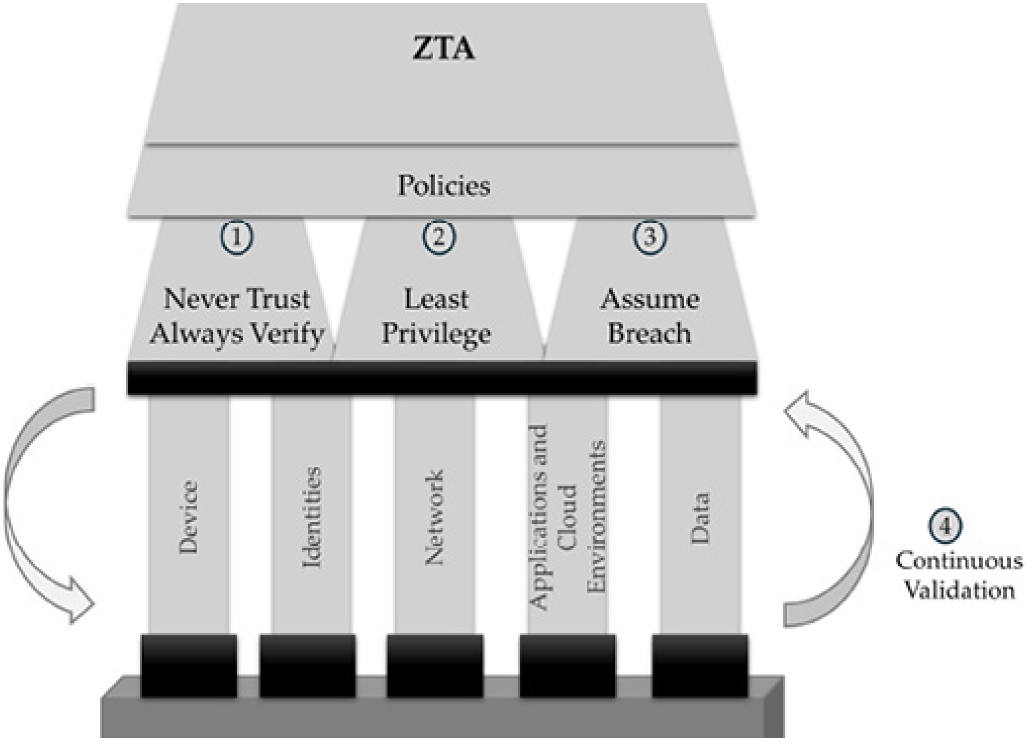


***Figure 9.2**: ZTA's principles and pillars*

Tip: A global consulting firm with 300K+ employees worldwide was hit by ransomware, prompting it to quickly choose and deploy an agent-based micro-segmentation solution to isolate, prevent lateral spread of future attacks, and create a resilient environment. They were able to roll out this technology with baseline and future-proof rapid response protective policies, and all that within a two-week timeline. The capability is designed to examine access continually. This improved their security posture and the board's confidence in the organization's ability to deal with ransomware effectively. This may be one of the several cases in point for the growing success of the ZT adaptation.

Applying Zero Trust

In today's digital world, a successful cyber attack is more a measure of when versus if. Implying there is a fair bit of likelihood that every organization will experience some successful cyber-attack. It is the robustness of the organization's controls that would be able to quickly detect such an event, because of its detective controls, and also manage and respond to such adverse events. In this light, **assuming breach** as a way of thinking about controls and access to information assets becomes a very strong reason to plan and implement a ZTA.

Zero Trust journey

It is important to remember that ZT is not a tool; it is a mindset and multiple processes, people, and technology components that need to come together. The following approach may be considered for applying ZT principles in an organization:

- **Defining the acceptable objectives from a ZT implementation**: Organizations must consciously make an informed choice on what they wish to achieve from the implementation of Zero Trust and what that would mean to the day-to-day work of the users. In the frenzy of implementing ZT, user experience may take a hit. For instance, having to authenticate session by session to the same corporate application may become cumbersome. The organization must define the period of intervals when such a validation should be done.
- **Gain management consent and sponsorship**: ZT is not just about the IT environment or cloud environment; it is also about the culture of working. The general human tendency to be averse to change will need some management support. Needless to say, ZT will also cost money.
- **Perform a gap analysis**: To determine the inventory, current business requirements, current issues, and investments needed for:
    - **Devices**: Details of devices used, such as models, types, and counts, their security configuration baselines, BIOS versions, etc., should be inventoried. Often, such information is available on the **configuration management database** (**CMDB**), but it is often not up-to-date. The mechanisms the organization uses to differentiate the devices it owns/manages from devices such as BYOD. Any known deviations in security baselines must also be cataloged. For instance, the field staff of sales and marketing may have been approved to use USB storage media.

      Devices include other computing environments such as servers, databases, load balancers, network security equipment like firewalls, routers, and so on.
    - **Identities**: The strategy of how various types of identities (covered in *Chapter 7, Identity and Access Management*) are used in the environment must be reviewed, and any threats must be identified.

For instance, if the IAM currently does not govern the use of service accounts, the ZT policies implementation will either be made to block all such access, causing business disruption, or will be defined to allow all such IDs, giving a loophole for the attacker to capitalize on. Both scenarios must not arise.

Additionally, relevant authorization and user activity accounting must also be examined.

  - **Network**: The network flow in the context of segmentation must be understood and documented, and any remedial plan for ZT must be prepared. For instance, if the server VLAN allows **Remote Desktop Protocol** (**RDP**) traffic from several regions without a check, malicious attackers can infiltrate the network. The request for resource access may not even get screened by the ZT policies.

    The north-south traffic and the east-west traffic (covered in *Chapter 6, Key Security Technologies*) should be evaluated for current relevance and anomalies that might indicate malicious activity. For instance, a user attempting to access an online code repository when the role does not demand it. Or a user frequently attempting to send local files to webmail but is blocked by a DLP tool.

  - **Application and cloud environments**: The business applications and the supporting infrastructure may either be on-prem, cloud, or both. Irrespective of that, the functionality needed by these business applications must adhere to the principles of least privilege. Managing user sessions and validating continual access is also required. Cloud workloads should be governed appropriately such that no misconfiguration happens. For instance, cloud storage must not be set to be publicly accessible by default.
  - **Data**: The organization may use data discovery tools to determine where it hosts data beyond known applications and servers. Oftentimes, sensitive data may be found on user devices.

- **Define a plan of action to prioritize and remediate gaps and the future state**: Risk management principles, covered in *Chapter 2, About Managing Risks*, should be applied to ensure appropriate time-bound decisions are made. For instance, if the management decides not to segment its server network for a year, as the migration to the cloud is pending, the ZT implementation will need to be adjusted accordingly. Monitoring practices already implemented may be focused on until such time.

The organization may choose to implement ZT principles for its identities first, implying there would be strong policy-driven authentication, authorization, and accounting for all identities. Adaptive, or context-driven MFA, may additionally be triggered, as elucidated earlier in the chapter.

It is also common for organizations to start by implementing Zero Trust for their network pillar as well. The core idea of this approach is to make the network location irrelevant and focus on who accesses what, when, and why. **Principle-of-least-privilege/need-to-know/need-to-have** are used to define the approach.

Organizations may draft a roadmap of their future state of some controls in ZT even though the current practices meet current organization requirements. For instance, the organization may choose to implement requirements of an involuntary standard (for instance, PCI-DSS ver.4) that may require server file monitoring to be enhanced. Similarly, it may choose to start implementing approved algorithms to protect data in the **post-quantum-cryptography** (**PQC**) world. We touched upon the PQC in *Chapter 6, Key Security Technologies*.

- **Manage and monitor key success parameters of ZT against agreed objectives:** ZT implementation is a journey, and each stage of its progress must be validated and checked against the agreed objectives. Using the plan-do-check-act cycle/continual improvements, feedback from current implementation, threat feeds, etc., may be used to continually refine the ZTA policies.

ZT implementation takes time, but the ZTA must be agile to remain aligned to the organization's business requirements.

Tip: Pokémon, the Tokyo, Japan-based media and entertainment company, has a large set of roaming users. It also uses a lot of external third-party partner organization specialists that connect to Pokémon's environment. Apart from the threat to its environment from the Internet, they also had to ensure the revenue-sensitive intellectual property rights (IPR) of its products, such as online games, are well protected. In their approach to ZT, the organization implemented a cloud-based product to inspect malicious DNS traffic and take appropriate

action. In doing so, they were able to secure their largely on-prem environment and yet reduce overheads around monitoring and investigating.

Zero Trust Network Architecture

Zero Trust Network Architecture (**ZTNA**) is a popular approach to start on ZT, especially with the advancement of network capabilities on the cloud (some of which were covered in *Chapter 8*, *Cloud Security*). The ZTNA is also called a **software-defined perimeter** (**SDP**). Here are some key aspects:

- It uses software to define policies and dynamically adjust based on context to permit access.
- Unlike traditional VPN-based remote access, which would generally allow the entire subnet to be available to the requester, the ZTNA makes it possible for only the least required resources in the request to be made available.
- Additionally, in the ZTNA implementation, every authorization request may prompt reauthentication or revalidation, such as an MFA prompt. For instance, in the ZTNA, all privileged users accessing a remote server for performing administrative tasks would authenticate to a ZTNA client and get authorization based on device context/device security posture.
- The physical location of the user, such as in the office or if connecting from remote sites, will not be relevant. A user within a corporate LAN would also need to authenticate and get authorization, much like a remote user with the same access permissions would.
- It is possible to limit access from certain geographies if the organization requires it. For instance, an organization in India may limit remote access to its data center or cloud control pane only when the requester originates from an India-based **internet service provider** (**ISP**).
- **ZTNA** allows for continual validation and even access monitoring to ensure the core principle of never trust, always verify is refined with not just identity, device, and application context, but also with threat intelligence. For instance, if the threat intelligence indicates an attack based on a particular vulnerability, the policy can be so defined that the traffic from those IPs and or insecure devices is blocked.

Tip: JAMF, a leading security configuration management tool for Apple devices, is used by scores of organizations. To maintain effective customer service and experience for their own workforce, they implemented a cloud-based identity solution as the primary means to control access. As their journey progressed, they coupled it with device validation (Okta Fastpass) and protection against identity/credential-based intelligence (Okta's ThreatInsight) to permit/deny access. Additionally, it examines the network IP to gain device, identity, and network context. Adaptation of these ZT pillars is a good example of the success of this ZT approach.

Some of the popular ZT technologies are:

- **Identities**: Microsoft Entra (formerly Azure AD), Okta single-sign-on, and Ping Identity.
- **Network**: **Zscaler Private Access** (**ZPA**), Google Beyond Corp, PaloAlto's Prisma Access.
- **Data**: Zscaler DLP, Forcepoint DLP.

Tip: LIXIL, a popular manufacturer with well-known brands such as GROHE and American Standard installed in thousands of restrooms the world over, had a complex ecosystem with disparate identities on around 600 applications. To streamline operations and to support a growing remote user base, LIXIL embarked on its ZTNA approach by using the Enterprise Application Access solution. This cloud-based solution from Akamai provided a seamless and secure access to the same set of applications using a connector and without changing any code on the applications. The same technology was configured to provide access to even the contractors without having to provide the traditional VPN like access.

Conclusion

In this chapter, we covered Zero Trust, its architecture, its principles, tenets, and pillars. We also covered some suggested aspects of the ZT journey that can be approached.

In the next chapter, we will explore core concepts of **cyber threat intelligence** (**CTI**) and some related aspects of detecting and managing threats.

Key takeaways

As we explored the concepts of ZT, the following were some of the key takeaways:

- The traditional security model, often referred to as the castle-and-the-moat model or **trust-but-verify** model, treats an authenticated user/device as largely trusted. Incidents after incidents in the world have shown that the model may have shortcomings if an attacker already gains access.
- Much like the Greek mythology of the Trojan horse, the ZT model professes a **never-trust–always-verify** approach.
- Using its seven tenets, the ZT model requires organizations to continually validate access and ensure principles such as the *least privilege* are adequately implemented.
- Given how cyber attacks are inevitable, it proposes measures to assume a breach.
- The ZT model requires a rethink to secure aspects, namely devices, identities, networks, applications, cloud environments, and data itself. Every aspect has been evolving rapidly with newer tools and offerings to enable organizations to adapt to ZT.
- The ZT model is an exciting and interesting possibility that can use principles of project management and risk management to progress towards it.
- The journey to ZT is fraught with several technical and process changes, but more importantly, a mindset change to accessing information systems. While it seems more restrictive, the positive aspect of the approach gives a fillip to the **anywhere work** of operations.
- Several organizations have made incremental progress in various aspects of ZT, especially ZTNA.

References

- **https://www.forrester.com/zero-trust/#business**
- **https://www.cisa.gov/zero-trust-maturity-model**
- **https://www.zscaler.com/blogs/product-insights/journey-zero-trust**
- **https://www.zscaler.com/customers/borgwarner**
- **https://www.akamai.com/resources/customer-story/global-consulting-firm**

- **https://help.okta.com/en-us/content/topics/security/threat-insight/about-threatinsight.htm**
- **https://www.akamai.com/resources/customer-story/pokemon**

Join our Discord space

Join our Discord workspace for latest updates, offers, tech happenings around the world, new releases, and sessions with the authors:

https://discord.bpbonline.com

CHAPTER 10

Threats and Exposure Management

Introduction

Advancements in technology have made it easier for cybercriminals to cause harm to organizations and even individuals, especially high-net-worth celebrities. Driven by motives such as financial gains, espionage, mischief, or even simply proving a point, cybercriminals continue to innovate. In recent times, cybercriminals even operate as organized groups with a specific modus operandi. The CISO and his/her team are required to be aware of these evolving developments and adequately protect their organization from such exposures.

Structure

The chapter covers the following topics:

- Foundations of threat and exposure management
- Threat intelligence
- Continuous threat exposure management

Objectives

By the end of this chapter, you will be able to understand threats, **cyber threat intelligence** (**CTI**), and how to use them effectively in protective and detective controls in the organization. We will also explore some of the tactics used by cyber attackers and how to take a risk-based approach to protect your information systems.

Foundations of threat and exposure management

The CISO and his/her team are required to keep an eye on the changing environment and prepare the organization for managing any untoward exposure from any such changes. For instance, cyber attackers have used cloud computing to launch email-based attacks on organizations (covered in *Chapter 7, Identity and Access Management*). Organizations would need to implement controls to protect themselves from exposure to such attacks.

In *Chapter 2, About Managing Risks*, we briefly introduced a **threat** as an agent that causes the risk to materialize. NIST has a few definitions in several of its publications. We will use the following two of those to understand the term **threat** better:

- *The potential source of an adverse event.*
- *Any circumstance or event with the potential to adversely impact organizational operations, organizational assets, individuals, other organizations, or the Nation through a system via unauthorized access, destruction, disclosure, modification of information, or denial of service.*

Vulnerabilities are exploited by a person or group of people called **threat actors,** malicious attackers, **threat agents,** or **adversaries**. The term **hacker** is also commonly used. Threat actors leverage technology extensively, especially cloud computing, to cause harm faster and wider. The threat actor could be internal or external to the organization and may have differing motives for their action, such as financial gain.

In *Chapter 3, Role of Standards and Controls*, we introduced **the lifecycle of controls** and the domain of **threat and vulnerability management** (**TVM**). A vulnerability is a weakness of an information system that a threat actor can exploit to cause an adverse event. These adverse events, in turn, can bring risks to the organization. We covered risks in detail in *Chapter 2, About Managing Risks*. Organizations implement a TVM process where such vulnerabilities are continually identified, prioritized, and remediated. Similarly, organizations define and deploy controls to detect behavior anomalies and take appropriate action. *Figure 10.1* illustrates the relationship between threat actors, threats, vulnerabilities, and risk:

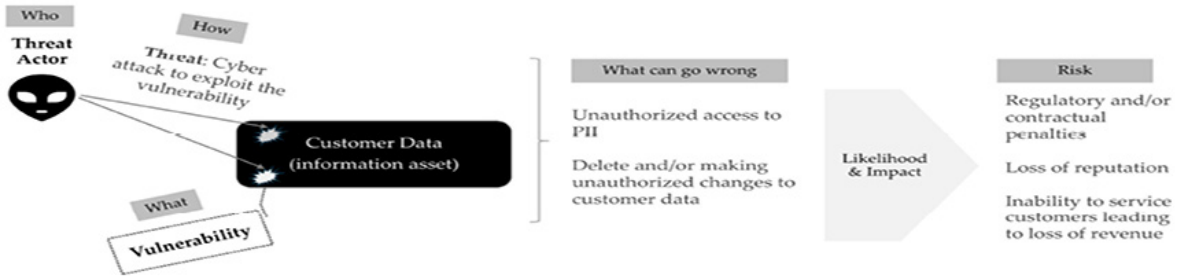

Who
Threat Actor
How
Threat: Cyber attack to exploit the vulnerability
Customer Data (information asset)
What
Vulnerability
What can go wrong
Unauthorized access to PII
Delete and/or making unauthorized changes to customer data
Likelihood & Impact
Risk
Regulatory and/or contractual penalties
Loss of reputation
Inability to service customers leading to loss of revenue

***Figure 10.1**: Threat actors, threats, vulnerabilities, and risks*

The process of gathering, processing, and analyzing information about threats and using it for reducing applicable risks via controls is **cyber threat intelligence** (**CTI**). The outcome of this process is actionable information, also called CTI. In other words, CTI is both a process and an outcome.

Tip: CTI and threat intelligence (TI) are used interchangeably. However, TI may additionally imply enrichment of intelligence with factors other than just cyber. For instance, a geo-political event that indicates enhanced government-backed activities may be relevant intelligence to monitor cyber activity from that region to an organization's information assets.

As per NIST, *Threat intelligence is threat information that has been aggregated, transformed, analyzed, interpreted, or enriched to provide the necessary context for decision-making processes.*

CTI and TVM are related concepts. CTI provides actionable insights into the existing and emerging threats that the TVM process can use to manage vulnerabilities and threats to it. *Figure 10.2* illustrates some of the suggested controls for detecting and mitigating the risks from threat actors, threats, and vulnerabilities:

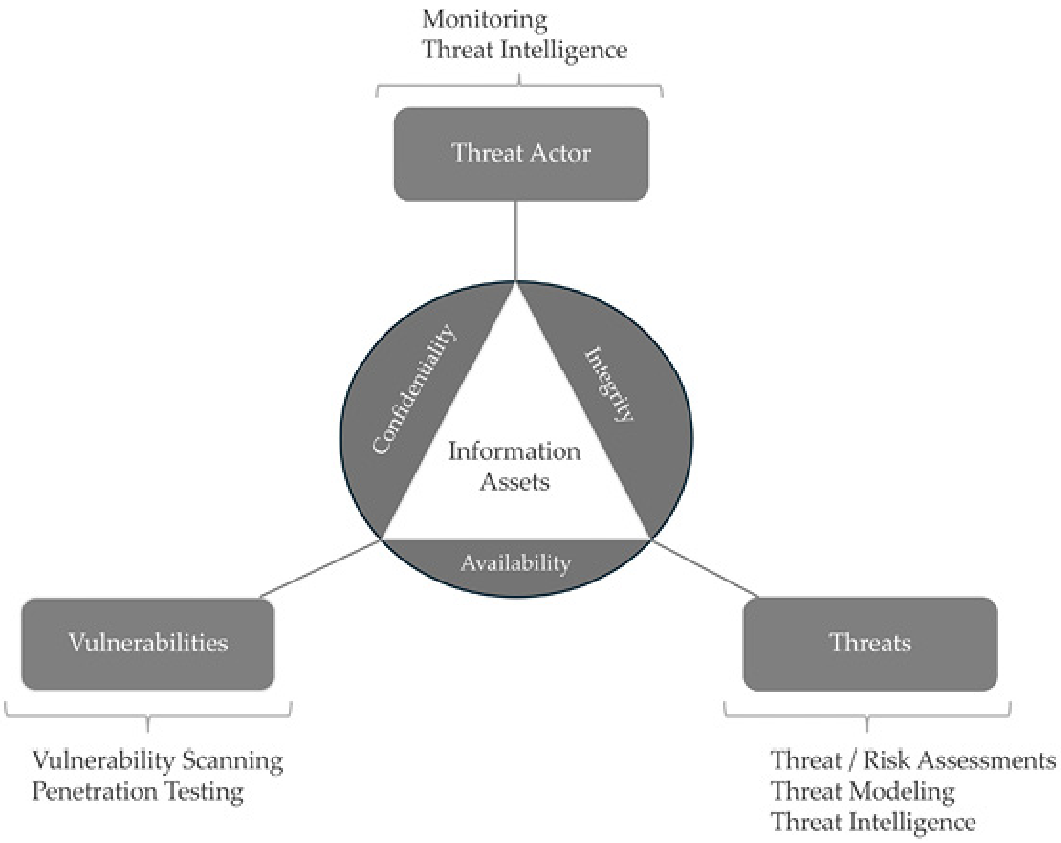


***Figure 10.2**: Some of the controls for managing threats*

Managing threats

To understand threats better, a concept called **threat modeling** is used to chalk out what could go wrong, how it may already be thwarted, and how and what additional controls might be needed. Threat modeling is closely connected to risk management practices, as the choice of controls and their success criteria should be a risk-based decision. NIST SP 800-154 elucidates the principles of a data-centric approach to threat modeling.

Tip: Microsoft provides a threat modeling tool on its website for developers.

For instance, if a pharmaceutical organization is keen to protect the **intellectual property rights** (**IPR**) of its patented but upcoming new drug/medicine, it can embark on a threat modeling exercise to cover aspects in *Table 10.1*:

What to worry about?	Are there some controls as of now?
How is the formulation stored and where?	Formulation is stored in a strong room with extremely limited access in the office of the CEO.
Who has access to it?	CEO, CEO’s office staff, scientists and pharmacologists working

	on the medicine. There is no camera overlooking the door to CEO’s office.
How can that formulation be accessed and copied?	By using the key to the lock (kept in CEO’s drawer) and copied using the photocopier on same floor.
Can the manufacturing happen while limiting the access to actual formulation?	Yes, by having different teams mix different components and a different team performing the final formulation and packaging.

***Table 10.1**: Example of a simple threat model*

The above model is a crude way to examine what can go wrong, where, and what needs to be additionally done. Organizations apply the same logic to perform threat modeling for their data stores and for various attack scenarios. For instance, an organization may perform threat modeling on the applications it has hosted on a PaaS provider.

About vulnerabilities

In today’s digital world, everything is a piece of code, and those may have weaknesses/vulnerabilities for the following reasons:

- Design flaws such as:
    - Improper error handling.
    - Improper user input validation.
    - Not encrypting username/password and/or data during transmission.
    - Insecure authentication, authorization, or access management.
- Configuration weakness such as:
    - Using deprecated protocols such as TLS 1.0 (covered in *Chapter 6, Key Security Technologies*).
    - Enabling access to information assets without proper controls. For instance, provisioning an S3 cloud bucket for public access.
    - Not enabling strong password requirements.
    - Allowing any-any on the firewall.

Vulnerabilities can also be of the physical access kind; for the purpose of the book, we will focus on software vulnerabilities only. Hardware devices like firewalls, routers, servers, **Closed Circuit TV** (**CCTV**) systems, or **building management systems** (**BMS**) all have software on them.

Cyber researchers examine/evaluate various technologies to determine their weaknesses that can be compromised. They may stumble upon a vulnerability while evaluating or testing features of the software as well. Notifying the right stakeholders or members of the public of a software vulnerability is **vulnerability disclosure**. These researchers, also called **white hat hackers**, would first disclose the existence of a vulnerability to the respective publishers of the software, such as *Microsoft, RedHat, Apache, Apple, Google, VMware, Cisco, Ivanti,* and *Palo Alto Networks*. This way of **responsible disclosure** allows the vendor(s) to develop fixes for those weaknesses/bugs, test them, and make the fix available for their consumers/public. These fixes are commonly called **patches**. The intent of researchers to delay public disclosure is to ensure bad actors/threat actors do not misuse and bring harm to scores of organizations.

Info: Microsoft operating systems and their software are probably the most used worldwide. Every second Tuesday of the month, called Patch Tuesday, Microsoft releases the list of vulnerabilities in its software and its patches.

To take advantage of the vulnerability, either a method is used, or an additional piece of code (called an **exploits**) is written. For instance, if the e-commerce portal has not appropriately validated inputs in the order form, the researcher/attacker may be able to pass parameters in the HTTP request to change the quantity, price per unit, bill value, and any such impacting value. Similarly, if the healthcare portal does not do session validation, the attacker can write a script to gain unauthorized access to patient medication, diagnostic results, or any PHI. They may also inject malware to make the devices dysfunctional and disrupt healthcare services.

Zero-day vulnerabilities (**0-day**) are those vulnerabilities that may have existed in software since the beginning but have not come to notice until a cyber researcher finds them. The software vendor may not have a means to quickly fix those. A lot of vulnerabilities identified today are 0-day ones.

Cyber attackers/threat actors with malicious intent are called **black hat hackers,** who do not generally go through responsible disclosure. Their motivation is often money, and thus, they may sell the vulnerability information to cybercriminals. In most countries, hacking any computer system is punishable.

Tip: Many organizations run bug-bounty programs to encourage such researchers to identify vulnerabilities in their software and responsibly notify them. The researchers get paid a bounty when a vulnerability is successfully proven to exist.

Tip: For a threat to be relevant/applicable, one should examine the probable intent of a threat actor, a weakness/vulnerability (an opportunity) for the adversary to use, and the capability (such as an exploit) to cause harm. This philosophy enables a risk-based approach to be followed.

All disclosed vulnerabilities are cataloged and are available in public forums in databases like the **National Vulnerability Database** (**NVD**) with a unique **Common Vulnerability and Exposure** (**CVE**) number. The CVE number is issued by **MITRE** and/or **CVE Numbering Authority** (**CNA**) and is the same across all such vulnerability databases. The vulnerability is also assigned a severity rating on a 10-point scale using a system called the **Common Vulnerability Scoring System** (**CVSS**). 10 is the most severe, and 1 is the least severe. The 0-day vulnerability may not have a CVE number initially. Some vulnerabilities are easily exploited or are quite commonly exploited. Such a list, called **Known Exploitable Vulnerability** (**KEV**), helps organizations prioritize their fix/patches. One of the most authoritative sources of vulnerabilities is the US Government's **Cybersecurity and Infrastructure Security Agency** (**CISA**). *Figure 10.3* illustrates a typical vulnerability disclosure process and timeline:

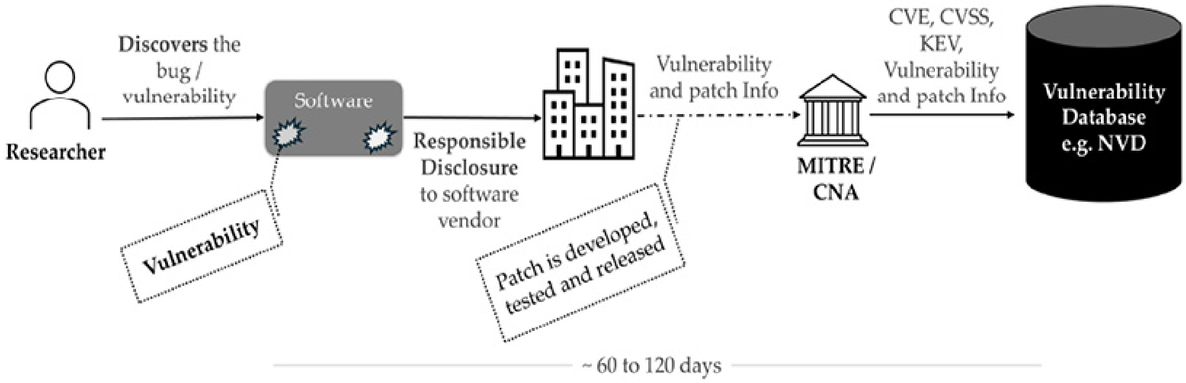

Researcher
Discovers the bug / vulnerability
Software
Vulnerability
Responsible Disclosure to software vendor
Vulnerability and patch Info
Patch is developed, tested and released
MITRE / CNA
CVE, CVSS, KEV, Vulnerability and patch Info
Vulnerability Database e.g. NVD
~ 60 to 120 days

***Figure 10.3**: Vulnerability disclosure*

Vulnerability management process

The **vulnerability management** (**VM**) process may be considered a cyclical and phased process, as shown in *Figure 10.4*:

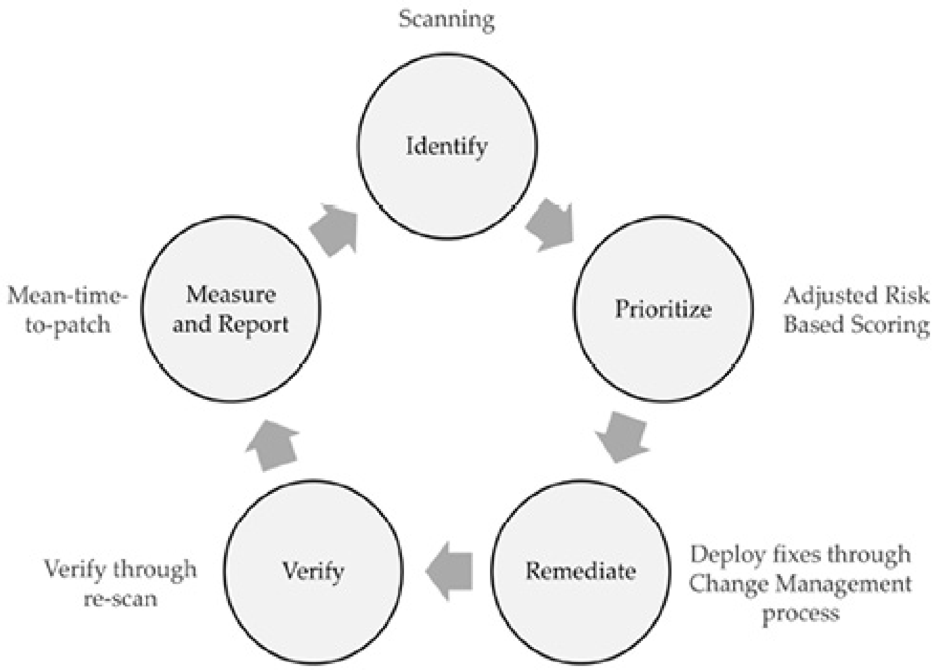


***Figure 10.4**: Phases of VM*

The details are as follows:

- **Identify**: Organizations must use specialized software called **vulnerability scanners**, also referred to as **VM scanners**, to periodically identify the vulnerabilities within their environment. The scanners essentially determine the version of the software and compare it with feeds or sources such as the NVD to determine if the version has an active and/or exploitable vulnerability. The scanners may also determine vulnerabilities based on missing configuration or the existence of risky. For instance, VM scanners can determine which version of the *Apache Tomcat* web service is running/installed, and if there are any known vulnerabilities. It can also determine if ports such as SMTP (25), Telnet (23), or RDP (3389) are open. Scans can be run:
    - In an authenticated fashion:
        - Use the VM scanner agent on each compatible device.
        - And/or run from the VLAN of the network as a network scanner.
    - Unauthenticated scans where the scanner attempts to find any vulnerability it can. This **reconnaissance** is useful to discover unknown assets, their vulnerabilities, and also determine the exposure

that an attacker might also see.

Tip: The organization's architecture and risk appetite are some of the deciding factors in choosing a VM strategy that includes what to scan, when, and how. The VM scanner may be deployed on-premises in data centers or may leverage SaaS solutions.

In this book, we will focus on the capability of VM scanners to detect vulnerabilities and create a program to govern them.

- **Prioritize**: Typically, every organization will have innumerable vulnerabilities, but neither are all exploitable nor is it practical to patch everything. Therefore, a strategy needs to be developed to take a risk-based approach to help prioritize the remediation/patching. The CVSS rating can be used to help prioritize. Additionally, an adjusted risk-based risk scoring method may be used. For instance, CVE-2024-56337, a vulnerability on Apache Tomcat, may be found on a server that is exposed to the internet and also on an internal test/Dev server. The external-facing asset must be assigned a higher asset criticality weightage than the internal one. Most modern-day VM scanners use CVSS rating, assess the context of the asset, allow asset criticality to be specified, and compute a vulnerability risk rating. This vulnerability risk rating is more useful in prioritizing assets.
- **Remediate**: Organizations would prefer to test the applicable patches prior to deployment to avoid any disruption. They do so on test environments and then observe. Once the list of prioritized assets and their vulnerabilities is available, they can start deploying patches. Several organizations still prefer a patching cycle based on the number of days taken to patch, versus focusing on risk. In an adjusted risk score-based approach, organizations would attempt to maintain the risk score of an asset below the risk appetite. Remediation on endpoint computers may be easier than on servers or network equipment, as patches sometimes need a reboot. In case a CVE exists without patches and/or fixes, the remediation steps would involve implementing compensating controls. For instance, the organization may restrict access to the Apache Tomcat external server from the internet or allow only limited known sources to connect. While it may appear to be disrupting the business, it may be a necessary call to ensure the organization does not face cyber-attacks that may use the vulnerabilities.

Tip: Given the rise of 0-day and the critical vulnerabilities, organizations also have an out-of-band expedited patching process. NIST recommends that applicable critical severity patches be deployed within 7 days of notification.

Tip: Changes to the IT environment, including patching, should ideally follow the organization's formal change review and approval process before deployment.

- **Verify**: This typically implies rescanning the environment to check if all the applicable patches were successfully deployed and the environment is no longer vulnerable to the chosen CVEs.
- **Measure and report**: The success of the VM program is in early detection, and ensure the remediation is as fast as possible, yet as least disruptive to the business as possible. The progress of the VM should be shared with the right stakeholders in IT, information security, and even business/functional teams. Some of the interesting measures could be:
  - Number of hosts in scope but not scanned.
  - Number of open vulnerabilities and their age.
  - Number of **end of life** (**EOL**)/**end of support** (**EOS**) environments.
  - Time taken to patch in days.

  We will cover metrics later in the book.

Info: As per CVE.org, there were over 40,000 new CVEs declared just in 2024, an increase of ov 38% from 2023.

VM scanners have introduced the capability to scan even the AI environments, including the **large language models** (**LLMs**). Much like the NVD, a dedicated **AI Vulnerability Database** (**AIVD**) is being maintained. It catalogs vulnerability, its severity, and suggested remediation.

Tip: Garak is an open-source tool to scan LLMs. There are other similar scanners, too.

Application security testing

The digital environment uses applications, APIs, containers, and so on. As part of determining weaknesses in applications, a formal method to test such weaknesses is needed. These tests should be included at the design stage of the application and during coding. Most software development environments do provide the capability of scanning the code for vulnerability/bad coding practice as the code is being developed. Even code generated using AI can produce code free from common vulnerabilities. For instance, hard-coded passwords in clear text in the code will be picked up by scanners. Similarly, it can detect coding errors like weak input handling or buffer overflow management.

The process to identify and analyze weaknesses in the source code of an application at any stage of the software development lifecycle is called **Static Application Security Testing** (**SAST**). The core idea for using SAST is to ensure the final application does not have coding and simple configuration-related vulnerabilities. Several tools are available for SAST, even integrated with **continuous integration/continuous delivery or deployment** (**CI/CD**) agile development practices, where validated code can be deployed frequently and reliably. Here, reliability from a security standpoint is achieved due to positive outcomes from an SAST scan on the application. If a code does not pass the minimum threshold of a SAST scan and/or does not meet the minimum security requirements, it will not get compiled/released for production.

SAST tools can evaluate even APIs and binaries for similar errors. SAST scanners are adept at understanding and determining code weaknesses to avoid the **OWASP top risks** (covered in *Chapter 5, Security and Privacy by Design*).

Applications can and should be tested at runtime as well. The **Dynamic Application Security Testing** (**DAST**) tools automate several tests, like SQL injection and cross-site scripting, to be run and determine if the application securely handles those appropriately. **OWASP's top 10 risks** have not had a significant change in the methods attackers use; these risks are commonly tested by a DAST tool. DAST is a black box testing that simulates attacks by a user. It can also test memory usage and encryption mechanisms. DAST is typically performed on an application already released or just ready to be released for production. Sometimes DAST is a contractual requirement from customers prior to the deployment. Issues/vulnerabilities that are identified should follow a risk-based remediation approach. The learnings from these scans and tests are fed back into the application development teams. Relevant changes should also be made to code repositories, such as GitHub, so that the future use of the code does not have those vulnerabilities. Using AI, the capabilities of DAST, the things it can test, and how to reduce false positives are becoming more and more mature. The leverage of AI in DAST may prove to be efficient and increase the velocity of the release of applications to the market.

Most mature software development organizations combine SAST and DAST to be deployed as a strong assurance mechanism for the security quality of their code and product. Several security certifications, covered in *Chapter 3, Role of Standards and Controls*, require SAST/DAST to be deployed by organizations.

Info: Some of the popular and effective SAST tools include Checkmarx, SonarQube (open source), Fortify Source Code Analyzer, and GitLab. Similarly, DAST ones are: Burp Suite, OWASP ZAP (open source), Acunetix, and Rapid 7.

Penetration testing

Penetrating testing (**PT**) is a controlled and formally approved intrusive test of the efficacy of layers of controls on an information asset/group of information assets. The PT's scope can be an entire server network, just an application, or even just a smaller ecosystem such as wireless access points. The exercise is highly skilled, is usually manual, and takes time and effort.

The core objective of a PT is to determine as many gaps as possible and see if an attacker can break through the defenses and use some vulnerabilities to cause harm/compromise the CIA. For instance, check if the attacker can gain access to data or information assets, exfiltrate data, bring down a server, and/or compromise information systems in any manner. Generally, an internal penetration test is conducted by specialists called **white hat hackers,** and they are usually given a mandate to prove if the information system can be compromised. They usually do not proceed to inflict actual harm (such as deleting data/sending data out). This is also a test of how efficiently and quickly the monitoring team can notice such an attack and take evasive/corrective action. We briefly covered the logging in *Chapter 3*, *Role of Standards and Controls,* as a detective control.

Mature organizations may empanel specialist penetration testers/organizations to evaluate their security posture. Such a need may be mandated by customer contracts as well.

Tip: AI will likely make more inroads in PT exercises. The LLMs may get better trained to suitably generate context-based scripts to test the information asset.

PT is a point-in-time exercise and will not be able to reflect the changing environment of vulnerabilities and the impact those have on the information assets. For instance, if an application that used the popular code library Log4J version 2.15 underwent a PT prior to December 2021, the security issue would not have shown. However, with the CVE-2021-44228 for Log4j announced in Dec 2021, the attack surface changed dramatically. The same application became vulnerable.

The term **vulnerability assessment penetration testing** (**VAPT**) is commonly used to denote the process of detecting vulnerabilities and performing penetration tests as a composite program. The approach depends much on the organization's security objectives and approach.

Figure 10.5 enumerates the broad steps that may be considered for a PT exercise:

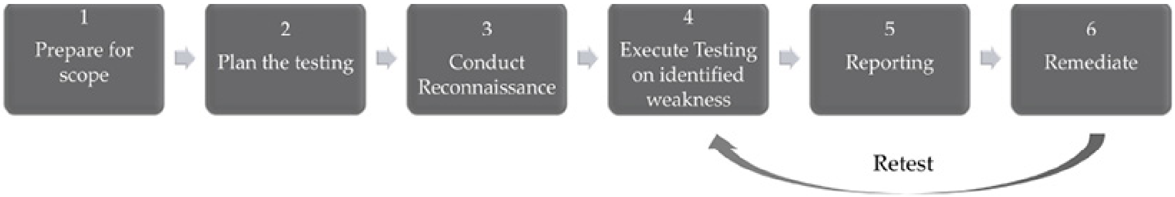


***Figure 10.5**: Steps for a PT exercise*

Red, blue, and purple teaming

Attackers just need to find one weakness and use that to bring harm to the organization. Once the attacker can perform a successful recon and break in, they can move around the network to compromise several information assets. The organization's information security team would ideally tackle this intrusion as fast as possible. A lot depends on what is logged, how, and the rules for altering.

As an analogy, a burglar who breaks into the house using a window (the vulnerability) will attempt to take any/all valuables (the risk—loss of asset) he can in any room of the house. Similarly, the guard must focus on the weakness that matters to thwart such a burglary.

Taking a cue from military practices, organizations use skilled teams to simulate attack and defense to test their security posture and readiness.

Red team is a specialized penetration testing team that aims to evaluate the organization's entire information ecosystem. This team's focus is to emulate what an attacker might do/be able to do after gaining access to the organization or to any of its information assets. It aims to determine the possible harm that an attacker can cause once a weakness has been identified and compromised. The core objective of this team and their intrusion exercise is to build better defenses and help improve the assurance of the controls.

Blue team is focused on detecting and thwarting the attempts of bad guys/threat actors (such as a red team). The alacrity of their detection and response is key to controlling cyber-attacks.

Purple team brings collaboration between the attack and defense sides of the equation, with each learning and sharing information about tactics. The premise is that the attackers and defenders think differently, and by learning from each other, each side can get better and thus improve the organization's security posture.

Threat intelligence

The knowledge of what the enemy can do and is currently doing is of immense value in warfare or even in times of peace. This knowledge/information can be assimilated and analyzed to drive actionable insights to create intelligence (used as a noun).

Info: India's Research and Analysis Wing (R&AW), the USA's Central Investigation Agency (CIA), Russia's Foreign Intelligence Service (SVR), and Israel's Mossad are some examples of intelligence agencies. The members of these specialized units are also called spies and are tasked with gaining actionable insights into other nations.

In the digital world, organizations can also benefit from awareness of threat data, threat information, and threat intelligence on what a cyber attacker is up to and prepare for defensive and offensive tactics. Threat intelligence has unique characteristics as it is an actionable piece of information, a process, a mindset to investigate, and an output of threat analysis itself. Threat intelligence covers the following aspects:

- **Tactic, technique, and procedure (TTP)** is a grouping of the ways the adversary uses to meet their intrusion objectives.
    - Tactic is what is to be done.
    - Technique is the choice of which method to use.
    - Procedure is the way it is to be done, i.e., step-by-step actions to take.
- **Tradecraft**: This can be thought of as the signature style of the adversary in an intrusion. It includes the methods, infrastructure (such as a particular cloud technology), and capabilities (such as password spraying attacks). In other words, it is the m*odus operandi*. It includes the **techniques** and the required **capability** and **infrastructure**. A Tradecraft is usually adversary-specific and thus is a valuable clue for a potential source/identity of an attacker. However, such **attribution**, i.e., conclusive declaration of the identity of the adversary or the **adversary group** (**AG**), cannot be made just on tradecraft.

A CTI program can be viewed from two aspects:

- **To enhance controls based on threat information**: To use the threat data, information, and intelligence available to spruce up the applicable security controls. The CTI process primarily aims to understand the threat-actor TTPs and determine the relevant gaps in the organization. This effort is proactive in nature. The process to determine the existence of intrusions provided by threat intel is called **threat hunting**. Threat hunting involves the use of pattern-matching rules called **Yet Another Recursive Algorithm** (**YARA**) rules. Using threat hunting, organizations can determine if an attack has taken place or is taking place based on the TTP. *Figure 10.6* illustrates how threat intelligence may be derived from threat data/information and applied to an organizational context to determine if any attacker activity can be identified:

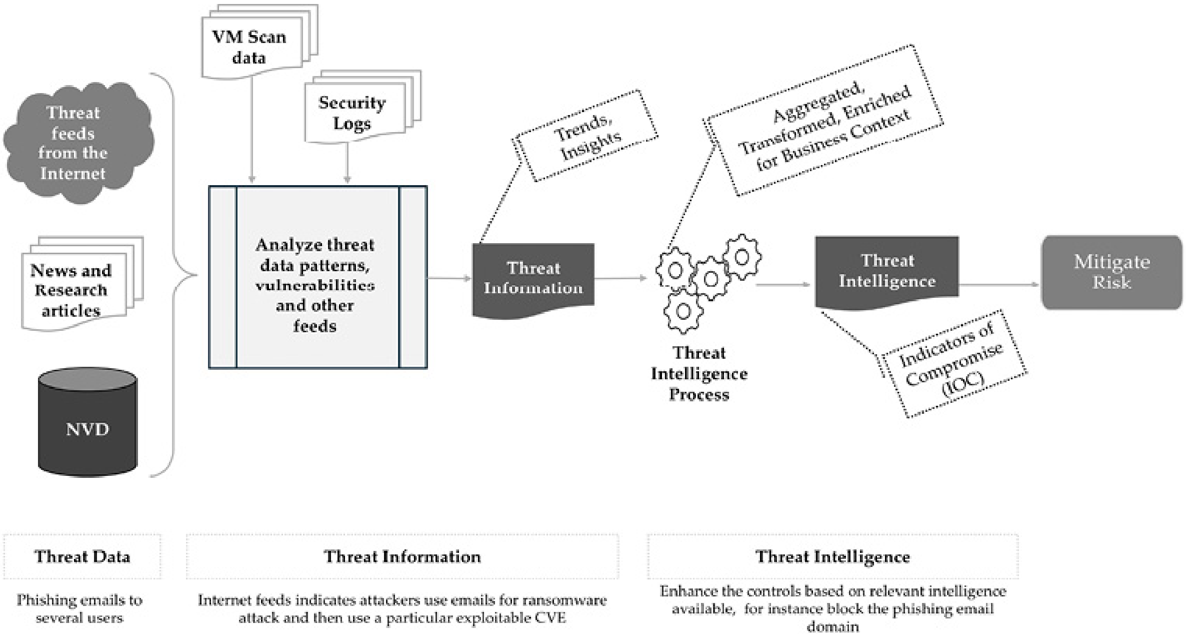

VM Scan data
Security Logs
Threat feeds from the Internet
News and Research articles
NVD
Analyze threat data patterns, vulnerabilities and other feeds
Trends, Insights
Threat Information
Aggregated, Transformed, Enriched for Business Context
Threat Intelligence Process
Threat Intelligence
Indicators of Compromise (IOC)
Mitigate Risk
Threat Data
Phishing emails to several users
Threat Information
Internet feeds indicates attackers use emails for ransomware attack and then use a particular exploitable CVE
Threat Intelligence
Enhance the controls based on relevant intelligence available, for instance block the phishing email domain

Figure 10.6*: Threat data, information, and linking intelligence*

- **To determine if an attack has happened and its extent**: This CTI lifecycle starts as a reaction to a trigger of an anomalous log, a suspicion or intelligence of a compromise, or the CISO's/management's direction. This could also be triggered by the attacker notifying the organization of a successful attack/intrusion. The organization's management would require validation of any such claim. This is called the **intelligence requirement** (**IR**). When a cyber-attack happens, the organization's foremost step is to contain the incident, and then spruce up controls, notify regulators and customers if needed, restore business functioning securely, and maybe even notify cyber insurance. The lifecycle and key aspects of each phase of such a CTI process are illustrated in *Figure 10.7:*

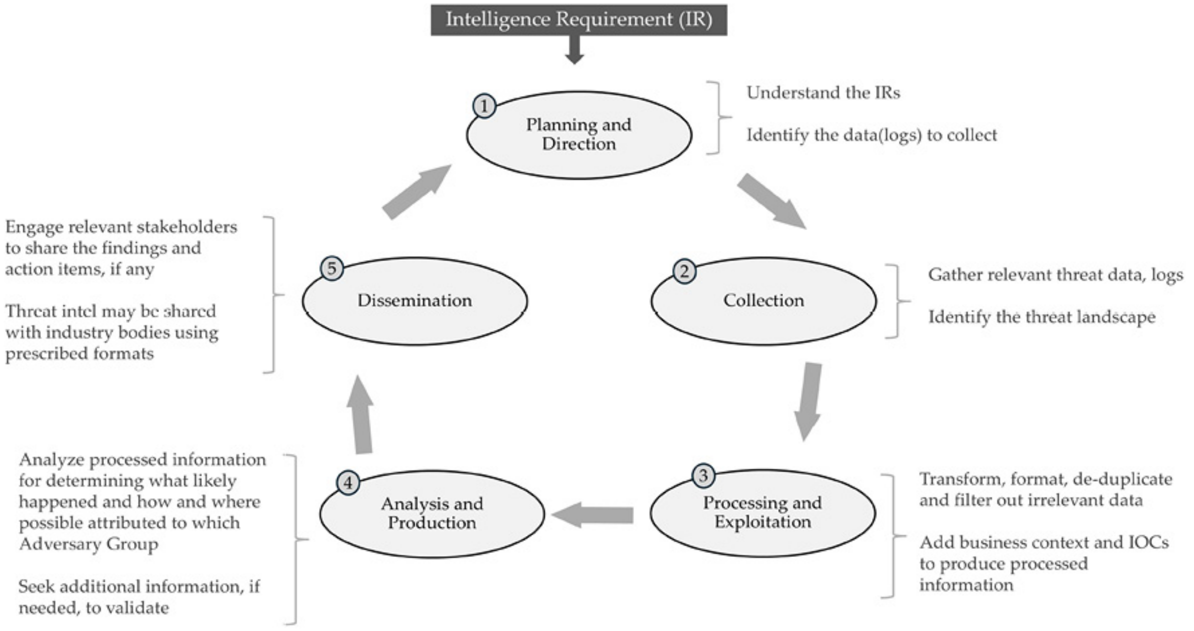

Intelligence Requirement (IR)
1
Planning and Direction
Understand the IRs
Identify the data(logs) to collect
2
Collection
Gather relevant threat data, logs
Identify the threat landscape
3
Processing and Exploitation
Transform, format, de-duplicate and filter out irrelevant data
Add business context and IOCs to produce processed information
4
Analysis and Production
Analyze processed information for determining what likely happened and how and where possible attributed to which Adversary Group
Seek additional information, if needed, to validate
5
Dissemination
Engage relevant stakeholders to share the findings and action items, if any
Threat intel may be shared with industry bodies using prescribed formats

***Figure 10.7**: Lifecycle of CTI*

A CTI analyst often **pivots** to a line of investigation based on the data/logs presented. This helps the analyst validate a hypothesis.

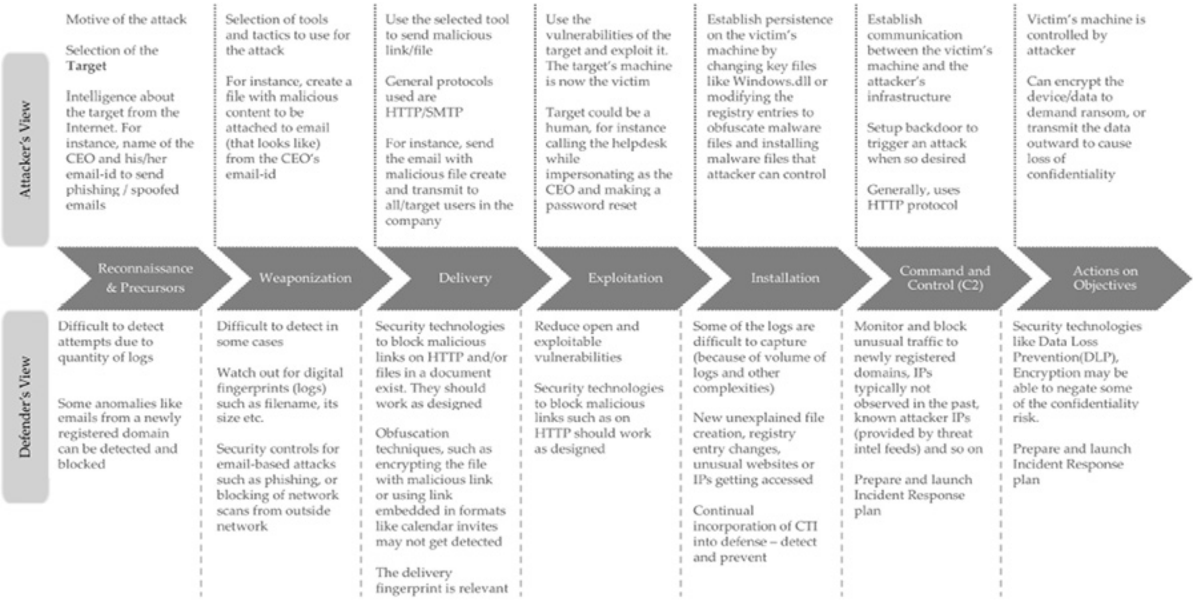
Tip: A threat analyst must ensure a structured and unbiased analysis based on facts to determine the outcomes. Pre-conceived notions and attempts to prove only what they believe to be true introduce confirmation bias and become unreliable in CTI.

There are several formal industry standards for sharing threat information. Two of the most popular and widely used are **Structured Threat Information eXpression** (**STIX**) and **Trusted Automated eXchange of Indicator Information** (**TAXII**). There are commercial and non-commercial services that allow the exchange of threat information between organizations called **threat intel platforms** (**TIPs**). The **computer emergency response teams** (**CERT**), a nodal agency in most countries, actively share threat intelligence with each other.

Tip: Commercial organizations generally do not share intelligence as commonly as one might expect for the fear of exposing their threats; however, several informal forums exist where such an exchange happens.

Attack kill chain

A structured and deterministic way to understand the attack in its various stages is through a concept called a **kill chain**. This model was introduced by three employees of *Lockheed Martin*, a large US defense supplier. Their model attempts to model the steps the attacker would have to take to intrude on a network information asset. The concept of Zero Trust basically does away with the concept of a traditional network, the kill chain model of CTI is still relevant to understanding the intrusion or attempts of intrusion better. This knowledge can then be used to enhance the security controls if needed. It is important to note that the steps are conceptual in nature; the attacker may execute several of the phases in tandem. *Figure 10.8* illustrates the Kill chain model and key highlights of each phase:



	Reconnaissance & Precursors	Weaponization	Delivery	Exploitation	Installation	Command and Control (C2)	Actions on Objectives
Attacker's View	Motive of the attack Selection of the **Target** Intelligence about the target from the Internet. For instance, name of the CEO and his/her email-id to send phishing / spoofed emails	Selection of tools and tactics to use for the attack For instance, create a file with malicious content to be attached to email (that looks like) from the CEO's email-id	Use the selected tool to send malicious link/file General protocols used are HTTP/SMTP For instance, send the email with malicious file create and transmit to all/target users in the company	Use the vulnerabilities of the target and exploit it. The target's machine is now the victim Target could be a human, for instance calling the helpdesk while impersonating as the CEO and making a password reset	Establish persistence on the victim's machine by changing key files like Windows.dll or modifying the registry entries to obfuscate malware files and installing malware files that attacker can control	Establish communication between the victim's machine and the attacker's infrastructure Setup backdoor to trigger an attack when so desired Generally, uses HTTP protocol	Victim's machine is controlled by attacker Can encrypt the device/data to demand ransom, or transmit the data outward to cause loss of confidentiality
Defender's View	Difficult to detect attempts due to quantity of logs Some anomalies like emails from a newly registered domain can be detected and blocked	Difficult to detect in some cases Watch out for digital fingerprints (logs) such as filename, its size etc. Security controls for email-based attacks such as phishing, or blocking of network scans from outside network	Security technologies to block malicious links on HTTP and/or files in a document exist. They should work as designed Obfuscation techniques, such as encrypting the file with malicious link or using link embedded in formats like calendar invites may not get detected The delivery fingerprint is relevant	Reduce open and exploitable vulnerabilities Security technologies to block malicious links such as on HTTP should work as designed	Some of the logs are difficult to capture (because of volume of logs and other complexities) New unexplained file creation, registry entry changes, unusual websites or IPs getting accessed Continual incorporation of CTI into defense – detect and prevent	Monitor and block unusual traffic to newly registered domains, IPs typically not observed in the past, known attacker IPs (provided by threat intel feeds) and so on Prepare and launch Incident Response plan	Security technologies like Data Loss Prevention(DLP), Encryption may be able to negate some of the confidentiality risk. Prepare and launch Incident Response plan

***Figure 10.8**: Kill Chain stages from the eyes of attackers and defenders*

Once the attacker gains control of the victim's computer, he will attempt to move to other information assets. This is called **lateral movement** and is a common tactic used by attackers to expand their footprint and inflict damage on as many information assets as they can. Refer back to the burglar analogy mentioned earlier.

MITRE, one of the world's leading technology and research and development companies, especially on topics related to cyber security, maintains and publishes:

- **MITRE ATT&CK**: A global knowledgebase of TTP that is useful for the **blue team** and **red team** in their exercises. The details of the tactics and the techniques used by attackers are an important piece of intelligence. In *Chapter 3, Role of Standards and Controls,* we referred to MITRE as a key input to the selection of controls. MITRE ATT&CK enumerates real attacks that happen and uses them in the knowledgebase. Therefore, this knowledgebase may be used to prioritize the top risks that an organization must focus on.
- **MITRE D4FEND**: It is a great matrix to understand the suggested countermeasures to implement against attacker TTPs. For instance, an attacker may manipulate the access token (MITRE ATT&CK T1134) to make it appear as a legitimate token to a malicious process. Using MITRE D3FEND, we can choose to deploy the countermeasure called **token binding** that binds the issued token only for the specific purpose. Any additional copy of the token would not work.

Continuous threat exposure management

In *Chapter 6, Key Security Technologies,* we covered some cyber-attacks that made their mark in history, some of the motives of attackers, and aspects like points of exit from a network (**egress**) and points of entry to a network (**ingress**). Any information asset that is exposed to any threat is part of an **attack surface** that an attacker can exploit. To take an analogy, in harsh winters of the upper Himalayan region, homes would ensure windows are shut and maybe one or two doors to enter and exit the house. This allows heat within to remain trapped and outside cold to remain out. In the same fashion, the organization would attempt to keep its attack surface as limited as possible. Organizations focus on **External Attack Surface Management** (**EASM**), i.e., reducing the number of assets exposed to the Internet, such as its website, mail servers, VPN gateways, etc. However, in today's business environment, organizations will continue to increase their digital presence. Therefore, the security objective should be to reduce the duration of exposure by quickly and efficiently remediating the relevant vulnerabilities.

Continuous Threat Exposure Management (**CTEM**), a term coined by *Gartner*, can be considered a proactive and continual means to detect and reduce the attack surface and/or the time an information asset might be exposed to vulnerabilities. CTEM program framework aims to determine the:

- Weaknesses of the information assets, external and internal.
- The extent to which that weakness can be exploited.
- Existing and emerging threats to the information assets.
- Prioritization of the remediation of such weaknesses as quickly as reasonably possible.

Gartner proposed the CTEM framework as a cyclic/continual five-stage process, namely:

1. **Scoping**: To define the scope of assets and the types of threats to be managed. For instance, an e-commerce company may choose to focus only on its merchant interface, customer shopping cart, and payment gateway for year one and exclude the SaaS services it uses for managing its code until year 2.
2. **Discovery**: Set up a mechanism to **increase the visibility** of the organization's digital footprint, i.e., the attack surface and its vulnerabilities.
3. **Prioritization**: Prioritize the **remediation** using principles of risk management for the in-scope assets based on their criticality. For instance, a healthcare organization would most likely ensure the vulnerable patient care systems are patched earlier than its building management system application.
4. **Validation**: Determine how the threat/attack may materialize and how soon the organization can respond effectively should the risk arise (**incident response**). For instance, a bank handling customers' financial transactions would likely ensure any patch to its core banking system is planned and executed in emergency mode if needed, while engaging relevant stakeholders, such as the **Chief Operating Officer** (**COO**), to manage any risk to CIA that may have materialized.
5. **Mobilization**: Ensure that adequately empowered cross-functional teams can move fast and remediate quickly for any exploitable vulnerability. For instance, in the banking example above, the COO and other stakeholders in IT should be aware that emergency measures may be taken by information security and IT teams without needing additional approvals as long as the CTEM program has established its relevance and prioritized the fix.

External scoring: Mature security-focused organizations use external parties (usually SaaS) to continually scan their external environment and score them for security performance and benchmarking. This provides a reasonably good assurance for customers in their third-party risk management program. Some of the common external non-intrusive scans performed are for open ports, vulnerabilities on external-facing servers, intelligence on the security of email systems, look-alike domain names, and the status of SSL certificates, etc.

The objectives of the CTEM include:

- Increase visibility.
- Reduce exposure.
- Improve external (and where possible internal) score.
- Ensure a strong incident response process to recover and respond to cyber-attacks.

Figure 10.9 illustrates the most common components and objectives of a CTEM program:

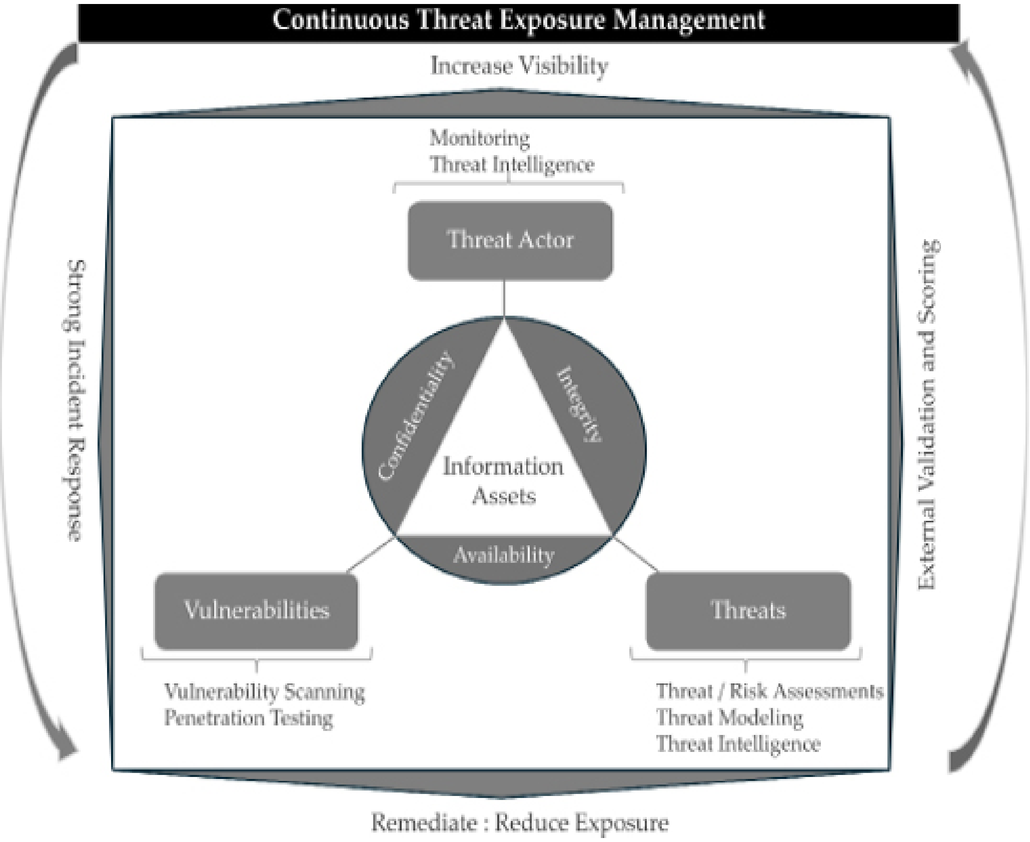


***Figure 10.9**: Components and objectives of a CTEM program*

Conclusion

In this chapter, we explored various techniques to understand attackers' evolving actions that bring about threats and explored strategies to mitigate them. Using VAPT and application security testing methods, we can determine and reduce several aspects of cyber threats. A CTEM program can help manage threats by using relevant intel, increasing the visibility of assets in scope, determining existing vulnerabilities, and remediating them on time using a risk-based strategy to help reduce the chances of security incidents.

In the next chapter, we will explore core concepts of logging and using them as a key detective control and for ensuring a robust and effective incident response is put into operation.

Key takeaways

Some of the key learnings from this chapter include:

- The CISO and the security team need to maintain a hawk's eye on the new and emerging threats.
- The attack TTP being used by the TA deployed today is complex and multi-dimensional.
- Irrespective of their motivation, their net goal is to bring harm to organizations and/or individuals.
- Several sources, such as the MITRE D3fend and MITRE Att@ck, are great sources to learn and create preventative and corrective mechanisms.
- With cloud computing, the ability of the TAs to leverage their tradecraft at scale also grew, and that brought in an additional element of cyber crimes from distant places with little or no jurisdiction for law enforcement.
- Enforcement agencies around the world may need to make concerted and coordinated efforts to prevent any widespread cyber disruption.
- The **threat intelligence** (**TI**) available from a variety of renowned industry sources, industry forums, and CERTs in various countries may serve as a good lead indicator of the attack vectors.
- Using them, programs like VM and PT should be enriched with such TI to detect existing gaps in the environment and drive risk-based remediation. Such efforts are critical to handle the emerging zero-day vulnerabilities.

- Using ESAM, i.e., monitoring and reduction of vulnerabilities and configuration weaknesses on internet-exposed environments, should be done continually.
- **Threat modelling** (**TM**) on applications and critical process workflows is a great way to ensure the CIA of information assets is protected.
- Several external services rank and rate the organization's external exposure and should be considered to get an independent review and score of those environments.

References

- **https://csrc.nist.gov/glossary/term/cyber_threat**
- **https://unit42.paloaltonetworks.com/north-korean-it-workers/**
- **https://www.gartner.com/en/articles/how-to-manage-cybersecurity-threats-not-episodes**
- **https://www.cisa.gov/known-exploited-vulnerabilities-catalog**
- **https://avidml.org/**
- **https://attack.mitre.org/**

CHAPTER 11
Incident Response and Planning

Introduction

The mission of a CISO and his/her information security team is to enable the secure functioning of the information assets of the organization. Several risk-based choices decide the controls that are deployed to protect the CIA's information assets. However, several organizations continue to be targets of successful attacks by the adversary. It becomes pertinent for organizations to gain visibility around access and use of their information assets and have mechanisms to detect, contain, and respond to any anomaly as fast as possible. In this chapter, we will explore the concepts of logging and monitoring and their relevance to incident response.

Structure

The chapter covers the following topics:

- Logging and monitoring
- Uses of logs
- Managing incidents and breaches

Objectives

By the end of this chapter, you will be able to understand the importance of logging and strategies to use to detect anomalous events and correlate them. You will also learn about incidents and breaches, and how to investigate and manage them.

Logging and monitoring

In *Chapter 3*, *Role of Standards and Controls,* we covered the NIST **Cyber Security Framework** (**CSF**) and a few other standards. All of them required detective controls for any attempts or successful compromise of the CIA. The ability to know which information asset is at risk through logs is called **visibility**.

Figure 11.1 illustrates the NIST CSF 2.0's Detect function, its categories, and some of the controls:

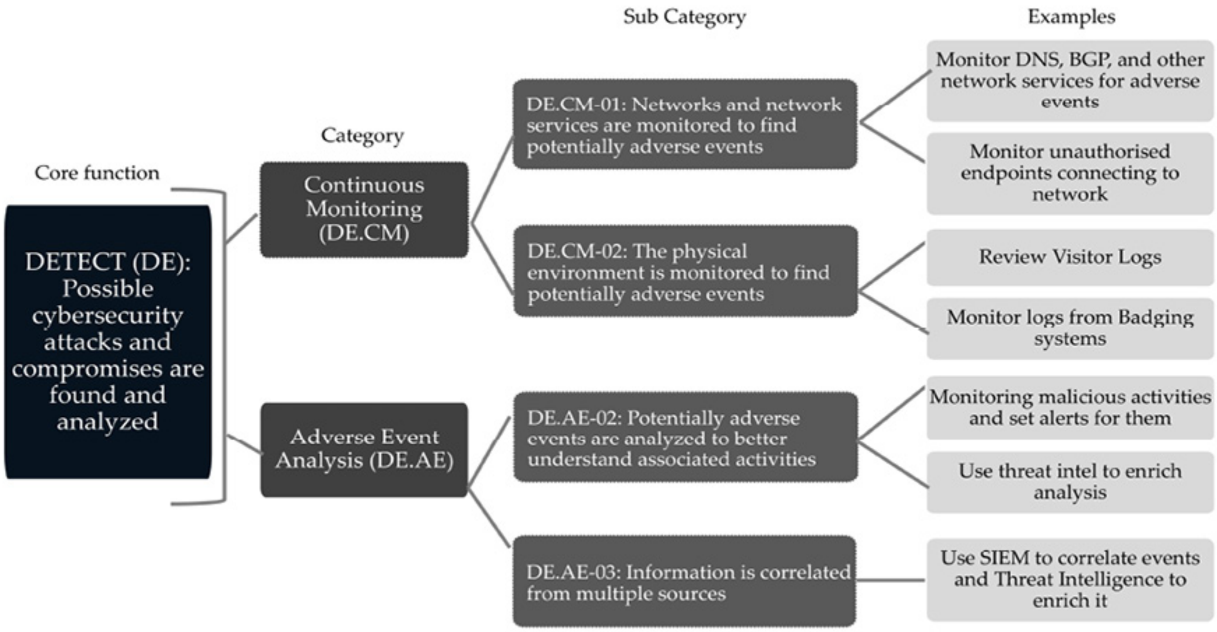

Sub Category
Examples
Core function
Category
DETECT (DE): Possible cybersecurity attacks and compromises are found and analyzed
Continuous Monitoring (DE.CM)
Adverse Event Analysis (DE.AE)
DE.CM-01: Networks and network services are monitored to find potentially adverse events
DE.CM-02: The physical environment is monitored to find potentially adverse events
DE.AE-02: Potentially adverse events are analyzed to better understand associated activities
DE.AE-03: Information is correlated from multiple sources
Monitor DNS, BGP, and other network services for adverse events
Monitor unauthorised endpoints connecting to network
Review Visitor Logs
Monitor logs from Badging systems
Monitoring malicious activities and set alerts for them
Use threat intel to enrich analysis
Use SIEM to correlate events and Threat Intelligence to enrich it

Figure 11.1*: NIST CSF 2.0 and DETECT function*

Often in buildings and offices, a physical register such as a visitor entry logbook is maintained at the office gate or the reception area to document the details of the visitors. This is a form of log. In this book, we will focus on logs of digital information systems.

Logs are the digital footprint on the network. Every device, application, IT, and **operations technology** (**OT**) environment can generate logs with varying degrees of detail. Logs reveal several stories, such as who logged in from where, when, how, and what. These are all records of an event or activity, and they help stitch the puzzle together when things are about to go wrong, or things have gone wrong, or you may just want to know some history.

Uses of logs

Logs serve the following primary purposes:

- **Management**: To maintain the availability of the device and ensure it is kept in good health. For instance, a server with high memory utilization may require a cleanup of its temporary files, unwanted files, and maybe even a simple reboot.
- **Troubleshooting**: To understand what could be going wrong and fix it. For instance, the VPN connection between the host and the end-user device may often be due to some inefficient timeout settings. Similarly, the logs are used to debug the applications or devices' logic flow and bugs (these are called **debug logs**).
- **Investigation**: To evaluate and stitch the chain of events that may indicate a security incident or to establish the evidence of an event. An event that may indicate a compromise of the CIA of any information system should be treated as an **incident**. NIST defines an incident as *an occurrence that actually or potentially jeopardizes the confidentiality, integrity, or availability of an information system or the information the system processes, stores, or transmits, or that constitutes a violation or imminent threat of violation of security policies, security procedures, or acceptable use policies*. In this book, we largely focus on security incidents; however, investigations may also be for aspects like transaction fraud. The terms incident and security incident have been used interchangeably.
- **Audit evidence**: To satisfy the assurance requirements of standards like **Control Objectives for Information and Related Technologies** (**COBIT**) or NIST CSF. For instance, a log of user provisioning and de-provisioning on **Active Directory** (**AD**) is important for user

management, covered in *Chapter 7, Identity and Access Management*.

- **Record for history**: Regulations may require records of financial transactions to be maintained/stored for a certain number of years. For instance, income tax authorities may require a taxpayer to keep relevant records of tax exemptions claimed for a period of 7 years.

Categories of logs

Table 11.1 explains the focus of various categories of logs and their purpose, with some examples:

Category of logs	Focus/Key questions it may answer	Examples
Performance	Is the system (device, application, cloud resource, CCTV etc. functioning appropriately within desired parameters? Is it time to plan and execute a maintenance activity?	Logs of the hourly reading on thermostat (temperature) installed in data center. Boot time, shut down time, Last reboot time and uptime of a machine. Memory and CPU utilization of servers. System error logs such as **Blue Screen of Death** (**BSOD**).
Security	Logs that would likely impact CIA due to a security threat. Were there any unauthorized changes to security devices such as firewalls or servers like AD? Was there any malicious traffic or file that came in or was sent out? Can we determine who accessed which system or file?	Visitor log for entry/exit time to office premises. Login and logout time from the building—using proximity cards and/or biometric. Login and logout time from applications or devices. User access logs to application/system. Attempts to change passwords. Attempts to flood perimeter devices with **denial of service** (**DoS**) requests. Log of which file was accessed by whom and what was changed. Log of VPN connection. Log of file quarantined by the anti-malware tool. Firewall rule that processed an incoming packet. Log of phishing emails received and processed. Log of files with PII sent out via email or chat forums. Logs of new permission assigned or revoked

		for a user.
Transaction/Activity	Generally, the most voluminous it captures most important transactions and activities relevant to the software/device.	Last date and status of vulnerability scan. Log of user accesses to a website. Log of when an invoice was processed. Log of payment made. Log of upgrade of operating system. Log of risk decision on a **governance, risk and compliance** (**GRC**) tool.

Table 11.1*: Major log categories, their purpose, and examples*

Tip: A server or endpoint reporting a higher and unexpectedly high CPU utilization are performance logs that may indicate a security event. WannaCry (2017), a ransomware, caused unexplainably high CPU utilization.

Sources of logs

Every information system, ideally, should generate logs for the categories listed above. We can classify logs by the sources, such as:

- **Network**: These are about network performance and traffic, such as DNS, DHCP, ICMP, and so on.
- **Servers**: These are about the server's performance and actions.
- **Security devices**: Logs from devices such as firewalls.
- **Application logs**: These are about an application's execution and/or error situations.
- **Database**: These are about actions within and to/from a database, for instance, running a SQL query, data, or field-related changes to the database.
- **Cloud**: Contains logs of activity on the cloud environment.
- **Endpoints**: Logs from end-user devices and security technologies like **endpoint detection and response** (**EDR**).
- **Identity**: Logs related to identity, sometimes the ID is centralized, such as an AD, but may be specific to applications (local user IDs). Similarly, logs of **Remote Authentication Dial-In Service** (**RADIUS**) logs to perimeter devices. Access identity logs cover all such scenarios. Identity logs are important for monitoring the behaviors of a privileged user.
- **Email**: Logs of email sent, and received at the email gateway, and at user mailboxes, and any security-related actions taken, for instance, blocking spam emails.

- **Web**: Logs that relate to websites accessed or attempted to be accessed.

Tip: Major operating systems like Windows, Unix/Linux, and iOS have a built-in event viewer that can be used to review localized event logs. Security tools and devices, such as EDR and firewalls, also have a built-in log viewer to sort, filter, and search.

Tip: Threat Hunting, i.e., the determination of any ongoing or past malicious activity, is done by searching through appropriate logs and analyzing them. It can be considered a part of the investigation.

The characteristics of a good log are as follows:

- **Relevant**: The log must contain relevant details with respect to the objective of the log. Any additional avoidable details should follow the need-to-have and data minimization principle. For instance, a log of user access to an application may not need to capture the file size of the html page that was loaded for the user. Similarly, a log of the firewall rule that was applied to an incoming packet need not capture which firewall admin authored the rule.
- **Synchronized time**: Logs must be aligned to a standardized time reference, such as by using a **Network Time Protocol** (**NTP**) server. Timestamp is probably the most important detail of the log to chain the events together during an investigation. For instance, the timestamp of a log of the user's login to AD must be chronologically relatable to the log of the same activity from the local machine. The log must mention the time zone of the log (machine time).

Note: It is possible that the firewall may have time in Universal Time Coordinated (UTC), whil the machine may have time recorded in Indian Standard Time (IST). As long as these two time are in sync with the NTP service/server, it is fine, though analyzing may become difficult.

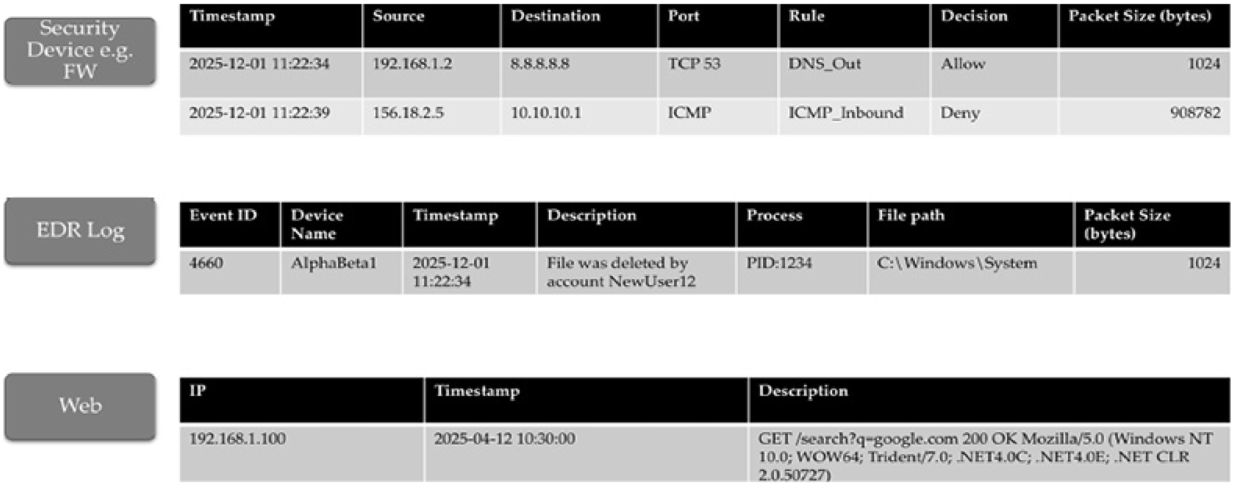
Tip: It is recommended to disable manual change of time of any machine, though allowing change of time zone on the machine is fine, especially for users who travel. This means when user travels to a country, he/she may be allowed to change the local machine's time zone. Th time itself will remain in sync.

- **Simple format**: Logs must be easy to parse/read to understand. For instance, the router log of outbound traffic to the Internet should make it easy to understand the origin, destination, packet/data size, and time of the activity.
- **Standardized format**: Ideally, all digital devices and applications must produce logs in a format as per industry standards such as **JavaScript**

Object Notation (**JSON**), **Common Event Format** (**CEF**), or **Extended Log Format** (**ELF**). This helps in log portability and provides easier ways to analyze.

- **Sufficient**: The details logged must be sufficient for the purpose it will be used for. For instance, the name of a service account that pushes patches to the machines is not logged, which can lead to troubleshooting failure and issues when investigating which ID may have caused a device to fail. Similarly, an application's debug log must be able to indicate which function call or code library generated the error, and if feasible, at which line of code.
- **Immutable**: The integrity of logs, i.e., any restriction on any modification, is key to ensuring any relevant use. Logs must be reliable, and their source, storage, and access must be of unquestionable integrity. Organizations may transform a copy of the log for analysis. For instance, if the logs of various devices are in different formats, or their time zone is different (while they sync to an NTP), the log transformation may be applied to get them to be consistent.

Figure 11.2 illustrates some examples of formats and logs:



Security Device e.g. FW

Timestamp	Source	Destination	Port	Rule	Decision	Packet Size (bytes)
2025-12-01 11:22:34	192.168.1.2	8.8.8.8	TCP 53	DNS_Out	Allow	1024
2025-12-01 11:22:39	156.18.2.5	10.10.10.1	ICMP	ICMP_Inbound	Deny	908782

EDR Log

Event ID	Device Name	Timestamp	Description	Process	File path	Packet Size (bytes)
4660	AlphaBeta1	2025-12-01 11:22:34	File was deleted by account NewUser12	PID:1234	C:\Windows\System	1024

Web

IP	Timestamp	Description
192.168.1.100	2025-04-12 10:30:00	GET /search?q=google.com 200 OK Mozilla/5.0 (Windows NT 10.0; WOW64; Trident/7.0; .NET4.0C; .NET4.0E; .NET CLR 2.0.50727)

***Figure 11.2**: Examples of some logs*

Considerations for logging strategy

Just about anything can be logged, and therefore, there must be careful thought of what to log, why, and how it will be used. Otherwise, it is very easy to have a plethora of logs that can become very noisy, i.e., too many distracting rows of information to peruse. Some of the considerations are:

- **Type of events to log**: An event may have at least two outcomes, such as **successful and unsuccessful**. Organizations must choose the right mix of logging based on the purpose it will serve. For instance, AD logs should log a user's successful logins and unsuccessful attempts as well. The unsuccessful login tracking is useful to protect against account compromise by setting a policy, such as on AD's **Group Policy Object** (**GPO**), to allow only five failed attempts. We covered this in *Chapter 7, Identity and Access Management*.
- **Type of personas to log**: This aspect focuses on what all identities are of interest for logging, such as normal users, admin users, and even adversaries, and to what extent. For instance, organizations would log more events of an admin user than of a normal user.
- **Type of transactions to log**: To ensure only relevant activities are logged. For instance, an e-commerce shopping cart's shipping-related logs need not capture the items added or removed during the time the customer is perusing through the shopping catalog. It must primarily capture the final items chosen, their unit price, and shipping details, etc. It may be a good-to-have log of the items perused to build customer experience and nudge the shopping for an item not selected for later use.
- **Log collation**: Logs are generated by various sources and are usually sent to a common system logging server using a protocol called **syslog**. Devices such as firewalls and routers have very limited storage on them, and thus sending the event logs to an external system allowed them to keep their memory optimally used. Additionally, collating all related logs in one place allowed for better analysis and freeing up the processing power on respective devices. These common log servers for security events have evolved to be **security information and event management** (**SIEM**) servers. A normal log server may have several performance and transactional logs that may not be directly relevant to a security event. A SIEM may be a module/service provided by a **cloud service provider** (**CSP**) as well.
- **Retention period**: Logs consume a lot of space and may outlive their value with time. The duration for which logs should be maintained is

called their **retention period** and is determined by aspects like regulatory requirements, contractual requirements, and management direction. Organizations would usually keep logs for 3+9 months, i.e., 3 months' live storage and 9 months in cheaper storage but accessible through queries. Certain logs, such as those parts of any investigation or litigation, are retained for a longer time. A **data lake** is a centralized repository of large volumes of data in low-cost, searchable, and scalable storage. Logs are an ideal candidate for using a data lake.

- **Access to logs**: Security event logs are confidential; organizations must apply the principles of *least privilege* and need-*to-have/need-to-know* to decide on access models. No one should be able to modify the logs.

Tip: No one, especially admins, should be able to delete logs. If any transaction log may have to be deleted, a log of such deletion should also be generated.

Figure 11.3 illustrates the log strategy and the various aspects discussed above:

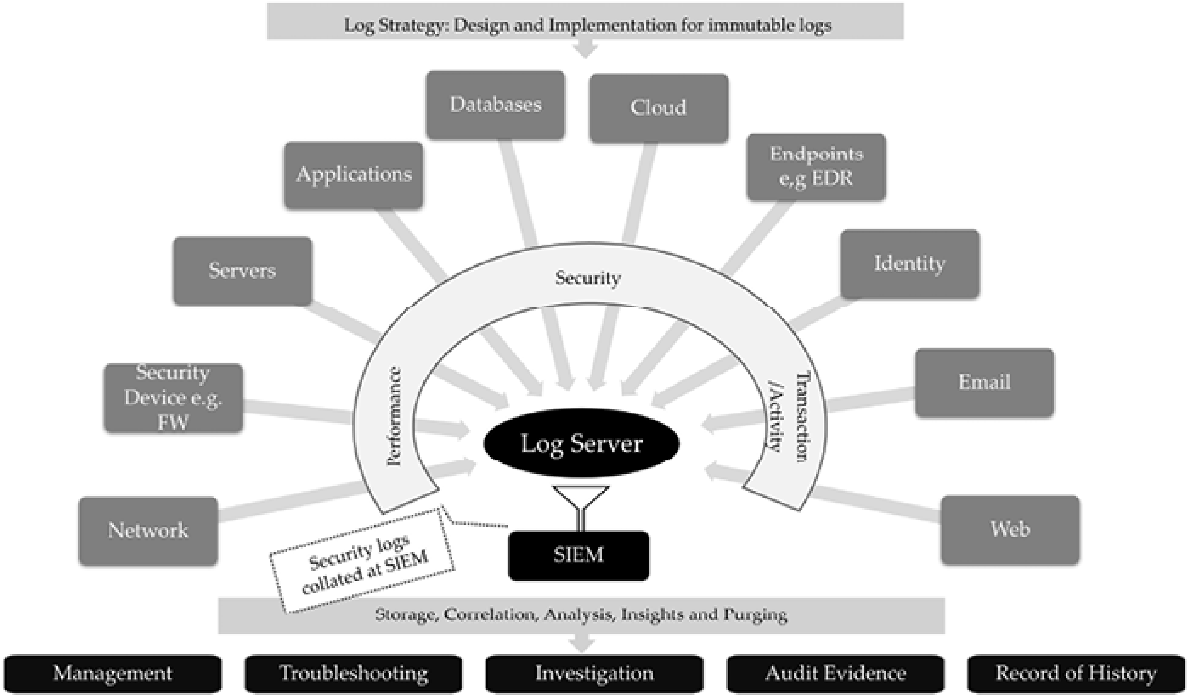


***Figure 11.3**: All about logs*

Using SIEM effectively

Logs are generated by different devices/applications; to stitch the story together, we have to examine several logs. For instance, to determine how and which malicious file attached to an email successfully came through the defenses and was delivered to the end-user's machine. We will examine EDR logs, users' email system logs, email gateway logs, and maybe even login-related logs.

All security event logs are collected/collated at a SIEM. The **capabilities of a SIEM** are:

- Mechanisms to have built-in **connectors** or methods to pull the logs from various systems. Connectors are software code that interfaces with other software. Logs may also be pulled using an **application programming interface** (**API**).
- Collect and aggregate logs.
- Create alerts based on predefined rules, for instance:
    - The administrator attempts to log in to a server without using the approved **privileged access management** (**PAM**) tools.
    - A userID's login attempts come from geographically distant locations in quick succession, such as one from Chennai, India, and within 5 minutes in Washington, DC, United States. This is also called **impossible travel.**
    - Firewall's detection of the use of the `curl` command from an internal host to an external unknown IP.
- Create an alert based on modern-day **machine learning** (**ML**) and threat intelligence feeds to automatically detect and report **indicators of compromise** (**IOCs**).
- Provide means to search for IOCs and any patterns in logs.

A SIEM can be on-premises, on the cloud, or a combination of both. Some of the popular SIEMs are *IBM's Qradar*, *Splunk Enterprise Security*, *LogRythm*, *Palo Alto's Nextgen SIEM*, and *Crowdstrike's Falcon*.

Note: In September 2024, Palo Alto Networks (PANW) acquired IBM's cloud-based Qradar SIEM. As of August 2025, the on-premise version of Qradar is owned by IBM.

The team that defines and implements the logging strategy in partnership with several asset owners, such as the IT admin of a server, the admin of the applications, and the manager of the building access, is called the **security operations center** (**SOC**) or **cyber defense center** (**CDC**). Such a team is usually staffed to work round-the-clock (24x7x365).

The key skills of a SOC analyst/CDC analyst are:

- Comprehension of networking concepts and protocols.
- Ability to understand applications, API, data flows, and identities.
- Hands-on experience with SIEM technologies and logging in general.
- Ability to write queries and scripts that help in threat hunting and log analysis.
- Apply analytical thinking, ask the *what-if* questions, and pivot to relevant clues.
- Use analytical judgement, i.e., avoid jumping to conclusions and avoid inferences based on pre-conceived notions (such a bias in judgement is called **confirmation bias**).
- Ability to understand and apply threat intelligence.
- Ability to think about the big picture and connect the dots.
- Curiosity is necessary.
- Communicate effectively.

Skills in a SOC team are layered in a hierarchy. Entry-level analysts, called L1, are the **eyes-on-the-glass**. Their role is to be glued to the screen and watch for any anomalous alert logs thrown up and triage them. For instance, a misdirected email with confidential company documents to an external recipient, or a **distributed denial of service** (**DDoS**) on a perimeter device. L2 analysts may be tasked to analyze the root cause and patterns of such attacks or control failures. L3, on the other hand, generally is the authority on the corrective and preventive actions to take based on logs or the available threat intelligence. For instance, L3 would direct the network security team to disable the WAN connection to a partner if the network traffic indicates a malicious packet coming from it. Each organization may structure its SOC/CDC team differently, with different roles and responsibilities. They may also choose to use a **managed service**, i.e., they may outsource the SOC team’s work to a third party who will have the expertise to monitor and triage. Organizations may also do this in hybrid mode, where the L1 or part of the L2 is outsourced by L3 in-house. There are several **managed SOC** services on the cloud that are bound by **service level agreements** (**SLA**) to triage, handle, and respond to events. Managed SOC is also called a **managed security service provider** (**MSSP**), though an MSSP may offer other services like continual penetration testing, too.

When a SIEM, like IBM Qradar, generates alerts for **impossible travel**, a CDC specialist may need to examine the network log, the identity log, the previous history of any such login, and so on. The sequence of things to do in these investigations may be a standard procedure and thus can be automated. Even corrective actions to take, such as notifying the concerned user, modifying the ID's password expiry state, or disabling access to privileged systems, can also be automated. This concept involves the ability to orchestrate across various relevant sources of logs, automate the analysis, and trigger a response action, which is called **security orchestration, automation, and response** (**SOAR**). SOAR significantly reduces the L1 efforts and even automates routine, standardized steps based on a concept called **playbook.** A playbook is basically the sequence of steps to take for a particular alert. A CDC analyst may need to get engaged in situations where a playbook does not exist or the threat level is elevated.

Some scenarios to consider for the SOAR playbook are:

- Automatic deployment of patching for a vulnerability with a **Common Vulnerability Enumeration** (**CVE**) rating of 9 and above.
- Quarantine of the file by EDR when a malicious attempt to modify the kernel is detected.
- Leverage threat intel and perform hunting operations automatically.
- Turning off the access to a newly configured publicly accessible cloud storage.
- Automatic renewal of the SSL certificate of a public website.

SOAR can also create a trigger for a required process, such as an IT change management ticket or a workflow for other teams, such as initiating user awareness training for handling phishing emails.

SIEM and SOAR are complementary to each other. The term SIEM, especially next-generation SIEMs, often includes SOAR capabilities by default.

Logging and monitoring provide actionable insight to deploy any additional controls to contain the incident, respond to it, and/or recover from it.

A key term used in logging and minoring for the timely detection of incidents is **mean time to detect** (**MTTD**). It is the mean time to detect an incident, a threat actor's activity. The shorter the MTTD, the better it is for appropriate triage and subsequent handling.

Managing incidents and breaches

Earlier in the chapter, we defined an incident as *an occurrence that actually or potentially jeopardizes the confidentiality, integrity, or availability of an information system or the information the system processes, stores, or transmits, or that constitutes a violation or imminent threat of violation of security policies, security procedures, or acceptable use policies.*

SIEM-SOARs handle security events and even security incidents to some extent.

A security event may be a relevant deviation from the norm to observe, validate, and maybe even respond. For instance, the access block due to impossible travel is a security event, and because it was proactively restricted, it is not a security incident. However, a continual attempt of DoS, defacement of the corporate website, unauthorized entry to a restricted area, unauthorized access to information system(s), and unauthorized or even incorrect transmission of PII would all be treated as security incidents. Organizations take a risk-based approach to classify and tier security incidents as well. The terms incidents, security incidents, and cyber security incidents are used interchangeably. An incident that involves unauthorized access, disclosure, or loss of PII/PHI is called a **data breach** or **breach**. Even unintentional access to PII/PHI where confidentiality is lost can be treated as a breach. All breaches are incidents, but all incidents need not be breaches. In an organization, teams such as privacy professionals and members of the legal team are the right authority to declare a breach in consultation with investigating teams. The term has many implications, including regulatory ones, and hence, adequate care must be taken.

Regulations like the USA's **Health Insurance Portability and Accountability Act** (**HIPAA**), India's Information Technology Act (2000) and its IT rules (2021), and the **General Data Privacy Regulation** (**GDPR**) require security incidents to be reported to regulators. Regulations such as these also have a stringent requirement to notify breaches.

Organizations usually have a dedicated incident response team (IR team: also called **incident handlers**) to manage the incidents, their triage, validation, analysis, containment, communication, and even recovery. IR handlers are specialists trained to handle not just technical aspects, but also have business acumen and skills in psychology and communication. Typically, a lead investigator is assigned to lead the overall incident and collaborate with a cross-functional team such as CDC, other cyber security teams, IT teams, legal teams, physical security, employee relations, and corporate branding. The senior management of the organization may be engaged during critical incidents and for breaches.

NIST CSF 2.0 provides very good guidance on how to manage incidents. There are broadly four categories:

- **Incident management**: This involves documenting and using a playbook, i.e., an operating procedure to validate and handle an incident. The playbook articulates the steps to take to contain, respond, and recover from incidents. It also articulates the roles and responsibilities of the key personnel across the cross-functional team. The **incident response plan** (**IRP**) must be periodically evaluated via a simulation to build muscle memory for incidents and to ensure there is less chaos and more planful action to contain the incident and recover from it. The scenarios chosen must be a good representation of incidents observed previously, the updates coming from threat intelligence, and any such learning from industry peers that may be applicable. For instance, a playbook should be updated based on new threat intelligence of increasing ransomware being spread using emails. As part of the plan, the IR team should model what can go wrong and then what to do. It may need to involve a cross-functional team to even build the playbook.
- **Incident analysis**: This focuses on conducting investigations based on logs, interviews, and other intelligence received or collected. The evidence must be appropriately maintained such that there is a clear handoff of collection and handling, and its storage to rule out any unauthorized access and/or tampering with the evidence. The process of managing the custody of evidence and artifacts collected securely to

relevant stakeholders is called a chain of custody. Some legal counsels advise marking all evidence as *privileged and confidential* to protect the evidence disclosure. The evidence collection and evaluation may involve digital forensics and analysis of the attacked system(s). In forensics, an image of the live system is acquired, and all analysis is done on that image. This is relevant to preserving the logs, such as the last file access. For instance, if an IR analyst directly accesses the infected file, the timestamp of the file and its last access log will change; this may mean that the metadata about the last person who actually accessed the file will not be determinable. Ideally, this effort of forensics should be outsourced to expert organizations that can be held on a retainer basis. This provision helps establish independent and unbiased reviews and augurs well with external stakeholders. The third-party forensics team shall be under a **non-disclosure agreement** (**NDA).** Similarly, an external legal counsel can help position the holding statement and navigate the regulatory conversations.

- **Incident response reporting and communication**: This category of the Respond function guides stakeholder communication. Stakeholders could be internal users, management, business partners, customers, end-consumers, and even regulators. The call on what to report, to whom, and how much is typically guided by the Legal function in partnership with corporate communications. They draft and approve a statement called the **holding statement**, which is the only detail that can be shared. Such external notification, especially when an organization has suffered a breach or a significant incident, requires due caution to only share what the organization authorizes. Listed companies are required to file such notifications to the market regulators. For instance, the US regulator requires listed companies to file a 10-K statement that captures risks and any material security incidents. In the event of public disclosure, it is important to ensure that updated and correct security contact information is maintained.
- **Incident mitigation**: This category focuses on actions to execute to contain the incident and enable measures to protect the information systems. For instance, leverage the spam filter feature of email security gateway, threat prevention modules on firewalls, kernel protection features of EDRs, and **network access control** (**NAC**) to allow only

approved devices on the network. Such leverage of built-in capabilities reduces the occurrence of incidents. Mitigation may need to be actioned by the incident handler(s) or the SOAR, based on the threat intel and log analysis. For instance, if the email security gateway has made an exclusion to some permitted sender domain and may have received a malicious email from that domain, the incident handler may have to remove such an exception and contain the email boxes that received the email.

MTTR, the mean time to respond to an incident and contain is a key means to gauge the effectiveness of handling an incident. Naturally, there would be some incidents that may take time for remediation and some that are quick fixes. MTTR is therefore often measured against the criticality of the incident. The objective for an IR team is to handle the incident effectively while keeping the MTTR as low as possible.

Figure 11.4 illustrates some of the examples of categories, subcategories, and controls for the core function **RESPOND**:

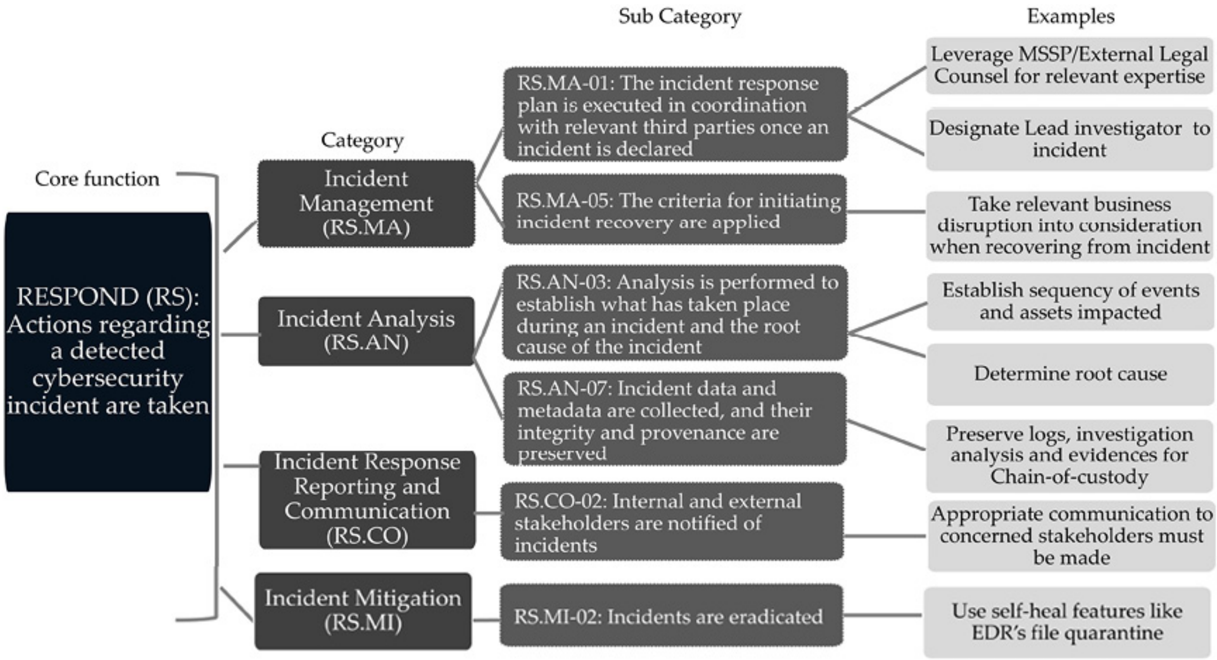

Sub Category
Examples
Core function
Category
RESPOND (RS): Actions regarding a detected cybersecurity incident are taken
Incident Management (RS.MA)
Incident Analysis (RS.AN)
Incident Response Reporting and Communication (RS.CO)
Incident Mitigation (RS.MI)
RS.MA-01: The incident response plan is executed in coordination with relevant third parties once an incident is declared
RS.MA-05: The criteria for initiating incident recovery are applied
RS.AN-03: Analysis is performed to establish what has taken place during an incident and the root cause of the incident
RS.AN-07: Incident data and metadata are collected, and their integrity and provenance are preserved
RS.CO-02: Internal and external stakeholders are notified of incidents
RS.MI-02: Incidents are eradicated
Leverage MSSP/External Legal Counsel for relevant expertise
Designate Lead investigator to incident
Take relevant business disruption into consideration when recovering from incident
Establish sequency of events and assets impacted
Determine root cause
Preserve logs, investigation analysis and evidences for Chain-of-custody
Appropriate communication to concerned stakeholders must be made
Use self-heal features like EDR's file quarantine

***Figure 11.4**: NIST CSF2 Respond function with examples of controls*

Recovering from an incident

Organizations should plan to recover and restore normal operations as soon as possible after any incident. The success of any restoration efforts is dependent on the quality, integrity, and completeness of backups. Before an incident hits, the organization must plan and implement a solid backup strategy to ensure all its chosen critical environments have immutable and usable backups that can be restored.

There are two key aspects to restoration:

- **Backup restoration**: Using backups to restore to a previously known healthy state.

  Backup restoration must be tested at periodic intervals to gain assurance that it will work and be able to restore the business operations as soon as possible. Generally, the point at which backup is taken may not be up to the current time. For instance, when working on a Word document, if the endpoint crashes and reboots, you may notice that an auto-recover copy of the same file may be available. It may still not have the last few word types. The point up to which the backup is available is recoverable. Organizations choose to define their **recovery point objective** (**RPO**), which is the maximum amount of data the organizations can tolerate as their effort. For instance, an e-commerce portal will prefer to have a strong backup so that no transaction is lost and everything related to it is immediately backed up. This is cost-prohibitive and may not always be practical. Remember, a lot of failed/successful transactions may still be there. Secondly, organizations may also need to define their **recovery time objectives** (**RTO**), i.e., the maximum downtime (in minutes/hours) that an organization may be able to tolerate. For instance, a financial services company allowing online stock trading will have zero downtime vs an organization that provides email services. A few minutes of delayed emails may not be detrimental to service. RPO and RTO become relevant as incident recovery is being considered to prioritize business functions that have shorter RTOs earlier.

- **Recovery communication**: Relevant communication to key stakeholders on the progress of restoration and its expectations must be continually made. The stakeholders may be internal or external. Much like incident communication, even restoration efforts may require using a Legal function-approved holding statement.

  Figure 11.5 illustrates some of the examples of categories, subcategories, and controls for the core function **RECOVER**:

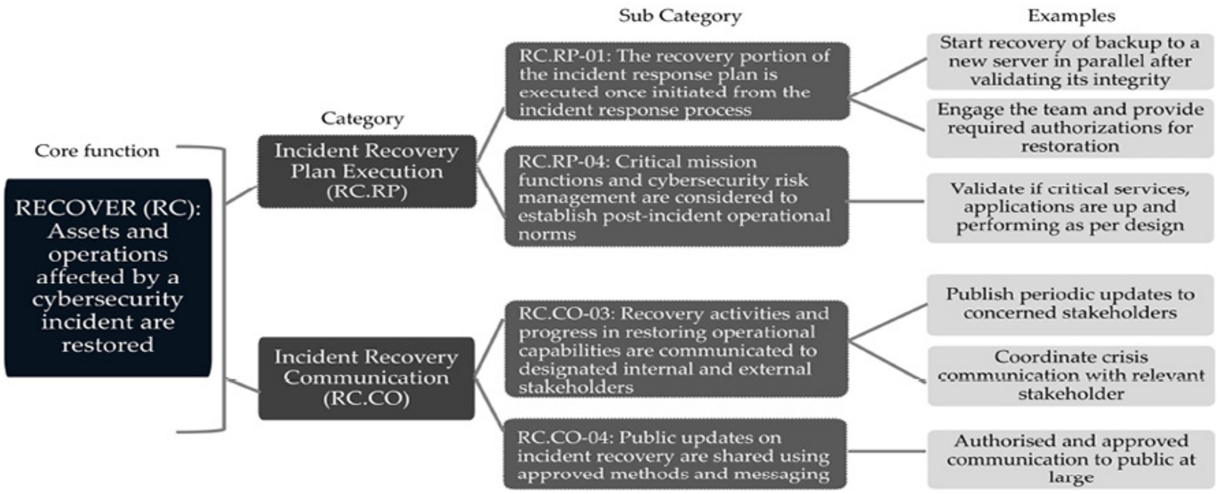

Sub Category
Examples
Core function
Category
RECOVER (RC): Assets and operations affected by a cybersecurity incident are restored
Incident Recovery Plan Execution (RC.RP)
Incident Recovery Communication (RC.CO)
RC.RP-01: The recovery portion of the incident response plan is executed once initiated from the incident response process
RC.RP-04: Critical mission functions and cybersecurity risk management are considered to establish post-incident operational norms
RC.CO-03: Recovery activities and progress in restoring operational capabilities are communicated to designated internal and external stakeholders
RC.CO-04: Public updates on incident recovery are shared using approved methods and messaging
Start recovery of backup to a new server in parallel after validating its integrity
Engage the team and provide required authorizations for restoration
Validate if critical services, applications are up and performing as per design
Publish periodic updates to concerned stakeholders
Coordinate crisis communication with relevant stakeholder
Authorised and approved communication to public at large

***Figure 11.5**: NIST CSF2 Recover function with examples of controls*

The trigger for an incident is usually an alert from SIEM, but it could also come from a user-reported issue and/or external sources, for instance, an SIEM tool cannot pick up a web defacement, but a researcher might notice and responsibly disclose. Using the NIST CSF 2.0 and the general industry practice, **incident response** (**IR**) is a process with the following phases:

- **Plan**: This phase closely represents the Respond (RS) function and the category *Incident Management (RS.MA): Responses to detected cybersecurity incidents are managed*. The outcome of this phase is an IRP.
- **Validate**: This phase closely represents the Respond (RS) function and the category *Incident Analysis (RS.AN)*.
- **Contain and isolate**: This phase closely represents the Respond (RS) function and the category *Incident Analysis (RS.AN)*.
- **Eradicate root cause**: This phase closely represents the Respond (RS) function and the category *Incident Mitigation (RS.MI)*.
- **Recover**: This phase closely represents the Recover (RC) function and the category *Incident Recovery Plan Execution (RC.RP)*.

All these phases include appropriate communication to relevant stakeholders and a feedback loop to the following parts of the security program:

- Employee awareness material to prepare them to avoid falling prey to phishing emails, etc.
- Policies such as the **acceptable use policy** (**AUP**).

- Changes to configurations, such as updates to the firewall threat protection module and SOAR.
- Updates to the IRP for any learnings from the incident.
- Updates to threat intelligence sources and integration with SIEM-SOAR.

Figure 11.6 illustrates the sources of incident triggers, various phases of IR, and their interlinkages to NIST CSF 2.0 functions:

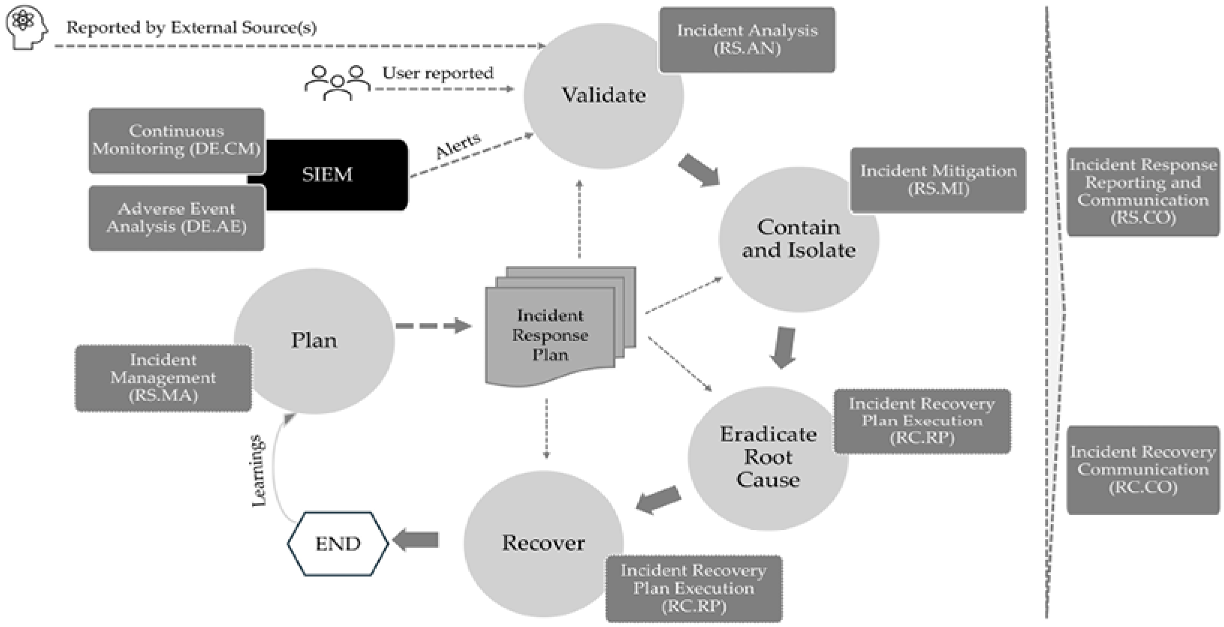

Reported by External Source(s)
User reported
Incident Analysis (RS.AN)
Validate
Continuous Monitoring (DE.CM)
SIEM
Alerts
Adverse Event Analysis (DE.AE)
Incident Mitigation (RS.MI)
Contain and Isolate
Incident Response Reporting and Communication (RS.CO)
Plan
Incident Response Plan
Incident Management (RS.MA)
Learnings
Eradicate Root Cause
Incident Recovery Plan Execution (RC.RP)
Incident Recovery Communication (RC.CO)
END
Recover
Incident Recovery Plan Execution (RC.RP)

Figure 11.6*: IR triggers, its phases, and interlinkages with NIST CSF2.0*

Learnings from past incidents

Almost every large organization has had security incidents over the years. Some have even had breaches. The list, unfortunately, is exhaustive and cannot be covered here. We will examine a few such breaches that were publicly announced in recent years:

- **Mar 2025, Oracle**: A hacker declared that the **Oracle Cloud** environment was breached, and about 6milion users and passwords were compromised. The company continued to downplay the incident, claiming no impact on its cloud infrastructure known as **Oracle Cloud Infrastructure** (**OCI**). Oracle later claimed two obsolete servers were indeed compromised, but still maintained no impact on OCI.
    - **Key learning so far**: Incident-related communication must be

handled well to reinforce customer confidence.

- **Aug 2021, Log4j**: A critical vulnerability CVE-2021-44228 with a CVSS of 10 was announced. Several organizations and applications use this Java library, bringing possible exposure to the entire world.
- **June 2023, MOVEit**: It is a file transfer tool that had a SQL Injection vulnerability (CVE-2023-34362) that was allegedly compromised by a Russia-based **Adversary Group** (**AG**) called Clop. Approximately 200 organizations and 17.5 million users worldwide were affected.
    - **Key learning**: Tools like MOVEit were used by multiple software, bringing the importance of managing supply chain software and risks from them to the fore. Organizations started focusing on requiring their key software providers to share their **software-bill-of-material** (**SBOM**).
- **Feb 2024, Change Healthcare breach**: An AG named ALPHV, a ransomware group, gained access to Change Healthcare and crippled the healthcare services. Change Health is a USA-based chain of primary healthcare centers and a company owned by one of the largest insurers in that country, UnitedHealth Group. Change Healthcare had to resort to manual methods to provide care to patients. Several of the patients could not get time approvals for their treatment, causing widespread hardship.
    - **Key learning**: Cyber resilience and the organization's robust planning to recover from incidents, especially cyber incidents, must be done proactively.
- **Jan 2024, Microsoft Azure test environment**: An AG named **Midnight Blizzard** conducted a password spray attack on Microsoft Azure's legacy test account that did not have any **multi-factor authentication** (**MFA**). The AG then gained access to code repositories, cryptographic keys, and emails of some key personnel.
    - **Key learning**: Even test environments must be duly protected, if needed by MFA, and access to confidential information must be restricted.
- **Dec 2023, X (formerly Twitter)**: PII of about 200 million X users was compromised by using a security misconfiguration of an old Twitter

API.

  - **Key learning**: API and their governance are important for controlling data exposure and data sprawl.

- **Oct 2022, Microsoft Azure Blob**: A misconfiguration in Microsoft Azure blob exposed the PII of around 158K users. Microsoft did not make any formal notifications to the affected consumers.
  - **Key learning**: Whether right or wrong, judgment on materiality must be made, clearly and by a competent authority in the company.

- **Dec 2020, SolarWinds**: A massive supply chain attack wherein the source code of a major IT software player was injected with malicious code that could impact millions of customers. The malicious code was inserted into the Orian software of SolarWinds, allegedly by a Russian threat actor group, **Cozy Bear**.
  - **Key learning**: Organizations must ensure proper control over their source code and its release management. More importantly, the USA's **Security and Exchange Commission** (**SEC**) directly named Tim Brown, the CISO of SolarWinds, in a lawsuit for misdirecting the investors about the security posture. While not going into the merits/demerits of this case, the need for CISOs to take personal liability cover to protect against such lawsuits came into vogue.

- **May 2017, Equifax**: PII for about 159 million users and about 200K credit card numbers were breached, allegedly by Chinese perpetrators by leveraging an unpatched vulnerability in the Apache Struts. The patch for vulnerability CVE-2017-5638 with a CVSS 8.7 has been available since Mar'17.
  - **Key learning**: Ensure a robust risk-based patching process is developed and implemented.

The time an attacker is on the organization's network is called **dwell time**. During this time, the attacker is likely to have the capability to launch an attack, expand the attack footprint, and cause harm to the organization. The organization must make all possible attempts to reduce the dwell time by faster detection and response mechanisms.

Lessons for CISOs and IR team

The IR team and the CISO undergo a lot of stress during and even days following an incident. Some of the tactics that may come in handy while under a cyber-attack include:

- Keep calm and do not panic.
- Stick to basics, especially when under attack.
- Consciously ensure rotation of personnel at IR to help reduce burnout and to ensure fresh perspectives are possible.
- Ensure a standby team, trained earlier, is available to support.
- Ensure a retainer service with the external incident handler, forensic examiners, and external legal counsel is available.
- Keep the contact details for suppliers and customer security teams as updated as possible.
- Ensure to take some breaks and, if possible, use practices like meditation to center your thoughts and to ensure stress does not negatively impact your health.
- The crisis communication plan, along with the person responsible, must be well prepared, rehearsed, and even simulated during normal times. For instance, in the event of ransomware, the organization's strategy to pay or not, who can make the transaction and how, and up to what extent, must be carefully discussed.
- Incident and breach disclosure has a significant implication with regulators; the IR team, in partnership with senior management, must ensure a clear plan for how the materiality of the incident shall be determined and disclosed to appropriate authorities.
- Be careful of what is worded in risk statements, like the 10-K.
- Consider taking personal liability cover for cyber decisions or ensure employment terms/benefits duly cover such risks.

Conclusion

In this chapter, we explored various aspects of logging and monitoring as a key control in detecting security events and incidents. We also examined three 3-core functions of NIST CSF 2.0, namely detect, respond, and recover, as key components of incident response. We studied several examples of logs and how to analyze them, including using MSSP and external digital forensics and legal counsel. A key component of handling incidents is to create an incident response program with cross-functional representation to document an IRP and evaluate and update it continually. Learnings from security incidents in the organization or those with other organizations may be used as key inputs to enhance the IRP and ensure the SIEM-SOAR is accordingly updated too.

In the next chapter, we will cover concepts around cyber resilience, its growing importance with Zero Trust, and growing trends of disruptive cyber incidents like ransomware attacks.

Key takeaways

Some of the key takeaways from this chapter are:

- In spite of security and privacy controls implemented, organizations may experience security incidents or even breaches depending on the maturity of the security controls in preventing, detecting, and quickly responding, as well as the motivation and caliber of the TA.
- Organizations must have trained and capable teams to ensure relevant security events are logged, monitored, triaged, and analyzed.
- Communicating relevant aspects of the incident to concerned internal and, where applicable, external stakeholders is a key element of incident management. Regulations around the world require the timely notification and handling of such adverse events.
- The sources of logs include applications, APIs, networks, cloud resources, identities, etc., and have relevant details to trace the chronology and accountability of *what happened* or even *what is happening*. It is pertinent that these logs are immutable for the chain of custody.
- Using SIEM and SOAR are foundations for automating triage, alerting, and taking immediate actions.

- Investigation may require performing **threat hunting** (**TH**) to determine any IOC.
- Using *Lockheed Martin*'s **kill chain** (**KC**) approach, the various stages of attack can be examined and relevant interventions planned.
- NIST CSF may be used to structure the logging, monitoring, and incident response plans. The core focus of an IR plan is to document and simulate/test the adverse scenarios, action plans for them, and connected communication elements so well that the concerned teams and stakeholders build muscle memory.
- Using clean and immutable backups, RTO and RPO may be met to continue the business as quickly as possible after a disruption.
- CISOs are now being required to answer to congressional hearings and are held accountable for cyber risk-related regulatory filings (for listed companies).

References

- **https://nvlpubs.nist.gov/nistpubs**
- **https://www.sans.org/digital-forensics-incident-response/**
- **https://www.changehealthcare.com/hipaa-substitute-notice.html**

CHAPTER 12
Cyber Resilience

Introduction

The digital world is under increasing trends of cyber-attacks that can cripple the operations of organizations. Therefore, they must implement appropriate controls to prevent security incidents, but also have the ability to detect and respond to cyber incidents. They must also be able to ensure a return to normal functioning as quickly as possible and as strongly as possible.

Structure

The chapter covers the following topics:

- Introducing cyber resilience
- Implementing a cyber resilience program

Objectives

By the end of this chapter, you will explore the concepts of cyber resilience and differentiate them from the traditional **business continuity and planning** (**BCP**) and **disaster recovery** (**DR**) programs. Resilience ties very strongly with fundamental accounting principles that expect a business to continue to run under all conditions and continue to demonstrate appropriate financial discipline. For this chapter, we will focus on cyber resilience.

Introducing cyber resilience

In the previous chapters, we explored several security technologies and the various threats they can face. These threats may cause security incidents and require careful and planned action to respond to and recover from them.

Case study of MGM Grand breach

MGM Grand, a 1.4 trillion annual revenue company, runs several high-end hotels and casinos. In September '23, its Las Vegas, USA properties were targeted by two **attacker groups** (**AG**) called **Scattered Spider** and **ALPHV**. The threat actors from Scattered Spider used the following method:

- Called the company's helpdesk impersonating as an employee and tricked them into resetting the access of the employee's account. The helpdesk did not realize they had been subject to social engineering.
- Used that access to gain elevated privileges to MGM's IT environment, including its **cloud service providers** (**CSPs**): **Okta** and **Azure**.
- Attackers deployed ransomware using services from the AG called ALPHV, encrypting 100s of VMWare ESXi servers that were core to the MGM operations. MGM refuses to pay ransom, as confirmed by the AG.
- MGM IT shuts down its Okta sync servers, causing further authentication issues, though the attacker already had access to the servers in MGM's Azure cloud.
- MGM Grand slot machine, check-in/check-out, and digital access to its hotel rooms, digital apps, and corporate website were all disrupted/offline for over 10 days.
- MGM suffered a direct estimated loss of about $100 million.

MGM had been subject to a cyber-attack in 2019 as well.

Note: The case suggests that basic security hygiene, such as awareness, still plays an important role in security. Possessing the capability to use threat intelligence better and to have a strong incident response playbook may have helped MGM react and respond better. MGM may have

been able to recover to its normalcy sooner.

Case study of CrowdStrike disruption

In July'24, customers using CrowdStrike Falcon **endpoint detection and response** (**EDR**) on Windows machines experienced the infamous **Blue Screen of Death** (**BSOD**). CrowdStrike periodically updates a sensor configuration file, which it calls **channel files**, and deploys it to all Falcon-managed endpoints. Channel file 291 with a timestamp of 2024-07-19 0409 UTC had a validation logic error. Customers who used Falcon version 7.11 and received this faulty version experienced the BSOD. The Falcon sensor, which runs as a kernel process, triggered a read of the memory location that was out of bounds for it and caused the crash.

CrowdStrike, within a couple of hours, fixed the faulty channel file 291, and several computers got the update and survived the BSOD. CrowdStrike and Microsoft provided workarounds to bypass by deleting the problematic channel file and allowing the Falcon sensor to get the corrected file after reboot. Some computers running encryption technology like BitLocker faced further issues in accessing the channel file. The worst-affected were servers running Windows with EDR installed.

Organizations across all sectors were impacted, such as hospitality, where hotel check-in and checkout were not working; Airlines, where check-in and boarding pass issues were disrupted; schools with digital classrooms could not function; manufacturing; professional services companies, and so on. Most organizations and institutes had to temporarily run operations manually. Several computers were dysfunctional for days.

Several fake/hoax fixes emerged, and threat actors pretending to be IT companies emerged to support organizations in this time of crisis. There is not much data available on how successful those hoaxes were. Both CrowdStrike and Microsoft provided good updates on their sites and their support network, and that could have been the reason.

Note: In an interconnected digital world, disruption can come from any side. Organizations must plan and carefully choose their digital environment and have mechanisms to build relevant scenarios into their incident management and threat intelligence.

CrowdStrike subsequently improved its channel testing and release process and allowed the organizations to choose to defer the auto-update by a few hours to prevent such disruptions.

The word **resilience** has its origins in the Latin language and means jump back or spring back. In the digital world, it would mean the organization should be able to bounce back to its functioning. According to NIST, *Cyber resilience is the ability to anticipate, withstand, recover from, and adapt to adverse conditions, stresses, attacks, or compromises on systems that use or are enabled by cyber resources*.

Both the case studies introduced earlier in the chapter highlight the need to anticipate adverse situations and prepare to deal with them quickly and effectively. We will discuss more methods to achieve that in this chapter.

In the increasing digital footprint, the attack surface has increased as well. There are more possible avenues for cyber-attackers to target and bring harm to the organization. The reputation of an organization takes a beating when a cyber-attack happens, and thus, the longer it takes to return to normalcy, the larger the dent on customers' trust. It is difficult to maintain and build trust. A key financial accounting principle that organizations follow is that, notwithstanding any adverse events, the organization must continue to run as favorably as possible and continue to grow.

There are several types of cyberattacks against which organizations need to prepare to protect, detect, respond to, and recover from. These attacks include:

- Ransomware
- **Distributed denial of service** (**DDoS**)
- Supply chain attacks/risks

About ransomware

One of the most common threats that looms in today's world is **ransomware**. A ransomware attack tricks the user, using techniques called **social engineering,** into providing access to an information asset. Subsequently, the attacker moves laterally within the network while encrypting the devices with his/her own keys. The attacker's objective is to encrypt all data on as many machines and servers as possible so that the business is unable to use them. The attacker then makes a ransom demand, usually in **cryptocurrency** such as **Bitcoin,** to unlock the data**.** The organization has a choice of getting the data decrypted by paying the ransom. Unless a thorough analysis, including a forensic one, is done, it is difficult to determine if the attacker accessed and/or **exfiltrated**, i.e., took any data out**.** The MGM case study highlighted how a social engineering tactic was used by the TA to bring MGM operations to a grinding halt and cause financial loss.

Tip: In some countries, such as India, there may be legal implications to make ransom payments, especially payment by cryptocurrencies.

In the past, threat actors have used the ransom received to perpetrate crimes against humanity, such as terrorist activity, drug trafficking, and interfering in local governance. They have also used the ransom for financial gain for themselves.

Once the ransomware hits, it may take several days/weeks/months before normalcy can be restored, and yet the question of the integrity of data may continue to loom large. Organizations may choose to reinstate their good copies of backup to get back to normalcy; however, the integrity of the backup itself needs to be examined. In an **advanced persistent threat** (**APT**) type of ransomware attack, the backup itself may also have the malware that the attacker can trigger. This possibility ties very well with the **Zero Trust** (**ZT**) model, where the approach is to **never-trust–always-verify** and **assume breach**. We covered the ZT model in *Chapter 9, Zero Trust.* The time an attacker remains in the organization's environment is called **dwell time.** Organizations must have strong detection and response capabilities to keep any such potential dwell time to a minimum. In the case of MGM, the attacker's dwell time was reportedly 10 days. If MGM had a more efficient IR process and plan, this dwell time might have been reduced.

The **MITRE ATT@CK** framework defines the types of ransomware based on **techniques, tools, and procedures** (**TTPs**) as:

- **Data encrypted for impact**: Wherein the data files, system files, etc., are encrypted to render them unusable. The encryption may be triggered by a malicious payload on a website or an email link. The threat actor may also use **ransomware-as-a-service** (**RaaS**), such as **Conti**, to rapidly encrypt files, including executable files. In such cases, the information asset's **availability** (CIA triad) is most commonly impacted. The recent and infamous examples where the **adversary**

group (**AG**) used Conti are:

  - In Costa Rica (Apr 2022), several of the government's departments were compromised. The government refused to pay the initial ransom, and subsequently, the data of its citizens were exposed.
  - Even private organizations like *Shutterfly* in December '21 and Snap-on in March '22 were targets of AGs using Conti. The PII, including partial credit card information, was compromised. They subsequently notified their respective customers.

  Conti uses phishing emails with malicious attachments to target its victims and then uses weak protocols like SMB to move laterally. Using security controls and principles covered previously in the book, the Conti attack can be detected, prevented, and responded to.

  Some of the other widely used techniques include:

  - Selling/supplying an infected **point of sale** (**POS**) device, which then connects to the network and is compromised for ransom.
  - The AG Scattered Spider uses techniques such as enumerating AD accounts and then attempting password spraying, resetting MFA where set, to gain initial access, and then deploying their ransomware, especially on VMWare ESXi.

  Other RaaS include:

  - **Hive**: The AG, allegedly targeted organizations running Microsoft Exchange Server around April'22. They also have a portal for other AGs to post their attack campaigns and directly leverage already compromised victims for a fee. They also mediate the ransom payment on behalf of the attacker.
  - **Darkside**: It was used by the AG in the Colonial Pipeline breach in May'20, disrupting power supply to several parts of the East Coast of the USA.

- **Inhibiting recovery**: Often, software has the ability to recover from sudden outages/machine reboots. For instance, the Word or Excel files, when abruptly shut down, provide a recovery option when the application is opened again. Similarly, when a machine abruptly crashes, the built-in boot process may check and reload the required libraries, making the device functional again. The AG will target disabling any such ability to recover files. The AG may simply delete online backups and demand a ransom for their restoration.

- **Locker ransomware**: The AG completely locks out the ability of the user to log in to any system, thus targeting the availability of the information system. The AG may or may not encrypt the computer's storage. In 2019, a Norwegian energy company was targeted. **LockerPin** is a ransomware for mobiles that locks Android devices out by changing their PINs.

In cases of these successful ransom attacks, the AG demands money in exchange for restoring normalcy. They often extort money by threatening to make the unencrypted copy of their victim organizations' data public. Such a disclosure may immensely disrupt the organization, its customer trust, and even its stock price.

When a ransomware hits an organization, they would need to quickly contain the damage and ensure the resiliency plan is activated. This plan may involve quickly creating and/or moving to a safe and clean environment using verified clean backups. The other main question is if and how the ransomware will be paid, and how and who will communicate with the impacted stakeholders.

Distributed denial of service

The threat actor shall deploy means to disrupt the normal functioning of an organization, such as preventing its services from being available to legitimate users and thereby damaging the **availability**. The usual targets for a **denial of service** (**DoS**) attack are email systems, DNS, IAM systems, perimeter devices like routers and firewalls, and e-commerce portals. This DoS attack aims to achieve its objective by overwhelming the processing and handling capacity of the system by:

- Sending several data packets in quick succession, such that the system cannot handle them anymore. Often, if the appropriate protection and error handling are not done, the system becomes vulnerable and may stop functioning correctly. For instance, a flood of ICMP attacks (called a Syn-flood attack) on a firewall may overwhelm it such that it starts to allow any-any traffic. This would give the threat actor an unauthorized entry into the organization (affecting **confidentiality**). Alternatively, the firewall stops processing any further requests and thus disallows even legitimate requests from users. This impacts **availability**. Similarly, VPN devices may be constantly subjected to negotiating the SSL connection, thus disrupting the users' use of it.

  Modern-day firewalls have built-in capabilities to handle any such flooding of data packets, like ICMP, DNS, etc.

- Exploiting unpatched vulnerabilities and disrupting the services, or by gaining privileged access (privilege elevation), and then gaining unauthorized access. For instance, by compromising.
- Endpoint DoS attacks target user devices and disrupt their ability to consume IT services. For instance, an attacker can gain access to a management executive's computer and delete the device certificate, inhibiting the machine's ability to connect to corporate email/network, etc.

A DoS attack is triggered by a single source. A DDoS wherein multiple sources or attacks spread across the network are used to attack. Such a method is used for reasons such as:

- Increase the sources of attack from geographically spread locations. For instance, an organization may have protected its inbound traffic to its e-commerce portal from specific countries/IP pools of bad reputation, but it may have inadvertently allowed traffic from a region without proper checks.
- To leverage botnets that can generate a large amount of overwhelming traffic.

Modern-day security defense in depth layers already provide for DDoS, including its enrichment of active threats using threat intelligence. Organizations should consider building and maintaining an independent resilience environment or **isolated recovery environment** (**IRE**), which would be isolated and ready to be activated in the event of a cyber attack. The IRE typically uses a different Active Directory and all the bells and whistles of controls needed for key functions. As an analogy, bunkers built around the war zone are equivalent to an IRE. These bunkers are on standby with minimal facilities, but are immensely useful for the safety of people/soldiers. Similarly, airplanes are equipped with **Auxiliary power**

units (**APU**), should all their engines fail in mid-air. The APU powers the basic air, ventilation, light, and key navigation functionalities of the plane in the event of power failure.

Supply chain attacks/risks

The CrowdStrike case study is a good example of a supply chain-related risk where a malicious threat actor was not involved, and yet millions of computers were disrupted, causing hardships to even air travel. The digital ecosystem of several organizations, the world over, was disrupted within minutes. To CrowdStrike's credit, they quickly corrected their gap and made their internal process stronger for the future.

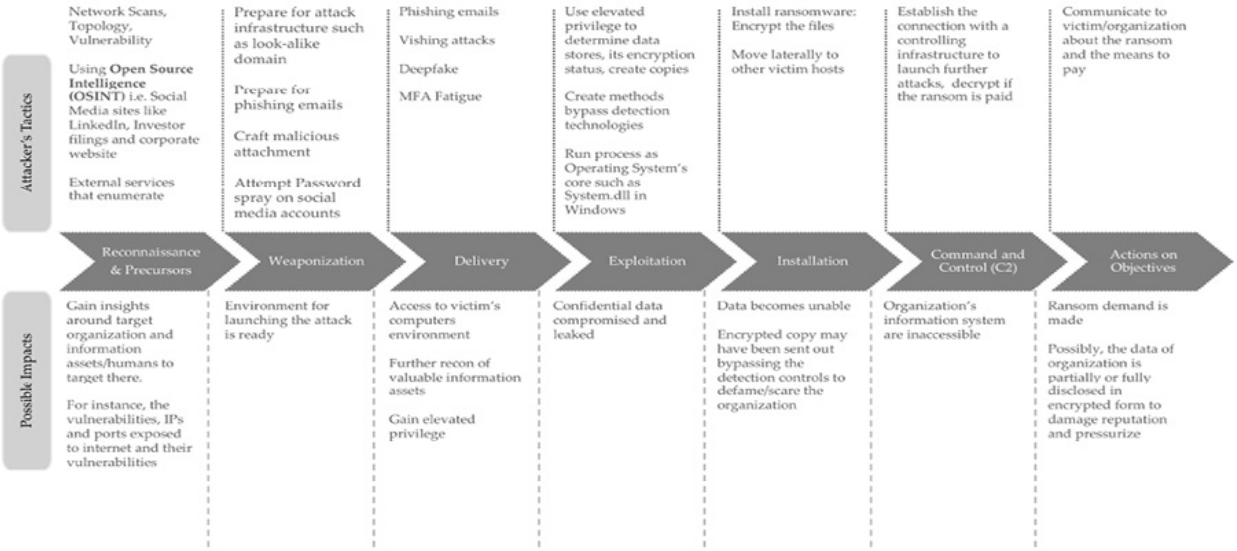
The SolarWinds breach was a significant event in cybersecurity history. The organizations, such as software vendors, professional service providers, and those connecting to customer organizations, came to realize that they may be subject to a cyber-attack or may be used as a conduit for a cyber-attack. Organizations started focusing on the **software-bill-of-materials** (**SBOM**) to ensure visibility of possible threats.

Some of the breaches, like the Dec'13 one of Target Corporation, caused by a compromised system of a third-party **heating, ventilation, and air conditioning** (**HVAC**) maintenance supplier, made organizations start focusing with greater intent on their connected network/remote access users and on **third-party risk management** (**TPRM**). The visibility into the suppliers used, their security practices, and the risks from them is a key ingredient to understanding the threat landscape.

Organizations continue to get breached, and how that impacts your organization is another element to consider in checking, validating, and taking corrective/preventive actions to safeguard your organization's environment.

Business processing organizations (**BPOs**) and such service providers often are in the front and center of several customer organizations. These organizations are suppliers to other organizations. Ensuring the customer organizations can trust the security of your service is key to ensuring their resilience.

Figure 12.1 illustrates a successful ransomware attack using the kill chain representation:



	Reconnaissance & Precursors	Weaponization	Delivery	Exploitation	Installation	Command and Control (C2)	Actions on Objectives
Attacker's Tactics	Network Scans, Topology, Vulnerability Using **Open Source Intelligence (OSINT)** i.e. Social Media sites like LinkedIn, Investor filings and corporate website External services that enumerate	Prepare for attack infrastructure such as look-alike domain Prepare for phishing emails Craft malicious attachment Attempt Password spray on social media accounts	Phishing emails Vishing attacks Deepfake MFA Fatigue	Use elevated privilege to determine data stores, its encryption status, create copies Create methods bypass detection technologies Run process as Operating System's core such as System.dll in Windows	Install ransomware: Encrypt the files Move laterally to other victim hosts	Establish the connection with a controlling infrastructure to launch further attacks, decrypt if the ransom is paid	Communicate to victim/organization about the ransom and the means to pay
Possible Impacts	Gain insights around target organization and information assets/humans to target there. For instance, the vulnerabilities, IPs and ports exposed to internet and their vulnerabilities	Environment for launching the attack is ready	Access to victim's computers environment Further recon of valuable information assets Gain elevated privilege	Confidential data compromised and leaked	Data becomes unable Encrypted copy may have been sent out bypassing the detection controls to defame/scare the organization	Organization's information system are inaccessible	Ransom demand is made Possibly, the data of organization is partially or fully disclosed in encrypted form to damage reputation and pressurize

Figure 12.1: *Kill chain representation of a ransomware*

An organization should prepare for the various attack types defined and build controls around them. However, to be resilient, it would also need to be prepared for the worst and focus on bouncing back with a clear, tested, and practiced resiliency plan. It would help the organization to adapt to the changing threat landscape and restore normalcy quickly.

BCP and DR

Cyber resilience is complementary to a BCP and DR. Several natural and man-made events can cause disruptions to business and thus a short-term or long-term crisis. Some of such events include:

- Natural disasters like floods, earthquakes, lightning strikes, outbreaks of pandemic/epidemic diseases, and fires.
- Man-made ones like arson, riots, strikes, political unrest, and cyber-attacks.

A DR plan is built in such a way that there are alternate IT data centers/recovery possibilities, backup, etc., that will be initiated in the event of a disaster/crisis. The disaster/crisis implies the business can no longer fully function from its current place/environment. The DR Plan would include aspects like what to fail over to, when, and how. It will also cover aspects like how to build. We previously explored the terms **recovery time objective** (**RTO**) and **recovery point objective** (**RPO**) in the context of backups.

A BCP caters to all aspects such as:

- **People**: This involves identifying personnel critical to functioning, their core skills, their ability to relocate to alternate sites, their contact information, etc.
- **Process**: This involves identifying key business functions, their operating procedures, service levels, and so on. A formal method called the **business impact analysis** (**BIA**) is generally used to determine the critical processes and aspects, such as RTO/RPO. It is important to understand that shorter RPO/RTOs may be desired, but may come at a high cost.
- **Technology**: This involves identifying key technology tools/systems/applications that might be required, such as emails, e-commerce portals, shopping carts, and storage systems.

A BCP plan should be periodically tested to understand how well the organization is prepared to deal with any untoward events. A BCP/DR can be triggered due to a cyber event as well, such as a widespread attack disrupting all endpoints with malware.

A common practice to contact the key personnel listed in a simulated scenario, called a call tree test, is often performed to gauge the reachability of personnel and if there should be alternate personnel identified for emergencies.

Cyber resilience can be thought of as an overarching program that ensures proper detection of cyber incidents, the wherewithal to respond to them, and, if needed, recover from them as quickly as possible, causing minimal disruption. The cyber-related parts of the BCP/DR plan may be subsumed into a cyber resilience plan. However, the BCP/DR plan may still be relevant for non-cyber-related events.

Figure 12.2 illustrates the relationship between the BCP, DR, and cyber resilience:

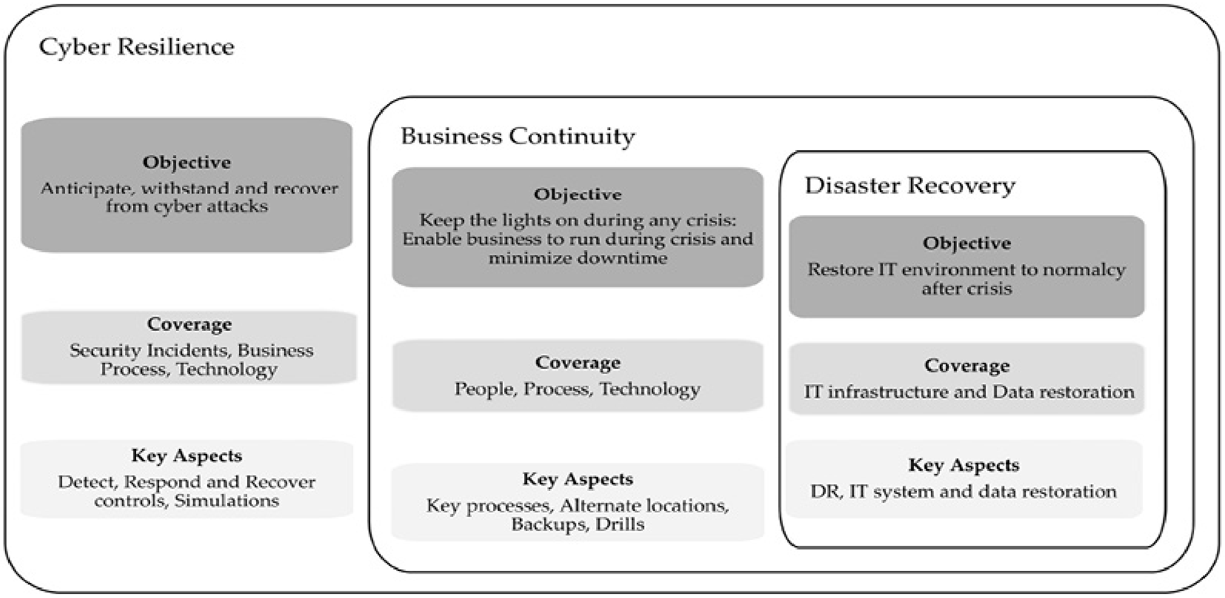

Cyber Resilience
Objective
Anticipate, withstand and recover from cyber attacks
Coverage
Security Incidents, Business Process, Technology
Key Aspects
Detect, Respond and Recover controls, Simulations
Business Continuity
Objective
Keep the lights on during any crisis: Enable business to run during crisis and minimize downtime
Coverage
People, Process, Technology
Key Aspects
Key processes, Alternate locations, Backups, Drills
Disaster Recovery
Objective
Restore IT environment to normalcy after crisis
Coverage
IT infrastructure and Data restoration
Key Aspects
DR, IT system and data restoration

***Figure 12.2**: Relationship of cyber resilience to BCP and DR*

Implementing a cyber resilience program

A cyber resilience program is fast becoming a key focus of business strategy. Following an incident such as a supply chain disruption or an attack, millions of $ are at stake. A cyber resilience program should focus on the following aspects:

- **Planning and assessment:**
  - Identifying and managing risks appropriately: This implies being deliberate about controls, such as:
    - Backup and recovery of key information assets, such as configuration files and data files.
    - Using ZT principles-based architecture for identity, phishing-resistant MFA, network segmentation, data access, etc.
    - Performing a BIA to proactively identify the key risks of disruption to business operations, key functions that would need to run even during such a disruption (or shortly after), and gather relevant insights for planning the recovery strategy from a capacity, capability, cost, time, interdependencies, and technology perspective. For instance, RTO and/or **maximum**

tolerable period of disruption (**MTPD**) can drive the determination of how soon the recovery must happen. The BIA should cover impacts due to regulatory, reputational, social, environmental, business output, and financial dimensions. The approach to performing a BIA is like that of **risk assessment** (**RA**) covered in *Chapter 2, About Managing Risks*. BIA aims to identify critical processes and their interdependencies, while the RA focuses on likelihood and probability.

- **Detection and response**: Strengthening logging, monitoring, and threat intelligence to ensure detection capabilities are strong and quick. For instance, if MGM had better means to detect and report the social engineering attack, it might have quickly been able to contain the event.
- **Recovery and adaptation**:
    - Defining, documenting, and testing:
        - **An IR plan**: To enable an expedited response time in managing an incident and containing its fallout. For instance, if MGM had an IR plan for ransomware that it may have simulated and practiced, there may have been chances of faster recovery to normalcy.
        - **A BCP/DR plan**: To enable the business to recover its operations in parallel, even when a crisis is going on.
- **Testing, training, and simulating**:
    - Perform control validation and execute required refinements to ensure weaknesses in controls have not occurred from a design and implementation standpoint. We covered these in *Chapter 3, Role of Standards and Controls*. For instance, the SolarWinds breach, covered in *Chapter 11, Incident Response and Planning,* highlights the need for stronger controls on an organization's source code.
    - Defining the scope and strategy for crisis management. The organization must be clear in articulating and knowing what constitutes a crisis and what business processes, infrastructure, and region will be in scope at that time. For instance, during the Change

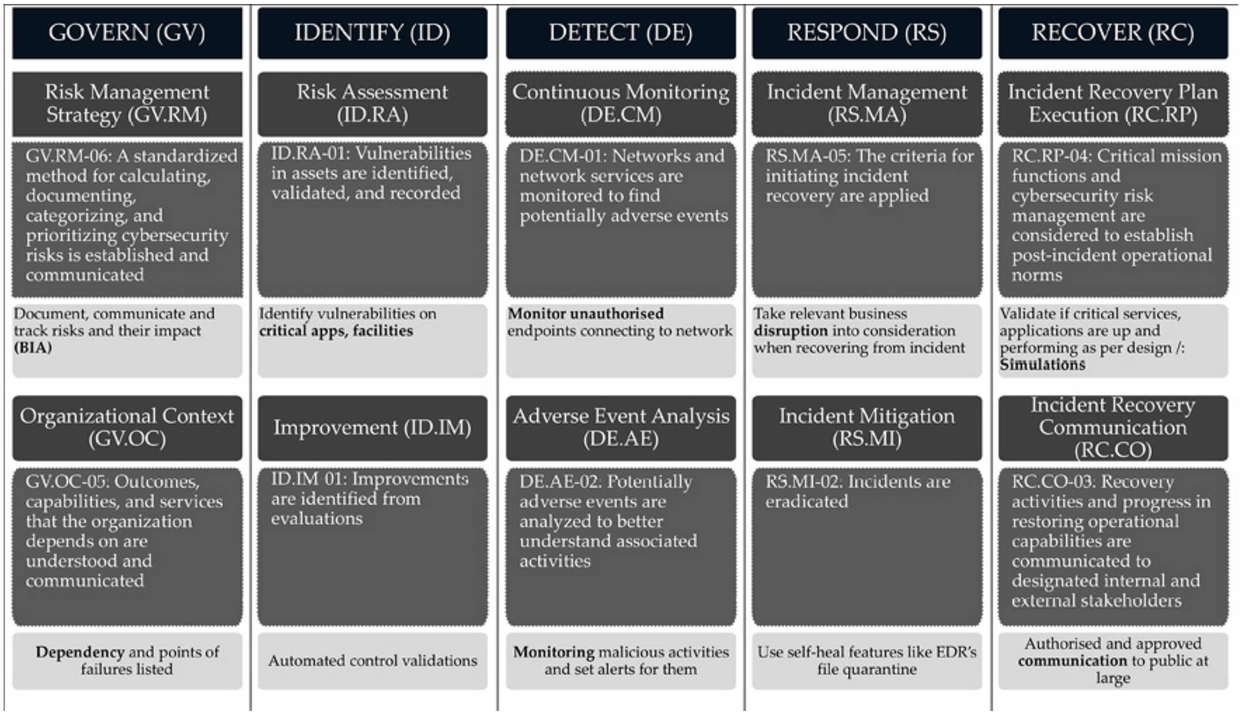
Health's breach in February '24, the organization may have benefited from better documentation and rehearsal for proactive anticipation of a breach.

- The crisis management team should be cross-functional with active representation of top management, HR, legal, finance, corporate communications, apart from real estate service, business operations, IT, and information security. The structure should have a clear line of primary and secondary authority to enable faster decisions and reduce confusion. Ideally, various types of crisis simulations should be done across the organization. For instance, simulating a ransomware attack:
    - With the top management of the company, build a muscle memory that the executives will need to think and act.
    - With IT and IR teams, it may involve what to do in case of ransomware, how an alternate clean IT environment will be set up, and how soon that can happen.
    - With the business operations teams to gauge which process needs to recover first, and how. Does the team have the relevant capability to recover in the absence of key personnel or IT setup?
    - With investor relations, corporate communications, and branding teams to prepare and plan what should be communicated, how, and what the impacts are to key considerations like regulations and the stock market.
- Ensure stakeholder engagement across the organization, including employee awareness efforts. Often, humans are the weakest link in a cyber-attack; by strengthening their awareness of topics like social engineering, which is covered in the next chapter, organizations can reduce the chance of an incident/breach.

Note: A strong cyber resilience program is increasingly being seen favorably by insurance providers and the boards of companies.

Organizations considering a strong resilience program may have to even think and prepare for alternate AD, single sign-on, email systems, and alternate out-of-band communication platforms in the event of a crisis, especially with ransomware. Such an environment constitutes the IRE introduced earlier in the chapter. All of this will cost money, and the organization will need to take a risk-based approach that suits them.

Organizations may adapt any framework for their resilience objective as long as the focus is on anticipating threats and being ready to detect, respond, and recover from them. *Figure 12.3* illustrates a suggested resilience framework using NIST CSF2:



GOVERN (GV)	IDENTIFY (ID)	DETECT (DE)	RESPOND (RS)	RECOVER (RC)
Risk Management Strategy (GV.RM)	Risk Assessment (ID.RA)	Continuous Monitoring (DE.CM)	Incident Management (RS.MA)	Incident Recovery Plan Execution (RC.RP)
GV.RM-06: A standardized method for calculating, documenting, categorizing, and prioritizing cybersecurity risks is established and communicated	ID.RA-01: Vulnerabilities in assets are identified, validated, and recorded	DE.CM-01: Networks and network services are monitored to find potentially adverse events	RS.MA-05: The criteria for initiating incident recovery are applied	RC.RP-04: Critical mission functions and cybersecurity risk management are considered to establish post-incident operational norms
Document, communicate and track risks and their impact **(BIA)**	Identify vulnerabilities on **critical apps, facilities**	**Monitor unauthorised** endpoints connecting to network	Take relevant business **disruption** into consideration when recovering from incident	Validate if critical services, applications are up and performing as per design /: **Simulations**
Organizational Context (GV.OC)	Improvement (ID.IM)	Adverse Event Analysis (DE.AE)	Incident Mitigation (RS.MI)	Incident Recovery Communication (RC.CO)
GV.OC-05: Outcomes, capabilities, and services that the organization depends on are understood and communicated	ID.IM 01: Improvements are identified from evaluations	DE.AE-02: Potentially adverse events are analyzed to better understand associated activities	RS.MI-02: Incidents are eradicated	RC.CO-03: Recovery activities and progress in restoring operational capabilities are communicated to designated internal and external stakeholders
Dependency and points of failures listed	Automated control validations	**Monitoring** malicious activities and set alerts for them	Use self-heal features like EDR's file quarantine	Authorised and approved **communication** to public at large

Figure 12.3*: NIST CSF 2 and some examples for a cyber resilience program*

Emerging trends for improving resiliency

As discussed in earlier chapters, AI is likely to play a huge role, especially in monitoring, validating controls, detecting, and even responding and recovering from cyber threats. Several of the key players in cybersecurity products are offering features enabled by AI, such as:

- **Planning and assessment**: AI may also play a role in using historical trends to suggest critical processes and help expedite steps like BIA.
- **Detection**:
  - Palo Alto Networks offers an AI-powered capability to perform threat hunting and determine security risks based on the

log/telemetry from various sources.

  - Other security products like *SentinelOne*, *CrowdStrike*, *Rapid7*, and *Qualys* also have AI-powered capabilities to detect security anomalies and configuration weaknesses. They use AI to connect the threat intelligence feeds to the contextual relevance to the organization. Similarly, email security products like Mimecast and Proofpoint are advancing their AI/ML engines to power malicious email detection.

- **Training and simulation**: AI is also used to generate more contextual phishing emails to help train employees using simulations.

AI will also be used by threat actors; therefore, it is key to continually focus on adapting threat intelligence.

Quantum computing is also expected to radically change the world, especially how encryption will need to be done. This has a direct impact on the resilience as well because the current algorithms may no longer be safe. For instance, the backups would now need to use quantum-safe cryptography algorithms.

Conclusion

In this chapter, we explored the concept of resilience in depth through case studies and examples of cyberattacks such as ransomware. We examined the relevance of resilience to customer trust, insurance, and foundational accounting principles of continued growth. While we leveraged the NIST CSF 2.0 framework to base our comprehension of the concept, other approaches could be adopted as well.

In the next chapter, we explore the importance and relevance of controls around people and their awareness of security-related behaviors. We will also explore some of the common security attacks focused on employees and contingent workers.

Key takeaways

The key to resilience is in identifying, responding to, and recovering from adverse events. Some of the learnings include:

- The better we can anticipate and/or simulate, the better the foundation of resilience.
- Real-life breaches like that at MGM Resorts, or the outage caused by the CrowdStrike channel file, are evidence that disruption may happen anytime to anyone.
- Disruption may be due to cyber attacks, operational issues, and/or supply chain-related disruptions.
- As trends of ransomware and other such threat scenarios grow, organizations must prepare for the worst.
- They must create and maintain the infrastructure and capabilities in technology, process, and people to bounce back as quickly as possible, such as an IRE.
- Some of the efforts in resilience would include an out-of-band communication channel, as the current potentially infected ones may be untrustworthy or unavailable.
- To plan for bouncing back effectively, organizations must perform a **business impact assessment** (**BIA**) and determine the criticality and other priorities of processes to enable.
- A clear line of thought with an empowered and trained cross-functional team should be identified to call the shots in times of crisis.

References

- **https://csrc.nist.gov/glossary/term/cyber_resiliency**
- **https://www.recordedfuture.com/threat-intelligence-101/cyber-threats/types-of-ransomware**
- **https://www.rsa.com/resources/videos/cyber-resilience-in-the-age-of-ai/**

https://discord.bpbonline.com

CHAPTER 13
Human Centric Security

Introduction

Notwithstanding the advancements in technology, such as quantum computing, artificial intelligence/machine learning, autonomous vehicles, and practically anything available under software-as-a-service, the human mind and behavior continue to be at the forefront of cyberattacks that happen. Often, humans are the weakest link that becomes a conduit for the success of the attack. The psychology of how the human mind thinks and reacts to an external stimulus is an interesting dimension in protecting an information asset. There are some regulatory or business requirements for security awareness. In this chapter, we will examine a human centric approach to driving a security-focused culture.

Structure

The chapter covers the following topics:

- Human centricity and role of security culture
- Building security culture

Objectives

By the end of this chapter, you will explore the concepts of security culture, i.e., enabling the organization to build muscle memory for identifying and dealing with security threats at all levels. We will go beyond the security awareness messages and training and explore additional avenues to engage the workforce as a line of defense in detecting and responding to security incidents. We will attempt to understand the basic psychology of the human mind and use that to build a security culture. We shall also use the learnings of some of the past security incidents in this approach.

Human centricity and role of security culture

Every human is wired differently; they tend to think and behave differently. In a largely interconnected world, powered by technology, the choices we make as humans may impact the entire ecosystem. For instance, when a user clicks on a phishing email with a malware attachment, he/she might compromise not just the organization but also its connected networks. The extent of the damage depends on the layered security controls that we have previously discussed.

As the digital revolution continues to grow, it is pertinent for organizations to build a culture of security with not just their workforce, i.e., employees and contractors across all levels, so that security does not remain as a set of **Do's and Don'ts** but becomes a habit with rationality. Similar reasonable efforts may also be considered for suppliers and customers, wherever there is a direct impact.

Note: Several large banks run awareness messaging campaigns on television, print media, and even digital platforms for the public at large. This contributes to building a culture within the society and comes in handy. Even some regulators, such as the Reserve Bank of India (RBI), also run messaging on the safe use of digital apps and banking systems.

The culture of security aspires to a state where everyone understands the relevance of the CIA and participates with responsibility, which is key to reducing adverse security events and ensuring better cyber resilience. An unattended *chink in the armor* is all that the attacker needs to capitalize on. **Zero Trust** (**ZT**), though a technical term, can be thought of as a concept for the human mind as well, to **never-trust–always-verify** before making any judgment.

Organizations usually onboard a team or outsource the efforts to implement and drive the culture of security. These teams are commonly called the **security awareness team** or **human centric security team**. Traditionally, the security awareness team focuses only on awareness training and messaging, etc., whereas the human centric team expands its remit to build interlinkages of human behavior, threat intelligence, learning from incidents, and participating in policy making, such as **acceptable use policy** (**AUP**), running simulations, and so on.

Some facets of human behavior that can be used by the human centric team in modeling a security culture are detailed below.

Solomon Asch conformity line experiment

In the 1950s, *Solomon Asch* experimented with several groups of male students and asked them to call out which among the three lines he showed was the longest. As part of his experiment, he had planted all but one student in the group to give the same answer, though the answer was incorrect. His attempt was to see if the participant who had not been instructed decided for themselves and chose the actual long line or conformed with the majority opinion, even if the answer was incorrect. Upon repeating the experiment with multiple groups multiple times, he noticed that almost 75% of participants went with the group's opinion, even though the answer was visually obvious as incorrect.

While the experiment may have lacked aspects like diversity of gender and age, etc., it did indicate that a significant majority of people in groups may make choices based on what their peers did. We will refer to this practice as **conformity behavior**.

Note: The human centric team can use this psychology to design security awareness and culture efforts, for instance, by driving a practice of always reporting any suspicious email, or by always encouraging the use of ID cards to badge-in, and dissuading tailgating.

Role of biases

The preference, prejudice, or inclination for choices is called a **bias**. Some examples of bias in judgment are:

- **Confirmation bias**: Where the individual seeks to reaffirm their judgment based on their experience or interpretation:

  - We often almost give in to **chain-mail**, where an email is recommended to be forwarded to 10 more people to ward off bad luck or to bring in special luck. Out of fear or greed, we tend to conform, often believing there is at least nothing to lose. It may not dawn upon us that we may be participating in the spread of misinformation, increasing unwanted traffic on the network, and providing a threat actor with validated emails to target.
  - As a Security log analyst or a **cyber threat intelligence** (**CTI**) analyst or even a risk analyst, you may only focus on evidence/data that you feel aligns with your thinking. They irrationally ignore the contradictory evidence. This implies that the analyst is not able to pivot to other possible root causes of the incident, possibilities of attack vectors, and avenues of risk.

- **Framing effect bias**: Where the choices/judgment made are based on the way the information is presented:
  - In the previous chapter, we covered the CrowdStrike outage. Fraudsters attempted to frame their email outreach to their victims as a critical upgrade to CrowdStrike, and may have fallen prey to allowing the threat actor to gain access to the network.
  - If the AI's **large language model** (**LLM**) is presented with incorrect data or decision patterns, it will learn the action incorrectly and therefore produce negative outcomes that will go undetected.

- **Optimism bias**: Where the individual does not believe the harm can come their way:
  - We may believe that identity compromise happens only to the rich, affluent, and the most important people, and average Joes do not. Therefore, it is ok to post one's own PII on social media for anyone to see. Ideally, such information should be visible only to the validated network, if at all.
  - The belief that there are so many security technologies implemented that no harm shall befall the individual personally. For instance, a normal user tends to believe the firewall will stop everything that is bad.

- **Negative bias:** Where the tendency is to pay attention to negative information:

  - For instance, an email that claims to disable user access in 24 hours unless a link is clicked to validate an active ID is likely to receive more clicks than an email that informs of an upcoming regular maintenance schedule with a link. Phishing emails leverage such an approach.
  - The ability to think rationally and make informed choices is sometimes overrun by **fear, uncertainty, and doubt** (**FUD**). During the COVID-19 pandemic, several fraudsters attempted to sell vaccines, medicines, and expedited medical assistance and supplies upon advance payment. Some fraudsters even posed as financial institutions, guaranteeing a safer **return on investment** (**ROI**) to next of kin upon the investor's passing away due to COVID-19. Fraudsters also attempted to collect funds to assist the underprivileged sections of society by posing as a local municipal body, Government departments, and reputable corporations. In all these scenarios of FUD, several people unfortunately did fall prey, as they did not validate the veracity of such approaches in spite of sources of truth being available in different media.

- **Information bias**: Where, due to a lack of awareness of information or a lack of it, may trigger flawed judgment:

  - Believing that sharing passwords is not a problem for business requirements, because it is the company's network and identity. We do not realize that apart from issues like accountability and repudiation, we might be giving a clue to how we design our password/passphrase, and thus making it susceptible to cyber-attacks.
  - As a CTI analyst, not being aware of the latest **techniques, tactics, and procedures** (**TTPs**) of threat actors, important detection, response, and recovery gaps may occur, causing a security incident or a breach.

- An **unconscious bias**, i.e., choices made as an impulse subconsciously, like an involuntary action, may additionally impact outcomes. For instance, clicking on the link in an email if it comes from the boss.

Figure 13.1 illustrates some of the bias that impacts security culture:

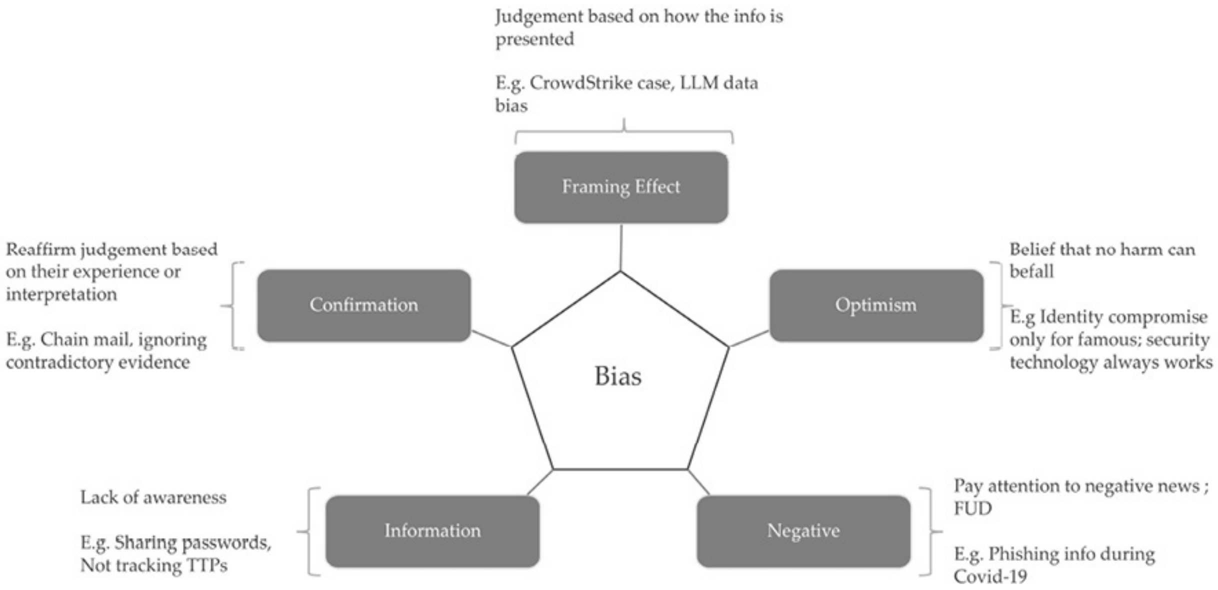

Judgement based on how the info is presented
E.g. CrowdStrike case, LLM data bias
Framing Effect
Reaffirm judgement based on their experience or interpretation
E.g. Chain mail, ignoring contradictory evidence
Confirmation
Bias
Optimism
Belief that no harm can befall
E.g Identity compromise only for famous; security technology always works
Lack of awareness
E.g. Sharing passwords, Not tracking TTPs
Information
Negative
Pay attention to negative news ; FUD
E.g. Phishing info during Covid-19

Figure 13.1: *Bias and effect on security culture*

Note: The human centric team can use such clues of biases to define the content and method of delivery of their efforts in shaping the security culture. For instance, to reduce FUD, provide more information with clarity and responsibility, or encourage a CTI analyst to gain more insights.

Attackers use these biases and behaviors to craft their attack techniques.

Other factors impacting security culture

Some of the other aspects that influence human actions are:

- **Upbringing**: The value system instilled during formative years is more likely to carry forward to adult life, too. For instance, a child who takes accountability for his/her actions, such as being punctual for class assignment submission, is more likely to report a security event. The human centric team can use such individuals, who report often, to champion the cause of timely reporting of suspicious and actual security events.
- **Choosing the easier route**: Humans gravitate to find easier ways that suit them, irrespective of policy. For instance:
  - An administrator may directly log in to the server to make changes without going through the **privilege access management** (**PAM**) system.
  - A user may choose to send a confidential PowerPoint deck to their

personal email ID to work from home.

    - A user working on multiple client engagements may keep passwords in a text file on the computer's desktop.

    In spite of the intent, such actions cause security incidents and are not representative of a good security culture.

- **Stress**: The increase in stress levels at work or at home may impact rational thinking and thus can be a risk to the CIA. For instance, an incident handler often needs to engage in conversations with employees to determine the root cause of the problems in a security incident. They may experience a range of responses, ranging from hostile and belligerent to emotional outbursts. This may cause stress in the individual and may result in them making biased decisions. The human centric team can partner with the HR team to provide interventions for de-stressing.
- **Lack of process or risk awareness**: The users may lack awareness of the right way to do things. For instance, a developer may not be aware that the organization's software code must not be uploaded to code repositories like GitHub. Similarly, a finance user may not be aware of how to classify documents so that the **data loss prevention** (**DLP**) tool can detect and block any inadvertent or willful transmission. The human centric team can study the persona of a typical user and design relevant byte-sized content for users to learn and use.
- **Disinterest/Disenchanted**: While several factors may contribute to a de-motivated or disinterested employee, there may be warning signs that can be used to nudge positive behavior or at least monitor it with the **cyber defense center** (**CDC**) team. For instance, an employee serving notice period is quite likely to make attempts to send project templates outwards, the human centric team can infuse appropriate messaging of acceptable use and on intellectual property.
- **Sense of power**: Organizations struggle with shadow IT, an IT team that is not part of the formal IT team and yet codes applications/bots or manages servers, etc. For purposes of convenience and speed to outcomes, these approaches work great. However, such teams often lack the discipline or awareness of aspects like secure configurations, proper change controls, strong passwords, etc. For instance, some in the

marketing team register a domain name and host campaign information containing customer outreach emails and contact details without properly protecting the application with strong authentication and/or even a **Secure Socket Layer** (**SSL**) certificate. Even the IT may choose to ignore the requirement of standards and hardening approaches, and have a misconfigured infrastructure from a security standpoint. Such activities increase and expose the attack surface.

The human centric team can partner with relevant internal teams to coach proper behavior and, if needed, remind such users of the disciplinary action process.

Some of the actions our workforce takes may be signs of **insider threat** and should be an aspect to drive better **visibility** using logs and relevant alerting through tools like SIEM-SOAR.

Cyber-attacks using the human element

Attackers use bias, gullibility, and other such behaviors to trick humans into divulging confidential or sensitive information and then use it. This is called **social engineering**. The commonly used social engineering techniques include:

- **Phishing**: Where emails are used to trick the user. Some examples of phishing email content include:
    - Prompting user to click on a link in an email to validate their identities or face email ID deactivation. When the user clicks on such a link, a look-alike website opens where the user enters the user ID and password; these are transmitted to the attacker for further use.
    - Prompting the user to click on a link and receive the latest security patch/or a free voucher for shopping. Upon clicking the link, malware may be installed and be the conduit of a ransomware attack.
    - Prompting the user to link and verify the correctness of the bank account address/nominee details or tax refunds. Upon clicking the link, which is usually a fake/look-alike website of the user's bank, the attackers get the credentials and misuse them.

  Human centric teams can use this to coach and train users not to click on any such links. There are technical solutions possible to warn users of such malicious links and prevent harm. However, such technical interventions may not be available, including their personal email ID.

- **Spear phishing**: In this form of phishing, the attacker targets a specific set of users in an organization, individuals with power, such as IT admins or individuals authorized to make payments. The attackers create a sense of urgency and secrecy while establishing trust with the recipient based on some facts. For instance, the mail would appear to be from a customer asking about outstanding payments, citing some invoice numbers, and asking the balance to be urgently paid to a new account.

  A spear phishing attack where senior executives are the target of the attack is called **Whaling**. For instance, the mail would appear to be from the **Chief Executive Officer** (**CEO**) directing the **Chief Financial Officer** (**CFO**) to urgently transfer funds to foreign (unknown) accounts for the merger being pursued by the organization. Or the CEO reminding the **Chief HR Officer** (**CHRO**) to click on the link to book their cricket match tickets.
- **Business email compromise (BEC)**: In this form of email-based social engineering, attackers pose as a senior executive and use a look-alike email ID to instruct the victim to divulge confidential information and/or make urgent payments to a different bank account. For instance, an email to the *accounts manager from the director of finance* in the company will look like *DirFin@MainC0mpany.com*, whereas the actual ID would be *DirFin@MainCompany.com*. Notice the use of the digit 0 in lieu of the letter o in the email domain name. Such nuances are not easy to pick. The attacker may also use the same email display name and signature style in the email to make it appear very authentic.
- **Smishing**: Phishing may also happen over SMS or other app-based messaging platforms, such as *Telegram*, *WhatsApp*, or *Signal*. The message content would be like the email-based ones.
- **Vishing**: In this type of social engineering attack, the voice channel is used to trick the user. In the MGM Grand breach of 2023, covered in the previous chapter, the attacker used a voice call to the helpdesk to gain access to the credentials of an employee.
- **Deepfake**: With advancements in technology, especially in AI, the possibilities of creating near-original video messages and targeting users

for action have emerged. For instance, the attacker would use AI to create a video of the CEO directing the director of accounts to make payments to the account mentioned in the video. The video would look quite real and convincing. Deepfakes have also been used to impersonate candidates at job interviews. The AI engine would respond to the interview questions with facial movements aligned. At the moment, it is still possible for the user to distinguish a fake video because of facial movements and the quality of imagery.

- **Multi-factor authentication (MFA) fatigue**: To overcome **account takeovers** (**ATO**) or the loss of credentials, organizations started implementing MFA for applications and other access, like VPN. Most MFA implementations require users to accept prompts on the MFA app and authenticate. Attackers bombarded the user of a compromised account with so many MFA requests, triggering their unconscious bias or the optimism bias, and thus, they accept all such requests.

Note: Phishing-resistant MFA technologies are now in vogue to counter the possibilities of MFA fatigue.

- **Other attack vectors**: In the urge to stay always connected, we may often be leveraging the free Wi-Fi at airports, railway stations, hotel lobbies, and so on. Malicious attackers can set up their Wi-Fi and/or compromise these Wi-Fi setups to capture the user activity, including their login to social media and bank accounts, and then misuse it. Similarly, technologies like **Bluetooth** (**BT**) and **Near Field Communications** (**NFC**) could also be used by attackers to compromise the laptop/handheld devices.

Note: As per the Verizon Data Breach Report (VDBR) 2024, 68% of reported breaches involved a human element.

Human centric teams can design messaging for the workforce to be aware of such attack vectors and tips to deal with such attacks, irrespective of the available technology controls. For instance, the organization may have controls to restrict inbound emails from IPs with a bad reputation or to protect email impersonation of key personnel; even then, the organization should still equip its workforce with the knowledge of such risks.

Building security culture

As the human centric navigates the various existing and emerging attack vectors, they need to ensure the efforts to build and manage a security culture over time. The efforts must be thought of as a program that aims to tweak the culture. Changes to culture are successful only when the workforce internalizes them.

Factors impacting a human centric security program

The factors impacting a human centric security program are as follows:

- **Top management support**: This program must be treated with importance and needs sponsorship at the highest levels to be successful. For instance, if the top management does not pay heed to attack vectors like phishing, the chances are the next layers of the organization will also be careless.
- **Scope**: While the interconnected digital ecosystem is a huge ecosystem, management direction and expectation are key in determining if the scope of the efforts must go beyond the organization's workforce to supplier-related awareness communication and the public at large. For instance, large software players do have training material for the public to use and also have targeted security messaging within the organization. In today's competitive world, organizations often have to expedite the release of their new service/product as quickly as possible. Given the scarcity of talent and the talent skill they often need to engage experts who just execute that specialized piece, such as training the LLM properly, and then go away. These **gig-workers** have access to information assets and must be treated cautiously and at par with contractors.
- **Environmental constraints**: The organization's regulatory environment will play a key role in the requirements. For instance, a financial institution such as a bank, a stock trading company, or a healthcare company will have far more regulatory requirements on what to educate the user and how to report incidents. In comparison, a standalone retail store chain would not be as regulated.
- **Organizational work culture**: It is important to recognize that every

organization is built differently and needs different things. For instance, the culture of invocation at pharmaceutical companies will be different from that at a startup in AI. It is likely that the technical controls will be more relaxed at a startup when compared to a pharma company. Similarly, some organizations may require restrictions for users to send and receive emails from outside.

- **User experience**: The workforce and the associated partner ecosystem are there to support the business objectives. The way they interact with technology/use it is a key determinant of how culture can be built. For instance, as the world continues to adopt AI, it is important for security teams to ensure that such access to AI models and platforms is possible. The human centric would need to ensure awareness of the appropriate and ethical use of AI.
- **Learnings from incidents and threat intelligence (TI)**: Incidents and TI are good sources to continually evolve the components impacting the culture. For instance, the human centric efforts must determine what biases are contributing to the DLP incidents and how to nudge the users to be cautious.
- **Technical changes**: It is also recommended to adapt technology for building culture as well. For instance, providing the developers with a platform to learn security technologies, and for the security risk team to practice log analysis and learn from it. Similarly, with phishing-resistant MFA gaining prominence, familiarize users with the tool.
- **Effort and budgets**: Building culture is a multi-year journey and would require time, effort, and money. It is important to structure the strategy accordingly to make incremental and yet meaningful progress. Driving accountability, especially security responsibilities, will take time because of different priorities, a changing workforce due to additions, attrition, and even role rotation.

Note: Remote working setups are increasingly being used by organizations. There have been a few cases of the same individuals working for more than one company at the same time without a declaration. This is called moonlighting and brings about additional risks to the CIA of information assets, and even potentially fraud. Human centric team, in partnership with the HR and other security teams, should consider appropriate awareness messages.

Components of the human centric security program

Each organization, based on the factors detailed above, would need to choose the way forward from the components listed here.

The suggested components of the human centric program are:

- **Training**:
    - **New hire orientation (NHO)**: Appraise new employees and contractors to cover topics such as:
        - Data classification and safe handling of data.
        - Acceptable use policy.
        - Reporting incidents.
        - Social engineering and physical security-related topics.
    - **Annual refresher training**: At an annual cycle, remind users of security concepts. Many standards and regulations require this as well. Suggested topics:
        - Data classification and safe handling of data.
        - Acceptable use policy.
        - Reporting incidents.
        - Social engineering and physical security-related topics.
        - Recent contextual topics, such as detecting and avoiding Deepfakes.
    - **Role-based training**: To equip the personnel in key roles with role-specific security concepts, such as:
        - **IT helpdesk:** Establishing identity of the requester.
        - **Administrators**: Importance of PAM.
        - **CDC personnel**: Human behavior impacting incident trends.
        - **Developers**: OWASP Top 10, GitHub security.
        - **Security administrators**: **Segregation of duties (SOD)**, identifying emerging threats.
        - **HR personnel**: Safe handling of HR files, **data subject rights (DSR)**.
        - **Accounts payable personnel**: Identifying phishing and BEC.

        - **Executive assistants**: Using the delegated rights responsibly, phishing.

    - **Topical security training**: To make users aware of key concepts with recent industry or environmental events, covering topics such as:
        - Relevance of cyber resilience.
        - Importance of encryption.
        - Importance of using AI responsibly.
        - Key learnings.
        - Phishing.

- **Awareness messaging**: Appraise the personnel using messages on key topics such as:
    - Reporting security incidents (why, how, and what to report).
        - Warning users against COVID-19 or Tax refund-related fraud.
        - Impact of weak passwords and how cyber attackers use them.

- **Simulations**: Simulate some key threats to prepare the personnel to be able to make decisive and secure choices.
    - **Phishing**: Use topical events or upcoming events to craft phishing emails and measure how many users click (**click rate**) and how many report (**report rate**), and measure it quarterly. The creativity in how phishing email is crafted may reflect in the click rate. Suggested topics are:
        - File tax returns.
        - Updating HR data for receiving an extra bonus or extra vacation days.
        - Urging users to click on a link to update software.
        - Registering for the lottery.

    - **Vishing**: Use voice-based impersonation to test key personnel. For instance, test the helpdesk for password resets and notice if they validate the identity appropriately.

  - **Deepfake**: Test key personnel or top management.
  - **Incident response and crisis communication**: Scenario-based tabletop test. We covered some of this in *Chapter 12, Cyber Resilience.*

- **Security Champions program**: Identity, coach, and enable personnel in functions and business operations to be the flag bearers in their teams. The Champions program could be made and positioned as a badge of honor. The champions can drive the culture in their teams, because they know their day-to-day work life and being conversant with security objectives, may be able to connect the dots and improve security by training them on key concepts of CIA, incident vs. breach, defense in depth, ZT, and MFA.

  Organizations often allot formal time on employee goal-sheets for being such a champion.

- **Awareness week**: To augment awareness efforts and create a buzz, celebrate an awareness week/day with curated content. Suggested topics include:

  - Identity theft.
  - Security in daily life.
  - Social engineering.
  - Acceptable use.
  - Risks of using public Wi-Fi.
  - Reporting security incidents.

Tip: October is observed as the International Cybersecurity Awareness Month.

Suggested delivery modes

It is important to design the content and the mode of delivery as well. The mode of delivery can be determined by factors such as:

- Geographical spread of the target audience: When the organization has distant locations and a geographically dispersed workforce, it may be practical to leverage only digital means, such as webinars or just recorded sessions.

- Consider replicating content with regional/country-specific language.
- Gamification is a great way to make the content interesting. The learners can be given bits of learning and then asked to choose the best possible answer. For instance, after introducing the concept of confidentiality and safe data handling, the learner could be asked to choose the best answer from the multiple choices for the secure transmission of confidential data. Reward points can be awarded for the correct answer.
- Generally, the attention span of a human brain is short; therefore, the content must be short, crisp, to the point, and with usable examples. Ideally, the learning content to view (or even gamify) should be within 20 minutes.

The suggested **modes of delivery** of security messaging include:

- Classroom or online sessions.
- On-demand recorded sessions.
- Leverage content from **Massive Open Online Content** (**MOOC**) platforms.
- **Byte-sized online self-learning** is akin to the reels that have become popular. A short duration, say 2 to 5 minutes, of a quick and specific topical learning module on platforms like podcasts, YouTube channels, and the Microsoft Stream platform (for internal use).
- Gamified learning.
- **Cyber ranges**: Online SaaS providers that allow users to set up their own labs to simulate real-world threats and learn from them. Developers can use it to practice secure coding, cyber defenders can use it for understanding smarter ways to detect, risk analysts can learn and practice to understand new-age risks faster, and even employees in other functions can experience how to detect and deal with malware attacks, phishing, etc.
- Emails from the security team’s mailbox.
- Messaging from top management.
- Messaging app-based nudges, where permitted by law.
- Posters.
- Screensavers on machines.
- **Method-based nudge**: For instance, an external email banner warning

of potential malicious links.

- Blogs/Articles on intranet.
- Corporate swags (gifts) with security messaging, such as pens, T-shirts, mouse pads, privacy screens, etc.
- Cafeteria events such as *Ask me anything*, *quiz*, *Bingo-type games* focused on security concepts, and/or even competition around best crafted security messaging.

Tip: The user may be assigned a byte-sized training on social engineering and phishing if they click on a phishing simulation email. Repeat clickers, those who continue to be careless and fall prey, may be considered for additional monitoring and or classroom interventions.

Figure 13.2 illustrates the suggested strategy framework for the human centric security program:

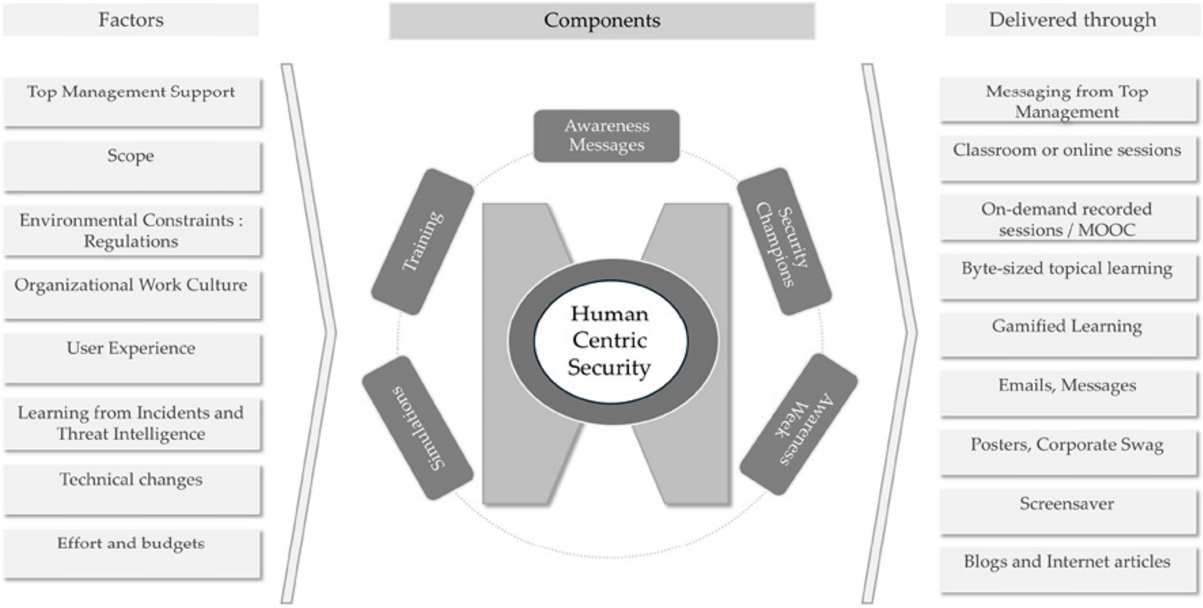


***Figure 13.2**: Human centric security program framework*

Conclusion

In this chapter, we examined the various facets of how the human mind may make choices based on conformity needs, biases, personality, and thus impact safe cyber decisions. The human centric security team drives the program to build and /or augment the security culture using some suggested component, their focus, and the methods to deliver them. The multi-year security culture shift should attempt to make everyone accountable to security.

In the next chapter, we explore the importance and relevance of controls around people and their awareness of security-related behaviors. We will also explore some of the common security attacks focused on employees and contingent workers.

Key takeaways

Some of the key lessons learnt include:

- The human centric approach is above and beyond the traditional security awareness approach of emails, security posters, and annual training.
- Notwithstanding the advancements in technology, such as quantum computing, artificial intelligence/machine learning, autonomous vehicles, and practically anything available under software-as-a-service, the human mind and behavior continue to be at the front and center of cyberattacks that happen.
- The human centric efforts weave in how the human mind thinks and reacts to an external stimulus, and use that to create a culture of empowered and accountable security choice, enriched aspects like **threat intelligence** (**TI**), learnings from incidents, regulatory, and contractual requirements.
- A multi-modal approach, using nudges, emails, case studies, role-based learning, gamification, byte-sized learning, etc., to infuse contextual knowledge and awareness, is successful in helping create the culture versus preaching about security.
- Global organizations will benefit from localized language and the regional cultural nuance infused in human centric efforts.

References

- **https://www.swarthmore.edu/a-brief-history/1951-psychologist-solomon-aschs-famous-experiments#**
- **https://www.sans.org/blog/solving-the-human-risk-problem-**

engaging-and-securing-your-workforce/

Join our Discord space

Join our Discord workspace for latest updates, offers, tech happenings around the world, new releases, and sessions with the authors:

https://discord.bpbonline.com

CHAPTER 14
Managing Security Talent

Introduction

In this chapter, we will cover the talent gap in the field of cybersecurity and the options the **Chief Information Security Officer** (**CISO**) may have. This chapter will also aid a security practitioner in using the suggested competency framework for upskilling, especially in light of advancements in **artificial intelligence** (**AI**) and cloud computing.

Structure

The chapter covers the following topics:

- Talent gap
- Strategies to use

Objectives

By the end of this chapter, you will be able to understand the gaps that exist in the space of cybersecurity talent and how organizations may need a focused strategy to upskill the workforce. We will also cover the role of security service providers, their risks, and suggestions to handle them. We will explore possibilities of AI in security processes to not just automate and modernize, but also manage the talent deficit. Lastly, we will cover how to define and leverage a competency framework for the security team.

Talent gap

Cybersecurity is an exciting field that continually deals with challenges introduced by evolving regulations, the varied **techniques, tactics, and procedures** (**TTPs**) deployed by **threat actors** (**TA**), and the advancement in technologies like AI, cloud computing, and quantum computing.

Professionals in IT and **enterprise risk management** (**ERM**) functions often aspire for a role in cybersecurity. The domains of cloud security, application security, and forensics seem to attract people. For instance, every person wants to know how to hack into something, mostly for fun, and has been enamored by such a possibility.

The industry is faced with a lack of skilled, experienced people to cater to the requirements of the organizations. The talent gap can be considered in three dimensions, namely:

- **Skill gap**: The security professionals lack sufficient know-how/knowledge of using the security controls. For instance, organizations may get a strong workforce skilled in tools like vulnerability management or in architecting cloud resources, but lack the ability to understand and articulate risks. Similarly, a lot of cybersecurity auditors lack technical acumen and focus on a checklist mode of validating controls, and thus miss out on comprehending and accounting for compensating controls.
- **Experience gap**: The cybersecurity team often lacks experience in applying broader knowledge of security domains in a variety of different situations. The complexity of the business environment, technology environment, and regulatory environment, coupled with fast-changing threat actor tactics, makes it difficult for professionals to stay abreast. Additionally, some of the decision maturity comes from experience and time. Every security team member will have varying degrees of expertise in each of the security domains.

  A security professional is expected to possess technical know-how and be able to connect the dots between business processes, management directives, and frame security outcomes in a way that the various levels of the organization can comprehend and make informed choices.

For instance, the CISO and the information security team are expected to understand cybersecurity concepts well apart from accounting processes (such as accounts payable), HR processes (such as background verifications), IT, cloud, **operations technology** (**OT**) security, physical security, and the implications of **Digital Personal Data Protection Act** (**DPDPA**), 2023 and GDPR, HIPAA, etc. The organization also expects cyber risk to be a domain of cybersecurity personnel only, and thus, any cybersecurity risk is expected to be owned by the security team.

The cybersecurity team is also required to be quick in deciding actions in times of crisis, such as a security incident, and must garner top management support. Building such credibility and confidence takes time.

- **Capacity gap**: Organizations worldwide are experiencing a shortage of people to handle security tasks and initiatives. Some of the factors impacting the capacity are:
    - There is far *more demand than the supply* of skilled people available. Though the security discipline has been around for decades, there are not many vocational, undergraduate, or postgraduate courses that focus on cybersecurity. Many courses in computer science at good universities do cover the topics of cybersecurity, but not in enough numbers to support the needs of the market. There are institutes that run professional certification courses that are widely recognized to help bridge domain-specific knowledge.
    - Conversely, because there are more jobs available in the market, generally at higher compensation, the workforce often moves jobs for greener pastures. This attrition impacts the sustained pace of organizational progress.
    - Security professionals are expected to be well-rounded, and thus their hiring process, with a deeper background check, acceptance of a job offer, and contractual requirements of notice period in several geographies, may take time.

- On one hand, the profession is exciting and fascinating, but on the other hand, the teams in **security operations center** (**SOC**), **incident response** (**IR**), cyber resilience, and **continuous threat exposure management** (**CTEM**) often work under a lot of pressure. These defender teams do not get a second chance and must continually ensure the attacker activity is monitored, detected, protected from, responded to, and recovered from **as soon as possible** (**ASAP**). Cybersecurity teams thus deal with the **Brittle, Anxious, Non-linear, and Incomprehensible** (**BANI**) world. The mental/physical exhaustion for a prolonged period may cause **burnout**. The productivity, quality of decision making, impact on an individual's health, and the motivation to bring the a-game are impacted. The burnouts may also cause attrition (or turnover).

In spite of the challenges listed, the cybersecurity team, especially the CISO, is well-positioned to drive organizational direction across the length and breadth of the company.

Figure 14.1 illustrates the talent gap in cybersecurity:

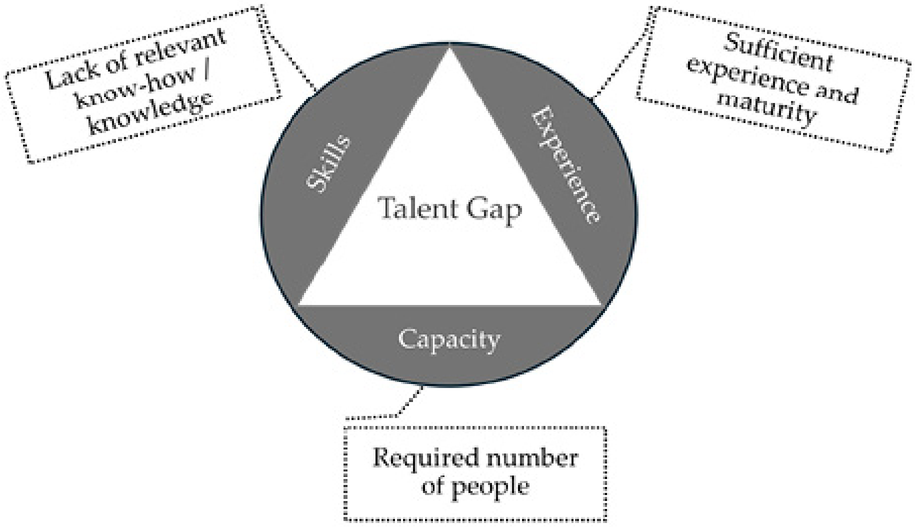

Lack of relevant know-how / knowledge
Sufficient experience and maturity
Skills
Experience
Talent Gap
Capacity
Required number of people

Figure 14.1*: Talent gap*

Strategies to use

Cybersecurity talent management needs careful planning and consideration.

As a CISO/Security leader, one would have to make difficult choices to ensure the security program can progress. In *Chapter 3, Role of Standards and Controls,* we articulated the strategy for using NITS CSF 2.0 and drafted the path for the future of the security program. This journey would need skilled cybersecurity talent in sufficient numbers. *Figure 14.4*, pictorially demonstrates an example of the journey to cover against *the* desired state (shown in dark shade), as compared to the industry (depicted in a darker shade), to optically convey the gap:

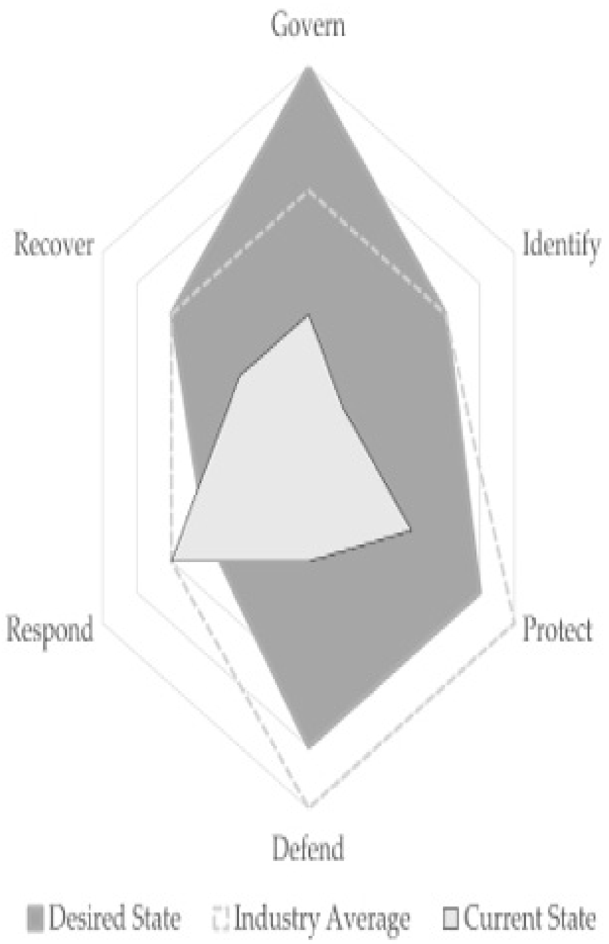


***Figure 14.2**: CSF: Current and desired state benchmarked against industry average*

Considerations for the skill gap

A cybersecurity professional needs to have:

- Functional skills (cybersecurity domain skills).
- Business acumen.

- Articulation and influencing skills.

Insofar as it relates to functional skills, the security requires not just security domain expertise but also IT knowledge, for instance, to understand network security, a thorough knowledge of ports, protocols, and IP addresses is needed. Therefore, the security team should ideally possess a T-shaped or a Pi-shaped skill set. Some of these skills are as follows:

- **T-shaped skill**: In which the individual has a wider understanding of a variety of security domains like **threat and vulnerability management** (**TVM**), Network security, endpoint security, policies and standards, and incident management and response, but is very skilled in risk assessments at depth. More importantly, the personnel can connect the dots between various processes. Middle-level managers are more successful if they possess such skill width and depth. *Figure 14.3* illustrates the T-shaped skill:

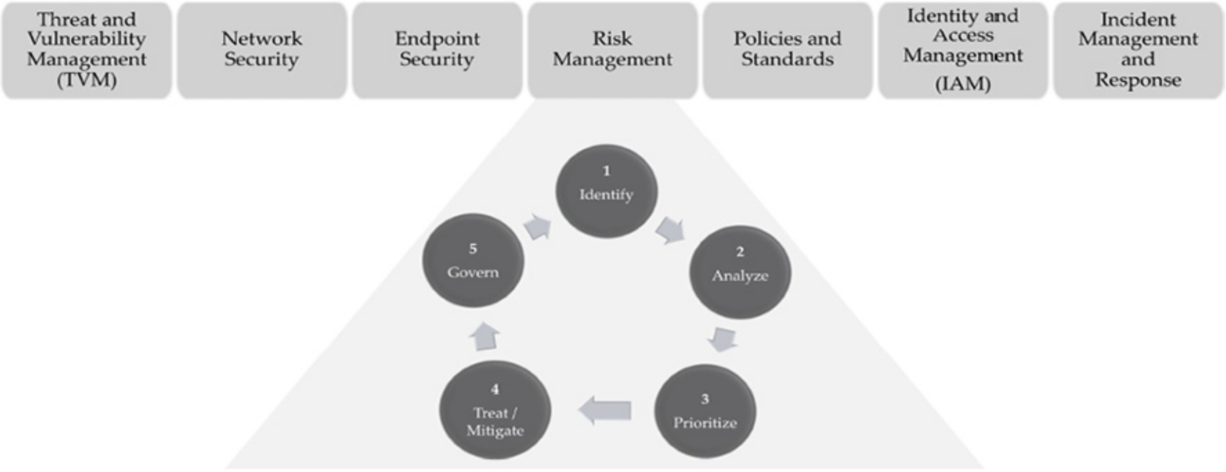


***Figure 14.3**: T-shaped domain skills*

- **Pi-shaped skill**: In which the individual has a deep understanding of more than one dimension of security controls and yet has a wider understanding of the overall security domain. It is named after the mathematical symbol π (pi) because the individual has in-depth skills in more than one security domain. Ideally, the leaders of security domains would benefit from such cross-domain knowledge. Often, the expertise in the domains may be complementary. *Figure 14.4* illustrates π-shaped expertise for cyber resilience that is aided immensely by cloud computing security:

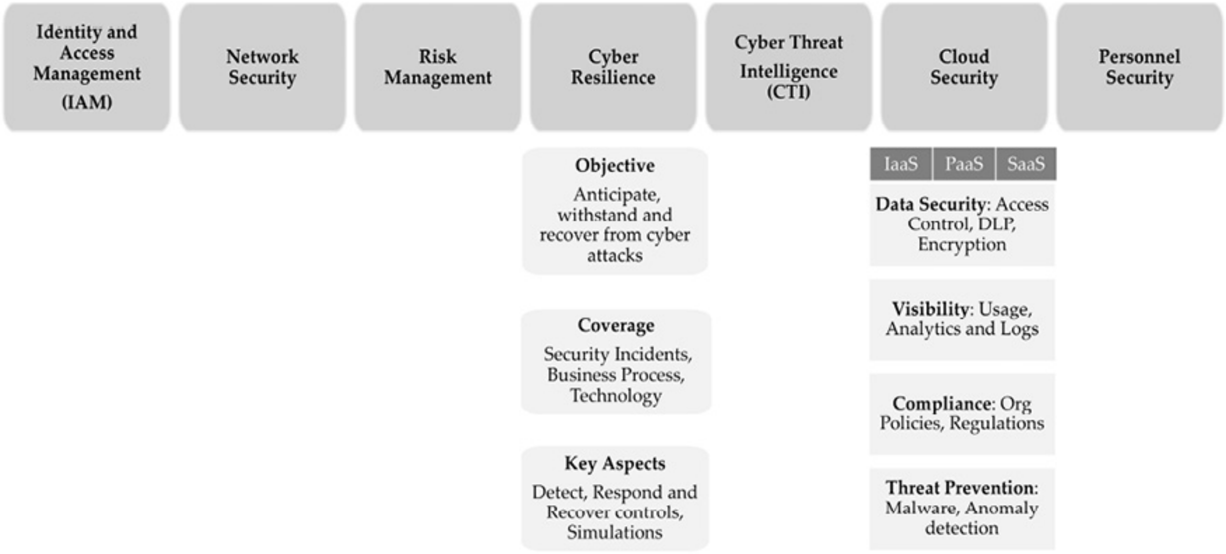

Identity and Access Management (IAM)
Network Security
Risk Management
Cyber Resilience
Cyber Threat Intelligence (CTI)
Cloud Security
Personnel Security
Objective
Anticipate, withstand and recover from cyber attacks
Coverage
Security Incidents, Business Process, Technology
Key Aspects
Detect, Respond and Recover controls, Simulations
IaaS
PaaS
SaaS
Data Security: Access Control, DLP, Encryption
Visibility: Usage, Analytics and Logs
Compliance: Org Policies, Regulations
Threat Prevention: Malware, Anomaly detection

Figure 14.4: *Pi-shaped domain skills*

Security professionals may also have skills across all domains, but not enough depth in any one domain. This is called **mile-wide** and **inch-deep**. This approach to integrating several security domains is quite useful for senior cyber security leaders, such as a **Chief Information Security Officer** (**CISO**) or **Chief Risk Officer** (**CRO**). It is not implied that the CISO must only have surface knowledge of all domains.

The requirements of the CISO's role and skills are dependent on the organization's needs and industry. For instance, a CISO of a healthcare organization must be familiar with the functioning of healthcare systems and the regulations that impact them. Similarly, the CISO of an e-commerce company would need to be aware of systems like inventory management, warehousing, logistics, supply chain, cloud computing, and its security, as well as secure coding.

Each security domain is a detailed study, and to build competency of the security talent, skilling and certifying are key. *Table 14.1* tabulates some of the popular and widely accredited certification courses:

Institute/Organization	Course/Certification (®) / (SM) / (™)	Key concepts covered
Information Systems Audit and Control Association (ISACA)	**Certified Information Security Manager (CISM)**	Importance of aligning security to business objectives. Managing security program.

Institute/Organization	Course/Certification (®) / (SM) / (™)	Key concepts covered
	Certified Information Security Auditor (CISA)	Auditing security controls. Risk Assessment process.
	Certified in Risk and Information Systems Control (CRISC)	Security audit, control validation, risk management.
International Information System Security Certification Consortium (ISC2)	**Certified Information Systems Security Professional (CISSP)**	Multiple security domains across network, endpoints, Identity management, asset management, change control, personnel security etc.
	Certified Cloud Security Professional (CCSP)	Principles of cloud computing, and securing cloud.
	Certified in Cybersecurity (CC)	Entry level certifications laying foundations of CIA.
	Certified Secure Software Lifecycle Professional (CSSLP)	Methods of secure coding, protecting against attacks such as cross-site scripting, SQL injection, session hijack etc.
Cloud Security Alliance (CSA)	**Certificate of Cloud Security Knowledge (CCSK)**	Principles of cloud computing and securing it.
	Certificate of Competence in Zero Trust (CCZT)	Principles of ZT and strategy for implementing it.
EC-Council	**Certified Ethical Hacker (C|EH)**	Determine vulnerabilities of an information asset and tricks and methods compromise it.
SANS Institute	**SEC497**: Practical **open-source intelligence** (**OSINT**).	Using various intelligence sources such as Internet, to gain insights about IP, network, people, photographs etc.
	SEC504: Hacker Tools, Techniques, and Incident Handling.	Range of cyberattacks and applying methods to protect from them.
	FOR578: **Cyber threat intelligence** (**CTI**).	The life cycle of CTI, ways to gather and analyze intelligence and understand TTPs.

Institute/Organization	Course/Certification ($^{®}$) / (SM) / (™)	Key concepts covered
Security Product Vendor Trainings	**Cisco Certified Internetwork Expert** (**CCIE**). **Cisco certified specialist**: Threat hunting and defending.	Product specific training on networking and threat hunting etc.
	Proofpoint: Certified Security Awareness Specialist.	Building and maintaining security culture.
	Palo Alto: Cybersecurity practitioner.	Cybersecurity technologies largely specific to the vendor's product capabilities.

Table 14.1*: Examples of cybersecurity certifications*

Most of the cybersecurity exams are online and **multiple-choice questions** (**MCQ**) where the aspirant must select the right answer from the choices given. Generally, 70% or more is required to pass. The list above is representative and is provided as a suggestion; the aspirant may need to choose based on requirement, cost, and mode (classroom, online, self-study, etc.). There are several paid/free **massive open online course** (**MOOC**) platforms available. Similarly, there are certification courses as implementer or auditor of standards such as ISO27001:2022.

Often, organizations would need to hire certified professionals and/or support funding of such for upskilling their security team to meet the directions of the board of the company/regulators/customer contracts. These certifications are one of the means to establish knowledge, but do not imply the level of expertise and experience.

Note: Most of the certifications require acquiring Continuing Professional Education (CPE) points, primarily to continue to learn and apply cybersecurity knowledge. Points can be acquired by attending webinars/seminars, publishing blog posts or articles, taking other exams, etc. Usually, internal projects and work applications do not count towards CPE.

Considerations for experience gap

The cybersecurity professional needs to have analytical thinking, the ability to look at big pictures, and be able to make unbiased judgments. These skills are honed over years of experience. The rapid change in the environment requires the talent to constantly be prepared to gain experience. For instance, between late 2023 and early 2025, the awareness and adaptation to AI had significantly grown. Similarly, the TTPs of the threat actor continue to evolve at a rapid pace. Some of the strategies to use in enabling experience to be built for functional skills are:

- **Shadowing**: Where the individual(s) are groomed to observe a senior practitioner and learn from them. For instance, a firewall admin is assigned to review the firewall changes under the guidance of an experienced practitioner. This is akin to being an intern within the team for a short duration. Such models are commonly used in the Legal fraternity, practitioners of medicine, and even Airline crews. This approach is very useful in training someone to take on higher-order tasks, usually in the same line of work.
- **Job rotation**: The specific tasks, responsibilities, and outcomes for an individual(s) are changed to provide a different workstream exposure to an individual(s). For instance, the Application security specialist may be assigned to perform SaaS reviews, or an infrastructure security architect may be assigned to manage TVM tools. Alternatively, a risk assessor for the banking business unit may be assigned to be a risk assessor for the insurance business. Both these businesses have different nuances and help the risk assessor gain a wider perspective. Often, job rotation from other adjacent functions, like IT, is very useful. For instance, the IT system admin can be a good bet on making and implementing strategies to govern user privileges (PAM systems).
- **Role rotation**: The larger expectation of the position/role is augmented or modified to provide the platform to build exposure. For instance, a **vulnerability management** (**VM**) manager handling VM scanners is assigned to learn and perform **penetration testing** (**PT**). Or takes on the role of being a cyber threat intelligence personnel.

  The job/role rotation is a good means to move people laterally, i.e., their designation /salary band remains the same; this is called **lateral movement.**
- **Project-based/Stretch assignments**: Special time-bound projects or assignments are given to individuals. For instance, the infrastructure security analyst is additionally assigned the project to streamline the application software inventory. Or a cloud security specialist is assigned to delve deeper into vulnerability management of cloud setups.

The efforts to coach and train in other aspects are also called **cross-training**. Additionally, the experience is also required to be built on aspects such as:

- **Articulating risk**: Irrespective of the security role, a cybersecurity professional is embedded in the process of risk management. To be able to articulate the risk for the information asset owner is key to enabling proper decisions. Consider that an organization's e-commerce application has an improper code that exposes credit card information, but only from the portal's admin console. While the risk to the CIA of that credit card information exists, it may be less of a panic for the organization when the console is protected by a **phishing-resistant MFA** versus when it is not. The application tester must be able to correlate the context of the data in business terms to the application/business owner, its relevance, and impact to CIA, essentially answering the *so-what* in simple terms. Similarly, to protect against **Business Email Compromise** (**BEC**), the security specialist should be able to explain in simple terms how anyone's email ID can be spoofed to make it look legitimate and how it can be protected.
- **Presenting to stakeholders**: A security personnel's stakeholder range from highly technical IT admins, Cloud architects, Application developers, common users, Finance, Sales, and HR teams, and even within the cybersecurity team, various specialists such as application security, infrastructure security, cyber defense specialists, and risk assessors. Each of these people would consume the information differently. The art and science of communicating relevant and yet crisp information is key. Much of the refinement comes from practice. As security professionals grow with experience and take on wider leadership roles, they are expected to have the ability to speak on their pain points, achievements, and funding requirements. In an **elevator pitch** mode, the practitioners should be able to capture the attention of the senior leader in a very short time (such as a ride on an elevator) and yet leave a hook for a detailed discussion if needed. Senior C-suite executives have the experience to pick up such 2-sentence (or so) long insights and engage when needed. We remember brand tag lines from advertisements and often make the purchase later. Similarly, the security certifications person can succinctly give the status of certification and

the positive impact it creates on customer trust. Or communicate the purpose and timeline to roll out MFA across the organization to the **Chief Operating Officer** (**COO**).

- **Other interventions**: Organizations may also use other tools to coach on skills like public speaking, how to use the stage, body posture, etc. Similarly, tools and techniques to use in critical thinking, problem solving, root cause analysis, what-if analysis, design thinking, and behavioral analytics.

While a good measure of experience is time, even a shorter duration of exposure to multiple complex environments can prepare one to the level of expertise. A security professional may become a proficient penetration tester by virtue of the exposure, and not only because of the years spent doing penetration testing.

Considerations for fulfilling capacity gap

The growing deficit of available talent and various unhired roles poses a challenge for organizations to make progress on their security journey. Automation is a good way to reduce several of the manual and undesirable tasks that keep security professionals busy. Some of the other strategies to use for managing the capacity challenge include the following:

- **Building capacity with universities**: Organizations should participate in building industry-relevant course curricula at colleges. This enables the academia-industry collaboration to use the right pedagogy to learn skills needed for a future job. The student would become more confident and relevantly skilled to be hired. Several universities across the globe facilitate such interactions. Many of the colleges even invite industry professionals to periodically deliver lectures and coach the students in their classrooms, too. Industry forums, such as the NASSCOM in India, **Data Security Council of India** (**DSCI**), and Government bodies like the **Cybersecurity and Infrastructure Security Agency** (**CISA**) in the USA, often have targeted programs for skill development and/or scholarships. Similarly, universities work with organizations to place their students as interns for a short duration to gain hands-on work experience.
- **Hiring strategies**: The hiring strategy in today’s time may need to

evolve for aspects such as using:

  - Contractors when needed.
  - Specialized retainer services, such as IR and forensic investigation.
  - Gig-workers, i.e., hiring specialists on short-term outcome-based assignments, for instance, a CTI expert to do threat hunting.
  - Using college interns.
  - Internal job postings provide the platform for other users or skilled personnel to apply and interview internally/within the organization.
  - Often, talent may be available in smaller cities with comparable skills, and expanding the search pool to locations can help.
  - Social media networks and other professional networks are a good source to seek candidates.

The following considerations may be applied while drafting the job description and selecting candidates:

  - **Hire for attitude**: Curiosity to learn, demeanor/general conduct, accountability, transparency, and communication skills.
  - Being clear on required skills versus desired skills.
  - Aptitude can be built over time. It is not possible to find a perfect match, so be prepared to make choices.
  - Culture and team fit (behaviorally).

- **Retention strategies**: There may be several reasons for an employee to move out of the team/organization. Some of the following suggestions may help with the attrition/turnover:

  - Role/job rotation (lateral movement).
  - **Upskilling**: Where the incumbent talent is taught higher-level/more complex skills. For instance, the system administrator handling ID creation and deletion may be additionally trained for managing the **Active Directory** (**AD**) or security group administration.
  - Ensuring appropriate compensation and/or benefits are aligned with market conditions.
  - Learning/skilling/upskilling opportunities: funding and allowing time for employees to learn newer skills and be ready for the next

  job.
  - Vertical movement/promotion.
  - **Employee Stock Options** (**ESOP**) are generally used by listed companies and startups to give the employees a stake in the company by awarding equity shares of the company.
  - **Location movement**: People may also like to explore a role/job in a different city/country.

- **Talent burnouts**: To avoid burnout and keep these teams motivated, several interventions should be actively considered and implemented, such as the following:
  - **Expectation setting**: The **objectives and key results** (**OKR**) could be used to help structure the outcomes expected, their timelines, and a means to prioritize.
  - Manager/supervisor engagement with the team: to keep them motivated, show them the big picture, and explain how their efforts keep the company from harm.
  - **Rotation of shifts**: Teams such as SOC usually work 24x7; it is important that the mind and body are adequately rested. Rotation of shifts and their shift timings must be carefully planned to make it equitable.
  - Where it is possible to work from home, may be considered.
  - Adequately lit, comfortable seating with cafeteria services at the office should be provided.
  - Opportunities to learn other aspects and upskill in any area. For instance, the importance of AI and how it can help CTI will help a CTI analyst; similarly, ways to automate risk assessment flow will help a risk analyst.
  - **Reward programs**: Monthly/quarterly should be considered to recognize stellar contributions/having taken initiatives, etc.
  - Team building exercises, such as sports events.

Competency framework

The level of expertise of a skill may be differentiated by scales of expertise, such as beginner, intermediate, and expert. The levels of skills can be built across the security domains using the strategies defined above. By defining the career path of progression, laterally or vertically, employees can not only upskill but also fulfill the current and future capacity needs of the organization. For instance, an employee with beginner-level AD administration skills can pursue beginner-level skills in the SOC or pursue higher skills, such as managing the **identity governance and administration** (**IGA**). Similarly, the IT Helpdesk can be trained and assigned a role to handle **data loss protection** (**DLP**) alerts as a lateral move or on security risk and governance as an IT auditor.

A competency framework can illustrate the possible options, based on current skills and areas of interest, to suggest the next possible career move for an employee. The path may also be built to visually show the next 10-15 years of the competency path (career path). The career path should indicate the expertise level required at each security domain and for each role.

Such enablement is a great retention strategy as well, and an aid to enable employees to plan their careers and fulfill the talent gaps defined above.

Organizations often use the trained skills repository to fill the available open positions.

Figure 14.5 is an illustration showing lateral and vertical options from a particular security domain to the other:

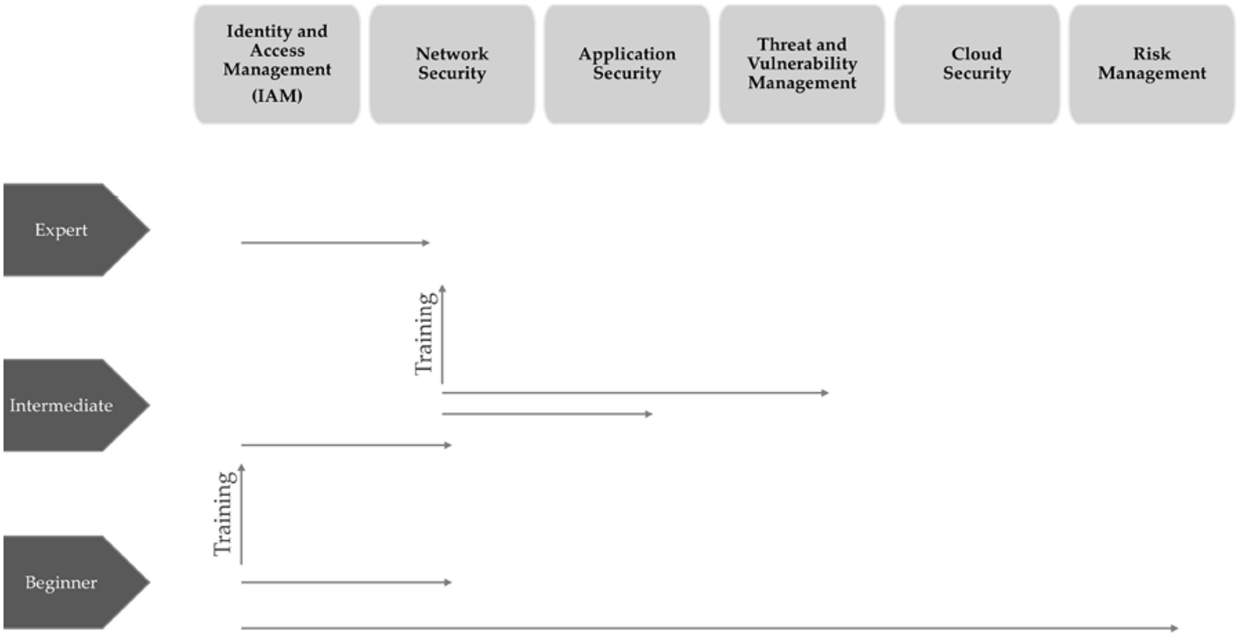


***Figure 14.5**: Illustration of career paths mapped against security domains and expertise level*

Note: As AI evolves, security roles are likely to evolve as well. The security practitioner's role and levels of competence shall evolve with it. AI is likely to be able to do the triage of incidents and logs, and enable a higher order of expertise with humans.

NIST has a framework called the **National Institute of Cybersecurity Education** (**NICE**), which is focused on enabling organizations to define and work on competencies using common terminologies.

Conclusion

In this chapter, we examined the importance of security talent and how to focus on the talent gap across three broad dimensions. We also defined the strategies to plug those gaps. A trained and ready talent is worth a gold mine for an organization. By deploying retention strategies, organizations can help retain the institutional knowledge within the organization and provide career paths. The competency framework is a great way to enable the employees of the organization to develop the right skills needed for their next job/role and be prepared to deliver in it.

In the next chapter, we will explore the dimensions of a security program, bring together the concepts we have learnt so far in this book, and provide guidance to meet the organizational objectives. We will discuss the how-to of defining, implementing, measuring, and continually improving a security program.

Key takeaways

Some of the learnings from this chapter are:

- Cyber talent is key to an organization's ability to build and run an effective security program.
- As part of the strategy, T-shaped, pi-shaped skills need to be built, and/or bought (outsourced), and those individuals provided platforms to experiment and learn.
- While traditionally security teams think of only technology and tools, it is pertinent that they pause, think, and equip themselves with business acumen and communicate clearly.
- The empowerment and retention strategy for security teams may include aspects like stretch assignments, job rotation, geographical relocation, external training, and adequate/relevant compensation.

References

- **https://www.isc2.org/Insights/2024/09/Employers-Must-Act-Cybersecurity-Workforce-Growth-Stalls-as-Skills-Gaps-Widen**
- **https://www.bcg.com/publications/2024/cybersecurity-talent-shortage-close-the-gap**
- **https://initiatives.weforum.org/bridging-the-cyber-skills-gap/home**
- **https://niccs.cisa.gov/workforce-development/nice-framework**

CHAPTER 15
Managing a Security Program

Introduction

In this chapter, we will bring together the concepts previously covered in the book, such as the technical controls, principles of security and privacy, **Zero Trust** (**ZT**), cyber resilience, threat intelligence, and security culture. We will explore how to measure the sufficiency and success of the security program and the impact of the budget on the program.

Structure

The chapter covers the following topics:

- Elements of a security program
- Team structure for the security team
- Measuring the security program

Objectives

By the end of this chapter, you will be able to appreciate the complexity and depth of structuring the various components of the security program, driven by organizational objectives, budgets, and culture. You would become familiar with various metrics and benchmarks that can be used to gauge the effectiveness of the security program.

Elements of a security program

The primary goal of all functions in an organization, especially the cybersecurity function, is to enable the organization's objectives to be met. As we learnt in *Chapter 2, About Managing Risks,* the management of the organization must be able to make informed risk-based choices.

The security program of an organization is an all-encompassing approach to protecting the CIA of its information assets by weaving together the organizational needs, security philosophy (such as Zero Trust), risk appetite, and the security objectives into actionable and measurable aspects.

Design considerations

Some of the foundational aspects to consider in designing, developing/augmenting a security program are:

- **Top management direction and support**: One of the most crucial elements of the program is the consistent and clear top management support for security objectives. Their strategic direction on acceptable risk (appetite) and the broader approach, such as ZT, security by design, or privacy by design, or the security outcomes expected, is a key foundation for defining the security program.
- **Adequately funded**: Security programs are an investment targeted at protecting the CIA's information assets. To do so, several investments might be required for selecting and implementing technology controls in the previous chapters covered, for instance, next-generation firewalls, **endpoint detection and response** (**EDR**), **threat and vulnerability management** (**TVM**), **security orchestration, automation, and response** (**SOAR**), cyber resilience, etc. Similarly, the security talent is scarce and expensive. All that requires money and therefore should be appropriately budgeted.

  An organization may have differing methods of budgeting. The security leadership, especially the CISO, would need to ensure the **Chief Financial Officer** (**CFO**)'s team is duly informed of the budgetary needs and their rationale and strong justification.

Tip: Cybersecurity budgets are often reviewed as a % of the IT budget. Generally, a mid-size organization spends about 5-7% of its IT budget for cybersecurity. The percentage may var based on the structure of the team and their responsibilities.

- **Empowered and capable teams**: As discussed in *Chapter 14, Managing Security Talent*, a plan to manage cybersecurity talent is key to ensuring the success of the program. The strategy must include capacity and capability aspects. The cybersecurity team of experts should be empowered to engage with the organizational stakeholders and enable secure choices for them. A moot question is the relevant empowerment of the **Chief Information Security Officer** (**CISO**).

Tip: In *Chapter 11*, *Incident Response and Planning*, we examined the growing focus o regulators on the role of CISO, especially after a breach, like the SolarWinds one.

Desired features/attributes

The following desired features/attributes may be considered for a successful security program to build upon the foundation:

- **Aligned to organizational objectives**: The security objectives of the organization must be aligned with the larger organizational objectives, such as:
    - Being participative and accommodating to an organization's planned **cloud migration** journey. For instance, by March 2023, Unilever, the global retail giant, had made aggressive progress in its digitization efforts by leveraging Microsoft's cloud offerings.
    - Supporting **mergers and acquisitions** (**M&A**), prompting newer capabilities to be secured. For instance, *Mastercard*, well known for its credit cards, acquired the threat intelligence company *Recorded Future* in 2024. This was meant to boost Mastercard's cybersecurity services, including fraud detection and prevention. Their teams would have had to quickly leverage the power of intelligence into their services.
    - **Enabling a business strategy** to expand products, services, and/or locations. For instance, *CitiFinancial India*, part of *Citigroup*, expanded rapidly, adding almost two branches a day in 2006. This required several teams, such as IT, cybersecurity, and operations, to rapidly scale talent and application accessibility in an era where

cloud computing had not picked up.

The vision and mission of security and its alignment with the organizational objective should be clearly defined and communicated by the CISO.

- **Adaptable**: As previously discussed in this book, the operating environment is constantly changing, and therefore, the security program must quickly adapt and prepare for applicable efforts to stay relevant. The core philosophy of control selection, covered in *Chapter 3, Role of Standards and Controls,* and its implementation, is useful in adapting to the changing environment. The main aspects to consider are:
  - Technological advancements, such as the emerging role of AI (generative and agentic), the **post-quantum cryptography** (**PQC**), and the general adoption of cloud computing. Some of these were covered in *Chapter 6, Key Security Technologies*.
  - Changes to the business environment, to ensure differentiation, aka **unique selling proposition** (**USP**), are maintained in lock step with competition. For instance, cloud-delivered entertainment content like *Netflix*, *Amazon Prime*, *Jio Hotstar*, or *Sony Liv,* and the options of services available in ride-hailing applications like *Uber*, *Lyft*, *Ola*, and *Meru*.
  - Regulatory changes, such as the DPDPA of India.
- **Proactive**: As rapid developments occur across the threat landscape, a security program must be proactive in anticipating and being ready. For instance, to counter the rapid increase in ransomware, detective and protective controls should be implemented and monitored. It is recommended to use a concept called **external breathing**, wherein formal and informal peer-level conversations may be leveraged to share and learn from each other. Some of the external sources include:
  - Customer's security and/or IT teams.
  - Product and security teams of key vendors.
  - Industry forums such as the *Data Security Council of India (DSCI)*, or the regional chapters of institutions like *ISC2/ISACA*.
  - Feeds from intelligence and other cybersecurity news sources.

    - CISO to CISO connects are a great way of engaging within the industry.

- **Reliable:** As previously covered in the book, organizations will benefit from adopting a consistent approach for a security framework, such as *NIST CSF*, and standards such as *ISO27001:2022*. Similarly, the philosophy of risk management and Zero Trust should be consistently applied so that the relevant teams progress with clarity.
- **Continual improvement**: A good security program should focus on continually evaluating the security posture and using a risk-based approach to fix gaps and incorporate new controls. An organization's risk appetite, covered in *Chapter 2, About Managing Risks*, will also be a deciding factor. A key aspect of continual learning includes learning from industry and organizational security incidents/breaches.
- **Measurable**: Consider using approaches like defined **objectives and key results** (**OKR**) to be definitive and progressive on outcomes and their measure of performance. We explored frameworks such as the NIST CSF to not only determine the current state but also the desired future state. Similarly, organizations may use the **Cybersecurity Capability Maturity Model** (**C2M2**) to define the current and future state of the security program. The OKR, or simply the measurable and actionable goals, must be clearly articulated and communicated. This includes the priorities for the upcoming months/year(s). For instance, the security team of an organization migrating to the cloud must ensure adequate priority and focus are provided for it with adequate technical resources and controls.
- **Communication**: Timely and periodic communication is essential to foster stakeholder engagement and to help build the security culture. The security program should engage stakeholders in various functions of the organization to enable business objectives, and their performance must be duly and appropriately reported to the stakeholders. For instance, the **Chief Technology Officer** (**CTO**) should be briefed about the status of closed and open gaps/vulnerabilities, whereas the business leader may be periodically informed about the current high risks in the business and emerging trends of risk.

Figure 15.1 illustrates the various desired features of a security program:

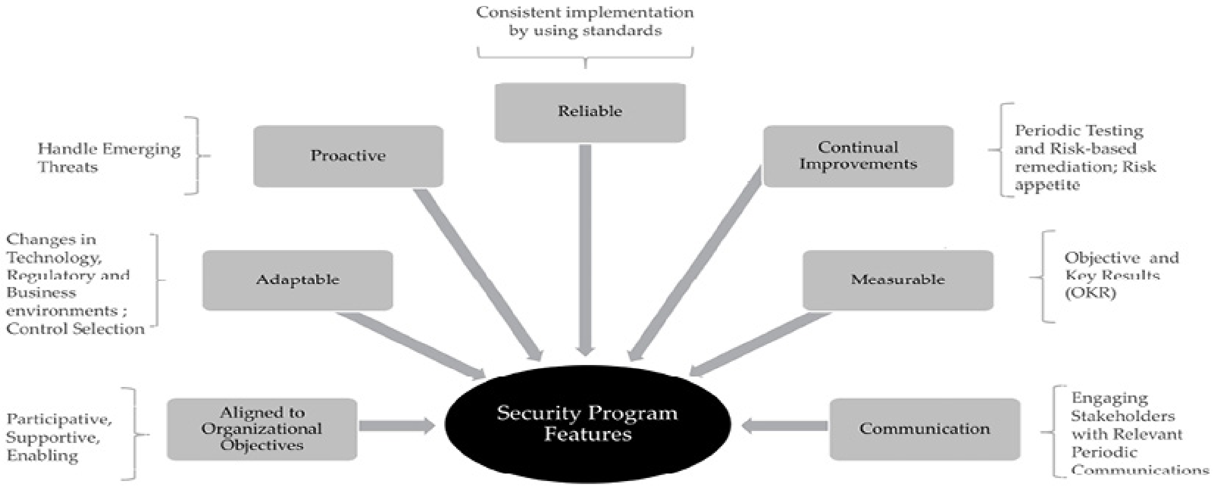

Consistent implementation by using standards
Reliable
Handle Emerging Threats
Proactive
Continual Improvements
Periodic Testing and Risk-based remediation; Risk appetite
Changes in Technology, Regulatory and Business environments ; Control Selection
Adaptable
Measurable
Objective and Key Results (OKR)
Participative, Supportive, Enabling
Aligned to Organizational Objectives
Security Program Features
Communication
Engaging Stakeholders with Relevant Periodic Communications

***Figure 15.1**: Security program features*

Components of a security program

Having set the foundation and determined the attributes/features of the security program, we will now examine some of its key components. An organization may choose to adapt the suggested structure and/or its variant based on its needs. *Figure 15.2* illustrates the overall framework for a security program with some representative components:

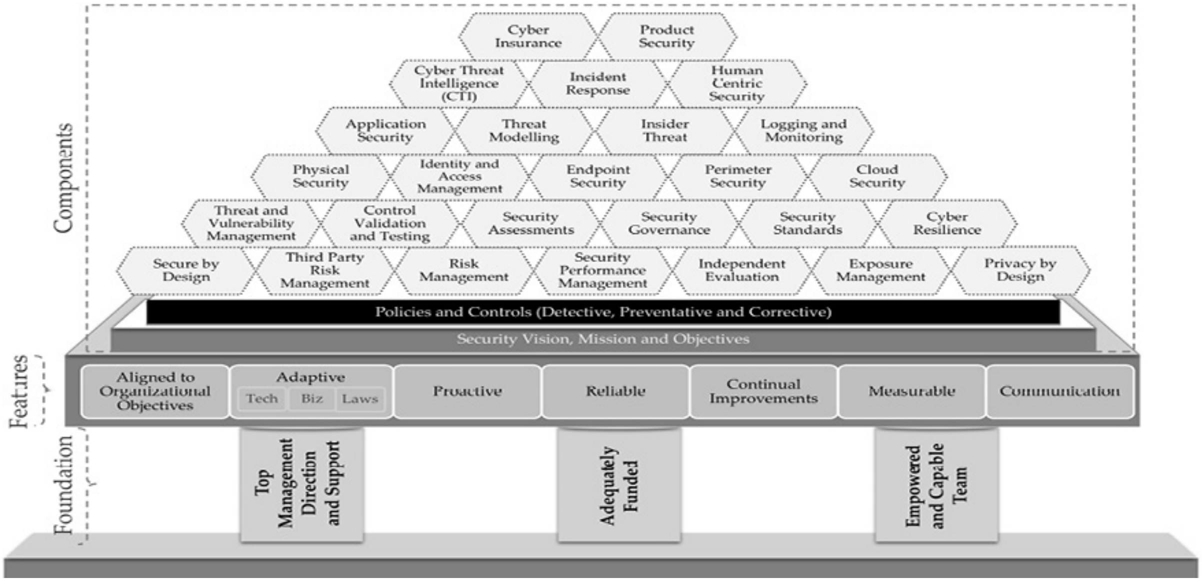


***Figure 15.2**: Security program framework with representative components*

We have covered several of the components shown in *Figure 15.2* at length throughout the book and thus would not be covered in detail here. Some of the components shown above are explained as follows:

- **Security vision**: The vision for the security program sets the tone for the long-term direction and mindset of the security program. A vision is generally broader and usually does not need changes even after decades. For instance, the security vision of an e-commerce platform may be *to provide for a safe and seamless integrated secure commerce experience across all delivery channels.*
- **Security mission**: The mission helps channelize the focus on how the vision will be achieved over the short term. The security mission may continue to be adapted as time progresses. The mission for the e-commerce organization may be *to become the trusted partner of choice for the suppliers and consumers over the e-commerce interface.*
- **Security objectives**: They are tactical goals that are defined and implemented to meet the security mission. For instance, the e-commerce organization may choose to ensure its domain name-related risks are monitored and expeditiously remediated for risks like look-alike domains, cybersquatting, and fake websites.
- **Security policies**: In *Chapter 3, Role of Standards and Controls*, we explored the various frameworks, standards, and considerations for an organization's security policy and the types of controls.
- **Security performance management**: Validating, measuring, and tracking the key results of the security objectives periodically shall help make continual progress. The measurement will not only be for internally visible controls, such as the efficacy of encryption on the endpoints, but also on external presence, such as the exposed patching on the internet.

Figure 15.3 illustrates some common fields that an issue tracker may use to document the control validation/other assessments:

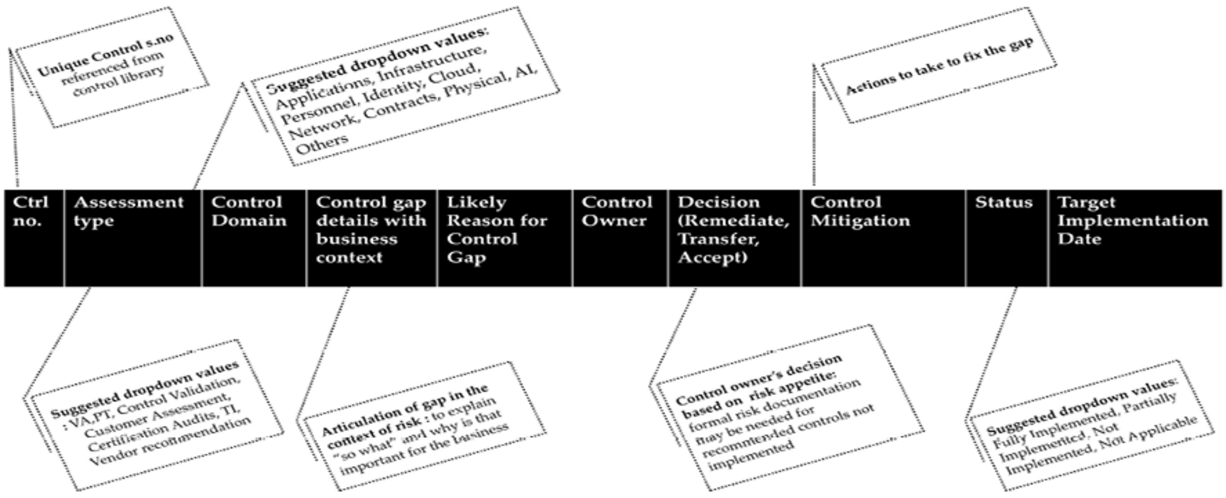

Unique Control s.no referenced from control library
Suggested dropdown values: Applications, Infrastructure, Personnel, Identity, Cloud, Network, Contracts, Physical, AI, Others
Actions to take to fix the gap
Ctrl no.
Assessment type
Control Domain
Control gap details with business context
Likely Reason for Control Gap
Control Owner
Decision (Remediate, Transfer, Accept)
Control Mitigation
Status
Target Implementation Date
Suggested dropdown values : VA,PT, Control Validation, Customer Assessment, Certification Audits, TI, Vendor recommendation
Articulation of gap in the context of risk : to explain "so what" and why is that important for the business
Control owner's decision based on risk appetite: formal risk documentation may be needed for recommended controls not implemented
Suggested dropdown values: Fully Implemented, Partially Implemented, Not Implemented, Not Applicable

***Figure 15.3**: Issue tracker*

Note: Organizations have started moving to a cybersecurity mesh architecture, which brings together security and IT technologies, their interoperability, and reporting. For instance, the Windows event logs are automatically correlated by an EDR with any inbound suspicious activity and then forwarded to proxy and firewall control. The technologies share the relevant insights like device posture, risk profile, and user persona context using open standards like the Open Cybersecurity Scheme Framework (OCSF). This architectural thinking helps in ensuring that future tools and technologies can integrate easily.

Budgeting for security outcomes

Earlier in the chapter, we explored the key pillar for the security program to be adequately funded. One of the popular budgeting methods is **zero-based budgeting**, wherein the required expenditure for the next period is thoughtfully planned and requisitioned without citing the previous period. This forces the teams to think hard about the money needed against current priorities and avoid wasteful expenditure just because it was spent in the past. Organizations sometimes have multi-year deals with CSP for unit rates; even in this case, the budget is required to be assumed zero and requisitioned fresh after careful evaluation of current demand. Over the last few years, several organizations have also started doing a quarterly budgeting instead of a yearly one, though the expenditure for the entire year may still be planned. *Table 15.1* illustrates a zero-based budget with the rationale for the spend being evaluated every period:

Category	Budget needed	Likely end of year spend projection	Justification

	Q2		
Human resources. Compensation and benefits for 27 people.	$1.2MM	$3.69MM	Compensation related expenses for our current and projected staff are needed to sustain the agreed outcomes.
Committed security technology spends.	$300K	$900K	These technologies are deployed to protect our email, internet and user devices. The actual user base and the license needed are rationalized quarterly.
Expansion of MFA deployment.	$10K	$25K	Our MFA deployment is continuing in phases. This quarter we will expand to financial applications users
Managed spend.	$150K	$400K	This is a retainer payable to the external incident response and forensics analysis team as per customer contracts and for cyber insurance purposes.

***Table 15.1**: Example of zero-based budgeting*

Some organizations continue to use **incremental budgeting**, wherein the past expenses of the period and the incremental amount are requisitioned to factor in growth. This may not accommodate the deceleration and does not give an incentive for cutting unwanted expenditure by default. *Table 15.2* illustrates an incremental annual budgeting example:

Category	Last Year's budget	Proposed budget for next year	Justification (assumes 5% growth as projected in annual plan)
Human Resources. Compensation and benefits for 27 people.	$3.51MM	$3.69MM	Compensation related expenses for our current and projected staff are needed to sustain the agreed outcomes.
Committed security technology spends.	$900K	$945K	These technologies are deployed to protect our email, internet and user devices. The actual user base and the license needed are rationalized quarterly.
Expansion of MFA deployment.	$32K	$33.6K	Our MFA deployment is continuing in phases. This year we will expand to financial and HR applications users.
Managed spend.	$400K	$420K	This is a retainer payable to the external incident response and forensics analysis team as per customer contracts and for

			cyber insurance purposes.

***Table 15.2**: Example of annual incremental budgeting*

There may be other methods (and formats) that may be directed by the finance organization. Irrespective of the CISO, they need to be clear about **capital expenditure** (**capex**) and **operational expenditure** (**opex**) required with solid justification. S/he must be clear on what outcomes may be at risk and by how much, if the required budget is not allocated or is curtailed. For instance, if, for whatever reason, the CFO declines the spend projected for the **cloud security posture management** (**CSPM**) tool, the CISO would need to communicate with the relevant stakeholders on the gap in governance and prepare for alternate controls. Similarly, if the expenditure for ISO27001 certification is put in abeyance, there may be contractual implications.

As market conditions change, the CFO may need business leaders like the CISO to continually adjust budgets, including aspects like making cuts. Those choices of what to cut and where must be thoughtfully evaluated, much like one does in a risk management approach.

In large and mature organizations, budgeting may happen on **enterprise resource planning** (**ERP**) platforms such as Oracle, Workday, and SAP. Otherwise, it can also be done on Excel sheets.

Team structure for the security team

One of the most debated questions in the cybersecurity and risk fraternity is the organizational hierarchy for reporting for roles like CISO and **Chief Risk Officer** (**CRO**). While optically it may be important for a CISO to report to the **Chief Executive Officer** (**CEO**), it very strongly depends on the culture of the company, the industry they operate in, and the general regional culture.

What matters more, instead of the hierarchy, are the following:

- Empowerment as an executive leader; if the CISO and the team are provided the right environment of independence and the ability to stand up and enable the organization to make the right decisions, the reporting hierarchy should not matter.
- The platform to engage the relevant mid and senior-level management to enable secure operations and growth of the business.

- The C-suite level executive's time and attention to the CISO's work.

It is also required for the CISO and the team to have the following characteristics, some of which were covered in *Chapter 14, Managing Security Talent*:

- **Growth mindset**: The CISO and the security team should ensure the secure growth of the business by providing solutions that reduce the risk. The mindset should not be that of a naysayer. For instance, instead of disallowing a business-required intranet application to be hosted on the Internet, the security analyst may suggest the risks of doing so or require **multi-factor authentication** (**MFA**) and constraints from unmanaged devices.
- **Roles and responsibilities**: Each organization may structure its security team differently. The roles and responsibilities must be adequately documented and communicated. For instance, in many organizations, vulnerability scans are the responsibility of the security team, whereas the downstream process of patching the vulnerable systems may be handled by IT. Similarly, the responsibility of identifying control gaps and remediating them would usually be a different security team from those who manage **incident response** (**IR**).
- **Strong stakeholder engagement**: The CISO role is emerging as one of the most influential enablers of business and growth. It is therefore important that the CISO and his/her office build and leverage the network of stakeholders. For instance, the CISO conversing with the sales team and the business operations team on how things happen will be more productive than a CISO who only works with the IT team. A CISO would need to be adept at de-cluttering technical/functional jargon to layperson level. For instance, explaining the need to introduce MFA as a layer of protection against password attacks.

In today's world, the CISO is a digital leader facilitating the rapid adoption of technologies securely. It must be noted that, irrespective of the hierarchy, a CISO would be expected to exercise influence without authority. *Table 15.3* illustrates the possible reporting hierarchy for the CISO and general advantages and disadvantages:

CISO reporting to	Advantages	Disadvantages

CEO	Visibility within the organization. Likely to be heard better.	CEO is busy and becomes the bottleneck for decisions. CEO's may lack interest in the CISO's work.
CFO	Likely to get better support in budgeting processes. Can get help in quantifying the business risk.	Difficult to explain technical terms to a deeply analytical and generally number focused professional.
Chief Operating Officer (COO)	Get better insights into operations and the potential gaps to fix. Larger network of stakeholders to tap on and get things done.	Difficult to explain technical jargon and make progress with ambiguities. COO may primarily be only focused on revenue and may overrule risk decisions.
CRO	Understand risk and can help management decisions around risk acceptance and mitigation. Can help elevate the risk to right layers, such as Audit Committee of the board to help expedite remediation.	Difficult to explain technical jargon and make progress with ambiguities. Sometimes CRO may be only focused on risk and may overrule the possibility of growth because of perceived risk.
CIO/CTO	Security by design and Privacy by design principles may be easy to incorporate. Security changes are easier to get rolled out. For instance, implementing controls to harden **Active Directory** (**AD**) based on threat intelligence.	Differing priorities and accountability. CIO/CTO are usually focused on availability but not integrity and confidentiality and thus may overrule the CISO's recommendations.
Chief Security Officer (CSO)	Most of the work is related to logical access but it can be correlated to other insider aspects with physical badging process.	CSO usually are focused on physical security and do not necessarily comprehend the technical layers of controls.

Table 15.3*: CISO and reporting hierarchies*

Sometimes organizations may also have the CISO report to the **Chief Human Resource Officer** (**CHRO**) or **General Counsel** (**GC**). Both have strong functional roles and may not appreciate the pace of implementation of controls, such as patching zero-day vulnerabilities. Both these roles may worry about human experience and/or liability primarily, and thus, their cognitive bias may interfere with decision-making.

TIP: Traditionally, a CISO reports to a CTO or CIO, as most of the CISOs come from technology backgrounds. There are also several CISOs, especially in regulated businesses like

healthcare services, where the CISO may report to the General Counsel or Chief Risk Officer.

Security team structure

The CISO's team, in turn, may be organized by security domains and/or regional presence. In most cases, the teams have a matrix organization with reporting to more than one boss. For instance, large organizations have a **Regional Information Security Officer** (**RISO**) or **Business Information Security Officer** (**BISO**) to enable the region/business unit to get local and personalized security expertise. A lot also depends on the organizational structure. Some organizations are organized for competencies across geographies. For instance, the risk management program is centrally run irrespective of the geography; similarly, the threat hunting and logging-related aspects are run centrally. Organizations may have a hybrid approach as well, such that the operations are decentralized at the regional level, while the strategy, planning, and budgeting are central. For instance, each business unit may have an information security officer(s) to manage several aspects of the security program horizontally for a region.

Figure 15.4 illustrates an example of a CISO's team organized by competencies and based on the CISO's focus:

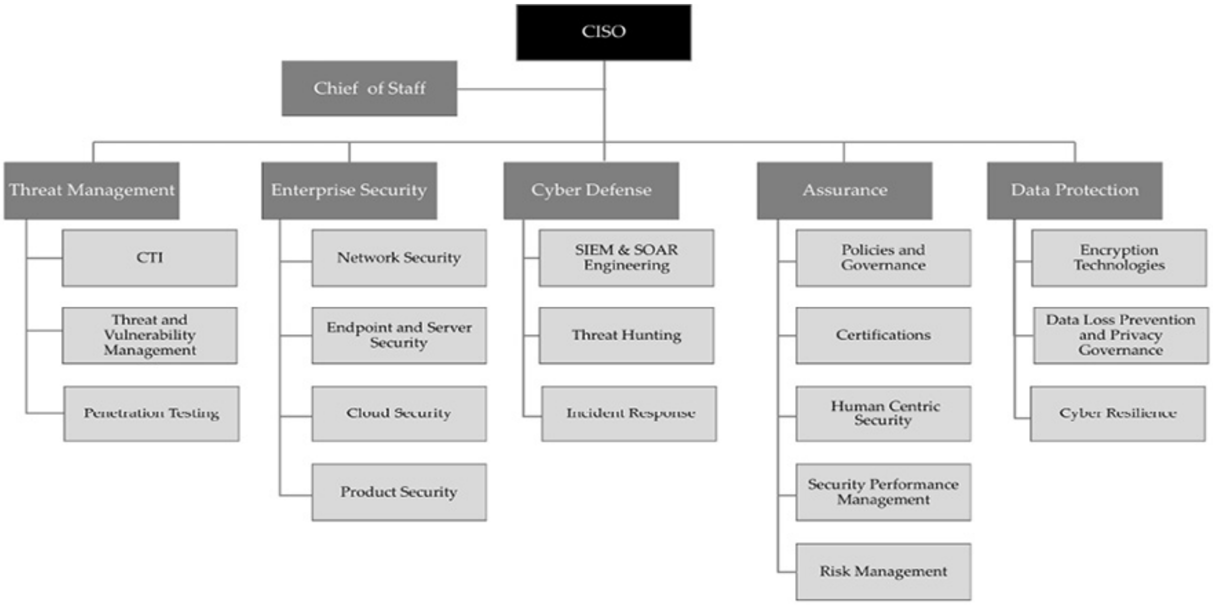


Figure 15.4*: CISO's team organized by competencies*

The illustration does not imply the number of people and/or direct reports of the CISO. The same person in the CISO's team may be handling more than one aspect. For instance, threat management and cyber defense may be handled by the same individual. Some aspects, such as control validation and security assessments, may be done at the functional level or by teams that oversee security certifications or security performance management. Similarly, external benchmarking may be handled by the risk management team. The illustration should not be treated as guidance for the names of the groups within the CISO's team.

Measuring the security program

The success and maturity of a security program are contingent on how the outcomes are tracked, measured, and reported. Security performance management, or simply security metrics, is a key means to determine if the security objectives are meaningfully contributing to the organizational objectives. The metric should be able to help prioritize remediation, help make budgeting decisions, and even help with reducing insurance premiums.

Security performance management may be seen at two levels, namely:

- **External scoring**: Several SaaS based services exist that can rank and rate the security posture of all external-facing assets. Notable among such services are *BitSight*, *SecurityScorecard*, *UpGuard*, and *ISS Cyber Risk Score*. Each of these platforms uses different attack vectors to perform reconnaissance and rate the security posture. For instance, if a *SecurityScorecard* uses a patching cadence, BitSight may focus on DNS records more. No one method is right. The CISO's team should consider these feeds as an externally exposed environment that might need a focused and time-bound remediation. The team may even set a target to keep a certain range of scores and use that as a guiding principle to drive the security culture.

TIP: Such external scores are visible to customer organizations, regulators, and cybe insurance teams. A good security program may benefit from better customer trust and fewe chances of catastrophic incidents.

SecurityScorecard uses an alphabetic rating A through F, A being the topmost positive rating, and provides a comparison on its portal for industry trends as well. Similarly, BitSight uses a numeric score range and provides an interface to predict the future state of the performance.

These external scores are a good input for an organization's third-party risk management program.

- **Internal metrics**: The OKRs should be measurable and, where possible, tied directly to the organizational objectives. Some of the metrics to track may be required for hygiene for the security program, for instance, the coverage of deployment of EDR and its effectiveness. The more important and relevant metric may be how the EDR tool has effectively thwarted security events and contained a problem automatically. We covered some of the concepts in previous chapters. Across all security domains, the practitioner should focus on the so-what of the metric and help articulate the impact on/enablement of the business. For instance, a metric on the number of phishing emails blocked at the gateway does not add much value to a business leader, but the same metric, when presented as the loss avoided and productivity gained (because incident investigation was avoided), may be more fruitful.

Outcome-driven metrics

Gartner introduced the term **outcome-driven metrics** (**ODM**) to enable the CISO and team's effective and meaningful engagement with business and functional stakeholders in line with organizational objectives. Fundamentally, approaches like ODM aim to demonstrate the preparedness of the security program to aid and augment the business objectives. For instance, if an e-commerce retailer aims to expand products and services into differently regulated markets like *China*, *Australia*, or *Europe*, the security ODM around the ability to detect and prevent data loss events as measured using the requirements of those countries. Similarly, by monitoring the blocked and allowed traffic to the web server, the CISO should be able to identify and take remedial action for the e-commerce portal to be available for safe use, especially during the period of sales:

- **Demonstrate risks avoided in $-value**: For instance, by proactively monitoring and controlling the look-alike domains, the organization would benefit from its sales and reputation not being compromised. This has a direct impact on the top line and bottom line of the organization.
- **Manage and monitor tech debt**: Technology updates continue to

happen over the years, and often, organizations struggle to keep pace. They may continue to maximize the life of a server just because it works well, notwithstanding the lack of support for new-age technologies like encryption. The organization's aversion may also stem from its reluctance to disrupt the existing functioning, often because there is little or no documentation of the dependencies the current systems use. The tech debt is the cost of having to maintain aging or unsupported versions of hardware or software, adding resilience-related risks to the organization. Attackers often look for badly configured, old, and weak platforms to compromise. Security teams can take a measured, risk-based approach to engage the stakeholders and reduce those risks over time.

Suggested success factors for a metric

A metric may consider both the coverage and the effectiveness of the control/program's operation and create a composite score as a device's posture:

- Coverage of EDR across all in-scope assets as a % of total assets.
- Effectiveness of EDR on all assets, for instance, functioning and updated EDR software.

An example of an ODM for this metric is: *the number of devices that do not meet a prescribed score and are most likely not able to support revenue operations, such as a client's accounts payable process*.

The security outcomes must be measured against at least two levels, namely:

- **Minimum level**: This is the minimum level a control or a program is expected to operate. For instance, the CISO may mandate that at least 95% of total assets must have a functioning EDR.
- **Desired level**: This is the level that the organization wishes to achieve over a period. For instance, the click rate to phishing emails (simulation or otherwise) is near zero. It is important not to have unreasonable targets like zero click rate or 100% machine learning, to always have the most up-to-date EDR. This is because in a complex organization, security culture is difficult to build and sustain. For instance, the users' travel and their machines would not be reporting back to the console, and should not be treated as a defect.

Figure 15.5 depicts one of the ways to demonstrate the trend in click rate of phishing simulation tests to drive the outcome of reducing it:

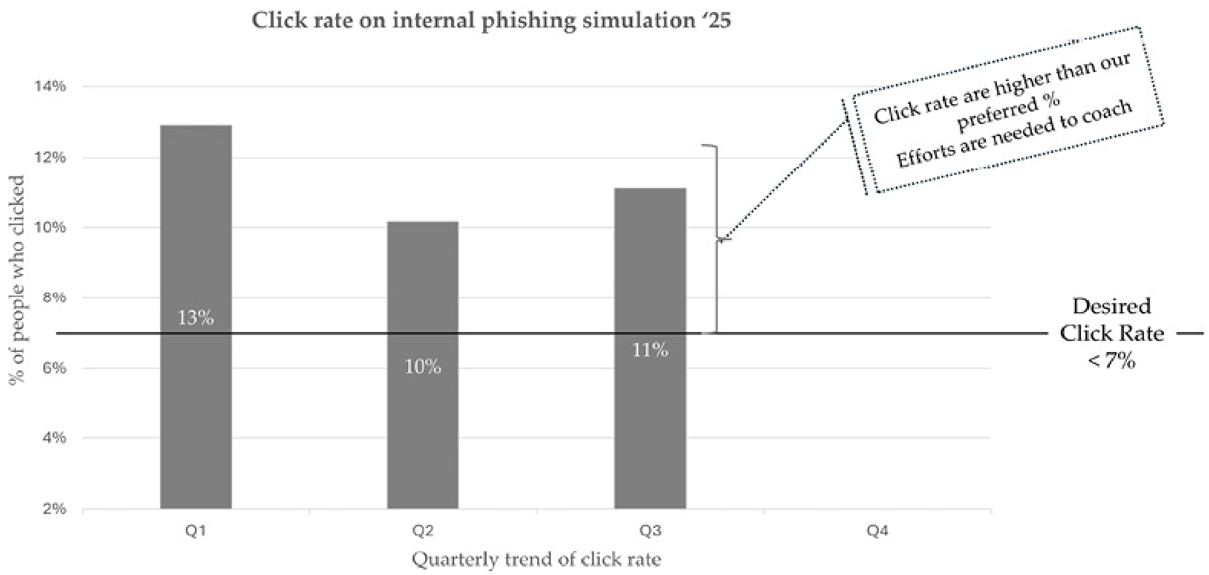


***Figure 15.5**: Example of phishing simulation results*

In *Chapter 3*, *Role of Standards and Controls,* we introduced the possibility of a journey of control maturity. That aspect is relevant in defining the thresholds, like minimum and desired levels.

Metrics often need a baseline/benchmark to determine the industry's best practice to identify the success criteria. For instance, the mean-time-to-patch, i.e., time between detection and remediation, for critical vulnerabilities may vary in the industry between 7 and 15 days. Based on the complexity of the organization, the compensating controls, the organization may choose to define the measure it wants to track.

Like any piece of information, a metric must be able to:

- **Provide actionable insights**: For instance, the number of machines not getting scanned periodically, their device score, and thus are likely to be disallowed from connecting to the network (due to device posture checks). This may indicate a likely loss of revenue due to the connectivity to the business environment. This can be used as a call to action to drive a time-bound remediation.
- **Relevant to business objectives**: Security teams must use their OKRs and define metrics relevant to measuring against business objectives.

- **Consistent and sufficient**: The information and insights in a metric should be consistent in their derivation and should be sufficient by themselves. For instance, the method to measure the click rate of phishing simulation should be consistently applied, and the relevant stakeholders must be furnished with adequate details needed for their role. For instance, the business manager would need to know the colleagues who failed the phishing simulation, whereas the COO would only need a top-level count or percentages.
- **Provide a trend of progress**: For instance, the number of privileged users using only the **privileged access management** (**PAM**) console for their administrative work. This increasing trend of metrics would indicate stronger governance of admin access.
- **Timely view**: Metrics should be published on time, including automatic alerting mechanisms. For instance, a metric on firewall rules with any source to any destination is configured but published only at the end of the month, which may have allowed undue risks to the network. Similarly, if the desired threshold of patching a critical vulnerability was seven days, then a report published only at the end of 30 days would render it useless.

  Using API's and a data lake, several of these metrics can be produced on demand with real-time/near real-time data.
- **Role-based access**: Metrics and their insights are a treasure trove. Organizations must apply required controls to ensure that only need-to-know principles are applied in creating and sharing such metrics.
- **Intuitive visualization**: Security teams often use a **Red, Amber, Green** (**RAG**) status to visually depict the status of the security program and its various dimensions. Irrespective of the colors used, the metrics and their visualization must be intuitive and should be simple to follow.

Figure 15.6 illustrates the expectations from metrics:

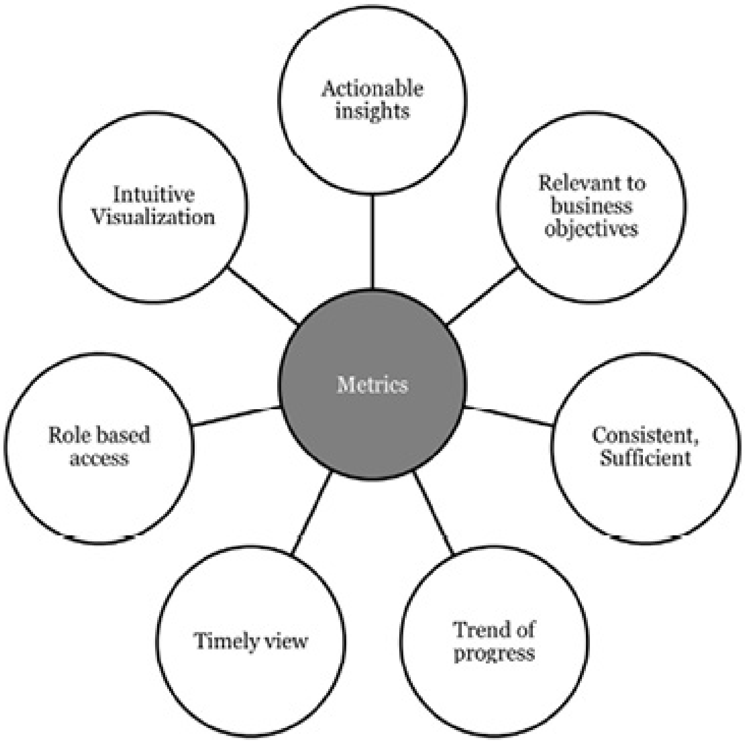


Figure 15.6*: Expectations from metrics*

Types of metrics

Every organization would have a variety of stakeholders that would need the metrics on the security program from the IT engineer to the CxO layer. Each of them would expect a different value and detail from the metrics. Organizations should attempt to follow the create-once-use-many times approach. It implies building on granular metrics needed for operational use (by IT engineers) and then adding insights for tactical consumption by teams such as risk managers, cloud security SME, or security architect, and then enriching it with business context and outcomes for the CxO layer. Several of the metrics are possible to create using API calls to various environments and then using visualization technologies like *Power BI*. F*igure 15.7* illustrates an example of the metrics that can be created based on their expectations:

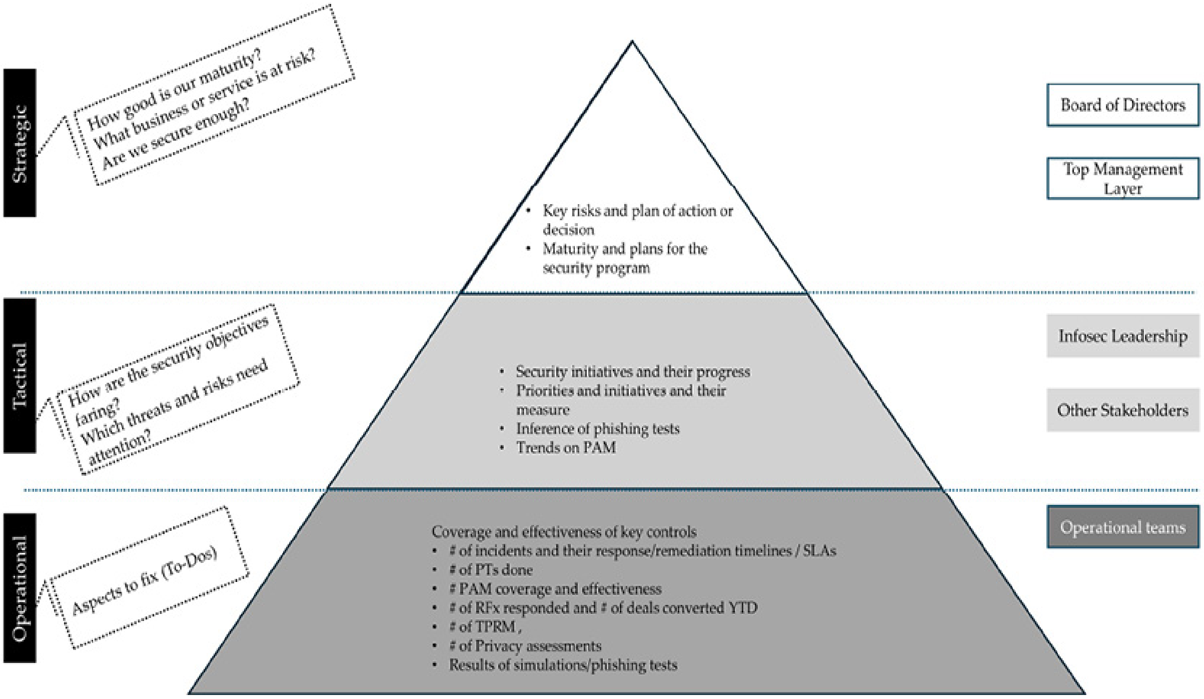

Strategic
How good is our maturity?
What business or service is at risk?
Are we secure enough?
Key risks and plan of action or decision
Maturity and plans for the security program
Board of Directors
Top Management Layer
Tactical
How are the security objectives faring?
Which threats and risks need attention?
Security initiatives and their progress
Priorities and initiatives and their measure
Inference of phishing tests
Trends on PAM
Infosec Leadership
Other Stakeholders
Operational
Aspects to fix (To-Dos)
Coverage and effectiveness of key controls
# of incidents and their response/remediation timelines / SLAs
# of PTs done
# PAM coverage and effectiveness
# of RFx responded and # of deals converted YTD
# of TPRM ,
# of Privacy assessments
Results of simulations/phishing tests
Operational teams

***Figure 15.7**: View of different metrics for various stakeholders*

Conclusion

In this chapter, we looked at a lot of detail about the security program and how to define, document, and implement it. We examined the various design considerations and the components of the security program, and how to measure them. We also covered the CISO's reporting hierarchies and a representative illustration of how a CISO can structure the security team. We can only improve things we can measure; with that light, we explore external and internal metrics at length, and how to define them in the context of outcomes.

In the next chapter, we will look at how organizations strategize their business and how security teams can participate and enable it. We will explore some case studies to learn from their approaches and define a recommended security strategy.

Key takeaways

Some of the key takeaways from this chapter are:

- The security program of an organization is an all-encompassing approach to protecting the CIA of its information assets by weaving together the organizational needs, security philosophy (such as Zero Trust), risk appetite, and the security objectives into actionable and measurable aspects.
- The program must adapt to the organizational strategy and objectives to spell out a clear security mission and objectives. These objectives should be measured and communicated to the required stakeholders at defined periodicity in a manner that creates relevance for them.
- Frameworks such as the NIST CSF and/or the **C2M2** may be used to define the vision and track the progress toward it.
- External infusion (and sharing) of ideas, thoughts, and conversations with peers, etc., i.e., **external breathing**, goes a long way in maturing the security program.
- An effective program requires continually testing the effectiveness of its controls and remediating issues.
- Using **ODM** helps the organization track the progress of the current state and journey to the future.
- While the details of the metric may vary by its intended audience, spelling out scales such as minimum required, optimal, and desired is helpful.
- Such programs need strong management support and adequate funding for tools and staff.
- The CISO and his/her leadership team must be familiar with foundational finance concepts and be able to draft and manage their budgets.
- Organizational dynamics, including the hierarchy, may influence the way a CISO might operate; however, a CISO would be expected to be the digital enabler and exercise influence without authority. Similarly, the CISO’s team structure may be influenced by organizational design elements, culture, etc. However, the focus should primarily be on clear goals and business outcomes.

References

- **https://www.unilever.com/news/press-and-media/press-releases/2023/unilever-goes-cloudonly-with-one-of-the-largest-cloud-migrations/**
- **https://m.rediff.com/money/2007/jun/23citi.htm**
- **https://www.gartner.com/en/documents/5138231**
- **https://securityscorecard.com/why-securityscorecard/security-ratings/**
- **https://help.bitsighttech.com/hc/en-us**

CHAPTER 16

Business and Security Strategy

Introduction

In this chapter, we will explore the meaning, purpose, and relevance of business strategy in the functioning of an organization. We learnt in the earlier chapters that the security objectives must be supportive of the business strategy. We will attempt to examine some of those strategies and derive a possible security strategy.

Structure

The chapter covers the following topics:

- Exploring business strategy
- Deriving security strategy

Objectives

By the end of this chapter, you will be able to interpret and understand the organizational vision and mission and their influence on the business strategy. You will be able to understand how a strategy can be formed and put into action. You will also learn to derive the security strategy from the business strategy. The approach will serve as an example for the way forward because each organization is different, and its operating constraints are different.

Exploring business strategy

Organizations, whether for-profit or not-for-profit, are formed and exist to serve at least one need (requirement). They attempt to fulfill a need/gap in the market with a product or service, or both. For instance, in the late 19th century, Switzerland was transitioning from a largely agrarian country to a rapidly industrializing giant across various sectors. This transformation meant a lot of people became busy with work. Nutritious and easy-to-cook food became an urgent need. That is where *Julius Maggi* introduced a ready-to-prepare concentrated soup. Similarly, the Tata Group founder, *Jamsetji N Tata*, was one of the foremost supporters of the cholera and plague vaccine that proved to be a huge lifesaver for the residents of Bombay (now Mumbai) in the late 19th century.

It may be noted that an organization may choose to, rightfully, make money and be profitable, and yet could also contribute effectively to social causes. In fact, in countries such as India, organizations are encouraged to spend a portion of their earnings on social causes, like education, healthcare, and sanitation. Such programs are called **corporate social responsibility** (**CSR**). Some other examples include:

- The *Mahindra Group*, a big conglomerate in India, has been on a mission since 2007 to plant a million trees annually (Project Hariyali).
- Reliance Industries, through its *Reliance Foundations*, has empowered 1 million plus women over several decades through various programs.

The foresight of the founders of the organization lends itself to the vision for the organization. i.e., the long-term aspiration for the organization (or even the group of companies) and what it wishes to be known for. For instance, the globally famous Tata Group is known for its philanthropy, worker-friendly, profitable businesses that are run ethically. The founding fathers of this centuries-old organization drafted the vision for the business to create positive social impact as a core of how they run their businesses and run them profitably.

The plan to achieve such a vision is built on foundational values, i.e., the guardrails and operating discipline. For instance, *Amazon*, the global e-commerce and cloud services giant, has customer obsession as one of its principles. This is exemplified by their e-commerce returns policy, their same-day delivery models, and so on. We previously introduced the importance of culture in *Chapter 13, Human Centric Security.* The vision and the values shape that culture. Culture tends to define the character of the organization and thus reflects on its operations.

While the vision drives the path to the destination, the mission helps make progress towards it with the short-term specification of what needs to be done. For instance, UnitedHealth Group, one of America's largest healthcare organizations, has set a mission for itself as *Helping people live healthier lives and helping make the health system work better for everyone*. This mission drives the energy and passion of everything the company does.

It is common for organizations to use vision, mission, and purpose interchangeably. The key to remember is that these are long-term, future-oriented, and directional expressions (statements) that drive the organization. While it is rare for a drastic change in an organization's vision, it is not uncommon for an organization to change its vision and mission statements with a changing ecosystem. For instance, in the 19th century, the Tata Group may have focused more on a mission to industrialize, while today they may focus more on services in several sectors like Aviation, Software services, and so on. The values or principles usually do not change over the years.

The word **strategy** has Greek origins and signified the *"art of the generals"*. In today's time, it may be defined as *a plan or method devised with steps to get to a goal (mission)*. The path to meeting the mission may have numerous options and various considerations, such as time, money, and skill. A strategy aims to define the path to achieve the stated mission with available resources and by when. As part of the strategy, the organization may make additional resources available. For instance, with the AI in focus, the organization may include data scientists and AI experts in its workforce to augment their **go-to-market (GTM)** plan and keep abreast with ensuring their product and services are in line with market needs. The strategic plan should continue to focus on creating business value.

Figure 16.1 illustrates the foundation's strategy built on a vision and mission guided by principles (values):

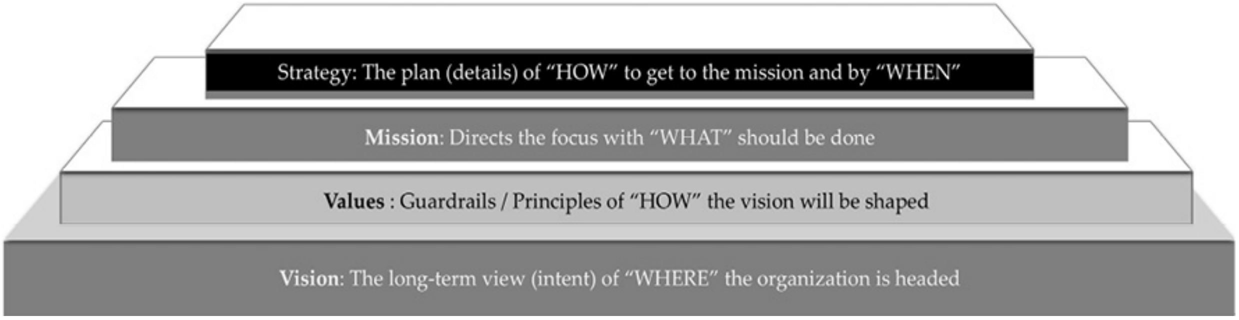

Strategy: The plan (details) of "HOW" to get to the mission and by "WHEN"
Mission: Directs the focus with "WHAT" should be done
Values : Guardrails / Principles of "HOW" the vision will be shaped
Vision: The long-term view (intent) of "WHERE" the organization is headed

***Figure 16.1**: Strategy built over vision and mission*

Armed forces adopt specific mission objectives for which several usable strategies are drafted, and one or more are chosen to execute. Many times, these strategies are a result of several war games and simulations, much like the incident response simulations in the cyber world. However, environments and scenarios are not predictable in armed conflicts. A strategy needs to be thought through quickly and implemented. Even in daily life, scenarios change, and the strategy needs to be tweaked and designed quickly. In March 2020, when the entire world was captivated by the devastating coronavirus (COVID-19), every nation had to quickly think about isolating citizens and treating the affected ones, developing/administering anti-virus, curtail travel and human contact, and many more. Utility organizations, such as power and water supplies, had to strategize on how essential services would run while keeping their workforce safe. Educational institutions had to ensure the children could continue to learn, leveraging technology like *Zoom*, *Teams*, and others. In the same light, platforms such as Zoom had to ensure that they had the leverage of cloud technologies to rapidly scale for the increasing customer demands.

It may be apparent that for an organization to be truly successful, innovation plays a key role. It is important to nurture innovation and allow it to help drive the transformation. We will cover an example of such an innovation-supported transformation later in this chapter.

Creating a strategy

There are many management approaches in defining and finalizing a strategy, and we introduce one of them here to leverage the same for the security strategy later on. As we covered earlier, the strategy is built on a clear vision and mission with defined values/principles to guide them.

Figure 16.2 illustrates one of the possible approaches to creating and implementing a strategy:

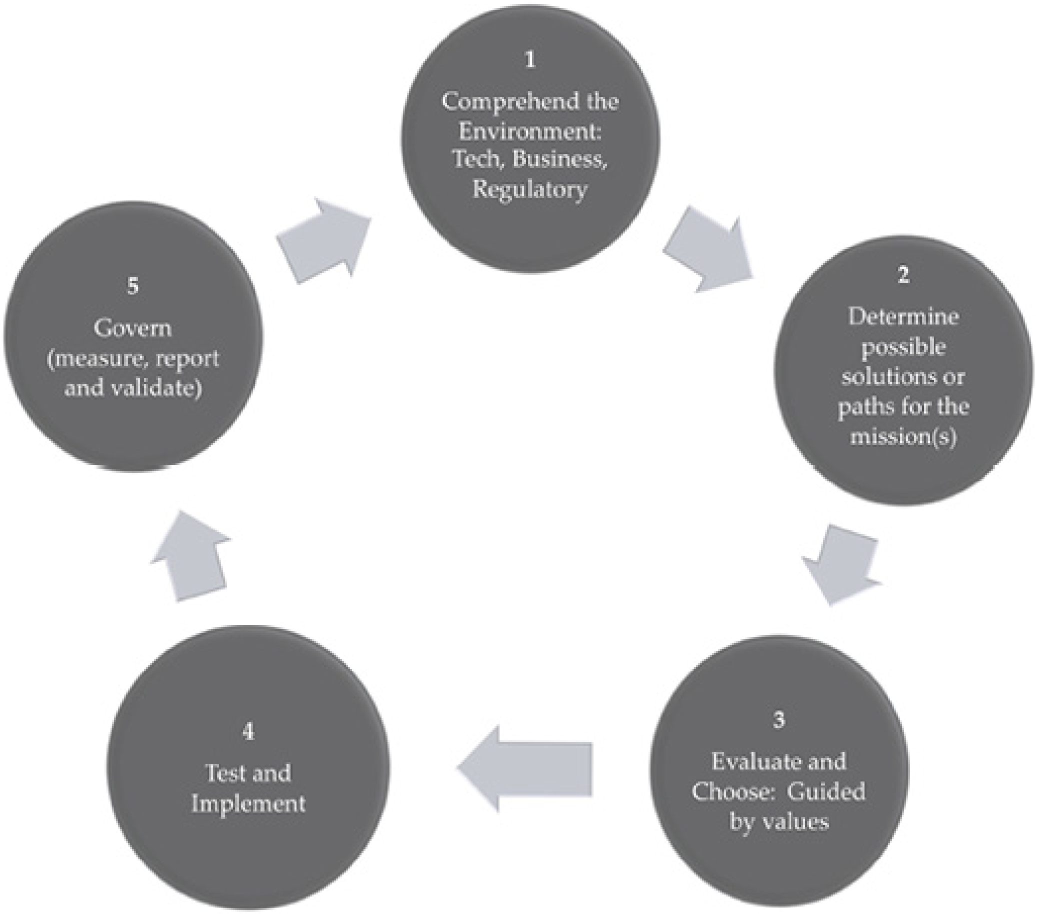

1
Comprehend the Environment: Tech, Business, Regulatory
2
Determine possible solutions or paths for the mission(s)
3
Evaluate and Choose: Guided by values
4
Test and Implement
5
Govern (measure, report and validate)

***Figure 16.2**: Process of strategy*

The process of strategy may use principles like those of *design thinking* to define, ideate, and prototype the solutions to finalize the possible solutions. The process would still need to be continually evaluated, using metrics and reports, and tweaked where needed. A good way to identify solutions/paths to the mission may be to use brainstorming. In a brainstorming exercise, the ideas are collected without any judgment or pre-conditioning of actual possibility. The core theme of brainstorming is to let ideas flow unfiltered and unconditionally in a time-bound exercise.

Management would choose the strategy from the options based on their risk appetite by factoring in complexity in execution, cost, capability, time, chances of success, etc. We covered risk management in detail in *Chapter 2,* *About Managing Risks*. *Figure 16.3* illustrates options and their criteria developed during a strategic planning exercise:

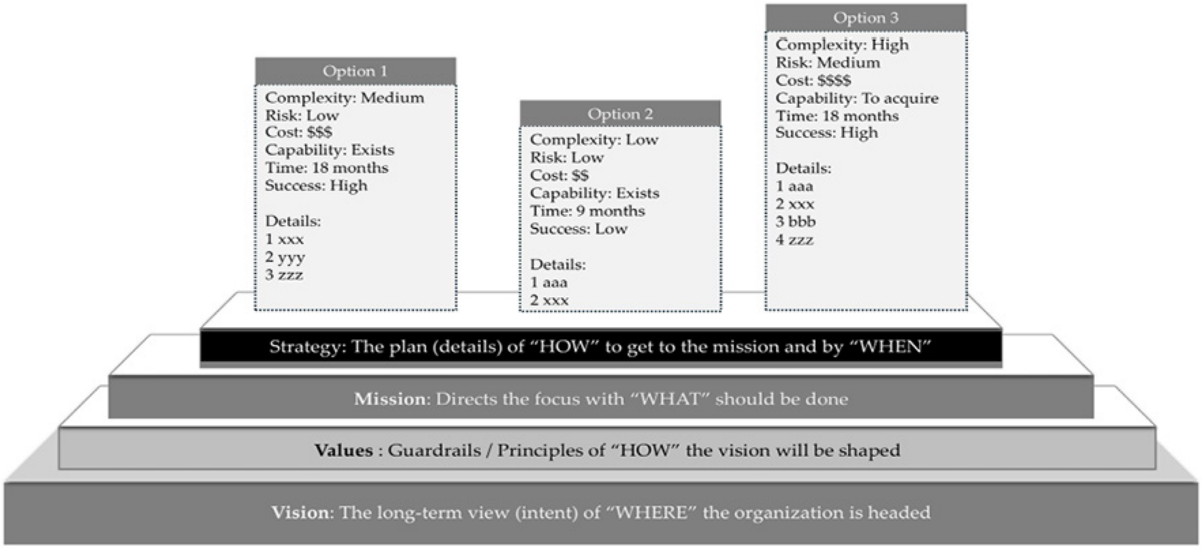


***Figure 16.3**: Strategic plans and their options*

Here are a few examples of strategies with the lens of the process defined above:

- In 2006, the **Government of India** (**GOI**) approved a project to identify and provide social schemes to service its citizens **below the poverty line** (**BPL**). With this mission, it went about leveraging technology to create a biometric-enabled identity that could be linked to provide amenities like foolproof/pilferage-proof delivery of essential food grains periodically. Within a decade, this identity system, called **Aadhar**, evolved to be one of the stellar identification mechanisms for taxpayers, applying telecom connections and even a driver's license, etc. The GOI had a clear mission, and its empowered team went about creating an ecosystem of process and technology to deliver that mission. The GOI also uses such means to facilitate **direct benefit transfer** (**DBT**) to the individual's bank account and thereby eradicating delays and potential fraud.
- Wiz.io, a cloud security company founded in 2020, has rapidly become a marquee name in cloud security. Their stated mission *to help organizations create secure cloud environments that accelerate their businesses* helped them accurately fill the gap of ensuring the rapidly

exploding cloud computing environment could be secured through a single pane of glass. Their leadership team was quick to identify a market gap, i.e., the need for cloud security platforms, and was able to quickly deliver a solution to determine the security posture of the cloud environments and secure them. We covered **cloud security posture management** (**CSPM**) and related aspects in *Chapter 8, Cloud Security*. Wiz's product team actively engaged with customers to continually refine and enhance their offering to make it relevant for the customer.

- Netflix started in 1997 as a DVD rental service. Within a decade, it had started transforming into a global streaming entertainment services company. It is a powerful and clear focus on *We are here to entertain the world, one fan at a time*. Probably drives their approach to new content creation (Netflix series), providing streaming content; latest movies globally, etc. Netflix had a fair share of failures in its journey, but what is notable is that it kept adapting and looking to the future. For instance, in the early 2000s, when the content on television with programmed content was the norm, they had probably already started thinking about the cloud and leveraging it to stream content using the Internet. The advent of smart TVs and enhancements to iPhones and Android phones has only expanded the delivery modes for their content. Clearly, innovation and making bold bets helped Netflix, and should serve as a lesson on continual adaptation.

There are scores of companies that might have been hugely successful by adapting to a changing environment and transforming. And yet there are many companies like *Kodak*, *Blockbuster*, *Kingfisher Airlines*, *ZebPay* that failed for a variety of reasons, including regulatory. Often, companies may get acquired or merged, and their unique proposition is eliminated by the acquiring company. For instance, Broadcom acquired *VMware* and, as of now, has eliminated key VMware offerings like *Workspace One* and *vSphere+*. Such changes have implications for IT and cybersecurity as well.

Note: In April 2025, Google Cloud announced the definitive agreement to acquire Wiz for an unprecedented $32 billion. The process of acquisition is subject to regulatory approval.

Deriving security strategy

In the previous chapters, we have covered the changing environment and its implications for the security program. In *Chapter 15*, *Managing a Security Program*, we also covered the desired features of a security program and highlighted the importance of management support. In *Chapter 3*, *Role of Standards and Controls*, we covered the process of control selection and implementation.

In today's interconnected world, organizations leverage capabilities and services provided by others to augment their own unique offerings. For instance, Wiz.io uses all major **cloud service providers** (**CSPs**) like *Azure*, *AWS*, and **Google Cloud Platform** (**GCP**) to deliver its offerings. This capability is relevant for its strategy to capture and retain the market.

Table 16.1 elucidates some possibilities of security strategies. These examples are neither meant to be exhaustive nor are they meant to reflect the posture of these organizations:

Organization and its mission	Suggested business strategy (partial)	Security strategy (partial)	Reference (Chapter number)
Aadhar program—GOI: *To empower Aadhar number holders of India with a unique identity and a digital platform to authenticate anytime, anywhere.*	Enrolment of and authorized modifications to an individual's Aadhar records using digital technologies.	CIA of the authentication records of Aadhar holders. Availability of the technology infrastructure. Protecting the Aadhar portal using appropriate controls like MFA.	Security and Privacy by Design (5) Identity and Access Management (7) Cloud Security (8) Zero Trust (9) Cyber Resilience (12)
Wiz: *To help organizations create secure cloud environments that accelerate their businesses.*	Securing customer's cloud environments. Visual representation of **attack path mapping** (**APM**).	**Application program interface** (**API**) security. Protecting cloud assets. Role based access. Shared security model (as a CSP). STAR certification.	Role of Standards and Controls (2) Cloud Security (8)
Netflix *We are here to entertain the world, one fan at a time.*	Create digital content. Provide access to digital content on demand.	Digital content security. **Intellectual property** (**IP**) rights. Access management Privacy by design.	Cloud Security (8) About Managing Risks (2) Continuous Threat and Exposure Management (10)

A large technology driven services organization using AI.	Responsible use of AI to solve real business problems such as faster and efficient accounts payable.	Network security. Endpoint security. AI including **large language models** (**LLM**). Security culture. Supply chain security.	Key Security Technologies (6) Continuous Threat and Exposure Management (10) Incident Response and Planning (11) Human Centric Security (13) Managing a Security Program (15)
A mid-sized manufacturing organization aiming to expand into consumer goods globally.	Manufacture the consumer goods in geo closer to the consumer markets.	**Operational technology** (**OT**) security. Physical security of manufacturing facilities. Regulatory considerations for privacy.	Key Security Technologies (6) About Managing Risks (2)

Table 16.1*: Table of security strategy derived from the organization’s strategy*

It is important to remember that the security strategy is a subset of the organization’s strategy and not an isolated one. The CISO may own the security strategy, but it must be tuned to deliver business value in line with the organization’s direction.

The selection of controls, which we learnt in *Chapter 3*, *Role of Standards and Controls*, will serve as a way to make informed decisions based on risk (covered in *Chapter 2*, *About Managing Risks*).

The security strategy will also cover the journey from the current status to the future state of maturity. We covered these gap assessments and the journey maps specifically in *Chapter 3*, *Role of Standards and Controls,* and *Chapter 15*, *Managing a Security Program.*

Role of management in security strategy

The role of the security team is to enable the business to run securely and within the confines of the regulatory and customer contracts. In *Chapter 15*, *Managing a Security Program*, we emphasized the need for management to be engaged in shaping and directing the security program. The security program is a sum-total of the security strategy derived from the business, regulatory, and technology environment. The top management’s participation in the security program is a foundational requirement for standards such as ISO27001. The management would need to act on the risk advice provided by teams such as that of the CISO to make informed choices. This symbiotic relationship augurs well for the security posture of the company and could be a key differentiator in the extremely vulnerable threat scenario.

It is possible that the CISO's professional opinion may be overruled by management. For instance, the CISO may have asked for all cloud accounts to be governed by CSPM. However, the management may decide against it for cost or convenience reasons. As a CISO, we would apply risk management principles to document the decision appropriately. It may also be food for thought on whether, as a CISO, the problem statement and solution were explained clearly and convincingly.

In some situations, the CISO may find himself/herself being asked to take an action that does not seem to align with organizational values and/or ethics. For instance, the management may prohibit reporting a material breach to regulators. Such a decision may not only be detrimental to the organization and its management, but also to the CISO's role. The CISO, in such a situation, must take the appropriate course of action that aligns best with his/her ethics.

Conclusion

In this chapter, we explored the definition and importance of strategies. We defined a framework based on an organization's vision, mission, and its core values, and how they shape not just the organizational strategy (business strategy) but even the security strategy part of it. We examined some examples of strategies and how the security elements may be derived from them.

In the next chapter, we will cover the importance of effective written and spoken communication that will help in the journey of managing risks, and also communicating the security posture of the company.

Key takeaways

Some of the key learnings from this chapter include:

- An organizational strategy aims to define the path of how it may achieve the stated mission with available resources and by when. This input is key to deriving the security objectives to enable the business to reach that objective.
- The security objectives cannot be divergent from organizational objectives and must cater to meet all regulatory and contractual requirements.
- Approaches such as design thinking can be used to define, ideate, and prototype the solutions, and then use a risk-based approach to finalize them as a way forward for the most optimal security objective.
- Management may exercise its choice, which may be different from that preferred by the CISO. It is important to apply the principles of risk management and ensure the decisions are recorded and communicated appropriately.
- Communicating the security strategy and its objectives to relevant stakeholders, including team members, is key to aligning and delivering positive outcomes on time.

References

- **https://web.stanford.edu**
- **https://uidai.gov.in/**
- **https://www.wiz.io/**
- **https://about.netflix.com/en**

CHAPTER 17
Effective Communication

Introduction

In this chapter, we will explore the importance of effective communication in a security program. We will understand and explore key concepts to ensure our communication is relevant, impactful, and easy to comprehend for our readers/audience. We will also explore some communication strategies for different internal and external stakeholders.

Structure

The chapter covers the following topics:

- Communicating effectively and with purpose
- Using communication modes
- CISO and the CxO engagement

Objectives

By the end of this chapter, you will be able to understand the importance of communication, the mindset of various stakeholders, and the methods of communicating with them. The success of a security program at various stages would require the CISO and the team to use these methods to be effective. One of the primary stakeholders of the CISO is the CxOs in the organization. We will examine how a CISO can engage with those stakeholders and be successful.

Communicating effectively and with purpose

Communication, i.e., the ability for us to express ideas, thoughts, feelings/emotions, decisions, or even just converse with another individual, could be a bane and boon. Effective communication is an art. It takes effort and intent to be heard (or seen) and understood correctly. A key dependency of communication to be effective is that the sender and recipient can interpret and understand the words correctly. For instance, a sentence spoken/written in English may not be understood by someone who speaks only German. In this book, we have considered the language of communication to be English.

Modes of communication

At the workplace, communication happens in a variety of ways, such as:

- **Written**: Such as over email, chat forums like MS-Teams, or social media posts like LinkedIn.
- **Verbal**: Spoken words over phone calls, video-conference technologies like Webex or Zoom.
- **Non-verbal**: These include visual signs of dejection, elation, and/or engagement, such as through body language, hand-eye movements, pacing the room, fidgeting with objects, using hand gestures, etc.

The focus of this chapter will be on formal communications by a CISO (or his/her team), i.e., day-to-day business interactions for security-related topics. While we consider that to be English, the concepts apply to all other languages, too. In this chapter, the terms message and communication may be used interchangeably.

Aspects of effective communication

In December 2021, a critical vulnerability, enumerated as CVE-2021-44228, was found in Apache's widely used Log4j software library. The vulnerability impacted practically every organization in the world across all sectors, like banking, manufacturing, retail, healthcare, and services. A CISO would be ineffective in communicating only the criticality of the vulnerability and the need to patch immediately without explaining the context to the relevant stakeholders. For instance, the patching of IT and operations technology, namely a manufacturing controller, may have caused a significant production downtime.

We will use the 5W1H (who, what, why, when, where, and how) approach to think about effective communication. This approach should be considered as a way to think about it, and not necessarily as a formula or fill-in-the-blanks. The 5W1H approach covers the following:

- **Who are we communicating to**: One of the most important facets of communication is the audience, i.e., the receiver of the communication/to whom you are communicating. Generally, the CISO and the security team would be in communication with internal teams and management. Factors such as the seniority of the receiver, their relevance to your message, and the means of communication will impact the content of the message. For instance, as discussed in *Chapter 15*, *Managing a Security Program*, a metric published to the board/CxO level leader vs that of the IT remediation manager/operator for the same vulnerability exposure topic would be different. Similarly, communications in general will vary in depth based on the audience. The Board level executives would be more interested in the big picture and topics like whether the CISO has things under control and/or needs any help. On the contrary, for an incident, the IT, HR, and legal department representatives need more tactical information. We can imagine knowing your audience being akin to **Know Your Customer** (**KYC**), but without the regulatory norms of needing identity proofs and such.

  Standards like the ISO27001 expect the security team to maintain contact with relevant authorities like the local cybercrime police and/or industry forums. In *Chapter 14, Managing Security Talent*, we covered a concept called **external breathing** that elucidated the need for an outside-in view into the security program.

  Public speaking, such as at seminars, webinars, and conferences, is an art, and it is not difficult to master. When speaking/presenting at industry forums or other public forums, it is fair to expect an audience with a mixed level of expertise and comprehension of the topic. Organizations often have policies on who can post what type of content on public forums and social media platforms. This is done largely to limit and contain the public content on organizational matters only to authorized individuals to manage branding and corporate communication aspects.

Tip: The content of the communication to external parties like regulators and law enforcem would generally be handled and dictated by the legal or public relations team.

- **Why is the message important to the audience**: It is more important to understand your audience and then draft your content to communicate *why your audience should care or relevance*. This *why* is also the purpose of your communication to enable you to leave your audience to take cognizance of your **point-of-view** (**POV**), understand and execute action, and/or get direction. For instance, when communicating information about a security incident to the CRO, the CISO would be better served to explain the risk, while briefing the same incident to the CIO, he/she may be explicit about which part of the application or infrastructure needs fixing.
- **What to communicate**: Generally, the *what* of the message, i.e., the content of the message, is likely to be one of the following:
  - **Progress/status report**: On a key program/control, or initiative. For instance, progress on the rollout of vulnerability scanning technology, or that of **multi-factor authentication** (**MFA**), or **Fast Identity Online** (**FIDO**).

    Other aspects include risk statements for cyber insurance and/or the stock market regulator filings, such as the 10 K in the USA.
  - **Awareness**: As cyber threats emerge from emails, messaging platforms, and even automated calls, the CISO's team would need to continually keep its workforce informed. For instance, communication about how to identify and not fall for phishing campaigns, deepfakes, and so on. We covered several of these in *Chapter 13*, *Human Centric Security*.
  - **Additional controls**: The CISO may need to periodically inform of upcoming changes/additions to security controls, especially if they impact end-user service. For instance, adding **business email compromise** (**BEC**) controls to the email security gateway may cause some emails to be blocked. The CISO's team would not only need to explain the change, the risks it solves, but also how the user experience may change.
  - **Seek direction**: The CISO may present his/her POV on the developments in the area of quantum computing and how the organization should look to manage the **post-quantum cryptography**

(**PQC**).

Irrespective of the content, to be effective, the CISO's team may consider the following aspects of the message:

- **Clarity**: Be able to relay the message such that the recipient is able to clearly understand and take the required next steps, wherever needed. There is no confusion on the ask/message. The message should be actionable, where applicable. For instance, the incident response manager would need to clearly instruct actions like isolating the impacted machine(s).
- **Context**: Communication must provide, even if briefly, context for the outreach. For instance, asking the IT team to perform urgent patching on key servers must be done for a specific reason, such as that the vulnerability being patched is being exploited in the wild. Such an inclusive context setting helps build trust and engagement with stakeholders.
- **No jargon or slang**: As much as possible, communication should not use jargon, especially when the audience is not expected to know it. For instance, while communicating the benefits of multi-factor authentication to the Board, it is not recommended to use jargon like phishing-resistant MFA or FIDO. Instead, explain the purpose using phrases like *an additional layer of security for the password*. Organizations with graphically spread teams may consider avoiding slang. For instance, speaking about a situation in American slang may only be enjoyed and understood by American colleagues. It is recommended to use normal spoken language.

- **Using no/fewer abbreviations**: IT and security have a lot of abbreviations, and not everyone in the audience may be familiar with all. It is recommended to use no or only very commonly used terms. For instance, while MFA may be commonly known or understood, internal stakeholders may not know what **endpoint detection and response** (**EDR**) is. In written and even in verbal communications, it would be a good idea to at least mention the expanded form at least once.

Tip: IT would mean information technology to most people, but it may mean income tax to som in the audience. It is recommended to set the context.

  - **Concise**: With fast-changing and busy work schedules, time is at a premium. It is important to quickly communicate impactfully using as minimal an amount of words as possible. The narrative should be such that the message is brief, yet complete. It should trigger interest in the recipient to engage, ask questions, or seek clarification. For instance, while explaining *how vulnerable are we*, it would help to succinctly communicate which environments are at risk and how much. The statement may be like *our e-commerce portal has no vulnerabilities, but some of our internal environments were vulnerable and are being fixed by this weekend*. We covered the **elevator pitch** in *Chapter 14, Managing Security Talent.*
  - **Factual**: The role of the security team, at least, is to provide a clear, unambiguous status of cyber risk as known at that time. The content should neither be alarmist nor unduly reassuring. It should be factual. For instance, when a one-off data loss incident happens, the security team is not expected to report that instance as a larger problem, nor is it expected to play it down as if nothing has happened.

  The CISO and his/her team are expected to have researched the topic and gathered several perspectives to bring a holistic picture. For instance, when a CISO proposes to ensure all machines must adhere to a particular encryption standard, it would be more effective to gather perspectives from IT on how they will implement, the support they need, and any technical limitations (such as obsolete laptop assets in use). It is important to point out that the CISO's job is to enable the secure running of the business, and thus, a conscious risk-based evaluation is key. We covered several such examples in *Chapter 16, Business and Security Strategy.*

- **When to communicate**: Information loses its relevance and impact if it is not communicated in time. We covered some of these aspects in *Chapter 15*, *Managing a Security Program,* in the context of metrics. CISO and the team would need to exercise cyber judgment to ensure relevant information reaches the relevant stakeholders on time. For instance, several regulations require breach notification to be within a certain time from detection. Similarly, when a **security operations center** (**SOC**) analyst observes a sophisticated attack on the organization's firewalls, it should promptly inform the stakeholders concerned in the organization. If such information is shared as part of the monthly metrics, it may be too late.

- **Where to communicate**: Organizations provide several communication channels to reach out to internal teams and stakeholders. Every organization's culture differs, and so does the most effective medium to communicate. Written communication channels serve purposes such as audience recall, reaching out to many people at once, cross-border communication covering time zone differences, and documenting decisions. However, the power of verbal communication, i.e., picking up the phone and talking or even meeting stakeholder(s), is underestimated.
- **How to communicate**: Everyone thinks and communicates differently. In an organization, the most popular means of communication are meetings, presentations, and emails. Several communications also happen verbally, such as on the phone and videoconferencing calls. Appropriate etiquette and respectful demeanor are expected, especially from the CISO. As an enterprise/organizational leader, it is important that the CISO casts the right leader's shadow for others to emulate.

  A CISO often exercises **influence without authority**, i.e., directs work and requirements from other teams. For instance, the changes to password length to be implemented on the **Active Directory** (**AD**) are typically done by IT teams. The security team does not manage the IT personnel. It is all the more important for the CISO and his/her team to be assertive yet respectful in engaging with colleagues in other functions. It is important to understand their perspective in implementing security controls and find a reasonable and yet secure way. For instance, the IT manager may indicate challenges in domain sync for the password policy to be effective and may need more time to tweak the architecture. Such perspectives should be respected, and any required support extended.

Using the 5W1H framework, effective email communication on the log4j vulnerability by the CISO or the security team may be as follows.

We have 210 critical servers that currently use the highly vulnerable software library called log4j. We have confirmed that cyber attackers are actively using this vulnerability to attack organizations to gain access to the organization's network, and/or launch a ransomware attack. While we have taken immediate preventive steps and have fixed our internet-facing servers, we will start deploying the permanent fix on other IT equipment on an urgent basis across regions this weekend using the follow-the-sun model. A downtime of 3 hours per region is expected, and relevant teams are on standby. Additionally, arrangements have been made for alternate production schedules.

Organizations used a variety of methods, formal and informal, to communicate. Some of the formal channels available in organizations and their suggested content type are tabulated in *Table 17.1*:

Mode of communication (Organization provided)	**Suggested content types**
Emails, such as on outlook application.	• Awareness messages. • Intimation of key change. • Summary of discussions. • Directional messages to the team/stakeholder (with asks to get something done) or take risk based decisions.
Chat applications, such as Slack or MS Teams.	• Quick decision by management—such a final direction of CFO for the ongoing purchase. • Quick update on some ongoing issue—for instance, the status of interview employee involved in a security incident. • Gaining context of a request/issue.
Collaboration forums on applications like MS Teams.	• Awareness messages (see *Chapter 13*, *Human Centric Security*). • Key information such as organizational plans and related security objectives (see *Chapter 16, Business and Security Strategy*). • Explaining security objectives for the team (covered in *Chapter 16, Business and Security Strategy*). • Operational matters of team, for instance status of vulnerability scans or coverage of EDR.
Intranet	• Awareness messages. • Thought leadership articles. • How-To articles, for instance, how to request an application security assessment.
Out-of-band (OOB) communication methods.	• Emergency or resiliency situations (see *Chapter 12, Cyber Resilience*).
Formal meeting/governance calls.	• Project status. • New proposals, for instance to introduce BEC controls. • Decision points based on data analysis, for instance whether to migrate from one technology to another. • Risk decisions. • Cyber threat landscape. • Program review.
One-to-One (1-0-1) meetings.	• Ideal for sounding an idea to colleague/peer or leader. • Individual performance feedback. • Inspiring team member for delivering outcomes.

Table 17.1*: Modes of communication*

Some of the modes of communication suggested above may be used in tandem or for specific uses during the same event. For instance, during a major security incident, the security team may use the OOB channel to initially inform key stakeholders and keep them abreast with crisp and yet short updates. Detailed action instructions, just enough context, may be provided as a directive to IT teams and tracked over platforms like MS Teams. While the HR/legal team is engaged for specific asks via email or formal meetings. Similarly, collaboration forums may be used for storing the artifacts as per the chain-of-custody requirements and/or the incident response plan. The CISO may brief the CRO/CEO on a one-to-one basis. A formal summary on a periodic basis may be published via email. Such a multi-modal communication approach is practical, effective, and provides for rapid and continual progress. It also supports the principle of need-to-know we learnt in *Chapter 1, The Triad of Security*.

People may prefer to use communication modes of chat applications like *WhatsApp*, *Viper*, *Signal*, *Discord*, etc. Though such modes are convenient, organizations would generally disallow their use in the acceptable use policy. Matters confidential to the organization, especially incident-related, must not be discussed or communicated over such modes.

Tip: An IT administrator may post indicating he/she is an administrator for a large complex set of servers and how wonderfully the environment is managed by a certain tool. Any and all such information may be used by attackers for social engineering and thus must not be disclosed.

Using communication modes

CISO and the security team may often use presentations to drive home the point. Some of the following tips may be considered for creating an effective presentation:

- **Avoid a lot of text or visuals**: A presentation with a lot of content will distract the audience, much like a slide with excessive visuals. The CISO and team would need to find the right balance. We already covered some suggestions for the facets of the content earlier in the chapter.
- **Use corporate-approved branding templates where feasible**: It is important to pay attention to approved font, color, size of text, and consistency of layout themes.
- **Use storyboard**: Set the context, aka the problem being discussed, the proposed solution, the ask from them (if any), and the next steps with clarity and in easy-to-understand words.

Here are some important tips for delivering a presentation:

- **Sending it in advance**: Senior management executives generally prepare well for their meetings. They like to see the content ahead of time so as to focus better and to get to their point quickly, and offer their decision.
- **Do not read the slide** word for word.
- **Engaging**: Attention spans, especially of the senior management, are shorter; their time is at a premium. Additionally, most senior folks understand things quickly, based on experience and general topical awareness. As a security manager/leader, it is of utmost importance to capture the attention and hold it, quickly communicating relevant information. Using storytelling techniques might be a good idea to effectively communicate.
- **Anchor slide approach**: The CISO may use an anchor slide with a few words or visuals to anchor the conversation. For instance, showing a stock price visual trending downwards as an impact because of an incident will more likely get attention than the technical details of a Kill-chain view (see *Chapter 10, Threats and Exposure Management*). This anchor slide will serve as a visual reminder of the cause and effect of the incident.
- **Using gestures**: At in-person meetings, using hand gestures and movements in moderation can create a positive impact on the conversation.
- **Maintain eye contact**: Establishing eye contact with the audience is effective in establishing credibility.
- **Prepare for the unexpected**: Like the cyber world, anything in a presentation can go wrong. The CISO's last edited change may unexplainably vanish, the projector may conk off, or the CxO may ask the CISO a tangential question or even a provoking question that was not prepared for. It is important to maintain composure and confidence. Often, it is not necessary to provide an answer then and there. Organizational systems generally allow for time to get the right facts versus guesswork.

Tip: A CISO may consider formal training on executive presence and public speaking to be effective, and using stage/webinar time, etc. Organizations often enable their leadership on such topics by leveraging leading educational institutes and experts.

Note: Over the years, the CxOs of organizations have increasingly invested time and money in hiring a personal coach. These personal coaches cover a variety of topics such as emotional intelligence, gravitas, executive communication, and personal branding.

Similarly, following email writing etiquette and being cognizant of the organization's culture are relevant. For instance, some organizations prefer a hierarchical mode to communicate both upwards and downwards, i.e., communicating as per the chain of command. While it may have benefits, in today's agile world with so many ambiguities, such an approach may be fraught with delays. Conversely, some organizations focus on outcomes and empowerment, and it is not required to deliver the message through a hierarchy. The CISO would need to navigate any such norms. It is not common for emails and their text to be misunderstood and/or misinterpreted. The tone of the email should be factual, assertive, but not aggressive or accusatory. For instance, a delay by the application team in implementing an appropriate digital certificate on the e-commerce portal should be communicated to the accountable business leader clearly, requiring immediate attention and action, but without using a tone that accuses that team of always delaying. The CISO can, however, articulate the number of times the team has delayed such an implementation, making it obvious to the application leader about the repeated nature of the issue and its impact on reputation and maybe even revenue loss.

Communication approaches may also be determined by nationality/geography. For instance, communicating with an American colleague may be less formal than with a British colleague. It may be important to recognize that one method is not better than the other, but is just different.

Tip: A security professional will be well served to understand and respect the cultural nuances of his/her stakeholders. There are good training courses like Globesmart available to leverage.

Figure 17.1 illustrates the 5W1H aspects of effective communication:

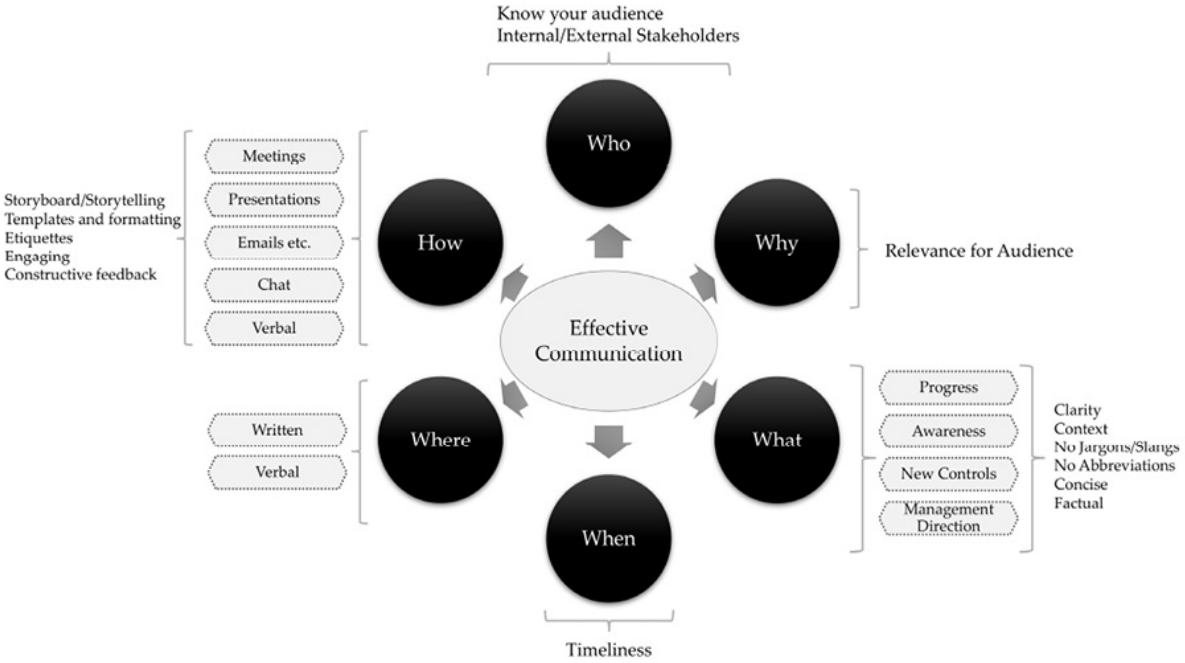


Figure 17.1*: 5w1h aspects of effective communication*

Perils of ineffective communication

Communication, when not effective, may lead to operational problems, security objectives not being met, and may even cripple the business objectives. It is pertinent that the CISO pays specific attention to the communication strategies, as it has a direct bearing on the team and on the organizational objectives. Some scenarios with ineffective communication are:

- **Internal stakeholders, IT/HR partners**: For instance, in many geographies, working on more than one job (moonlighting) without proper declaration to concerned employers is not permitted. To prevent such instances, a CISO may need to implement controls on validating the employee's time and focus on machines, use geo-location of connections, and so on. To implement such controls, the legal, privacy, IT, and business teams need to engage. Today, such technology probably offers less than 90% accuracy rates. The CISO would need to make sure the right tool and process are implemented, such that the objective is met and the expectation of the false positive is clearly communicated.
- **Organizational politics and undercurrents**: In modern organizations, where outcomes matter more, the hierarchy and reporting structures barely matter. However, when in conflict or doubt, humans may gravitate to

paying heed to the organizational hierarchy. It is possible that the IT manager may not want to execute the change control prescribed by the security manager.

This is where relationships at all levels of teams come in handy. The CISO and the security team must make a conscious effort to build and maintain relationships to ensure work can progress. Building relationships enables each other to be heard amicably, find solutions, and keep a growth mindset (see *Chapter 15, Managing a Security Program*). It would not be practical for the CISO to be engaged transactionally every time; hence, it is imperative that the security team also step in. One of the most important reasons why IT and security teams may be at loggerheads or in conflict is communication. Often, security teams may come across as aggressive or bossy and do not take IT partners into confidence. We discussed some of these concepts in *Chapters 14 and 15*.

- **Team members on the security program**: If the CISO and his/her leadership are unable to communicate the larger picture, the vision, and the security objectives clearly and/or in simple terms, the security team lead may not deliver the objectives. For instance, if the CISO does not communicate the need to run network discovery scans on a periodic basis, the vulnerability management team may continue to rely only on agent-based authenticated scans and not discover any rogue machines. Often, the security team members may only be technical and do not get the business context. It is imperative for the CISO and the CISO's leadership to help nudge such colleagues to think about the big picture. It is true that some individuals would not scale up to such thinking, and HR interventions may be needed.

  Organizations are complex ecosystems, and the CISO is part of that. Not all the decisions and/or requests of the security team members may be met. For instance, the security researcher on the team realizes the need to put a program together for **application program interface** (**API**), but the same may not be a priority for the organization based on its business model and funding available. In such cases, as well, the CISO should be able to communicate to the team in a cordial, professional, and accountable manner that the API governance may be prioritized later.

  In both examples, the team may get demotivated and not meet performance objectives as well. A CISO, like any other leader, would need to work through such situations and inspire his/her team.

- **Leadership engagement**: Like any other role, A CISO is expected to bring value and thought leadership to the organization. He/She should be able to communicate the effectiveness of the security program and how it meets the business objectives clearly, failing which the credibility and the relevance of the CISO and the security team will be in jeopardy. It may also impact the budget available for executing security objectives.

Tip: Security teams with a progressive and positive communication culture adopt a "yes and" approach instead of a "but and no" approach. In the former approach, the security professional provides a secure mid-path to the requester in lieu of saying no. For instance, if the marketing team wants to use Google Drive for file sharing with its external partners for all publicity materials, the security professional could approve this while requiring access controls, MFA, and data purging requirements based on risk.

For instance, if the CMO wishes to quickly launch a new digital marketing capability available from a **cloud service provider** (**CSP**) without putting the capability through review, the CISO should be able to communicate a *yes and* approach to set reasonable expectations, provide a clear **point-of-view** (**POV**), and, if needed, provide for a fast-track review. This would augur better than declining the request without investing the time to get the CMO's perspective.

CISO and the CxO engagement

In previous chapters, we have explored the role of top management in security strategy and the reporting structure of the CISO. We also learnt that the CISO often creates influence without direct authority. Each of the executives (the CxO) has a specific role to play and specific priorities to deal with. The world of information security is like an enigma, and yet, they have expectations from the CISO. On the other hand, the CISO is expected to comprehend the business nuances, the finance or HR practices, and so on, with similar clarity. This aspect, in fact, makes a CISO's role interesting and a true enabler that cuts across functions.

Table 17.2 illustrates the probable expectations of the CxOs/questions they expect the CISO to know and how the CISO might be able to use that in his/her strategy and communications:

CxO	Expectations from CISO's role	Some suggestions for CISO to communicate

CxO	Expectations from CISO's role	Some suggestions for CISO to communicate
Chief Executive Office (CEO)	Keep lights on and to keep the organization safe.	Cyber resilience and overall health of security program, including customer feedback. Linkage between organizational strategy and security.
Chief Financial Officer (CFO)	**Return on security investment (ROSI)**. Are we spending too much?	Strong zero-based budgeting. Strong financial discipline, Benchmarking on key objectives.
Chief Operating Officer (COO)	Keeping customers happy with no security issues. Ensuring operations are safe.	Governance on security programs, like phishing assessment results and customer audits. Success of security controls (on CIA) and initiatives in flight.
Chief Risk Officer (CRO)	Keeping cyber risk within risk appetite.	Progress on risk remediation.
Chief Information Office (CIO)/ Chief Technology Officer (CTO)	Keeping information and its related infrastructure safe and available.	Technologies and innovation enabled and issues fixed. Review of key security controls to augment availability. Applications reviewed.
Chief Innovation Officer **Chief Digital Officer (CDO)**	Be a partner to enable innovation at scale. Time to outcome. Data transformation security.	Technologies and innovation enabled. Turn around time for reviews. Patterns of data transformation enabled.
Chief Human Resource Officer (CHRO)	Ensuring personnel have access to required data. Keeping personnel data safe.	Initiatives to enable availability and yet security of CIA of personnel data. Participation in key HR data transformation initiatives.
General Counsel/Legal Head/ **Data Protection Officer (DPO)**	Keeping the organization compliant to regulatory requirements, for instance data subject rights requests.	Review of DSRs completed on time. Review of contractual security obligations. Overall security posture.
Chief Procurement Officer	Keeping access of third parties under governance.	Review of third-party security controls.
Chief Product Officer (CPO)	Securing the organization's product security.	Examples of partnering on security and privacy by design. Security issues prevented in product prior to release.

***Table 17.2**: CxO expectations and suggested strategy for CISOs*

It is recommended for the CISO to have an executive face-to-face with at least some of the CxOs. It may be highly dependent on the organization's culture.

Most of the time, the CxO would need to understand and appreciate the need to engage with the CISO from their busy schedules. This is where the quality of content, business acumen, and sharpness of the communication of the CISO will come in handy. For instance, if a manufacturing rival announces moving facilities from a war-torn nation to a seemingly safe nation, the CISO can bring in threat intel around how safe the operations are likely to be and what threats the competition might face. The CISO may offer such insights to the COO on how the organization should brace for that change by the competition while offering specific help that the CISO's team can provide. The intent is not to boast of knowledge, nor should it be to engage just for the sake of it. The content should add value to the recipient. As a CISO, it is suggested to be ready with an elevator pitch of one or two of the most important points to share with the CxO should they meet you in an elevator or in an office corridor. It is important for a CISO to recognize the time and place to start such a topic. It may not be appropriate to talk about security all the time. For instance, when a CRO happens to be in the same elevator as a CISO, talking about the weather and/or sports may equally be engaging. When asked about security topics, the CISO should be ready with the elevator pitch.

Conclusion

In this chapter, we explored the importance of business communication. We examined some typical use cases of the modes used in different types of communication. The importance of culture, geography, and choice of words in communication is quite underrated. Efforts must be made to be aware of them, appreciate them, and accept them. We also referenced a few topics from previous chapters to link the concepts. With the concepts covered here, we hope to have the CISO prepped up to be an effective communicator, which is a key ingredient to success.

In the next chapter, we will revisit the key learnings from all the previous chapters of the book and consider some suggestions for templates to use.

Key takeaways

Some of the key takeaways of this chapter are:

- Effective communication is an art. The sender and recipient should be able to interpret and understand the words correctly.

- The 5W1H (who, what, why, when, where, and how) approach may be used for effective communication with relevant internal and external stakeholders.
- The CISO and team may have to communicate on a variety of topics, at various periodicities, to a variety of audiences (IT personnel to senior management), and thus a multi-modal approach may be used.
- The CISO engages with top management, and each member of the management has a distinct role and focus.
- CISO would need to ensure they can speak the language the CxOs speak and make the content relevant.
- Practicing elevator pitches and being contextual and concise helps build credibility and stronger relationships.

References

- **https://www.cisa.gov/news-events/news/apache-log4j-vulnerability-guidance**

CHAPTER 18

Preparing For and Presenting to the Board

Introduction

The **board of directors** (**BOD**), or board for short, plays a key role in corporate governance, satisfying several regulatory requirements. Over the last 5-6 years, cyber security has increasingly become a core topic of the board's interest. Security leaders are often required to present to the board. This chapter will help you, as a CISO, to create a compelling narrative of your security program for the consumption of the board. It will help you think from a board's perspective and bring to bear the things that matter at that level of strategic focus.

Structure

The chapter covers the following topics:

- Purpose of the board
- Composition of the board
- Board's expectations
- Board and its oversight committees

- Important aspects for CISO to consider

Objectives

By the end of this chapter, you will be able to understand how a typical board of the company operates, how to think about what is important to them, and prepare well for a meeting with them. A lot of practice and a lot of preparation go into bringing up a crisp summary for the topmost strategic layer of the company—the board. Some of the same concepts may be applied to other strategic meetings, like the company's governance forums that have a CEO, COO, CFO, etc. We will try to understand personality types briefly and focus our preparation on that.

Purpose of the board

In an organization, the role of the board, i.e., its purpose, is largely focused on things such as:

- Ensuring the company's strategic direction is in line with the interests of the stakeholders. Stakeholders include mainly the investors, creditors, customers, employees, and such interested parties. While there is much more to it, let us call this **return on investment** (**RoI**).
- The business is run in an ethical fashion, and the board has oversight on ensuring the company's financial processes and its risk management practices are appropriate and fair. Typically, they have the power to question and provide guidance on these topics.
- The board endorses or even selects key executives like the CEO. And their remuneration.

The board should not and would not be engaged in the day-to-day operations of the company. The overall running of the business and its accountability lies with the CEO and her/his team. This team has people such as the **Chief Executive Officer** (**CEO**), **Chief Technology Officer** (**CTO**), **Chief Information Officer** (**CIO**), **Chief Financial Officer** (**CFO**), **Chief Marketing Officer** (**CMO**), **Chief Growth Officer** (**CGO**), **General Counsel** (**GC**), **Chief Human Resources Officer** (**CHRO**), and so on. Sometimes the **Chief Risk Officer** (**CRO**) and/or the Chief Information Security Officer may also be part of this leadership team.

Tip: This group of executives is often referred to as CxOs, where the x is substituted by the positions, such as finance for Chief Finance Officer.

Composition of the board

The board member of a company is appointed or selected to oversee that the functioning of the company is sound and ethical, and keeps the interests of the stakeholders at the top. The size of the board is determined by regulators and/or the **memorandum of understanding** (**MoU**) or the **Articles of Association** (**AoA**) that the company's owners/founders have signed. For this book, we just take it that there is a board of directors as part of corporate governance.

The company may have criteria for qualifications for its board members to ensure the quality of leadership oversight is appropriate. For instance, a company in the business of manufacturing autonomous cars may require 30% of its board to be from an engineering background. Similarly, companies may require a minimum number of independent directors who bring in the diversity of gender or any other factor.

Tip: Companies across the globe have started hiring/selecting members to their boards to represent the diversity of gender, race/ethnicity as part of their diversity and inclusion (D&I) program.

The board may have members internally selected and externally hired as well. Internal ones are employees who are named directors of the company and might be in any role, like operations or finance. The CEO is a part of the board.

External experts with key skills may be appointed to the board as independent directors. For instance, a manufacturing company with an interest in leveraging new-age technology like AI may approach an external expert to guide the strategy of the company's management. The expert will also be providing overall governance to aspects like proper/ethical leverage of AI.

Additionally, key investors with their money in the company may also tend to be members of the board. External board members and independent board members also bring confidence to the stakeholders of the proven track record these members may have. Generally, external members of the board may have their own businesses or jobs and may also serve on boards of other companies, bringing in cross-industry and cross-organizational infusion of ideas or checks.

Board's expectations

Remember, the board is there to oversee the stakeholder interests and ensure the company is run appropriately in compliance with laws and industry best practices. Here are some of the expectations of the board from the company's top management.

- The company has invested sufficient effort to define and support security programs. For instance, the CEO/CTO of an e-commerce company should have invested in processes and technologies to protect the payment systems.
- There are trained security personnel, hired or contracted, to ensure the systems work securely as required. For instance, the CISO to be hired by the company must have specific certifications or proven expertise in their field.
- There are appropriate risk management practices for important risks to surface to the top management. For instance, if the company's payment gateway system is unreliable and is impacting the financial reporting, the processes in the company are designed so that control failures are detected and reported up the hierarchy. Remedial measures, also called management response (refer to *Chapter 2*, *About Managing Risks*), are put in place.
- Transparency within the company in reporting facts on time. Notwithstanding the requirements spelled by regulators, the board may expect and even require that they be appraised of the significant deviations from expected norms. For instance, in the e-commerce example quoted above, the board may require the management to notify them within a certain time if the revenue is at risk beyond a

certain limit.

- Handling cyber security incidents appropriately. For instance, the board would expect the top management and the CISO to be prepared and conversant with the company's incident management plan. It would be expected that the leadership would be aware of how the business runs and what to do in case of a cyber incident. Effectively, they would think about various what-if scenarios of incidents and the playbook the company can deploy to deal with those adverse situations.

Tip: It is common for organizations to conduct an incident simulation not just for the top management of the company but also for the members of the board at annual or bi-annual frequency.

- Flexibility in adapting to changing geopolitical, market, environmental, technological, or customer requirements. For instance, when COVID-19 broke out in the first quarter of 2020, many things were impacted. Several airlines and the majority of their flights remained grounded for months. The airline had to quickly work out a process to minimally allow travel with a lot of thermal screening at airports, requiring passengers to wear **personal protection equipment** (**PPE**), etc. This nature of adaptability is key to survival and future growth. The board would look for such a mindset and culture of dealing with uncertainties.

Board and its oversight committees

To function effectively, the board provides its oversight through its committees. In this book, we will remain scoped to the committee that oversees cyber or information security risk. Such a committee is usually called an **audit committee** (**AC**) and is responsible for governing (among other things) the quality and depth of financial controls and coverage of topics of risk.

This committee typically has members of the board who are experts in enterprise risk, financial controls, and cyber. This is usually the committee the CRO/CISO interacts most with, depending upon the company; the AC may have several members to bring in the expertise. Some boards have even required former CISOs or serving CISOs of other organizations to be independent directors on the AC.

For the remainder of this chapter, we will use AC and board interchangeably.

Important aspects for CISO to consider

Cybersecurity/information security risks are increasingly becoming very hot topics in boardroom discussions, not just at the AC level but even at the full board level.

Unfortunately, there is no straight answer to stitch the CISO's update for such boardroom requirements. It hugely depends on the following:

- **Organizational culture**: Some organizations like things hierarchical, and some may be unstructured. Some organizations like to celebrate wins, while some focus on opportunities (losses).
- **Personality types**: Like any team, individual dynamics or personality traits may play out. For the record, the intention is not to judge anyone. The idea is to know the members of the board that you would interact with, and accordingly prepare. For instance, if the board member is analytical, s/he would need a lot more data to back up your narrative. They will be confident of your narrative if you can relate to the depth of analysis done to reach your conclusion. Similarly, if the member of your board is a technological geek, you would need to communicate appropriately, some of which was covered in *Chapter 17*, *Effective Communication*. The trick is in the right balance.
- **Industry trends**: For instance, higher adoption of **artificial intelligence** (**AI**) technologies, at least by organizations in services vs banking.
- **Recent geopolitical or industry incidents or events** may trigger the board's interest. For instance, at the time of writing this, a few regional conflicts have ensued, and these conflicts also bring about risks to data. Similarly, an attack on the company's competition may also trigger a deeper dive session.
- **The expertise of the members of the committee/board**: Sometimes,

the board members may be too hands-on and thus may have very specific questions or ask.

Preparing for the board meeting

To prepare for your meeting with the board, which may be quarterly or any such duration, you need to:

- **Align to management's positioning**: The positioning here implies what the topics or updates management (say the CEO/CFO) wants to inform the board about. You must be empowered to independently report on things that you professionally must. For instance, your board should know of a significant security incident that might have occurred. In the same breath, it is not appropriate to inundate the board with regular events like x number of cases of outbound data loss having been prevented. Depending upon the culture of the company, you would work with your management on the appropriate words to be used in communicating your thoughts without diluting or exaggerating the facts. This is also the time to be aligned on the future state of security objectives in the company's plans.

  In engaging with the board, you must be careful not to undermine the top management or your teams. For instance, if the board questions whether the amount of money being spent on security is enough, but you feel that you did not get that money from your CFO, a board meeting is not the place to bring that difference up. You must back your management and find smarter ways to get the funding you need from them. For instance, you wanted to implement a new and better proxy solution at an investment of $3 per user, but your management turned the proposal down as too expensive. It would be for you as the CISO to articulate how the technology brings about value, maybe contributes to operational cost reduction, or enhances revenue due to better productivity, etc.

  To be amply clear, as a CISO, your role requires you to maintain the highest levels of ethics and higher-order conduct. In fact, there have been cases in the past where CISOs have been part of litigation by regulators for not reporting things transparently.

Tip: In 2023, the CISO of SolarWinds was charged by the US Securities and Exchange

Commission (SEC) for his role in the cyberattack that happened on the company in 2019. According to the SEC, the company and the CISO made materially misleading disclosures about the impact of the cyberattack on the stakeholders. This was probably the very first case where a CISO was officially named in a regulator's investigation pertaining to materiality disclosure. This case triggered a lot of questions, such as the role of CISO in disclosures, the coverage of liability insurance for CISO, and the common definition of materiality of incidents. We will not go into the merits of the case and the outcome it saw.

- **Align to the company's objectives**: It is pertinent to ensure that your narration is contributing to the company's objectives and that you can articulate them well. For instance, if your company's plan is to expand its operations in regions in Africa, your security initiative around CNAPP should be able to explain the positives in protecting the application as the expansion happens there, and thus securing revenues or reducing regulatory risk. For those of you familiar with ISACA's courses and certifications on cybersecurity (like CISM), you would recall this concept as quite old and yet very relevant.
- **Know your board (KYB)**: You would need to spend time to understand from your management leadership and or platforms like LinkedIn who your board members are and what their focus areas are. It is almost like a recce you do before you buy a new real estate for yourself. This investment of time and effort is intended to help them think about their perspectives and see if your plan or your narrative is in the same direction. For instance, if the member of the board is a technology geek, and you are updating the board on the new age **cloud-native application protection platform** (**CNAPP**) technology being introduced, you would need to ensure you have enough technical knowledge to explain to the member if asked.
- **Use simple language**: Do not use jargon or abbreviations, especially from the security world. For instance, CNAPP may not be a topic everyone on the board is conversant with. Your focus should be to convey the value the technology brings in simple language. In case you are required to have a 1x1 conversation with a technologically conversant board member, you may use such lingo with their permission.
- **Prepare, prepare, and prepare a bit more**: You should target to have your materials ready and even your talking points ready much before

the meeting, and you must practice them by yourself or in front of your management. The more you prepare, the better it shows off in the meeting. Your audience will walk away knowing you have the right skills and depth to carry out the role, and more importantly, you took this meeting seriously. That does not mean you cannot use humor or have a smiling demeanor at the meeting. The practice of your talking points shall also help you in logical sequencing, choice of words, length of your narration, and voice modulation. Remember, often and highly recommended as well, the prepared materials (presentation slides, Word documents, etc.) are sent to the board members well in advance, something like at least a fortnight.

- **Communicate effectively by**:
    - Being focused on the areas of interest, the board wants to cover.
    - Bringing in a required business context, and if possible, impact on revenue, growth, and returns.
    - Being brief—to the point.
    - Active listening:
        - Pay attention to what is asked and answer accordingly.
        - Ask clarifying questions if needed.
    - Using simpler language and day-to-day analogies where possible.
    - Truthful—representing things as they are.
- A few of the themes the board may be looking for:
    - How are the security controls working?
    - How do we compare with the rest of our peer group or geography?
    - Is there a satisfactory return on investment on the security spending?
    - What are the key security initiatives in flight or being planned, and why? How would the organization benefit?
    - How do the security initiatives/elements of the program contribute to the company’s objectives?
    - How does a security incident damage the reputation or brand of the

    company?
    - Are you funded for your programs appropriately? Or are we over-investing in security programs?
    - Do you have the right team?

In answering all the above and similar questions, think about the question behind the question. The deeper intent behind asking the question and your response may accelerate or impede the company's strategy. For Instance, in answering the question, *Do you have the right team,* there could be any of the following reasons they want to evaluate:

- Is the team trained and capable of doing what is needed? For instance, without a trained accountant, you cannot create your financial statements; similarly, without a trained security professional, you may not be able to manage incidents and the damage that may follow.
- Are you, as a CISO, thinking about your succession planning? That is a key leadership element. Your next line should be groomed to take on higher-order tasks.
- Do you have enough people to manage the workload? Your answer must be thoughtful and constructed to reflect upon you and your team's ability to do all things that matter. There may be some wish list items that are not getting done for lack of workforce, and that may be fine.

Conclusion

In this chapter, we covered the composition, purpose, and functioning of the board. As a CISO, you may be presenting or engaging with the board; it is important to have the right domain/functional knowledge (information security), the right context of the business, and the right means to communicate your point of view effectively.

Key takeaways

The key learnings from this chapter include:

- Over the last few years, security has become a core topic of interest to the board.

- As a CISO, presenting or engaging with the board might be required; it is important to have the right domain/functional knowledge (information security), the right context of the business, and the right means to communicate the point of view effectively. This may be done using the 5w1h framework defined earlier.
- A CISO would need to get in the shoes of a board member to comprehend what their interest and focus areas are and then prepare for the periodic briefings and engagement.

Join our Discord space

Join our Discord workspace for latest updates, offers, tech happenings around the world, new releases, and sessions with the authors:

https://discord.bpbonline.com

Index