

term *key space* also refers to the number of keys available, based upon the key length; so, key space = number of keys available.

The attack times noted in the table apply to a system built in 1999 at a cost of \$250,000. The system was called “Deep Crack” and was developed to break DES. Modern computing technology and capabilities would reduce the attack time for a 56-bit length key to minutes and the other key lengths to a fraction of the years required by Deep Crack, but even then, the attack times would not decrease enough to make a significant difference.

Key Length	Key Space	Attack Time
56	7.2×10^{16}	20 hours
80	1.2×10^{24}	54,800 years
128	3.4×10^{38}	1.5×10^{19} years
256	1.15×10^{77}	5.2×10^{57} years

Table 3-55: Brute-Force Time Relative to Key Lengths

Additional Cryptanalytic Attacks

Understand the different cryptanalytic attacks

Because even with modern computers certain key lengths would require too many years to successfully brute-force, other ways to determine the key can be utilized. A comparison of the most common cryptanalytic attacks can be found in [Table 3-56](#).

Known by Cryptanalyst					
Attack	Algorithm	Ciphertext	Plaintext	Device	Details
Ciphertext only		✓			Most difficult attack
Known plaintext		✓	✓		The plaintext/ciphertext pair is available
Chosen plaintext	✓	✓	✓	✓	All elements except the key are known, and the plaintext is attacked
Chosen ciphertext	✓	✓	✓	✓	All elements except the key are known, and the ciphertext is attacked

Table 3-56: Cryptanalytic Attack Comparison

From the four types of attacks in Table 3-56, some may be more effective than others. A **ciphertext-only attack is the most difficult, because the attacker ONLY has access to the ciphertext**. In other words, the attacker has no real way to determine the actual plaintext except through enormous and timely effort.

With a **known-plaintext attack**, the attacker has access to both the ciphertext and plaintext, and both sources of information can be used to determine the key. More importantly, once the key is determined, two things can happen. One, other pieces of ciphertext can be easily decoded, and two, the key can be used to forge messages.

By far, the easiest cryptanalytic attacks are chosen plaintext and chosen ciphertext. In each case, everything except the key is known, which makes determining it much easier. In the case of a chosen-plaintext attack, the attacker can feed plaintext into the device and examine the resulting ciphertext to determine the key; in the case of a chosen-ciphertext attack, the opposite happens, and ciphertext is fed into the machine with the resulting plaintext being scrutinized. It's important to note that even with access to a device and the algorithm, the actual key must still be deduced through examination of the resulting outputs.

In addition to the attack types noted above, other attacks exist, like:

■ Linear and differential cryptanalysis

Both types of attacks use complicated math to deduce the key. In each case, multiple iterations of the attack are conducted to determine probability values of a given key being the key used for encryption. Differential cryptanalysis employs a form of chosen-plaintext attack, while linear cryptanalysis uses a known-plaintext attack approach.

■ Factoring cryptanalysis

In this attack, the attacker is trying to factor a very large number to determine the private key. In other words, the equation used to create the key is fairly easy to perform in one direction—multiplying two very large prime numbers together—but the reverse is not true; going from the very large number to the two original numbers is virtually impossible and

could likely only be accomplished over a very long period of time. **This type of attack is specifically focused on the RSA algorithm, which uses factoring as the underlying hard math problem.**

3.7.3 Cryptographic Attacks

CORE CONCEPTS

- The goal of cryptographic attacks can vary.
- Multiple types of cryptographic attacks, including popular ones like man-in-the-middle, replay, side-channel, social engineering, and ransomware.

As noted above, the goal of cryptanalytic attacks is to deduce the key. With cryptographic attacks, the same is not always true, as will be discussed in the various cryptographic attacks noted in the rest of this section.

Man-in-the-Middle Attack

Understand the different cryptographic attacks and key characteristics of each

In this attack, denoted in [Figure 3-67](#), the attacker pretends to be both parties in relation to the communication. The attacker places themselves in the middle of the conversation between Alice and Bob. When Alice sends a message to Bob, she's really communicating with the attacker, who will respond and act as if he's Bob; likewise, if Bob initiates the communication, the attacker will pretend to be Alice. Via this type of attack, the attacker can control the flow of information very carefully and potentially learn valuable information that can be used in other attacks.

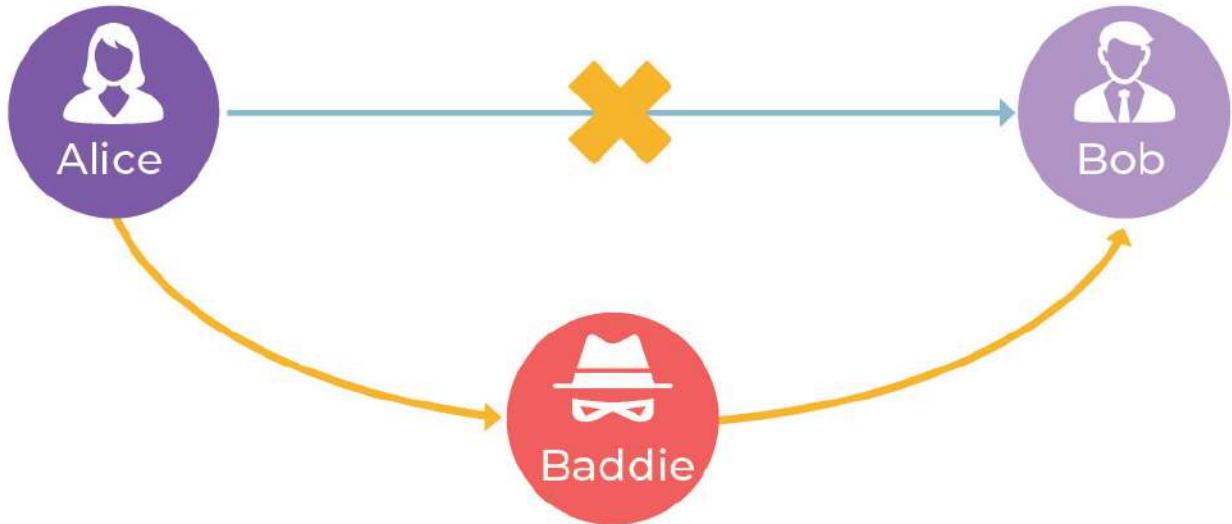


Figure 3-67: **Man-in-the-Middle Attack**

Replay Attack

A replay attack, also presented in [Figure 3-68](#), is like a man-in-the-middle attack, as the attacker is again able to monitor traffic flowing between two or more parties. However, this time the attacker aims at capturing useful information (like session identification details or authentication information) and then “replaying” it later to gain access to a target system. An example of that can be when a user logs into a system, and the password is hashed for security purposes; so, the username and hashed password are sent across the network, and if the hashed password matches what is stored in the system, the user will be successfully authenticated, commonly establishing a session with a specific identifier.

An attacker performing a replay attack can obtain a copy of that session identifier as it was communicated via the network and try to replay it to the server to denote the legitimate user.

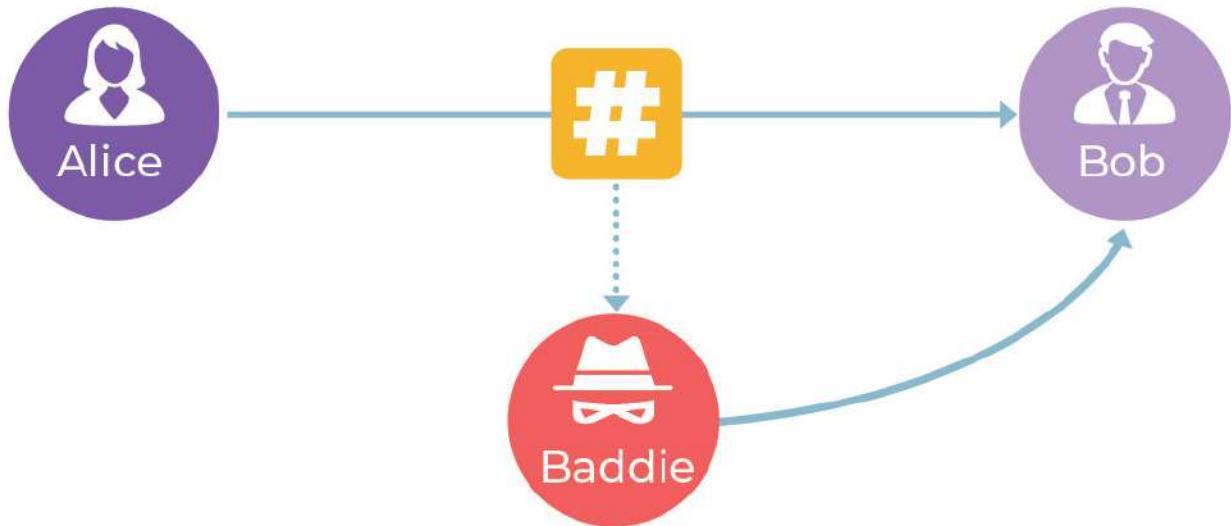


Figure 3-68: **Replay Attack**

Pass-the-Hash Attack

The goal of the attacker during this attack is to gain access to valid password hashes that can then be used to bypass standard authentication steps and authenticate to a system as a legitimate user. Since the attacker doesn't know a valid password for a user, they can try to intercept the password hash and then, instead of brute forcing that to try and obtain a password, they can present the hash to a target resource in the hope that it will be tricked into thinking that the attacker is the legitimate user to which this hashed password relates.

Temporary Files Attack

For any decryption or encryption activity to take place on a system, certain things must be present on the system, like the plaintext or ciphertext, as well as the algorithm and the encryption/decryption key. During the encryption or decryption process, the key might be obtained from a secure storage area and temporarily stored in RAM or another volatile memory location to facilitate quick decryption/encryption calculations. At this point, if an attacker can gain access to the system, they might be able to read the memory space and gain access to the key, which could then be used for much broader purposes.

Implementation Attack

Implementation attacks focus on an inherent weakness in how an algorithm is implemented rather than a weakness with the algorithm itself. An example of this is WEP, which is fundamentally broken. The algorithm underpinning WEP—RC4—is actually a great algorithm, but in the case of WEP, RC4 is implemented very poorly. Specifically, there is a problem with the initialization vector (IV), among other things, that allows RC4 to be cracked very easily and quickly. The IVs used are too short, and due to that, they're repeated too often.

So the issue is not the algorithm itself, but rather the way the algorithm is implemented, which makes WEP inherently insecure.

Side-Channel Attack

Side-channel attacks are very sophisticated attacks and are therefore only used by equally sophisticated organizations, like intelligence agencies, advanced persistent threat groups, and security researchers.

With a side-channel attack, the target is not the system or the algorithm itself. Rather, using complex tools, a system's operations can be monitored and measured. Based upon these observations and measurements of items like timing (the length of time to perform an activity), power used (how much power is consumed during an activity), and radiation emissions (emissions made by all devices and systems), significant insight can be gleaned.

Table 3-57 summarizes the different types of side-channel attacks.

Timing	Radiation Emissions	Power
Focuses on the length of time of an activity	Focuses on the emissions made by all devices and systems	Focuses on how much power is consumed during an activity

Table 3-57: Side-Channel Attacks

Dictionary Attack

The goal of a dictionary attack is to determine somebody's password. Rather than trying every possible combination of letters and numbers, an attacker will utilize a much **more efficient method** and **try the most likely possibilities/combinations of words** to determine a password. In the past, actual dictionaries were used. Now, these attacks can be enhanced by using databases of leaked passwords that exist and are available on the internet as the result of past breaches.

Rainbow Tables

Passwords should never be stored as plaintext in a system database. Rather, the hashed value of the password should be stored, so in case of a system breach, an attacker would only have hashed values instead of actual passwords. From the perspective of an attacker, a rainbow table anticipates such a situation and is a **precomputed table of hash values, based upon the most popular passwords used in the world and the most popular hashing algorithms**.

Essentially, a rainbow table is a big database that contains passwords and hash values, and it can be used to quickly determine specific hash values used for authentication. An attacker can steal a password file (which contains hashed passwords) and use the rainbow table to identify matches of those hashed representations.

Reducing the Risk of Rainbow Tables

The best way to reduce the risk related to rainbow table attacks is to use what's known as "salt." Salt is a random value appended to a password, which is then hashed. A rainbow table attack will not work against a database that has been salted because now instead of just the password being hashed, there's a random string appended to it before hashing takes place. This also has the added advantage that it doesn't allow the attacker to identify two users who have the same password. If both Alice and Bob set a password of "password123," then a random salt will be appended to each password before getting hashes, thus resulting in two totally unrelated hashing representations. Each user is assigned their own unique random salt value.



Peppers serve a similar purpose as salts, but instead of having a unique random value for each user, the same pepper (the same random value) is used for all users in a system. Salts are therefore considered to be more secure than peppers.

Birthday Attack

A birthday attack is used to identify collisions in hashing algorithms. Remember the birthday paradox statistics mentioned earlier, denoting that in a group of twenty-three people there's 50 percent chance of two of them having the same day and month of birth. That fact could represent a collision. **Collisions, hashing, integrity, and birthday attacks typically point to each other.**

Social Engineering

In addition to the more tech-centric attacks noted above, social engineering attacks can be used to obtain a cryptographic key as well. Two such attacks are: **purchase key attack** (bribing someone to obtain a copy of the key) and **rubber hose attack** (use of duress or torture to obtain the key).

Kerberos Attacks

Kerberos exploitations may take on several different forms, and each involves a password hash related to specific elements of Kerberos. One such exploitation is actually an extension of the pass the hash attack noted above. In this case, after an attacker uses a stolen password hash to authenticate as a legitimate user, the password hash is then used to create a valid Kerberos ticket.

Other Kerberos attacks involve what are known as golden and silver tickets. In the former case, an intruder who can obtain the KRBTGT account password hash has access to what's known as a golden ticket. The KRBTGT account is the Kerberos Key Distribution Center (KDC) service account and is responsible for encrypting and signing all Kerberos tickets. Through this access, an intruder can forge ticket granting tickets (TGT), which means any account in the active directory can be exploited. With a golden ticket and through interaction with the KDC, an intruder can request ticket granting service (TGS) tickets, which give access to system-specific resources, like a SQL server or an application server.

The latter case involves the forging of Kerberos TGS tickets, by virtue of an attacker gaining access to the password hash of a target service account, for example, for a service like SharePoint, MSSQL, and so on. In this context, TGS tickets are known as silver tickets, because the scope of access is limited compared to the access of a golden ticket—the KRBTGT service account password hash. Silver tickets only allow for access to a particular resource, including the host system. However, the ability to forge silver tickets also means that an attacker can create TGS tickets undetected, because interaction with the KDC is not required.

Ransomware Attack

Ransomware exploits have grown significantly in recent years, and one reason has been the corresponding growth in popularity of cryptocurrency. By nature, cryptocurrency is shrouded in secrecy, which aligns perfectly with the goal of actors involved in ransomware attacks—to remain anonymous or untraceable. Ransomware actors could be acting on behalf of nation-states or simply cybercriminals seeking to enrich themselves, and the target of attack could be anything from critical infrastructure to valuable intellectual property, and anything in between that could lead to a significant ransom being paid. When victim files are encrypted, sadly even paying the ransom doesn't guarantee the victim will get access, as that totally depends on the attacker's whim. Also, sometimes the file decryption may not work, which means that file access can't be restored. Lastly, various victim companies hire third parties to negotiate ransom and get file access back for them or work with law enforcement and security companies to see if they can identify any encryption vulnerabilities that would allow them to identify the encryption key and obtain their coveted files.

Fault Injection Attack

Fault injection is a technique that can be used to determine where vulnerabilities might exist. Typically, fault injection involves deliberately injecting a fault into hardware or software to modify its normal behavior, which then allows identified vulnerabilities to be corrected. Fault injection is often used during hardware and software testing. With a fault injection attack (FIA), however, an attacker is attempting to change normal behavior in order to exploit something, for example, an access control mechanism. Fault injection attacks are sometimes used in conjunction with side-channel attacks (like a differential power analysis) where the fault injection attack can be used to reduce countermeasures, and then the differential power analysis attack can take place.



3.8 Apply security principles to site and facility design

3.8.1 Intro to Physical Security

CORE CONCEPTS

- Physical security seeks to protect an organization from the outside to the inside.
- The primary goal of physical security is the safety and protection of human life.
- Physical security goals
- Threats to physical security

Goals of Physical Security

Physical security, like logical security, ultimately focuses on increasing the value of an organization by providing protection from outside the perimeter to all assets within, including protection related to confidentiality, availability, and integrity as depicted in [Figure 3-69](#). For example, let's consider

availability. Physical security is responsible for things like ensuring that the building's temperature is consistent and that there is a good clean supply of electricity.



Figure 3-69: **Information Security Goals**

Computer systems can't be up and running and provide good availability without proper physical security controls in place. Physical security is about protection of the organization from around the building all the way to the inside and with the same overarching goals of providing confidentiality, integrity, and availability.

Know the primary goal of physical security

Like logical security, physical security employs several controls, but the names are slightly different. With logical security, preventive, detective, and corrective controls are used; with physical security, deter, detect, and correct are used. Deterrence is focused on preventing. Detecting is focused on

identifying something. Assessing and responding is focused on correcting. More detail on these can be found in [Table 3-58](#).

Primary Goal of Physical Security

Safety and protection of human life is the most important goal and focus of physical security, and this fact drives many of the decisions related to physical security. In simple terms, this simply means that the implementation of any physical security control should never put people in danger.

Deter/Prevent	This type of control serves to deter or prevent an intruder from taking certain action. A sign in front of someone's property stating "All trespassers will be shot" or a sign stating "Danger: Mines" is a great example of a deterrent. A fence is a good example of a preventive control. The overall idea with preventive/deterrent controls is to try to prevent an attacker from attacking. While it would be ideal to prevent an attack from being successful, this is not a reasonable expectation, as all controls can be defeated.
Delay	Delay controls function to hinder activity being pursued by an intruder. Locks are a great example of delay controls; they serve to delay an intruder's activities. As such, they should never be counted upon by themselves and should be used in combination with other controls such as cameras to detect an attacker attempting to pick a lock and security guards who can assess and respond.
Detect	Detective controls help detect or alert to an intrusion. A barking dog is a good example of detective control; when the dog is quiet, all is well, and when the dog is barking, something might be amiss and needs to be investigated. A CCTV camera is another good example of a detective control, as it can detect and capture activity and help drive a proper response.
Assess	Assessment can lead to a proper response to a given situation. Assessment functions in the same manner as a corrective control.
Respond	Aligned with "assess" is "respond," which also functions in the same manner as a corrective control and denotes taking appropriate action to an event that has taken place.

Table 3-58: **Physical Security Controls**

Threats to Physical Security

Threats to physical security might take any of a number of forms, and a list of them could grow very long. Below are a few examples:

- **Theft:** The attacker steals items from the premises of the target.

- **Espionage:** Usually sensitive or proprietary information (like a new drug research) is targeted by the attacker, who aims to obtain that and sell it to a competitor or on the dark web for monetary gain.
- **Dumpster diving:** Inspecting a company's trash to obtain sensitive information that might not have been disposed in a secure manner.
- **Social engineering:** Leveraging the human element and trying to persuade a company employee to perform an action or provide information to an unauthorized individual.
- **Shoulder surfing:** A form of social engineering where the attacker is standing over someone's shoulder as they perform an activity (e.g., logging in to a sensitive resource) in the hope to gain sensitive information.
- **HVAC:** Compromise the heating, ventilation, and air conditioning system for either access or damaging company equipment.



3.8.2 Layered Defense Model

CORE CONCEPTS

- **The best physical security involves a layered, or defense-in-depth, approach.**
- **Physical security must always consider the safety and protection of people.**

One of the main concepts applied in physical security is the concept of layered defense. Essentially, this is defense-in-depth. Multiple layers of defense are put in place, starting outside the building of an organization. A typical first layer of physical security defense could be a fence. Then the next layer would be the perimeter of the building. But remember that human life is always at the forefront of any security system. Let's consider a question in relation to that. What's the optimal number of doors to have on a building's perimeter? The most optimal number of doors from a security standpoint is zero; however, this is not functional. So what's the next best answer? The next best number would be one, but that may impact personnel safety because buildings need to have multiple egress points so people can easily exit the premises in case of emergency. So that makes the ideal answer "as close to zero as possible."

3.9 Design site and facility security controls

3.9.1 Security Survey

CORE CONCEPTS

- **Security, or site, surveys are an extension of risk management and involve: threat definition, target identification, and facility characteristics.**
- **Crime Prevention Through Environmental Design (CPTED) is a specific professional practice that outlines guidelines and best practices regarding the design of buildings and surrounding structures, considering the environment as well as nearby infrastructure and facilities.**

How do we decide which physical security controls to put in place?

First, the most valuable assets and their associated risks must be identified; then, risk treatments can be evaluated and the most appropriate—based upon value—can be put in place. Because the focus is

physical security, the risk management process is sometimes called a **security survey** or site survey; still, the goal remains the same: identify the most valuable assets and then consider risk treatments, based upon the vulnerabilities and threats to those assets.

This is the process in security, or site, survey: threat definition, target identification, facility characteristics. Then, based on survey findings, site planning can be conducted. Remember that when identifying appropriate controls, the goal of site planning is to protect people.

Though not specifically mentioned above, one of the primary driving questions behind physical security is “What and where are the most valuable areas?” Inside a building, for example, what’s a wiring closet? Inside every office, hotel, hospital, there are rooms—often one on each floor and usually a main room—where network cables are pulled and other network equipment, like switches and routers, are housed. Should people have access to these areas? Of course not, unless they’re specifically authorized to access these areas, because they’re considered sensitive or high-value areas. The point is this: high-value areas within a facility or campus need to be identified and appropriate, and cost-effective controls to protect these areas should be implemented. Some key terms surrounding security surveying and high-value areas have been provided in [Table 3-59](#) and [Table 3-60](#).

Threat Definition	This is where applicable threats are identified. These could be any type of threat that might impact the site.
Target Identification	What are the assets that might be targeted by threats (identified as a result of “threat definition”)?
Facility Characteristics	Identify each asset’s vulnerabilities.
Life	This is the number one priority of physical security—to protect life, to protect people.
Property	Closely aligned with the number one goal of protecting life is the goal of creating a safe physical work environment that anticipates ideal working conditions as well as conditions that might exist in the event of a disaster. This includes exterior and interior elements and should support the top priority of protecting people.
Operations	Items like an organizational Business Continuity Plan (BCP) and Disaster Recovery Plan (DRP) should be developed and utilized for this purpose.

Table 3-59: **Security Survey**

Wiring Closets	Typically found on each floor in a building, where wires and other important networking equipment might be located
-----------------------	--

Media Storage	The area where physical/digital media that store sensitive and valuable assets is located
Evidence Storage	The area where important evidence is stored
Server Rooms	The room(s) where the most important and valuable network assets and infrastructure are located
Restricted Work Areas	Any work locations that require additional security, due to the nature of the work being performed, the people performing the work, or any combination thereof

Table 3-60: High-Value Areas

3.9.2 Perimeter

CORE CONCEPTS

- Perimeter controls are an important element of physical security, including things such as landscaping, grading, fences, gates, and bollards.
- Like access points (doors) to a building, the less access points through a fence, the better the security of the perimeter.

A few specific controls that can be placed around the perimeter of a building for protection include things like landscaping. **Landscaping** is actually a very important control. Things like trees and bushes can be placed around a building in such a manner to prevent close access to certain parts of the building and to direct how people move around the exterior of the building. Large trees should not be directly adjacent to a building, because people could potentially use them as cover. The same applies to dense foliage, which could block CCTV sightlines. With the proper landscaping and other well-placed perimeter controls, like lighting, around a building, people with bad intentions will have a very difficult time achieving their goals.

Grading is another control and refers to the slope of the ground. Does it make sense to slope the ground away from or toward a multimillion-dollar data center? Away, so if a massive rainstorm hits and causes flooding, the data center remains bone dry.

Bollards are another great physical security control. They're pop-up barriers, about the size of a thick pipe, that prevent cars from driving on certain streets or into restricted areas at certain times; they're also stationary and are often found in front of office buildings and especially in front of US State and Federal government buildings as well as in front of military base entry points. Stationary bollards in front of these buildings are usually quite large, and they are positioned in a manner to prevent vehicles,

even large ones, from driving straight into the building or past a checkpoint and blowing up, which is what happened in the past at the FBI office in Oklahoma and resulted in significant loss of life. Many organizations that employ bollards will use large concrete flower planters around their buildings—this has the effect of looking attractive and defeating vehicle-based attacks.

3.9.3 Closed-circuit TV (CCTV)

CORE CONCEPTS

- **CCTV cameras serve primarily as a detective control.**
- **CCTV cameras also deter and monitor, and it can be used for security audits.**

What primary type of control are CCTV cameras?

Closed-circuit television cameras, better known as CCTV cameras, **serve primarily as a detective control**, though they also function as a deterrent and can be used for security audits. **A few things to consider related to camera systems:** First, **placement**. Cameras should cover major entrance and exit ways as well as other areas of importance, and they should be placed and positioned in a manner that allows a person's face and other important distinguishing features to be captured. Second, **image quality**. It's imperative that the quality of the captured images be considered, whether those images are captured during the day or at night and in any conditions. If the cameras are placed correctly, but the quality of the image is such that people are unrecognizable, then the system is worthless. Third, **transmission media**. Among other things, you must identify how will the images be transmitted back to the monitoring stations and recording systems, if there are people watching the camera feeds 24x7x365 or, if everything is going to be captured to tapes/hard drives, how long will the recordings be kept before being archived. Also note that local law may dictate some of these decisions, as in many areas of the world, strict privacy laws define for what purpose and how long images may be kept, who may view them, what areas the cameras cover, and whether they are recording public areas, and so on.

3.9.4 Passive Infrared Devices

CORE CONCEPTS

- **A passive infrared device is a motion sensor that detects motion by picking up on infrared light.**
- **Passive infrared devices are very sensitive to temperature changes.**
- **Passive infrared devices must continually recalibrate themselves based on ambient temperature.**

How a passive infrared device works

What's a passive infrared device? It's a **motion detector**, and the way it detects motion is by picking up on infrared light. It's essentially a low-resolution camera that's able to detect infrared light by taking a picture of a room on a continual basis and then comparing the pictures. Human beings are usually warmer than a room, because they give off heat; so, when someone enters a room, the passive infrared device is comparing one image to the next. Very quickly, the device will see that the image has changed and send an alert, if programmed to do so, or otherwise make note of the presence of the person in the room. Passive infrared devices are very sensitive to temperature changes, and even in a manner contrary to the way most people would consider, especially in places, for example, like Texas. Texas temperatures often get warmer than the temperature of the human body. Once it's too hot outside, many passive infrared devices will note the change of something cooler instead of something hotter. As a result of this fact, passive infrared devices must constantly recalibrate themselves based on ambient temperature so they can detect what's happening.

3.9.5 Lighting

CORE CONCEPTS

- External lighting serves as a deterrent and as a safety precaution.
- External lighting allows camera systems to have better visibility of the surrounding area.

External lighting is a very good security control, especially as a deterrent. If a building is well lit, it's very hard to sneak around it at night. It's also good from a safety perspective. A well-lit exterior allows people to see where they're walking, and statistics show that attacks are far less prevalent in a well-lit parking lot. So external monitoring/lighting helps keep people safe, and it also allows camera systems to work optimally and to detect things.

3.9.6 Doors and Mantraps

CORE CONCEPTS

- Door composition and frame construction impact security of doors significantly.
- Mantraps typically consist of a double set of doors or a turnstile and prevent tailgating/piggybacking.

What do mantraps prevent?

On the surface, the topic of doors does not typically receive much acclaim. However, in the realm of security and controls, there are a few things about doors that deserve attention. One of them concerns the composition of the door. In other words, how the door is made, as its composition makes a very big difference as to its security. The construction of the door makes a difference, and so does the construction of the frame. If the world's most secure steel door is hung on a wooden frame, what is an intruder going to do? They're going to bust the frame and kick in the door. So the door frame also matters. The final thing that needs to be considered is the location of the hinges. If the hinges are on the outside and accessible, the pins can be knocked out and the door easily removed. Many exterior doors on buildings have exterior hinges. Does this make the door more or less secure? It makes it less secure, but why are the hinges on the outside? For the safety of people, so when the door is pushed it swings outward. Unfortunately, there have been tragedies at places like night clubs when a fire breaks out or another emergency occurs, and everybody runs to escape and hit doors that open inward. Even though it's less secure, for the safety of people in places like nightclubs or movie theaters, doors should hinge outward.

A **mantrap**, on the other hand, is usually one of two things: a double set of doors, with space for one or two people in between, or, a turnstile, with enough room usually for one person. In either case, a mantrap is designed to prevent **tailgating—an unauthorized person following an authorized person into a building or other secure location**. In the case of a double set of doors, the authorized person enters the first door, and after the first door closes, the second door will open. Usually, two forms of authentication are involved: a badge to enter the first door and perhaps biometrics to open the second. If a tailgater somehow makes it through the first door, they'll immediately be noticed and will be trapped in the room. The other type of mantrap, a turnstile, is quite common in buildings around the world and only allows one person at a time. The person steps into a small chamber and the walls move around the person, giving them access to the inside.

3.9.7 Locks

CORE CONCEPTS

- **Locks are a delay control; they do not prevent access.**
- **Two types of locks: mechanical and electronic.**
- **Biometric locks are more prevalent, but employees may not desire to use them due to privacy concerns.**
- **The security of a combination lock is dependent on the complexity of the combination.**

What pieces of hardware are usually found on doors and door handles or around doors? Locks, like the traditional keyed lock, or, more likely in an office setting, some type of electronic lock—card reader, keypad, or biometric. **Regardless of the type, locks are a delay control.** Given enough time, just about any lock can be defeated; thus, they delay versus prevent.

Understand inherent weaknesses with some types of locks and precautions that should be taken

Depending upon the type of lock in use, a few precautions should be noted. Some locks are susceptible to shoulder-surfing, others to brute-force attacks, and many to multiple types of attack. Take a standard combination lock, for instance, where a dial is turned in alternating directions and stopped at a specific number at the end of each turn. Depending on the number of internal mechanisms, locks like this are often easily defeated. Furthermore, if the composition of the metal making up the lock is weak, it can easily be compromised. Keypad locks, where the digits are typically easy to see, can easily be shoulder-surfed. In cases like this, it makes sense to install some type of cover or shield to prevent others from seeing what code combination is being entered. And in cases where the keypad code is shared among people, the code should be changed, and access privileges reviewed frequently. Another means by which access to or within a building can be granted is using biometric systems. They're another way of gaining physical access.



Types of Locks

[Table 3-61](#) contains some of the most common lock types. Each type of lock, whether mechanical or electrical, can be extremely complex and multiple variations can exist.

Mechanical	 Key	A key operates a lock by being inserted into the keyway, for example, of a lock on a door, turned, and based upon the key design, moving tumblers inside the lock and thereby releasing the bolt and securing the door shut. There are many types of physical key locks.
	 Combination	A combination lock typically involves some type of rotating dial, which interacts with cams and the locking mechanism. The dial is usually populated with numbers, and the unlocking process commonly involves lining up the numbers in the proper sequence through a series of right/left turns, or directly on the cams.
	 Magnetic	Magnetic locks are often referred to simply as maglocks and are a combination of a powerful electromagnet and a steel or metal plate to which the electromagnet engages. Maglocks are typically operated through a button or via an attached card reader or motion sensor system—the type of operation often dependent upon where a maglock is utilized in a facility.
Electronic	 Proximity / RFID	Proximity/RFID cards are also referred to as key cards, and a defining feature is that they can be read simply by holding the card near the associated card reader. Proximity/RFID cards are often used as part of access control systems in buildings.
	 Combination	Combination, or keypad, locks are often used to secure entrances. They usually consist of a simple keypad containing numbers and perhaps a * and # symbol. The lock is typically tied to a central system for purposes of logging activity and managing users and associated passcodes/combinations are needed to unlock the entrance. Combination locks are susceptible to shoulder-surfing as well as to brute-force attacks.
	 Biometric	The use of biometrics as a form of identification has extended to include access control in buildings. Whether through fingerprint or palm scanning, or iris or retina scans, the use of biometrics has grown steadily over the years. They are very accurate and hard to defeat, but their primary drawback relates to concerns about privacy.

Table 3-61: **Lock Types**

What determines the security of a combination lock?

With one type of lock in particular—combination—security is dependent upon one thing: *the complexity of the combination.* If the combination consists of only a few numbers or characters, it can easily be determined. By requiring a longer combination, complexity is increased exponentially and the possibility of a brute-force attack succeeding drops significantly.

3.9.8 Card Access/Biometrics

CORE CONCEPTS

- **Card access control systems are inexpensive, relative to biometric access control systems.**
- **Card access control systems are more prone to abuse and are not foolproof.**
- **Biometric access control systems are very accurate, but employees may not desire to use them due to privacy concerns.**

How card access/badge systems can be used to ensure people's safety

In the realm of physical security, card access and biometric security control mechanisms are often used and quite helpful with regards to automating passage of people through security checkpoints, doors, elevators, and other physical barriers. Often enough, card access control systems require a card to be read upon entry to a building or space and again on exit. These movements are typically logged by the access control system and can provide a level of safety for employees by being able to determine who may still be in a building and potentially their location, in the event of an emergency.

With card access, whomever possesses an authorized card would be able to gain access to secured areas, so in and of themselves, they may not be the most secure form of access control. Cards can be lost, an employee may loan their card to a colleague, and so on. Coupled with biometrics, however, access can be much more strictly enforced, and biometric access control can be used very successfully by itself. Biometric checks of this type often involve facial recognition, a palm scan, or a retina or iris scan.

Relative to each other, biometric access control systems are significantly more expensive than card access systems. Additionally, as privacy is more and more a topic of concern around the world, employees may not be comfortable with the use of biometric access control systems. Thus, they are best used where security is most important and the cost benefit of doing so makes sense.

3.9.9 Windows

CORE CONCEPTS

- Windows often represent a major point of vulnerability in a structure.
- Shock and glass break sensors can help detect window breaches.
- Sensors that detect sound and frequencies are useful in quiet environments.
- Shock sensors detect vibrations related to glass breaks and are useful in noisy environments.
- Glass break sensors are essentially microphones that listen for the sound of glass breaking.

What type of sensor is effective in a noisy environment?

Another important physical security control topic is windows. Most people love having lots of windows in their homes and at their place of work. They allow for wonderful views, lots of natural light, and fresh air. But windows represent a major weak point.

To mitigate that weakness, shock or glass break sensors can be used. **Shock sensors** must be installed on each pane of glass and are designed to detect the vibrations related to glass breaking. **Glass break sensors** are different, they are essentially microphones that are tuned to listen for the sound of glass breaking. The advantage of glass break sensors is typically fewer sensors are required as one glass break sensor could be listening for the sound of multiple windows breaking in a room.

Shock sensors are particularly useful and effective in noisy environments, as they can effectively detect glass breaking in loud occupied rooms (e.g., if there is a loud party).

3.9.10 Walls

CORE CONCEPTS

- The composition and height of walls is important for the sake of strong physical security.

Walls are another important aspect of physical security. Like doors, the **composition** of walls is the primary concern. A standard wood or steel frame wall is fairly easy to breach. Along with the composition, the **height** of a wall also contributes to its security. Does it extend from the floor to the ceiling? If so, is the ceiling a drop ceiling, where—in most cases—the wall ends at the drop ceiling? To

ensure a secure room, the wall should extend from the floor to the true ceiling, not to a drop ceiling or anything similar. Similarly, in data centers, raised floors are often found, because this supports the flow of cool air from the HVAC system and routing of electric cables; in this case, walls should extend from the actual floor to the true ceiling.

3.9.11 Automated Teller Machine (ATM) Skimming

CORE CONCEPTS

- **Skimming utilizes disguised technology to steal debit and credit card information from people.**
- **Automated teller machines (ATM) and gas station pump card readers are frequently targeted for placement of a skimmer.**
- **Antitampering technology is the best way to prevent skimmers from being placed on card readers, like those used on gas pumps and ATMs.**

Skimming is the act of placing cleverly disguised technology at a point of sale—gas station pumps and Automated Teller Machines (ATM) are two frequent targets—and capturing debit and credit card information from unsuspecting customers. Gas station pumps, especially newer ones, include card reader technology that allows people to pay for gas at the pump, instead of going inside a building or otherwise needing to interact with an attendant. Similarly, ATMs can often be found outside of banks, inside convenience and grocery stores, and other places where people often need cash. ATMs bring the convenience of banking to each of these locations and allow people to withdraw cash, check balances, and even make deposits to checking and savings accounts without stepping foot inside a bank or interacting with a teller. As a result of the convenience and widespread use afforded by these card reader technologies, they're often the target of hackers and criminals.

Ways to detect or prevent skimming

Oftentimes, the skimming technology looks and functions like the real technology. The stolen information can be stored on the skimmer device, and more frequently it is transmitted to criminals via real-time methods that utilize mobile data or wireless connections. Even though skimmers look like the real thing, they can be detected by a few methods: 1. Pull on the keypad or card insert/reader mechanism to see if it detaches easily. Card readers in any location—gas pump, ATM, retail check-out, and so on—should be tamperproof. **Ideally, readers include antitampering protection to prevent the placement of skimmers.**

2. Look at the spelling on the device itself. Many criminals do not have strong command of the English language, and misspellings are often indicative of something amiss.
3. If errors occur during the transaction process, this could be a sign that a skimmer is in use. Usually, a gas pump reader or ATM will indicate an unavailable system at first glance. If the

system appears to be available and then throws errors after a card has been entered and PIN code entered, a skimmer may have just captured the information.

Infrastructure Support Systems

The term “infrastructure support systems” refers to the three major services or utilities that are always a focus of physical security; they’re often referred to as power, ping, and pipe. Power is pretty obvious—electricity. Ping refers to the internet. Pipe refers to cooling water and how it is piped into a facility. Most buildings, especially data centers, have a significant cooling water need; without it, buildings would quickly overheat, leading to equipment failure and disgruntled people. Physical security is concerned with providing a good, clean, consistent supply of all three utilities.

3.9.12 Power

CORE CONCEPTS

- **Power disruption of any type can seriously impact the viability of an organization.**
- **Uninterruptible power supply (UPS) units, systems, and generators can provide clean, predictable, and consistent power in the event of an outage or other power degradation.**
- **Different types of power disruption/degradation: blackout, fault, brownout, sags and dips, surges.**

Within any organization, disruptions in electrical power can seriously impact the business. The goal is clean, steady power. Two of the major systems used to provide clean, consistent power are Uninterruptible Power Supply (UPS) systems and generators.

What’s a **UPS**? Essentially, it’s a giant battery—it could be one battery or multiple batteries linked together—and it is designed to provide power when regular electrical power goes out. However, most UPSs are designed to only provide power for a short period of time lasting in minutes. UPSs also provide power conditioning, which simply means the power that enters the UPS is cleaned up—there are no sags, dips, or other power issues with what is flowing to important devices.

Generators are typically large diesel engines hooked up to an alternator to produce electricity. Generators can run for long periods of time, depending upon the supply of fuel. When the power fails, a signal is sent to the generators to turn on, but they can take some time—maybe up to a minute—to become fully functional. UPSs provide short-term power in the interim, while the generator engine becomes fully operational to provide long-term-power.

Within the area of power, situations can arise where there is no power available, not enough power, or too much power. The various types of power degradations and disruptions are summarized in [Table 3-62](#).

	Short period of time (Milliseconds)	Long period of time (Seconds, minutes, hours, days)
No Power	Fault	Blackout
Not enough power (e.g., low voltage)	Sag / Dip	Brownout
Too much power (e.g., high voltage)	Spike	Surge

Table 3-62: **Power**

3.9.13 Heating Ventilation and Air Conditioning (HVAC)

CORE CONCEPTS

- An important element of physical security is temperature and humidity control, for the sake of people, equipment, and areas of a building that may require specific temperature and humidity settings.
- Air quality is an equally important consideration.
- Positive pressurization helps ensure that contaminants do not enter a building, room, or space.

As mentioned earlier, one aspect of physical security is the maintenance of correct temperatures within a building. People, equipment, and other areas of a building all require optimal temperatures in order to function most effectively and efficiently. Heating, ventilation, and air conditioning (HVAC) systems provide air that is the right temperature, the right humidity, and the right air quality. Variations in any or all these variables can significantly impact a work environment.

The American Society of Heating, Refrigeration, and Air-Conditioning Engineers (ASHRAE) Technical Committee 9.9 created widely accepted guidelines for optimal temperature and humidity set-points in data centers, and these guidelines are summarized in [Table 3-63](#).

	Low	High
Temperature	64.4°F / 18°C	80.6°F / 27°C

Humidity	40% Relative humidity	60% Relative humidity
----------	-----------------------	-----------------------

Table 3-63: ASHRAE Temperature and Humidity Guidelines

If the temperature is too hot or cold, people get cranky, and equipment may not function correctly. If the humidity is too low, static electricity can develop and potentially cause electrical shorts; if humidity is too high, condensation can develop, which may cause corrosion and shorts as well. Air quality, especially in data centers, is also important, as this is the air used to cool servers. If the air is filled with dust and debris, over time, servers and other equipment will become filled with the same things. This can lead to overheating and eventually failure. So, air should be filtered, and contaminants removed.

Description of “positive pressurization” and the benefit it provides

One term related to air quality is “positive pressurization.” **Positive pressurization** refers to air that is pumped into a server room or data center at a slightly above ambient pressure. It’s not much, but when higher than the ambient pressure and a door to the server room or data center opens, clean air will rush out of the room rather than potentially dirty air rushing in. Even if cracks exist in the walls or someone cracks a window, with positive pressurization all that will happen is clean air will flow out - preventing dirty air from infiltrating. In summary, the reason to use positive pressurization is to **keep the air in the data center cleaner—free of contaminants** that would collect inside equipment causing it to overheat and fail.

The above recommendations are summarized in Table 3-64.

Temperature	Humidity	Air Quality
Avoid too hot or too cold	Too low = static electricity Too high = condensation may develop and lead to corrosion or shorts	Clean air; contaminants should be filtered out Positive pressurization

Table 3-64: HVAC Recommendations

3.9.14 Fire

CORE CONCEPTS

- **Fire requires fuel, oxygen, and heat—if any of the three are removed, the fire goes out.**
- **Fire and the risk of fire should be prevented via a complete control of**

prevention, detection, and correction.

- **Fire detectors—flame, smoke, and heat**
- **Water-based fire suppression systems—wet pipe, dry pipe, pre-action, deluge**
- **Gas-based fire suppression systems—INERGEN, Argonite, FM-200, Aero-K**
- **Fire extinguishers**
- **CO₂ is noncorrosive, but must be used with caution if people are around.**

Understand types of fire detection systems and the most effective for early detection

Think of fire as a three-legged stool. If one of the legs is removed, the fire—like the stool—fails; it goes out. The three legs of the fire stool are: fuel, oxygen, and heat. All three are required for a fire to burn. Take any one of the three away, and the fire goes out. This is where it makes sense to think about fire as a risk. When mitigating any risk, preventive, detective, and corrective controls should be put in place. Ideally, where fire is concerned, the goal is prevention, and one of the best ways to prevent fire is not to have combustible materials in the vicinity of what is being protected.

If fire can't be prevented, it should be easily detected. And once fire is detected, it should be possible to immediately correct it, in other words, to quickly extinguish it. This is where fire detection systems as well as water-based and gas-based fire suppression systems come into play.



Even in the best-case scenarios, fires will still happen, and in those cases it's critical that they can be quickly detected. Three primary types of fire detection systems exist: flame, heat, and smoke, as also seen in [Table 3-65](#).

	Flame	Flame detectors are actually video cameras that detect infrared and UV light created by a flame . The camera is pointed at an area, and if infrared or UV light is detected, a fire is likely present. If a fire is far enough along that flames are present, it's probably pretty far along. So flame detectors are not a good early detection tool.
	Ionization	Ionization detectors have a small amount of radioactive material embedded in the sensor that ionizes particles that flow between two metal plates. If smoke particles enter the chamber that houses the plates and radioactive material, the ionization process will be disrupted, allowing fire to be detected. This type of detector responds more quickly to flaming/fast fires.
	Photoelectric / Optical	Photoelectric sensors are made up of a light source and a sensor, and the light source is pointed slightly off-angle relative to the sensor. If smoke particles pass through the sensing chamber, the off-angle light will be refracted more directly into the sensor. Like Ionization, this change in the sensing chamber is indicative

		of a fire. This type of detector responds more quickly to smoldering fires.
	Dual	Most sensors today are dual sensors—they incorporate both ionization and optical sensors as part of their functionality.
	VESDA	Of the types of smoke detection systems noted, the best is known as VESDA—it is the most expensive and the best at detecting a fire very early. VESDA, which stands for Very Early Smoke Detection Apparatus, is a smoke detector that can detect a fire at the smoldering, or what is known as incipient, stage. This type of device is usually found in data centers, clean rooms, art galleries, and other locations that house very high-value equipment or assets.
	Heat (rate of rise)	Heat detectors are literally just temperature sensors that are placed around a building or data center, and they look for rapid rises in temperature. If the temperature in a given area rises very quickly, there's likely a fire present. However, like flame detectors, if a fire has grown to the point where the ambient temperature is spiking, the fire is likely quite advanced. Heat detectors, therefore, are also very ineffective early detection tools.

Table 3-65: Fire Detection Systems

Best Way to Prevent or Limit Damage from a Fire

As already mentioned, the best way to prevent a fire is to minimize combustible materials in proximity to valuable assets. For example, in a data center, where server and other network equipment is routinely added and removed, the new pieces of equipment should be unpacked and unboxed on the loading dock, leaving all the combustible cardboard and other material there and only transporting the bare metal hardware into the room. However, as noted above, in the event a fire does occur, the best way to limit damage is through early detection, and the best way to do this is through implementation of a VESDA smoke detector that is capable of detecting fire at the incipient stage.

Fire Suppression

After a fire has been detected, the goal is to correct it, or put it out. Fire correction systems fall into two categories: water-based and gas-based. Water-based systems use water to extinguish the fire. They remove the heat from a fire, which causes it to cease burning.

Water-Based Fire Suppression Systems

Understand the pros and cons of water-based and gas-based suppression systems and which type is best for a data center and other places where expensive equipment and other items are housed

Four **primary types of water-based systems** exist: **wet pipe, dry pipe, pre-action, and deluge**, as described in [Table 3-66](#). Water-based fire suppression systems work by removing heat from the fire equation. Relative to gas-based systems, they are much less expensive, and they are much simpler to operate. However, a major disadvantage of water-based systems is that they use water, which could potentially destroy millions of dollars of equipment in the event of a fire. So water-based systems are not typically found in data centers or other areas where their use would cause significant damage. This is where gas-based systems come into play, though they are significantly more expensive than water-based systems.

Wet Pipe	<p>Wet pipe means the pipes in a sprinkler system are “wet”—filled with pressurized water at all times. In the event of activation, water will flow until the water source is shut off. This can result in significant excess water after a fire has been extinguished.</p> <p>Though inexpensive, wet systems include some significant disadvantages:</p> <ul style="list-style-type: none">■ risk of leaks■ when used in locations where freezing is a possibility, pipes can freeze and burst—water expands when frozen—and once thawed, flooding will occur
Dry Pipe	<p>Dry pipe systems appear very similar to wet-based systems, with the big difference being that dry pipe systems do not always have pressurized water in the pipes. They’re dry and typically filled with some type of pressured gas (e.g., air or nitrogen).</p> <p>Dry pipe systems are often combined with some type of pre-action system (defined below) that, when triggered, activates a valve that allows water to fill the pipes very quickly and extinguish the fire.</p>
Pre-action	<p>Pre-action means the fire suppression is armed when something happens. For example, in the context of a dry pipe system, a pre-action system—which incorporates a detection system—will only activate valves that release water when a fire is detected. Many pre-action systems are tied to the primary “standpipe”—the main water pipe—in a building. A standpipe is tied to piping on each floor, with a corresponding valve off the standpipe on each floor.</p> <p>With a pre-action system in place, if a fire is detected on the third floor, for example, the system will activate the valve connected to the piping on the third</p>

	<p>floor, releasing water only to that floor. Additionally, only the sprinkler heads that have been activated by heat on the that floor will shower water.</p> <p>This type of system offers great advantages:</p> <ul style="list-style-type: none"> ■ due to the detection system, concerns of water damage due to false activations can be eliminated ■ water is held back until detectors in the area are activated; thus, in the example above, other floors would not be showered with water
Deluge	<p>A deluge system involves massive amounts of water flowing at once. With it, all sprinkler heads are in the open position. Thus, if a fire is detected, when the pre-action system activates the water valves, water will immediately flood the pipes and flow out of every sprinkler head.</p> <p>A deluge system should only be used where immediate extinguishment of a fire is required, like in a fireworks or explosives factory, where a fire could cause a catastrophic explosion.</p>

Table 3-66: Water-Based Fire Suppression Systems

Gas-Based Fire Suppression Systems

Gas-based systems are more expensive to install and more expensive to maintain, among other things, but they possess one significant advantage: they don't use water. Instead, they use various types of gases, and these gases are specifically selected because they don't typically damage expensive equipment. Gas-based systems extinguish fires by one of two means: they remove the oxygen from the impacted location or interrupt the chemical process that the fire consists of. Note that care must be taken with gas-based systems because although they're great at putting out a fire they can also kill everybody in a room. Typically, prior to a gas-based system deploying, an alarm sounds, emergency lights blink, and people are provided with ample time to evacuate the area before the gas is released.

Gases most commonly used in gas-based systems are INERGEN, Argonite, FM200, and Aero-K, as shown in [Table 3-67](#). **One gas that used to be popular but is now illegal is Halon; it should never be used in a gas-based system and was outlawed because of its significant negative impact on the environment, specifically on the ozone layer that protects the Earth's atmosphere.**

INERGEN	<p>INERGEN is a gas-based system that became popular in the early 1990s in response to the banning and subsequent replacement of Halon-based systems. INERGEN works like CO₂ in that it reduces the oxygen concentration where a fire exists, thus helping extinguish it. At the same time, and unlike CO₂-based systems, the use of INERGEN still allows for a breathable atmosphere, which is critical for the sake of personnel who may still be in the area when an INERGEN system is discharged.</p>
----------------	---

Argonite	Argonite is a gas-based system consisting of a mixture of argon and nitrogen. Similar to INERGEN, its use does not endanger human life, and it too is safe for the environment and therefore a suitable replacement for Halon.
FM-200	Like INERGEN and Argonite, FM-200 is considered a “clean agent,” meaning it does not endanger human life when used, and it does not leave residue, which makes it safe for high-value computing equipment and other hardware typically found in a data center.
Aero-K	Aero-K is a fire suppression system that disperses an ultrafine, potassium-based aerosol that can quickly suppress a fire. Additionally, the aerosol mist remains suspended in the air for a period of time, which allows it to prevent reignition of the fire as well as to disperse naturally and not leave a residue on equipment. It too is considered environmentally friendly and a safe alternative to Halon.

Table 3-67: **Gas Fire Suppression Systems**

Fire Extinguishers

Because different types of fire exist, different types of fire extinguishers also exist, as shown in [Table 3-68](#). Fire extinguishers are denoted by “class,” and different class extinguishers put out different types of fire:

- Class A extinguishers are used for putting out fires fueled by common combustibles
- Class B extinguishers put out liquid-based fires, like those fueled by gasoline
- Class C extinguishers put out electrical fires
- Class D extinguishers put out fires fueled by combustible metals; they use dry powder as a suppression agent
- Class K extinguishers are usually found in commercial kitchens, and they use wet chemicals as suppression agents.

Class	Type of Fire	Suppression Agents
A	Common combustibles	Water, foam, dry chemicals
B	Liquid	Gas, CO ₂ , foam, dry chemicals
C	Electrical	Gas, CO ₂ , dry chemicals
D	Combustible metals	Dry powders
K	Commercial kitchens	Wet chemicals

Table 3-68: **Fire Extinguisher Types**

CO₂

Understand why CO₂ is an effective suppression agent

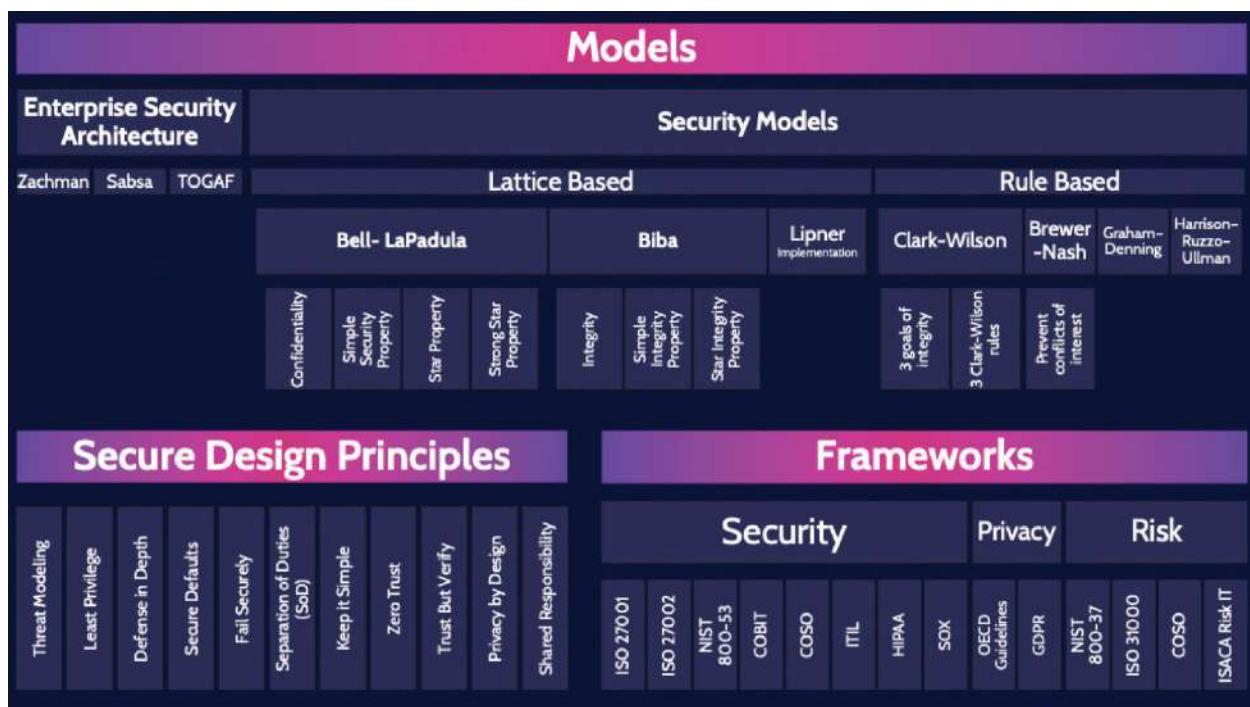
CO₂, better known as carbon dioxide, is something commonly found in nature and is also commonly used as a suppression agent. More importantly, it's not corrosive to expensive equipment. Some other suppression agents corrode equipment. *However, the one caveat with CO₂ is this: if too much is used, and people are present, those people might die, because it removes oxygen from the air.* If using a gas-based system, a better alternative would be one of the systems that utilizes the gases noted above, like Aero-K, FM-200, Argonite, or INERGEN.

3.10 Manage the Information System Lifecycle

The information system lifecycle is the entire lifespan of an information system, from its initial conceptualization to its eventual decommissioning. **It is essentially the same as the system life cycle, which we discuss in section 8.1.2.** However, you may notice a slight difference in the steps and their names. Integration testing is part of the development phase, while verification and validation are parts of the overall testing process.



MINDMAP REVIEW VIDEOS



Secure Design, Models and Frameworks

dcgo.ca/CISSPmm3-1

Evaluation Criteria							
Certification							
TCSEC (Orange Book)		ITSEC		Common Criteria			
Functional Levels		Assurance Levels		Assign EAL			
Confidentiality only	Single Box only	Confidentiality + Integrity Networked devices	Same Functional levels as TCSEC	ISO15408	Protection Profile	Target of Evaluation	Security Targets
D1 - failed or not tested	C1 - Weak protection mechanisms	E0	E1	E2	E3	E4	E5
C2 - Strict login procedures	B1 - Security labels	E5	E6				
B2 - Security labels and verification of no covert channels	B3 - Security labels verification of no covert channels, and must stay secure during start-up						
A1 - Verified design[0]							

Evaluation Criteria

dcgo.ca/CISSPmm3-2

Trusted Computing Base (TCB)

dcgo.ca/CISSPmm3-3

Vulnerabilities in Systems

Mobile Devices									
OWASP Mobile Top 10									
Single Point of Failure					Mobile Devices				
Mobile Controls					Mobile Devices				
Single Point of Failure	Bypass Controls	TOCTOU (Race Condition)	Emanations	Covert Channels	Aggregation & Inference	Policy training & procedures	Remote access security	End-point security	Mobile Devices
Redundancy	Mitigating Controls	Increase frequency of Re-authentication	Shielding (TEMPEST)	White Noise	Control Zones	Analysis & Design	Polyinstantiation	M1: Improper Platform Usage	OWASP Mobile Top 10
Redundancy	Mitigating Controls	Increase frequency of Re-authentication	Shielding (TEMPEST)	White Noise	Control Zones	Analysis & Design	Polyinstantiation	M2: Insecure Data Storage	OWASP Mobile Top 10
Redundancy	Mitigating Controls	Increase frequency of Re-authentication	Shielding (TEMPEST)	White Noise	Control Zones	Analysis & Design	Polyinstantiation	M3: Insecure Communication	OWASP Mobile Top 10
Redundancy	Mitigating Controls	Increase frequency of Re-authentication	Shielding (TEMPEST)	White Noise	Control Zones	Analysis & Design	Polyinstantiation	M4: Insecure Authentication	OWASP Mobile Top 10
Redundancy	Mitigating Controls	Increase frequency of Re-authentication	Shielding (TEMPEST)	White Noise	Control Zones	Analysis & Design	Polyinstantiation	M5: Insufficient Cryptography	OWASP Mobile Top 10
Redundancy	Mitigating Controls	Increase frequency of Re-authentication	Shielding (TEMPEST)	White Noise	Control Zones	Analysis & Design	Polyinstantiation	M6: Insecure Authorization	OWASP Mobile Top 10
Redundancy	Mitigating Controls	Increase frequency of Re-authentication	Shielding (TEMPEST)	White Noise	Control Zones	Analysis & Design	Polyinstantiation	M7: Client Code Quality	OWASP Mobile Top 10
Redundancy	Mitigating Controls	Increase frequency of Re-authentication	Shielding (TEMPEST)	White Noise	Control Zones	Analysis & Design	Polyinstantiation	M8: Code Tampering	OWASP Mobile Top 10
Redundancy	Mitigating Controls	Increase frequency of Re-authentication	Shielding (TEMPEST)	White Noise	Control Zones	Analysis & Design	Polyinstantiation	M9: Reverse Engineering	OWASP Mobile Top 10
Redundancy	Mitigating Controls	Increase frequency of Re-authentication	Shielding (TEMPEST)	White Noise	Control Zones	Analysis & Design	Polyinstantiation	M10: Extraneous Functionality	OWASP Mobile Top 10

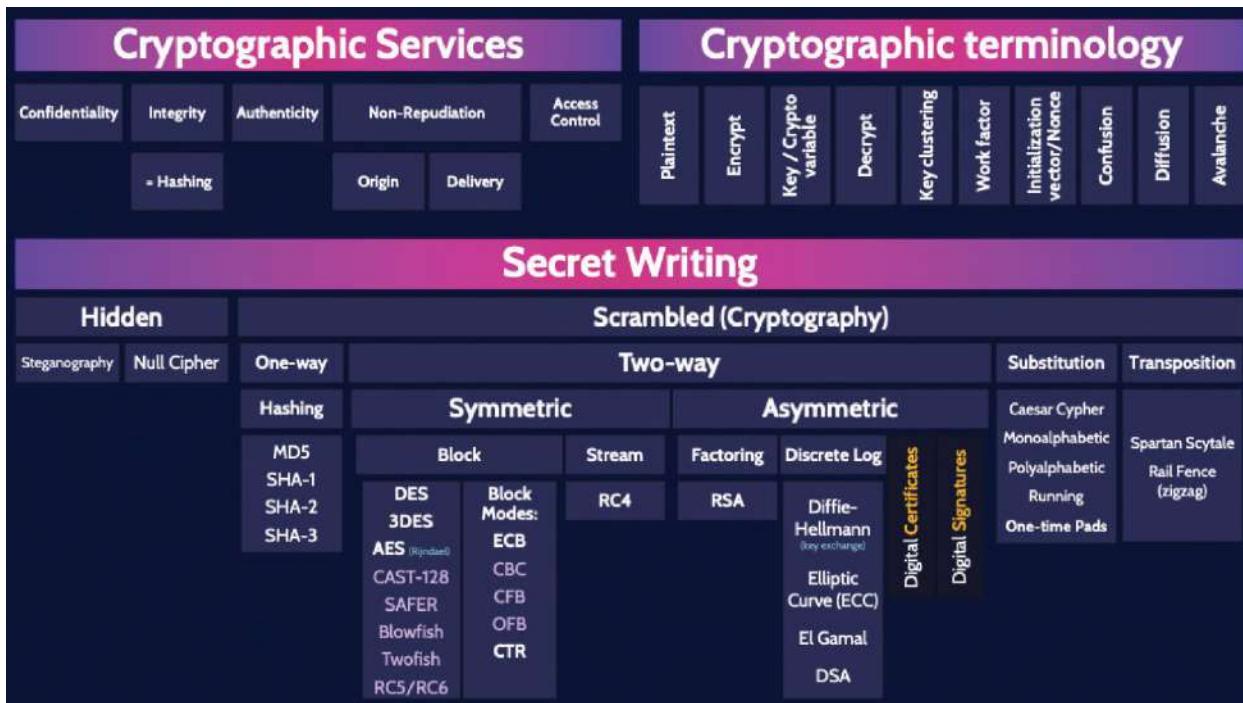
Vulnerabilities

dcgo.ca/CISSPmm3-4

Cloud Computing											
Characteristics	Service Models	Deployment Models	Virtualized Compute	Identity Provider	Cloud Identity	Roles	Protocols	Migration	Forensics	Data Destruction	
On-Demand Self Service	IaaS	PaaS	SaaS	Public	Private	Community	Hybrid	Virtual Machine	Containers	Serverless	Local
Broad Network Access								Hypervisor	Container Engine		Cloud
Resource Pooling											Cloud
Rapid Elasticity											Linked
Measured Service											Synced
											Federated
											Accountable
											Responsible
											SPML
											SAML
											OpenID
											OAuth
											Data Centric
											SLA
											Snapshot, Virtual Disk, Image
											Crypto Shredding / Crypto Erase

Cloud Computing

dcgo.ca/CISSPmm3-5



Cryptography

dcgo.ca/CISSPmm3-6



Digital Certificates and Signatures, PKI, and Key Management

dcgo.ca/CISSPmm3-7

Cryptanalysis

Cryptanalytic Attacks

Brute Force Ciphertext Only Known Plaintext Chosen Plaintext Chosen Ciphertext Linear & Differential Factoring

Cryptographic Attacks

Man-in-the-middle Replay Pass the Hash Temporary Files Implementation Side Channel Dictionary Attack Rainbow Tables Birthday Attack Social Engineering

Power Timing Radiation Emissions

Purchase Key Rubber Hose

Cryptanalysis

dcgo.ca/CISSPmm3-8

Physical Security

Safety of people

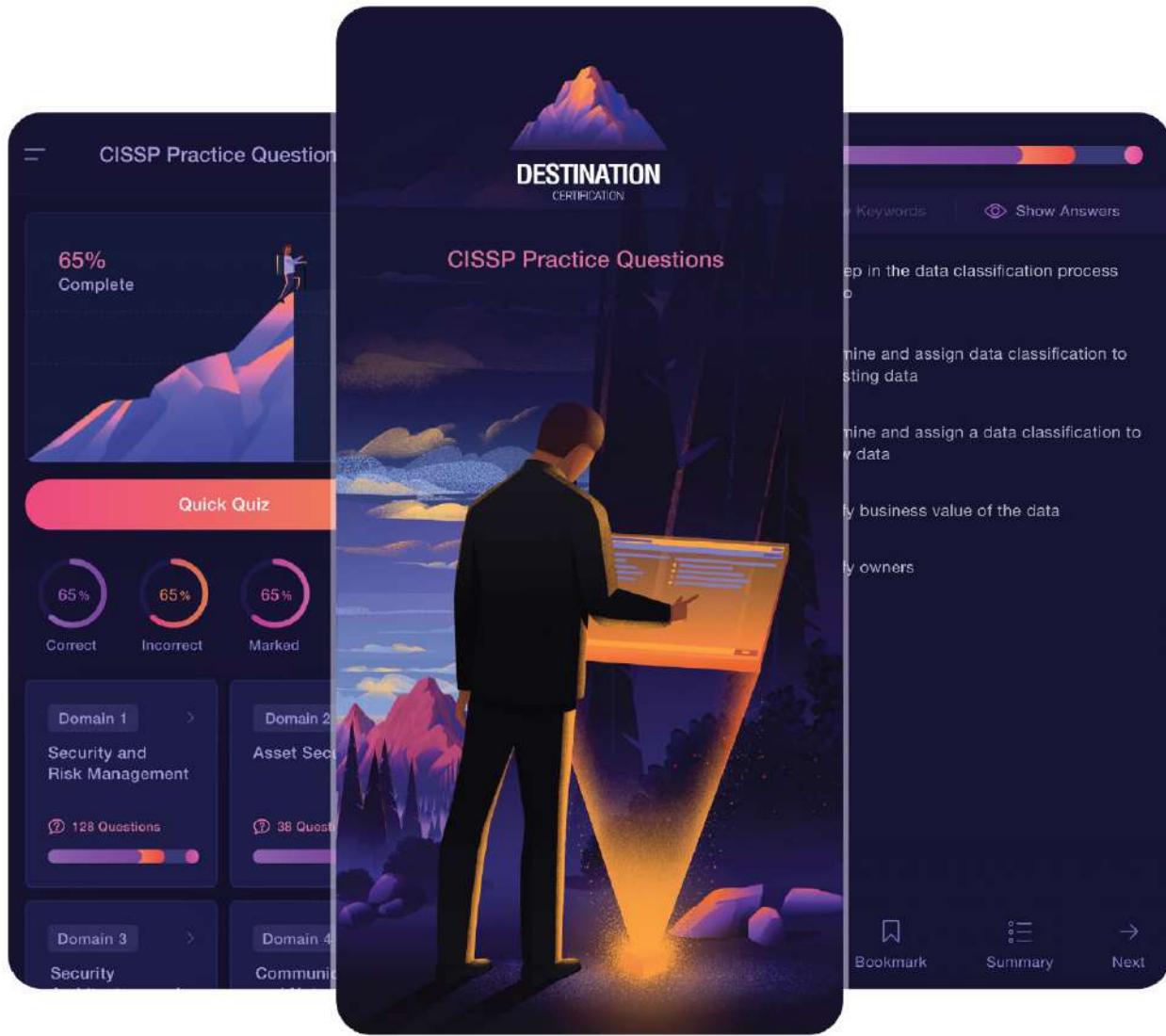
Categories of Controls		Layered Defense														
Deter	Delay	Perimeter			Infrastructure			Fire Detection			Fire Suppression					
		Cameras	Passive Infrared Devices	Lighting	Locks	Windows	Walls	Skimming	Network	Power	HVAC	Flame (Infrared)	Smoke	Heat (Thermal)	Water	Gas
Landscape	Grading			Mechanical	Digital	Shock	Glass break		UPS	Generator	Power Outages Power Degradation Temperature Humidity Air Quality	Ionization Photo-electric Dual		Wet Dry Pre-action Deluge	INERGEN Argonite FM-200 Aero-K	

Physical Security

dcgo.ca/CISSPmm3-9

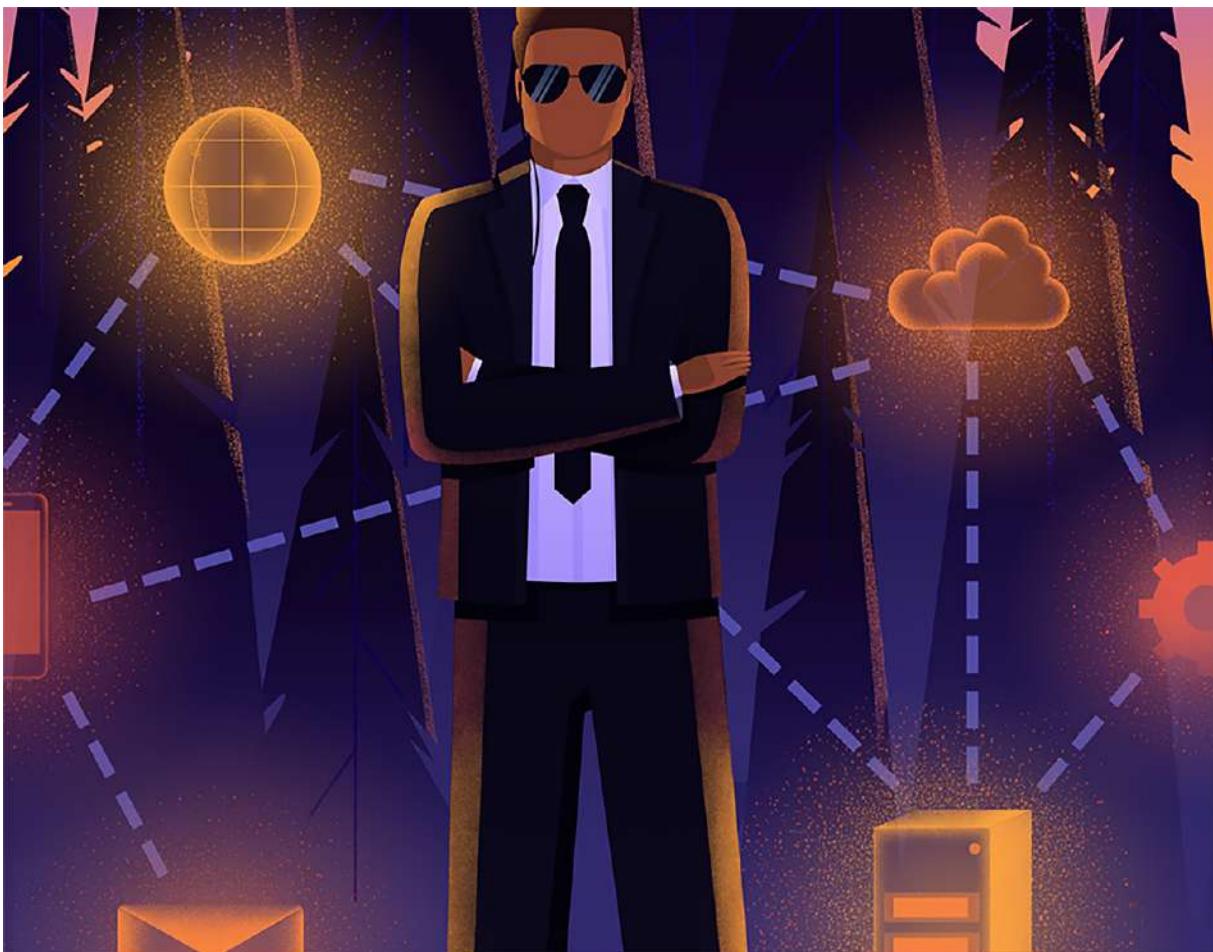


CISSP PRACTICE QUESTION APP



Download the Destination CISSP Practice Question app for Domain 3 practice questions

dcgo.ca/PracQues



Domain 4

Communication and Network Security

4.1 Implement secure design principles in network architectures

4.1.1 Open System Interconnection (OSI) Model

CORE CONCEPTS

- **Open System Interconnection (OSI) and TCP/IP models**
- **General understanding of what happens at each layer**
- **General understanding of common devices and protocols found at each layer**
- **Concepts of encapsulation and decapsulation**

Though the CISSP exam is not a networking or purely technical exam, it does contain technical elements that can lead to overthinking beyond the general competence required to answer a question. If this statement applies anywhere, it especially applies with regards to Domain 4, because it tends to be more technical than the other domains. However, the approach to questions should still be from a management perspective. First, the concept of network and where security fits must be explored. Every organization utilizes one or more networks. That allows it to meet its goals and objectives cost-effectively and efficiently. Networks allow organizations to identify and realize revenue opportunities and communicate and interact with clients. Networks are a really valuable asset of any company and therefore require protection.

What Is a Network?

A network is at least two devices that are connected to each other. Like people, in order to communicate, these devices must be able to speak a common language (which is what a protocol does), and common rules of communication must be followed.

What Is a Protocol?

The common rules of network communication are called protocols. A protocol is simply a standard set of rules that are understood, conformed to, and abided by so that two or more devices on a network can communicate. Protocols allow messages to be sent and received, interpreted, and acted upon, and all of this takes place in the context of what's known as the Open System Interconnection (OSI) model.

OSI (Open System Interconnection) Model

Many people know the OSI model as simply a seven-word mnemonic that corresponds to its seven layers as depicted in [Table 4-1](#). It's important to know the seven layers, what happens at each of them, and where security fits in. OSI stands for Open Systems Interconnection, which implies that the OSI model is about open systems that can interconnect and communicate with each other, using protocols. The OSI model is a structured, layered architecture comprising seven layers. Because it is a layered architecture, think of the seven layers of the OSI model as team members. Each member has

responsibilities that allow the ultimate goal of communication to be accomplished. No layer can work on its own and accomplish this ultimate goal.

Although a lot of people simply refer to every type of information as “packets,” that is actually incorrect. Information within the uppermost three OSI layers (Application, Presentation, and Session) is referred to as “data.” When that reaches the Transport layer it is referred to as “segments” or “datagrams.” At Layer 3 (Network), the term “packets” is used. Layer 2 (Data Link) uses the term “frames” while at Layer 1 (Physical) everything is just referred to as bits (or 0s and 1s).

7 Application	
6 Presentation	Data
5 Session	
4 Transport	Segments or Datagrams
3 Network	Packets
2 Data Link	Frames
1 Physical	Bits

Table 4-1: OSI Layers

Now, you may be wondering how you will be able to remember these seven layers. The two most commonly used mnemonics for OSI are: **A**ll **P**eople **S**eem **T**o **N**eed **D**ata **P**rocessing and

Please Do Not Throw Sausage Pizza Away. The first mnemonic refers to the OSI model from a top-down approach (Layer 7 to Layer 1). The second mnemonic refers to the OSI model from a bottom-up perspective (Layer 1 to Layer 7). Both are depicted in [Figure 4-1](#).

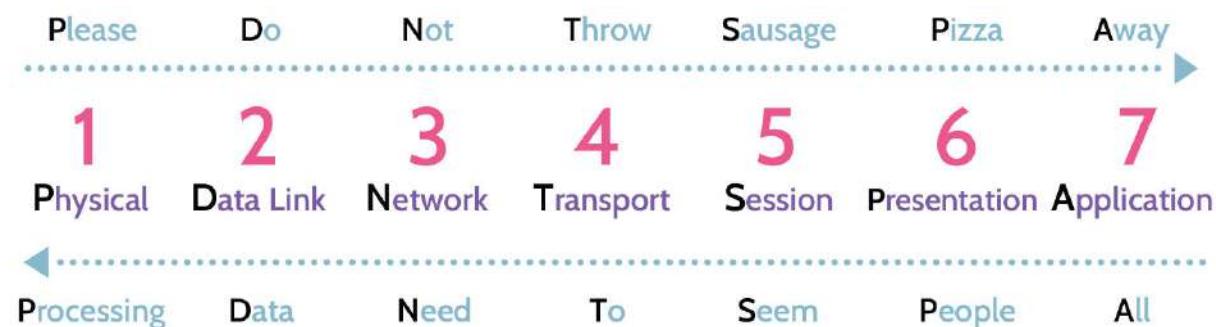


Figure 4-1: **OSI Mnemonics**

It's very important to know what security-specific features exist at different layers of the OSI model. The higher the layer, the more functional the security features become, and more comprehensive controls can be implemented; the lower the layer, the opposite is true. However, at the higher layers, the functionality is accompanied with complexity, which has an expense at speed and efficiency. Similarly, at the lower layers, where complexity is minimized, speed and efficiency are a given. At the lowest layer—the Physical layer—no intelligence exists. This is simply the layer that interconnects devices, and it is extremely fast.

Something else that should be kept in mind is that some types of devices—firewalls, for example—can be found at multiple

layers in the OSI model. Firewalls can exist at three of the seven layers—Network, Session, and Application—and functionality at each layer varies significantly. At the Network layer, packet filtering firewalls can be found. Decision-making here is very limited, but processing efficiency is high. On the upper end, at the Application layer, application or application proxy firewalls can be found. They offer significant decision-making capabilities, but at the cost of substantial overhead and processing efficiency.

Clearly, the technology industry long-ago recognized that making security decisions requires striking a balance between functionality and efficiency. This explains the importance of security protocols, and why so many exist and operate at different layers of the OSI model.

Let's relate this and take it a step further. The OSI model is exactly that—it's a model, and a model is simply a representation of something and the accompanying rules. The OSI model is a representation of communication rules and to work the OSI model must be implemented. The actual implementation is through TCP/IP. The internet protocol suite (TCP/IP) consists of many protocols—a family of protocols. TCP and IP and the other members of the protocol family run at different layers of the OSI model to support the underlying tasks of a given layer.

For example, the Network layer is primarily responsible for taking information and routing that, breaking it—

fragmentation—into manageable chunks called **packets** and providing addressing so those chunks can be communicated across a network using a logical addressing scheme called IP addresses.

The Transport layer is responsible for transporting information that is being exchanged between devices. Think of it as the truck that carries information between two people. To provide reliable transportation, a road should be built that facilitates ordered flow of traffic, including consistent reporting of road conditions. Within the family of protocols mentioned above, TCP is the protocol that provides ordered and reliable transport service.

Conversely, unordered, and unreliable transport services can also be provided by simply pointing the trucks toward the destination and not relying on a specific road. In this case, the trucks are simply loaded, and the drivers are told to go find the destination. The drivers will likely follow different routes, arrive in different order, and some may not arrive at all. All the trucks may travel quickly, but otherwise they'll be completely unreliable. Within the family of protocols, UDP is the protocol that operates in a similar manner—unordered and unreliable, but very quick, transport.

Let's consider this from the perspective of two people connected by devices across a network. If one person wants to send the other an email, an application like Outlook will be used to compose and send the message. Once ready to go, the

person will hit “send.” At this point, the Application layer will take the message, perform certain activities, and then pass it to the next layer, the Presentation layer. This process will continue all the way down to the Physical layer, which is where devices actually connect and bits (0s and 1s) are encoded as voltage and sent across the wire. Seemingly magically, those bits will travel the wire until they reach the other person’s device via the Physical layer. At this point, the Physical layer will perform certain functions and pass the information up to the next layer, the Data Link. This process will continue all the way up to the Application layer, at which point the other person can read the email.

The description above describes two processes known as **encapsulation** and **decapsulation**. As information moves down, from the Application layer to the Physical layer, encapsulation is taking place. Each layer will add its own header and trailer information to the existing data and then pass it down to the next layer. On the other side, when the fully encapsulated information arrives, a process called decapsulation takes place, where the header and trailer information is removed layer by layer all the way up to the Application layer.

What happens and what protocols are found at each layer?

Note that although the OSI model consists of seven layers, the TCP/IP implementation consists of four layers as shown in [Table 4-2](#). Like the OSI model, rules must still be followed, but just a bit differently. For instance, the top three layers of the OSI model are handled by the Application layer of the TCP/IP model. The Transport layer is the same in both models. The OSI Network layer is called the internet layer in TCP/IP and then the bottom two layers of the OSI model are handled by TCP/IP's link layer.

OSI	Description	Devices & Protocols	TCP/IP
7 Application	Identify capabilities of applications and resource availability	Application Firewall HTTP/S, DNS, SSH, SNMP, FTP	4 Application
6 Presentation	Formatting of data	XML, JPEG, ANSI	
5 Session	Interhost communication and session management	Circuit Proxy Firewall	
4 Transport	End-to-end connection with error correction and detection	TCP/UDP, iSCSI (SAN)	3 Transport
3 Network	Logical addressing, routing and delivery of packets	Routers, Packet Filtering Firewalls, IP addresses, ICMP, NAT	2 Internet
2 Data Link	Physical addressing, and reliable point-to-	Switches, bridges, MAC addresses, L2TP,	1 Link

	point connection	PPTP	
1 Physical	Binary transmission of data across physical media (wire, fiber, etc.)	Hubs, NICs, Network media	

Table 4-2: OSI vs. TCP/IP

4.1.2 Layer 1: Physical

CORE CONCEPTS

- Data at Physical layer exists as bits—0s and 1s
- Transmission media—wired, wireless, and so on
- Network topologies: Bus, Tree, Star, Mesh, Ring
- Collisions and collision avoidance; CSMA
- Transmission methods: unicast, multicast, broadcast
- Layer 1 devices: hubs, repeaters, concentrators, network interface cards (NICs)

Understand what happens at Layer 1 and how the primary wired transmission media types differ from one another

Layer 1, the Physical layer, focuses on how devices interconnect as well as encoding the bits, the 0s and 1s, that Layer 1 understands. Devices can connect using wired or wireless technologies. For now, let's focus on wired technologies, with the three most used media being **twisted pair**, **coaxial**, and **fiber optic**. The use of one versus another should focus first and foremost on security and then speed and cost, although

different organizations may have a varying approach on this.

Table 4-3 summarizes the major types of wired and wireless transmission media.

Wired	Wireless
<ul style="list-style-type: none">■ Twisted Pair■ Coaxial■ Fiber Optic	<ul style="list-style-type: none">■ Radio Frequency■ Infrared/Optical■ Microwave

Table 4-3: Transmission Media

Twisted pair cable refers to the fact that it is a pair of wires twisted together in a specific way that creates a magnetic field, which allows the signal traveling across the wire to remain within the magnetic field. Additionally, twisted pair cable can be shielded (STP) or unshielded (UTP), with shielded twisted pair offering additional protection from cross talk and interference.

Whether they realize it or not, most people are familiar with **coaxial cable**. This is the cable often used by cable companies to bring television, telephone, and high-speed internet access to homes. Coaxial cable consists of a single strand of copper wire sheathed in a protective coating, and a technology called multiplexing allows the wire to provide all the services mentioned. Multiplexing allows the information carried along the wire to be split into different frequencies, waves, and time slices at the same time, and it does so at incredible speeds.

Unlike twisted pair and coaxial cable, which use voltage for communication, **fiber optic** utilizes light pulses to represent 0s and 1s. Both speed and security are great advantages of using fiber optic. Among other things, twisted pair and coaxial cable are both subject to what's known as **cross talk**—interference—because copper, by design, conducts electricity. Thus, nearby electrical equipment, lightning, and deliberate signal scrambling can severely disrupt communications. Additionally, each type of cable can be tapped to intercept or eavesdrop the signal. These types of issues can impact the integrity of communications as well as confidentiality of information being passed along the wires. Though fiber is not immune to issues like tapping, it's certainly not as easy. Using security requirements as the primary decision driver, fiber is the best choice among the three. In addition to security, other criteria that should also be considered when choosing among transmission media options include bandwidth, distance, geographic location, interference levels, cost, and so on. For example, relative to twisted pair and coaxial cable, fiber optic cable offers significantly better signal quality over long distances.

Cabling is one part of the equation; another part relates to **topologies**, or how the cables are laid out. The most used network topologies are depicted in [Figure 4-2](#).

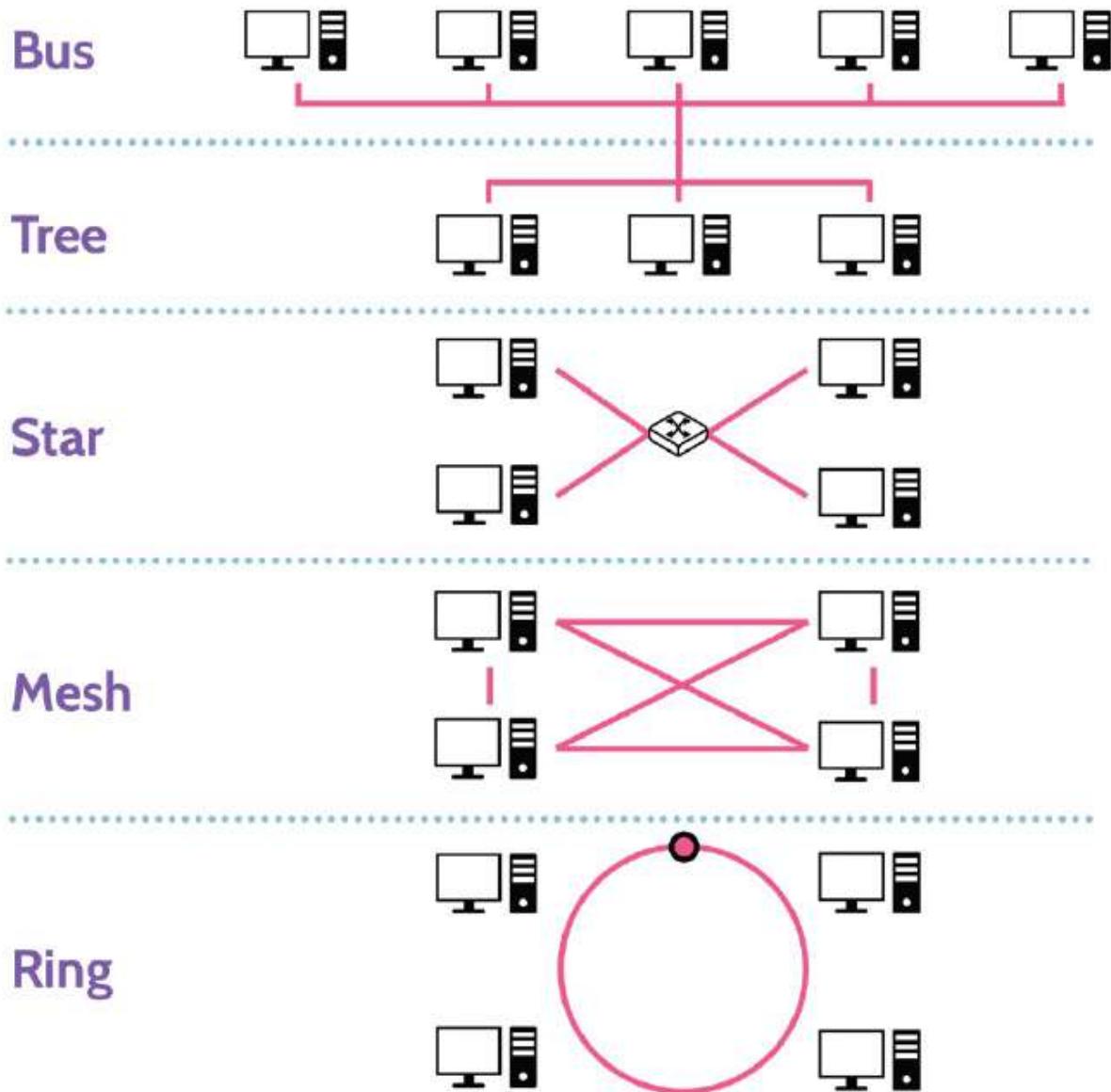


Figure 4-2: Network Topologies

A common network topology is known as a **bus topology**. Bus simply means all devices are connected to a central wire, called a bus. One great advantage of the bus topology is that the failure of one node does not affect the rest of the network; all other devices can still communicate with each other. So a bus topology allows for node failure. From a security point of view, a

bus topology has several weaknesses. For one, the bus represents a single point of failure. For another, all devices are connected to a single wire, so by default every device can intercept all the information being transmitted across the wire. These transmissions are known as broadcasts. As the network grows, adding more devices to the network is as simple as extending the bus, but doing so leads to an issue known as attenuation. Attenuation is a loss of signal strength over distance, but this can be mitigated using signal boosters and repeaters. Looking back at broadcasts for a moment, another disadvantage of a bus topology is that when two or more devices send information at the same time, something called a collision can occur.

Another topology is known as a **tree topology**, which somewhat resembles a tree with different branches. By virtue of this structure, one of the immediate benefits is that transmissions can be isolated to certain branches of the tree, thereby limiting transmissions from being seen by the entire network. In addition, if any of the tree leaves become damaged or unavailable, the issue doesn't affect the whole network, as that node can be easily repaired or removed.

One of the benefits of these topologies is that they can be mixed and matched when building out networks. By doing so, functional and security needs can be achieved. This is one reason why a bus topology is the most implemented topology, because it easily supports other topologies and can be combined with them.

A **star topology**, as the name suggests, resembles a star. All devices are connected to a central device, like a switch or a hub. One significant disadvantage of a star topology is that the central device represents a single point of failure. If that device goes down, communication between all devices is effectively halted. However, if switches are being used, functionality exists that allows network segmentation to be implemented by connecting certain devices to certain ports. Segmentation allows a network to be segmented based on value, which leads to the creation of a more sophisticated network, as elements of a bus topology are expanded to include elements of other topologies, like tree and star. Though more complexity is introduced, additional functionality and security are added too.

A **mesh topology** interconnects every device with every other device. This is excellent for purposes of redundancy—if one device goes down, communication with other devices is not impacted. Mesh topologies can be implemented as full (where every device is connected to every other device) or partial (where only the most critical devices are connected to each other for purposes of redundancy).

A **ring topology** looks like a ring. Devices are connected to a closed loop, and in a sense the loop is still essentially a bus, which can lead to issues like collisions. However, particular ring topology implementations (like a token ring) can include a mechanism called a token that is designed to prevent collisions. The token is passed around the ring from device to device, and when the token is in possession of a device, the

device can send information. This approach mitigates the issue of collisions, but it also introduces one particular problem, which led to the demise of token ring networks. Namely, if any node in the ring went down, it prevented the token from being passed to the next node. The direction of travel of the token could be reversed, but the problem of one node impacting the entire network remained. That's why specific implementations use a redundant ring, which can replace the primary ring in the event it becomes unavailable.

Topologies can be either physical or logical, with **physical topologies dictating how devices are physically linked** and how they communicate over these physical connections.

Logical topologies define how data actually flows within the network, regardless of the physical layout. The important thing to note is that even though a network may physically have a star topology, logically, it could actually function as a bus, with all communications being broadcast to all nodes.

The **data plane** is the doer. It's what does the work of transferring packets across the network based upon directions from the control plane. The **control plane** is the intelligence that tells the **data plane** what to do. Above this is the **management plane**, which manages and monitors the network's operations.

Cut-through involves a switch starting to forward a packet as soon as it reads the destination address, without waiting for the

entire packet to be received. It reduces latency and does not allow error checking for the entire packet.

Store-and-forward involves a switch waiting until it receives the entire packet. It then checks it for errors before forwarding it on to its destination. This process adds latency but ensures that the packet is error-free prior to forwarding.

Dealing with Collisions

As described, being able to mix and match different topologies can provide significant functionality and security benefits.

However, one of the major problems found in all the topologies except token ring is collisions. Three primary methods exist to handle collisions.

1. **Token-based collision avoidance:** A token is passed from device to device, and only the device holding the token can transmit information. This is what token ring networks use.
2. **Polling:** Interconnected devices poll each other to learn if any information needs to be transmitted. This method obviously implies a significant amount of network traffic, which explains why it is not popular or used often, if at all.
3. **Carrier Sense Multiple Access (CSMA):** Modern networks use what's known as Carrier-Sense Multiple Access. Devices are connected to the same carrier, the

same wire, and therefore each device can sense the wire to identify if another device is transmitting. To send information, the wire must be available, and devices can sense this availability based upon what travels across the wire—voltage. Voltage represents 1s and 0s. +5 volts = 1 and -5 volts = 0. To send information, no voltage should be on the wire. Once the wire is free, data can be transmitted, which should then be acknowledged by the receiving device. If the acknowledgment comes back in a reasonable amount of time, everything is okay. If not, information might need to be resent.

Even with CSMA, collisions are still going to occur. For example, in the case of a wireless network, there's no media to sense if there's someone already transmitting, which makes it rather difficult to use. In addition, CSMA tends to require larger transmission times due to its operation and as such can cause delays. This is why two flavors of CSMA exist—CSMA with **Collision Avoidance** (CSMA/**CA**) and CSMA with collision detection (CSMA/CD). CSMA/CA completely avoids collisions. CSMA/CA is used in the context of wireless networks and employs the use of two lanes of communication. One lane is used to receive information, and the other is used to send information. When a device communicates with a wireless access point, CSMA/CA is used.

Wired networks, like Ethernet networks, used to use the other flavor of CSMA. That's CSMA/CD - Collision Detection, which

detects collisions after information has been transmitted. CSMA-CD was popular in the past when Ethernet used twisted pair cables that needed hubs and repeaters. Modern Ethernet uses switches in full duplex mode, which avoids collisions, so we don't use CSMA/CD much anymore.

To understand how CSMA/CD works, think about it this way: when data is transmitted, voltage is running across the wire. If there's a spike in voltage after transmission, it means a collision has likely occurred. Thus, the information that was just transmitted would likely need to be sent again.

The CSMA/CD process for sending frames is as follows:

1. Before sending a frame, it detects whether the line is idle. If it is, it sends the frame. If it isn't, it waits until the line is ready.
2. When it sends frames, it monitors transmissions for collisions. If a collision is detected, it sends a jam signal instead of the frames. It then waits for a random period of time so that the receivers can detect the collision.
3. It then resumes transmission.

Transmission Methods

As a quick review, a network is two or more devices that are connected to each other. How these devices communicate is through transmission methods. One method is one-to-one,

which is called **unicast** and is used when a specific device needs to be reached. Another method is one-to-many, which is called **multicast** and is used when a group of devices need to be reached. The final method, **broadcast**, is one-to-all and is used when all the devices, e.g., on a specific subnet, need to be reached. From a security perspective, unicast is the most secure method, because communication is limited to a specific destination device.

In addition, [Table 4-4](#) contains a summary of the various transmission methods that can be used to communicate.

Unicast	Multicast	Broadcast	Anycast
One-to-One	One-to-Many	One-to-All	Nearest or Best

Table 4-4: Transmission Methods

Let's dig a little deeper into what "**anycast**" means. Anycast is an addressing and routing method where incoming requests can be sent to any, or a variety of locations, depending on what the objective is. A good example of where anycast transmissions would be ideal is a CDN. A content distribution network (CDN) is a tool for getting data as physically close to the user as possible. In this context, anycast transmission technologies help you connect to the server that is closest to you, as opposed to one on the other side of the planet. Anycast also allows you to connect to the server that offers the best performance or the best security.

Another important transmission method is **geocast**. It involves delivering messages to nodes in a specific geographical region.

Performance Metrics

[Table 4-5](#) runs through some key performance metrics that you should be aware of.

Bandwidth	The maximum amount of data that can be transmitted over a network or internet connection in a given amount of time.
Throughput	The actual rate of successful data transfer achieved, measured over a period of time . Note that bandwidth is the maximum amount, while throughput is the actual amount.
Signal-to-noise ratio	The level of the desired signal in comparison with the amount of background noise . A higher signal-to-noise ratio can allow for higher data transfer rates, because it means that there is less chance of lost packets and corrupt data.
Latency	The amount of time it takes for a signal to travel from its source to its destination and back . It's the time it takes for a round trip, measured in milliseconds.
Jitter	The variation in time delay between data packets over a network . It is a measure of the inconsistency of latency over time, measured in milliseconds. Ideally, we want consistent latency, so we want low jitter.

Table 4-5: Key Performance Metrics

Traffic Flows

It's important to understand two types of data center traffic flow. We explain both north-south and east-west traffic in [Table](#)

4-6 and Figure 4-3.

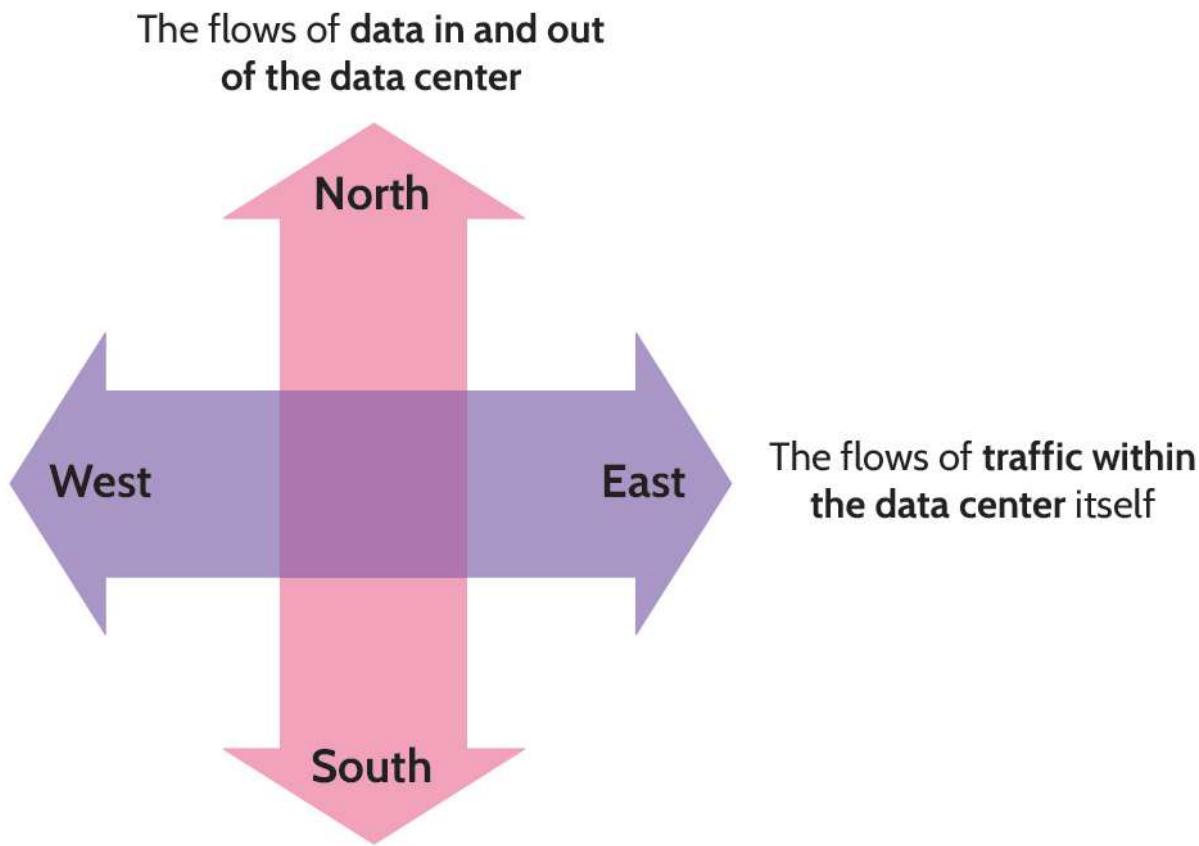


Figure 4-3: A Diagram of North-South and East-West Traffic Flows

North-south traffic	The flows of data in and out of the data center . Traffic sent to the data center's servers from clients on the Internet is considered southbound, while traffic from the data center's servers to the clients is considered northbound.
East-west traffic	East-west traffic is data moving between devices within a data center .

Table 4-6: North-south and East-west Traffic Flows

The differences between the two are critical, especially when making decisions about network architecture. The traffic pattern can impact the choice of topology, routing protocols, and security strategies.

Physical Segmentation

Physical segmentation involves creating a physically separate network or network segment. It can be useful for things like isolating network segments from one another. If we choose to physically segment a network, it can have impacts on how we manage the devices on it. [Table 4-7](#) runs through some of the different options.

In-band management	In-band management does not involve any physical segmentation. It involves managing network devices through the same network that is used to transmit user or application data. The lack of physical segmentation makes it a less secure option. Let's take the management of a switch as an example of in-band management. If you connect to and administer the switch via the same network that transmits user traffic, then this is considered in-band management.
Out-of-band management	Out-of-band management contrasts with in-band management because it involves a separate network. Network devices are managed from a dedicated network that is separate from the main network over which user traffic travels. Due to this separate network out-of-band management is considered more secure.
Air-gapped management	Air-gapped networks are an extreme form of segmentation that involves physically isolating a network from all others. This is the most secure approach, but having a completely isolated network also introduces a bunch of limitations. To manage an air-gapped network, someone will have to be physically onsite to

log in to it, because air-gapped networks are inaccessible from other networks. We tend to air-gap networks that contain important and sensitive systems, such as industrial control systems

Table 4-7: Management and Segmentation

Logical Segmentation

We can use logical segmentation in a number of different ways. This is often cheaper, easier, and more flexible than physical segmentation. However, if we don't use logical isolation properly, the network may not actually be truly isolated. We outline some logical networking concepts in [Table 4-8](#).

Virtual local area networks (VLANs)	Through VLANs, a single physical network can be logically partitioned into multiple smaller networks . We discuss VLANs in more depth in section 4.1.15
Virtual private networks (VPNs)	VPNs allow you to securely connect to a private network across public network infrastructure . We often use VPNs to connect an organization's different offices, and to connect remote users to the main network. We discuss VPNs in more depth in section 4.3.1
Virtual routing and forwarding (VRF)	VRF allows us to make many virtual networks with just a single network component.
Virtual domains	Virtual domains provide the ability to create multiple separate security domains within a single physical device . As an

example, they allow you to create multiple virtual firewall instances on a single device.

Table 4-8: Common Logical Networking Concepts

Monitoring and Management

Monitoring and management are critical for ensuring the performance, availability and reliability of systems, services and networks.

Network observability	This refers to the ability to gain insights and understanding into a network and its internal workings.
Traffic flow and shaping	Traffic shaping involves controlling packets and their movement within a network. We use it to enforce policies, optimize performance, and prioritize important traffic. As an example, corporate networks often prioritize VoIP traffic so that employees can be heard clearly during calls.
Capacity management	This involves monitoring current network resource usage and planning for the future. We want to ensure that we can meet demands both now and in the future. In the cloud, we have rapid elasticity, which can help reduce the complications associated with capacity monitoring.
Fault detection and handling	Fault detection and handling is the process of identifying, diagnosing and handling issues in an appropriate manner. We often use manual intervention, automatic remediation, incident response and other processes.

Table 4-9: Important Monitoring and Management Concepts

Layer 1 Devices

Several important devices operate at Layer 1, among them hubs, repeaters, and concentrators. Because they operate at Layer 1, they’re very fast, but they’re also not considered intelligent devices, as they don’t possess any decision-making abilities. [Table 4-10](#) contains a list of all of those.

Hubs	A hub is a Physical layer device with multiple ports and is used to connect multiple devices in a network. A hub is a very simplistic, passive device, which receives a data frame at one port and broadcasts that to all the other ports on the hub. Hubs are very “noisy” as a result, and data collisions often occur since they are a part of the same collision domain.
Repeaters	A repeater is a Physical layer device that is sometimes referred to as a signal booster and is often used to mitigate the issue of signal attenuation when data is being transmitted over a long distance. A repeater regenerates the signal and then sends it along to the destination.
Concentrators	Concentrators are like hubs in that both devices deal with multiple signals. However, while hubs repeat signals to all connected devices, concentrators <i>combine</i> all signals for transmission down a single line—they concentrate signals together.

Table 4-10: **Layer 1 Network Devices**

4.1.3 Layer 2: Data Link

CORE CONCEPTS

- Data at the Data Link layer exists as frames.

- Physical addressing via MAC addresses uniquely identifies devices on a network.
- Two types of networks: circuit-switched and packet-switched
- Common location to implement link encryption
- Layer 2 devices: bridges and switches
- Layer 2 protocols: L2TP, PPTP, ARP

Looking at the OSI model, Layer 2—the Data Link layer—acts as a conduit between Layer 1—the Physical layer—and Layer 3—the Network layer. The Physical layer only works with bits, and the Network layer works with packets. Between them, the Data Link layer takes packets from the Network layer and formats them in a manner that allows the Physical layer to work with them as bits. Likewise, the Data Link layer takes bits from the Physical layer and formats them in a manner that allows the Network layer to work with them as packets.

Physical Addressing

Layer 2 is also the layer where devices that operate across a network are physically and uniquely identified and separated from each other. This makes sense, because for a network to work, the devices on it need to have unique physical addresses. This unique physical address exists and is known as a **Media Access Control (MAC)** address. A MAC address is simply bits—0s and 1s—that uniquely identify and distinguish every device on a network, and this unique identifier is specified via a device's network card. To ensure unique MAC addresses, an

industry standard was adopted that specifies that each address should comprise 48 bits of information, with the first set of bits being specific to the vendor (first 24 bits or 3 bytes, which constitute the Organizational Unique Identifier, OUI), and the remaining bits (last 24 bits or 3 bytes) being specific to the device and assigned by its manufacturer, as shown in [Figure 4-4](#). Note that, based on this information, duplicate MAC addresses should never exist under normal circumstances.

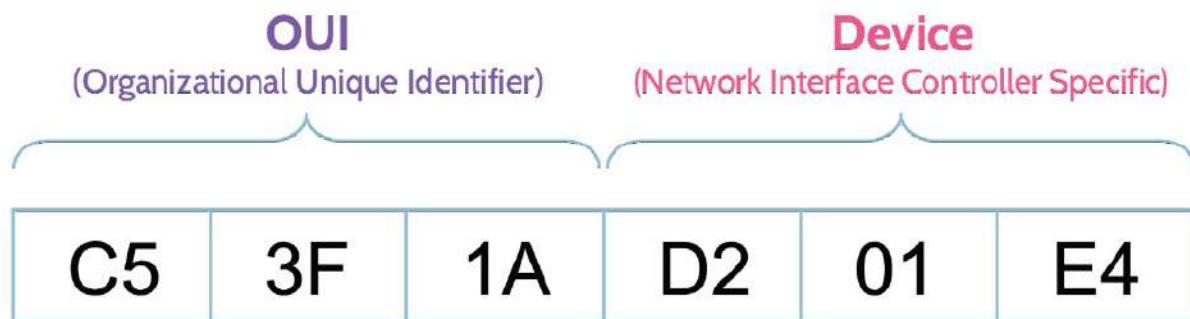


Figure 4-4: MAC Address

Networks work by virtue of logical address schemes that facilitate communication between devices. This is done at Layer 3 and is known as IP addressing, which can take IP addresses and convert them to MAC addresses and vice versa. Specifically, two protocols handle these needs: **Address Resolution Protocol (ARP)** and **Reverse Address Resolution Protocol (RARP)**. ARP allows IP addresses to be mapped to physical addresses, while RARP allows physical addresses to be mapped to IP addresses.

ARP poisoning is a topic that will be examined in more detail later, but it's good to introduce now. ARP poisoning is a spoofing or masquerading type of attack, where one or more devices on a network pretend to be other, legitimate devices. Then information intended for a legitimate device will be sent to the device masquerading as the legitimate one. ARP poisoning can be accomplished quite simply, because every device contains an ARP table, which facilitates request resolutions. The ARP table can be modified to point to a rogue device, thus spoofing and masquerading the legitimate device. DNS, routers, and switches all use tables as well and could therefore be subject to unauthorized modification too.

In networking, there are two types of network communications that mainly exist: circuit-switched and packet-switched networks.

Circuit-Switched Network

A great example of a circuit-switched network is the **Public Switched Telephone Network (PSTN)**, which has been in existence for many, many years. Connecting across the PSTN requires another person's telephone number, which can then be dialed, and a series of devices that comprise the PSTN will establish the circuit—the connection, as shown in [Figure 4-5](#). Thus, the name circuit-switched network. Parties on each end of the circuit can speak and hear at the same time, which illustrates the fact that full-duplex transmission is in place. With a circuit-switched network, a *connection is established*

permanently or on demand and is maintained between switches in order to route traffic to the correct destination.

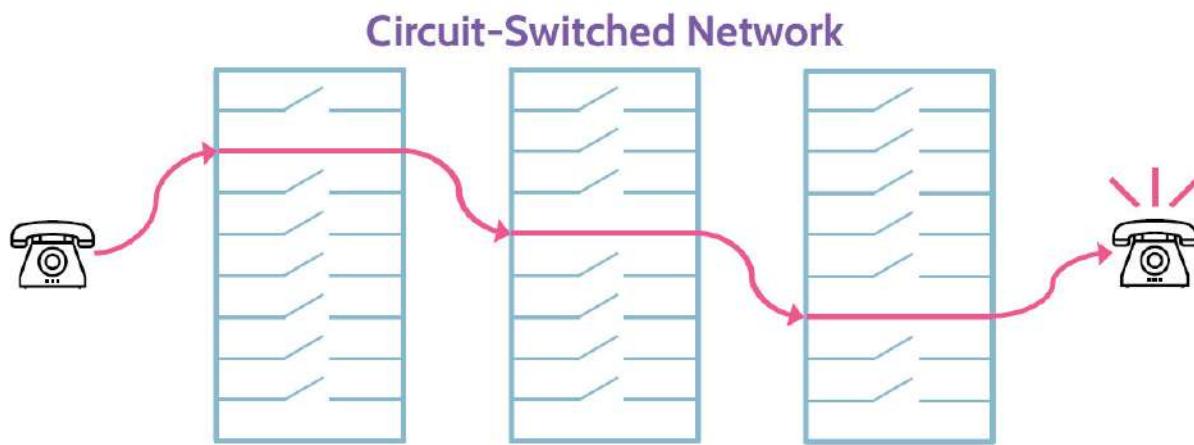


Figure 4-5: **Circuit-Switched Network**

Transmission of Digital Data over Analog Connections

When telephone communication was first being envisioned, scientists and engineers determined that because the human voice is analog, analog frequencies should be used for telephone communication. However, as technology advanced and data demands grew, the need for better communication technology also increased. Data doesn't travel well over analog frequencies, though a temporary fix came in the form of modems. A modem, which stands for modulator/demodulator, takes data and converts it to analog, so that it can be transmitted across the analog telephone network. Once it arrives at its destination, the analog signal is converted back to data by the receiving modem as depicted in [Figure 4-6](#). As was already noted, data does not travel well across analog networks, and part of the issue lies in the fact that regardless of the

surrounding technology, communication speed over analog networks is limited to 65,000 bits per second. This explains why separate data networks were built, including the global network known as the internet. Data networks were built for speed and bandwidth, and it wasn't long before voice communications using data transmission was perfected. We know this as Voice over IP, or VoIP or IP telephony, which *encapsulates the internet protocol to enable transmission of digital data over analog connections.* Though the ability to do telephony across data networks offers several advantages, it also presents a number of security risks.

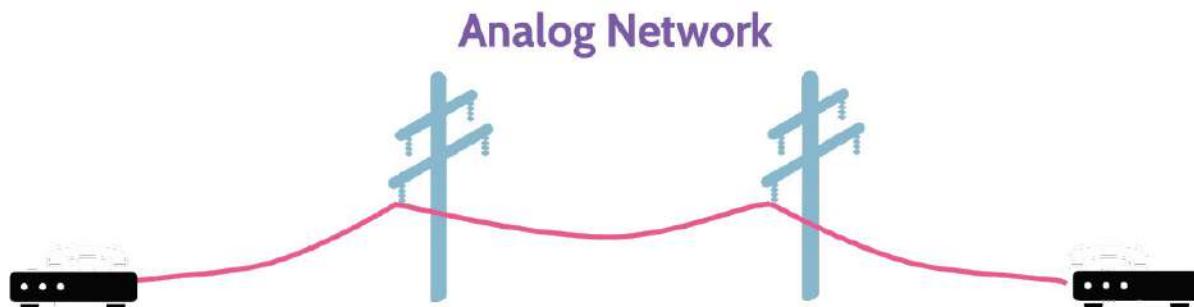


Figure 4-6: Modem Operation

Packet-Switched Network

Packet-switched networks function by taking data that needs to be communicated from one device to another and breaking it into packets. Each data packet contains information, such as addresses and sequence numbers. As [Figure 4-7](#) illustrates, the packets may travel along different routes to the final destination, and they might even arrive in a different order than how they were sent. Switches switch the packets to the

final destination, based on the header information and network conditions. Some datagrams might not even arrive, as there is no guarantee of delivery with data networks.

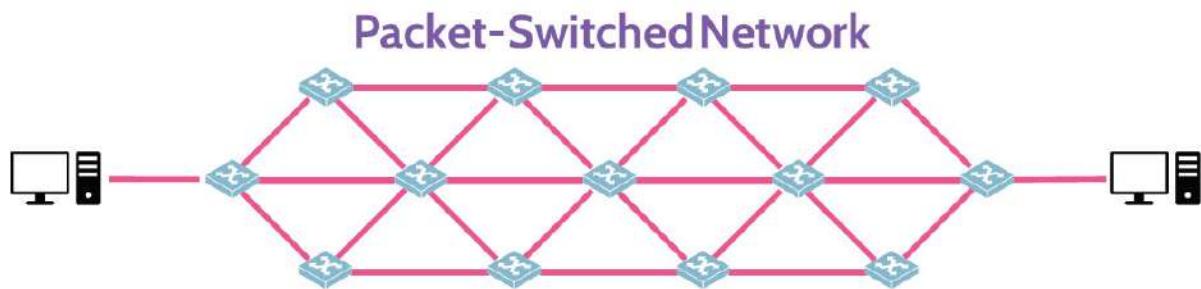


Figure 4-7: **Packet-Switched Network**

Layer 2 Protocols

Table 4-11 outlines Layer 2 protocols that are of the TCP/IP family of protocols. The first three are tunneling protocols, which are required to create virtual private networks (VPN), which are encrypted tunnels.

L2F	Layer 2 Forwarding tunneling protocol
PPTP	Point-to-Point Tunneling Protocol. Uses three distinctive authentication protocols: <ol style="list-style-type: none">1. Password Authentication Protocol (PAP): Simplest but least secure of the three. Uses a static plaintext password for authentication.2. Challenge Handshake Authentication Protocol (CHAP): More secure than PAP, as password is encrypted before being sent over the wire.

	3. Extensible Authentication Protocol (EAP): Considered the most robust of the three due to its increased level of flexibility, allowing it to be combined with other protocols.
L2TP	Layer 2 Tunneling Protocol
SLIP	Serial Line Internet Protocol—an older protocol used for remote access via serial ports and modem connections
ARP IP to MAC	Address Resolution Protocol—used to map IP addresses to MAC addresses
RARP MAC to IP	Reverse ARP—used to map MAC addresses to IP address

Table 4-11: Layer 2 Protocols

Layer 2 Devices

Significant devices that operate at Layer 2 are bridges and switches, as also depicted in [Table 4-12](#). Like Layer 1 devices, Layer 2 devices are efficient and fast, with just a bit more functionality, like for example the ability to form VPNs. Switches have a bit more functionality, because they have ports that can be used to create network segments.

Bridges	Connect different networks together, with no concern for what traffic is going across the bridge
Switches (Layer 2 switches)	Connect multiple network devices together. A frame sent to the switch is forwarded only to the intended recipient, based on the destination MAC address in the frame header. There are also switches that can operate and do certain functions at other layers, for example, there are also Layer 3 switches. One important note to remember is that exam questions will be very

specific as to whether they are talking about regular switches that operate at L2, or L3 switches.

Table 4-12: Layer 2 Network Devices

4.1.4 Authentication Protocols

CORE CONCEPTS

- As remote authentication needs have matured, authentication protocols have also matured to meet those needs.
- Extensible Authentication Protocol (EAP) is the best authentication protocol available, and its capabilities have been extended as Protected Extensible Authentication Protocol (PEAP).

Understand the basic difference between PPP, PAP, CHAP, and EAP

To understand the topic of authentication protocols, it's important to understand a bit about how the need came to be. Not too long ago, the only way for an organization to host remote access was with the use of modems. Larger organizations would typically have modem banks—a bunch of modems in a room, connected to the network. Employees would dial a phone number, and if a modem was available, a remote connection to the network could be made. Of course, this was before the internet grew to what it is today, and VPN solutions became the rule rather than the exception. So, the PSTN was used and a protocol that allowed TCP/IP to run across dial-up networks needed to be created. This led to **Serial**

Line Internet Protocol (SLIP) being developed. Though SLIP worked, it didn't work well, and as remote access grew in popularity, a better protocol called **Point-to-Point Protocol (PPP)** replaced it.

PPP is a Layer 2 protocol that is used to allow remote access, typically via VPN solutions today. With the advent of PPP and knowing that remote communication most often supported people, the creators of PPP determined that including authentication protocols made sense. As a result, three authentication protocols were developed and supported as part of PPP—PAP, CHAP, and EAP—with each protocol achieving different levels of popularity and each offering different degrees of functionality and security. These are depicted in [Figure 4-8](#). Even though we are discussing the PPP protocol as the basis for remote access, these authentication protocols (PAP, CHAP, EAP) are actually operating at OSI Layer 5. This might be confusing to understand, but if we remember that in order for communications to happen as per the OSI model, **all** layers need to work together, and certain things, like authentication, need to be done at the higher layers.

Password Authentication Protocol (PAP) simply prompts for a user ID and password when establishing a connection. Sadly, passwords are transmitted in plaintext, and the user will never be prompted to change their password, as it will be static in nature.

As a result of the issues with PAP, **Challenge Handshake Authentication Protocol (CHAP)** came to the foreground. CHAP is an improved version of PAP. Passwords are encrypted during transmission, while challenges are sent in regular intervals behind the scenes to ensure that an intruder has not hijacked or otherwise compromised a session.

The most robust and flexible authentication protocol of the three is **Extensible Authentication Protocol (EAP)**. The word *extensible* means being able to be extended or designed to allow new capabilities and functionality to be added. EAP allows vendors to adapt the latest authentication technologies, like smart keys and digital certificates, to their products. In fact, due to its inherent strengths, EAP can also be embedded into other things, like wireless security, where it's used with WPA2 for purposes of connecting to wireless networks and authenticating users at the same time.

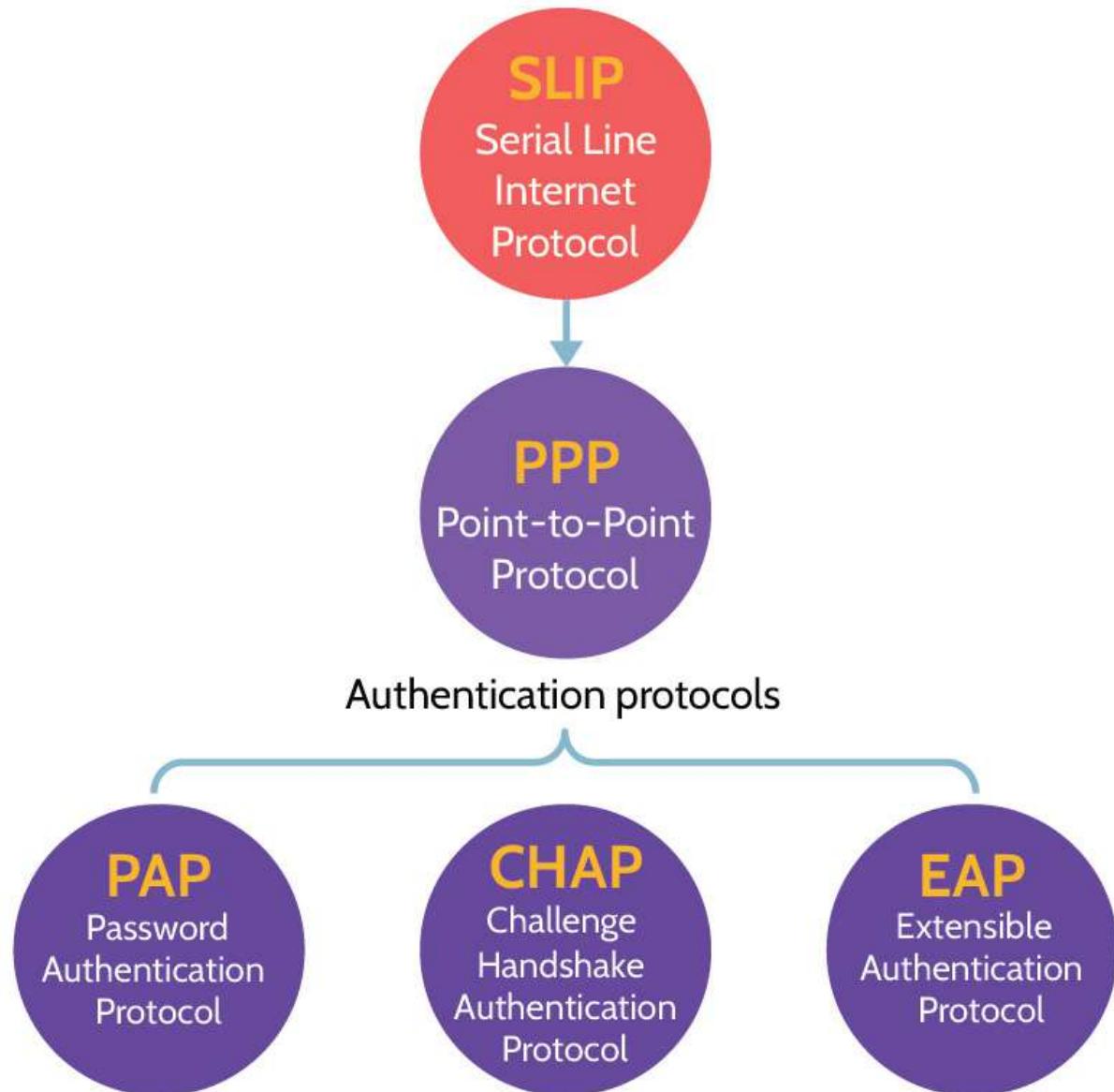


Figure 4-8: **SLIP and PPP Protocols**

Protected Extensible Authentication Protocol (PEAP)

What differentiates PEAP from EAP?

Note that PEAP is a more improved version of EAP, as it encapsulates EAP within an encrypted and authenticated **TLS tunnel** as seen in [Figure 4-9](#). A comparison between the most common types of EAP can be seen in [Table 4-13](#).

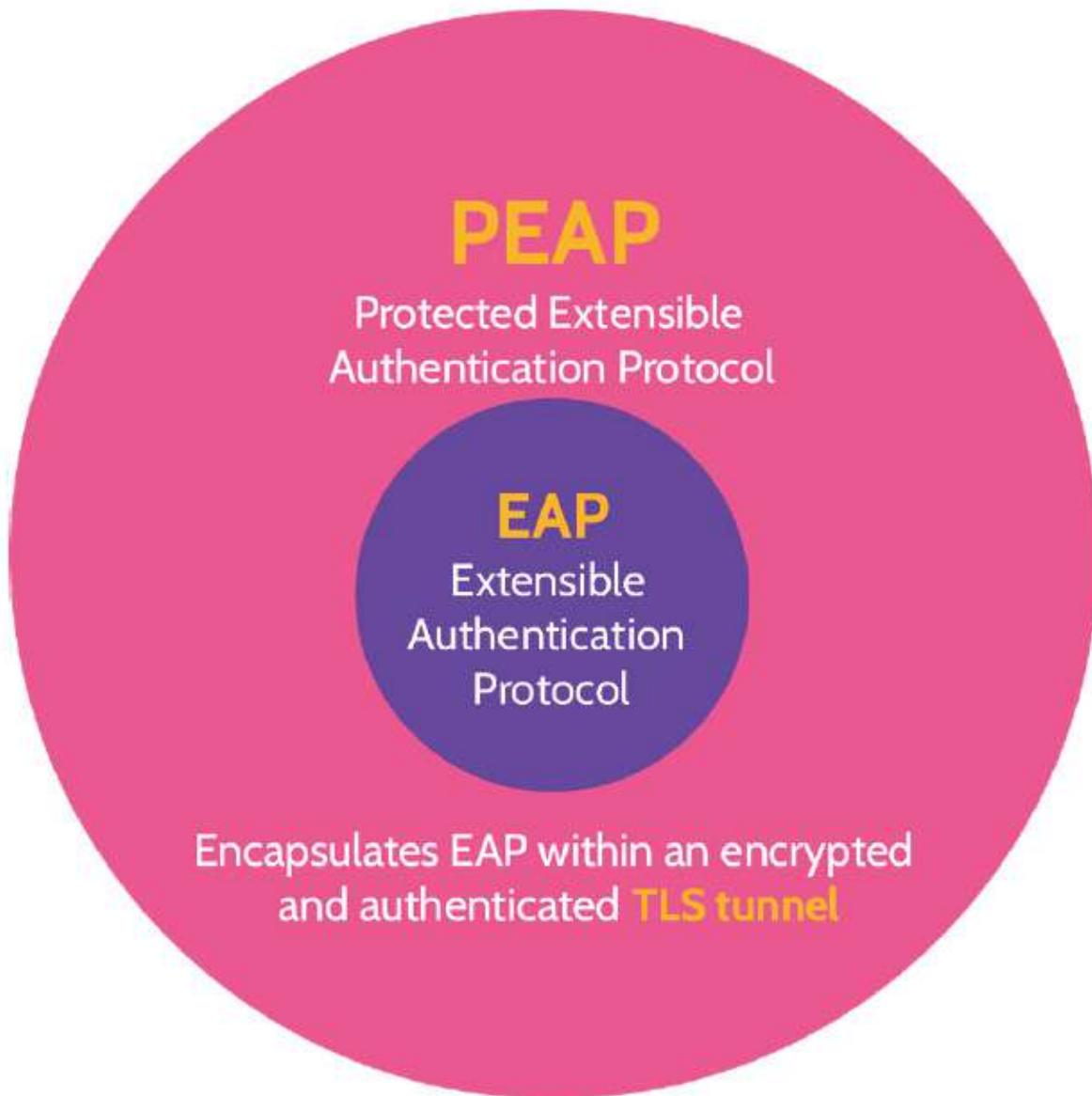


Figure 4-9: **PEAP**

Understand what type of authentication (client, server, or both) each type of EAP provides

Type	Client Authentication	Server Authentication	Security	Industry Support	Proprietary
EAP-TLS	Certificate	Certificate	High	High	No
EAP-TTLS	ID & Password	Certificate	Medium	Medium	Yes (Funk Software & Certicom)
EAP-PEAP	ID & Password	Certificate	Medium	High	Yes (Cisco, RSA, & Microsoft)
LEAP	ID & Password	ID & Password	Low	High	Yes (Cisco)
EAP-MD5	ID & Password	–	Low	Low	No

Table 4-13: EAP Types

4.1.5 Layer 3: Network

CORE CONCEPTS

- Data at the Network layer exists as packets.
- Logical addressing is used to map IP addresses to MAC addresses—ARP; Reverse ARP (RARP) maps MAC addresses to IP addresses.

- Route selection, including alternate routes to avoid congestion or node failure
- Layer 3 devices: packet filtering firewalls, routers, and Layer 3 switches
- Layer 3 protocols: ICMP, IGMP, IPsec, and routing protocols, including OSPF

The Network layer handles two very basic but important responsibilities—fragmentation and IP addressing.

Fragmentation provides the ability to take data and break it up into packets—chunks of data—and then allow those packets to be delivered to their destination via IP addressing. The IP protocol within TCP/IP specifically handles most of the responsibilities found at the Network layer. In this context, ARP and RARP are heavily relied upon. If data is leaving the network, the sending device's MAC address is mapped to an IP address; if data is entering the network, the IP address is mapped to the receiving device's MAC address.

Layer 3 Protocols

Layer 3 is home to several significant protocols that are part of the family of protocols within the suite of TCP/IP protocols.

Layer 3 protocols include IP, IGMP, and IPsec, as well as routing protocols like BGP, OSPF, and RIP. Some of these protocols are discussed in [Table 4-14](#).

ICMP	Internet Control Message Protocol is used for messaging and specifically provides feedback about problems in the network communication environment. Ping and traceroute are two important commands that utilize ICMP. Ping attempts to see if a
-------------	---

	host device is reachable; traceroute tries to map the path of traffic. Together or alone, both command tools can provide significant information about a network.
IGMP	Internet Group Management Protocol is used to establish and manage group memberships for hosts, routers, and similar devices.
IPsec	IPsec is a tunneling protocol that supports authentication of other Layer 3 devices as well as encryption. It's discussed in more detail in section 4.3.2 .
OSPF	Open Shortest Path First is a routing protocol used by routers to manage and direct network traffic properly and efficiently. OSPF includes security features that make it a more secure routing protocol than others.

Table 4-14: **Layer 3 Network Protocols**

Internet Group Management Protocol (IGMP) is used to establish and manage group memberships for hosts, routers, and similar devices. One important thing to note is that authentication in this context does not mean access control authentication. Rather, it refers to authentication of other devices that operate at Layer 3, like routers and Layer 3 switches.

Routers are the most important devices that operate at Layer 3. They route data based on information included in the header portion of the packet, and those routing decisions are based upon a routing protocol. A routing protocol is simply a language and set of rules that routers understand and use to make decisions about routing. **Border Gateway Protocol**

(BGP), Open Shortest Path First (OSPF), and Routing Internet Protocol (RIP) are common routing protocols.

What protocol is often used to quickly determine if network communication problems exist?

In addition to the Layer 3 protocols mentioned above, **Internet Control Message Protocol (ICMP)** is another very significant Layer 3 protocol. ICMP is used for the messaging aspects of networking and acts as a helper protocol by facilitating reporting issues over ICMP messages. For example, an ICMP “destination host unreachable” message denotes that the packet destination couldn’t be reached. However, note that ICMP can also be used maliciously. An attacker can use the ping utility to identify if a host is “alive” and reachable in any given network. Traceroute is a utility that maps the hosts between a source and destination and provides the traffic path taken. Both commands are often utilized as part of the reconnaissance phase of an attack, and this explains why it is standard practice for organizations to filter incoming ICMP packets at the firewall.

Layer 3 Devices

Several significant devices operate at the Network layer, Layer 3. The most obvious are routers, but Layer 3 switches can also be found at the Network layer. In relation to the exam, unless a

question specifically mentions a Layer 3 switch, assume that this refers to a Layer 2 device.

In addition to routers and **Layer 3 switches**, packet filtering firewalls also operate at Layer 3. Firewalls help protect networks from a number of different types of unwanted traffic, and they can operate at three of the seven layers of the OSI model. As was previously mentioned, the speed and efficiency of devices is inversely proportional to their intelligence as you go up the OSI model. Layer 3 firewalls are still fast, but they possess only limited decision-making capabilities. In fact, Layer 3 firewalls are only able to interpret the header portion of a packet. Simple concepts like source and destination IP addresses and ports (which equate to services) can be used to make decisions. If source IP addresses from certain countries are listed as malicious, the packet filtering firewall can be instructed to drop the packet and not allow it to traverse the network. Additional intelligence exists in firewalls operating at higher OSI layers (like the Application layer, where application proxy firewalls can filter traffic based upon web content, deep packet inspection, stateful inspection, antivirus, intrusion detection signatures, and so on). Of course, this increased intelligence and functionality comes with a price: slower processing speed and increased device cost. [Table 4-15](#) contains an indicative summary of Layer 3 network devices.

Packet Filtering Firewalls	Devices that make decisions based upon the header portion of a packet, such as the source and destination IP addresses. Due to low processing overhead, they are very fast.
-----------------------------------	---

Routers	Devices that connect and route network traffic between networks, based upon IP information included in received data packets. Routers utilize routing protocols and typically dynamically maintain routing tables to determine the optimal route for data packet forwarding.
Switches (Layer 3 switches)	Layer 3 switches function very similarly to routers and are often used to connect devices on the same virtual local area network (VLAN).

Table 4-15: Layer 3 Network Devices

Know what layer routers operate at and what layers routers operate between

4.1.6 Logical Addressing

CORE CONCEPTS

- Internet protocol packets consist of data, also known as the payload.
- IPv4 = 32 bits divided into four groups of 8 bits each versus IPv6 = 128 bits divided into eight groups of 16 bits
- Private versus public IP addresses; private IP addresses are not routable on the internet
- Network classes (subnetting) allows for the creation of networks of varying sizes.

IPv4 addresses are comprised of four numbers separated by dots, e.g., 192.168.1.254. The IP address is a 32-bit value, and each number represents an 8-bit octet. The valid range for each of

them is 0–255, with .0 used to denote a specific network, e.g., 192.168.1.0, and .255 to denote a network's broadcast address.

Each packet header contains routing details, like source and destination IP addresses. These fields can each hold 32 bits of information, which are divided into four groups of 8 bits (known as bytes). This is why IPv4 addresses consist of four decimal numbers. If every bit in a given octet is turned on, the highest number that can be achieved is 255. This fact points to a limitation with IPv4 addresses—the highest number in any octet can only be 255.

With a bit of math, it can quickly be seen that the number of valid, unique 32-bit IP addresses is limited. In fact, the number is just under 4.3 billion— 2^{32} . When this was originally put into production, no one imagined we would need more addresses. However, with the rapid expansion of the internet and our need to be constantly connected, the IPv4 address space quickly wasn't enough. To allow for better addressing consumption and management, Network Address Translation (NAT) came into effect. Take the example of your home network, which has a router provided commonly by your ISP. That router has a publicly known routable address, provided by your ISP, but when you connect multiple devices to your local network (like your internet TV, tablet, mobile phone, laptop, and others) they all get private addresses assigned to them. NAT is what your router uses to change all those internal IP addresses to that single publicly routable ISP address each time traffic leaves your network and is directed externally. The fact that internal IP

addresses are changed is also a great security feature because it allows internal IP addressing schemes to be hidden. This prevents attackers from being able to easily perform reconnaissance and gather information about a potential target.

LAN Technologies

Institute of Electric and Electronic Engineers (IEEE) is an organization of very bright individuals who meet regularly to discuss and ratify technologies, especially new ones. Through the ratification process, standards and parameters can be stipulated, resulting in uniformity and use of the technologies by vendors. For example, different Wi-Fi standards and parameters are defined by the IEEE 802.11 family of standards, which over time grew from 802.11 to 802.11a, b, and g. Then n, ac and ad came along, and now (at the time of writing) 802.11ax is at the helm. The next version of the standard, IEEE 802.11be, is expected to be approved toward the end of 2024. [Table 4-16](#) summarizes three common IEEE standards from the 802 family.

Wired	Wireless	Virtual LAN (VLAN)
IEEE 802.3 defines a collection of communication standards for physical connections on a wired, Ethernet network.	IEEE 802.11 is a collection of communication standards specific to the implementation of WLAN communication.	IEEE 802.1Q defines the standard for virtual local area networks. VLANs are used to create isolated networks for purposes of security and to minimize

broadcast traffic on a network.

Table 4-16: LAN Technologies

Internet Protocol (IP)

IP is the principal communications protocol for **addressing and routing packets** of data, so that they can travel across networks and arrive at the correct destination.

Internet Protocol v4

Figure 4-10 depicts the IPv4 header. As can be seen, several fields besides the source and destination IP addresses exist in it. The significance here is the 32-bit source IP address and the 32-bit destination IP address, which means IPv4 is limited to a maximum number of 4.3 billion IP addresses. NAT, discussed earlier, mitigated the issue of IP addresses running out while experts in the industry who recognized the onrushing problem developed IPv6.

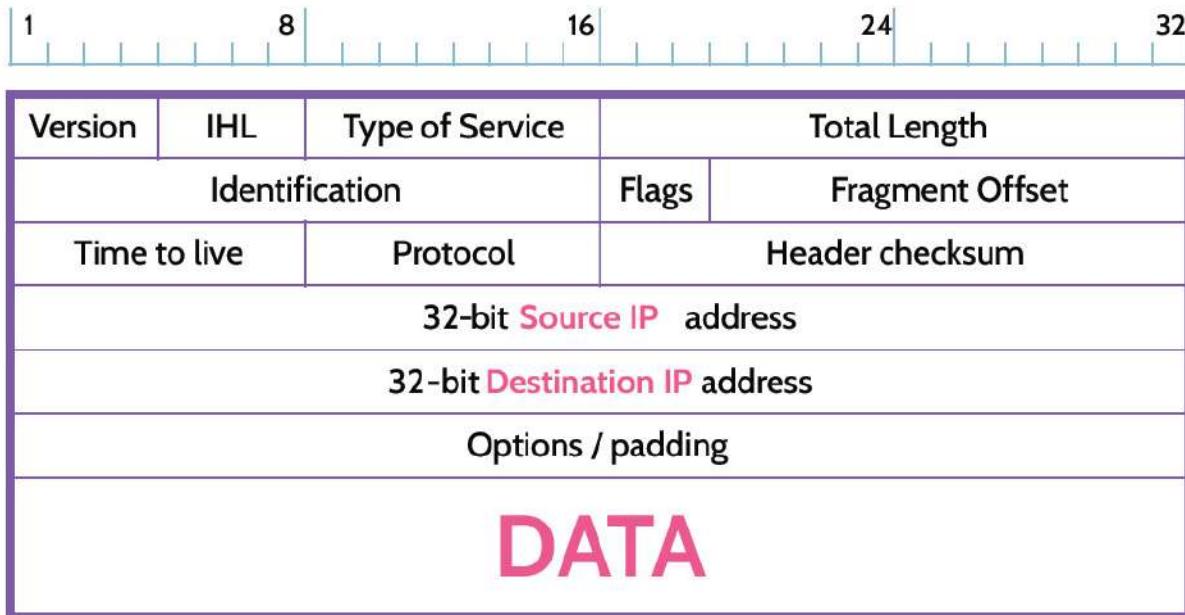


Figure 4-10: **IPv4 Header**

Internet Protocol v6

IPv6 (shown in [Figure 4-11](#)) expands IP addressing to 128 bits, and using the same math noted with regards to IPv4 points to the IPv6 address space being significantly larger—staggeringly so—as 2^{128} is a very big number. Additionally, unlike the four decimal numbers separated by dots that represent IPv4 addresses, an IPv6 address is represented in hexadecimal format, with each group separated by colons.

One benefit of IPv6 is backward compatibility, which explains why some organizations are still using IPv4 at the same time as they move to incorporate IPv6 more fully into their networking environment. At some point, all organizations will only be using IPv6, and experts have predicted that the IPv6 address space

will never be exhausted. But they likely said the same thing about IPv4.

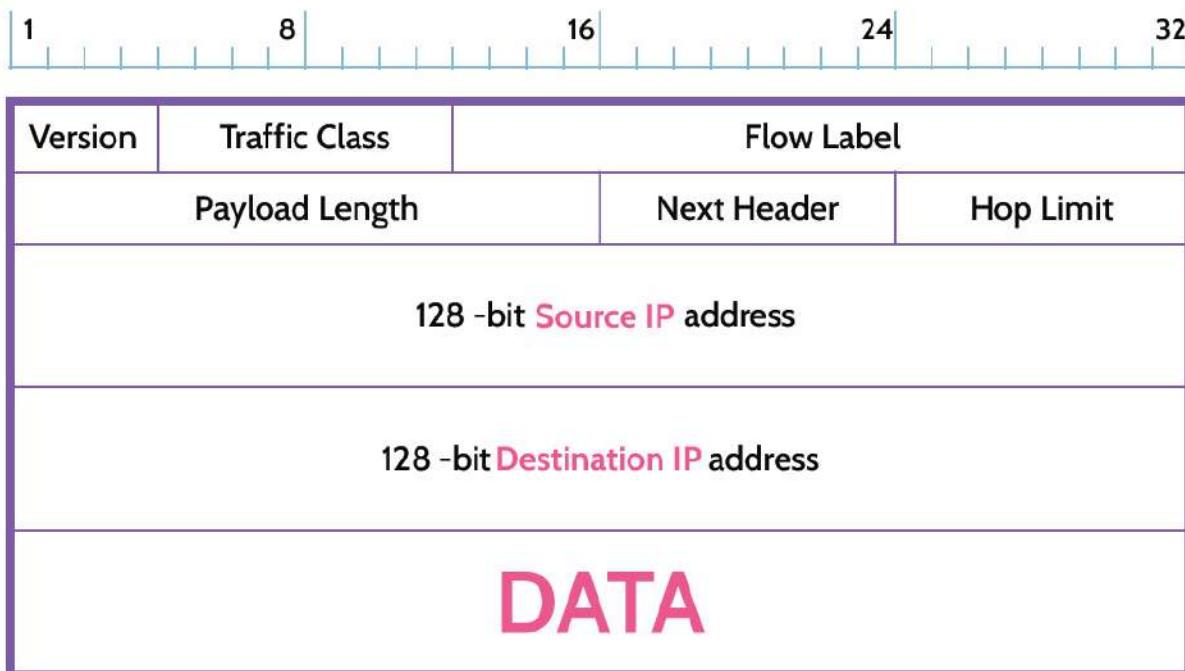


Figure 4-11: **IPv6 Header**

IPv4 versus IPv6

Why IPv6 was created

Table 4-17 contains a comparison between IPv4 and IPv6. Back in the day, nobody ever imagined the need for IP addresses would grow the way it did. With the introduction of IPv6, hopefully the need for IP addresses will never be exceeded again. One other thing that's important to note is that security,

in the form of the protocol IPsec, is supported in implementations of IPv6.

	IPv4	IPv6
Address Size	32-bit (4 bytes)	128-bit (16 bytes)
Address space	$2^{32} =$ 4,294,967,296	$2^{128} =$ 340,282,366,920,938,463,463,374,607,431,768,211,456
Address format (example)	10.0.0.1	0000:0000:0000:0000:0000:ffff:0a00:0001
IPsec	Supported	Supported

Table 4-17: **IPv4 and IPv6 Comparison**

Private IPv4 Addresses

Earlier, when NAT was mentioned, the concepts of public and private IP addresses were also discussed. Unlike public IP addresses, private addresses may be used by any organization or individual and are commonly used for local area networks in the context of large and small organization network environments as well as in home networks. Private IP addresses are non-routable, which means their use within corporate or home networks provides a layer of security between public-facing internet devices and internal devices. In fact, because private IP addresses are non-routable, two businesses next door to one another could use the exact same private IP range for their internal networks. And using NAT, conceivably an

organization with thousands of computers and peripherals could connect to the internet through a single router configured with one public IP address. [Table 4-18](#) lists the private IPv4 IP address ranges (according to RFC 1918), which are not to be used on public networks (like the internet).

From	To
10.0.0.0	10.255.255.255
172.16.0.0	172.31.255.255
192.168.0.0	192.168.255.255

Table 4-18: Private IPv4 Address Ranges (RFC 1918)

Be able to differentiate public IP address from private IP addresses

Network Classes (Subnetting)

In its simplest form, subnetting allows for the creation of networks with more logical host limits versus the limitations imposed by specific IP addressing classes. Looked at another way, if networks were limited to only Class A, B, or C ranges, every network would have only 254, 65,534, or 16+ million IP addresses for host devices, and these limitations could create huge inefficiencies, potential security issues, significant administrative overhead, and potential network-related performance and congestion issues. Through the use of

subnetting, which reflects awareness of the current environment and planning for the future, the proper-size logical host environment can be architected and deployed and can mitigate many, if not all, of the issues noted. [Table 4-19](#) summarizes the subnet classes and related IP address ranges. Neither the network address nor the broadcast address are included in the number of IP addresses, which is why each class is two addresses lower than you would expect from the math.

Know the maximum number of Class A, B, and C IP addresses

	Exponent	Result	Number of Addresses
Class A	2^{24}	16,777,216	16,777,214
Class B	2^{16}	65,536	65,534
Class C	2^8	256	254
Class D	Multicast address		
Class E	Reserved		

Table 4-19: Subnet Classes and Related IP Address Ranges

4.1.7 Layer 4: Transport

CORE CONCEPTS

- TCP and UDP are two transport protocols that reside at Layer 4.
- TCP three-way handshake: SYN, SYN-ACK, ACK
- Ports equate to services that provide specific functionality.
- Layer 4 protocols: TCP, UDP, SSL/TLS

TCP and UDP

As previously mentioned, Layer 4 (the Transport layer) is like a truck that transports information between devices. Delivery services can come in the form of reliable, ordered transmission (using TCP) or in unreliable, unordered transmission (using UDP). Many people refer to UDP as a “send and pray” protocol. This said, both protocols are still heavily relied upon, and each serves a purpose. For reliable, perhaps a bit slower, transmissions, TCP is the clear choice. However, UDP is fast, and for things like video streaming, which requires speed, as well as handling DNS requests, UDP is very efficient.

	
TCP Transmission Control Protocol	UDP User Datagram Protocol

Table 4-20: TCP and UDP

TCP versus UDP Headers

Figure 4-12 illustrates the differences between UDP and TCP headers. As you can see, the UDP header contains much less information than the TCP header, which makes it much easier and faster to process. The additional information contained in the TCP header provides the reliability needed for certain applications, which, however, requires considerable processing time.

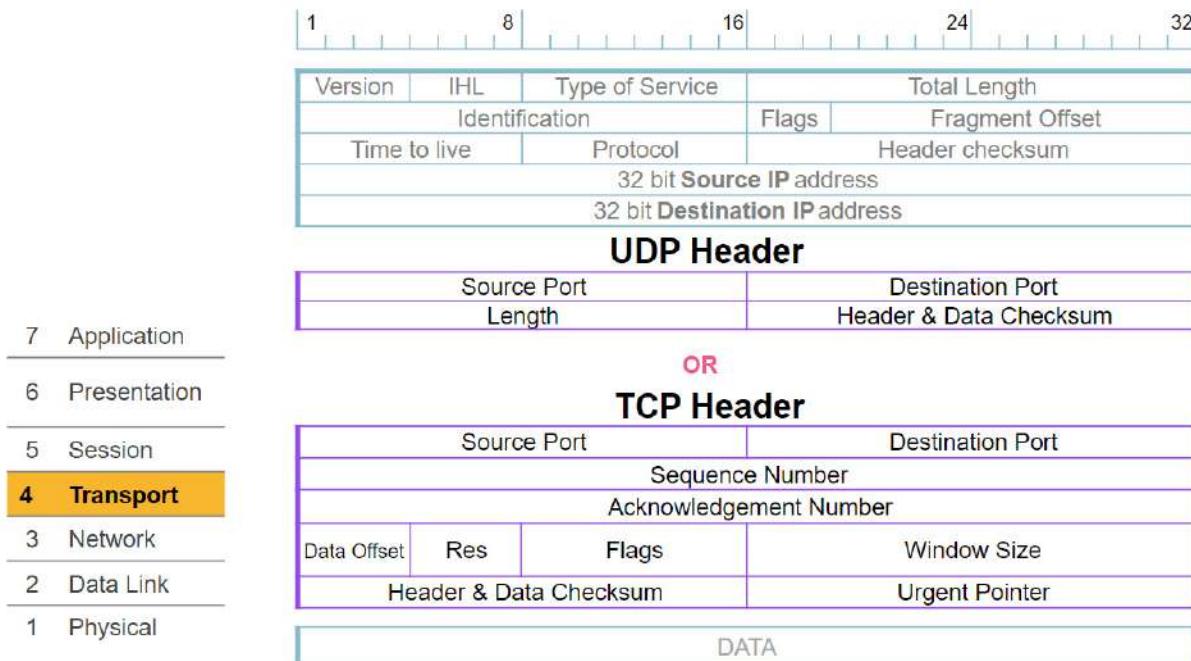


Figure 4-12: UDP and TCP Headers

TCP Three-Way Handshake

TCP provides reliable, ordered and sequenced transmissions. This reliability, however, comes at a cost. Looking back at the truck analogy, for this reliability to be realized, a road first needs

to be built. What's known as the TCP three-way handshake is the building of the technological road to reliably connect hosts across a network. Let's examine this more closely through an example depicted in [Figure 4-13](#).

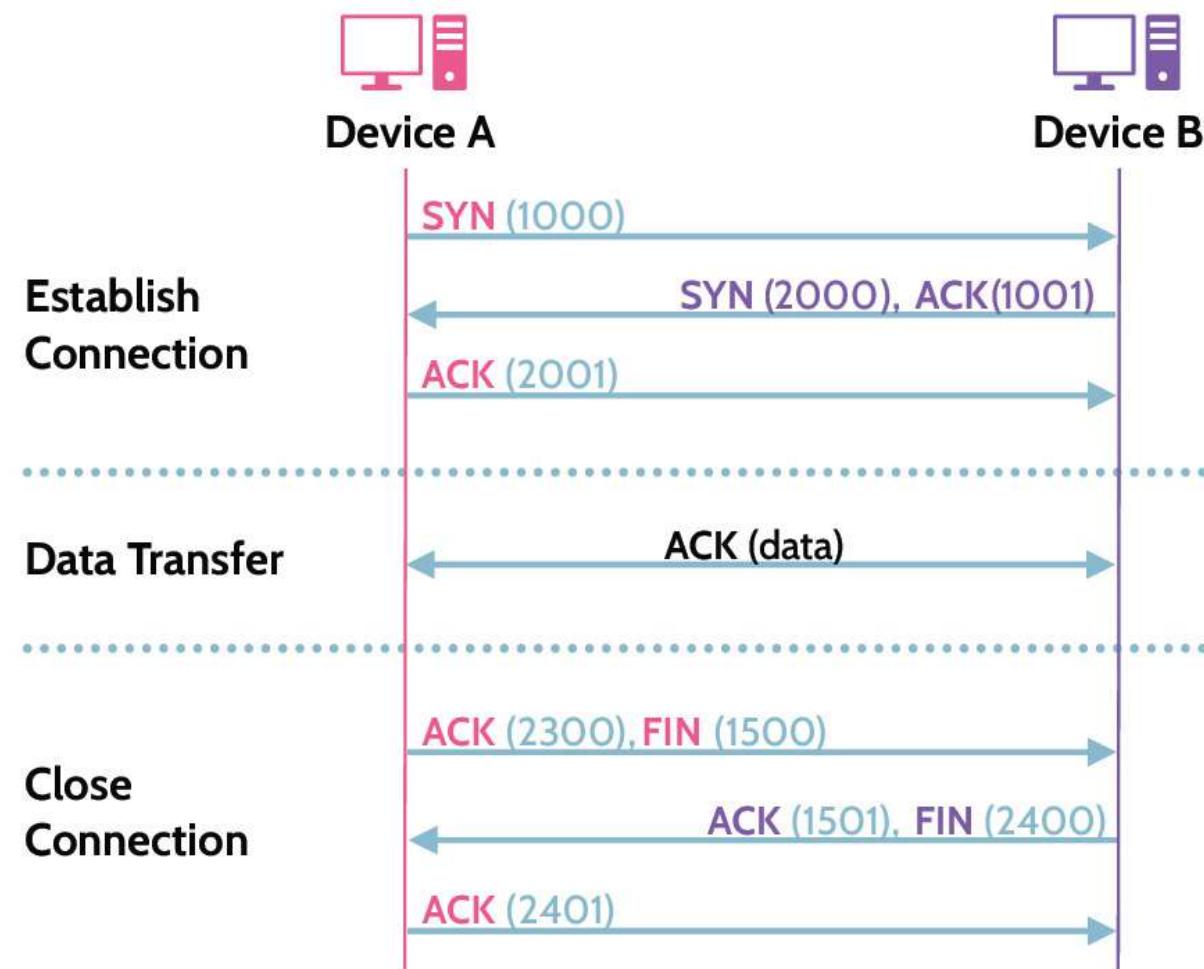


Figure 4-13: TCP Three-Way Handshake

Imagine two devices, A and B. Device A is initiating communication, and Device B is the receiving device. This is the sequence of steps that follows:

1. Device A first sends what's known as a synchronization (SYN) request along with a random session ID or synchronization number (in this case 1000). The session ID number is random to prevent an attacker from predicting the number, which if known could allow the session to be hijacked.
2. The receiving device is going to acknowledge the synchronization request by sending a packet with the ACK (acknowledge) flag set (in which it's going to increment the session ID it received in step #1 by one, making this 1001). In addition, because the goal is two-way communication, the receiving device will also send its own synchronize (SYN) request and random session ID number (in this example, 2000) to the initiating device. That means that the packet Device B sends in step #2, will have two flags set, SYN and ACK.
3. Device A sends a final packet with the ACK flag set, including an incremented session ID number (2001, which is incrementing the SYN value of 2000 that it received earlier by one). So the full final sequence is SYN, SYN – ACK, ACK.

Here's where the reliability of TCP comes in.

Each transmission must be acknowledged by the receiving device, and in the three earlier steps, a full-duplex connection is established; thus, the term three-way handshake.

Communication in both directions can take place and all transmissions are acknowledged.

Additionally, the synchronization number allows each party to know that all data has been received. If a synchronization of 1000 is sent, the number received in reply should be 1001. If something other than 1001 is received, it means one or more packets are missing and retransmission would then be requested. This functionality further ensures the reliability for which TCP is known.

To close the connection, finish requests and acknowledgment of them takes place. That is done in a pair of FIN-ACK packets. If a graceful disconnection is to take place, Device A can send a FIN request, to which Device B would respond with an ACK. In the same regard, Device B may then send a FIN request, to which Device A would respond with an ACK. As such, we would have two pairs of FIN-ACK sequences to gracefully terminate the session.

For all its inherent advantages, the TCP three-way handshake can also be exploited for malicious purposes. Consider this example: When a synchronize request is received, the receiving device tries to acknowledge it by sending an ACK, and the session ID incremented by one. What if many synchronize requests from different hosts are arriving at the same time? Of course, the receiving device will attempt to acknowledge each request. However, if the requests are coming faster than can be handled, the receiving device's process or connection queue is

going to fill up faster than it's emptying. Eventually, the queue is going to fill up, and the system may crash and/or become unresponsive. This is how a SYN flood attack works. To handle something like this, it's best to offload the processing of SYN requests and utilize hardware or software at the Application layer to handle the load. At the Application layer, some type of SYN proxy will be more intelligent and able to handle the incoming requests. It can determine very quickly that a SYN flood is present and simply drop the incoming requests, and the host continues to operate normally.

Ports

Ports equate to services, and services are small applications that provide specific functionality. For example, for the common web service, HTTP, port 80 is used by default. Some of the services associated with different ports are used quite extensively, while others are hardly used. Additionally, some of the most popular or most functional services are also the most often exploited services. Some examples of commonly used ports are denoted in [Table 4-21](#).

20	File Transfer Protocol (FTP) data transfer port
21	File Transfer Protocol (FTP) control port
22	Secure Shell (SSH) (remote login protocol)
23	Telnet (remote command line protocol)
25	Simple Mail Transfer Protocol (SMTP)

80	Hypertext Transfer Protocol (HTTP)
443	Hypertext Transfer Protocol Secure (HTTPS)

Table 4-21: Common TCP Ports

If a service is not needed, especially ones that are dangerous, the associated port should be closed, because attackers may abuse it. Techniques like packet filtering can be used to block packets that reference the associated ports in the header. The use of packet filtering in this sense is part of a process called *hardening*. In addition to using packet filtering, hardening might also involve shutting down services—effectively shutting down the ports—on host machines or within applications or architectures inside a network, because services in those contexts might also provide a communication channel that malicious actors can exploit. If security is needed across a network, hardening can be used to enforce the use of more secure versions of services, like HTTPS or SFTP, instead of HTTP and FTP (which are plaintext protocols). And in cases, where a service cannot be made more secure, a tunneling and encryption protocol like SSH can be used to protect the data transmission. Hardening focuses on security, whether through the removal of or disabling of services, blocking services, or applying patches and upgrades to address known vulnerabilities in an architecture or application.

Among the 65,536 available TCP and UDP ports, three classes exist, also shown in [Figure 4-14](#):

1. **Well known:** Used by specific system services like HTTP, HTTPS, SMTP and similar ones
2. **Registered:** Internet Assigned Numbers Authority (IANA) assigned ports after a specific request is submitted (e.g., UDP 4244 used by the famous Viber, which is a VoIP application)
3. **Dynamic/Private/Ephemeral:** These high ports are dynamically assigned and are often used by applications and other services. For example, when a host machine sends a SYN request to initiate a connection, the source port might be something like 52,367.

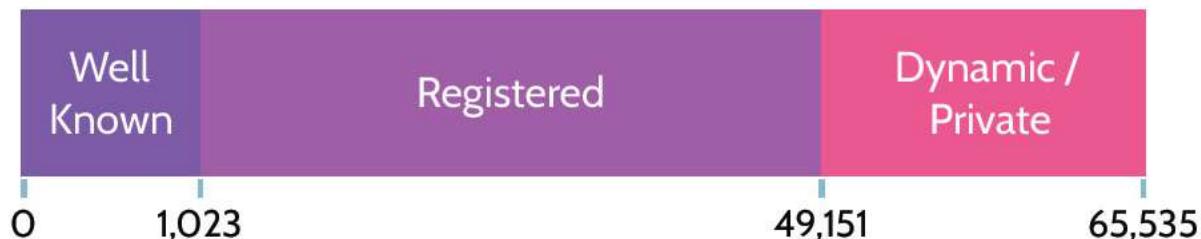


Figure 4-14: Port Ranges

Be familiar with port ranges and common ports and the services they provide

Layer 4 Protocols

Table 4-22 lists some of the most common Layer 4 protocols.

TCP	Transmission Control Protocol—provides reliable, ordered, connection-oriented, sequenced services. The cost of TCP's reliability is speed; it's a slower protocol. When reliability is needed, TCP is the protocol of choice.
UDP	User Datagram Protocol—unreliable, “I'll try my best, but there are no guarantees ...” services. “Send and pray” is a phrase often attributed to UDP. UDP's unreliability also results in it being a much faster protocol. For streaming and for things like DNS requests, UDP is good.
SSL/TLS	Secure Socket Layer (SSL)/Transport Layer Security (TLS) are essentially the same thing—TLS is basically just the name for the later versions of SSL. SSL/TLS is the most used security protocol across the Internet, and it facilitates secure connections between, for example, a browser and a secure server located anywhere in the world.

Table 4-22: Layer 4 Protocols

4.1.8 Layer 5: Session

CORE CONCEPTS

- Supports interhost communication and coordinates dialogue between cooperating application processes
- Maintains a logical connection between two processes on end hosts
- Ideal place for identification and authentication
- Layer 5 protocols: PAP, CHAP, EAP, NetBIOS, RPC
- Layer 5 devices: circuit proxy firewall (also referred to as circuit level gateway)

Understand what happens at the Session layer

Layer 5 (the Session layer) is responsible for establishing and maintaining connections. It can perform functions like establishing, maintaining, synchronizing, and tearing down a connection.

Layer 5 Protocols

Some of the protocols in the table below have already been mentioned. PAP, CHAP, and EAP are the authentication protocols that support the PPP protocol found at Layer 2, which as you may remember facilitates remote connections that are ultimately established at Layer 5, the Session layer. RPC and NetBIOS are additional protocols found at Layer 5, and they also help facilitate remote access. These have been listed in [Table 4-23](#), including the prementioned PAP, CHAP, and EAP as reminders.

PAP	Password Authentication Protocol simply prompts for a user ID and password when establishing a connection. However, passwords are transmitted in plaintext, and the user will never be prompted to change their password as it will be static in nature.
CHAP	Challenge Handshake Authentication Protocol was developed because of PAP's vulnerabilities. CHAP is an improved version of PAP. Passwords are encrypted during transmission, and challenges are routinely sent behind the scene to ensure that an attacker has not replaced the original client.

EAP	The best authentication protocol is called Extensible Authentication Protocol. The word <i>extensible</i> means being able to be extended or designed to allow new capabilities and functionality to be added. With this understanding, EAP allows vendors to adapt the latest authentication technologies, like smart keys and digital certificates, to their products. In fact, due to its inherent strengths, EAP can also be embedded into other things, like wireless security, where it's used with WPA2 for purposes of connecting to wireless networks and authenticating users at the same time.
NetBIOS	Network Basic Input/Output System is a legacy communications protocol that allows computers within a local network to communicate with each other and with various peripheral devices, like file shares and printers.
RPC	Remote Procedure Call is a communications protocol that facilitates communications between clients and servers to execute procedures.

Table 4-23: Layer 5 Protocols

Layer 5 Devices

The primary type of security device found at the Session layer is known as a circuit proxy firewall or gateway. Unlike other firewalls found at the Application layer, circuit proxy firewalls do not filter individual network packets; rather, they monitor and track traffic to determine if it is legitimate by inspecting TCP packet handshakes and sessions. One benefit of circuit proxy firewalls is that they hide details of an internal network from external users through the use of network address translation (NAT), so all outgoing traffic shows only the gateway IP address.

4.1.9 Layer 6: Presentation

CORE CONCEPTS

- **Formatting and encryption of data for end users**
- **Ensures compatible syntax in how the information is represented for exchange by applications**
- **Provides translation, encryption/decryption, and compression/decompression**

Layer 6, the Presentation layer, focuses on the visual elements of data. It focuses on graphics, character conversions, codecs, and other elements that relate to how data is actually presented to users at the Application layer. Codecs are a particularly interesting topic in this context. Codecs are small software programs that enable different types of video files to be viewed on a computer. Users are usually alerted to the need for a codec by their multimedia application, at which point the user will use Google, for example, to search for, download, and install the codec. Unfortunately, malware writers often use codecs to deliver malicious software too, so consumers should be very careful to confirm that a given codec is free from malware.

Like modem standing for modulator/demodulator, codec stands for coder/decoder, or better, compression/decompression, that represent a multitude of ways that digital media is handled. Digital media often entails very large files, and delivering these large files from a media source to consumers requires significant bandwidth and some

technical wizardry—like compression/decompression—to assist. This is where codecs came into play. However, because one codec standard was not adopted, a multitude of codecs were developed to support the multitude of video formats. As vendors wanted their video format to grow in popularity, supporting codecs were often made freely available and shared on the internet, and they became a hotbed of malware as a result. Today, content distribution networks (CDNs) handle the bulk of media distribution around the world. Vendors locate servers around the globe specifically to host things like YouTube videos closer to consumers.

4.1.10 Layer 7: Application

CORE CONCEPTS

- Layer where most functionality resides, also the layer where most security breaches and attacks occur
- Provides a user interface through which a user gains access to communication services
- Ideal place for end-to-end encryption and access control
- Layer 7 protocols: HTTP/S, FTP, DNS, Telnet, SSH, SMTP, SNMP
- Layer 7 devices: gateways and application firewalls

Be familiar with what happens at Layer 7

Layer 7 (the Application layer) is the layer where most functionality resides. For this reason, it's also the layer where

most security breaches and attacks occur. Think about it: applications are created to provide functionality, and this often requires significant amounts of coding. In fact, this points to another CISSP domain that focuses entirely on software development and application security (Domain 8).

Layer 7: Protocols

Be familiar with common Layer 7 protocols and their uses

Among a number of protocols, or services, that run at the Application layer, those listed in [Table 4-24](#) (along with their default ports) are some of the most popular. As can be seen, many of these protocols come in two versions: standard (HTTP) and secure (HTTPS).

HTTP/S	<p>Hypertext Transfer Protocol (HTTP) using TCP port 80 is essentially the language of the web and allows for communication between web browsers and web servers. However, it's not inherently secure. Thus, Hypertext Transfer Protocol Secure (HTTPS) using TCP port 443 was developed to meet the needs of situations where secure web communication is required. HTTPS incorporates SSL/TLS for purposes of tunneling and therefore protecting traffic from interception across the internet.</p>
FTP/FTPS/TFTP	<p>File Transfer Protocol (FTP) using TCP ports 20 and 21 is heavily utilized by companies, but it is also one of the most vulnerable services. Thus, a more secure version of FTP—Secure FTP (SFTP), also referred to as SSH FTP using TCP port 22—was developed, so companies can transmit and retrieve files securely. Note that Trivial File Transport Protocol (TFTP)</p>

	using UDP port 69 is highly insecure, and most companies tend to ban and disable this service as a result.
DNS/DNSSEC	Domain Name Service (DNS) using TCP and UDP port 53 maps hostnames to IP addresses. Surprisingly, DNS includes very little built-in security, and an initiative has been under way for a number of years to better protect DNS services through the use of Secure DNS (DNSSEC). DNSSEC provides for better protection of DNS data itself, so it can't be easily spoofed or otherwise misused.
Telnet	Telnet (using TCP port 23) is one of several protocols that allows a user to remotely connect from one computer to another computer on the same network. As it stands, it's a very insecure protocol, and it's best used in conjunction with something like Secure Shell (SSH), which protects the transfer of data between devices.
SSH	Secure Shell (SSH) using TCP port 22 is a network protocol that utilizes public-key cryptography for purposes of secure connections to a remote computer. Some uses of SSH include login to a shell on a remote server, execution of commands on a remote host, and securing file transfer protocols, among many others.
SMTP/POP3	Simple Mail Transport Protocol (SMTP) using TCP port 25 is the protocol used for sending email over the internet, while Post Office Protocol (POP3) using TCP over port 110 is the one used for receiving email.
SNMP	Simple Network Management Protocol (SNMP) using UDP ports 161 and 162 is a protocol most often used by network administrators to collect, organize, and manage information related to devices on a network. Over the years, SNMP has been revised, and earlier versions (1 and 2) are considered highly vulnerable. SNMPv3 is the latest and recommended version.

Table 4-24: **Layer 7 Protocols**

Layer 7 Devices

Be familiar with common Layer 7 devices and their uses

Adding security at Layer 7 typically involves deploying gateways and firewalls. Essentially, a firewall can be a software or hardware solution and is used to filter traffic between two (or more) networks. A firewall can be as simple as a router that makes filtering decisions on packets based on an access control list, or it can be an extremely sophisticated, specifically purposed device that looks at traffic headers and/or payload to make complicated decisions. Two common Layer 7 devices are summarized in [Table 4-25](#).

Gateways	Gateways are simply connections between domains or networks.
Firewalls	At Layer 7, firewalls are known as application-proxy firewalls. At this layer, they're extremely sophisticated and therefore able to make the most informed and intelligent decisions. Because of this fact, application-proxy firewalls tend to have more processing overhead, and they're typically slower than firewalls found at lower layers.

Table 4-25: Layer 7 Devices

4.1.11 Network Administrator

CORE CONCEPTS

- The term network administrator is often used interchangeably with system administrator.

- Network administrators are primarily responsible for technical matters related to a network.

A network administrator, sometimes also referred to as system administrator, is usually a member of the IT department. Among their responsibilities, they help support information systems and resources and the CIA triad through proper

- Configuration of the network, servers, desktop and mobile computers, and other computing devices,
- Patch and software update management, and
- Vulnerability management.

4.1.12 Convergence and Voice Over IP (VoIP)

CORE CONCEPTS

- Convergence refers to the ability of native IP networks to carry non-IP traffic via what are known as converged protocols.
- Popular converged protocols include Fibre Channel over Ethernet (FCoE), Internet Small Computer Systems Interface (iSCSI), Voice over Internet Protocol (VoIP)
- VoIP Protocols: SRTP (Secure Real-time Transport Protocol) and SIP (Session Initiation Protocol)
- Vishing is a specific type of attack that takes place in VoIP environments

Convergence

In recent years, data networks have been tasked with carrying multiple types of traffic in addition to data traffic. Now it's very common for a network to carry data, voice, multimedia, and other types of traffic. This ability is referred to as IP convergence, as depicted in [Figure 4-15](#).

Recall the earlier discussion about SCADA and ICS. Typically, those systems use non-IP means of communication, which then requires their traffic to be tunneled across IP networks. This points to IP convergence or the ability of native IP networks to carry much more than only IP data. For example, telephony is not a native part of an IP network. IP telephony requires data networks to be able to support IP telephony protocols like H.323 and SIP. This is IP convergence.

From a security perspective, functionality is being added to the basic requirements of a data network, and anytime functionality is added, more potential vulnerabilities and security concerns are also added.

Important converged protocols include Fibre Channel over Ethernet (FCoE), Internet Small Computer Systems Interface (iSCSI), and Voice over IP (VoIP).

Each type of converged protocol serves a purpose, but because data networks do not include native security, carrying a different protocol across a data network poses risks. If confidentiality is a need, encryption can be used, which adds

latency because data needs to be encrypted and then decrypted during transit.



Figure 4-15: **Convergence**

Three commonly used converged protocols are noted in [Table 4-26](#).

FCoE	Fibre Channel over Ethernet
iSCSI	Internet Small Computer Systems Interface
VoIP	Voice over Internet Protocol

Table 4-26: **Converged Protocols**

Be familiar with VoIP, common VoIP protocols, and the types of attack that takes place in VoIP environments

FCoE enables Fibre Channel protocol traffic to be encapsulated and carried over Ethernet networks. iSCSI enables the use of SCSI commands across a network, particularly in the context of

storage-related activities like backups. VoIP, which many people refer to as IP telephony, enables data networks to carry voice traffic using protocols such as H.323 and SIP. With each of these examples, security is not a built-in component, and adding it often results in lowered functionality and efficiency. Other common terms relating to voice are summarized in [Table 4-27](#).

PBX	A private branch exchange is a private telephone network that supports internal communications, usually in the context of an organization or a place like a hotel.
PSTN	A public switched telephone network is essentially the traditional, copper-wire-based telephone network that allows people and businesses to communicate with each other.
VoIP	Voice over Internet Protocol allows people and businesses to communicate via an internet connection, instead of a traditional copper-wire-based phone line.
InfiniBand	A protocol for remote direct memory access (RDMA). This protocol is designed to provide access to memory as quickly as possible across the network. It is commonly used in applications like machine learning.
Compute Express Link	A protocol for connecting CPUs to other components, such as devices and memory, as quickly as possible.

Table 4-27: Common Terms Related to Voice

Considering the way that IP networks operate—with data being broken up into packets and sent across the network, where they might travel in different directions and arrive in a different order than they were sent—it follows that supporting

voice communications in the fundamentally same manner is not as easy. For this reason, specific protocols were developed to ensure smooth and secure VoIP sessions and to mitigate problems that might otherwise result if voice and video session packets were treated the same as ordinary data packets. Two main VoIP protocols are summarized in [Table 4-28](#).

Secure Real-time Transport Protocol (SRTP)	Session Initiation Protocol (SIP)
This is the secure version of RTP, which supports encryption, authentication, integrity, and replay attack protection. Note that RTP is mainly used for streaming voice and video over IP, with no existing security in it. SRTP also provides good bandwidth optimization, low resource requirements, and is independent from underlying protocols. The full description of its operation is described in RFC 3711.	SIP is responsible for initiating, maintaining, and terminating voice and video sessions. It can also support a direct connection between PBX and public telephony networks.

Table 4-28: **VoIP Protocols**

Vishing

Vishing is a form of phishing (**voice phising**) that specifically takes place in the context of VoIP environments. The attacker can easily spoof known or familiar numbers and then obtain information that a victim might not otherwise provide. Note the difference between vishing and smishing. Vishing entails the attacker calling the victim and trying to obtain valuable information or make them perform a certain action (e.g., click

on a URL). Smishing, on the other hand, relates to the attacker sending SMS messages to the victim (**SMS phishing**).

4.1.13 Network Security Attacks

CORE CONCEPTS

- Network attacks resemble network assessments in terms of the steps/phases that each follows, with the exception that attacks also include an exploitation phase.
- Passive attacks do not change the environment or information; active attacks can change information.
- SYN scanning is an active attack that abuses the way the normal three-way TCP handshake operates and helps an attacker determine what services might be running on a target machine.
- SYN flooding abuses the normal three-way TCP handshake by rapidly “flooding” a target machine with multiple SYN requests that ultimately exhaust the target machine’s resources, causing it to crash.
- A Denial-of-Service (DoS) attack is any attack that attempts to impede or completely deny functionality of a system—one machine is being used to cause the loss of functionality. A distributed-denial-of-service (DDoS) attack involves multiple machines acting in unison.
- A man-in-the-middle attack is any attack that originates in between two connected devices.
- Spoofing and masquerading are essentially the same thing—pretending to be someone or something else.
- ARP poisoning involves a malicious user modifying their ARP table to direct network traffic meant for another device to their device.
- ARP uses tables to map IP addresses to physical addresses—the MAC addresses—of a device. Every device on a network has a physical address and an ARP table.

Understand how a network attack differs from a network assessment

With an understanding of the OSI model, relevant protocols and devices found at each layer, and a sense of the importance of security at each layer, let's dive a bit deeper and examine how to best prevent and defend against network attacks. With any attack, a series of steps or phases will be executed.

Network Attack Phases

Network attacks and network assessments (including penetration tests) are very similar. In fact, at a high level, they include identical steps. The main difference is that an attacker will not care if they inflict damage on the target network when they are launching an exploit, while the penetration tester will do that in a more controlled manner and while respecting the scope and rules of engagement.



Figure 4-16: Attack Phases

In virtually all cases, any successful attack will sequentially go through the phases outlined in [Figure 4-16](#). From a security perspective, making each of these phases very difficult or impossible to achieve should be the goal. Preventing attacks is

practically impossible, which is why organizations should focus on making potential attacks time consuming, expensive, and not worth the effort, so attackers will look elsewhere.

With any attack, **reconnaissance** will first be performed. Reconnaissance is the initial gathering of valuable information that will aid the attacker. IP address ranges used by an organization, domain names, services that are running on devices, or operating system in use are all things an attacker might seek as part of the reconnaissance phase.

From a security perspective and with this in mind, one of the best things an organization can do to minimize the amount of information an attacker might discover during the reconnaissance phase is to not publish things like IP address ranges or domain names. Additionally, much information about an organization can be gleaned from social media, from sites like LinkedIn, or from technical forums, where people from around the globe often convene to learn from and help each other. Individually, these bits and pieces of information might not mean much, but when gathered together by a determined attacker, they could very well become the key that opens the door. So, limiting publicly available information is one part of the equation, and creating policy and awareness for the sake of staff is another.

The phase that follows reconnaissance is known as **enumeration**. Usually before account enumeration, the attacker will perform scanning to look for open ports, especially

open ports for interesting services like FTP or HTTP servers, and other software versions of interest. Once those are identified, the attacker will try to enumerate the target for active accounts that can be used to gain access.

During **vulnerability analysis**, the attacker will look for vulnerabilities that can be exploited. That's why it's so helpful for an organization to have a vulnerability management program, as they can use the same tools as attackers and perform similar vulnerability analyses to identify vulnerabilities and gaps and implement controls to make it very difficult for an attack to be successful.

During **exploitation**, the attacker attempts to exploit the vulnerabilities identified during the vulnerability analysis phase. An organization can put detection mechanisms in place to provide alerts if some type of breach or attack does take place during exploitation. For example, if an attacker tries to perform brute-force password cracking against the administrator account, an alert can be generated to signify this activity.

In addition to understanding the network attack phases, it's important to understand the difference between passive and active attacks. In general, a **passive attack** means the attacker is passively performing an activity (e.g., intercepting traffic and inspecting it) without directly interacting with the target. However, during an **active attack**, the attacker will directly engage with target, e.g., during a Denial-of-Service attack.

Passively Eavesdropping

Passive attacks leave the environment unchanged, and the target has no idea that anything has taken or is taking place. Eavesdropping is a great example of a passive attack. Information is being reviewed, but nothing is changing, and the information gained can be used for purposes of exploitation later in the attack. Monitoring or reviewing information traveling across a network is another form of eavesdropping. This is also known as sniffing the network or network sniffing, as shown in [Figure 4-17](#). Data intended for others is intercepted, but no other action is taken.

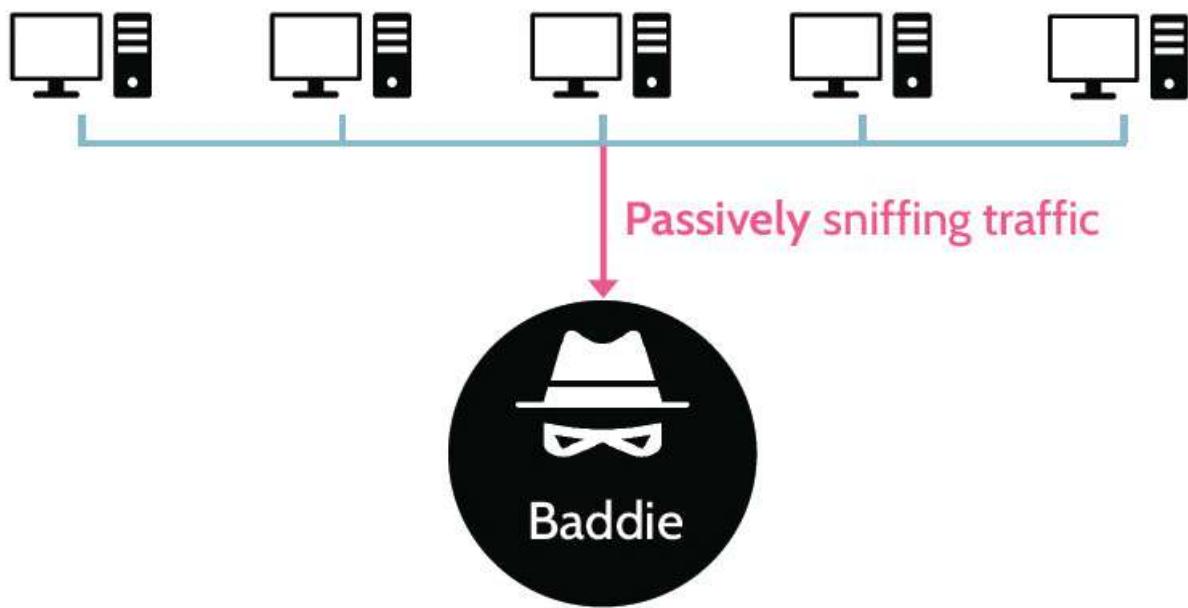


Figure 4-17: **Passive Traffic Sniffing**

Actively Scanning

Understand how SYN scanning and SYN flooding abuse the normal three-way TCP/IP handshake

Unlike passive attacks, active attacks change packet content. Masquerading and denial-of-service attacks are examples of active attacks. During scanning, the target gets alerted of the activity as there's interaction with it, so it's considered an active attack. A classic scanning tool for that is Nmap, and it only supports active operating system fingerprinting, because it sends specific packets and waits for responses to be provided by the target system, which are then reviewed to identify what the target OS may be.

SYN Scanning

Tools like Nmap can easily perform SYN scanning, which consists of the following steps and is also depicted in [Figure 4-18](#).

1. A client sends a SYN packet to a target machine's specific port (e.g., TCP port 80) to try and identify if it's open or closed.
2. Possible responses are:
 - a. If the port is open, the target replies with a SYN-ACK packet, and then the client responds with a final

ACK packet, and the session is established.

- b. If the port is closed, the target responds with a RST packet, and the session is terminated.

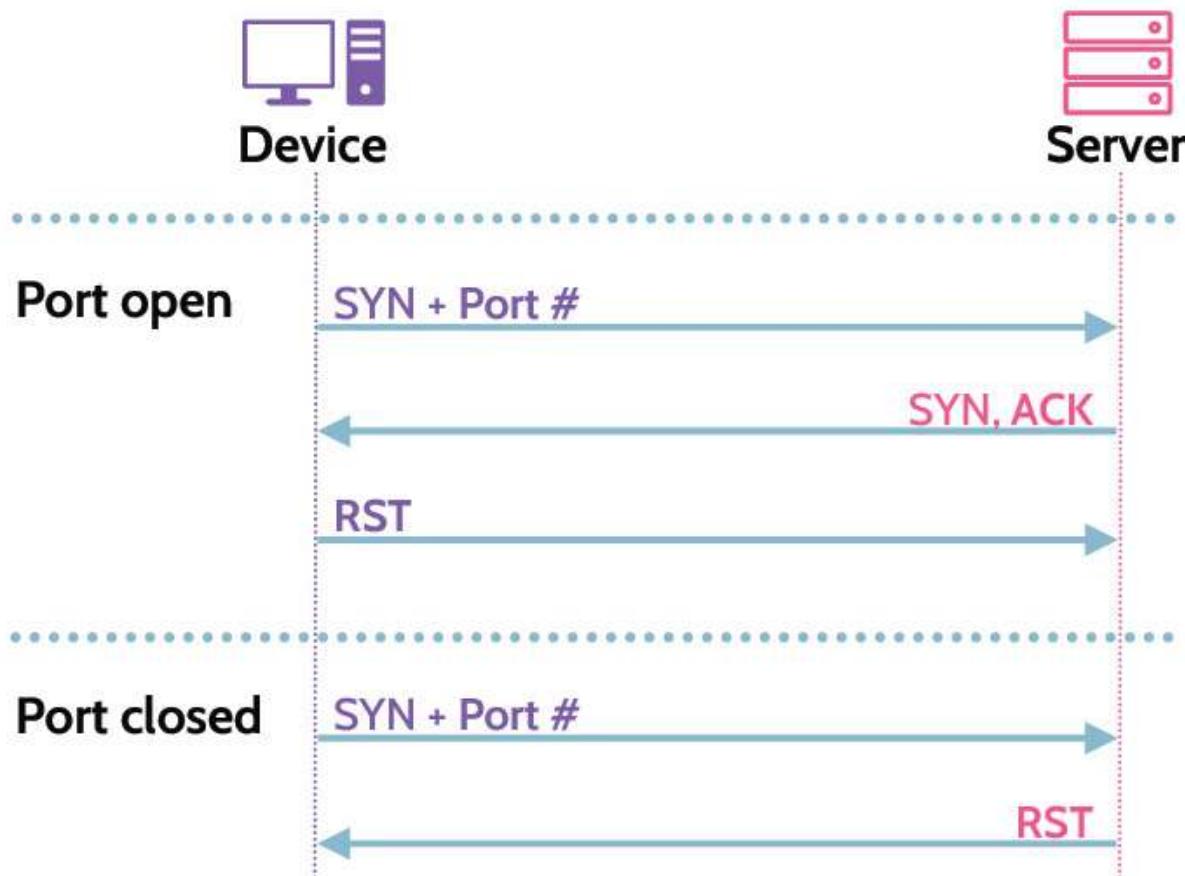


Figure 4-18: **SYN Scan**

Note that if the attacker wants to perform a stealth scan, e.g., using Nmap, they can do so by sending a RST packet at the third step of the three-way handshake and thus never sending the final ACK and formulating a full connection. This is known as a stealth or half-open scan.

SYN Flooding

As noted, with a normal three-way handshake, each open connection consumes a small amount of system resources. SYN flooding, depicted in [Figure 4-19](#), takes advantage of this fact when multiple SYN requests are sent in rapid succession to a target machine, which responds with a SYN-ACK packet, considering these valid connection requests (parts of the first step of the three-way handshake). However, as the volume increases, the target machine is unable to acknowledge new requests fast enough, as its connection table is now filled. Eventually, the target machine's system resources and ability to respond are exhausted, and the machine crashes or stops responding entirely. SYN floods are active attacks, because they impact the host by degrading its performance or bringing it down altogether.

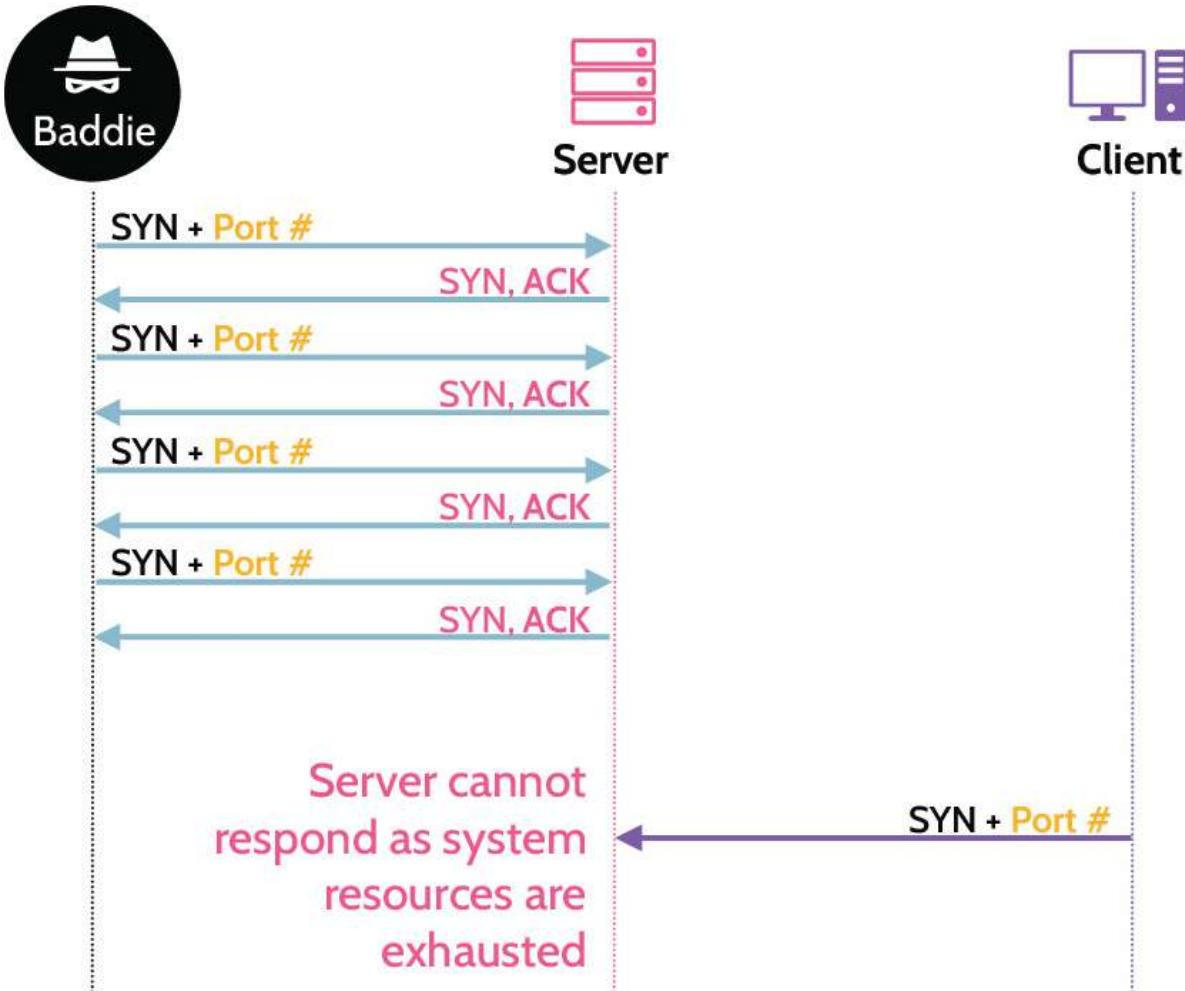


Figure 4-19: **SYN Flood**

One of the best ways to protect against SYN attacks is using a proxy, which acts on behalf of the devices on a network. The proxy has the ability and intelligence to understand whether a SYN scan or flood is taking place and then act accordingly to block the attack. Firewalls and IPS devices are also quite good at detecting SYN flood attacks and dropping offending traffic.

IP-Based Attacks

Understand different types of IP-based attacks

Examples of IP-based attacks include overlapping fragment and teardrop attacks, both of which are discussed in [Table 4-29](#).

Fragment attacks	Overlapping fragments	This attack uses overlapping fragments in an attempt to bypass firewalls and IDS/IPS tools to gain access to a target system. Part of the attack pattern is sent in fragments, along with other data that can pass through the firewall. The goal of the attacker is to pass the attack sequence through any security tools, and when reassembly takes place (at the destination resource) to have the attack sequence created for it to execute.
	Teardrop	With this TCP attack, the attacker sends fragments of packets of differing sizes, out of order, and with fake fragment sequence numbers to a target system. The target system cannot reassemble the packets, causing it to consume its resources and degrade its performance and even crash, thus causing a denial-of-service attack.
IP Spoofing attacks	Smurf	A Smurf attack is executed following steps below: <ol style="list-style-type: none">1. Attacker spoofs their IP address to match the victim's2. Attacker sends multiple ICMP echo request packets to intermediary network devices

		<p>3. Those devices respond with ICMP replies, which are directed to the victim (since the attacker spoofed their IP address to match the victim's in step #1)</p> <p>This causes a denial-of-service attack to take place at the victim's device, which is inundated with ICMP reply traffic.</p>
	Fraggle	<p>Like a Smurf attack, a Fraggle attack is also a denial-of-service attack that begins with IP spoofing. The attacker then sends a massive amount of UDP packets to the target system (destined to ports 7 and 19). Historically these ports relate to the Character Generator Protocol (CHARGEN) service. If either of these ports is open on a target, once they receive a request they will start generating a sequence of characters that will be directed to the victim thus overwhelming it with traffic.</p> <p><i>Note that both the Smurf and Fraggle attacks are considered old attacks, and it's unlikely to see them used today.</i></p>

Table 4-29: **Fragment and IP Spoofing Attacks**

DoS and DDoS

Understand DoS and DDoS, man-in-the-middle, and spoofing attacks

In the context of discussion about IP-based attacks, Denial-of-Service was mentioned. Let's examine this concept a bit more,

including what's known as a Distributed-Denial-of-Service attack. Both types can be seen in [Figure 4-20](#).

A **Denial-of-Service (DoS)** attack is any attack that attempts to impede or completely deny functionality of a system or network. In this sense, one machine is being used to cause the loss of functionality. A **Distributed-Denial-of-Service (DDoS)** attack involves multiple machines acting in unison. In this case, what first must happen is that a number of hosts must be compromised by a threat agent. Once this is accomplished, the compromised hosts can be programmed to work as one, aiming all their processing power and bandwidth at the target, resulting in a Distributed-Denial-of-Service attack.

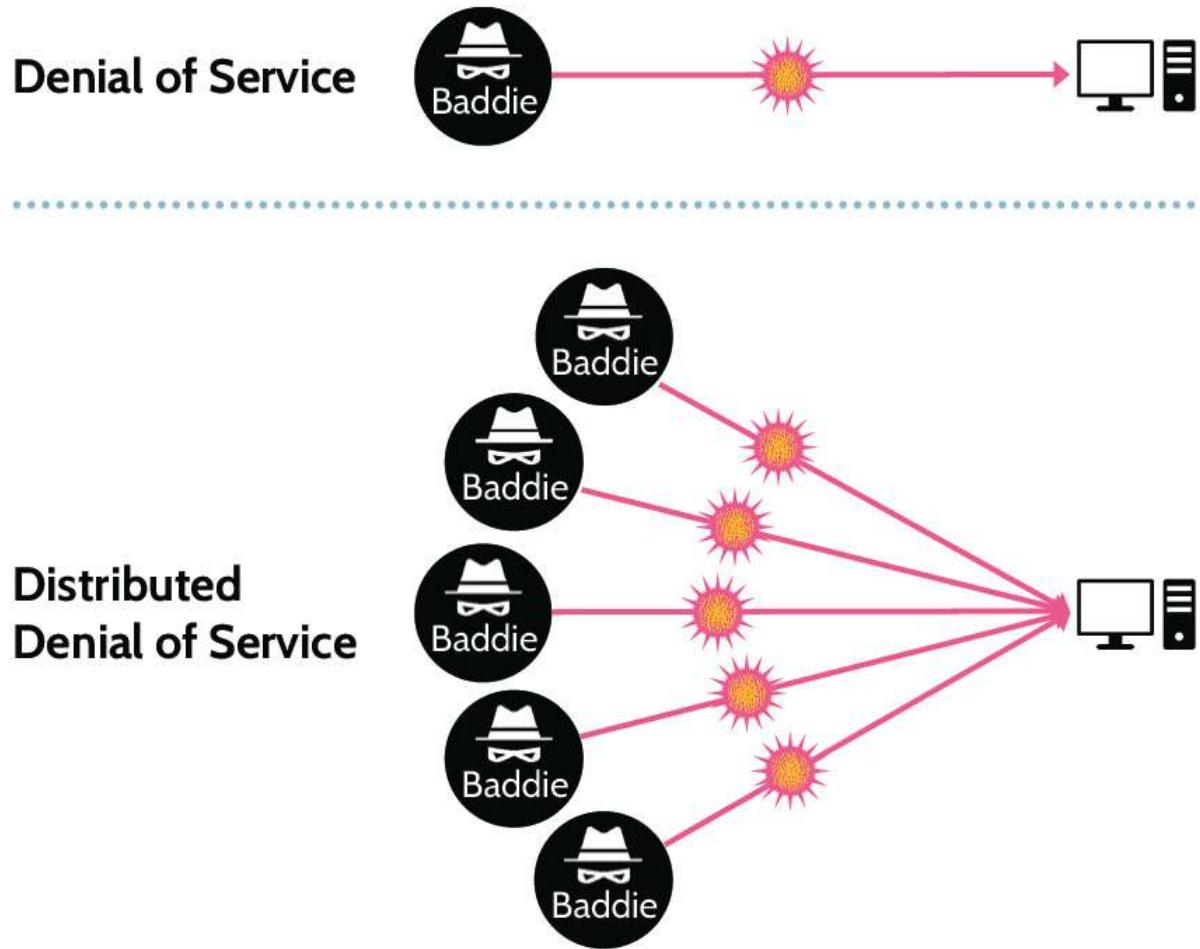


Figure 4-20: DoS and DDoS Attacks

Man-in-the-Middle

A man-in-the-middle attack, depicted in [Figure 4-21](#), manifests when the attacker inserts themselves in the communication path of two entities and thus has an opportunity to intercept and manipulate traffic between them.



Figure 4-21: **Man-in-the-Middle Attack**

Spoofing

Spoofing and masquerading, mentioned earlier, are essentially the same thing. Spoofing is pretending to be someone or something else, because that someone or something usually possesses more privileges or has access to a resource. For example, an attacker may spoof their IP address to bypass an access control list that only accepts traffic from a certain IP address. In that case, the attacker may assume a different IP to send offending traffic to a target. It's really important to remember that when an IP address is spoofed by the attacker, although they can send traffic to a target resource, they won't be able to accept something in response. Any response will be directed to the legitimate holder of that IP address.

Spoofing is not limited to one area of focus. Email, DNS entries, user IDs, IP and MAC addresses, and even biometrics can be spoofed. Really, anything that might prove valuable or useful can be spoofed.

Common Tools and Protocols

Table 4-30 summarizes some common tools and protocols that attackers use to take advantage of networks.

Ping	Ping is a TCP/IP-based utility that is used to determine if a network host is “alive” or available and to measure response time. It can be a very useful and quick network troubleshooting tool. However, because of the way it functions, it can also be a very useful reconnaissance tool, by allowing an attacker to identify potential targets.
Traceroute	Traceroute, like ping, is a TCP/IP-based utility that takes ping a step further and actually maps a network connection from one host to another. Its usefulness comes in that it shows every hop traversed between the two locations. An attacker can take advantage of that to map a target network.
ICMP	ICMP stands for Internet Control Message Protocol and supports TCP/IP utilities like ping and Windows traceroute, among others. ICMP messages come as different types that are differentiated by types and associated codes (think of them as subtypes). For example, when using the ping utility to see if a host is alive, ICMP might return a type 3 code 1 “Destination Unreachable” message. Type 3 means the destination is not reachable, while code 1 shows that this is due to the host not being reachable in particular. For example, a code 0 would indicate the target network was unreachable. This can be valuable information for an attacker.
DHCP	Dynamic Host Control Protocol is used to assign a valid IP address to a device when it first connects to a network. Whether at home, on a corporate network, or anywhere else, if a connection to a network is made a valid IP address is assigned to the connecting device. DHCP does this automatically. An attacker can use a tool to make them seem like they are a legitimate DHCP server on the network and thus trick victims

	into getting connection information, which may include setting up the attacker's machine as their default gateway. In that regard, the attacker would then be able to intercept all network traffic, as it would pass through their machine.
Ipconfig	Used mainly in Windows systems to display TCP/IP network configuration values and can refresh Dynamic Host Configuration Protocol and Domain Name System settings.
WHOIS	A query and response tool that is widely used for querying databases that store the registered users or assignees of an internet resource, such as a domain name, an IP address block, or an autonomous system public domain registration information.
Dig	A network administration command-line tool for querying the Domain Name System (DNS).
Putty	A free and open source terminal emulator, serial console, and network file transfer application. It supports several network protocols, including SCP, SSH, Telnet, rlogin, and raw socket connection.
Nmap	A free and open source network scanner used to discover hosts and services on a computer network by sending packets and analyzing the responses.
John the Ripper	John the Ripper (JtR) is an open source password cracking tool and is available for many operating systems.
Netstat	A command-line network utility that displays network connections for TCP and UDP (both incoming and outgoing), routing tables, and a number of network interface (network interface controller or software-defined network interface) and network protocol statistics.
Nslookup	A network administration command-line tool used for querying the Domain Name System (DNS) to obtain domain name or IP

address mapping or other DNS records.

Table 4-30: **Common Tools and Utilities Attackers Leverage**

ARP Poisoning

Understand what ARP tables do and ARP poisoning

We already mentioned how ARP works earlier. An attacker can try to leverage the protocol's operation to intercept traffic destined for a legitimate destination as depicted in [Figure 4-22](#). A switch has a content addressable memory (CAM) table, which holds mappings of MAC addresses and which ports the respective devices are connected to. When traffic for a device for which the switch doesn't have an entry reaches it, it will send a broadcast message that essentially asks, "What MAC address belongs to this IP address?" All of the devices will look at their ARP tables, and the device associated with that IP address will reply back, "That's me, here's my MAC address," at which point the switch will send the request to that device. It's actually quite simple, and this fact explains why it's equally simple for someone to modify their ARP table to direct network traffic meant for another device to their device. There's no authentication or security built into ARP tables, and the same holds true for many other devices and protocols that utilize tables. As such, if the attacker were to send an ARP reply to the switch denoting a specific MAC address, then the switch would consider that legitimate and log it in its ARP table.

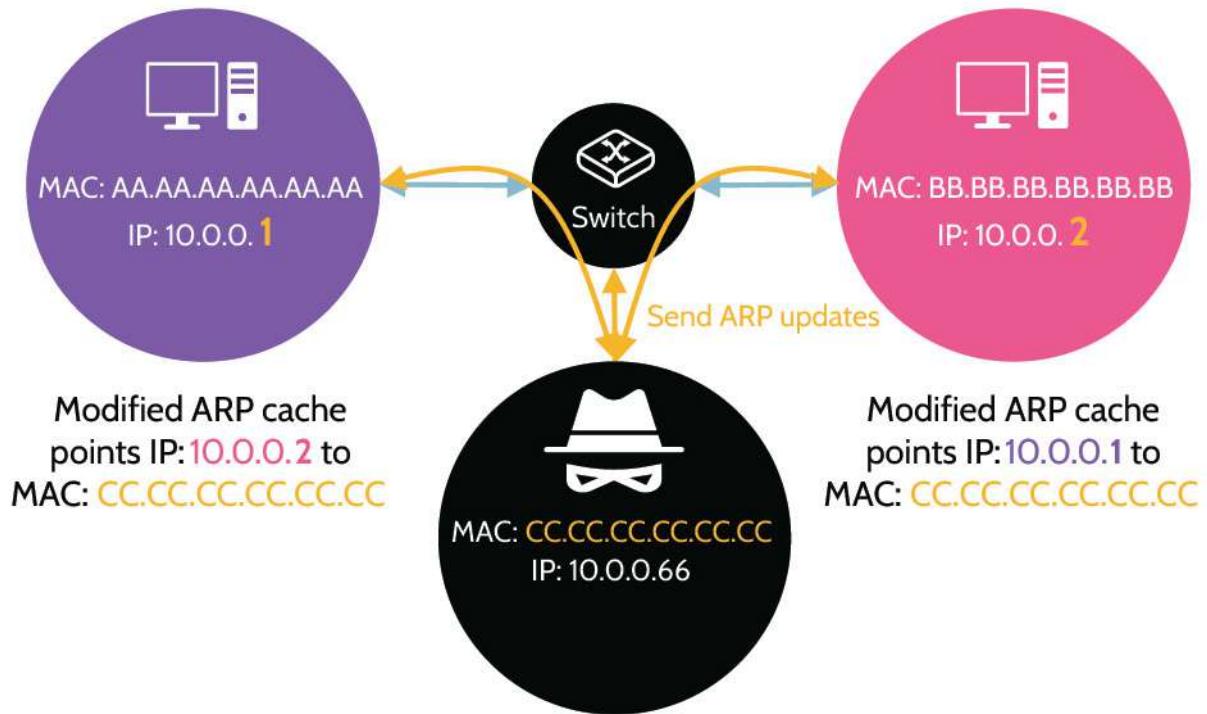


Figure 4-22: ARP Poisoning

Monitoring of activities across network segments is one of the best ways to prevent ARP poisoning. This holds true with regards to DNS too, though in recent years, DNSSEC has been developed and employed by many organizations to prevent DNS poisoning. With all these vulnerabilities, if good preventive or detective controls don't exist, good compensating controls should be implemented. This typically means more logging and monitoring, depending on the value of the asset in question.

4.1.14 Wireless

CORE CONCEPTS

- IEEE 802.11 specifications define wireless standards.
- Wireless authentication requires an authenticated key exchange mechanism.
- Temporal Key Integrity Protocol (TKIP) was designed to replace WEP without requiring the replacement of legacy hardware, due to a significant flaw found in WEP.
- To protect any wireless communication, four things are needed: access control, authentication, integrity protection, and encryption.

Early versions of Bluetooth, Wi-Fi and similar technologies had limited security features. Regardless of the wireless technology in use, one of the easiest ways to introduce security is through segregation. In other words, especially with Wi-Fi, different networks can be set up. Guests can be segregated to their own network, employees to an internal network, and vendors or external visitors to yet another network, with each network segment containing security appropriate to the need. All of this points to a network architecture as illustrated in [Figure 4-23](#).

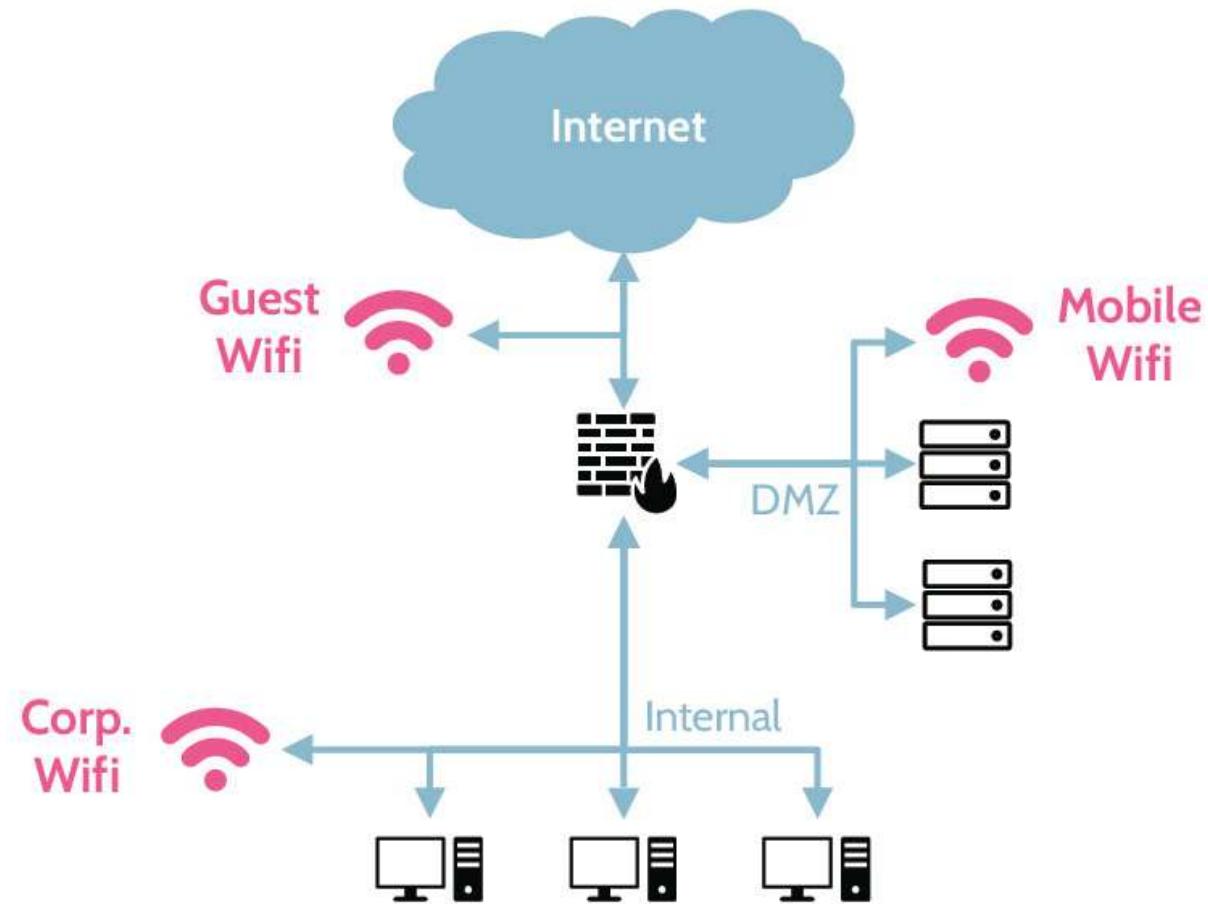
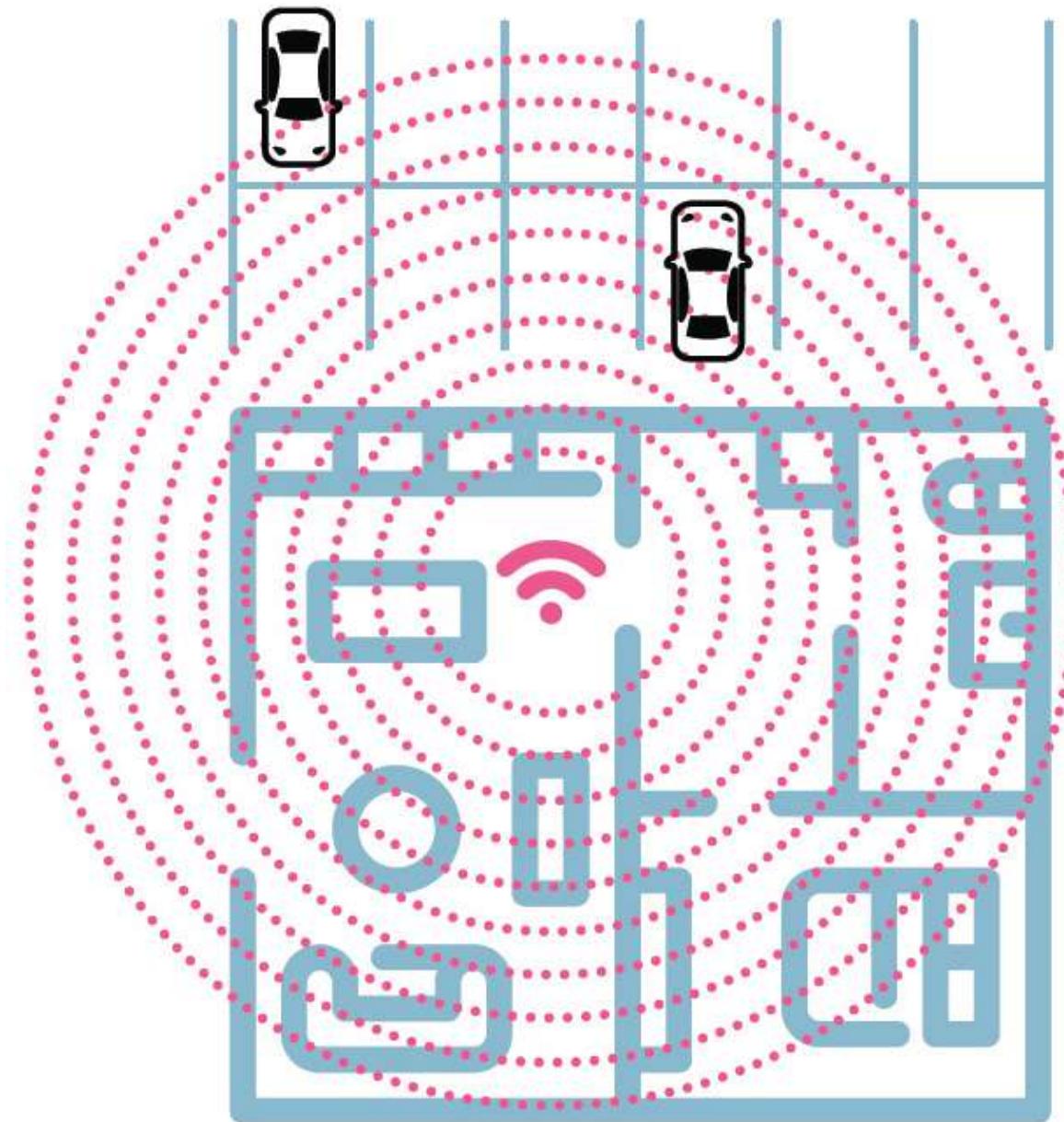


Figure 4-23: Network Architecture

Radio Frequency Management

Radio frequency management refers to the placement of devices that broadcast wireless traffic. For example, the placement of Wi-Fi access points within a building should carefully consider how far out wireless signals extend outside the building. If the signal extends to the parking lot, somebody could sit in a car for hours and attempt to break into the network. Ideally, access points are located in such a way that

authorized individuals can utilize the Wi-Fi service and untrusted people cannot.



Managing Wi-Fi signals is called radio frequency management, which can prove quite challenging at times as there are an abundance of different devices operating at different areas of

the radio spectrum, as also seen in [Figure 4-24](#). In addition to Wi-Fi, other technologies like Bluetooth, cellular, and RFID, can be found. Despite the differing technologies, they all share a common feature. They're all examples of emanations. Though a number of frequencies exist, three of them—2.4 GHz, 5 GHz, and 900 MHz—are of particular interest. These are unlicensed radio frequencies, which means that anybody can create a technology or device that operates within those frequencies.

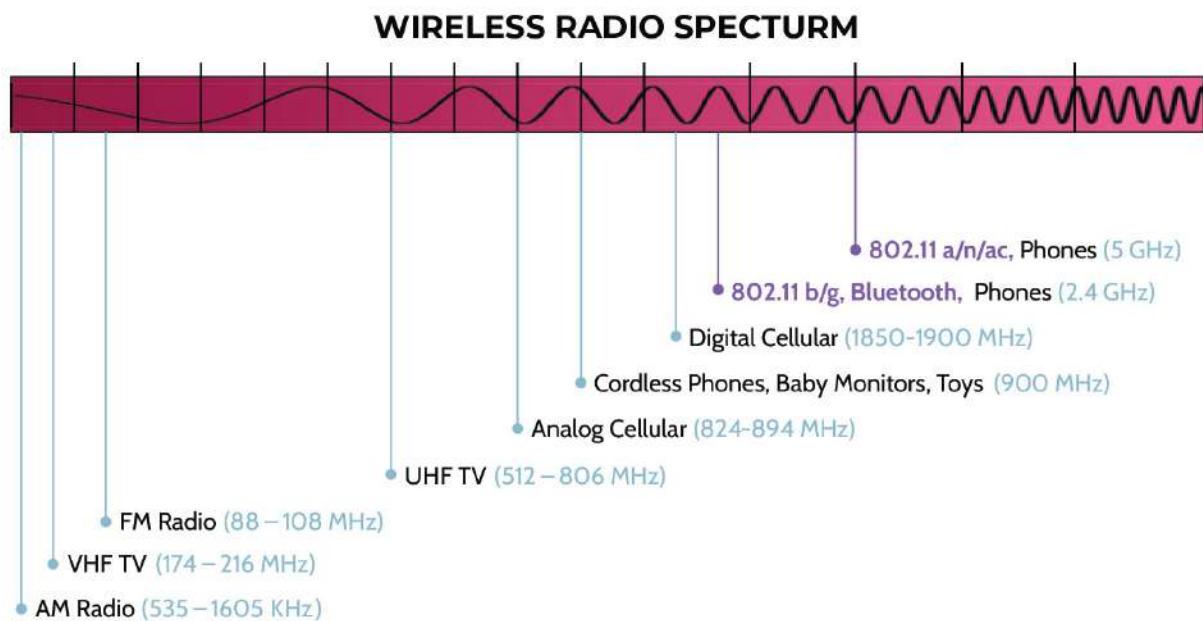


Figure 4-24: Wireless Radio Spectrum

Wireless Technologies

As noted in [Figure 4-24](#), the wireless radio spectrum includes a number of wireless technologies, some of which are summarized in [Table 4-31](#). In its most basic form, wireless refers to communication without using the physical medium of wires

or cables. It is the transmission of data over the air—via channels encompassed by the wireless radio spectrum.

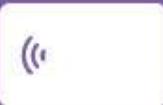
 Wi-Fi	Wi-Fi, described in more detail below, follows standards outlined by IEEE 802.11 specifications. Wi-Fi is used in many contexts, from printing to connecting to network resources and turning a mobile phone into a Wi-Fi hot spot for use by a computer or other devices.
 Bluetooth	Bluetooth is a wireless technology that supports devices in close proximity. For example, many people today use Bluetooth keyboards and mice with their computers instead of wired versions. Similarly, most modern automobile audio systems include Bluetooth, which allows a similarly equipped mobile phone to connect and play music stored on the phone through the automobile speakers.
 Cellular	Cellular is used to refer to wireless protocols and standards used by mobile phones. Within this spectrum, Code Division Multiple Access (CDMA) , Global System for Mobiles (GSM) , 3G, 4G, and now 5G all represent the specific ways the devices are designed to communicate with cellular networks.
 RFID	RFID, also known as Radio Frequency Identification, refers to a specific type of wireless system that is made up of readers and some type of tag, chip, or label. The reader emits radio waves and receives signals back from the object. The tag, chip, or label can be used for any of a number of purposes, such as asset management, inventory control, or similar type of tracking. RFID technology allows otherwise mundane tasks to be automated as well as to reduce errors that might otherwise occur when processes are performed manually.

Table 4-31: Important Wireless Technologies
802.11 Wireless Protocol Family

[Table 4-32](#) outlines different IEEE 802.11 specifications, starting with 802.11 and progressing to one of the latest: 802.11bn (Wi-Fi 8). This is essentially a look down wireless memory lane. Each variation of wireless is noted, along with the frequency range(s) a given specification covers as well as the maximum achievable speed. However, as noted above, each specification shares the same security concern—security is not native to any of them. Note that at the time of writing, 802.11be is expected to be approved late in 2024, while 802.11bn is expected to be approved in 2028.

Type	Frequency	Top speed
802.11	2.4 GHz	2 Mbps
802.11a	5 GHz	54 Mbps
802.11b	2.4 GHz	11 Mbps
802.11g	2.4 GHz	54 Mbps
802.11n	2.4 & 5 GHz	72 – 600 Mbps
802.11ac	5 GHz	422 – 1300 Mbps
802.11ax (Wi-Fi 6)	2.4 & 5 GHz	10 Gbps
802.11ax (Wi-Fi 6E)	2.4, 5 & 6 GHz	10 Gbps

802.11ad (WiGig)	60 GHz	7 Gbps
802.11be (Wi-Fi 7)	2.4, 5 & 6 GHz	40 Gbps
802.11bn (Wi-Fi 8)	?	?

Table 4-32: 802.11 Protocol Family

802.11 Security Solutions

Understand common wireless standards, from weakest to strongest, and elements of each

To secure wireless communication, four (4) security services are required: access control, authentication, encryption, and integrity protection. Over the years, different bundled security protocols have been developed that provide these services, and they're represented by three main standards in the wireless industry: 802.1X, WPA, and WPA2. It's worth noting that WPA3 has been released but is still not widely implemented.

Table 4-33 outlines how each of the security services are implemented. For example, WPA2 uses a pre-shared key (PSK) or 802.1X for access control. It also uses EAP or PSK for authentication, CCMP (which is essentially AES Counter Mode) for encryption, and CCMP-AES plus **Message Authentication Code (MAC)** to ensure integrity.

802.1x Dynamic	Wi-Fi Protected	Wi-Fi Protected	Wi-Fi Protected
---------------------------	----------------------------	----------------------------	----------------------------

	WEP	Access (WPA)	Access 2 (WPA2)	Access 3 (WPA3)
Released	1997	2003	2004	2018
Access Control	802.1X	802.1X or Pre-Shared Key	802.1X or Pre- Shared Key	802.1X or Pre-Shared Key
Authentication	EAP methods	EAP methods or Pre-Shared Key	EAP methods or Pre-Shared Key	EAP methods or Pre-Shared Key
Encryption	WEP	TKIP (RC4)	CCMP (AES Counter Mode)	CCMP or GCMP (Galois Counter Mode Protocol)
Integrity	None	Michael MIC	CCMP	CCMP or GCMP

Table 4-33: **WEP and WPA**

Wireless Authentication

There are three main ways used to authenticate to a wireless network:

- Open authentication (a device can connect by using the network's SSID, with no security enabled).

- Shared key (a pre-shared network key is used, which is shared across any device that needs to connect to the network).
- EAP is used (authentication requires an authenticated **key exchange** mechanism), which can lead to one- or two-factor authentication:
 - **One-factor:** EAP-MD5, LEAP, PEAP-MSCHAP, TTLS-MSCHAP, EAP-SIM
 - **Two-factor:** EAP-TLS, TTLS with OTP, and PEAP-GTC

Note that to achieve robust authentication in wireless networks, it's ideal to have mutual authentication enforced, which means that the client is asserted of the access point it's connecting to, and the access point can be assured that the client is valid.

Wireless Encryption

The main encryption technologies are:

- Temporal Key Integrity Protocol (**TKIP**): Was instituted as a fix of the WEP vulnerabilities and was used in WPA (with an RC4 stream cipher and 128 bit per-packet keys) but is vulnerable to attacks mainly due to the WPA requirement to support hardware compatibility with WEP.

- Counter-Mode-CBC-MAC Protocol (**CCMP**): Uses Advanced Encryption Standard (**AES**) with 128-bit keys, which is used in WPA2 and WPA3.

Understand the importance of TKIP relative to WEP and why WEP is considered weak

Wireless Integrity Protection

Main methods for integrity protection are:

- TKIP uses a Message Integrity Code called “Michael.”
- WPA2 uses CCMP (which uses AES in CBC-MAC mode).

A short summary of TKIP is provided below:

- Designed to **replace WEP** without requiring the replacement of legacy hardware
- Required due to significant flaw found in WEP—specifically, the use of a weak IV that allowed WEP to be easily cracked
- Sends each new packet with a unique encryption key (key mixing)
- TKIP is no longer considered secure and is superseded by AES

TKIP is an interesting protocol, as it served as a stopgap when significant flaws and vulnerabilities were discovered with WEP. Immediately, attempts were made to mitigate the vulnerabilities with WEP. TKIP, which uses a 128-bit key size symmetric stream cipher known as RC4, was developed. A better version of TKIP was developed, a counter mode version, that uses stronger encryption known as AES.

To protect any wireless communication, four things are needed: access control, authentication, integrity protection, and encryption. TKIP implements a **Message Integrity Check (MIC)**, often referred to as “Michael.” The details about how it works are not necessary for the sake of the CISSP exam; it’s enough to know that it provides integrity control, like a hashing algorithm. Additionally, TKIP was implemented in a manner almost as if applying a patch. Technology that already used WEP would still work, and the more secure encryption protocol TKIP would handle encryption needs.

4.1.15 VLAN and SDN

CORE CONCEPTS

- **VLAN = Virtual Local Area Network, which allows local networks to be created using hardware devices, like Layer 3 switches, and other technologies and software. A VLAN reduces the need for physical wiring.**
- **IEEE 802.1Q is the standard that supports VLANs on networks.**
- **SDN = Software-Defined Networks, which refers to networks created and managed using software.**
- **SDN architecture includes application, control, and data planes.**

■ Communication between application and control planes is facilitated by northbound APIs; communication between data and control planes is facilitated by southbound APIs.

Virtual Local Area Network (VLAN)

Understand the basic premise and use of VLANs

Virtualization technologies are the underpinnings of cloud computing. Though virtualization is more well-known now, virtualization capabilities have been around since the early days of computing. Mainframe computers could actually be broken up into virtualized environments, with each environment having its own operating system, devices, peripherals, and so on. From a security perspective, isolating things through this type of separation creates a more secure environment. Since the days of mainframe computers, virtualization has matured significantly, to the point where cloud computing is now at the forefront of most often used technologies.

Virtualizing networks is separating networks by creating segments that are separate from each other. Through this isolation, a more secure environment can be created.

VLAN stands for virtual local area network, and a VLAN can be created using devices, technologies, and software. A VLAN **reduces the need for physical rewiring by creating virtual tunnels through physical networks to connect devices.** VLAN

segments should be created and used based upon the value of those segments. Typically, a Layer 3 switch can be used to create VLANs, based upon needs and value. [Figure 4-25](#) shows a simple example of two VLANs (VLAN 1 and VLAN 2) that were created across two switches. Specific ports are configured to support a given VLAN, and any devices connected to those same ports will be on the same VLAN.

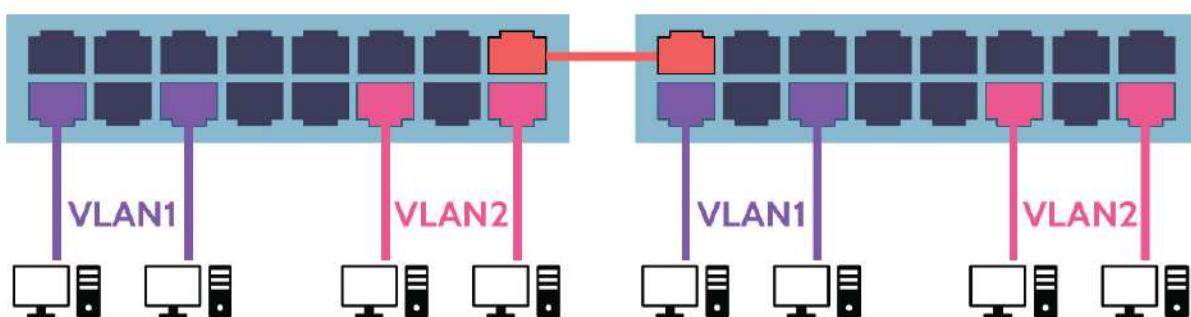


Figure 4-25: VLAN Operation

Software-Defined Networks (SDN)

Understand how a SDN differs from a traditional wired network

As the name suggests, software-defined networking (SDN) is basically creating and managing a network using software. Software applications can be written that are very intelligent and function to the point that they mimic hardware devices, like routers, switches, and firewalls. Of course, as has been

discussed, this software-based functionality comes at the cost of potential vulnerabilities that can be exploited.

[Figure 4-26](#) illustrates a traditional network and a software-defined network (SDN). In the latter, applications provide the functionality provided by hardware devices found in a traditional network. With a software-defined network, some traditional physical network elements—like cabling—are still required; otherwise, the primary network elements are virtualized, and the network functions exactly as if everything were hardware based.

A SDN provides abstraction from the underlying physical network to enable rapid reconfiguration with centralized control. The abstraction is accomplished in the context of “planes”—areas where certain processes are executed.

With a SDN, two primary planes are found—the data plane and the control plane. Looking closely at [Figure 4-26](#), there is one control plane for the SDN. The **control plane** acts as the “brain” of the SDN and understands everything about the network. More specifically, it is the part of the SDN that determines how the routing of packets across the network should take place. Based upon this information, it follows that routing tables and routing protocols would be part of the control plane.

Based upon how the control plane determines packets should be forwarded, the **data plane** handles the actual forwarding of packets to the proper destination.

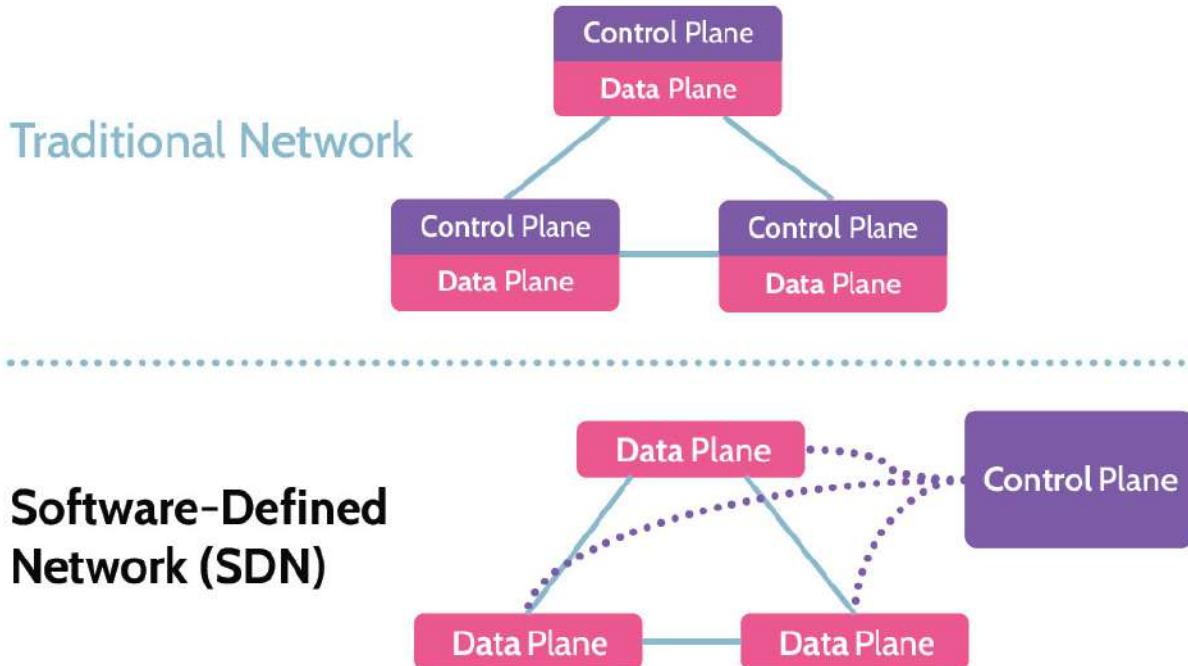


Figure 4-26: Traditional vs. SDN Network

SDN Architecture

Understand the basic components, including flow, of an SDN

Figure 4-27 further illustrates the architecture of an SDN and brings to light another plane—the Application plane, as well as two more terms: Northbound and Southbound APIs.

The **Application plane** in an SDN hosts the applications and services that make network function-related requests from the control plane via northbound APIs. Application plane applications and services could include things like firewall and

other security functionality, reporting, and network management, to name a few examples. Northbound APIs only pertain to communication between the Application and Control planes.

Similarly, **Southbound APIs** only pertain to communication between the Data and Control planes. The southbound interface facilitates communication between the SDN controller and the physical networking hardware.

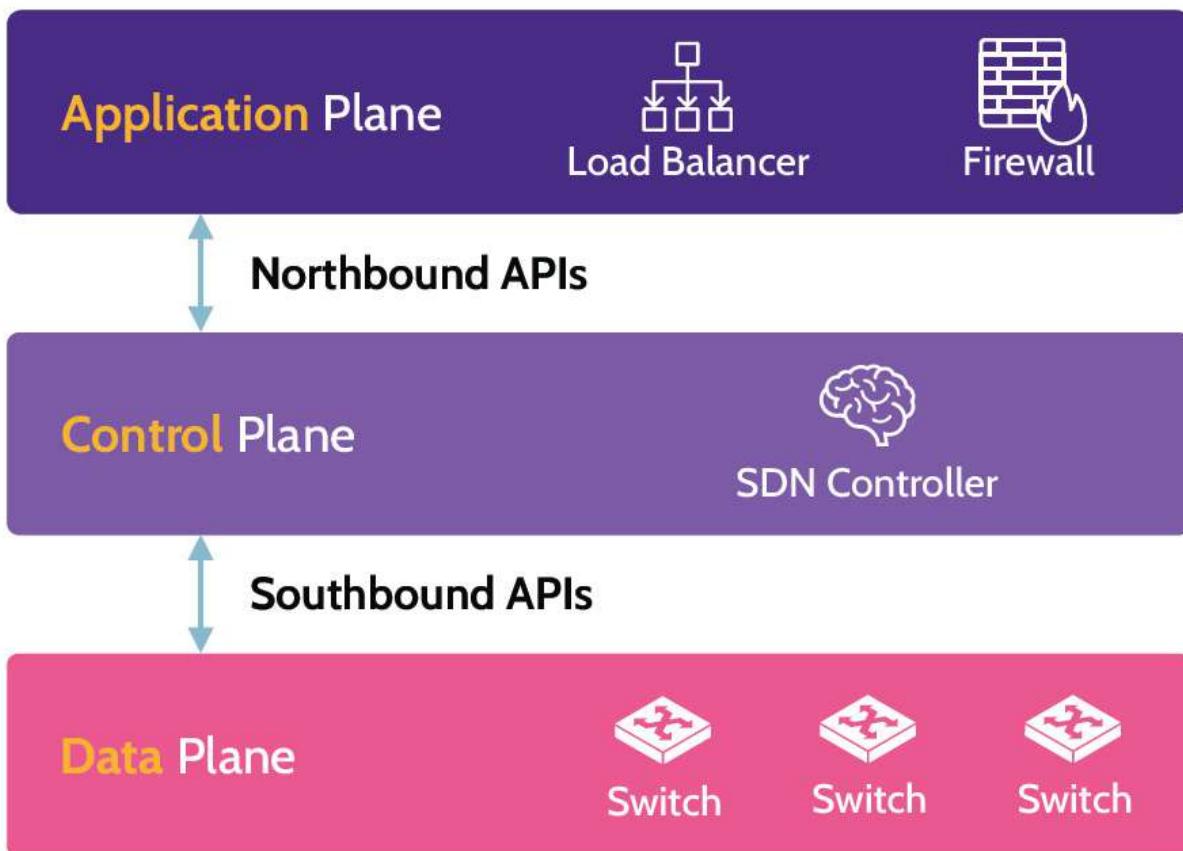


Figure 4-27: **SDN Architecture**

Virtual Private Clouds (VPCs)

A virtual private cloud (VPC) is a portion of a public cloud provider's infrastructure. It's a customizable and logically isolated cloud that providers offer to customers. VPCs do not involve separate physical hardware, which means that the isolation is purely virtual.

IEEE 802.1Q

Know the IEEE standard that supports VLANs and SDNs

IEEE 802.1Q refers to the IEEE standard that supports VLANs and SDNs. Among other things, the standard defines a system of what is known as "VLAN tagging" for network traffic as well as how bridges and switches should handle tagged frames. In effect it ensures that traffic destined for a specific VLAN, e.g., VLAN1, is able to only reach that VLAN and no other one.

4.1.16 Wide Area Networks (WAN)

CORE CONCEPTS

- **WANs connect LANs through technologies such as: dedicated leased lines, dial-up phone lines, satellite and other wireless links, and data packet carrier services.**
- **WAN protocols include: X.25, Frame Relay, Asynchronous Transfer Mode (ATM), and Multi-Protocol Label Switching (MPLS)**

To understand what constitutes a Wide Area Network (WAN), it helps to understand what constitutes a Local Area Network (LAN). A LAN is traditionally considered as a network that is confined to a small, local area, like a building. A WAN therefore is a network that extends far beyond a single location and usually involves connecting devices across wide geographical boundaries, like different cities or even different countries. Certain technologies enable these connections, and they're often sold or leased to consumers by service providers, like AT&T, Verizon, Sprint, among others. These providers have engineered WAN packet-switching networks that enable connectivity between locations, for example, a LAN in Texas to a LAN in Washington, DC, and they charge for things like data and bandwidth needs as well as for varying levels of quality of service (QOS). As the name implies, QOS focuses on error handling and the ability to provide services like IP convergence.

Know common WAN protocols and key features of each

Of course, these technologies have evolved over the years. Original WAN technologies connected devices across telecommunications networks—PSTN—using modems. Satellite and microwave-based technologies were also quite prevalent years ago. Today, however, data packet carrier services are most often used. These services include things like Frame Relay, ATM, and MPLS, and are procured through service providers as noted above.

Table 4-34 outlines some of the key evolutions of WAN technologies.

X.25	X.25 is one of the pioneers of WAN packet-switching protocols. In fact, it's still highly regarded for its error correction capabilities. At the same time, X.25 is not efficient and creates high overhead.
Frame Relay	Frame Relay followed X.25, and unlike X.25's focus on quality of service and error correction, Frame Relay's focus is on speed of transmission. In the context of Frame Relay, two types of circuits can be found: permanent virtual circuits (PVCs) and switched virtual circuits (SVCs). PVCs support end-to-end links over a physical network, and SVCs are similar to circuit-switched networks, like PSTN.
Asynchronous Transfer Mode (ATM)	ATM builds on many of the best features of Frame Relay and can support very high-speed transmission needs. ATM is connection-oriented and works through the creation of a virtual circuit between endpoints prior to data being exchanged. ATM virtual circuits can be permanent or on-demand.
Multi-Protocol Label Switching (MPLS)	MPLS is at the forefront of current WAN connectivity solutions, because it offers network connectivity that includes built-in security. Using features like forwarding tables and labeling schemes, MPLS providers can guarantee that customer's data transmissions cannot be touched by anybody else while traversing the MPLS provider network. However, even with this guarantee, the provider networks are still untrusted networks, right? Just because AT&T can guarantee that an organization's data transmissions will be safe while traveling across AT&T's network, somebody within AT&T could still access and read those transmissions. It's their network, so they can potentially access all data traveling within it. This explains why many organizations, even when using MPLS technology, will still encrypt their data—to prevent potential data snooping from the MPLS provider itself. Organizations should decide whether or

not to encrypt their data based on the value and sensitivity of the data.

Table 4-34: WAN Technologies

4.2 Secure network components

4.2.1 Network Architecture

CORE CONCEPTS

- Network architecture includes employing concepts such as defense in depth, partitioning, a well-protected network perimeter, network segmentation, and bastion hosts.
- Partitioning and network segmentation refer to the same concept where visibility of certain network traffic is limited through the use of networking devices, like switches and routers, and firewalls.
- Proxies are devices that act on behalf of someone else.
- NAT and PAT are examples of proxies that facilitate wide scale access to the Internet and serve as a layer of protection against eavesdropping of internal network structures.

Understand what is meant by the term “network architecture” and key elements of a good network architecture

Good network architecture can provide many benefits and help prevent and protect an organization if attacked by malicious outsiders or insiders. Key elements of network architecture include the concepts noted below.

Defense in Depth

The concept of defense in depth refers to combining multiple layers of security controls to protect a network. Think of defense in depth as multiple layers of circles, with items like policies and procedures being the outermost layer, environmental considerations being the next, physical infrastructure next, followed by operating systems, and finally software configurations. Put another way, the outermost layer consists of people and processes, then architecture controls, then cabling and switching, and finally operating system controls and things like firewall configurations.

Partitioning

Partitioning is the practice of controlling the flow of traffic between segments. Refer to the earlier network topology example of a bus (depicted in [Figure 4-2](#)), where all devices are connected to the same wire—each device can see all the traffic traveling across the wire.

To protect this traffic, partitioning—also known as network segmentation—can be utilized. Some areas of a network might require a higher level of security than others. Partitioning can be used to prevent traffic from those areas from being seen across the entire network. Switches, routers, and firewalls can all be used to create partitions, or segments, and then rules implemented to control the flow of traffic between segments.

By far, the most important partition to create is the one that separates an organization network from externally connected networks. The best example of an external network is the internet. Organizations do want, and often need, to be connected to the internet, but they also want to block certain network traffic—in both directions. Incoming traffic should certainly be scrutinized and confirmed to be legitimate, but outgoing traffic should also be scrutinized for purposes of things like data loss prevention and identification of any malicious traffic destined to the outside world. So, firewalls and devices that sit between the main network and the internet can enforce rules that look at incoming and outgoing traffic.

Network Perimeter

The network perimeter is the last point any organization can control. Like physical security, where controls should comprise preventive, detective, and corrective capabilities, the same should hold true for the network. Controls at the perimeter should comprise preventive, detective, and corrective capabilities. Similar to a physical perimeter, where the ideal number of entrance and exit points is exactly zero, a network perimeter should be equally hard to breach. Ideally, only one entrance and exit point should exist. Otherwise, like a building with multiple entrances and exits, monitoring the flow of traffic can be extremely difficult. Limiting the ingress and egress point of a network to one creates a **choke point**—a point where devices and technologies that enforce rules can be placed to ensure all incoming and outgoing traffic is analyzed. [Figure 4-](#)

28 depicts two choke points in a simple network, the first between the public network and the non-sensitive private network, and a second that any traffic must pass through to access the sensitive private network.

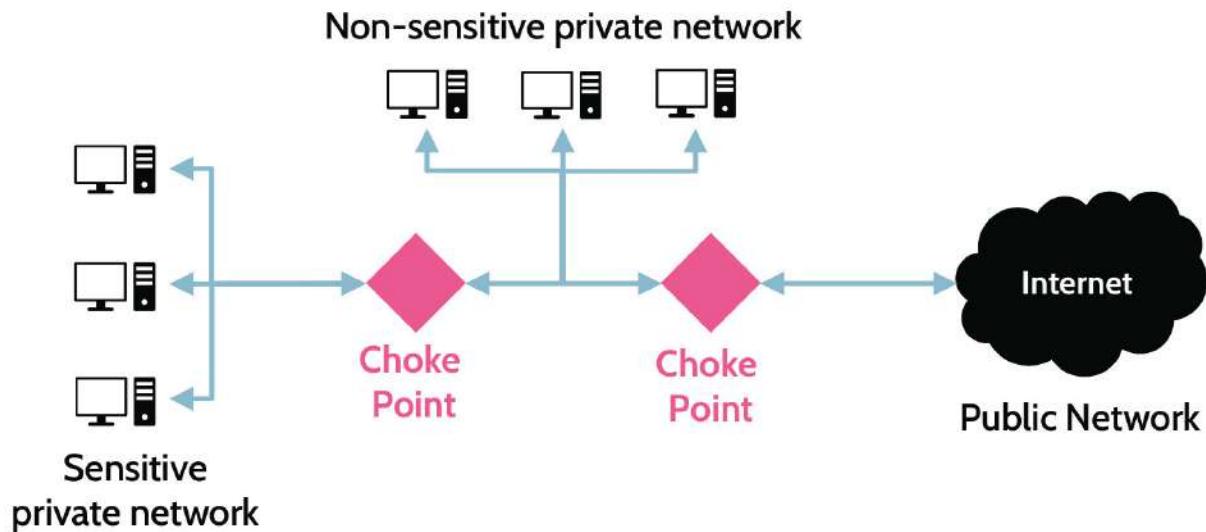


Figure 4-28: **Network Choke Points**

Network Segmentation

Referring to Figure 4-28, consider the public network to be the internet and the devices to the left to be an internal network (named sensitive private network). Access to the internet is important for so many reasons—to connect with clients, business partners, to market products via a website, and to exchange email, just to name a few. To make these connections and accomplish these goals, specific applications hosted on specific devices would need to be in place. For instance, if customers want to order products, an e-commerce application would need to exist. If anybody wants to send or

receive an email, an email application would need to be running. With this in mind, does it make sense for these applications to be hosted on the internal network? From a security perspective, this would be very unwise, because people from the public network—from the internet—would then be on the inside, and everything would be at risk.

Bastion Host

Understand what is meant by the term “bastion host” and where a bastion host might typically be found in a network architecture

This risk can be mitigated through the creation of a subnetwork, usually referred to as a **Demilitarized Zone (DMZ)**, where services and applications that require public access can be segregated. The DMZ is not part of the internal network nor is it part of the internet, it sits in between the two and it can be controlled by the organization. As alluded to above, any service or application that requires access from the outside—like web applications, email, DNS, and remote access—can be placed in the DMZ. Because the organization controls the DMZ, it can also provide necessary protection for each application. In this context, devices and applications within a DMZ are often referred to as **bastion hosts** and bastion applications. The word *bastion* is French and loosely translates to “fortress,” readied and strengthened for attack. Within the DMZ, hosts and applications that have been hardened and strengthened to

protect against exploits and attacks are referred to as bastions. Between the DMZ and the Internet is a boundary router. The simplest form of a firewall is a router which sits between two networks and controls the flow of traffic by analyzing each packet header for source and destination IP addresses and ports. A router used in this role is often referred to as a **boundary router**, as shown in [Figure 4-29](#).

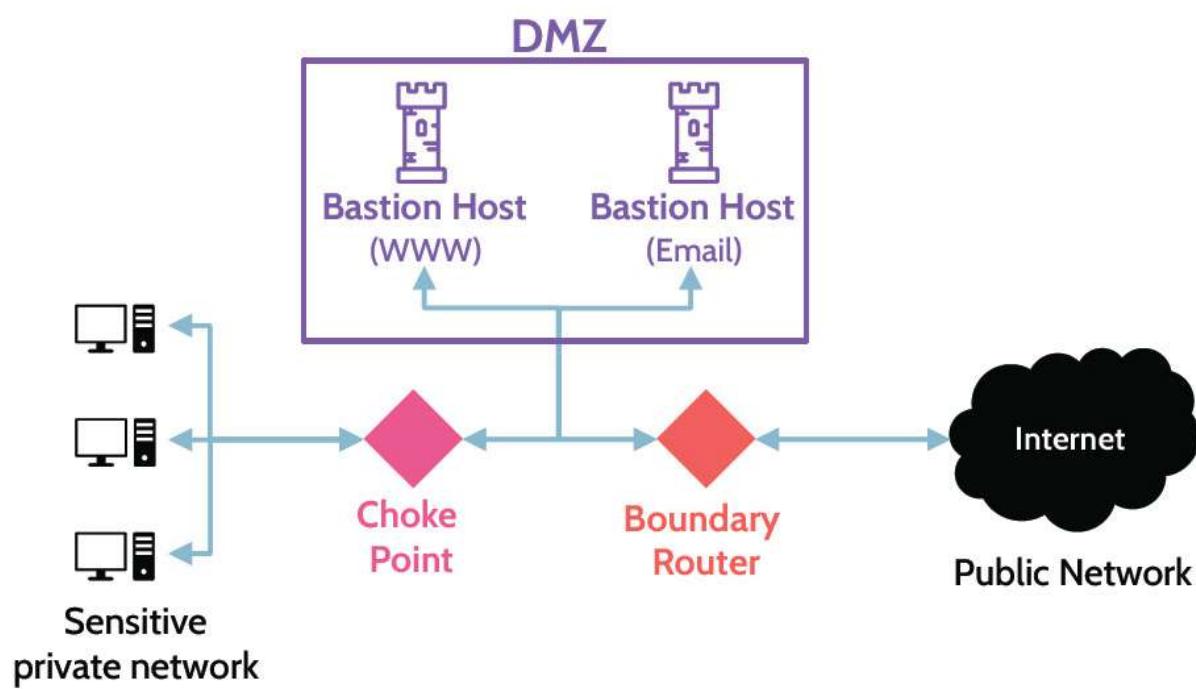


Figure 4-29: DMZ and Bastion Hosts

Microsegmentation

We often use virtualization to quickly and cheaply deploy logical networks. In contrast to using physical devices, the cost-effectiveness of virtualized networks allows us to segment our networks at a much more granular level. We refer to this as **microsegmentation**.

Let's give you a more concrete example to demonstrate just how advantageous this can be. First, let's say your organization has a traditional network. You would have the insecure Internet, a physical firewall, and then the DMZ, where you would have things like your web server, your FTP server and your mail server, as shown in [Figure 4-30](#). Under this setup, your firewall rules would need to be fairly loose to allow the web traffic, the FTP traffic and the SMTP traffic through to each of your servers. The downside of this configuration is that if the web server was compromised by an attacker, this would give them a foothold in your network that they could use to access your FTP server or your mail server. This is because all of these servers are on the same network segment.

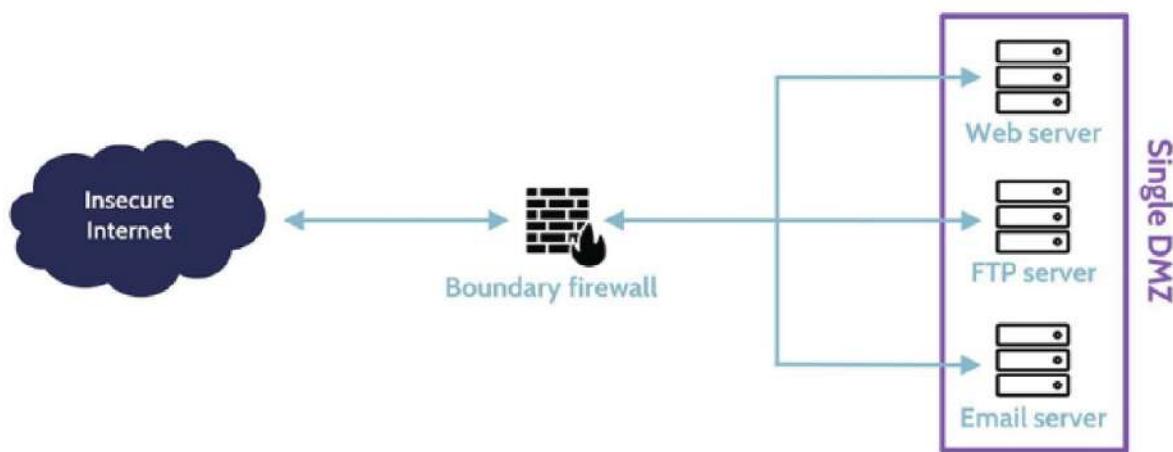


Figure 4-30: A Traditional Network

In contrast to a physical approach, **virtualized networks allow you to deploy virtual firewalls easily and at low cost**. You can easily put virtual firewalls in front of each server, creating three separate DMZs, as shown in [Figure 4-31](#). You could have much

tighter rules on the firewalls for each of these network segments because the firewall in front of your web server would only need to let through web traffic, the firewall in front of your FTP server would only need to let through FTP traffic, and so on.

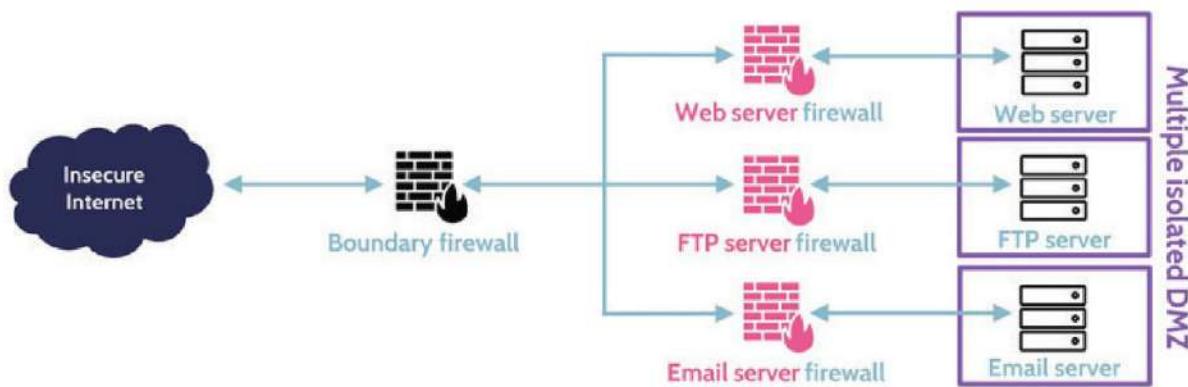


Figure 4-31: An Example of Microsegmentation

The benefit of having these virtualized segments with their own firewalls is that the much stricter rules limit the opportunities for malicious traffic to get through. In addition, if an attacker does manage to get a foothold on one of your servers, such as your web server, they would not be able to move laterally as easily. They would still need to get through the other firewalls if they wanted to reach your FTP or mail servers.

As this example has demonstrated, virtual networks allow for microsegmentation, sometimes with only a single asset in the network segment. Not only does this make it substantially more difficult for hackers to move across the network, but it

also allows you to set granular and precise firewall rules for each network segment. This compartmentalization is a significant security advantage.

Some technologies that we use alongside microsegmentation are outlined in [Table 4-35](#).

Network overlays/encapsulation	The creation of virtual networks that are abstracted or overlaid on top of the physical network,
Distributed firewalls	As we outlined in our example, microsegmentation can be used to easily and cheaply deploy multiple firewalls, allowing for much more granular firewall rules.
Distributed routers	Similar to distributed firewalls, distributed routers allow us to distribute routing rules to individual workloads as opposed to having one central router.
IDS/IPS	Intrusion detection systems and intrusion prevention systems can be strategically deployed to protect individual workloads or network segments. We discuss IDS/IPS in more detail in section 4.2.4 .
Zero trust architecture	Microsegmentation allows us to implement granular trust zones. We discuss ZTA in section 3.1.2 .

Table 4-35: Technologies Used Alongside Microsegmentation.

Proxy

One way to provide strong security across a network is through the utilization of what are known as proxies. A proxy is a device that acts on behalf of something else, commonly a user or

application. In the context of a network, as shown in [Figure 4-32](#), a proxy helps facilitate the connection between a client and a server, because it is better equipped to manage and direct outgoing and incoming traffic. A *proxy or proxy server is an intelligent application or hardware that acts as an intermediary and is placed between clients and a server*. As a result of the inherent intelligent nature of proxies, they're usually found at Layer 7—the Application layer—of the OSI model. In [Figure 4-32](#), the client perceives the connection as being direct to the server, though the server perceives otherwise—the connection is from the server to the proxy. In reality, the actual connection in both cases is from the client to the proxy and from the server to the proxy. All decision requests are routed through the proxy, which has the intelligence and ability to make decisions, enforce rules, and otherwise manage requests.

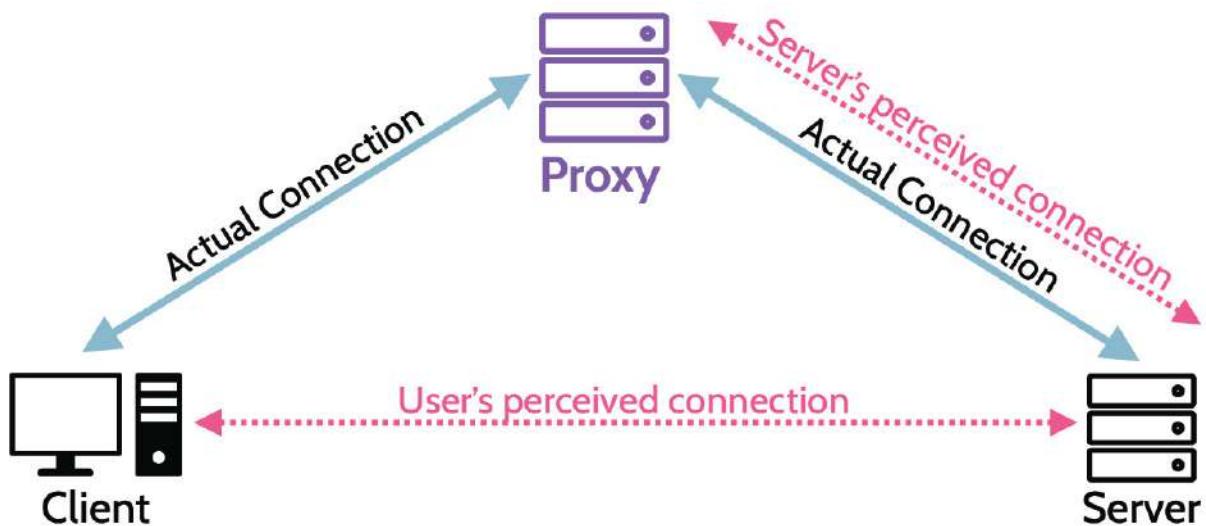


Figure 4-32: **Proxy Use**

This provides enhanced security because devices like web proxies are used to filter requests and discard any traffic that resolves to a known malicious destination. That can help keep the environment secure because if that proxy wasn't there to perform filtering, the user would have easily navigated to a malicious domain.

NAT and PAT

Understand NAT and PAT

An example of NAT was already mentioned in [section 4.1.6](#), where the home router scenario was discussed. As a refresher, NAT is the mechanism that allows us to translate private IP addresses to public ones and vice versa. PAT is another mechanism that can be used, which helps us perform port translation, in the same notion as IP address translation is performed. At the same time, NAT and PAT provide a layer of security to organizations by masking internal networking schemes, which can hinder reconnaissance efforts of attackers. NAT and PAT are summarized in [Table 4-36](#).

Network Address Translation (NAT)	Port Address Translation (PAT)
In its simplest form, NAT changes IP addresses to other IP addresses. For example, a private, internal, and non-routable IP address to a routable IP address.	PAT is an inherent part of NAT, and it helps keep track of individual internet requests using unique port assignments for each request.

Table 4-36: NAT and PAT

An example of NAT/PAT operation is depicted in [Figure 4-33](#).

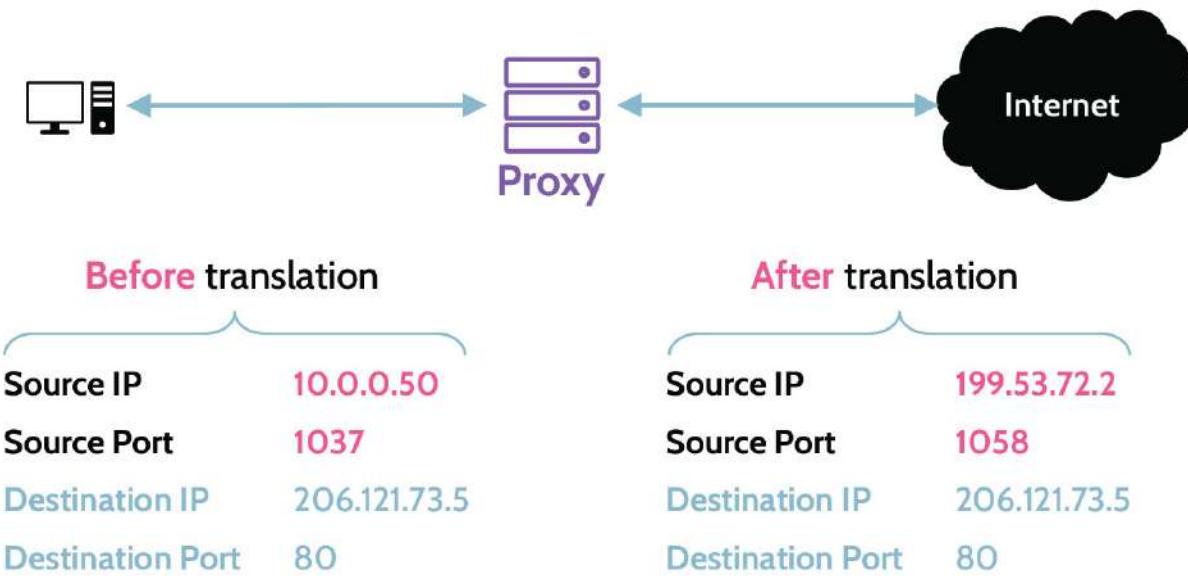


Figure 4-33: NAT/PAT Operation

It's easy to see that NAT is used to change the source IP address from 10.0.0.50 (a private, non-routable address) to 199.53.72.2, which is a publicly routable IP address that allows access to internet-based resources. At the same time, PAT is used to change the source port from 1037 to 1058, which is a unique port assigned only to the connection in this example. If other connection requests existed, the source IP after translation would likely be the same, but the source port would be different—it would be unique for each connection. Because of this uniqueness, when request results are returned, the proxy server can easily track and route those results to the device from which the original request was made.

4.2.2 Firewall Technologies

CORE CONCEPTS

- A firewall is a concept that enforces security rules between two or more networks.
- Firewall technologies include: simple packet filtering firewalls, stateful packet filtering firewalls, circuit-level proxy firewalls, and application-level firewalls.
- Simple and stateful packet filtering firewalls operate at Layer 3
- Circuit proxy firewalls operate at Layer 5
- Application proxy firewalls operate at Layer 7

What is a firewall?

Understand the different firewall technologies and pros/cons of each

A firewall is a concept that enforces security rules between two or more networks by performing traffic filtering. A firewall could be as simple as a router or as complex as a set of applications that work together to protect a network. In network security, a firewall either allows or blocks traffic, based upon predefined rules. Firewalls are typically found between an internal network and the Internet, but they're also frequently used internally to protect different network segments from each other and they are preventive controls.

Firewall Technologies

Table 4-37 summarizes the different firewall technologies in use today.

Packet Filtering	<ul style="list-style-type: none">■ Examines packet headers to either block or pass packets■ Uses access control lists (ACLs) that allow it to accept or deny access
Stateful Packet Filtering	<ul style="list-style-type: none">■ State and context data are stored and updated dynamically■ Provides information for tracking connectionless protocols; e.g., Remote Procedure Call (RPC) and UDP-based applications where source/destination ports and IP addresses are used to track state
Circuit-Level Proxy	<ul style="list-style-type: none">■ Create a circuit between client and server without requiring knowledge about the service■ Have no application-specific controls■ An example is a SOCKS server
Application-Level Proxy	<ul style="list-style-type: none">■ Able to inspect packet payload■ A different proxy is needed for each service■ Can be a performance bottleneck

Table 4-37: Firewall Technologies

Understand where different firewall technologies are found in the OSI model and implications of the same

Additionally, with the functionality described above, it makes sense that firewall technologies exist at different layers of the

OSI model. Recall from earlier discussion that efficiency and intelligence vary at each layer of the OSI model. At the lower layers, intelligence is very low, but efficiency is very high; at the higher layers, intelligence is very high, but efficiency is much lower. [Table 4-38](#) breaks down where firewalls live in the OSI model and key characteristics of each firewall technology.

Simple Packet Filtering	Stateful Packet Filtering	Circuit Proxy	Application Proxy
OSI Layer 3 Network	OSI Layer 3 & 4	OSI Layer 5 Session	OSI Layer 7 Application
Simplest	Complex	More Complex	Very Complex
Fastest	Fast	Higher latency	Highest latency
Filters based on source and destination IP address and port	Maintains state table and filters based on pattern matching	Filters sessions based on rules	Filters based on data (Payload)

Table 4-38: Firewalls and OSI Layers

Context-Based Access Control (CBAC)

CBAC is a feature of firewall software that intelligently filters TCP and UDP packets based on application layer protocol session information. It allows for deep traffic inspection and filtering to take place, that is, it's able to detect potential DDoS attacks or provide advanced statistics about the various connections and protocols used.

4.2.3 Firewall Architectures

CORE CONCEPTS

- Firewall architectures can vary, based upon multiple factors and needs.
- Common firewall architectures include: packet filtering, dual-homed host, screened host, screened subnet, three-legged firewall.
- Firewall architectures should always directly reflect the security needs of the organization.

With advancements in firewall technologies, firewall architectures have also evolved and become more sophisticated. Firewall architectures specifically focus on how firewall technologies can best be utilized and deployed to provide robust security for an organization.

Packet Filtering

A packet filtering firewall architecture, depicted in [Figure 4-34](#), is the simplest one. The internal network is represented by the computer icons, and a simple router is placed between the internal network and the internet or another untrusted network. The router, which operates at Layer 3, can only make decisions based upon information that exists at Layer 3—the header portion of the packet, which contains information like source IP, destination IP, service being requested, and so on. Thus, decision-making capabilities are very limited, but efficiency and speed of decision-making is very high.

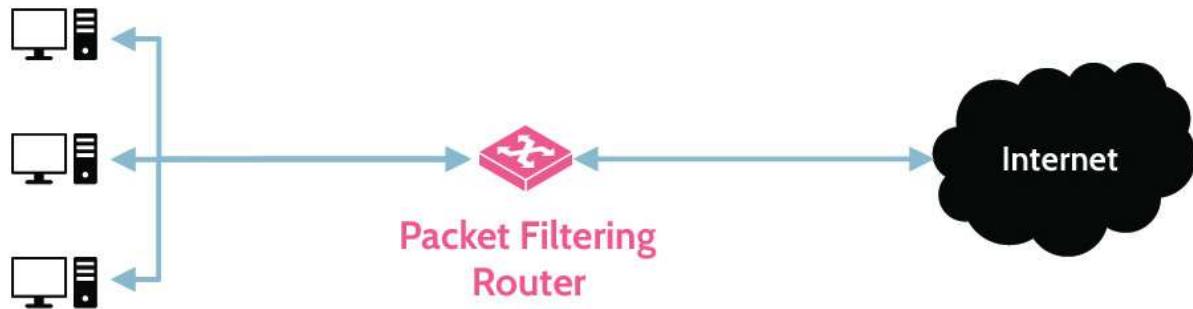


Figure 4-34: **Packet Filtering Firewall**

Dual-Homed Host

A dual-homed host improves upon a packet filtering router by replacing it with a more intelligent computer or host that contains two network cards. The host can understand *all* layers of the OSI model and can therefore support all types of firewall technologies described above. As a result, the host can make simple to very complex decisions. This architecture is depicted in [Figure 4-35](#).

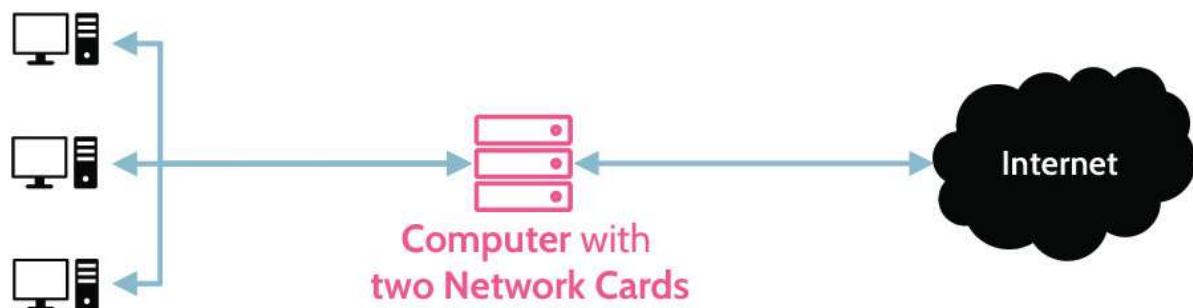


Figure 4-35: **Dual-Homed Host**

Screened Host

By combining the architectural elements of a packet filtering and dual-homed host firewall, a screened host firewall architecture results as shown in [Figure 4-36](#). The router can handle the first level of decision-making related to incoming packets, and any packets that are allowed through can then be further examined by the bastion host, which can be any type of firewall technology. In addition to providing initial filtering services, the packet filtering router can act as a screening device for the bastion host. In other words, before an attacker could target the bastion host, they would first need to compromise the router. This is why this type of firewall architecture is known as a screened host firewall architecture.

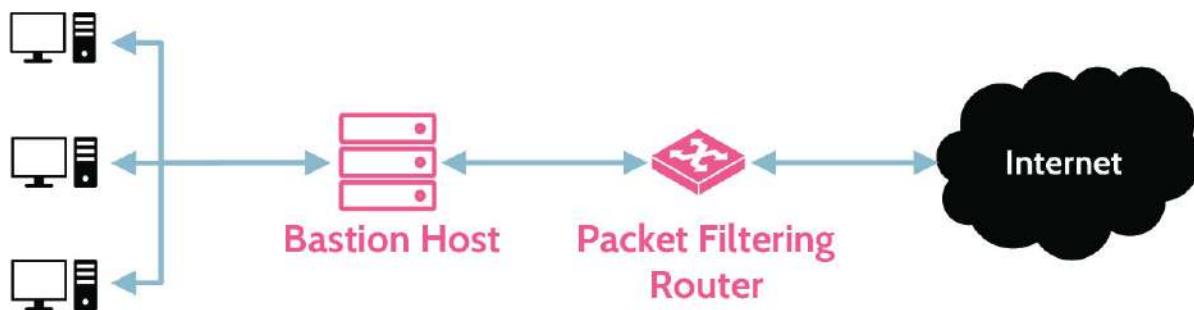


Figure 4-36: Screened Host

Screened Subnet

[Figure 4-37](#) shows a screened subnet architecture, where two firewalls are used and between them a subnet, such as a DMZ, can be created. Traffic from the outside can be specifically directed to the DMZ and thereby protect the internal network

from potential attacks. Screened subnet architectures are expensive as two firewalls are required; however, a potential advantage is that the two firewalls could be purchased from different vendors. Thus, if a vulnerability is discovered in one firewall, the same vulnerability is unlikely to exist in another vendor's firewall.

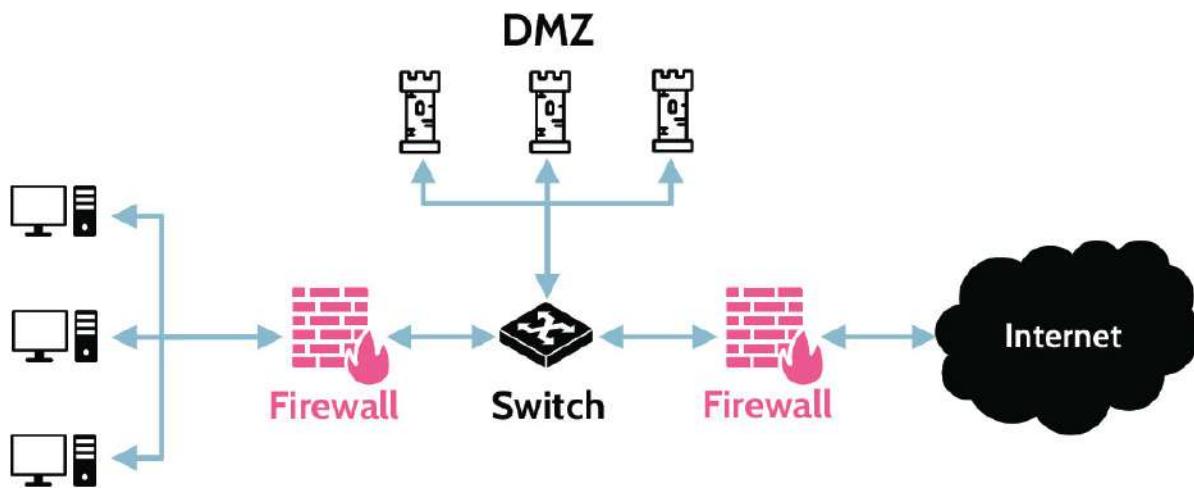


Figure 4-37: **Screened Subnet**

Three-Legged Firewall

Figure 4-38 depicts a three-legged firewall (or three-zoned firewall), by virtue of the three connection points, although any number of connection points could really exist. This is dependent on the needs and creativity of the organization. Any number of firewall technologies can be utilized in the context of this type of firewall, and the choices made in this regard should directly reflect the security needs of the organization and the assets being protected.

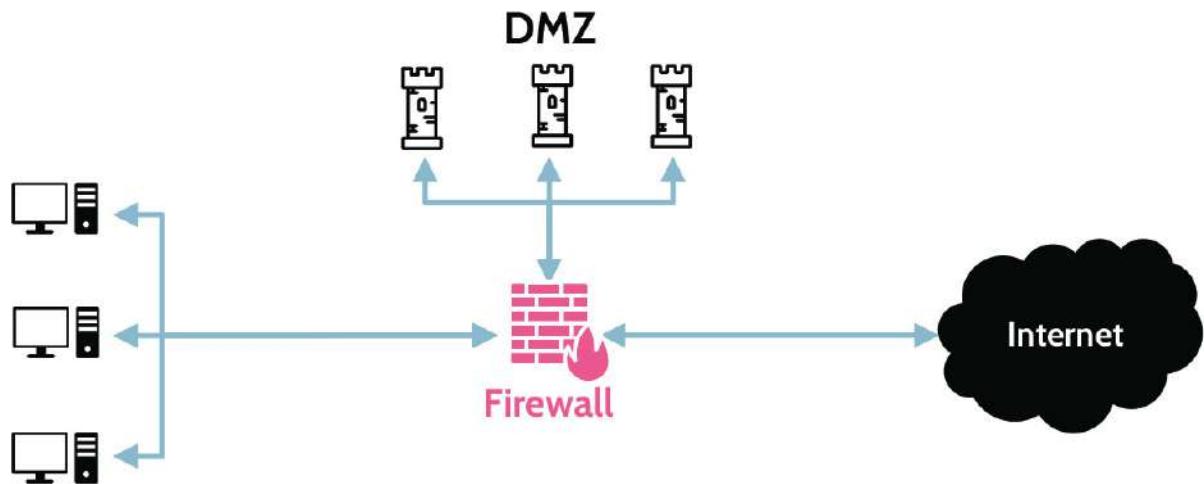


Figure 4-38: Three-Legged Firewall

4.2.4 IDS and IPS

CORE CONCEPTS

- Data inspection involves monitoring and examining data.
- Intrusion Detection System (IDS) performs data inspection and detects, logs, reports, and sometimes triggers other devices to take corrective action.
- Intrusion Prevention System (IPS) performs data inspection and additionally prevents or takes corrective action.
- Two types of IDS/IPS systems: network-based and host-based.
- Mirror/span/promiscuous port refers to setting a specific port on a network device (e.g., switch) to receive ALL traffic transiting the network device for monitoring purposes.
- Two IDS/IPS detection methods: pattern and anomaly.
- Ingress and egress monitoring refers to the specific flow of traffic; ingress = incoming traffic, and egress = outgoing traffic.
- Whitelisting and blacklisting: whitelisting refers to specifically allowed IP addresses—all others are blocked; blacklisting refers to specifically

blocked IP addresses—all other IP addresses are allowed by default.

Considering that firewalls are preventive controls, but a complete control encompasses prevention, detection, and correction, it follows that other systems should be in place and work with firewalls to provide the most protection to an organization. Specifically, these systems should provide detection and correction capabilities, and at the network level this involves data inspection.

Data Inspection

At a high level, data inspection involves monitoring and examining transmitted data and taking appropriate action if not allowed by security rules. Drilling down a bit, data inspection includes the activities summarized in [Table 4-39](#).

Virus scanning	Files are scanned against known signatures for malware.
Stateful inspection	Dynamic state/context table is maintained to track and analyze communications between systems.
Content inspection	Content of mobile code is scanned and inspected for compliance with specific security rules.

Table 4-39: Data Inspection Activities

**Understand the similarities and differences between
IDS/IPS**

IDS and IPS

Drilling even further, data inspection can be accomplished using devices created specifically to examine the header and data (or payload) portions of a packet.

An **Intrusion Detection System (IDS)** is just as the name suggests—a detection system. An IDS examines traffic at the network level or the host level, specifically looking for malicious activity, policy violations, or other signs of suspicious activity and can alert and/or log those events. An IDS does not take direct action against malicious activity; however, the activity could trigger the corrective steps that need to happen, for example, by being tied back to a firewall, the firewall could take corrective action like filtering packets or blocking IP addresses.

An **Intrusion Prevention System (IPS)**, on the other hand, can prevent, detect, and take corrective action when necessary. When an offending pattern is identified, the IPS can, for example, block the source IP address or terminate a connection.

IDS and IPS systems are summarized in [Table 4-40](#).

Intrusion Detection System (IDS)	Intrusion Prevention System (IPS)
Monitors a network or host for malicious activity or policy violations and reports events .	Monitors a network or host for malicious activity or policy violations, reports events , and attempts to block .

Table 4-40: **IDS and IPS**

Network-Based versus Host-Based

There are two types of IDS and IPS systems: network-based and host-based.

Understand the pros and cons of networkbased vs. host-based IDS/IPS

Network-based IDS/IPS require strategically placed sensors across a network and monitor network traffic to ensure that rules are applied. Host-based IDS/IPS run as agents on specific devices, like servers and other mission-critical systems, and do the same thing. From the perspective of which type of device provides the best protection, a combination of both types offers the best means of protection, detection, and correction for a network and associated critical systems. Network-based versus host-based IDS/IPS systems are summarized in [Table 4-41](#).

Network-Based IDS/IPS	Host-Based IDS/IPS
Monitors network traffic transiting a network segment .	Installed on a host (server) and monitors only that host .

Table 4-41: Network-Based vs. Host-Based IDS/IPS

IDS/IPS Network Architecture

Understand typical placement of IDS and IPS

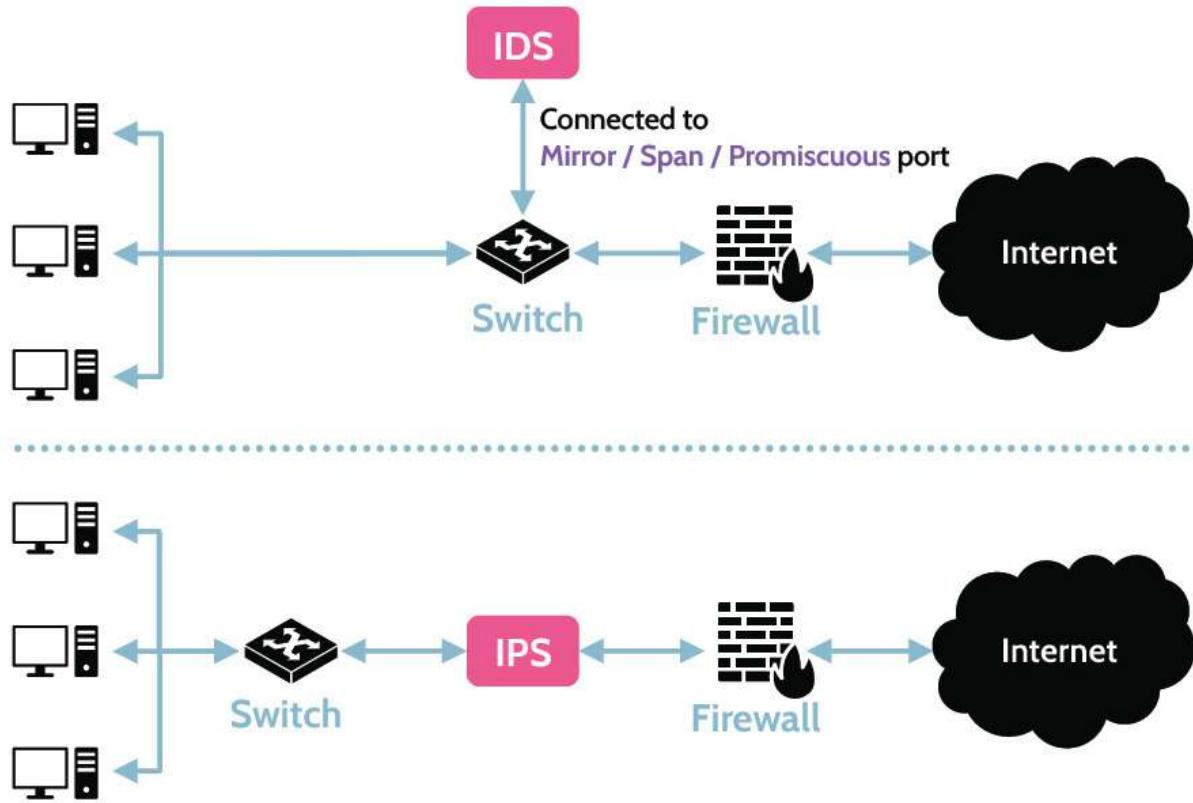


Figure 4-39: **IDS/IPS Architecture**

The sole purpose of an IDS is to detect, and to gain the benefits of a complete control, the IDS would need to be coupled with something else that could provide the additional capabilities of prevention and correction. As shown in the top half of [Figure 4-39](#), an IDS is connected to a network via a mirror or span port (also known as promiscuous port). This allows the IDS to get a copy of all network traffic. If potentially malicious traffic is identified, the IDS can communicate with the firewall, which can then take preventive and corrective actions.

In the lower half of [Figure 4-39](#) the IPS is placed in line with the network traffic, because it has the capability to detect, prevent,

and correct. As traffic comes into the network, it passes through the IPS. If a rule is triggered for malicious activity, the IPS can act and prevent that from traversing the rest of the network.

IDS & IPS devices can potentially be placed in numerous locations across a network depending on where detection and correction capabilities are desired. Typically, an IDS/IPS device needs to be placed in each network segment to be monitored. The pink boxes in [Figure 4-40](#) very roughly depict some of the potential IDS/IPS locations in a basic network architecture. The exact placement and selection of IDS or IPS devices is a complicated topic beyond the scope of the CISSP exam.

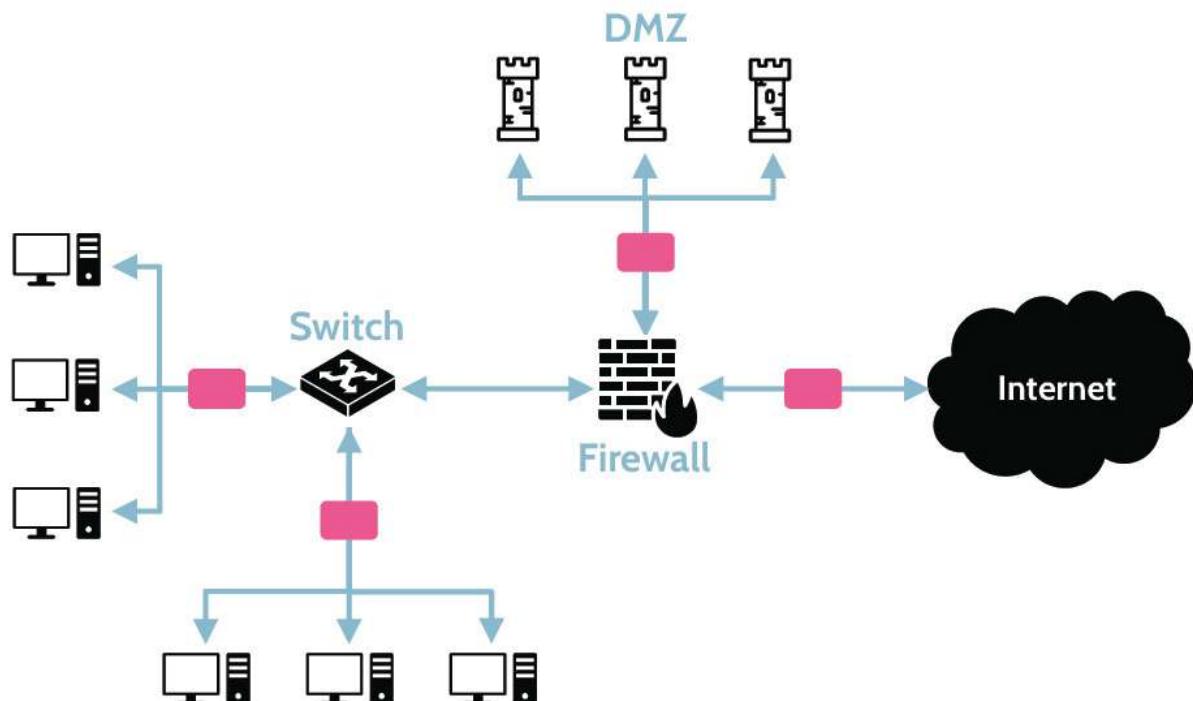


Figure 4-40: The Need for IDS/IPS

Mirror/Span/Promiscuous Port

Understand what is meant by any term that uses the words *mirror*, *span*, or *promiscuous* with the word *port*

When a port on a network device (e.g., switch) is described as **mirror**, **span**, or **promiscuous** it's meant that traffic passing through that device is copied to that port and any device connected to it, like an IDS, can obtain a copy of it for inspection. When you install Wireshark (one of the most common packet analyzers in the industry), at some point of the installation process it warns you that the network card will now be set in promiscuous mode so that the machine can capture surrounding network traffic.

Setting a device port to this mode allows broadcasts from any device on the network to be monitored, which can then allow the IDS/IPS to work most effectively.

IDS/IPS Detection Methods

Understand the two types of IDS/IPS analysis engines and how each works

Two types of analysis engines are used in IDS and IPS. These are signature-based and anomaly-based, each of which are discussed in [Table 4-42](#). Signature-based engines focus on

known types of attacks and use this information to build a database of patterns for detection purposes. Anomaly-based engines look for unusual, abnormal, and out-of-the-ordinary patterns, which implies they understand what is normal and predictable. Anomaly-based engines therefore go through a learning process that involves monitoring user activity and understanding what is considered normal, and anything that is out of the ordinary will be triggered as an anomaly.

Signature-based	<p>Signature-based detection involves looking for known signatures of malicious activity, such as file hashes, suspicious email subject lines, malicious IP addresses or byte sequences. These signatures are added into the analysis engine. When an IDS/IPS sees one of these signatures in the network traffic, it can send an alert, which makes signature-based detection useful for finding known threats. However, signature-based detection systems have no hope of detecting the latest threats if they don't have the signatures for them. This leaves us vulnerable to the latest attacks.</p>
Anomaly-based	<p>Anomaly-based detection seeks to complement signature-based detection by first creating a baseline of normal behavior within the system, and then sending alerts when it detects anomalous or suspicious behavior. The major downsides are that it is computationally expensive and it can generate a lot of false positives.</p> <p>There are four major ways to detect anomalies:</p> <ul style="list-style-type: none">■ Stateful matching – If the anomaly-based IDS/IPS detects communications that don't align with the expected stream of traffic (the state) it can block them.■ Statistical anomalies – This involves looking for a statistically significant deviations from the norm. These deviations can result in alerts or be blocked.

■ **Traffic anomalies** – These are differences from the normal flow of traffic. When the IDS/IPS detects them, they can trigger an alert or block them.

■ **Protocol anomalies** – If traffic from new protocols starts to appear, an anomaly-based IDS/IPS can send alerts or block it.

While anomaly-based detection isn't perfect, it works under the assumption that attacker traffic will often look quite different to normal traffic. **Anomaly-based detection can help us to detect threats even if it's a new attack and we don't already have signatures for it.**

Table 4-42: IDS/IPS Detection Technologies

Ingress and Egress Monitoring

The terms *ingress* and *egress* refer to the direction of flow or, in this case, network traffic, as shown in [Figure 4-41](#). Ingress is the act of going in or entering; egress is the act of going out or exiting. It follows that **ingress monitoring** is monitoring all traffic entering a network; **egress monitoring** is monitoring all traffic exiting a network. The best protection involves both types of monitoring. Ingress monitoring can help prevent malicious traffic from entering a network; egress monitoring can help prevent data loss, denial-of-service, and other types of malicious activity from originating from the corporate environment. It's important that IDS/IPS monitor traffic in both directions.

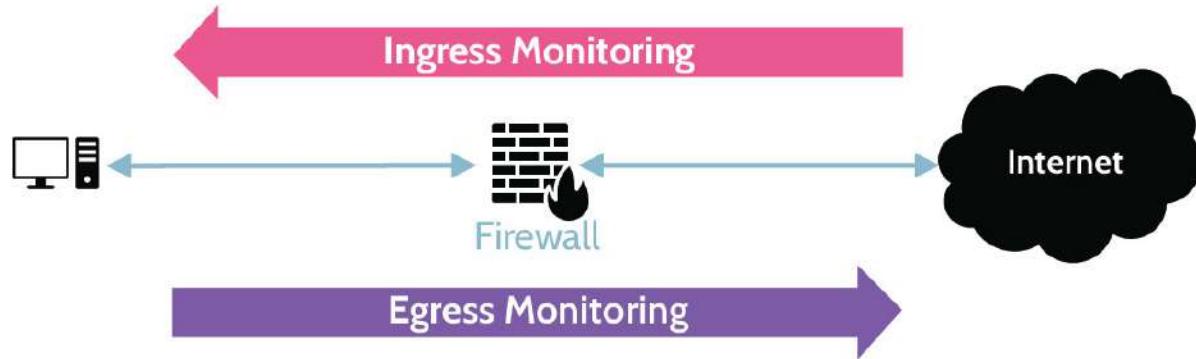


Figure 4-41: Ingress and Egress Monitoring

Allow List and Deny List (Whitelisting and Blacklisting)

Understand the difference between the terms “allow list” (white list) and “deny list” (black list)

A technique that IDS/IPS devices can use to detect and potentially block suspicious traffic is allow lists and deny lists.

Allow and deny lists are lists of IP addresses and specifically determine what action may or may not be performed with respect to the IP addresses in a given list. With an allow list, network traffic to listed IP addresses is allowed. Any other IP address is blocked by default. Deny lists are the exact opposite; any traffic to listed IP addresses is specifically blocked. Any other IP address is allowed by default.

Note, that “allow lists” and “deny lists” are much better terminology, and these terms are becoming more pervasively used in the industry. However, you may still see references to

whitelists and blacklists, including on the exam, which is why we have noted this terminology.

Allow lists and deny lists are summarized in [Table 4-43](#).

Allow List (Whitelist)	Deny List (Blacklist)
The following IPs may be visited. All other addresses are NOT permissible.	The following IPs may NOT be visited. Any other address is permissible.

Table 4-43: Allow List and Deny List

4.2.5 Sandbox

CORE CONCEPTS

- A sandbox is a safe area where untrusted code can be isolated and run.
- Four possible alert scenarios: true/false-positive, true/false-negative.
- False-negative is a worst-case scenario.

When an IDS/IPS identifies potentially malicious activity, something called a sandbox can be activated. A sandbox is a safe area, where unknown or untrusted code can be isolated and run and tested and determined to be malicious or not. Being able to run **potentially malicious software** in a sandbox environment is one of the corrective actions that an IDS/IPS can take. In addition, sandboxes are used quite often by malware analysts who run malicious code in them and try to identify indicators of compromise and gain an in-depth understanding of how malware operates.

Alert Statuses

Understand false-positives and false-negatives and the worst possible outcome

[Table 4-44](#) illustrates the possible conditions of alerts that may be received by security tools. These outcomes are quite important and can denote if tools are working optimally or if tuning is required to maximize efficacy.

1. **True Positive:** An attack is taking place and the security tool raises an alert to denote that fact. This indicates appropriate operation is in effect.
2. **True Negative:** No attack is present, and no alert is generated by a security tool. This indicates appropriate operation is in effect.
3. **False Positive:** An alert is generated by the security tool; however, there's no actual attack taking place (e.g., a suspicious login alert was generated for a user logging in from Colombia who had never logged in before from that location, but that person is legitimately there for work for the next three months). This is indicative of tuning required to eliminate unnecessary alerts from flooding the security team.

4. False Negative: An attack is ongoing, but the security tool failed to raise an alert. This is the worst scenario, since the security team isn't aware of malicious activity actually taking place in the environment. New rules being applied, policy redesigns, and new attack signatures are usually some of the things analysts use to resolve this issue.

	True	False
Positive	True-Positive Alarm is generated and attack is present	False-Positive Alarm is generated but no attack is taking place
Negative	True-Negative No alarm is generated and no attack is taking place	False-Negative No alarm is generated but an attack is actually taking place

Table 4-44: Possible Alert Statuses

Tuning our security tools is the process of adjusting their sensitivity. Ideally, we want as few false positives as possible, because we don't want our security teams to get overwhelmed with unnecessary alerts to the point that they become fatigued and end up ignoring a true positive. However, if we tune them too low, we run the risk of a false negative where an attack is actually occurring but the security team doesn't receive an alert. Appropriate tuning is a very delicate balance that will vary between organizations, their assets and the threats they face.

4.2.6 Honeypots and Honeynets

CORE CONCEPTS

- **Honeypots and honeynets are technical detective controls.**
- **Honeypots are individual computers and devices set up to appear as legitimate network resources.**
- **Honeynets are two or more networked honeypots.**
- **Honeypots and honeynets contain vulnerabilities that entice intruders into exploring further.**
- **Enticement is legal and pertains to situations where an intruder has already broken into a network.**
- **Entrapment is illegal and pertains to situations where somebody is persuaded to break into a network.**

What type of control do honeypots/honeynets represent?

Honeypots are individual computers (usually running a server OS posing as interesting targets for an attacker), but they contain no real data or value to the organization employing them. **Honeynets** are two or more honeypots networked together, and a sophisticated honeynet will also employ the use of routers, switches, or gateways. Honeypots and honeynets invariably always contain vulnerabilities—usually unpatched systems, applications, open ports or running services—that aim to entice potential attackers into exploring further. This exploration can then be detected. Honeypots and honeynets are usually located within the DMZ, or within a separate subnet and are usually built using virtualized systems.

Purpose of honeypots and honeynets

Honeypots and honeynets, depicted in [Figure 4-42](#), can serve a number of purposes, including:

- Detecting sophisticated cyberattacks, such as **Advanced Persistent Threats (APTs)**, where an attacker gains access to a network / system and stays there undetected for a long period of time. APTs are difficult to detect with traditional detective controls like an IDS system, as the attacker is very intentionally minimizing any activities that could be detected. For example, the attacker will use passive monitoring and data gathering techniques vs. actively scanning the network or systems.
- Helping trace how an attacker has traversed or moved through a network
- Distracting attackers away from valuable systems or resources
- Gathering valuable information that may help a security team better define their organization's security plan
- Conducting research by companies that serve the cybersecurity community

However, honeypots, and honeynets come with risks, including:

- Attackers might be able to leverage access to the honeypot or honeynet and gain access to real hosts or network resources, depending on the underlying architecture.
- If employed incorrectly, legal action against the company using the honeypot or honeynet could be the result, which is known as entrapment.

In either case, the ultimate responsibility for damages or monetary liability would rest upon the shoulders of senior management.

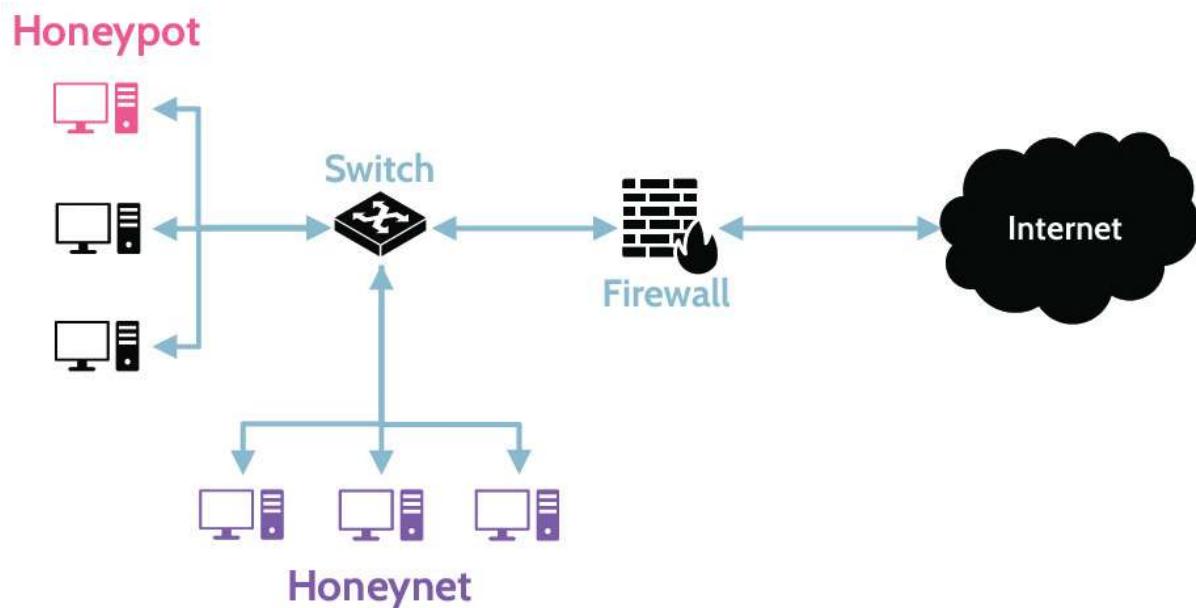


Figure 4-42: **Honeypot and Honeynet**

Enticement and Entrapment

Difference between enticement and entrapment

As noted above, when working with honeypots and honeynets, an organization must be careful to avoid entrapment of a potential attacker, which is illegal. [Table 4-45](#) provides the definitions of enticement and entrapment *in the context of honeypots*, including their core difference.

Enticement	Entrapment
Legal activity of persuading someone to commit a crime that they were already planning to commit.	Illegal activity of persuading someone to commit a crime that they would not otherwise have committed

Table 4-45: Enticement and Entrapment

4.2.7 Endpoint Security

CORE CONCEPTS

- **Endpoint security focuses on protection of devices found on corporate networks and seeks to minimize the attack surface and thereby prevent or minimize attacks.**
- **Network access control (NAC) solutions seek to unify endpoint security technology, user authentication, and overall network security.**

Primary focus of endpoint security and where commonly found

Endpoint security refers to the specific protection of client devices (endpoints) commonly found in corporate networks. Laptops, tablets, mobile devices, printers, IoT devices, and wireless devices are common examples of endpoints that are also potential attack points and paths to corporate networks. Endpoint security strives to minimize the attack surface and narrow the path to the corporate network as much as possible.

How network access control (NAC) solutions complement endpoint security

In the past, endpoint security was mainly accomplished through antivirus software. Today, endpoint security has evolved and become much more robust. In addition to antivirus software, many organizations deploy comprehensive device management policies and enforcement applications, endpoint data leak prevention (DLP), **Network Access Control (NAC)** solutions to restrict access (e.g., if an endpoint does not pass a health check) and platforms focused on endpoint threat detection, response, and monitoring.

4.3 Implement secure communication channels according to design

4.3.1 Tunneling and VPNs

CORE CONCEPTS

- **Remote access** is connecting to resources over an insecure network.
- **Tunneling** is the process of taking a packet and placing it inside the data portion of another packet.
- **Split tunneling** allows disparate remote resources to be used at the same time.
- **Generic Routing Encapsulation (GRE)** is a tunneling protocol that can be used to encapsulate a variety of protocols.
- **VPNs** are encrypted tunnels.
- **PPTP and L2TP** are Layer 2 tunneling protocols; L2TP is typically used in conjunction with Internet Protocol Security (IPsec).
- **SSL/TLS, SOCKS, and SSH** are tunneling protocols that include encryption capabilities.

Remote Access

Implementing secure communication channels is an important component of network security. In other words, putting protections in place to support network connections—especially remote access—is critical. Remote access is connecting to corporate resources over an insecure network (internet). Due to this fact, a method to protect traffic across that untrustworthy network must be utilized, and the best method is usually a VPN solution.

Tunneling

Before we can understand VPN solutions in depth, the concept of tunneling (depicted in [Figure 4-43](#)) must first be examined, because tunneling is the precursor to establishing a VPN. **VPN**

is tunneling plus encryption; without encryption, it can only be called a tunnel. Tunneling is simply the process of taking a packet and placing it inside the data portion of another packet. If the header and data portion of one packet are placed into the data portion of another packet, a tunnel is created. Some people also refer to this as encapsulation. The header and data from the one packet become the data portion of the other packet.

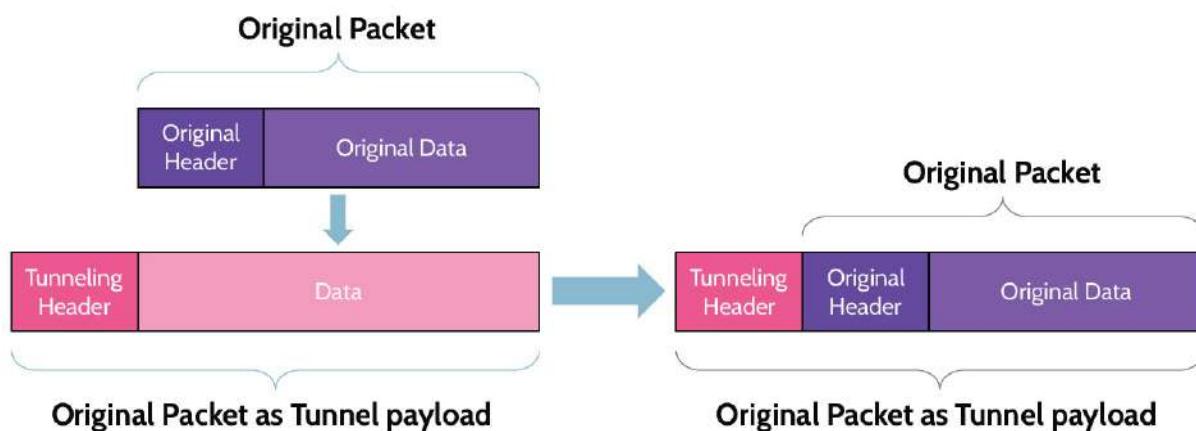


Figure 4-43: Tunneling Operation

This **encapsulation** process does not hide anything. It's simply placing the original packet inside the data portion of another one. If the information needs to be hidden, to be protected, encryption of the data portion of the new packet must take place. This results in tunneling plus encryption.

Why is tunneling used? The premise behind tunneling is that the encapsulated packet is being forced along the path specified in the header of the host packet. So, whatever path might be specified in the header of the encapsulated packet,

the header of the host packet will dictate the actual path both packets follow. However, the encapsulated packet can still be read because no encryption is used.

Tunneling protocols exist at several different layers of the OSI Model. This starts at Layer 2 and can go all the way up to Layer 7. At each layer, different functionality is available, and it's important to remember that the higher the layer, the bigger the trade-off between functionality and performance. At lower layers, performance is very efficient, but functionality is more limited. At higher layers, functionality is considerably enhanced, but performance is slower. Decisions about which tunneling protocol to use should take these factors into account.

Generic Routing Encapsulation (GRE)

Pros and cons of GRE and how it works

In a sense, Generic Routing Encapsulation (GRE), depicted in [Figure 4-44](#), is a tunneling protocol that can encapsulate a variety of protocols and route them over IP networks. GRE can transport traditional IPv4 traffic as well as multicast and IPv6 traffic, and it provides a means by which traffic can be exchanged between two networks using a network like the internet. GRE works by encapsulating the original packet—the payload that needs to be delivered—inside an outer packet. Once the receiving endpoint is reached, the GRE packet is removed and the original packet—the payload—is forwarded to

the ultimate destination. Unlike IPsec, which can use ESP to provide encryption and secure a payload, GRE is not considered a secure protocol, because it provides no encryption, just encapsulation.

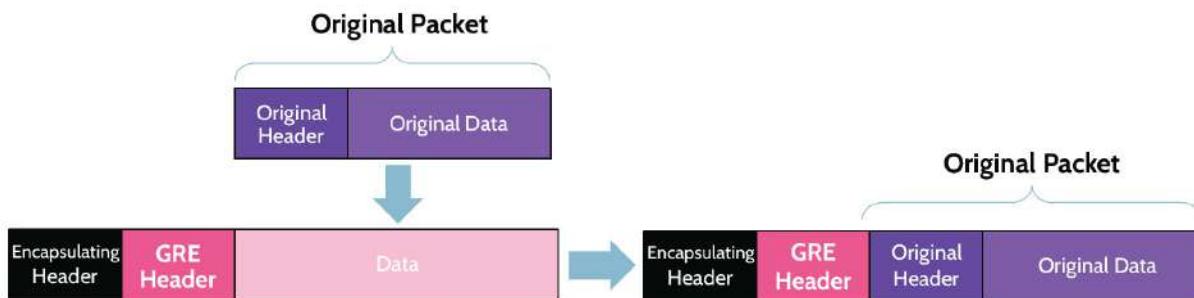


Figure 4-44: **GRE Tunneling**

Split Tunneling

Understand what a split tunnel is and what its inherent weaknesses are

Split tunneling allows a user to access disparate resources—the internet and a LAN, for example—at the same time, without all the traffic passing through the VPN. In the example shown in [Figure 4-45](#), in the upper half of the diagram, access to a corporate LAN can be achieved through use of a VPN (encrypted tunnel) established through a hotel network, while noncorporate resource access, illustrated in the lower half of the diagram, is achieved through a direct connection from the user's computer to the hotel network. It's important to note, however, that running internet traffic directly from the user's

computer through the hotel network can bypass organizational security controls, which can create significant risk for the organization.

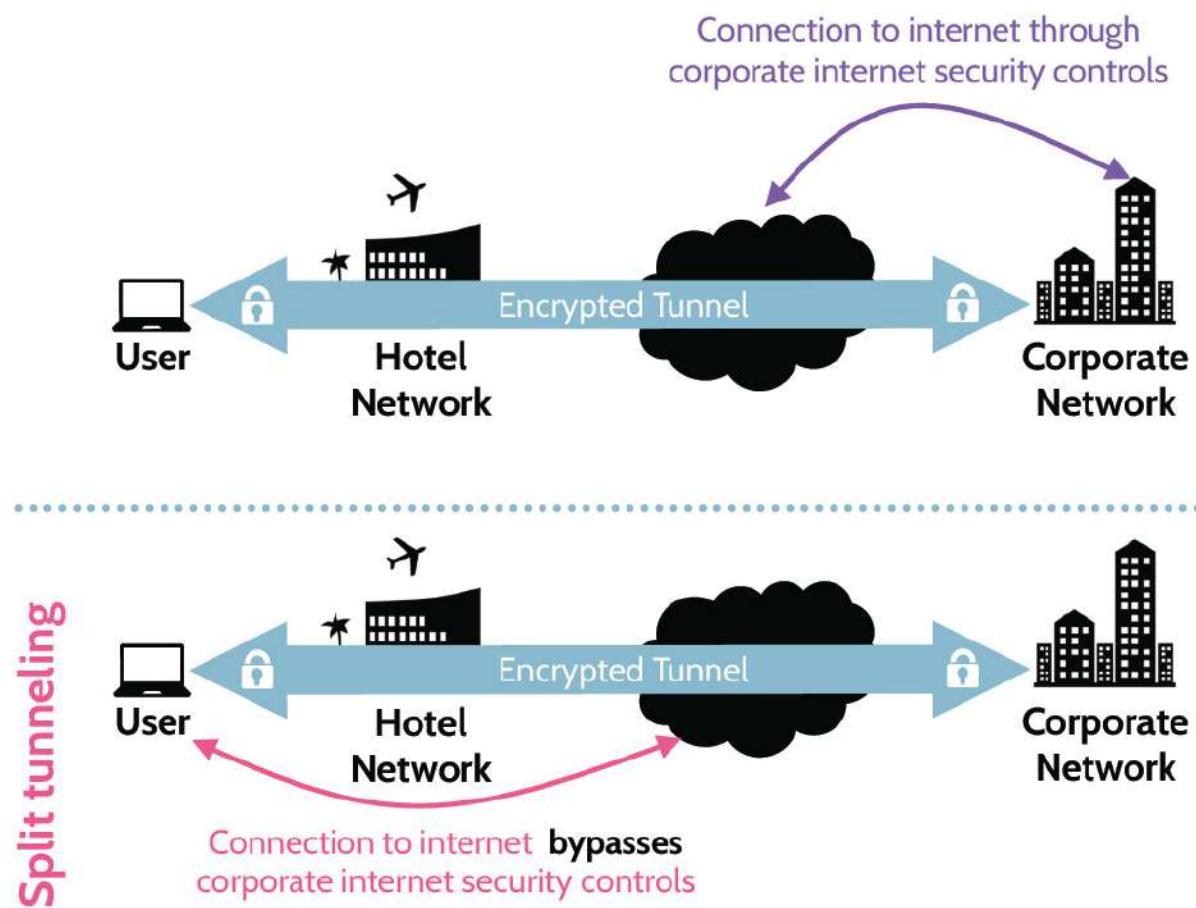


Figure 4-45: **Split Tunneling Operation**

At the same time, split tunneling helps reduce network bandwidth and resource consumption. For example, if the user is browsing to a publicly accessible resource, like Google drive, then there's no point in having that traffic traverse the corporate VPN tunnel and then be redirected to Google's servers. It could just be directed there to begin with, thus optimizing resource consumption. On the other hand, if a

company file server is being accessed, the VPN tunnel can be used.

Virtual Private Network (VPN)

Know the most reliable and cost-effective way to securely connect two networks

VPNs are one of the most reliable and cost-effective ways to securely connect two networks together.

To create a VPN, tunneling plus encryption is needed. A number of solutions exist that utilize SSL/TLS or IPsec VPN technologies. Layer 2 and Layer 3 VPN solutions typically provide both the functionality and efficiency required to host a VPN.

Tunneling and VPN Protocols

Know the most common Layer 2 protocols and understand the differences between PPTP and L2TP

PPTP, L2F, and L2TP are Layer 2 tunneling protocols. They're efficient but don't provide much in the way of functionality and security. Moving up layers allows these needs to be met, and much more functionality can be gained. For example, SSH—at Layer 7—can be used to host command line utilities that would

normally contain zero security. Telnet, FTP, and similar protocols don't contain security, but they become very secure when utilized in the context of an encrypted SSH tunnel. Remember, a VPN requires a tunnel and encryption.

Table 4-46 illustrates commonly used tunneling protocols that include encryption capabilities, with one exception: L2TP. The protocols that contain encryption capabilities can be used to create VPNs. In the case of L2TP, it is often deployed in conjunction with IPsec, which adds the needed encryption element to create a VPN.

Protocol	Tunneling	Encryption	OSI Layer
SSH Secure Shell	✓	✓	7 – Application
SOCKS Socket Secure	✓	✓	5 – Session
SSL/TLS Secure Sockets Layer/Transport Layer Security	✓	✓	4 – Transport
IPsec Internet Protocol Security	✓	✓	3 – Network
GRE Generic Routing Encapsulation	✓		Encapsulation at multiple layers
L2TP Layer 2 Tunneling Protocol	✓		2 – Data Link
L2F Layer 2 Forwarding Protocol	✓	✓	

PPTP	Point-to-Point Tunneling Protocol	✓	✓	
-------------	-----------------------------------	---	---	--

Table 4-46: Common Tunneling Protocols

4.3.2 IPsec

CORE CONCEPTS

- IPsec is preferred for establishing a VPN and is embedded in IPv6 as a default feature.
- IPsec offers authentication via Authentication Header (AH) and encryption via Encapsulating Security Payload (ESP).
- IPsec works in one of two modes: transport or tunnel.
- Internet Key Exchange (IKE) is used to generate the session key shared at each end of the VPN.
- Security Associations (SA) are used to establish components in each direction for an IPsec-based VPN. One SA is needed for each component in each direction.

Of the tunneling and VPN protocols listed in Table 4-46, IPsec is the preferred method for establishing a VPN. One thing to remember about IPsec is that it is natively supported in IPv6 and is therefore becoming a standard component of networking. IPsec offers two advantages over other protocols from a security perspective. It adds authentication of devices as well as encryption. Authentication is added through AH and encryption through ESP. AH provides integrity, data-origin authentication, and replay protection. ESP provides all the

functions AH does, in addition to ensuring confidentiality, as it provides payload encryption.

Think of each as subprotocols or elements of IPsec.

Understand elements and modes of IPsec and services provided by each mode

Additionally, IPsec can be utilized in two different modes: transport mode and tunnel mode. Transport mode uses the header of the original packet, whereas in tunnel mode the header of the new packet encapsulates and encrypts the AH or ESP header and original IP header in the data, or payload, portion of the new packet. Recall from earlier that the original packet was encapsulated in the data portion of a new packet. Now, a choice exists: the header and payload of a packet can be encrypted inside the new data portion. One encrypts only the payload; the other encrypts the header and the payload. [Table 4-47](#) and [Table 4-48](#) and [Figure 4-46](#) summarize the IPsec modes.

IPsec Modes

Authentication Header (AH)	Encapsulating Security Payload (ESP)
Provides integrity, data-origin authentication and replay protection	Provides integrity, data-origin authentication, replay protection, and confidentiality through encryption of payload

Table 4-47: Authentication Header (AH) and Encapsulating Security Payload (ESP)

Transport Mode	Tunnel Mode
Use header of original packet, followed by the AH or ESP header, then the payload	New header encapsulates the AH or ESP header and the original IP header and payload

Table 4-48: IPsec Modes

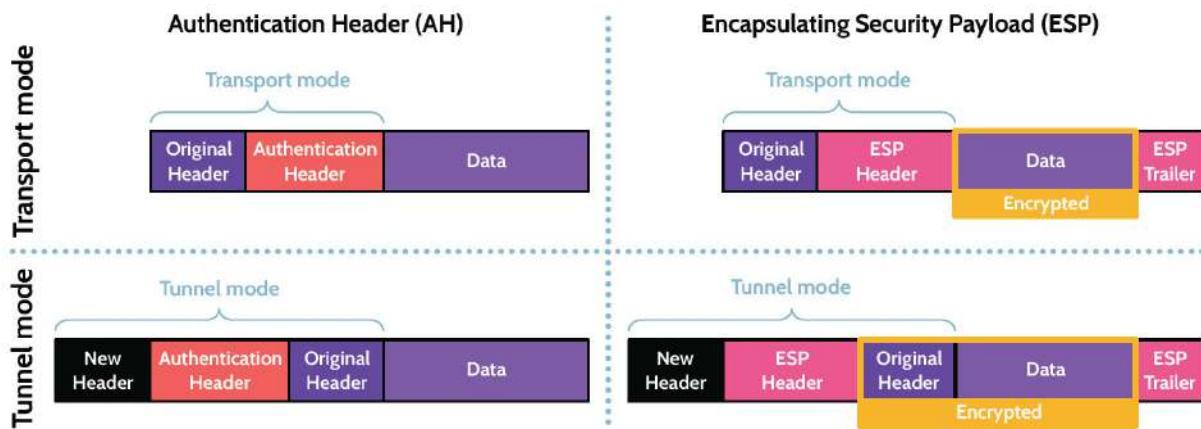


Figure 4-46: IPsec Modes

Internet Key Exchange (IKE)

What protocol is used to exchange keys with an IPsec VPN?

To create a VPN, two things are required: tunneling plus encryption. The only type of encryption that can be used for a VPN is symmetric, meaning the same key is utilized at each

end of the VPN connection. This implies that all VPN solutions must include a key management protocol that allows the same secret to be generated at each end. In the context of IPsec, the key management protocol used is known as Internet Key Exchange, or IKE. IKE is essentially a version of Diffie–Hellman and is used by IPsec to generate the same session key at each end of a VPN.

Security Association (SA)

IPsec tunnels are established through a Security Association (SA). *SA is a **one-way** establishment of attributes at the start of communication between two entities.* Imagine two people who want to connect and communicate. To do so, security associations would need to be established. Because an SA is one-way form of communication (like walkie-talkies), where only one person can communicate at a time, in order for two people to communicate, an SA would need to be established in each direction. If IPsec AH and ESP components are needed, SAs are needed in each direction for each component. So, if both AH and ESP are needed, a total of four SAs is required—two SAs in each direction.

Attributes include:

- Authentication algorithm
- Encryption algorithm

- Encryption keys
- Mode (transport or tunnel)
- Sequence number
- Expiry of the SA

4.3.3 SSL/TLS

CORE CONCEPTS

- **Secure Sockets Layer/Transport Layer Security (SSL/TLS) provide secure client to server connections; the two names refer to the same thing, but TLS is now the proper standard and most current version, as SSL is obsolete.**
- **DROWN attack is a major threat to SSLv2.**
- **SSL/TLS connection steps: client hello, server hello, creation and sharing of session key, establishment of secure session.**
- **Asymmetric cryptography is used to encrypt the symmetric session key created by the client.**
- **Unencrypted SSL sessions can exist, where a browser and server authenticate to one another, but the communications channel is not encrypted.**

Over the years, Secure Sockets Layer (SSL) has been instrumental as a means of securing communications channels. Of course, like most technologies, it's been revised over the years, with the latest revision being SSL 3.0. In 1999, SSL was revised yet again, but this time, to remove confusion about where this protocol operates in the OSI model, it was

renamed to Transport Layer Security (TLS), and the most recent revision of TLS was released in 2018 as TLS 1.3.

SSL/TLS is used extensively; in fact, most websites that require secure connections between a browser and a web server utilize SSL/TLS. And even though TLS is the proper name, many people still refer to it as SSL. Naming conventions aside, the key is that the latest version is being used, as use of SSLv2—even when used for purposes of backward compatibility—could result in what is known as the **DROWN attack** exploiting a specific vulnerability in SSLv2. If successful, a DROWN attack could result in a session being compromised, and sensitive information—passwords, credit card numbers, proprietary and other valuable data—could be read or stolen. Because disabling backward compatibility can be somewhat complicated, the best defense against DROWN is for server owners to ensure their private keys are not utilized with web, email, and other servers that allow SSLv2 connections.

Understand the foundational reason for using SSL/TLS

SSL/TLS provides secure client to server connections; for instance, when managing your bank account via the bank's online portal, or when looking at your insurance or medical information from providers of those services. A secure connection can be established between your browser and a secure server anywhere in the world, and the secure connection protects transmission of sensitive information and

transactions from eavesdropping. A secure connection can be established in four relatively simple steps, as outlined in [Figure 4-47](#).

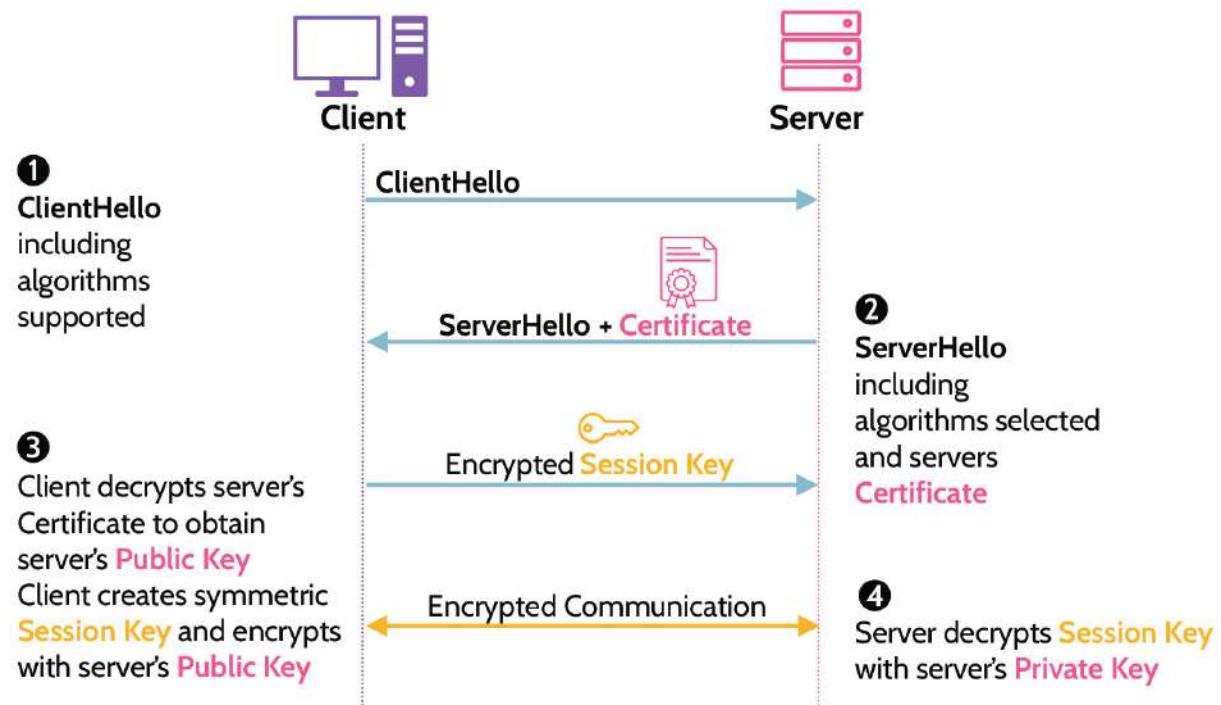


Figure 4-47: SSL/TLS Handshake

Understand the SSL/TLS handshake process and how asymmetric and symmetric cryptography both play a role in establishing a secure session

[Figure 4-47](#) shows an example of an SSL/TLS handshake.

You want to purchase something online from Amazon.

Initially, you'd visit Amazon's website by typing www.amazon.com in your browser's address bar. This

establishes the initial connection, and at the same time Amazon's server and your browser understand that a secure connection is required, which activates the SSL/TLS handshake process.

1. The client, your browser, will first send a client hello message to the server, which allows the server to determine whether it understands the client, the browser. If all is well, the server will essentially respond with a message saying it understands the browser. Otherwise, if the browser is outdated or doesn't have a compatible version of SSL/TLS installed, the server will respond with some type of error warning.
2. After the initial "hello," the second step requires the server to do something similar and send a server hello message to the client. However, with the hello message, the server will also send its certificate, which contains Amazon's public key.
3. With Amazon's public key, the client's browser understands what certificate authority issued Amazon's certificate. For this example, let's assume that VeriSign issued the certificate. The browser knows this because one of the fields on the certificate is the issuer ID, which the browser uses to look up the issuer, in this case, VeriSign. With this information and built-in functionality, the browser fetches VeriSign's public key and decrypts the certificate. If the certificate decrypts

properly, the public key on Amazon's certificate can be trusted and, in fact, Amazon's server has been authenticated. You know your browser is connected to Amazon's server and not a server pretending to be Amazon. Otherwise, to spoof Amazon, VeriSign's certificate would've first needed to be spoofed, which would require access to VeriSign's private key. And only VeriSign has access to its private key.

4. With the first three steps complete, step four involves the creation of a symmetric key (session key) that will be utilized from this point forward for the session. Again, utilizing functionality built into the browser, a session key is created and encrypted using Amazon's public key and then sent back to Amazon, where Amazon's private key will be used for decryption. At this point, the client's browser and Amazon will each hold the same session key, which will be used for further communication. Now safe transmission of authentication information, credit card numbers, and other sensitive information can take place, because only the client and Amazon have access to this specific session.

From the example above, let's highlight some key takeaways.

- There's always a need for the client to authenticate the server. In this case, it's much more important for the client to ensure it has authenticated to Amazon. Or

when banking, it's vital to know that the client has authenticated to the actual bank. There's really no need for Amazon or the bank to authenticate the client—whether the client is at home or somewhere around the world, the server really doesn't care. However, if this functionality was needed, it's built into SSL/TLS and is known as **mutual authentication**. In this case, the server can also authenticate the client, which would require the client to have a certificate too.

- Only the client can create the session key, which is then sent back to the server encrypted with the server's public key.

The SSL/TLS handshake process happens all the time and very quickly. Once complete, a small lock in the browser address bar will appear, and the URL will begin with HTTPS, which signifies a secure HTTP session is now enforced.

TLS VPN versus IPsec VPN

Understand high level differences between TLS VPN and IPsec VPN, including the layers where they operate and which one provides encryption by default

Both TLS and IPsec VPNs provide secure communication channels, but they do so a bit differently from one another. A comparison is provided in [Table 4-49](#).

TLS VPN	IPsec VPN
<ul style="list-style-type: none"> ■ Operates at the Transport layer and above ■ Can be used to encrypt traffic sent between any processes identified by port numbers ■ Encrypts connections by default ■ Easier to establish and manage and has more granular configuration options available ■ A successful attack could lead to compromise of specific systems and applications 	<ul style="list-style-type: none"> ■ Operates at the network layer ■ Can be used to encrypt traffic between any system that can be identified by an IP address ■ The use of IKE provides a layer of data authentication via an external protocol ■ Does not encrypt connections by default ■ Can be more complicated to establish, configure, and manage ■ A successful attack could lead to compromise of an entire network

Table 4-49: **TLS vs. IPsec VPN**

Choosing one type of VPN over another is very dependent upon the requirements of the organization and the specific goals and objectives that relate to security of information being communicated internally and externally. Factors like performance, security, data authentication, and others should be considered in this vein. The approach that might work for one organization could be completely wrong for another organization.

4.3.4 Remote Authentication

CORE CONCEPTS

- **Remote authentication protocols include: RADIUS, TACACS+, Diameter.**

- RADIUS was originally developed to support dial-in networking and provides authentication, authorization, and accounting.
- TACACS+ uses TCP and encrypts all packets.
- Diameter is the successor to RADIUS and includes much-improved security, including EAP.

Understand the importance and use of remote authentication protocols and their similarities and differences

One important consideration related to remote access is that a simple VPN connection does not prove who is sitting behind the VPN. If a company uses SSL/TLS for VPN capabilities, the VPN only protects the traffic between the device on one end and the server or device on the other end. Most organizations also want to know who is using the computer or browser and will therefore use some type of authentication—usually two-factor—for this purpose. Authentication protocols like RADIUS, TACACS, TACACS+, and Diameter, summarized in [Table 4-50](#), are typically the most used for this purpose.

Remote Authentication Protocols

RADIUS	Remote Authentication Dial-In User Service is an application-layer protocol that allows a user to connect to and access network resources. RADIUS was developed to support dial-in networking and provides authentication, authorization, and accounting (AAA).
---------------	---

TACACS+	Terminal Access Controller Access Control System Plus was developed by CISCO as an extension of TACACS and an improvement over RADIUS. Unlike RADIUS (which uses UDP) TACACS+ uses TCP and encrypts all the transmitted packets, while RADIUS only poorly obfuscates user passwords.
Diameter	Diameter is the successor to and an enhanced version of RADIUS. Diameter adds improved security features, such as EAP, which provides a much more secure and robust authentication of users.

Table 4-50: **RADIUS, TACACS+, and Diameter**

 MINDMAP REVIEW **VIDEOS**

Open Systems Interconnection (OSI) Model							
	1. Physical	2. Datalink	3. Network	4. Transport	5. Session	6. Presentation	7. Appl.
Media	Topologies	Collisions	MAC Address	IP Address	Ports = Services	Application Firewalls	
Wired: Twisted Pair, Coaxial, Fiber Optic Wireless: Radio Frequency, Infrared, Microwave	Bus, Tree, Star, Mesh, Ring	CSMA/CA, CSMA/CD Hubs, Repeaters, Concentrators	Devices Protocols	Devices Protocols	Switches & Bridges Routers & Packet Filtering Firewalls Common Ports	TCP/UDP, SSL/TLS & BGP Circuit Proxy Firewall NetBIOS & RPC	HTTPS, DNS, SSH, SNMP, LDAP, DHCP
		802.1x	ARP, PPTP, PPP, CHAP, EAP	ICMP (Ping), IPSec, IGMP	Protocols	Protocols	Protocols

OSI Model

dgo.ca/CISSPmm4-1

Networking							
WAN		Wireless				Network Attacks	
X.25	Frame Relay	ATM	MPLS	Wi-Fi	GSM / CDMA	Microwave	Internet Protocol (IP) Addresses
Protocols	802.11a, b, g, n, ac, ax	Encryption	802.16	WiMAX	IPv4 vs. IPv6	IPv4 Network Classes	Private IPv4 Addresses
WEP	TKIP	WPA / WPA2			iSCSI, FCoE	PAP	Converged Protocols
					CHAP	EAP	Network Authentication
					PEAP		
Phases				Reconnaissance			
Eavesdropping				Enumeration			
SYN Flooding				Vulnerability Analysis			
IP Spoofing				Exploitation			
DoS / DDoS				Man-in-the-Middle			
ARP poisoning				ARP poisoning			
VLAN				Virtualization			
SDN				Common Commands			
ipconfig				ping			
traceroute				whois			
dig							

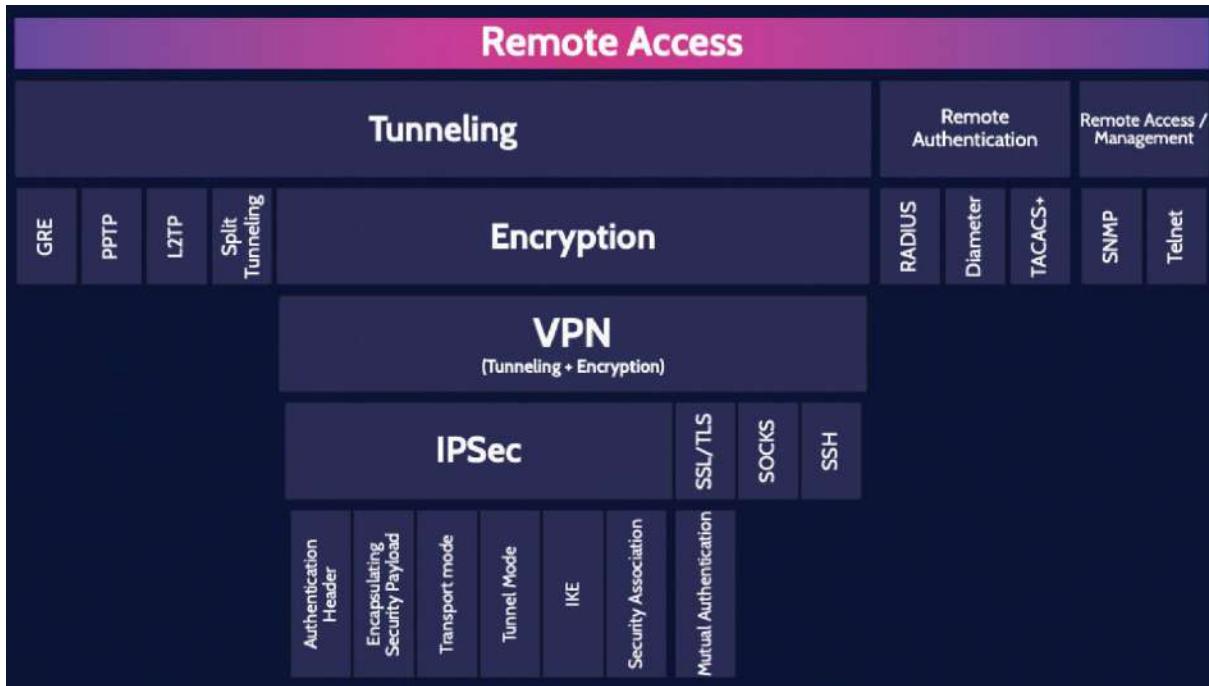
Networking

dcgo.ca/CISSPmm4-2

Network Defense											
Defense in Depth	Network Segmentation / Partitioning			Firewalls		Inspection					Endpoint Security
	Network Perimeter	DMZ	Bastion Host	Proxy	NAT / PAT	Types	IDS	IPS	IDS/IPS Location	IDS / IPS Detection Methods	
	Packet Filtering	Stateful Packet Filtering	Circuit Proxy	Application		Host Based	Network Based	Pattern			Honeypots & honeynets
						In-line	Mirror, Span, Promiscuous	Signature analysis	Stateful matching	Anomaly	Ingress vs. Egress
									Statistical	Traffic	
									Protocol	Sandbox	
										White & Black lists	

Network Defense

dcgo.ca/CISSPmm4-3

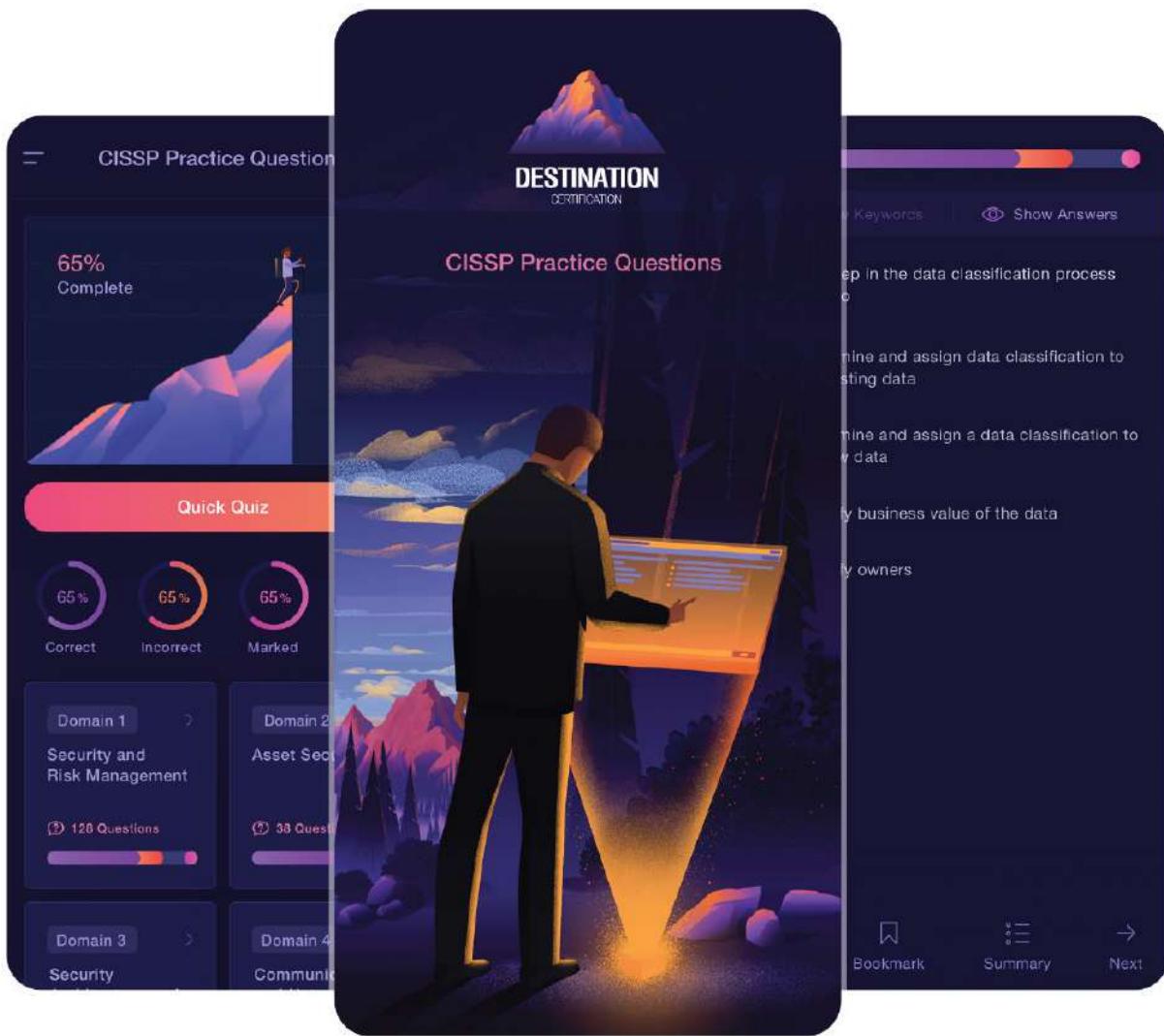


Remote Access

dcgo.ca/CISSPmm4-4

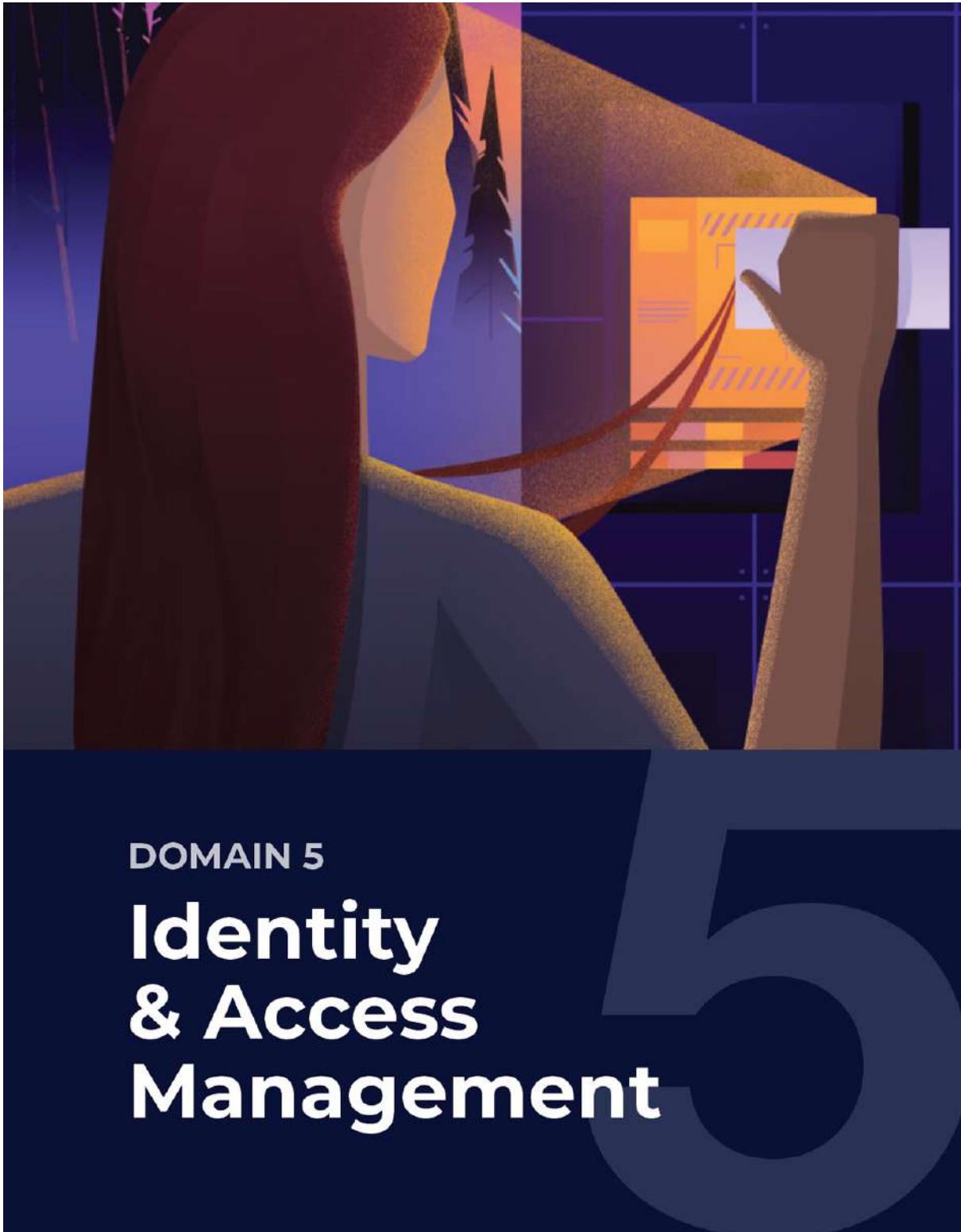


CISSP PRACTICE QUESTION APP



**Download the Destination CISSP Practice Question app for
Domain 4 practice questions**

dcgo.ca/PracQues



DOMAIN 5

Identity & Access Management

5.1 Control physical and logical access to assets

5.1.1 Access Control

CORE CONCEPTS

- **Access control is the concept that refers to the collection of mechanisms that work together to protect organizational assets while simultaneously allowing controlled access to authorized subjects.**
- **Fundamental access control principles include: need to know, least privilege, separation of duties.**
- **Access control is applicable at all levels of an organization and covers all types of assets.**

What is access control?

*Access control is the collection of mechanisms that work together to protect the **assets** of an organization and, at the same time, **allow controlled access to authorized subjects.***

Access control enables management to:

- Specify which **users** can access the system
- Specify what **resources** they can access
- Specify what **operations** they can perform
- Provide individual **accountability**—know who is doing what
- Specify what

Access Control Principles

The fundamental access control principles denoted in [Table 5-1](#) are important to understand because they are applied everywhere within access control. Individually and together, they help protect organizational assets by limiting what individuals can do with an asset in order to perform their job, and nothing more.

Need to Know	Least Privilege	Separation of Duties
Defending sensitive assets by restricting access to only required personnel who require access	Defending sensitive assets by granting only the minimum permissions required by the user or system	Defending sensitive assets by requiring more than one person to complete a task, to prevent errors and fraud

Table 5-1: Access Control Principles

Understand the fundamental access control principles and how they might be applied

Need to know: The concept of *need to know* can be applied in many ways. Imagine a law enforcement agent being undercover and investigating a case. Their true identity doesn't need to be known to anyone apart from their direct

supervisor and a handful of agents involved in the case. In short, only who needs to know will know to maintain operational security. Need to know simply means that you are given access to an asset that you absolutely need to access, based on your job role, function, or what you've been authorized to do.

Least privilege: A classic example of overprivileged accounts exists in a variety of companies today where several people (if not the whole company) just have local administrator permissions on their machines. This goes against least privilege. Most people in a company don't need to have local administrator permissions, and hence to apply least privilege, a group policy should state that everyone has standard accounts configured, apart from a handful of administrators. Least privilege means you are given only the absolute level of access that you need, and nothing more. If you only need read access, that's all you are given.

Separation of duties and responsibilities: Separation of duties and responsibilities refers to the concept that one person should not be responsible for all aspects of a process. Separation of duties is often employed in areas of an organization where money is received or disbursed. For example, when a new vendor is added to an accounts payable system, one person might enter the vendor

information and another person might confirm the validity or accuracy of the information. These two steps can help prevent fake vendors from being created in a system. In addition, when the vendor is paid, one person might enter the invoice and payment information, another person might generate the check, and yet another actually confirms the check amount against the invoice and then signs it. Another example relates to developers, they should not be the same people who push applications to production. There needs to be a separation of duties in place, so proper testing, validation, and approval can be conducted to prevent errors. *Separation of duties helps prevent fraud.*

Access Control Applicability

Access control includes *all aspects and levels* of an organization and covers all types of assets including:

- Facilities
- Systems/Devices
- Information
- Personnel
- Applications

Access Control System

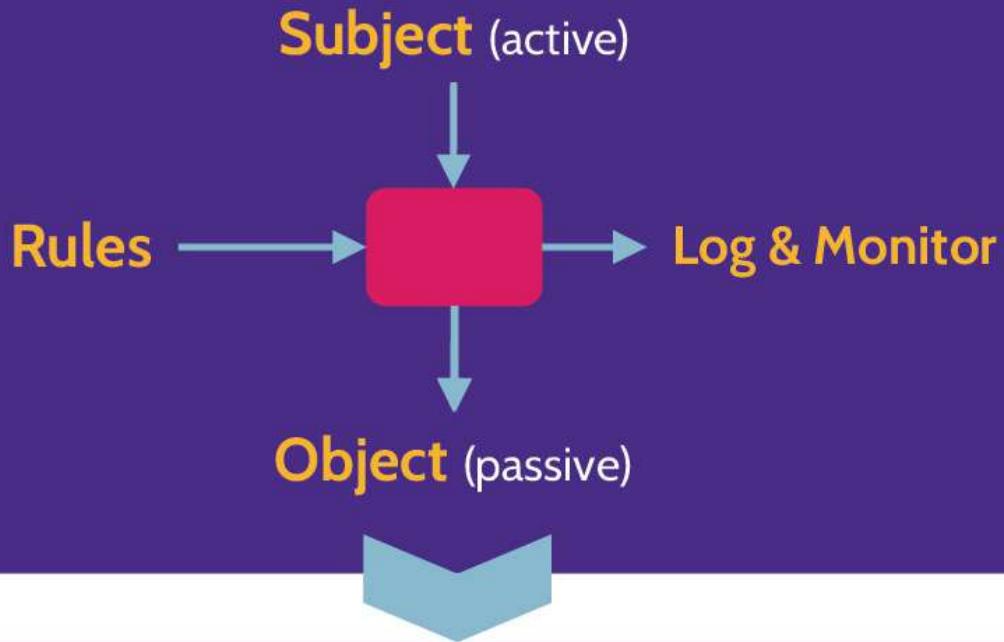
- The focus of access control is controlling a subject's access to an object through some form of mediation.
- Mediation is based upon a set of rules.

- All activity is logged and monitored to provide accountability and gain assurance that things are working properly.

Fundamentally, the above point to the use of the RMC, or Reference Monitor Concept, where some type of rules-based decision-making tool is placed in between subjects and objects to mediate access, and all activity is logged and monitored for the sake of accountability and assurance. Any implementation of the RMC is known as a security kernel.

[Figure 5-1](#) depicts the RMC and its various components.

REFERENCE MONITOR CONCEPT (RMC)



Implementation of RMC = **Security Kernel**

Figure 5-1: RMC Components

Logical Access Modes

- Access control is more granular than simply allowing subjects to access objects.
- Access control rules allow the access control mechanism to be much more granular.
- Specific access rules allow precision with regards to what subjects can access what objects and exactly what those subjects can do with those objects.

- Access should be based on the use of concepts like *need to know* and *least privilege*.

Access rules and related subject rights will vary from system to system, but in general, the following logical access modes can be found in most systems:

- Create
- Update
- Read
- Read/Write
- Execute
- Delete

Groups versus Roles

Groups and roles are essentially just two different approaches for assigning permissions to users. The differences are outlined in [Table 5-2](#).

Groups	Roles
<p>A collection of users that has been named. The users are generally not associated with a specific job in the organization. Instead, they could be part of a specific internal leadership team or focus area. As an example, they could be specific members of a business continuity management team.</p>	<p>A set of permissions that is usually associated with a specific job within an organization, such as a call center agent.</p>
<p>An admin can create groups of users and then assign permissions to the group, which can make it simpler to manage many users at once.</p>	<p>If a user is assigned to role, they are given the permissions of that role.</p>

<p>Groups are more flexible.</p>	<p>Roles are focused around the function of the job, the actions that must be performed, and therefore the permissions that are required to do the job.</p>
----------------------------------	--

Table 5-2: The Differences Between Groups and Roles.

5.1.2 Administration Approaches

CORE CONCEPTS

- Access control administration often takes one of two approaches: centralized or decentralized, and many organizations also utilize a hybrid approach.
- Each approach offers pros and cons.

Understand access control administration approaches and pros and cons of each approach

Access Control Administration Approaches

When administering access control, two primary approaches are often taken: centralized or decentralized, although more organizations are now also utilizing a hybrid approach that incorporates elements of centralized and decentralized approaches. These are depicted in [Table 5-3](#).

Centralized	Hybrid	Decentralized
-------------	--------	---------------

<ul style="list-style-type: none"> ■ One central system controls access to remote systems ■ One username and password in the central system used to access all systems ■ Central administrative point represents a single point of failure and potential target of attack 	<ul style="list-style-type: none"> ■ Approach taken by most organizations, and it means access control utilizes both approaches (centralized and decentralized) ■ This is often due to legacy systems in an organization that can't be integrated with newer, more modern, access control systems 	<ul style="list-style-type: none"> ■ Control is granted to the people closer to the resource ■ Access requests are not processed by one centralized entity; separate usernames and passwords exist on each resource ■ Lack of standardization, overlapping rights and security holes may exist ■ Peer-to-peer relationship

Table 5-3: Centralized, Decentralized, and Hybrid Access Control

Centralized Administration

Some advantages of a centralized approach are much easier administration, lower overhead, cost reduction, and greater flexibility. However, a major disadvantage is that the central administrative point represents a single point of failure as well as target of attack.

Decentralized Administration

Likewise, an advantage of a decentralized approach is the ability to manage access control at much more granular levels, though a corresponding disadvantage is that this ability creates much more administrative overhead. Another advantage is that decentralized administration of access control minimizes the risks associated with a single point of failure, as found with centralized administration. If one system goes down or is compromised, other systems remain viable.

5.2 Design identification and authentication strategy (e.g., people, devices, and services)

The Seven Laws of Identity

The seven laws of identity were developed by Kim Cameron and a range of other security experts. They are described in [Table 5-4](#).

User control and consent	Identity systems should be designed so that information identifying a user is only revealed with the consent of the user.
Minimal disclosure and constrained use	The most stable long-term identity solution is the one that reveals the least amount of identifying information and limits the use of this data.

Justifiable parties	Identity solutions should be designed to only disclose identifying information to parties that have a justifiable and necessary reason to be part of the identity relationship.
Directed identity	Identity systems should support omni-directional identifiers for public entities, as well as uni-directional identifiers for private entities. As an example, a public website should have omni-directional identifiers so that users can easily find and connect to it. These users should have uni-directional identifiers so that they can connect to the website while preserving their privacy from other parties.
Pluralism of operators and technologies	Universal identity systems should be interoperable with a range of identity providers and technologies.
Human integration	Identity systems must be designed under the consideration that human users are a vital component. They must be integrated into the system, with protections in place to also guard the communications between the human and the system itself. Identity systems need to ensure that the communication between the users and the system is extremely reliable.
Consistent experience across contexts	The identity system must provide users with a consistent and simple experience, even when they use different identity providers and technologies.

Table 5-4: The Seven Laws of Identity.

5.2.1 Access Control Services

CORE CONCEPTS

- **Access control services consist of: identification, authentication, authorization, accountability.**
- **Identification refers to the assertion of a user's identity or a process to a system.**
- **Authentication refers to the verification of an identity through knowledge, ownership, or characteristic.**
- **Authorization refers to the level of access defined for the identified and authenticated user or process.**
- **Accountability refers to proper identification, authentication, authorization that is logged and monitored.**
- **Accountability is also known as the Principle of Access Control.**

Access control and related services are fundamental elements of organizational security. Assets and users can best be protected and held accountable for their actions. In fact, this latter point—the notion of accountability—is the fundamental driver behind Access Control Services. As a security professional, it's imperative to understand the implications of nonexistent or weak and ineffective access control, especially as it relates to the Principle of Access Control.

The components of Access Control Services are depicted in [Figure 5-2](#) and explained in detail in the following section. These are Identification, Authentication, Authorization and Accountability. Sometimes, we refer to these last three processes as AAA.

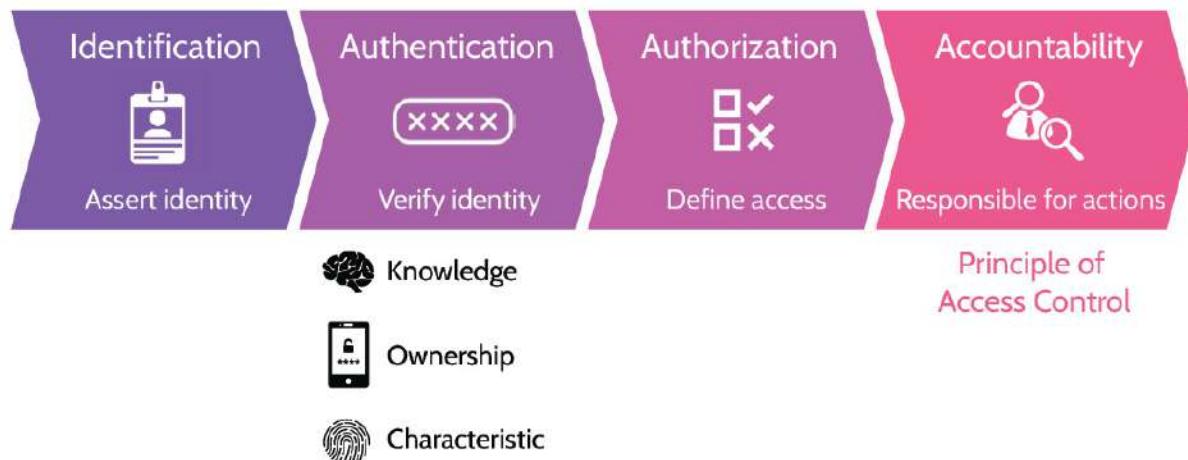


Figure 5-2: Access Controls Services Components

5.2.2 Identification

CORE CONCEPTS

- **Identification is the component of Access Control Services that uniquely asserts a user's identity or a process to a system.**
- **Identification guidelines include: unique identifiers, nondescriptive of job/role, secure issuance.**

Identification as the first component of Access Control Services, identification methods and guidelines

The notion of identification is simple: To gain access to a system, a *unique* identity should be presented, which can then be used to trace activity to an individual. Shared user accounts potentially allow circumvention of the Principle of Access Control and therefore should not be used.

A few examples of identification methods include a user's ID (for example, first name, last name, both, etc.), account ID, access card, biometrics.

User identification needs to be

■ **Unique** (relating to one individual or process) ■

Nondescriptive of role (e.g., admin accounts should not include the word “admin,” or finance-related user accounts should not point to a role or job related to the same) ■ **Issued and used securely**

(e.g., use a password manager to generate and store user passwords rather than writing them down to a notepad) As it pertains to **authentication**, there are three factors of authentication that can be used to verify a user's identity, which are summarized in

[Table 5-5](#).

Authentication by Knowledge	Authentication by Ownership	Authentication by Characteristic
Something you know	Something you have	How you behave or your physiology; something

Table 5-5: Factors of Authentication



5.2.3 Authentication by Knowledge

CORE CONCEPTS

- **Authentication by knowledge is the component of Access Control Services that refers to verification of an identity through something that is known, like a password, passphrase, or questions**

Authentication by knowledge as one type of authentication

Authentication by knowledge, or also known as something you know, simply means that a person uses a password, a passphrase, or a response to one or more security questions to authenticate to a system. A password could be as simple

as “password” or something more complex, like “m{BLB9FF#6hJ#U\$.” Complex passwords aren’t easy to remember, which is why many people write down their passwords on sticky notes and place them under their keyboard or on their monitor. A passphrase is typically as complex as the password example and much easier to remember. The thinking here is that a sentence or lyric from a song or book can be used to authenticate. A phrase is longer and can be as or more complex than a standard password. Finally, one or more security questions can be presented to a user, and upon answering them correctly the user is authenticated. The questions are usually chosen by the user, and the corresponding correct answers should be details that only the user knows. These are often known as cognitive questions. In addition, note that answers to these questions don’t have to be true. For example, if a question asks, “What’s your maiden name?” you can still answer “3487487glkjgokjo!(*&” (good luck pronouncing that!), but the point is that the answer doesn’t have to be true. This makes security questions impossible to guess from an attacker’s point of view.

Different forms of authentication by knowledge

Ideally, whether a password, passphrase, or questions are used, each should be unique to the user and not easily

guessed or otherwise determined.

5.2.4 Authentication by Ownership

CORE CONCEPTS

- **Authentication by ownership is the component of Access Control Services that refers to verification of an identity through something a user possesses.**
- **One-Time Passwords (OTP) are generated via a synchronous or asynchronous process.**
- **Soft tokens refer to software-based applications, like Google Authenticator, that generate one-time passwords.**
- **Hard tokens refer to small physical tokens, like RSA's SecureID device, that generate one-time passwords.**
- **Smart cards are typically credit-card-size plastic cards with an embedded semiconductor chip that stores, accepts, and sends information; they work in collaboration with a smart card reader.**
- **Memory cards are typically credit-card-size plastic cards with a narrow magnetic strip on the back of that card that contains information related to the card owner, issuing bank, account number, PIN, etc.**

Authentication by ownership as another type of authentication

One-Time Passwords (OTP) are exactly what the term says, they are passwords that can be used one time. Additionally, one-time passwords are *dynamic* in that they expire after

being used, or they expire after a certain period of time. One-time passwords can be generated via a **soft token**, a software app (like Google's Authenticator app), or they can be generated via a **hard token**, a dedicated hardware device (like RSA's SecureID device).

Understand the difference between asynchronous and synchronous password generation

One-time passwords can be generated via an asynchronous or synchronous process as depicted in [Figure 5-3](#). Of the two processes, asynchronous is much less common, as it involves a fair amount of complexity with regards to the synchronization elements involved between the authorization server and the user and system they're accessing. This complexity often comes with a hefty price tag, though it does offer a more robust layer of security. Relative to the asynchronous process, the synchronous one-time password authorization process is much more straightforward and less complex.

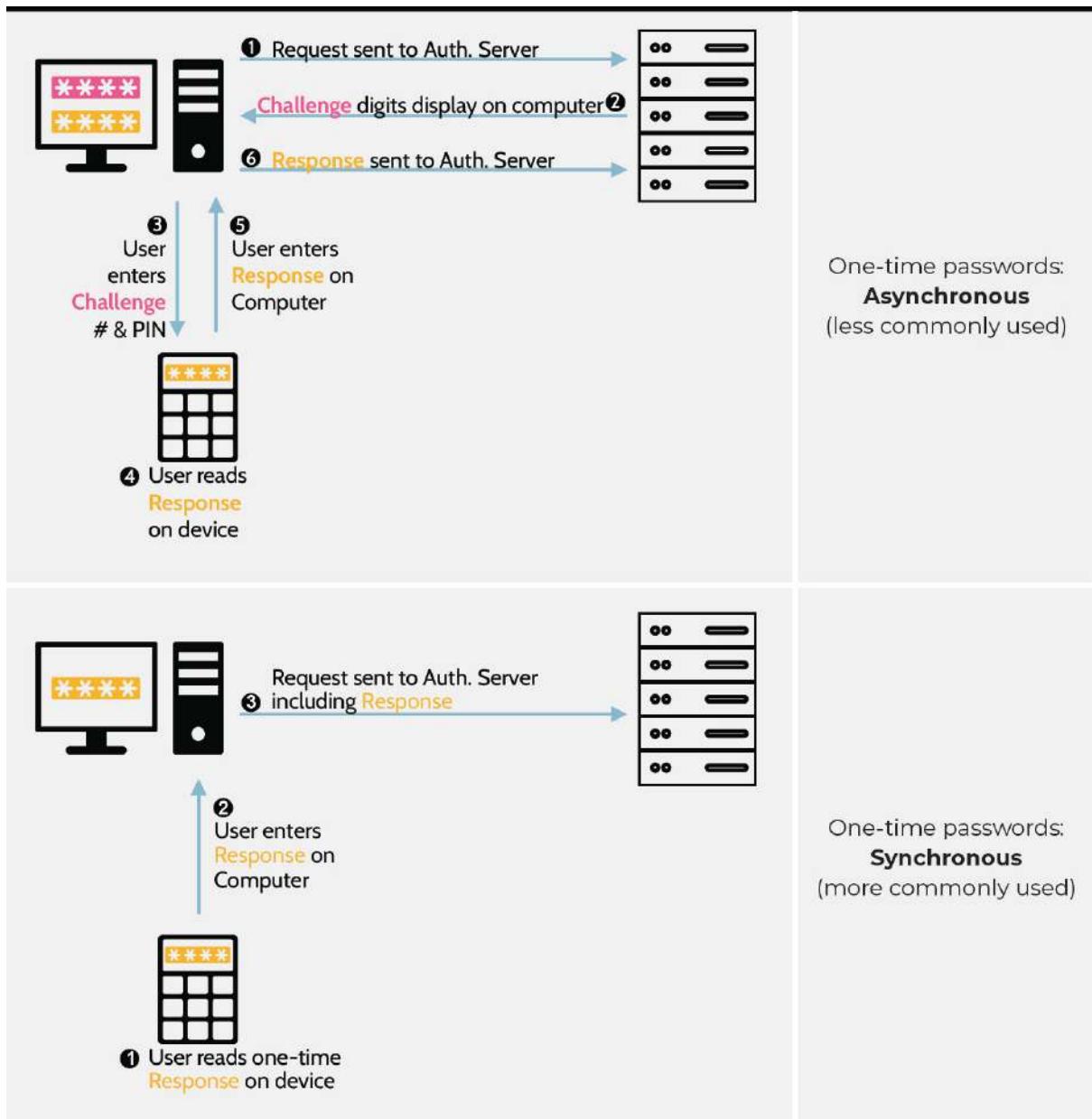


Figure 5-3: **Synchronous and Asynchronous One-Time Passwords**

Smart/Memory Cards

Understand the differences between smart and memory cards

Both smart and memory cards consist of an authentication factor by ownership, as you have a smart or memory card in your possession. Both cards store information about the card holder. Smart and memory cards are summarized in

[Table 5-6](#).

Smart Card	Memory Card
Called smart cards because they contain a small embedded integrated circuit (IC) chip that can perform calculations and generate unique authentication data with each transaction.	Implies a form of memory stored on a card, typically on a magnetic strip on the back of the card. The same data is read from the magnetic strip with every transaction.

Table 5-6: Smart and Memory Cards

A memory card will have a magnetic stripe on the back of the card, which is where the information is stored. Older cards were typically just memory cards, and this fact led to widespread growth of credit card fraud—specifically skimming (covered in 3.9.11). Newer cards (especially modern debit and credit cards) tend to be a combination of smart and memory card as they contain a chip (smart) and

a magnetic strip (memory). The chip in smart cards can act as a processing engine for them to accept, store, and send information as they communicate with a card reader.

There are two main methods that smart cards can communicate with a reader as noted in [Table 5-7](#).

Contact Smart Cards	Contactless Smart Cards
The chip in the card needs to contact the reader for the chip to be powered and allow transactions to be processed.	The reader sends out signals that are powerful enough to communicate with and power the chip in a smart card, allowing it to perform some calculations and wirelessly send a response to the reader.

Table 5-7: Smart Cards' Communication Methods

5.2.5 Authentication by Characteristics

CORE CONCEPTS

- **Authentication by characteristics refers to physiological and behavioral biometric types.**
- **Biometric device accuracy can vary and is not always 100 percent accurate.**
- **The use of biometric devices must consider: processing speed, user acceptance, protection of biometric data, accuracy.**
- **Crossover Error Rate (CER) represents the intersection between Type 1 (false reject) and Type 2 (false acceptance) errors, and it measures**

the accuracy of a biometric system.

The various physiological and behavioral authentication types are shown in [Table 5-8](#). Remember that physiological relates to the physical attributes of a person, while behavioral relates to how a person behaves.

Physiological Characteristics	Behavioral Characteristics
fingerprints	the way a person writes
hand geometry	the way a person walks—their gait
facial features	the way a person speaks—their voice
eyes (retina and iris)	the way a person types—keyboard dynamics

Table 5-8: Physiological and Behavioral Authentication Types

Biometric Device Considerations

Due to the nature of how biometric authentication works, factors like the ones noted below must be considered:

- Processing Speed
- User Acceptance
- Protection of Biometric Data
- Accuracy

Biometric systems can be much slower than other types of authentication systems. Because of this fact, users may be less willing to accept the implementation and use of biometric authentication, and because of the inherent uniqueness of biometric data, its

protection is of paramount importance. Finally, the accuracy of biometric systems must be carefully considered.

Biometric Device Accuracy/Types of Errors

Biometric error types and which is worse

Unlike other types of authentication systems, biometric systems are not 100% binary. In other words, they're not always 100% accurate. With a traditional username/password authentication system, for example, the username and password must be 100% correct for a user to be authenticated to a system. The same cannot be said of a system that uses physiological or behavioral attributes for authentication.

When considering biometric systems and their accuracy, two primary types of errors need to be understood as summarized in [Table 5-9](#).

Type 1—False Rejection	Type 2—False Acceptance
This is when a valid user is falsely rejected by the system. The false rejection rate (FRR) is the probability that a valid user will be rejected by the system. It is expressed as a percentage.	This is when an invalid user is given access to a system. It is a much more serious and potentially dangerous situation. The false acceptance rate (FRR) is the probability that an invalid

user will be accepted by the system.
It is expressed as a percentage

Table 5-9: Metrics for Biometric Systems Accuracy

With Type 1 errors, the result is usually just frustrated users that can't authenticate to the system. However, with Type 2 errors, a malicious individual could gain access to the environment, as the tool falsely thinks they are legitimate.

Hence, **Type 2 errors are much more dangerous than Type 1 errors**. Also note the terms False Rejection Rate (FRR), which expresses the Type 1 error rate, and False Acceptance Rate (FAR), which expresses the Type 2 error rate.

An interesting feature of biometric systems is that these error rates are inverse to one another, which relates to the system's tuning. In other words, by tuning a system and reducing Type 2 errors, Type 1 errors tend to increase. On the other hand, by reducing Type 1 errors, Type 2 errors will increase. This inverse relationship is depicted in [Figure 5-4](#).

Crossover Error Rate (CER)

Crossover error rate and what it represents

As [Figure 5-4](#) depicts, when Type 1 (false reject errors) are low, Type 2 (false accept errors are high) and vice versa. The

intersection of the two error plots is what's known as the Crossover Error Rate (CER). The crossover error rate is a useful metric for biometric systems, because it's a way to measure the overall accuracy of the system. No system is going to have a CER of zero, but a number closer to zero means the system is more accurate. CER is often used when looking for and comparing biometric systems and their functionality and effectiveness.

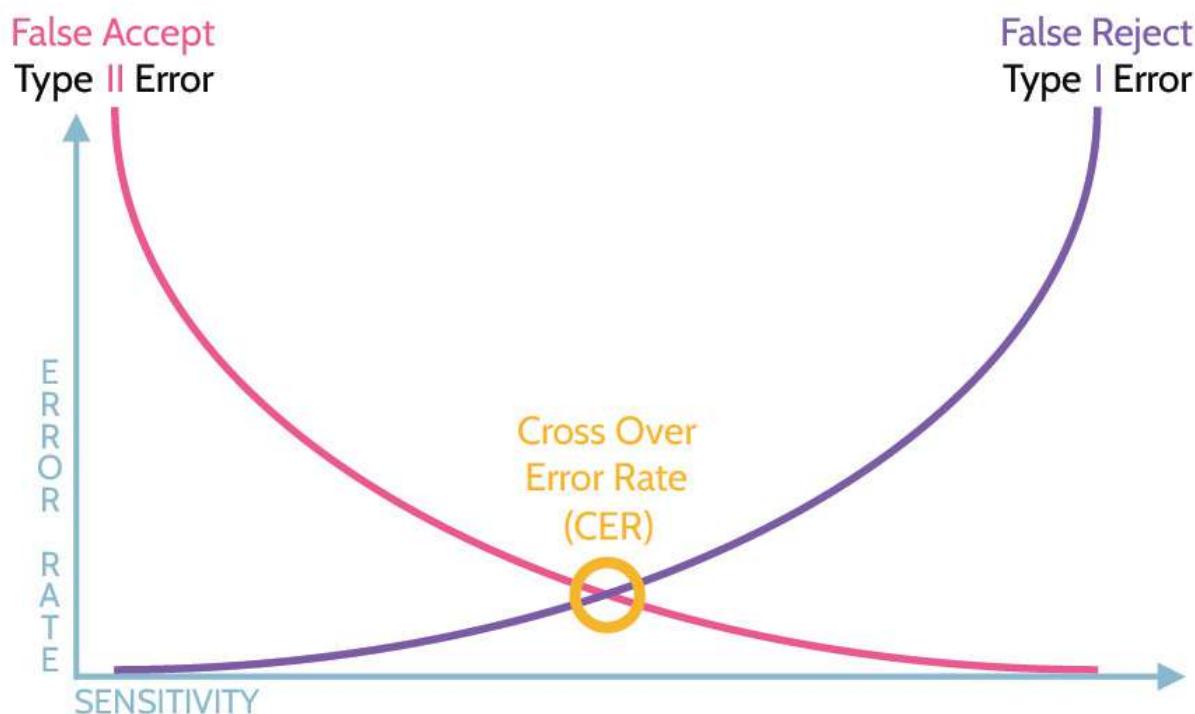


Figure 5-4: FAR, FRR, and CER

Biometric Templates

As noted above, the protection of biometric data is extremely important. If someone's biometric data is exposed, the consequences are much more severe than, for

example, their password being exposed. The reason is obvious: if someone's password is exposed, they can just change their password and memorize the new password. However, if their biometric data is exposed, they can't just grow a new finger or a new eyeball.

Accordingly, raw, or original biometric data should never be stored, such as a simple picture of a fingerprint. Instead, good biometric systems will use one-way mathematical functions to create a representation of the features or characteristics from the source data – this digital representation is called a **template**. A template is essentially a digital representation of someone's unique biometric features.

Templates can be used in two major ways as summarized in [Table 5-10](#).

1 : N for Identification	1 : 1 for Authentication
Imagine someone walks up to a locked door and places their finger on a biometric scanner to unlock the door. The scanner has no idea who this person is before they place their finger on the scanner. The scanner is capturing the person's biometric data, generating a new template, and then searching through a database of existing templates to try to find a	Imagine a user logs into their laptop by typing in their username and password. As a second factor of authentication, the laptop then asks the user to place their finger on the built-in biometric scanner. The scanner will capture the user's biometric data, generate a template, and because the user has already been identified with their username, the system will look up the user's

<p>reasonably close match to identify the person.</p> <p>A 1 : N scan of biometric templates is used for identification</p>	<p>existing template in a database and compare the user's existing template to the newly generated template. If the templates reasonably match, then the user is authenticated.</p> <p>A 1 : 1 comparison of biometric templates is used for authentication</p>
--	--

**Table 5-10: Uses of Biometric Templates
Biometric Devices**

As can be seen in [Table 5-11](#), biometric devices vary and focus on different characteristics.

<p>Physiological</p>	<p>Fingerprint</p>	<p>Fingerprint scanners examine a fingerprint. These are common, especially on devices like computers and mobile phones, and they're reasonably accurate. They're also often used at border crossings, for example, between the United States and Canada.</p>
	<p>Hand geometry</p>	<p>Hand geometry scanners are rarely used, but often portrayed in movies. The scanning device could be a small machine, with an outline of a hand and guideposts that help position the hand and fingers for scanning; or, the device could be a futuristic-looking screen upon which a user places his or her hand, and the geometry is read. Some devices</p>

		scan the ridges of the hand, while others examine the geometry.
	Vascular pattern	Vascular pattern scanners examine veins in a hand and are often used in testing centers, like the ones where CISSP and other exams are administered. The scan allows biometric information of the exam candidate to be captured prior to the exam, and if a break or visit to the restroom is needed, the hand can be rescanned and the results compared to the original. This ensures that the exam candidate remains the test taker, and not somebody else.
	Facial	Facial scanners examine an individual's facial features and pattern.
	Iris	Iris scanners examine the colored ring around an individual's eye.
	Retina	While iris scanners examine the colored ring around an eye, retina scanners look at the back of the eye and specifically at the vein pattern of the retina. Retina scanners are very accurate; in fact, they're the most accurate type of biometric system. They're also controversial. For one thing, they're invasive, due to the way they operate, and this explains why they're rarely used. In

		use, an individual has to place their eye to a rubber eye cup, and a bright light is flashed as part of the scan. Most people find this process unpleasant. Additionally, and more to the point of controversy, retina scans can lead to privacy issues, because the results of a retina scan can reveal medical issues related to the individual.
Behavioral	Voice	How a person speaks.
	Signature	How a person writes.
	Keystroke	How a person types on a keyboard.
	Gait	How a person walks.

Table 5-11: Biometric Authentication

Biometric device types and least/most accurate

5.2.6 Factors of Authentication

CORE CONCEPTS

- **Factors of authentication refers to the three types of authentication: authentication by knowledge, authentication by ownership, authentication by characteristic.**

- **Single-factor authentication refers to any one of the three types of authentication being used.**
- **Multifactor Authentication (MFA) refers to two (or more) of the three types of authentication being used.**

Refer back to the three types—factors—of authentication:

- Authentication by knowledge—something you know
 - Authentication by ownership—something you have
 - Authentication by characteristic—something you are (physiological or behavioral)

Each of these families on its own can be thought of as a type of single-factor authentication. If an authentication system uses any number of authentication types but all falling within a single factor (e.g., all belong “to something you are”), then single-factor authentication is in place. However, two (or more) of these factor families used in combination can be considered as multifactor authentication (e.g., using any authentication type from “something you are” and another belonging to “something you have”). Single-factor versus multifactor authentication are summarized in [Table 5-12](#).

Single-factor Authentication	Multifactor Authentication
One (1) factor of authentication used by itself	Two (2) or more different factors used in

combination

Table 5-12: Factors of Authentication

If a user logs in to a system using only an RSA ID key and a Microsoft token, they've authenticated using a single factor. However, if the authentication process involves entering a password and an RSA token, then two factors of authentication have been used, and therefore this would be considered multifactor authentication.

Also, consider this question. If a username/password combination *and* a challenge question is used, is this single-factor or multifactor authentication? Before you answer, consider what type of authentication each represents. A username/password is authentication by knowledge; likewise, so is a challenge question. So, in fact, even though two authentication objects are utilized, only a single-factor of authentication has been employed (something you know). It's important to understand the different factors of authentication and to be able to distinguish them from each other.

Password-less authentication

CORE CONCEPTS

- Password-less authentication options like passkeys can reduce friction, limit weak passwords and mitigate against phishing.

As the name implies, password-less authentication involves **authenticating users without needing them to input a password**. Passwords come with a range of issues—entering them can add friction, users can forget passwords

and get locked out of their accounts, users may choose weak passwords that can be cracked, or users may fall victim to a phishing attack and have their passwords stolen.

Password-less authentication options can include biometrics, mobile devices owned by the user, or security tokens. One of the major password-less options to emerge in recent years is passkeys, which involves users authenticating themselves by entering their PIN or biometrics into their device. While tools like passkeys can be more convenient to users and are resistant to phishing, they come with their own set of downsides. As discussed in the earlier sections of Domain 5, biometrics have their own flaws. If a user loses their mobile device or their hardware token, they may end up locked out of their account. On top of this, password-less options like hardware tokens have higher implementation costs, due to the cost of the tokens.

5.2.7 Credential Management Systems

Credential management systems allow organizations to effectively manage—at scale—access to assets by ensuring that all personnel, processes, and devices have unique credentials. Credential management systems are designed to manage, grant, and revoke credentials, typically using strong two-factor authentication that incorporates public key infrastructure. Credential management systems include the programs, processes, technologies, and personnel used to create trusted digital identity representations of individuals and nonperson entities (processes) and bind those identities to their credentials.

Password Vaults

CORE CONCEPTS

- **Password vaults (password managers) allow us to generate, store, sync and manage unique passwords for each of our accounts, while only requiring the user to remember the single master password that unlocks the password vault.**
- **They can improve security by making it practical to create strong and unique passwords for each account. However, they present a single point of failure.**

Password vaults, also known as password managers, **are apps that are designed to generate, store, sync and manage passwords**. The passwords are kept in an encrypted database, which is guarded by a master password. The theory behind password vaults stems from the problem that most people can only remember a few passwords at most. This is a significant issue because security best practices involve having strong and unique passwords for every account. With a password vault, users can have strong and unique passwords for the dozens of accounts they might regularly use, without having to know or remember any of them. All they have to do is remember the master password so that they can log in to their password manager whenever they want to sign into an account.

Password vaults can improve security, because they make it more practical for users to create strong and unique passwords for each account, which defends against attacks like credential stuffing. However, they do also present a single point of failure. If an attacker manages to access someone's password

vault, the attacker may be able to access all of their accounts unless MFA is in place for each account.

5.2.8 Single Sign-on (SSO)

CORE CONCEPTS

- **Single sign-on refers to authenticating one time and being able to access multiple systems.**
- **A disadvantage of single sign-on is the implication of centralized administration, which represents a single point of failure.**
- **Kerberos is one of the major single sign-on protocols, and it provides accounting, authentication, and auditing services.**
- **SESAME is an improved version of Kerberos, but it has not been widely adopted due to Kerberos being built into Microsoft Windows operating systems by default.**
- **Kerberos disadvantages include that it only supports symmetric encryption, and it is vulnerable to TOCTOU attack.**

Understand the underlying premise and pros and cons of single sign-on

The concept of Single Sign-On (SSO) is best illustrated through an example. From the perspective of a user, single sign-on takes place when the user types in their username and password, or username and Microsoft Authenticator code, for example, and is then authorized to access multiple systems. The user logs in one time and is authorized to access multiple systems.

Users typically love it, and one immediate advantage is they'll be more likely to use one stronger password to log in once versus using a bunch of weaker passwords to access multiple systems.

A big disadvantage of SSO was mentioned in the section about administration approaches. SSO implies centralized administration, and centralized administration represents a single point of failure from both an availability and a confidentiality perspective. If a SSO system is compromised, an attacker potentially has access to everything. Contrarily, if the system goes down, users have access to nothing.

At a high level, the single-sign-on process is depicted in [Figure 5-5](#).

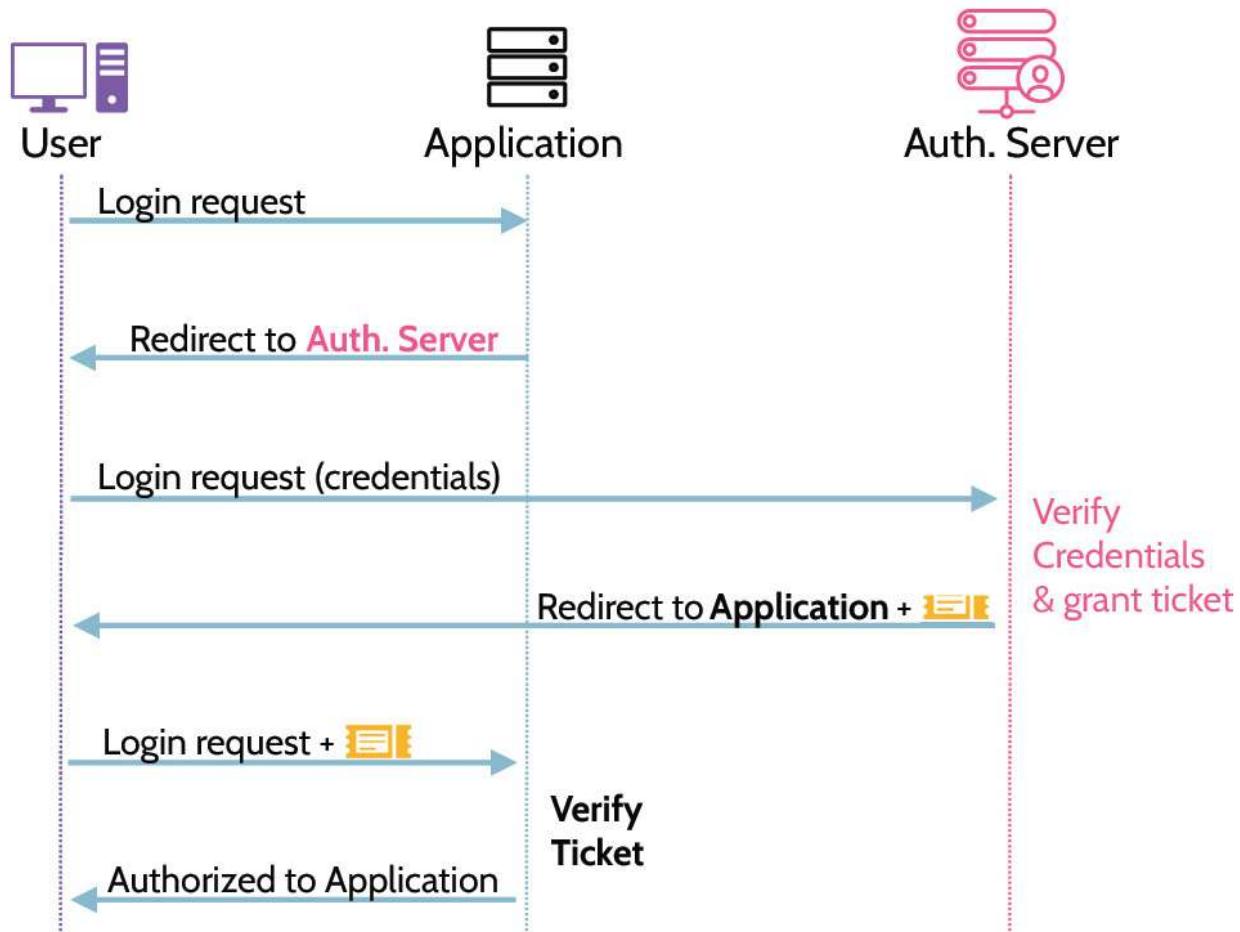


Figure 5-5: Single Sign-On Process

1. A user sends a login request to an application.
2. If the user has not already logged in or authenticated, the application will essentially say, “I don’t know who you are right now,” and redirect the user back to the authentication server, saying, “You’re not currently authenticated, I don’t know who you are, you need to go and authenticate.”

3. The user will identify themselves to the authentication server and authenticate via knowledge, ownership, or characteristic, or a combination of two or more types. Once authenticated and authorized, the user will be given some type of ticket or token.

4. The user is directed back to the application, and the ticket or token is presented for authorization to the application.

5. If the application grants authorization, the user will be able to access the application.

Pros and Cons of Single Sign-On

The pros and cons of single sign-on are summarized in [Table 5-13](#).

Pros	Cons
<ul style="list-style-type: none"> ■ User experience ■ Users may create stronger passwords ■ Timeout and attempt thresholds enforced ■ Centralized administration 	<ul style="list-style-type: none"> ■ Single point of failure for compromise and availability ■ Inclusion of unique/legacy systems

**Table 5-13: Pros and Cons of Single Sign-On
Kerberos**

Understand how Kerberos works as well as specific components of Kerberos

One of the major SSO authentication protocols is known as Kerberos. As a side reference, the name Kerberos stems from Greek Mythology and refers to the three-headed dog, Cerberus, that guarded the gates of Hell. Drawing from the myth, Kerberos protects access to resources and provides three primary functionalities:

- Accounting
- Authentication
- Auditing

Kerberos is an old and complicated protocol, and thankfully you do not need to be an expert on how the protocol works for the exam. [Figure 5-6](#) provides a simplified depiction of the major steps in the Kerberos authentication process. Alice is the client who wants to access a service. If Alice does not currently have a valid ticket, she must first be authenticated by the Kerberos service before she can access the desired service.

- Alice, the **Client**, sends a couple of initial messages to the **Kerberos Authentication Service (AS)**.
- The Authentication Service will verify that Alice is a valid user and, if so, will send a few messages back to Alice. One message is encrypted with Alice's password as the encryption key, and another message is the **Ticket Granting Ticket (TGT)**—a message that Alice cannot decrypt as it is encrypted with the **Ticket Granting Service's (TGS)** key.
- When Alice receives the messages from the authentication service, she decrypts one of the messages with her password as the encryption key. This is the way that Kerberos verifies that the user knows their password without having to send the user's password across the network. If the user knows their password, they can use

it to decrypt one of the messages and can therefore proceed with the process. If the user doesn't know their password, they can't decrypt the message and obtain the information they need to proceed.

- Assuming Alice knows her password and decrypts the message, she will create a couple of new tickets and send them along with the still encrypted TGT to the Ticket Granting Service.
- When the Ticket Granting Service receives the messages from Alice, it performs a number of verification steps, and if everything looks good, it will create some new messages to send back to Alice including the encrypted **Service Ticket** (encrypted with the service's key).
- When Alice receives the messages from the Ticket Granting Service, she creates some new messages and sends them to the **Service** along with the still encrypted service ticket.
- When the service receives the messages from Alice, it does a final few verification steps, and if everything looks good, the service will finally grant Alice access.

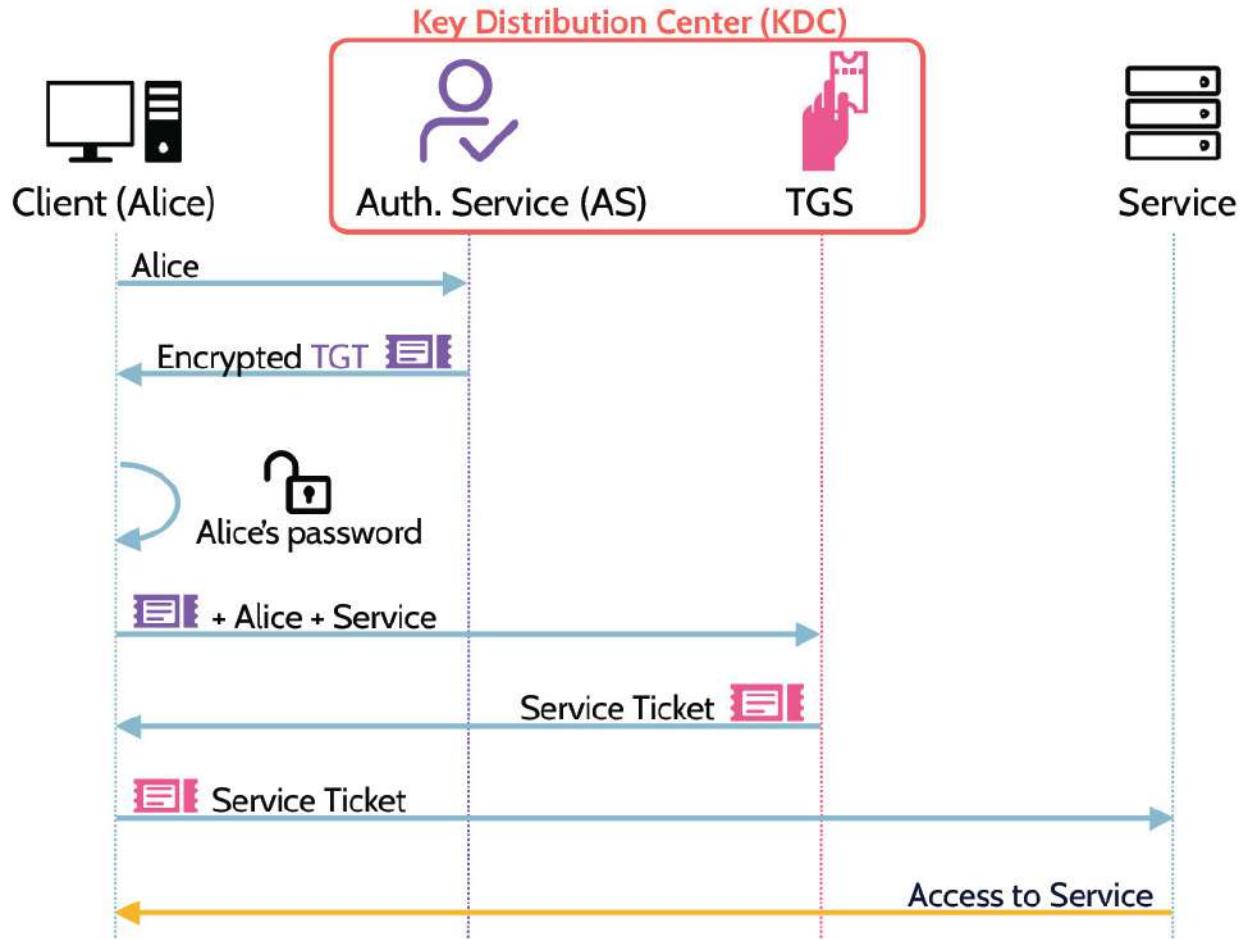


Figure 5-6: Kerberos Operation

This is a very incomplete description of all the various messages that are sent, the data they contain, and all the verification steps performed at each stage; however, it should provide a sufficient overview of the critical messages (TGT and Service Ticket), and the major services of Kerberos—the Authentication Service and the Ticket Granting Service—both of which are components within what is known as the **Key Distribution Center (KDC)**.

Despite Kerberos's inherent strengths, including that it enables single sign-on, some significant disadvantages exist too. For one thing, Kerberos only

supports symmetric encryption (e.g., RC4, DES, AES), which automatically implies symmetric key distribution challenges. For another, Kerberos only issues one major ticket, which is used to gain access to a system. As a result of using only one ticket, a system is vulnerable to a Time Of Check Time Of Use (TOCTOU) attack, which can be mitigated by increasing the frequency of authentication. In other words, to minimize the risk associated with a TOCTOU, or session-hijacking attack, users should be re-authenticated more frequently, especially if the system is a high-value system. Re-authentication means expiring the ticket, which means the user must login again. The point is this: for users accessing low-value systems all day, forcing them to re-login frequently, because of a high-value system in the environment, can create a significant burden and end-user resistance. In this scenario, it would be better to isolate the high-value system(s) and allow users to have longer-life-span tickets for the rest.

SESAME

Secure European System for Applications in a Multi-Vendor Environment, better known as SESAME, is an improved version of Kerberos. Like Kerberos, SESAME is a protocol for enabling single sign-on. Additionally, one of the big advantages of SESAME over Kerberos is that it supports symmetric and asymmetric cryptography, so it naturally solves the problem of key distribution, and it issues multiple tickets, which mitigates vulnerability to attacks like TOCTOU.

Even though SESAME is a better protocol, Kerberos is by far the more prevalent, because it's built into many prevalent systems including the Windows operating systems, MacOS, and various Linux and Unix distros.

Remember that to use Kerberos in a Windows environment, Active Directory must be enabled.

5.2.9 CAPTCHA

CORE CONCEPTS

- CAPTCHA is a security measure that works by asking a user—typically a visitor to a website or portal—to complete a simple test to prove they're human and not a robot or automated program.
- CAPTCHA is used to prevent automated account creation, spam, and brute-force password decryption attacks.

Understand what CAPTCHA is and why it is most often used

When accessing a website, vendors often employ the use of what is known as Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) as a security measure to protect against automated account creation and to protect users from spam and brute-force password decryption attacks. CAPTCHA works by asking a user to complete a simple test to prove they're human and not a robot or automated program trying to access or break into a protected account or area of a website. In its simplest form, CAPTCHA works by presenting a website visitor with an image of distorted letters and numbers and asking the user to type those letters and numbers into an empty field on the page. If the information is typed in correctly, the visitor gains access to the protected area; if not, the visitor is typically given another opportunity to do so. Bottom line: CAPTCHA is used

to prevent bots from creating multiple accounts on systems, spam, and unauthorized access.

5.2.10 Session Management

CORE CONCEPTS

- **Session management refers to management of sessions created through a successful user identification, authentication, and authorization process.**
- **A session represents the connection and interaction between a user and a system.**
- **Session hijacking is a risk where no session management exists.**
- **Session termination and re-authentication is the best way to prevent or mitigate session hijacking.**

Session management refers to sessions, and a **session** is what's created as the result of a successful user identification, authentication, and authorization process as shown in [Figure 5-7](#). Upon a user being authorized by a system, a session is created, and the session represents the connection and interaction between the user and the system. Until the user logs out—manually or automatically—a session remains intact. Session management is focused on managing sessions effectively and securely for the entire duration.

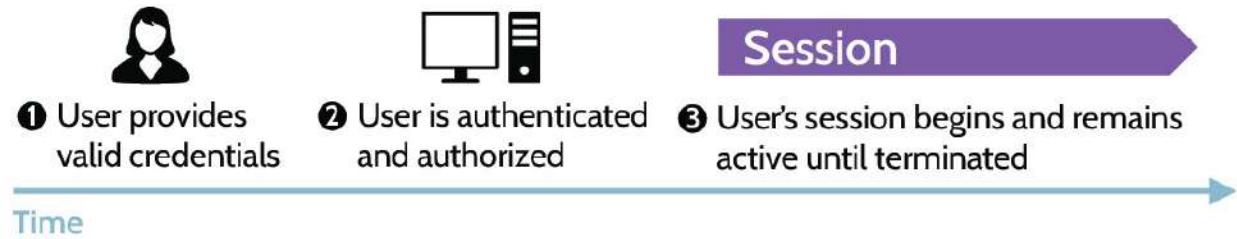


Figure 5-7: User Successful Authentication

Session Hijacking

What is session hijacking?

Session management is very important, because without it a major risk exists: session hijacking (shown in [Figure 5-8](#)). In other words, through simple carelessness or sophisticated technical means, somebody other than a legitimate user could gain access to a session and use it for malicious purposes. The preventive measure for session hijacking is session termination, which is an important component of session management.

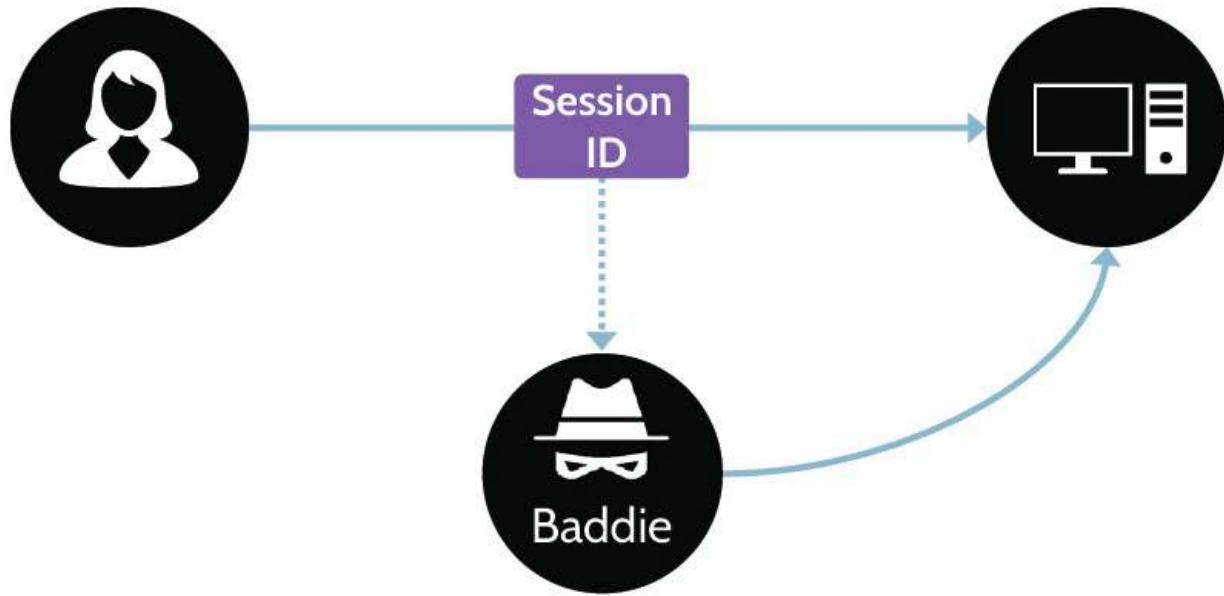


Figure 5-8: **Session Hijacking**

How do you prevent session hijacking?

How can session hijacking be prevented?

As noted, session hijacking can be mitigated through session termination, and several major ways exist to terminate a session. The primary and best way to prevent session hijacking is through frequent re-authentication. Many VPN solutions include continuous re-authentication as part of their security package. Session encryption keys are established at the beginning of a VPN session and then re-established in the background at certain time intervals. Additionally, a user is continually re-authenticated by the system in a manner that is transparent to the user. That makes it much more difficult for an attacker to compromise a user's active session.



Session Termination

In addition to continuous authentication, other major ways to terminate a session are outlined in [Table 5-14](#).

Schedule Limitations	Schedule limitations refers to a system administrative control that logs all users out of a system at a set time, for example, at 5 p.m. every evening, or perhaps the system does not allow logins during a weekend.
Login Limitation	Login limitation refers to preventing more than one simultaneous login using the same user ID. In other words, an account may not be shared and used at the same time by different people (or even the same person).
Time-outs	If there's inactivity, or after a set period of time, a session expires (it's timed out).
Screensavers	Screensavers are another popular session-management tool. After a screensaver pops up on a computer, typically the only way to gain access is through re-authentication.

Table 5-14: Session Termination Methods

5.2.11 Registration and Proofing of Identity

CORE CONCEPTS

- **Identity proofing—registration—is the process of confirming or establishing that somebody is who they claim to be.**
- **Identity proofing is a component of provisioning in the identity life cycle.**

What is identity proofing, and when does it take place?

Identity proofing, also sometimes called registration, is simply the process of confirming or establishing that somebody is who they claim to be before they are given access to a valuable resource or asset. Remember, for example, what happens before a certificate authority (CA) issues a digital certificate to somebody.

The registration authority (RA) proofs the identity of the certificate applicant. Similarly, prior to beginning employment and issuing an employee badge, account credentials, etc., an organization will proof the identity of a new employee. Using some type of government issued ID, a driver's license, or other form of identification unique to an individual, the company will confirm that the person is who they claim to be.

5.2.12 Authenticator Assurance Levels (AAL)

CORE CONCEPTS

- Authenticator Assurance Levels (AAL) refer to the strength of authentication processes and systems.
- AAL levels rank from AAL1 (least robust) to AAL3 (most robust).

Understand AAL ratings and elements of each

With regards to digital identities, the National Institute of Standards and Technology has produced a suite of documents that can be found here: <https://pages.nist.gov/800-63-3/>. One document, NIST SP 800-63B, entitled “Authentication and Lifecycle Management,” contains information about the AAL levels, which has been summarized in [Table 5-15](#).

AALs measure the robustness of the authentication process. The higher the number, the more robust the strength of the service provided.

AAL1	■ Some assurance ■ Single-factor authentication ■ Secure Authentication protocol
AAL2	■ High confidence ■ Multifactor authentication ■ Secure Authentication protocol ■ Approved cryptographic techniques
AAL3	■ Very high confidence ■ Multifactor authentication ■ Secure Authentication protocol ■ “Hard” cryptographic authenticator providing proof of possession of key and impersonation resistance

Table 5-15: AAL Levels

5.2.13 Federated Identity Management (FIM)

CORE CONCEPTS

- Single sign-on refers to one-time authentication to gain access to multiple systems in one organization; federated identity management (FIM) refers to one-time authentication to gain access to multiple systems, including systems associated with other organizations.
- Federated Identity Management (FIM) relies on trust relationships established between different entities.
- FIM trust relationships include three components: principal/user, identity provider, relying party.
- Principal/user = the person who wants to access a system.
- Identity provider = the entity that owns the identity and performs the authentication.
- Relying party is also known as the service provider.

Understand the basis of Federated Identity Management (FIM) and the three components that make up any federated access system

In comparison to single sign-on (SSO), where a user authenticates one time and gains access to multiple systems in the context of an organization, federated identity management (FIM) allows a user to authenticate one time and gain access to multiple disparate systems. In other words, a user gains access to company-owned systems as well as systems outside of the organization's control.

Microsoft's Active Directory is one example of the type of system used within an organization to provide SSO services.

Let's take a closer look at federated identity management, sometimes referred to as federated access, through an example. When a person travels via airplane, they must go through a security checkpoint before proceeding to the departure gate. Passing through this checkpoint means the traveler is in a secure zone. After traveling to another location, the person is still in a secure zone, because the new airport trusts the security check that was performed at the original airport. This fact highlights one of the most important and foundational aspects of federal access—*trust relationships between different entities*. In this example, the two airports are owned and operated by different organizations, but they share a trust relationship.

Let's look at federated access in the context of the logical world. With many websites today, when creating an account, two or more options are often available. One option is to create an account using a unique username and password; another option is to create an account using an already existing Facebook, Google, or account from a similar platform. For this example, imagine a user prefers to use their Google account, and they want to create an account on Pinterest, but Google doesn't own it. So the user visits Pinterest, and they're given the option to create an account or log in with Google (among several choices). They choose to log in via Google, and a small window pops up asking them to provide their Google username and password. Google is authenticating the user, but because a trust relationship exists between Google and Pinterest, Pinterest is trusting the authentication

being performed by Google. This is another, real-world example of federated access in the logical world.

With any federated access system, three major components exist as depicted in [Figure 5-9](#). First is the **user**, also referred to as the **principal**. The user or principal is the person who wants to log in or access the system. Second is the **identity provider**. The identity provider is the entity that owns the identity and performs the authentication. In the example above, Google is the identity provider. Third is the **relying party**, sometimes called the **service provider**. In the example noted above, Pinterest is the relying party. Federated identity management—federated access—relies on a trust relationship between the three entities.

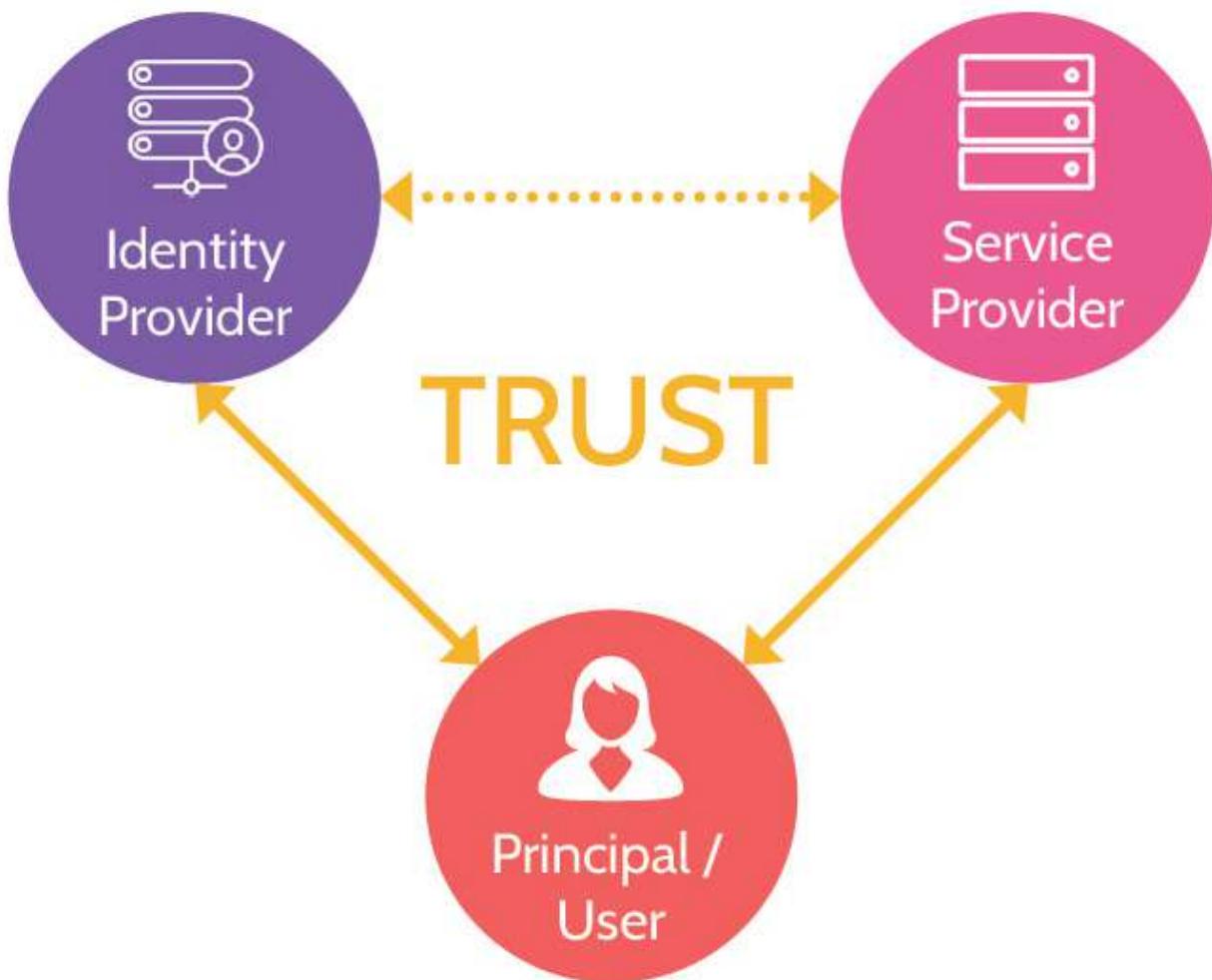


Figure 5-9: Federation Components

5.2.14 Federated Access Standards

CORE CONCEPTS

- Key Federated Access protocols include: Security Assertion Markup Language (SAML), WS-Federation, OpenID (for authentication), OAuth (for authorization).
- SAML is frequently used in Federated Identity Management (FIM) solutions and provides authentication and authorization.

- **OpenID and OAuth** are open-standard federated access protocols that provide authentication via OpenID and authorization via OAuth.
- **SAML assertions** are written in a language called XML, or Extensible Markup Language. XML is a way of communicating in a manner that is machine and human-readable.

Several major protocols enable federated access, with Security Assertion Markup Language (SAML) being one of the most important to understand. WS-Federation, OpenID, and OAuth are the others that should be known at a high level. [Figure 5-10](#) depicts these four federated access standards and whether they provide authentication or authorization services or both.

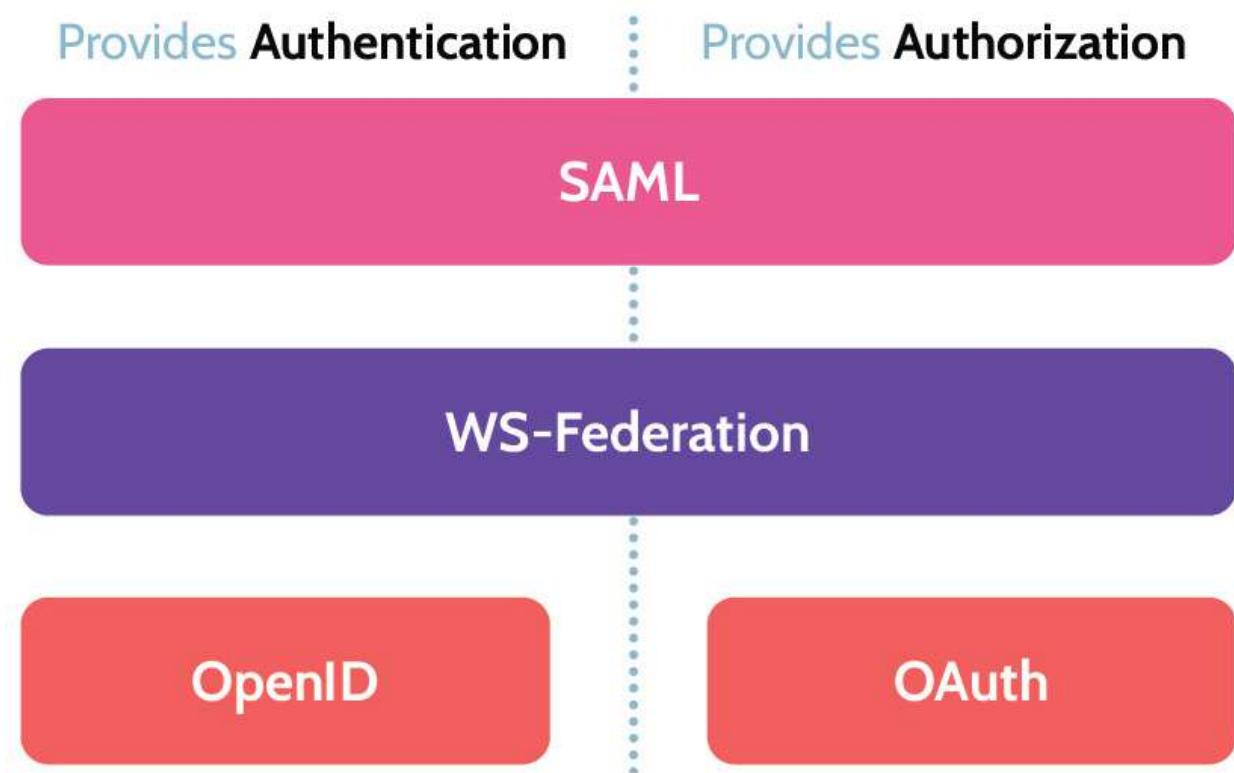


Figure 5-10: Federated Access Standards

WS-Federation (like SAML) offers authentication and authorization functionality. Like most federated access standards, the primary goal is enabling identity federation authentication and authorization. WS-Federation was created by a consortium of companies, including IBM, Microsoft, and Verisign, and it was codified as a standard by **OASIS**.

OpenID and OAuth are complementary protocols that often work together. **OpenID** provides the authentication component, and **OAuth** provides the authorization component to a federated access solution. In its simplest form, OpenID allows a user to use an existing account to identify and authenticate to multiple disparate resources—websites, systems, and so on—with the need to create new passwords for each resource. With OpenID, a user password is given only to the user's identity provider—Microsoft, for example—and the identity provider confirms the user's identity to sites the user visits. OAuth is the protocol or standard that allows users to gain access—to be authorized—to resources. Both OpenID and OAuth are open standards. While they can work independent of each other—especially OpenID—they're often deployed together, because of the richer functionality they provide as a unit.

Security Assertion Markup Language (SAML)

SAML's operation is depicted in [Figure 5-11](#).

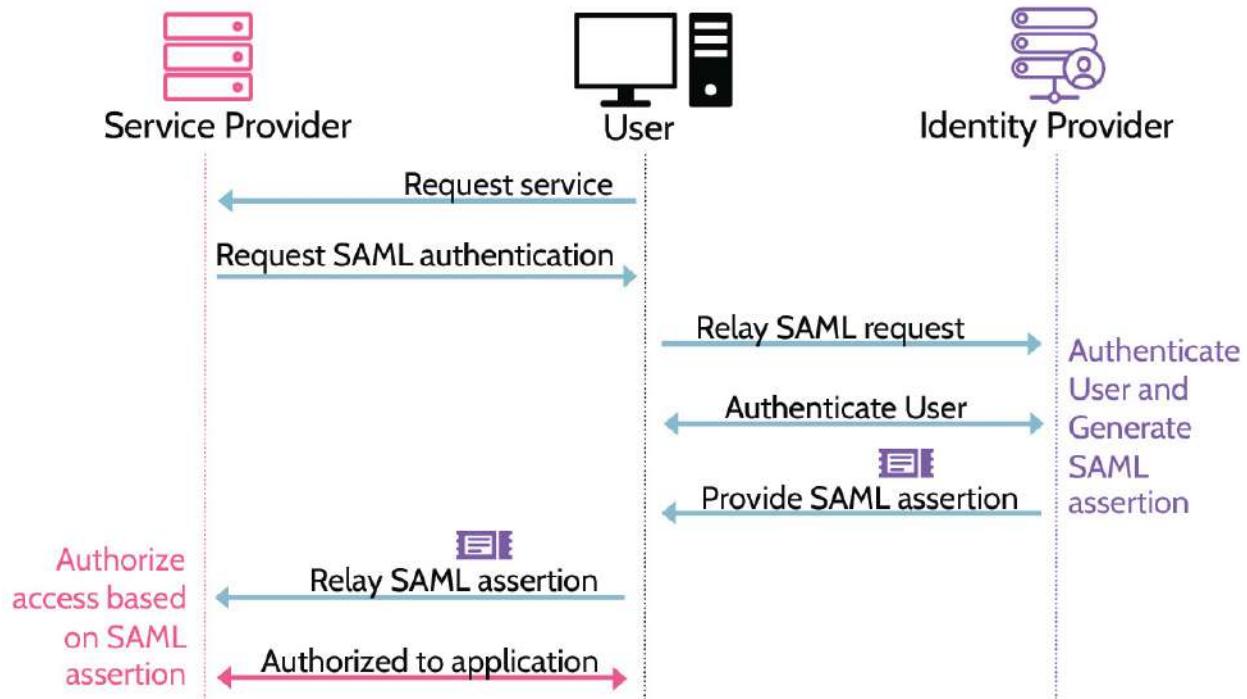


Figure 5-11: SAML Operation

Understand the importance of SAML & its relation to Federated Identity Management (FIM)

SAML provides two capabilities: authentication and authorization.

1. First, the user (**principal**) must authenticate via the identity provider. If the user is not logged in and requests access to a service (offered by the service provider), the request will get bounced to the identity provider, where the user can authenticate.
2. The **identity provider** will authenticate the user through the process of identification and authentication, at which point the user will be

issued a SAML assertion ticket. One critical fact to note here: the SAML assertion ticket does not contain the username and password of the user. Rather, as the name suggests, the ticket contains assertion statements that the service provider—the relying party—can use for authorization purposes or to determine the level of authorization granted to the user.

- Once the SAML assertion ticket is provided to the user, the user will pass it on to the **service provider**. The relying party is going to read the assertion statements contained within the SAML ticket and make an authorization decision. Similar to Kerberos, SAML uses tickets or tokens, usually denoted as SAML assertion tickets or tokens. The words are used interchangeably, and the critical thing to note is that assertions or statements about the user—username, role, level of access, etc.—are contained within them.

The four major components of SAML are summarized in [Table 5-16](#).

Component	Function
Assertion	Authentication, authorization, and other attributes
Protocol	Defines how entities request and respond to requests
Bindings	Mapping of SAML onto standard communication protocols (ex: HTTP)

Profiles

Define how SAML can be used for different business use cases (ex: Web SSO, LDAP, etc.)

Table 5-16: **SAML's Key Components**

In addition to the above, it's important to remember two key characteristics of SAML.

- SAML uses assertion tickets or tokens.
- Assertions are written in a language called **Extensible Markup Language (XML)**, which is a way of communicating in a manner that is machine and human-readable.

5.2.15 Accountability = Principle of Access Control

CORE CONCEPTS

- Accountability = the Principle of Access Control.

Understand what the phrase Principle of Access Control means; understand that accountability refers to the Principle of Access Control

The Principle of Access Control refers to accountability.

In order to achieve accountability, several things need to happen: 1. Users must be uniquely identified

2. Users must be properly authenticated
3. Users must be properly authorized
4. All actions should be logged and monitored

With all these components in place, then, and only then, can the Principle of Access Control be achieved.

5.2.16 Just-in-time (JIT) Access

CORE CONCEPTS

- Just-in-time access refers to the elevation of user privileges to an authorized user for a short period of time, so a user may complete necessary, but infrequent, tasks.
- Just-in-time access mitigates the need for long-term elevation of privileges, which minimizes potential security risks.

The term “just-in-time” is often used in the context of just-in-time delivery, meaning an organization—a manufacturer, for example—receives components needed for production at the time production commences. One of the huge benefits of this type of arrangement is that the manufacturer can focus on what it does best and not need to worry about managing and storing inventory. If they know they’re going to produce 100,000 widgets, they’ll get enough components (and some extra, just in case) to produce those 100,000 widgets at the time needed.

Just-in-time access works in a similar fashion, albeit from a security perspective. Imagine a user that needs to access a sensitive part of a database once a month to run a report. At a high level, just-in-time access allows the user to gain elevated privileges during this monthly time window to run the report. Just-in-time access mitigates the need for long-term elevation of privileges, and the way it is set up and administered oftentimes allows for the access to be granted in an automated fashion versus a manual process. In other words, it minimizes potential security risks, and it does so in a manner that is efficient and effective.



5.3 Federated identity with a third-party service

5.3.1 Identity as a Service (IDaaS)

CORE CONCEPTS

- Identity as a Service (IDaaS) refers to the implementation or integration of identity services in a cloud-based environment.
- Risks of IDaaS include those related to availability of service, protection of critical identity data, and trusting a third party with potentially sensitive or proprietary information.

Understand the basic premise underlying Identity as a Service and why it might be used as well as associated risks

Identity as a Service (IDaaS) is the implementation or integration of identity services in a cloud-based environment. In other words, identification, authentication, authorization, accounting, and federated access all take place in the cloud. IDaaS has a variety of capabilities, which are:

- Provisioning
- Administration
- Single Sign-on (SSO)
- Multifactor authentication (MFA)
- Directory Services
- On premises and in the cloud

It also supports multiple types of identities/accounts as listed in [Table 5-17](#).

	Account Stored	Authentication by
Cloud Identity	Created and managed in the cloud	Cloud service

Synced Identity	Created and managed in local store (e.g., active directory) and synced/copied to cloud or vice versa	Either local or cloud
Linked Identities	Two separate accounts which are linked . For example: one account in local store and second account in cloud service	Either local or cloud
Federated Identity	Identity Provider	Identity Provider

Table 5-17: IDaaS Identities

Identity and Access Management Solutions

Identity and Access Management (IAM) solutions can use any of the three models listed in [Table 5-18](#).

On Premise	<ul style="list-style-type: none"> ■ Systems controlled by a private organization ■ They are not reliant on the internet to function ■ Are typically very secure
Cloud	<ul style="list-style-type: none"> ■ Service and systems provided by a cloud service provider ■ Through the use of Federated Identity protocols (like SAML), organization user identities and credentials can be used ■ Can be subject to availability risk as well as security risks due to the multitenant nature of the public cloud
Hybrid	<ul style="list-style-type: none"> ■ Hybrid IAM solutions combine the best features of on premises and cloud, and offer the most flexibility for dynamic and growing organization

Table 5-18: IAM Models

IDaaS Risks

Potential risks relating to IDaaS include the following:

- **Availability of the service:** If the cloud service provider suffers an outage or the service is otherwise unavailable, the users will be unable to access systems.
- **Protection of critical identity data:** PII and other sensitive data will be in the control of the cloud service provider. That means adequate protection of data is based on protection mechanisms the provider has available.
- **Entrusting a third party with sensitive or proprietary data:** Based upon the identity data shared with the cloud service provider, other information about the organization might also be gained. Protections need to be in place to protect against this information from being leaked or shared with any unauthorized parties.

5.4 Implement and manage authorization mechanisms

CORE CONCEPTS

- **Discretionary Access Control (DAC)** means an asset owner determines who can access the asset; access is given at the *discretion* of the owner.
- **Rule-based access control** is based upon rules and can be utilized in a very granular manner, though it is very administrative-heavy as a

result.

- **Role-based access control** is based upon roles or job functions, and users can be assigned to one or more roles that include authorizations to perform duties.
- **Attribute-based access control** is very granular and is based upon user attributes, such as job function, type of device, working hours, asset classification, and so on.
- **Other access control approaches** include: context-based access control and risk-based access control. Context-based access control typically looks at the context (internal or external) of an initiating connection and is usually enforced via firewall rules. Risk-based access control looks at elements of a user connection—IP address, time of access request, and so on—to determine a risk profile associated with the request. Based upon the result, further authentication challenges may be presented to the user before access is granted.
- **eXtensible Access Control Markup Language (XACML)** is one tool that defines and enables attribute-based access control.

Within the realm of **authorization**, a number of different philosophies and methodologies exist, and these can be broadly categorized as:

- Discretionary
- Mandatory
- Non-discretionary

Each of these categories are analyzed in detail in the following section. They're also illustrated in [Figure 5-12](#), while their main characteristics are listed in [Table 5-19](#).

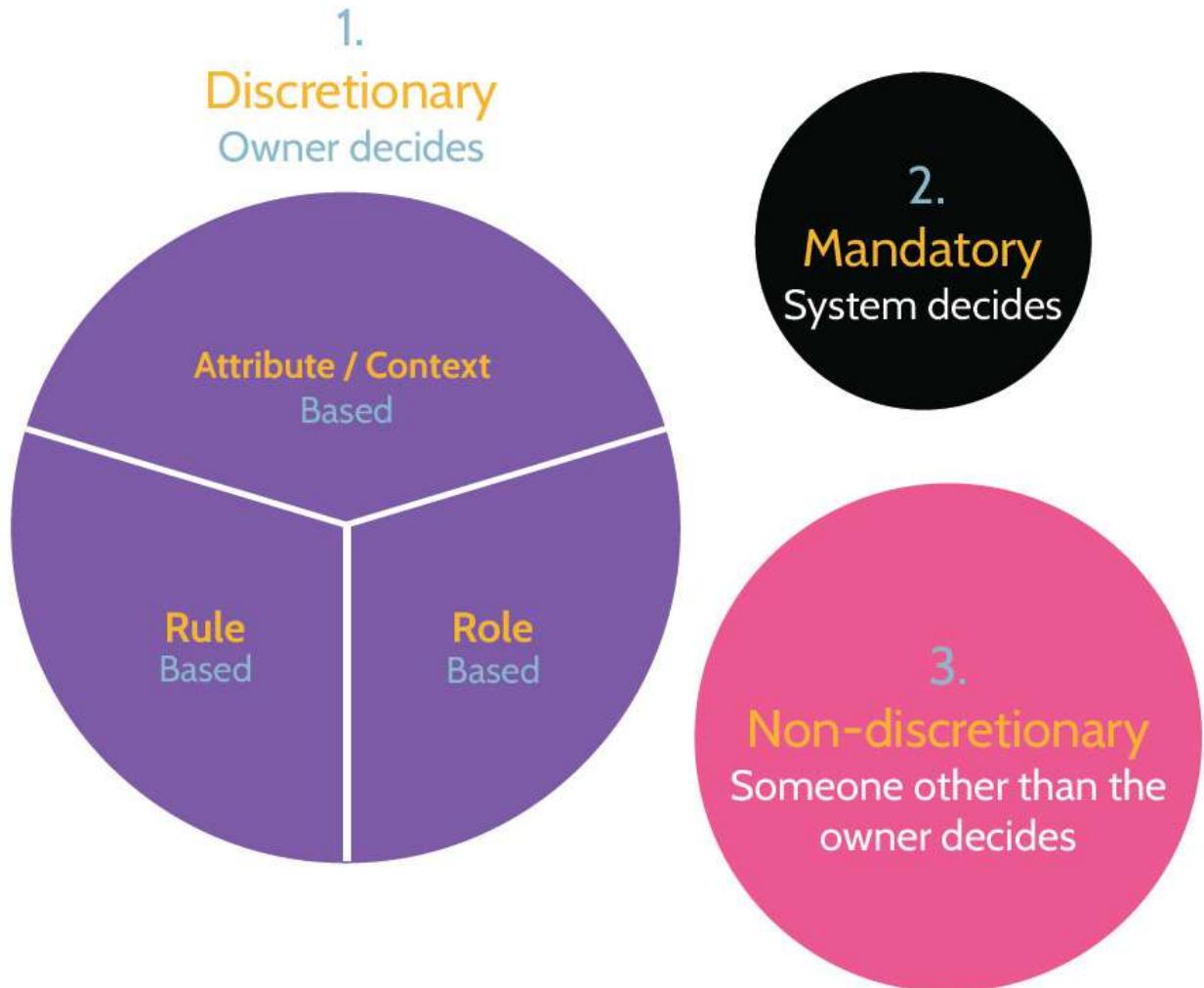


Figure 5-12: Access Control Types

Access Control Summary

Discretionary Access Control (DAC)	<ul style="list-style-type: none"> ■ Owner determines access rules
Role-Based Access Control (RBAC)	<ul style="list-style-type: none"> ■ Access to resources is based on user roles (e.g., firewall administrator, or accounts payable clerk)
Rule-Based Access Control	<ul style="list-style-type: none"> ■ Access to resources is based on a set of rules (e.g., an Access Control List, ACL)

Attribute-Based Access Controls (ABAC)	■ Access to resources is based on user attributes (e.g., OS, browser version, IP address)
Mandatory Access Control (MAC)	■ System determines access rules based on labels
Risk-Based Access Control	■ Considers elements of a user connection (IP address, time of access request) to determine a risk profile associated with the request. Based upon the result, further authentication challenges may be presented to the user.

Table 5-19: Main Characteristics of Access Control Types

With this foundation, let's explore each of these authorization mechanisms further.

5.4.1 Discretionary Access Control (DAC)

CORE CONCEPTS

- **Discretionary Access Control (DAC)** means an asset owner determines who can access the asset; access is given at the *discretion* of the owner.
- Three primary types of DAC exist: rule-based access control, role-based access control, and attribute-based access control.

Understand the premise of discretionary access control and the three primary types of DAC

Discretionary Access Control (DAC), as the word *discretionary* implies, means *somebody* determines who can access an asset. That *somebody* is the owner. *The defining characteristic of DAC is that access is given, based upon the owner's discretion* as also shown in [Figure 5-13](#). That's considered a great security best practice, since owners are accountable for and are therefore in the best position to determine who should access those assets.

REFERENCE MONITOR CONCEPT (RMC)

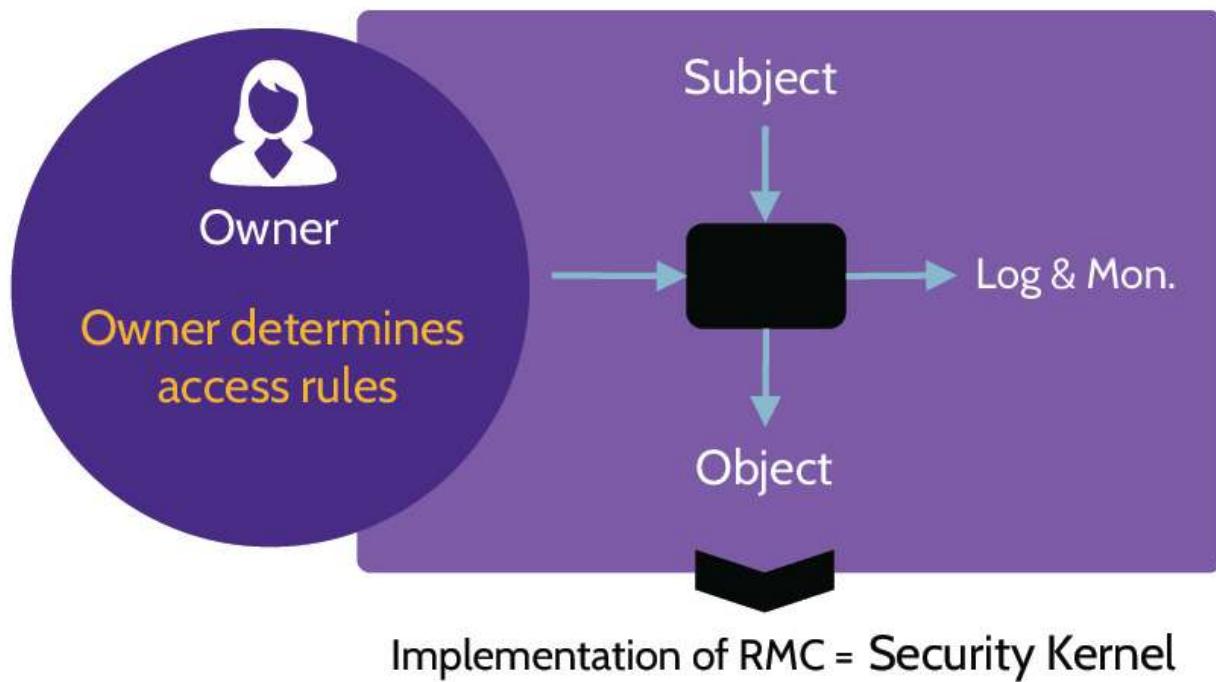


Figure 5-13: **DAC Operation**

Within the realm of discretionary access control, three primary types of DAC exist: 1. **Rule-Based Access Control:** Access to an object by a given subject is based upon one or more rules, determined by the owner.

- 2. Role-Based Access Control (RBAC):** Access to an object by a given subject is based upon the role, or job function, and related authorizations needed to perform duties.
- 3. Attribute-Based Access Control:** Access to an object by a subject is much more granularly controlled and based upon attributes, such as job function, type of device being used to access the object, time of day, classification of the asset, and so on.

Understand primary types of discretionary access control and their differences, why each might be used, and pros and cons of each

Rule-Based Access Control

Rule-based access control is simply a set of rules assigned to users. Look at [Table 5-20](#). The ruleset governs access to the numerous objects. As such, if a subject requires access to an object, a rule will be present to allow or deny that. For example, there's a rule in place to only allow Alice read access to Bob's directory and nothing more while she has both read and write access to her home directory.

User	Resource	Read	Write	Execute
Alice	24th floor printer			✓
Alice	Alice's home directory	✓	✓	

Alice	Bob's home directory	✓		
Alice	Historical finance data	✓		
Alice	Finance database	✓	✓	✓
Alice	CRM	✓		
Bob	24th floor printer			✓
Bob	Bob's home directory	✓	✓	
Bob	Marketing data	✓	✓	
Malory	Alice's home directory	✓		
Malory	Bob's home directory	✓		

Table 5-20: **Rule-Based Access Control**

Note that while rule-based access control offers much more granular control, it also requires much more administrative effort.

Role-Based Access (RBAC)

With role-based access control, users can be assigned to one or more roles and access is determined by a given role. The significant advantage gained by role-based access control is simplified and more manageable user and permissions administration. Instead of managing permissions at the user level, only permissions at the role level need to be considered, and then users with appropriate needs simply need to be assigned to the role or roles.

Typically, RBAC roles will mirror the structure or organizational chart of an organization and, based upon this fact, the use of RBAC is often considered a “best practice.” For example, in an organization with five hundred people who provide call center services, access needs are likely the same for each person. The use of RBAC in this case can eliminate much redundant and administrative overhead. Contrarily, in a more complex environment, where many more roles and cross-functional needs exist, an organization could easily end up with more roles than users and therefore much more complex RBAC needs.

[Figure 5-14](#) shows an RBAC operation example, but the primary point to take away from this is that RBAC provides the ability to assign privileges to users with minimal administrative overhead.

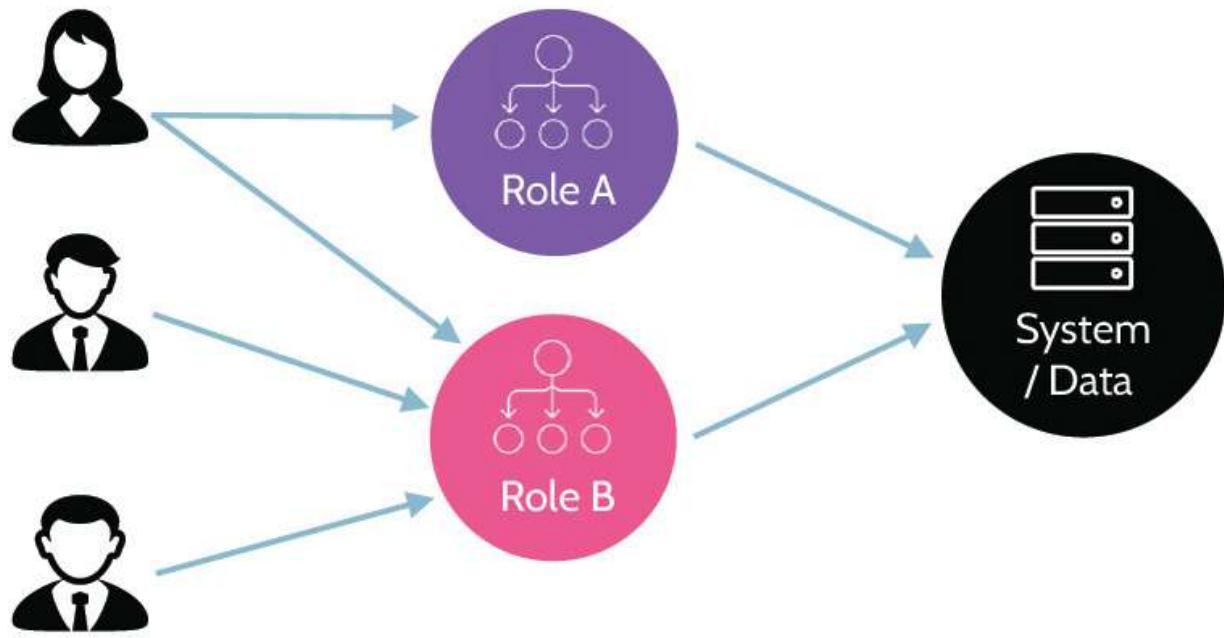


Figure 5-14: **RBAC Operation**

Implementing Full-RBAC across an entire organization is often very difficult and counterproductive. One of the main reasons to implement RBAC is to reduce access administration burden. When Full-RBAC is implemented across an entire organization for every system, it is common to end up with more roles than employees. Hence, Full-RBAC may increase the access administration burden.

This is one of the reasons most organizations implement Limited or Hybrid-RBAC, depicted in [Figure 5-15](#).

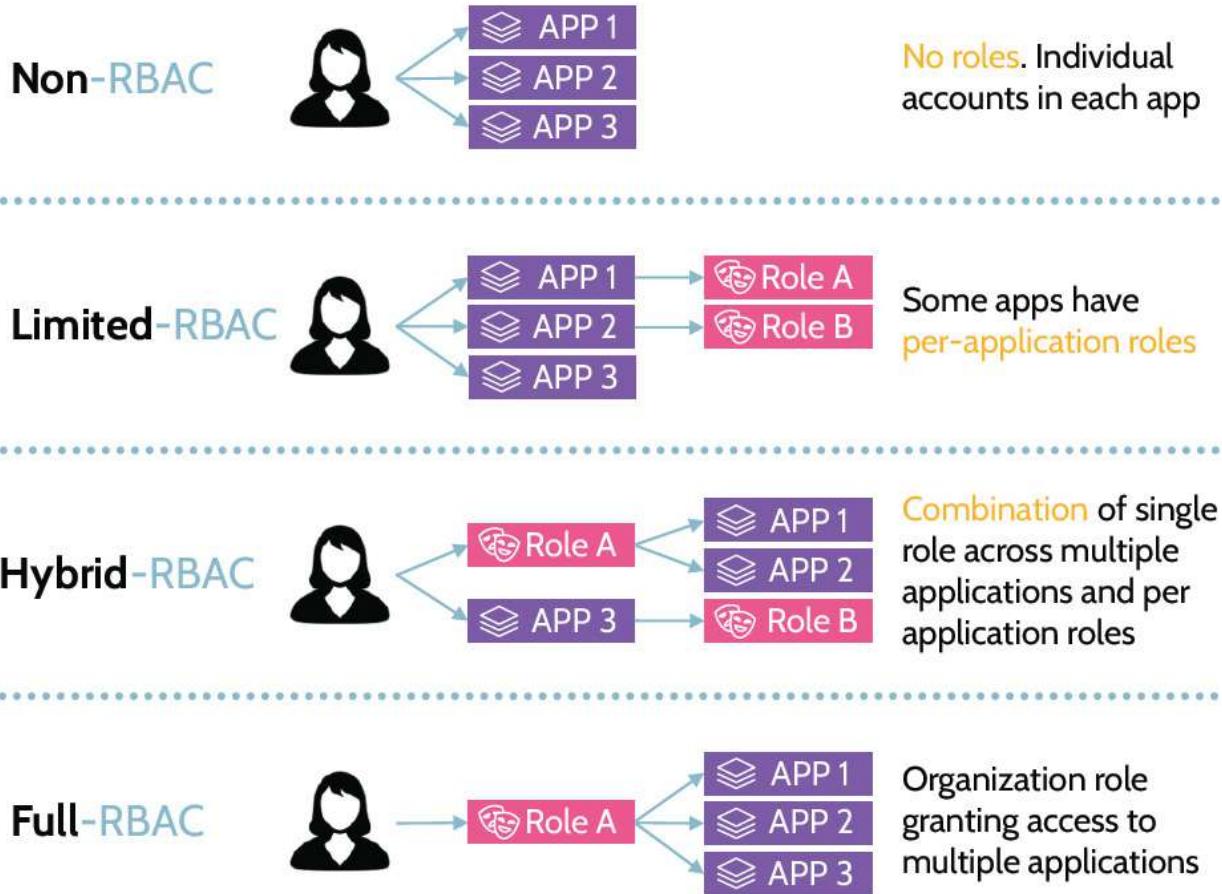


Figure 5-15: Different RBAC Models

Attribute/Context-Based Access Control

Attribute- or context-based access control is interesting and becoming quite prevalent. The premise behind the increased usage of attribute-based access control is the need to do a better job of authenticating user access with regards to the cloud, as most cloud-based applications are web applications. One of the defining characteristics of cloud computing is broad network access, which implies the ability to access cloud applications from anywhere with any type of device. If a company develops a web-based application and hosts it in the public cloud, is it protected by the corporate firewall and

hidden inside the corporate network? No, it's on the public internet and potentially accessible by anybody. Because of this fact and because so many companies are moving to the cloud and hosting important applications there, the need for better authentication and authorization to those applications is imperative. ABAC's operation is shown in [Figure 5-16](#), where a user needs to access a particular resource, and for that to happen, the authorization engine has to check the policy and match that against a variety of attributes that relate to the user and their environment.

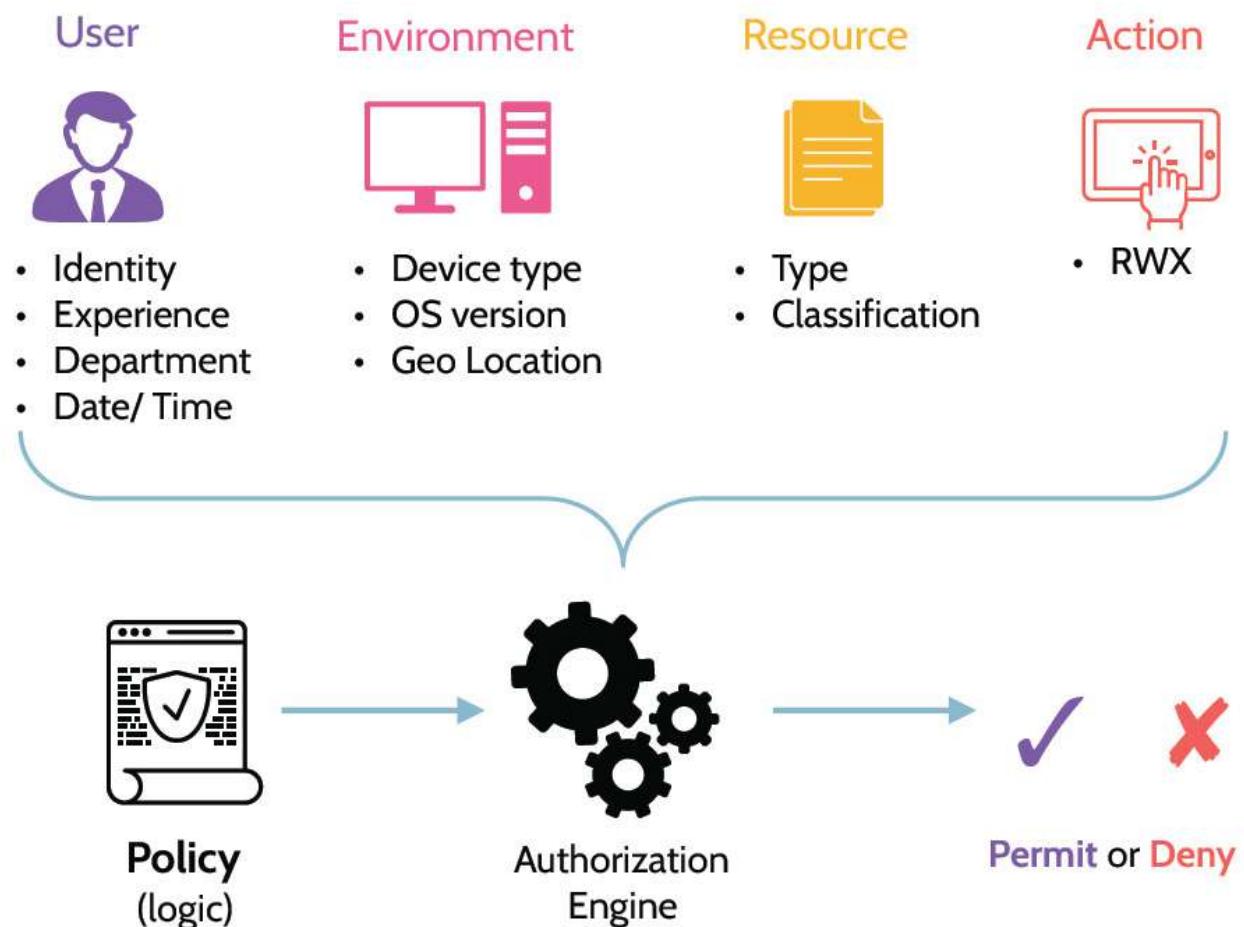


Figure 5-16: ABAC Operation

eXtensible Access Control Markup Language (XACML)

Understand the purpose of XACML

One of the tools that defines and enables attribute-based access control is a standard known as eXtensible Access Control Markup Language—XACML. This standard defines an attribute-based access control policy language, architecture, and processing model that allows attribute-based access to be implemented and utilized in a standardized manner.

Risk-Based Access Control

Risk-based access control looks at elements of a user connection, like the IP address, time of access request and more, to determine a risk profile associated with the request. Based upon the result, further authentication challenges may be presented to the user before access is granted.

5.4.2 Mandatory Access Control (MAC)

CORE CONCEPTS

- **Mandatory Access Control is very rare to see in use, and only typically used in government organizations, where confidentiality is often of primary importance.**
- **Mandatory Access Control requires every asset in an organization to have a classification and every user to be assigned a clearance level.**
- **A Mandatory Access Control system determines access based upon clearance level of the subject and classification, or sensitivity, of the**

object.

Understand defining characteristics of mandatory access control and where and why it is typically used

The key distinguishing feature of MAC (depicted in [Figure 5-17](#)) is that access is determined by a system, and most MAC systems are designed to protect confidentiality related to assets and information. The system itself makes access control decisions, based upon the classification of the objects being accessed and the clearance of the subject requesting access. In a MAC environment, every single object should be classified with a specific classification label, e.g., public, secret, top secret, and so on.

Correspondingly, all users should have a security clearance that aligns with the classification system used for objects. Within this framework, access will then be granted or denied accordingly. For example, if a user with “Public” clearance attempts to access an object that is classified as “Secret,” the system will deny the request. Note that MAC isn’t often implemented in private companies because in a typical organization it’s very rare to find employees with clearly defined levels of clearance and every asset with a clearly defined classification level. This explains why one or a combination of the previously discussed access control systems is used. However, in the context of government (specifically the military), MAC might be easily used. Even here, though, it’s not a given.

REFERENCE MONITOR CONCEPT (RMC)

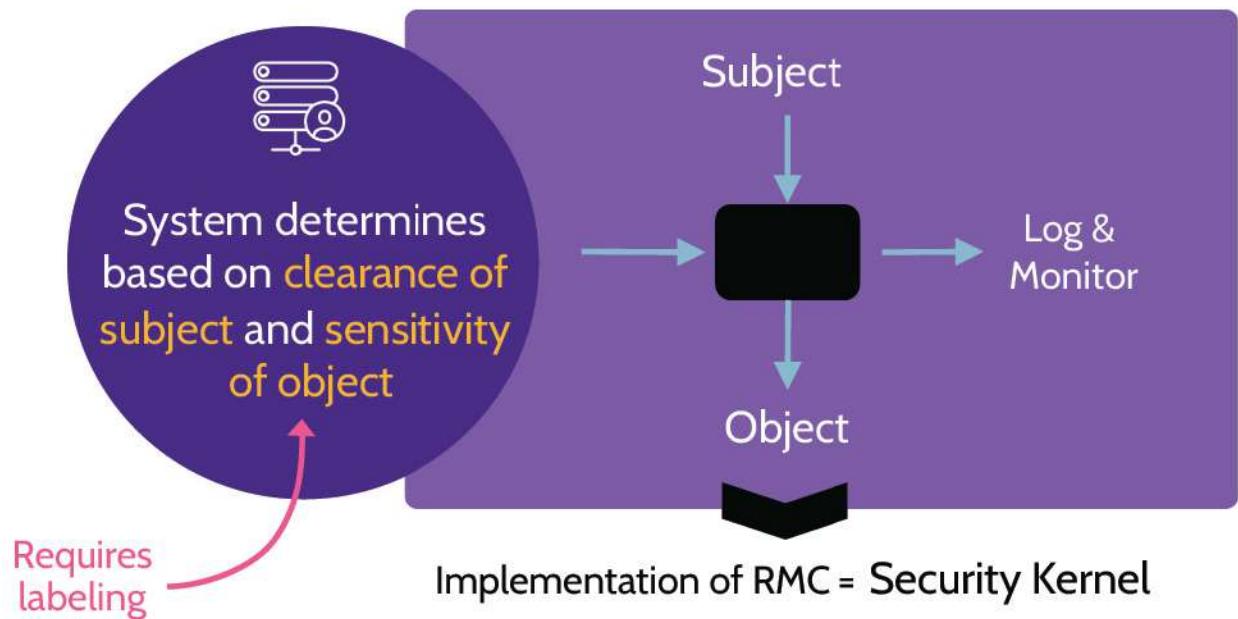


Figure 5-17: MAC Operation

5.4.3 Non-discretionary Access Control

CORE CONCEPTS

- **Non-discretionary Access Control means that somebody other than the asset owner determines who gets access.**
- **Non-discretionary Access Control should be avoided, if possible.**

Contrary to Discretionary Access Control (DAC) is **Non-discretionary Access Control** (depicted in [Figure 5-18](#)). If DAC means the owner decides who can access an asset, Non-discretionary Access Control means someone *other* than the owner determines access. Although this isn't a security best practice, it's an existing working practice in many companies and leads to someone in the IT department, for example, creating a user account and

granting access to numerous assets, whether access is needed or not. One reason is that an identified owner does not exist for a system; therefore, Discretionary Access Control can never be exercised. Additionally, situations may exist where the owner (accountable) delegates responsibility of access control to areas like IT, but then the owner doesn't offer input with regards to who should be given access. Rather, that's left up to the IT folks and therefore is another example of Non-discretionary Access Control being exercised.

REFERENCE MONITOR CONCEPT (RMC)

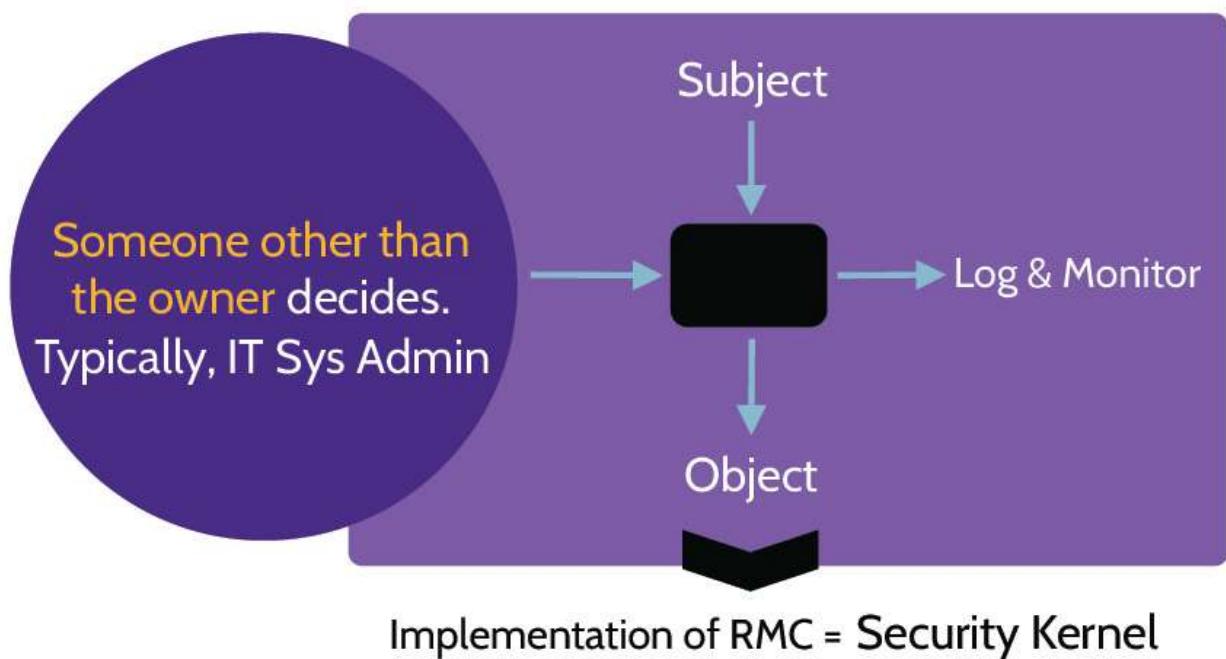


Figure 5-18: Non-discretionary Access Control

5.4.4 Access Policy Enforcement

There are two critical aspects for access policy enforcement in an application, A policy enforcement point (PEP) and a policy decision point (PDP). The

definitions are listed in [Table 5-21](#).

Policy enforcement point (PEP)	PEPs are parts of an application that receive authorization requests for protected systems and data. They function as gatekeepers and send the requests to the PDP where they are evaluated. When the PDP sends back the decision, the PEP enforces it by either granting or denying access. They are placed throughout an application's access points.
Policy decision point (PDP)	PDPs make decisions on the authorization requests that have been sent to them by PEPs. They make decisions based upon pre-defined rules and are generally centralized.

Table 5-21: **PEP vs. PDP.**P

5.5 Manage the identity and access provisioning life cycle

5.5.1 Vendor Access

CORE CONCEPTS

- **Vendor identity and access provisioning for systems and data should be considered with the same or more care than employee identity and access provisioning.**
- **Vendor provisioning might also include a security review component that includes a deeper review of the vendor or inspection of a vendor's facilities, systems, and other relationships.**

With any organization, certain functions and critical relationships require vendor access to some systems and data. Outsourced functions and relationships can vary and might include things like IT services, marketing,

finance and accounting, and supply chain suppliers, to name a few. As a result, third-party vendor relationships can represent significant risk to an organization, and identity and access provisioning for these relationships should be considered with as much or more care as identity and access provisioning for employees. In addition to normal provisioning, review and revocation activities, vendor access provisioning might include a security review of the vendor or even an onsite inspection of the vendor's facilities and systems.

5.5.2 Identity Life Cycle

CORE CONCEPTS

- **Identity Life Cycle is composed of three parts: provisioning, review, revocation.**
- **Provisioning = upon hire of new employee and when employee changes roles**
- **Review (also known as user access review) = should take place as often as necessary, and more frequently for higher privilege accounts**
- **Revocation = upon voluntary or involuntary termination**

Understand the identity life cycle and what happens at each stage

The Identity Life Cycle is simple in concept and refers to the creation or provisioning of user access, the periodic review of that access, and eventually

the revocation of that access. The three steps of the Identity Life Cycle are depicted in [Figure 5-19](#).



Figure 5-19: Identity Life Cycle

Provisioning activities include things like background checks, confirming skills, and identity proofing, among other things.

Periodically, this access should be **reviewed** to ensure continued appropriate access. Assets and systems to which a user has access should be reviewed by the asset or system owner to determine if ongoing access is necessary or if access should be modified. Timeliness of access reviews is dependent on a few variables, the most important being as often as necessary, based upon the value of the asset or system in question. Different assets and systems have different values to an organization and different risk profiles. This fact ultimately drives the need for more frequent or less frequent access reviews. Additionally, different types of user and system accounts also drive the timing of access reviews. High value accounts—system/admin/root—should be reviewed much more frequently than lower value accounts.

Eventually, and when necessary, an account should be *revoked*, or deprovisioned. **Revocation** typically takes place when an employee leaves the organization, through voluntary or involuntary separation, but it can also take place when an employee changes roles. This is very important to note. Otherwise, if an employee changes roles, they may gain additional privileges and rights as well as maintain existing privileges and rights, which can lead to increase in actual appropriate access. So, sometimes it is appropriate to revoke access for an employee and provision again, based upon new needs as set forth by the appropriate asset and system owners.

5.5.3 User Access Review

CORE CONCEPTS

- Account access review is an ongoing process, regardless of the type of account (user, system, service).
- Account access review frequency should be based upon the value of resources and associated risks.
- Privileged accounts should be reviewed more frequently.

Why access reviews should be conducted?

Once an account has been registered for a user, and the user is granted access to facilities, systems, and other resources, that doesn't mean that the access should remain forever. All user access should be reviewed on a periodic basis by the owner of the asset, because the owner is in the best position to conduct

this review and confirm that continued user access is appropriate. Additionally, user access reviews can mitigate access or privilege creep.

How often should access reviews be performed?

How often should access reviews be conducted?

The fact that user access should be reviewed at least annually raises additional questions. What about if a user changes role, leaves the company, or you're concerned about admin or "super user" roles? How often should reviews take place in these cases?

In the case of a user changing roles, their access should be reviewed at the time of the change. New access should be granted, as needed, and any access that is not needed should be removed. Of course, access should always be reviewed and approved by the owner. When someone leaves the company (through voluntary or involuntary termination) that user's access should be reviewed, and in most cases, all access should be removed. In the case of administrative and "super user" accounts, because they grant broader and more powerful access, access might need to be reviewed more frequently than annually; perhaps these reviews should occur as often as weekly.

Which accounts should be reviewed most frequently?

In every case, the value of the resources and the associated risks should drive access review timing. As noted above, it might be fine to review some user

access annually, while other reviews might need to take place more frequently.

5.5.4 Privilege Escalation (e.g., use of sudo, auditing its use)

In addition to more frequent reviews of privileged accounts, a recommended security practice is for system administrators (users with admin, root, and similar privileges) to only use their privileged accounts when strictly necessary.

Privileged users should utilize two accounts. They should use a standard user account for regular business purposes, such as checking email, participating in meetings, etc., and they should use a separate account with elevated privileges only when performing administrative tasks that require a higher level of access. A good example of this approach on Unix/Linux systems is the command sudo (“superuser do”) which is similar to the Windows RunAs command. An administrator can use a standard user account when logged into a system and only run specific commands or programs that require elevated privileges with the sudo command.

This way, when an administrator is performing activities that are most likely to compromise their account, such as checking email, browsing the web, etc., they are using their standard user account. The likelihood of their privileged account being compromised and used for malicious purposes is greatly reduced.

5.5.5 Service Account Management

CORE CONCEPTS

- Service account management involves managing the accounts used by services (not humans).
- We should limit these accounts to single purposes and reduce their privileges to limit the chances of compromise.

Service accounts are accounts that are used by services, workloads or applications. Service account management involves setting up, configuring, monitoring and maintaining these service accounts. Even though people don't use service accounts, human oversight must be in place to ensure that the accounts are appropriately secured. Otherwise, the accounts could be utilized for things like privilege escalation and spoofing. Best practices include limiting service accounts to single purposes and reducing their privileges to the minimum that is required to function effectively.

5.6 Implement authentication systems

5.6.1 Authentication Systems

CORE CONCEPTS

- Authentication systems are used to prove or verify an identity or system assertion.
- Popular authentication systems include: OpenID Connect (OIDC), Open Authorization (OAuth), Security Assertion Markup Language (SAML), Kerberos, Remote Authentication Dial-In User Service (RADIUS) and Terminal Access Controller Access Control System Plus (TACACS+).

As has been previously discussed, the implementation of authentication systems helps protect an organization, its users, and its critical assets from unauthorized access. At its core, authentication is the act of proving, or verifying, an identity or system assertion.

A number of authentication systems exist, and some of the more popular ones include:

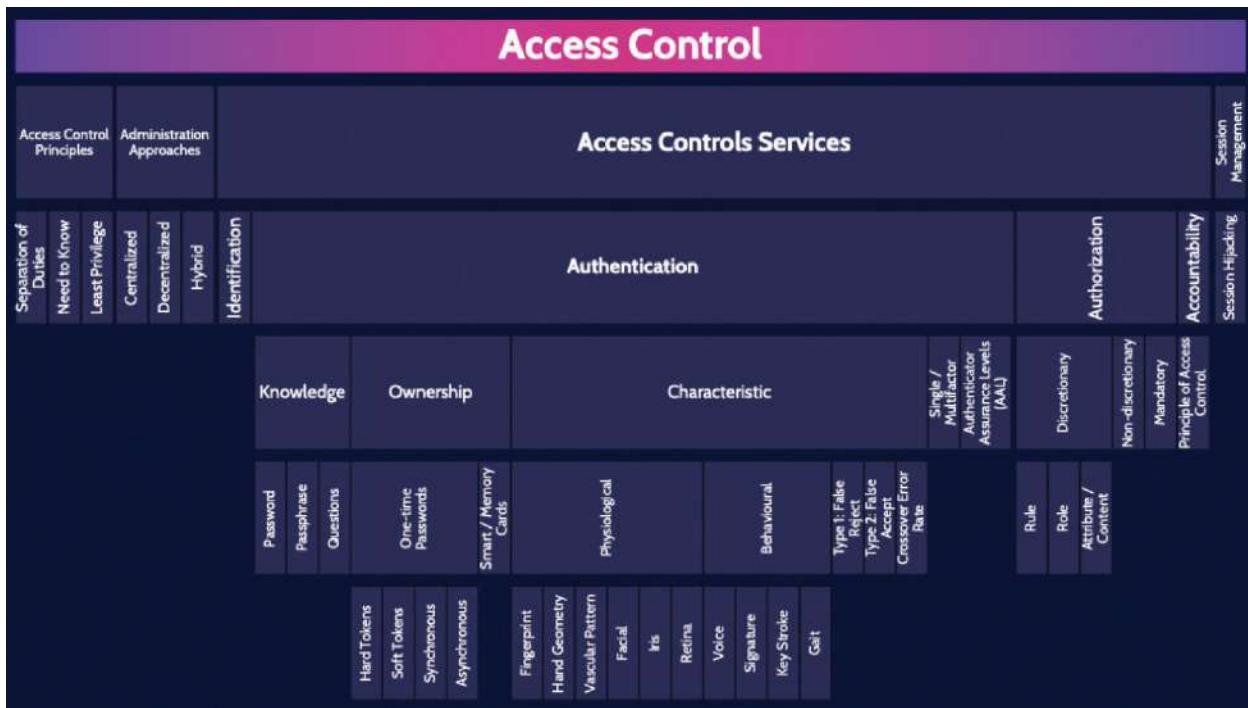
- OpenID Connect (OIDC)/Open Authorization (OAuth)
- Security Assertion Markup Language (SAML)
- Kerberos
- Remote Authentication Dial-In User Service (RADIUS) and Terminal Access Controller Access Control System Plus (TACACS+)

RADIUS, TACACS+, SAML, and Kerberos were both discussed in this chapter and in Domain 3. OIDC and OAuth are briefly discussed below:

■ OpenID Connect (OIDC)/Open Authorization (OAuth)—OAuth is an access delegation standard that target applications can use to provide client applications with secure delegated access over HTTPS. It authorizes devices, APIs, servers, and applications with access tokens rather than credentials. **OpenID Connect (OIDC)** is an identity layer built on top of the OAuth 2.0 framework. It allows third-party applications to verify the identity of the end user and to obtain basic user profile information. While OAuth 2.0 is about resource access and sharing, OIDC is about user authentication.



MINDMAP REVIEW VIDEOS



Access Control

dcgo.ca/CISSPmm5-1

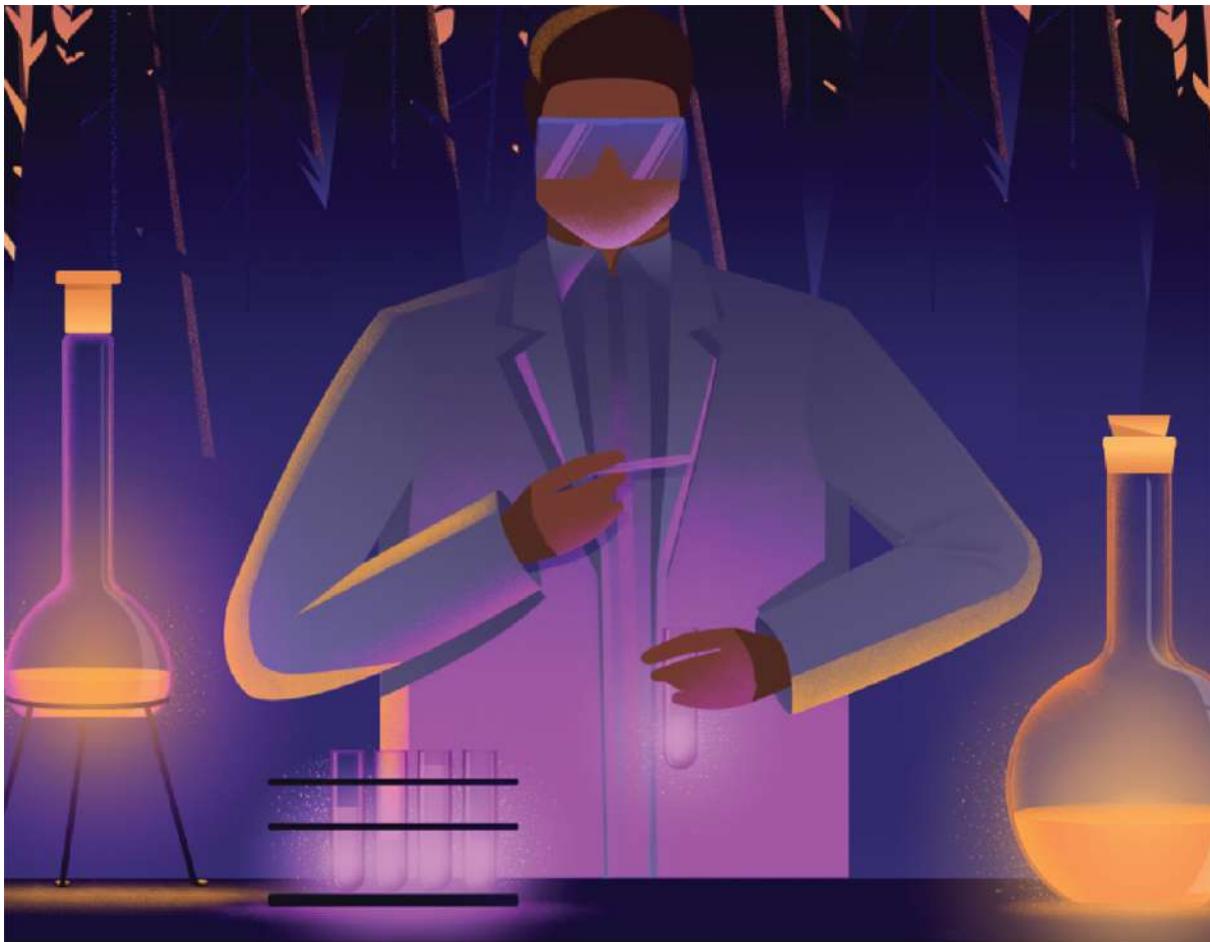
Single Sign-on / Federated Access

Allows users to access multiple systems with a single set of credentials

Single Sign-on					Federated Identity Management (FIM)					IDaaS				
Access systems within the same organization					Access systems across multiple entities					Identities				
Kerberos					SAML					WS-Federation				
User / Client	Key Distribution Center	Authentication Service	Ticket Granting Ticket (TGT)	Ticket Granting Service	Service Tickets	Service	Symmetric encryption only	Symmetric & Asymmetric encryption	Principal / User	Identity Provider	Relying Party / Service Provider	Tokens	Assertions written in XML	Components
			Components											Profiles
														Bindings
														Protocol
														Assertion

Single Sign-on/Federated Access

dgo.ca/CISSPmm5-2



DOMAIN 6

Security Assessment & Testing



6.1 Design and validate assessment, test, and audit strategies

Purpose of Security Assessment and Testing

Domain 6 focuses on security assessment and testing of architectures, assets, and systems. Information covered in earlier domains underscored the importance of every control to address security from the perspective of the two pillars—functional and assurance—that should support every control. Security assessment and testing specifically focuses on the latter—providing assurance to stakeholders, how security is contributing to goals and objectives, and that the right level of security is built into any architecture that provides value to the organization. In other words, *the purpose of security assessment and testing is to ensure that security requirements/controls are defined, tested, and operating effectively.* Additionally, security assessment and testing apply to the development of new applications and systems as well as the ongoing operations, including end-of-life, related to assets.

Consider this question for a minute: How many architectures and systems do we interact with daily and simply take for granted they will work? Think about all the planes in the sky, traffic control infrastructure in a major city, computer systems, and mobile phones, to name a few

examples. Though an imperfect measuring tool, the number of lines of computer code required to run and manage each of those systems can offer a glimpse as to their complexity, and some of these systems require millions or even billions of lines of code.

It is estimated that a modern operating system can contain something like fifty million lines of code. That's a massive number. Is there a chance that maybe a mistake or two exist in those fifty million lines of code? Yes, it's pretty much guaranteed that many mistakes, bugs, unknown exploits, and vulnerabilities may reside in fifty million lines of code. With each passing day, systems in use around the globe are becoming ever more complex. The more complex, the more opportunity and likelihood for errors to exist. Furthermore, not only are systems becoming more complex, but our dependence upon them is becoming more critical and pervasive, like planes using avionics systems with millions of lines of code. Most planes today can fly themselves, with very little manual intervention. However, air crashes still happen because of fatal errors. Obviously, these architectures are incredibly important, and it's imperative that they run well to ensure as much safety and order as possible. Rigorous testing and assessments must be done, and it's very important to understand that it's not just about the initial development of a system. Once deployed to production, a system should be monitored and tested

consistently for a number of reasons, i.e., making sure regulatory requirements are being met. When updates to a system are made, it should be tested to ensure the updates did not break something or create vulnerabilities. Even when a system is retired, testing and assessment should be done to confirm that data has been migrated to a new system properly and has been defensibly destroyed on the old.

6.1.1 Validation and Verification

CORE CONCEPTS

- **Validation answers one fundamental question: Are we building the right product?**
- **Verification follows validation and asks a related and equally important question: Are we building the product correctly?**

Understand validation & verification and the difference between the two

Two very important terms related to testing are: *validation* and *verification*. Both terms underscore the necessity and importance of early and ongoing testing, not simply testing after a product has been built.

Validation is the process that begins prior to an application or product being built. Validation is concerned with

answering one fundamental question: *Is the right product being built?* From the start, it's imperative that requirements are understood and documented correctly.

Understand when verification should be performed/stopped and what helps determine the stopping point

Verification follows validation and is the process that confirms an application or product is being built correctly. Like validation, verification answers another fundamental question: *Is the product being built correctly?* Domain 6 focuses a significant amount of attention on testing to ensure that an application or product is functioning properly, is secure, and is meeting business requirements. Testing can never offer 100 percent confidence that an application is working perfectly. However, verification testing can be used to develop a level of confidence, and the desired level is typically directly proportional to the organizational relevance and value of the application or system. Level of confidence is another way of saying "level of care" about how well the application or product is working. A high level of confidence would be desired regarding a plane's avionics systems but not about the workings of a company wiki.

Included with verification are three terms that deserve further examination: *completeness*, *correctness*, and *consistency* (also known as the 3 Cs).

- Completeness means ensuring all the use cases of an application, based upon all defined requirements related to functionality, have been covered.
- Correctness refers to each “use” case representing what’s supposed to be built.
- Consistency speaks to functionality being specified consistently in all areas.

These three words capture the essence and philosophy of application and system development.

Validation and verification are summarized in [Table 6-1](#):

Validation	Verification
<p>Are we building the right product?</p> <p>Develop a level of confidence that business requirements are clearly understood and have been validated with the business owner. Cannot build the right product if it is not clearly understood what the owner wants.</p>	<p>Are we building the product right?</p> <p>Develop a level of confidence that whatever is being built is meeting all the defined requirements. The product is being built correctly based on the requirements defined during the validation stage.</p>

Table 6-1: Validation and Verification

6.1.2 Effort to Invest in Testing

CORE CONCEPTS

- **The purpose of security assessment and testing is to provide assurance regarding the architecture, application, or system being assessed and tested.**
- **Assurance is provided through validation and verification.**
- **The effort to invest in testing should be proportionate to the value the application or system represents to the organization.**
- **Assessment, Testing, and Auditing strategies include: internal, external, third party.**
- **The role of a security professional is to: identify risk and advise testing processes to ensure risks are appropriately evaluated.**

How much testing is enough?

As with all things related to security in an organization, the time and effort invested in testing should be proportional to the value it represents to the organization. Value drives security, including the testing done to prove to stakeholders that security is contributing. Testing strategies flow from this value/security relationship.

Assessment, Testing, and Auditing Strategies

As noted earlier, testing strategies are used to provide assurance and can be considered from three broad contexts: ***internal*, *external*, and *third party***. Each can be used alone or in combination, based on the type/level of assurance sought.

Internal assessment/testing/auditing (e.g., within organization control) are conducted by someone internal to the organization, in other words, by an employee.

External assessment/testing/auditing (e.g., outside organization control) can be defined in two different ways, and both should be noted.

On one hand, a company that uses a certain application, for example, Microsoft Azure, might use a team of employees to look at Azure's hosting environment, which is external to the company. So, a company essentially audits or examines the environment of an external service provider.

On the other hand, imagine a company building an application using its own development team. Internal assessment/testing/auditing has been conducted, but the company also wants independent assessment/testing/auditing performed as a means of providing objective assurance that the application is well

designed and working properly. To meet this need, the company hires a major consulting firm to come in and audit the application. This consulting firm is external to the company that is building the application.

Third-party assessment/testing/auditing (e.g., outside enterprise control), as the name suggests, means three parties are involved in the process: the customer, the vendor, and then perhaps a consulting or similar company. Third-party assessment/testing/auditing is very prevalent in the context of cloud computing. For example, imagine a customer is interested in cloud computing, and they engage Amazon to provide one or more services offered by Amazon Web Services (AWS). The customer, rightly so, wants to know if AWS is secure. To provide independent and more objective assurance about the security of their services, Amazon may engage a consulting firm to come in, audit their environment and service offerings, and produce reports about their findings. Then, when a company asks Amazon about the security of their services, Amazon can point to the independently produced report to assure the customer. Thus, three parties—the consumer, the service provider, and the external auditor—make up third-party testing.

Assessment/testing/auditing strategies and the

implications of each

All three strategies are depicted in [Table 6-2](#) and can be used in combination based on the type/level of assurance sought.

Internal Audit	External Audit	Third-party Audit
Testing conducted by somebody internal to the organization	Testing via either of two scenarios: <ul style="list-style-type: none">■ somebody internal to the organization examining an external service provider's controls,or ■ an organization asking somebody external from the company to come in and provide an unbiased examination of an application or system	Three parties are involved: customer, vendor, independent audit firm

**Table 6-2: Testing Strategies
Location**

In this context, location refers to where the audit is being conducted: on-premise, in the cloud, or hybrid. This contrasts with internal, external and third-party audits,

which are classed according to who is doing the auditing.

Table 6-3 describes the three options in more detail.

On-premise	Cloud	Hybrid
An on-premise audit focuses on evaluating security within an organization's physical facilities and data centers.	A cloud audit focuses on evaluating the security of systems, data and applications hosted by a cloud provider.	A hybrid audit is for hybrid infrastructures and combines both on-premise and cloud evaluations.

Table 6-3: The Three Major Locations for an Audit

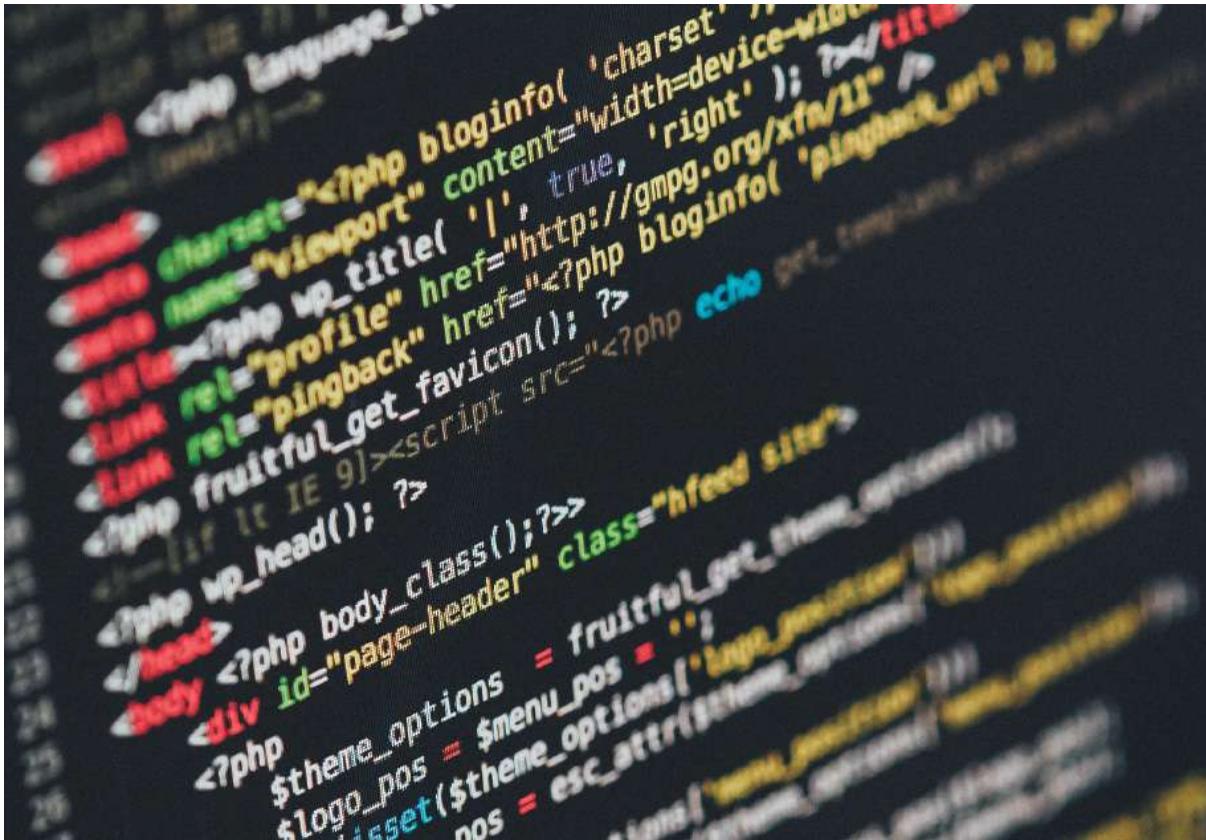
Role of Security Professional

Understand the role of a security professional

As might make sense, based upon everything written to this point, testing should include all relevant stakeholders. However, the role of security professionals is to do three things:

- Identify risk
- Advise testing processes to ensure risks are appropriately evaluated
- Provide advice and support to stakeholders

The security team should not actually perform the testing alone; however, security should advise, provide assurance, monitor, support, and evaluate results.



6.2 Conduct security control testing

6.2.0 Testing Overview

CORE CONCEPTS

- Security control testing typically includes steps that align with the phases of the application and systems development process.
- Software testing includes several types of testing that build upon one another: unit testing, interface testing, integration testing, system testing.

Examples of Testing Performed

[Figure 6-1](#) provides a high-level overview of some of the required software testing throughout the system life cycle, and [Table 6-4](#) summarizes the testing performed at each phase. This is not an official framework and accordingly shouldn't be memorized. It is simply meant to provide an overview and highlight that testing is required during every stage of a system's life cycle from the start and throughout.

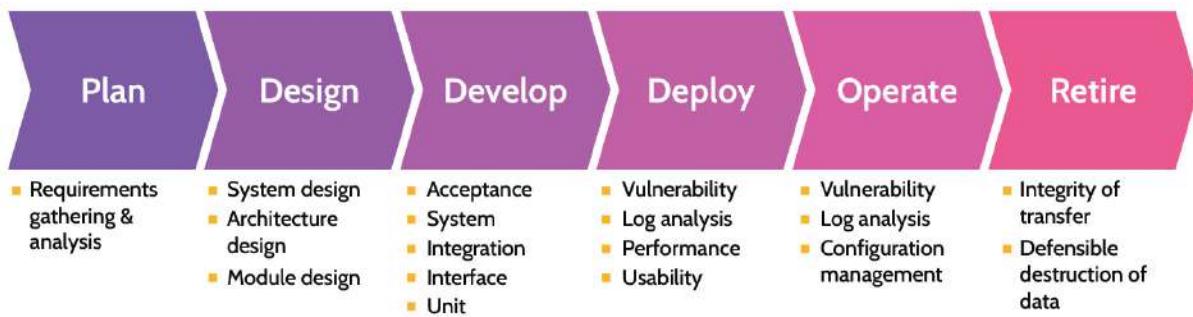


Figure 6-1: Software Testing Phases

Planning	Capture related requirements to a system before any design takes place and validate the requirements have been accurately captured.
Design	The security team can provide advice on what types of controls the system should have to protect fundamental security principles, like confidentiality, integrity, and availability. Test to confirm all required controls have been included in the system design, architecture design and module, and so on.
Develop	Numerous types of testing must be performed during the development stage to confirm all the required controls are being implemented correctly: unit testing, integration

	testing, system testing, acceptance testing, vulnerability assessments, and so on.
Deploy	Moving an application from a quality assurance/preproduction/testing environment to the actual production environment. Numerous types of testing must be performed during deployment: usability testing, performance testing, reviewing logs for errors and anomalies, vulnerability assessments, and so on.
Operate	Configuration management reviews can be performed to ensure the product is working as intended without its security being compromised. Vulnerability management and log analysis can continue being performed for that very purpose.
Retire	Testing and ensuring that data has been migrated into the new system in a secure manner in addition to safely disposing it from the old one.

Table 6-4: Software Testing Phases
Software Testing Overview

Most software development projects comprise multiple teams, with each team responsible for developing specific aspects of functionality that are then brought together as a complete software product. Software testing should be a comprehensive process that examines everything from the individual functional components to the integrated functional system, and the process is outlined in [Table 6-5](#).

 Unit Testing	<p>Examines and tests individual components of an application. As specific aspects (units) of functionality are finished, they can be tested.</p>
 Interface Testing	<p>As more and more individual components are built and tested, interface testing can take place. Interfaces are standardized, defined ways that units connect and communicate with each other. Interface testing serves to verify components connect properly.</p>
 Integration Testing	<p>Integration testing focuses on testing component groups (groups of software units) together.</p>
 System Testing	<p>System testing tests the <i>integrated system</i> (the whole system).</p>

Table 6-5: Software Testing Stages



6.2.1 Testing Techniques

CORE CONCEPTS

- Testing techniques are broken down broadly into two categories: manual and automated.
- Manual testing is performed by a person.
- Automated testing is performed by an automated tool.
- Static application security testing (SAST) looks at the underlying source code of an application while the application is not running; SAST is considered white box testing, because the code is visible.
- Dynamic application security testing (DAST) examines an application and system as the underlying code executes; DAST is considered black box testing, because the code is not visible.

- **Fuzz testing** is a form of dynamic testing and is premised upon chaos, to see how an application responds to complete randomness.
- **Code review** is considered from two perspectives: black box (zero knowledge about the code is available) and white box (full knowledge about the code is available).
- **Test types include:** positive (tests a system from a normal usage perspective), negative (tests a system from the perspective of normal errors), misuse (tests a system from the perspective of a malicious user or attacker); all three test types are valuable.
- **Equivalence partitioning** is testing, where specific input values are used to test from a grouping perspective (partitions of values and possible inputs).
- **Boundary value analysis** is testing from a bounds perspective (lower and upper bounds of groups or partitions).

Methods/Tools

When developing an application or other software, several testing techniques can and should be utilized, and they can be done via manual or automated means.

Manual testing means a person or team of people is performing the tests. They might be following a specific process, but they're actually sitting in front of a computer, looking at code, testing a form by entering input, and so on.

Automated testing means that test scripts and batch files are being automatically manipulated and executed by software. Thorough testing employs both approaches to

produce a multitude of outcomes and achieve the best results.

Manual and automated testing are summarized in [Table 6-6](#).

Manual Testing	Automated Testing
Done by a person—hands on keyboard	Done by an automated tool, like code scanning or vulnerability assessment software

Table 6-6: **Testing Methods/Tools**

Understand the key differences between SAST, DAST, and fuzz testing

Runtime

Runtime specifically refers to when an application is running. The runtime environment refers to the environment it is running in.

With **Static Application Security Testing (SAST)**, an application is not running, and it's the underlying source code that is being examined. Static testing is a form of white box testing, because the code is visible.

With **Dynamic Application Security Testing (DAST)**

testing, an application is running, and the focus is on the application and system as the underlying code executes. As opposed to static testing, dynamic testing is a form of black box testing, because the code is not visible. The entire focus is the application itself and how it behaves, based upon inputs.

Fuzz testing is a form of dynamic testing. The entire premise behind fuzz testing is chaos. In other words, fuzz testing involves throwing randomness at an application to see how it responds and where it might “break.” Fuzz testing is quite effective, because application developers and programmers tend to be very logical people, and the software they develop tends to reflect this fact. By testing from an illogical perspective—by throwing chaos at an application—issues not previously uncovered can be identified.

There are two major types of fuzz testing as defined in [Table 6-7](#).

Mutation (dumb fuzzers)	Generation (intelligent fuzzers)
The input to an application is randomly changed by flipping bits or appending/replacing additional random input.	New input to an application is generated from scratch based on an understanding of the file format or protocol.

This is often referred to as dumb fuzzing as the fuzzer has no understanding of the input structure	This is often referred to as smart / intelligent fuzzing as the fuzzer must understand the input structure
--	---

Table 6-7: Mutation and Generation Fuzzers

The three major types noted are summarized in [Table 6-8](#).

Static (SAST)	Dynamic (DAST)	Fuzz
<ul style="list-style-type: none"> ■ White box ■ Examines code 	<ul style="list-style-type: none"> ■ Black box (no access to code) ■ Examines application itself 	<ul style="list-style-type: none"> ■ Form of dynamic testing ■ Premise is chaos

Table 6-8: SAST, DAST, and Fuzz Testing

Understand the difference between black box and white box testing

Code Review/Access to Source Code

Code review and access to source code can be considered from two perspectives when testing:

- No access to the source code exists (also known as black box testing)
- Access to the source code does exist (also known as white box testing)

Extrapolating further, the concepts defined above can be mixed and matched. For example, automated static white box testing means a tool is used to automatically examine the available source code, looking for common

errors, undefined variables, and similar types of problems. Automatic dynamic black box testing, on the other hand, is like how vulnerability scanners operate. A vulnerability scanner does not have visibility or access to the underlying source code. Instead, it performs dynamic testing that seeks to identify common vulnerabilities and other issues with an application. Both types of testing provide value and can be mixed and matched as needed to provide comprehensive results.

Be able to differentiate between test types—positive, negative, and misuse

Test Types

When a system is running, it's possible to test it as if a user was using it. Specifically, the system can be tested a few different ways—via positive testing, negative testing, or misuse testing. These testing types are explained in [Table 6-9](#).



Positive Testing

Focuses on the response of a system, based upon normal usage and expectations. For example, under normal circumstances, if a login page requires a username and password, and the correct username and password are provided, the system should complete the log-in process. This is positive testing, checking if the system is working as expected and designed.

 Negative Testing	<p>Focuses on the response of a system when normal errors are introduced. Using the example above, if the incorrect username or password are entered, the system shouldn't crash. It simply should not log the subject in and should instead issue some type of error indicating an incorrect username or password was entered. This is normal, expected behavior, under negative testing.</p>
 Misuse Testing	<p>Unlike positive and negative testing, misuse testing is a bit more devious. With this type of testing, the perspective of someone trying to break or attack the system is applied. If a system can be tested and understood from the standpoint of normal expected usage (when everything is done correctly, or when errors are made), it should also be understood from the viewpoint of somebody who wants to break into or otherwise abuse the system in order to gain further access.</p>

Table 6-9: Positive, Negative, and Misuse Testing

Understand the difference between equivalence partitioning and boundary value analysis

Equivalence Partitioning and Boundary Value Analysis

These two testing techniques are designed to make testing **more efficient**.

Let's use the following example, which is also depicted in [Figure 6-2](#). An application provides a user with a password input box, which instructs them to enter a password between eight and sixteen characters. An inefficient way to

test would be to test passwords of a set number of characters one by one: ■ A password of 0 characters should be **rejected**

- 1 character should be rejected ■ 2, 3, 4, 5, 6 and 7 characters should all be rejected ■ 8 characters should be **accepted**

- 9, 10, 11, 12, 13, 14, 15 and 16 characters should all be accepted ■ 17 characters and above should be **rejected**

To test this password input box more efficiently, boundary value analysis could be used. **Boundary value analysis** requires first identifying where there are changes in behavior—these are called boundaries. In the given example, there is a change in behavior between 7- and 8-character lengths—7 should be rejected, and 8 should be accepted. There is a second boundary between 16- and 17-character lengths—16 should be accepted and 17 should be rejected. Once the boundaries have been identified, **testing can be focused on either side of the boundaries**, as this is where there are most likely to be bugs.

Equivalence partitioning starts with the same first step of boundary value analysis—identifying the boundaries, and then goes a step further to identify partitions. Partitions are **groups of inputs that exhibit the same behavior**. Based

on the example, there are three partitions: 1. Partition I: Password consisting of zero to seven characters (all rejected) 2. Partition II: Password consisting of eight to sixteen characters (all accepted) 3. Partition III: Password consisting of seventeen or more characters (all rejected) Once the partitions have been identified, some testing can be performed within each partition.

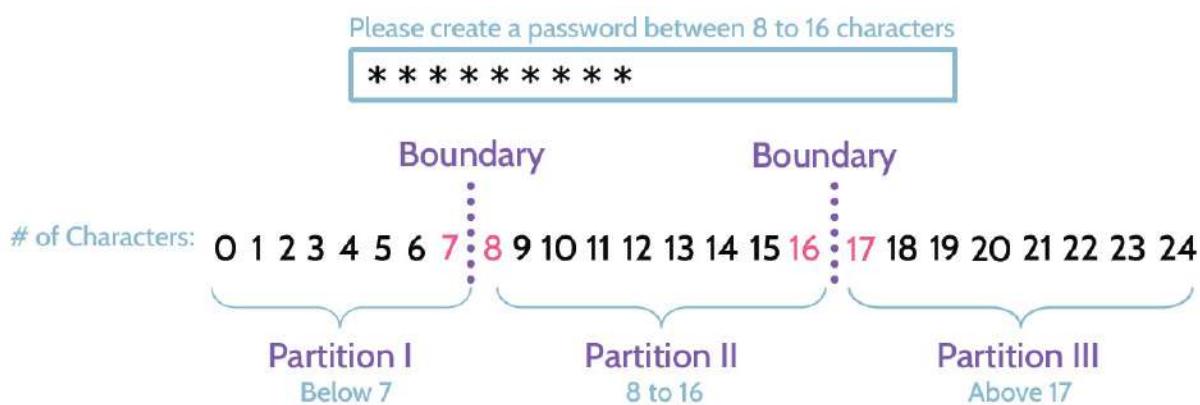


Figure 6-2: **Boundary and Partitioning Value Analysis**

A summary of equivalence partitioning and boundary value Analysis can be seen in [Table 6-10](#), while [Figure 6-3](#) contains some testing examples to demonstrate different types of testing just discussed.

Equivalence partitioning	Inputs are divided (partitioned) into groups that exhibit the same behavior with test cases covering each partition.
Boundary value	Focus on testing data at the boundaries with test cases covering extreme ends of the input values.

analysis

Table 6-10: Equivalence Partitioning and Boundary Value Analysis

Enter a number between 0 and 9 <input type="text" value="aaa' OR 1=1--"/>	Misuse	Entering a value of aaa' OR 1=1-- constitutes an attempt to perform SQL injection, which denotes a misuse test.
Enter a number between 0 and 9 <input type="text" value="10"/>	Negative	Entering 10 while the field expects a number between 0 and 9 denotes a legitimate user error, so it is classified as a negative test.
Enter a number between 0 and 9 <input type="text" value="2"/>	Positive	Entering the number 2 means an expected number is provided, denoting a positive test.
Enter a number between 0 and 9 <input type="text" value="-1 0 9 10"/>	Boundary Value Analysis	Entering numbers -1, 0, 9, and 10 (in successive attempts) when the system expects a number between 0 and 9 denotes a boundary value analysis test, with -1 and 0 testing the 0 boundary while 9 and 10 are testing the 9 boundary. -1 and 10 should be rejected, and 0 and 9 should be accepted by the system.
Enter a number between 0 and 9 <input type="text" value="-5 5 15"/>	Equivalence Partitioning	Entering -5, 5, and 15 (in successive attempts) denotes equivalence partitioning using those three numbers as parts of different partitions. -5 and 15 should be rejected, and 5 should be accepted by the system.

Figure 6-3: Testing Examples

Table 6-11 summarizes a couple of additional types of testing that can be performed in specific scenarios

Decision Table Analysis	Different input combinations and their corresponding system behavior/output are captured in a table (useful for complex software testing and requirements management).
-------------------------	---

State-Based Analysis	<p>A set of abstract states that a unit of software can take are defined, and then tests compare its actual state to the expected state (useful for testing GUIs and communications protocols).</p>
-----------------------------	--

Table 6-11: Decision Table and State-based Analysis

Test Coverage Analysis

Test coverage analysis or simply “coverage analysis” refers to the relationship between the amount of source code in a given application and the percentage of code that has been covered by the completed tests. Test coverage is a simple mathematical formula: amount of code covered/total amount of code in application = test coverage percent. To illustrate using a simple example, if an application contains one hundred lines of code and fifty lines have been tested, the test coverage would be calculated as follows: amount of code covered/total amount of code in application = 50/100 or 50 percent.

6.2.2 Vulnerability Assessment and Penetration Testing

CORE CONCEPTS

- Vulnerability testing techniques tend to be automated and can be performed in minutes, hours, or a few days; penetration testing techniques tend to be manual and can take several days, depending on the complexity involved.
- Testing stages include: reconnaissance, enumeration, vulnerability analysis, exploitation, reporting.

- Testing perspectives include: internal (inside a corporate network) and external (outside a corporate network).
- Testing approaches include: blind (tester knows little to nothing about the target) and double-blind (tester knows little to nothing about the target and internal security teams do not know the test is coming).
- Testing knowledge includes: zero, or black box (similar to blind approach, where tester knows nothing about a target), partial, or gray box (tester has some information about a target), full, or white box (tester has significant knowledge about a target).

Purpose of Vulnerability Assessment

What is the purpose of vulnerability assessment?

Vulnerability assessments and penetration testing (better known as pen testing) are important topics when discussing vulnerabilities and threats, and this points back to the more general topic of risk analysis. As a quick review, a vulnerability can be defined as *a weakness that exists in a system*, while a vulnerability assessment is an attempt to identify vulnerabilities in a system.

Understand the difference between vulnerability assessment and penetration testing

With any risk analysis, it is important to first know what assets exist. Next, the threats these assets face must be

identified, which can happen through threat modeling. Two well-known and often-used threat modeling methodologies are STRIDE (Spoofing, Tampering, Repudiation, Information disclosure, Denial-of-Service, Elevation of privilege) and PASTA (Process for Attack Simulation and Threat Analysis). Finally, to understand the full breadth of risk that exists, vulnerabilities must also be identified.

Vulnerability analysis helps in this regard, and two primary methods are used to identify vulnerabilities: vulnerability assessment and pen testing.

Vulnerability Assessment versus Penetration Test



Figure 6-4: **Vulnerability Assessment vs. Penetration Testing**

Figure 6-4 shows a comparison between vulnerability assessment and penetration testing.

Both processes start the same way, as they each seek to identify potential vulnerabilities. However, with a **vulnerability assessment**, once vulnerabilities are noted, no further action is taken apart from a report of findings being produced. A **penetration test** goes a very important step further: after the identification of vulnerabilities, an attempt

is made to exploit each vulnerability (breach attack simulations).

Various common characteristics exist between vulnerability assessments and penetration tests. A scope of what is being examined needs to be defined. Furthermore, an activity schedule needs to be set. In either case, if activity takes place without prior knowledge of the owners of the systems or network, alerts may unnecessarily be triggered, and a response set in motion. In other words, with either a vulnerability assessment or pen test, business impact can be a possible result (especially with a pen test). There is not an insignificant chance that production systems can be negatively impacted (e.g., knocked offline) because of these tests. Thus, a very clearly defined scope, schedule, and approval of activity must be in place.

However, some core differences between vulnerability analysis and penetration also exist.

For one, vulnerability analysis tends to be more automated. Tools like Nessus, Qualys, and InsightVM can be run and automatically gather significant information about vulnerabilities in a system or network. Pen tests, on the other hand, tend to be more manually driven, although pen testers will often use automated scanners. With a pen test, manual attempts to exploit vulnerabilities and breach a

system are made. The quality of results is often directly proportional to the skill level and experience of the pen tester.

Additionally, whereas a vulnerability assessment can be performed quickly (minutes, hours, or a handful of days usually), a pen test can take significantly longer (commonly several days), depending upon the complexity of the identified vulnerabilities and targets being exploited. Finally, it's worth noting that during a pen test, confidential or sensitive information might be accessed or identified. Along with the likelihood that a pen test could cause impact, this fact further underscores the need for approval and perhaps even something like an NDA being in place before taking any type of assessment or action against a network.

When performing a vulnerability assessment or a penetration test, a series of steps are followed, which are shown in [Figure 6-5](#).



Figure 6-5: Vulnerability Assessment/Penetration Testing Phases

Understand the vulnerability assessment and penetration testing process and which key step differentiates the two

1. **Reconnaissance:** Involves gathering publicly available data via activities like Domain Name System (DNS) and WHOIS queries, browsing social media sites like LinkedIn, browsing job listings on sites like Indeed or forum sites like Google Groups, where sensitive company information might be inadvertently posted by somebody looking for help with an issue. With just a small amount of effort, a large amount of publicly available information about a company can be gleaned, and the company won't know that this information is being sought nor that it is being compiled. Thus, *reconnaissance is considered a **passive activity*** because the target doesn't know any activity is taking place or can't detect that this information is being gathered, as there's no direct interaction between the tester and target.
2. **Enumeration:** Unlike reconnaissance, where information is passively gathered, *enumeration is considered an **active activity*** because the target can detect the scans. Typically, items enumerated

are IP addresses, ports, hostnames, and user accounts. If reconnaissance indicates that an organization has control of a certain IP range, enumeration will involve identifying which IP addresses and ports specifically are being used and potentially open. Ports equate to services, and 65,536 TCP and UDP ports exist (0–65,535) and can be enumerated. If a port is open, this means a service is running. For example, if port 80 is open, there's likely a web server running, assuming default ports are being used. Enumeration focuses on identifying a system and services behind a given IP address. Knowing this information can point to potential vulnerabilities specific to the system. A web server will be vulnerable to certain things that don't apply to a database or other type of server. Enumeration helps determine and narrow down this information. In addition, enumeration focuses on identifying hostnames and active user accounts on the various targets, which can be leveraged for access later.

3. **Vulnerability analysis:** This phase follows enumeration and helps determine which vulnerabilities exist within a target network or machine. However, this stage also represents a fork in the road where vulnerability and penetration

tests are concerned. If a vulnerability test is being performed, attempts to exploit are not conducted, and the next step is documentation of findings and compilation into a report.

4. Execution/Exploitation: *If a penetration test is being performed*, an attempt will be made to exploit the identified vulnerabilities and therefore confirm, definitively, if the vulnerability can be exploited and is a true-positive. The execution step is only performed as part of a Penetration Test 5.

Document findings/Reporting: This is where it all comes together, regardless of whether this is a vulnerability assessment or a penetration test. The tester will use a report to compile all their findings and provide a detailed record of all the techniques tested, which worked, and which didn't, associated tools, identified vulnerabilities, and most importantly mitigation steps required to be taken by the organization. Some vulnerabilities might require immediate attention, while others can be considered informational and less serious. It's important to clearly define and prioritize these vulnerabilities, so proper attention can be given to the most critical vulnerabilities first. Additionally, another important facet of documentation is trying to eliminate and remove as many false-positives as

possible. Otherwise, a vulnerability report that should be twenty pages in length is more likely to be two hundred pages, and the critical data is buried in a sea of otherwise non-essential information. As much as anything, compiling findings in a clear, concise, and relevant manner are as or more important than the efforts that preceded the documentation process.

Red Teams, Blue Teams, and Purple Teams

CORE CONCEPTS

- **Red teams simulate threats against an organization and report on how the organization can improve its defenses.**
- **Blue teams evaluate organizational security, provide mitigations, monitor systems, perform incident response, oversee security operations, and make recommendations.**
- **Red teams are attackers and blue teams are defenders.**
- **Purple teams are collaborations between red and blue teams that aim to promote working together and sharing information.**

Red teams are groups of ethical security professionals that simulate attacks against organizations. They use techniques like pen testing, social engineering, and threat intelligence. **Blue teams are the defenders, and they focus on things like evaluating an organization's current security, recommending mitigations, monitoring,**

incident response, digital forensics, and security operations.

Purple teams aren't distinct teams, but instead they are collaborations between red and blue teams. The idea behind purple teams is that organizations want red and blue teams to cooperate when appropriate. Purple teams can share information and ultimately make an organization much more secure through the collaboration. While competition between red and blue teams can be good, a situation that is solely adversarial between the two sides is detrimental. It could limit the sharing of critical information, which would ultimately make the organization less secure.

Testing Techniques

Vulnerability assessments and penetration testing can be quite nuanced, and several variables come into play with each. These variables include: perspective, approach, and knowledge.

Perspective

Understand testing technique perspectives, approaches, and knowledge types

Perspective refers to the perspective from which the assessment or test is being performed. Is the assessment or

test coming from an internal (inside the corporate network) or from an external (out on the internet) perspective? [Table 6-12](#) explains the difference between internal and external testing.

Internal Testing	External Testing
The test is being performed from inside the corporate network . This is important because threats can originate from inside a network (like a disgruntled employee or attacker already inside the network), and an internal test can help pinpoint exactly what an insider threat can access or what may have already been compromised.	Testing from an external perspective , where threats from outside the network can be considered and tested. Note that an outsider may need to circumvent multiple layers of defense (defense in depth) in order to access a resource which might be easily (or even directly) accessible if they were positioned internal to the network.

Table 6-12: Internal vs. External Testing Approach

In addition to testing from an internal or external perspective, testing can be approached differently. One approach is known as blind and the other as double-blind. [Table 6-13](#) explains the difference between blind and double-blind testing.

Blind Testing	Double-Blind Testing
---------------	----------------------

<p>The assessor is given little to no information about the target being tested. It might simply be the name of the company or an IP address provided; otherwise, the assessor is blind to network details and must use reconnaissance and enumeration techniques to gain more visibility about the target. With a blind approach, members of the target company's IT and Security Operations teams will likely know that some type of test is coming and can be better prepared to respond to alerts.</p>	<p>A double-blind approach goes one step further. In addition to the assessor being given little to no information about the target company, the target company's IT and Security Operations teams will not know of any upcoming tests. This type of approach tests the assessor's ability to identify vulnerabilities and other weaknesses as well as the target's internal team ability to respond. In this case, and to prevent the notion that hacking or anything illegal is taking place, usually only senior management will be aware of upcoming tests, because they commissioned the double-blind test.</p>
---	--

Table 6-13: Blind vs. Double-Blind Approach Knowledge

Knowledge pertains to how much insight or information an assessor has about a target. [Table 6-14](#) explains the difference between zero, partial, and full knowledge testing.

Zero Knowledge (black box)	Partial Knowledge (gray box)	Full Knowledge (white box)
<p>The assessor has zero knowledge—same as the blind approach noted above.</p>	<p>The assessor is given some information about the target network but not the full</p>	<p>The assessor is given full knowledge (including items like IP addresses/range,</p>

	<p>set that a white box test would have. It lies somewhere in between a white and black box test.</p>	<p>network diagrams, information about key systems, and perhaps even password policies).</p>
--	---	--

Table 6-14: Zero, Partial, and Full Knowledge

6.2.3 Vulnerability Management

CORE CONCEPTS

- **Vulnerability management is the cyclical process of identifying, classifying, prioritizing, and mitigating vulnerabilities**

The steps of vulnerability management

Vulnerability management is a critical element of the risk management process that aids with the determination and implementation of appropriate controls related to identified vulnerabilities. At its core, vulnerability management is the ongoing process of identifying vulnerabilities, understanding the potential organizational impact as part of risk analysis, and ensuring that vulnerabilities are mitigated.

At a high level, an effective vulnerability management process should include the following steps noted below.

- An understanding of all assets in an organization, which requires an accurate **asset inventory**.
- Identifying the **value of each asset** in the inventory, which requires:
 - A data/asset classification and categorization structure.
 - Identified owner for each asset.
 - Assigned classification and categorization for each asset.
- Identifying the **vulnerabilities for each asset** and remediation of identified vulnerabilities via patching, updating, and other means necessary to eliminate the vulnerabilities or reduce their risk (all remediation activities should be done as part of a patch or remediation management process, which is part of change management).
- **Ongoing review and assessment** of all steps to ensure the asset inventory is kept up to date, and new vulnerabilities are identified and remediated.

Always remember that vulnerability management is a cyclical and ongoing process, because organizational assets are constantly being added and removed and new vulnerabilities are constantly being identified.

6.2.4 Vulnerability Scanning

CORE CONCEPTS

- Automated vulnerability scanning can help identify vulnerabilities from an organizational perspective as well as from the perspective of an attacker.
- Two primary types of vulnerability scans: credentialed/authenticated scans and uncredentialed/unauthenticated scans.
- Banner grabbing is a process used to identify a system's operating system, applications, and versions.
- Fingerprinting works to identify the unique characteristics of a system through examination of how packets and other system-level information is formed.
- Interpretation and understanding of scan results is often achieved with the help of two tools: CVE and CVSS.
- CVE, also known as Common Vulnerability & Exposures dictionary, is “a list of records—each containing an identification number, a description, and at least one public reference—for publicly known cybersecurity vulnerabilities.”
- CVSS, known as Common Vulnerability Scoring System, reflects a method to characterize a vulnerability through a scoring system considering various characteristics.
- Two types of alerts: false-positives and false-negatives.

- **False-positives:** the system claims a vulnerability exists, but there is none.
- **False-negatives:** the system says everything is fine, but a vulnerability exists; false-negatives are bad.

Automated Vulnerability Scanners

Understand the different ways automated vulnerability scanning can be used and implications of each approach

Most vulnerability scans are performed using automated tools, like Nessus, Qualys, OpenVAS, InsightVM, or Retina. These tools can scan entire networks or individual machines, and even specific applications for *known* vulnerabilities. They can also perform two significantly different types of scans: credentialed or non-credentialed scans.

When run as a credentialed scan, the vulnerability scanner is given a username and password to log into the system it is scanning. Being able to authenticate allows Nessus to scan at a much deeper level and report more detailed information as a result. Additionally, credentialed scans often help eliminate false-positives (where the system claims a vulnerability exists but there is none), because specific configuration settings and similar details can be considered. Finally, credentialed scans can help ensure that

all systems are configured correctly relative to baseline configuration for a given system. Variances can easily be noted and addressed.

Non-credentialed scanning means the vulnerability scanner is *not* given the ability to login and connect to the network, system, or application being scanned. Used in this manner, the tool will help identify especially glaring vulnerabilities and weaknesses but without being able to perform deep scanning (due to the lack of credentials). False-positives may be identified in a larger volume, and they will need to be addressed with the appropriate teams. Both types of scans should be used, as each serves specific purposes that when combined can prove very insightful and useful.

Credentialed versus uncredentialed scanning is summarized in [Table 6-15](#).

Credentialed/Authenticated	Uncredentialed/Unauthenticated
<ul style="list-style-type: none">■ Automated scanning tool given username/password in order to authenticate to the system being scanned ■ Helps prevent false-positives	<ul style="list-style-type: none">■ Scanning tool is used from perspective of a hacker ■ May lead to false-positives

Table 6-15: Credentialed vs. Uncredentialed Scanning

An important thing to note is that a vulnerability scanner can only identify *known* vulnerabilities. These tools depend upon up-to-date catalogs or databases of all known vulnerabilities in order to identify them in the systems being scanned. As one might imagine, these catalogs and databases are continuously evolving and in constant need of updating. When Nessus scans a system, it compares details about the system with known vulnerabilities about it in the database. If a vulnerability is not known and therefore not in the database, the tool will report nothing.

Banner Grabbing and OS Fingerprinting

An important consideration when assessing a computer system is knowing exactly what operating system and version of the software is running. Is it a Linux system, or Windows 7, or Windows 10, and, if Windows 10, exactly what build of the OS? Knowing the operating system and exact version can help identify specific vulnerabilities.

Vulnerabilities that apply to one operating system and version likely differ from vulnerabilities that apply to other operating systems and versions.

*Banner grabbing and fingerprinting are methods whereby active or passive techniques are used to **identify** a system's specific operating system, applications, and associated versions.*

The more information that can be gained about a system, the easier it is to protect it or, conversely, attack it. **Banner grabbing** helps identify software and versions, while **fingerprinting** looks a bit more specifically at the way a packet is formed and similar attributes to identify the unique characteristics of a system—similar to the ridges of a person's fingerprint that make it unique.

Interpreting and Understanding Results

Understand the difference between CVE and CVSS and how they are used together to evaluate vulnerabilities

Closely coupled with identifying and reporting results from activities like vulnerability scanning, banner grabbing, and fingerprinting is identifying the severity and exact nature of the results. This is done using two tools: CVE and CVSS, which are explained in [Table 6-16](#).



CVE
(Common Vulnerability & Exposures)
Dictionary

CVE, also known as Common Vulnerability and Exposures directory, is a list of security flaws and vulnerabilities that are publicly disclosed for awareness and risk mitigation. Security and technology-related firms around the world keep an eye out for vulnerabilities in their products, services, and elsewhere. Whenever a new vulnerability is discovered, the company that made the discovery typically tends to publicize it and give the vulnerability a unique name. This can be challenging, because another company may identify the same vulnerability using a different tool set, and thus—on the surface—the same vulnerability may be identified using

	different names. The CVE mitigates this duplication by serving as a clearinghouse to ensure that each vulnerability is only identified and recorded one time . Through the use of a standardized and unique identification number as well as description for each vulnerability, companies around the world can be on the same page.
 CVSS (Common Vulnerability Scoring System)	CVSS, known as Common Vulnerability Scoring System, is a framework that uses common vulnerability metrics and characteristics to provide an average score of how severe a vulnerability is . It assigns a number between 0 and 10 to any vulnerability; the higher the number, the more severe and critical the vulnerability. When a company identifies a new vulnerability, a scoring methodology will be followed to determine a score and then reported for inclusion in the CVSS database.

Table 6-16: **CVE vs. CVSS**

When a vulnerability scanner is used to run a scan, a detailed report of findings will be generated. For each identified vulnerability, CVE and CVSS information will be included—the CVE identifying the vulnerability, and the CVSS scoring the severity of the vulnerability.

False-Positives and False-Negatives

With any type of monitoring system, two types of alerts often show up. One type, a false-positive, indicates a vulnerability exists, but in fact there is no vulnerability. The other type, a false-negative, indicates everything is fine, but in fact a vulnerability is lurking. While false-positives can

create a lot of “noise” and administrative overhead, false-negatives can potentially lead to serious harm and damage. Of the two, false-negatives are much worse, as they don’t allow us to identify specific vulnerabilities within the network.

False-positives versus false-negatives are summarized in [Table 6-17](#).

 False Positives	System claims a vulnerability exists, but there is none.
 False Negatives	System says everything is fine, but a vulnerability exists; false negatives are bad.

Table 6-17: Possible Alert Statuses

6.2.5 Log Review and Analysis

CORE CONCEPTS

- Log review and analysis is a best practice that should be used in every organization.

- Logs should include what is relevant, be proactively reviewed, and be especially scrutinized for errors and anomalies that point to problems, modifications, or breaches.
- Synchronized log event times is critical for activity correlation, especially if a breach occurs; NTP is typically used for purposes of time synchronization.

Understand the importance of timely log review and analysis

Review and analysis of log files is a best practice and can help an organization know if systems deployed in production are working properly. This said, as most systems can generate significant amounts of logged data, it's important that the points noted in [Table 6-18](#) be considered.

 Log what is relevant	Most systems produce a wealth of information, but not all of it is relevant. Using risk management as a guide, risks to assets can point to ways to detect if a risk were to occur. This in turn points to what is relevant to log.
 Review the logs	Whether done via automated or manual means, logs must be reviewed. In today's typical environments, an automated system (like a SIEM tool) is going to best facilitate the review of hundreds, thousands, or even millions of logged events.



Identify errors/anomalies

As log review is undertaken, focus on identifying errors or anomalies that may indicate attacks or suspicious activities. Examples include:

- **Errors:** Unexpected errors that might indicate a system is not working properly.

- **Modification:** Modification to a system, especially if unauthorized. This is usually a significant red flag and may indicate a breach.

- **Breach:** Actual penetration of a system or network that may lead to significant damage—monetary, reputation, or worse.

Table 6-18: Log Review and Analysis



Log Event Time

Ensuring consistent time stamps of log entries is very important. If an organization has deployed multiple servers and other network devices—like switches and firewalls—and each device is generating events that are logged, it's critical that the system time, and therefore the event log time, for each device is the same. Otherwise, if a breach occurs, for example, trying to correlate activities as an

attacker moves through a network becomes extremely difficult. Consistent time stamps mean all system and device clocks are set to the exact same time, and this is done by use of **Network Time Protocol (NTP)**. With most networks, at least one network device (and usually two or more, for purposes of redundancy) is synchronized with a publicly available nuclear clock managed by a government agency, like NIST. Then all other network devices are synchronized with the one device, therefore ensuring consistent time stamps across a network.

Log Data Generation

As already noted, log data is generated by virtually every system operating in an organization. Logging and monitoring processes go deeper and include:

- Generation
- Transmission
- Collection
- Normalization
- Analysis
- Retention
- Disposal

These topics will be covered in more detail in Domain 7.

6.2.6 Limiting Log Sizes

CORE CONCEPTS

- **Log file management is important.**
- **Two log file management methods are used: circular overwrite and clipping levels.**
- **Circular overwrite limits the maximum size of a log file by overwriting entries, starting from the earliest.**

■ Clipping levels focus on when to log a given event, based upon threshold settings, and log file sizes are limited as a result.

Understand the difference between circular overwrite and clipping levels and which could potentially provide more valuable information

Log file management in any organization is important, especially with regards to log file sizes, and two methods are often utilized for this purpose: *circular overwrite* and *clipping levels*. The purpose here to ensure that log files only contain relevant information and are not full of massive meaningless information.

The first method—**circular overwrite**—works as the name suggests. For example, if the log file size is set to 100 MB or perhaps ten thousand logged events, enabling circular overwrite means that once the log file reaches the maximum size or length, log entries start being overwritten, from the earliest to the most recent. Thus, the maximum file size or number of entries will never be exceeded. If disk or memory space is limited, circular overwrite can be very useful and potentially prevent a disk from filling up or the system from crashing.

The other method—**clipping levels**—is a bit more interesting and potentially more informative. Rather than

logging every bit of activity, clipping levels focus on when to begin logging. For example, logging every failed login attempt due to a wrong password makes no sense, as people mistype passwords all the time—they may just have the Caps Lock key on or have hit the wrong key. However, what if the wrong password is typed fifty times, or twenty-five times, or even fifteen times? This could indicate a system-related problem or password cracking attempt. This is where clipping levels can be effectively used. A threshold can be set to only allow logging of any activity once that threshold is met. So, clipping levels is another way to limit log file sizes and to narrow down the focus of logging to the most pertinent and meaningful details. Unlike circular logging, though, clipping levels do not actually delete data. So, if an organization is interested in identifying something like security breaches, using clipping levels is a better approach, because relevant information can be logged and preserved. If circular logging is used, breach-related log entries might be overwritten and deleted due to file size or number of entries limitation.

6.2.7 Operational Testing—Synthetic Transactions and RUM

CORE CONCEPTS

- Operational testing occurs while a system is operating.

- Operational testing techniques include: Real User Monitoring and Synthetic Performance Monitoring.
- Real User Monitoring (RUM) monitors user interactions and activity with a website or application.
- Synthetic performance monitoring examines functionality and performance under load.

Operational testing is performed when a system is in operation. Within the context of operational testing, two testing techniques are often employed.

- **Real User Monitoring** is a passive monitoring technique that monitors user interactions and activity with a website or application. Log files are often examined in real-time, and performance measures might also be fed from the website or application into a Real User Monitoring tool for more detailed analysis.
- **Synthetic Performance Monitoring** is a bit more complicated and more interesting. The word synthetic can also be referred to as fake, phony, or counterfeit, and so “synthetic performance monitoring” means making up transactions and subjecting them to an architecture or system to see how it reacts. Let’s look at an example to illustrate. Imagine a bank that operates an online banking system. Bank customers can log in to the system,

check account balances, pay bills, transfer funds, and perform other related actions. The online banking system provides a lot of functionality, and to best serve customers it's important that this functionality be available whenever a customer desires to use it. Thus, ongoing testing of the system's functionality is important. Testing can be done manually, but this can be a slow and painful process, and it might miss some tests. A better solution is using automation. Test scripts for each type of functionality can be created and then run at any time. This is what synthetic performance monitoring describes.

Along with these functional tests, synthetic performance monitoring can also test functionality and performance under load. What happens if thousands of the tests noted above were conducted simultaneously? Of course, this would indicate how well the system can handle load and speaks to the performance side of things. So, synthetic performance monitoring considers the functional and performance aspects of a system.

For sake of accuracy, the best environment to perform synthetic testing in is production. However, prior to testing in production, significant QA and related testing must be done in a test environment. Cyber Monday, Black Friday,

and similar online events place enormous loads on retail e-commerce environments, and retail organizations will often perform major functionality and load tests using their production environment prior to these big events.

Real user monitoring and synthetic performance monitoring are summarized in [Table 6-19](#).

Real User Monitoring	Synthetic Performance Monitoring
Monitoring user transactions in real time for usage, performance, and errors	Running scripted transactions to monitor functionality, availability, and response times

Table 6-19: Real User Monitoring and Synthetic Performance Monitoring

6.2.8 Regression Testing

CORE CONCEPTS

- **Regression testing is the process of verifying that previously tested and functional software still works after updates have been made.**
- **Regression testing should be performed after enhancements have been made or after patches to address vulnerabilities or problems have been issued.**
- **Results of regression testing should be captured and communicated in a manner that is specific and relevant to the party reading the results—this is done using “metrics that matter.”**

What is regression testing, and why is it important?

Regression testing is the process of verifying that previously tested and functional software still works after updates have been made. Updates can be in the form of enhancements or similar changes, or in the form of patches to address vulnerabilities or problems. Regression testing is used to verify that the software still works and that nothing else is broken.

Additionally, regression testing of a complex application can take time and might involve a significant number of tests. Once this testing is complete, the results need to be compiled and reported. If thousands of tests have been run, is it likely that the CEO is going to be interested in hearing or reading about all the details? No, of course not. The CEO is likely only going to be interested in a high-level summary report. However, the development team would likely be very interested in all the details contained in the results. This information could prove very useful.

As noted and coupled with things like regression testing and the related reporting of results, it's critical to consider who your audience is and what they need to see. This points to the need to build reports using "metrics that matter." For senior management, the report should include metrics that enable them to make informed decisions from an

organizational perspective; for members of an application development team, the report should include metrics that include detailed information regarding the application and the development process. As such, the points below should be considered when creating these reports:

- Objective pass/fail decisions
- Include the right detail for the right audience
- Metrics that matter

6.2.9 Compliance Checks

With everything noted above about security control testing in mind, it's imperative that compliance checking be an integral and ongoing part of the security control testing process. Through review and analysis of implemented controls and their output to confirm alignment with documented security requirements, compliance can be confirmed.

Furthermore, compliance checking can confirm that testing and controls are aligned with organizational policies, standards, procedures, and baselines, and it can effectively put a bow on all the activities described in this section.

6.3 Collect security process data (e.g., technical and administrative)

6.3.1 Key Risk and Performance Indicators

CORE CONCEPTS

- Key risk and performance indicators help inform goal setting, action planning, performance, and review, among other things.
- Key risk indicators (KRI) are forward-looking indicators and aid risk-related decision-making.
- Key performance indicators (KPI) are backward-looking indicators and look at historical data for purposes of evaluating if performance targets were achieved.
- SMART metrics are: Specific, Measurable, Achievable, Relevant, Timely.

The term *SMART metrics* has been around for some time and refers to data points that can be used for goal setting, taking action, and so on. More to the point, they should be specific, measurable, achievable, relevant, and timely.

- Specific—Result clearly stated and easy to understand?
- Measurable—Result can be measured/have the data?
- Achievable—Results can drive desired outcomes?
- Relevant—Aligned to business strategy?
- Timely—Results available when needed?

Oftentimes, SMART metrics are found in the context of employee performance and reviews. They can also be used in the context of security, and these metrics should be

specific to the audience—what do they want to see, what do they need to know, etc.

In this context, two types of metrics are very important: KPIs—Key Performance Indicators and KRIs—Key Risk Indicators.

KPIs are backward looking. They look at historical data and indicate whether performance targets were achieved. KRIs are forward looking. They help with decision-making about things like risk exposure, operational risk, and so on.

How do you decide what to focus metrics on?

This said, not everything can be measured. Organizations make decisions on which metrics to focus on based on their business goals and objectives, denoting what's most important to the organization.

Example Areas for Metrics

- Account management ■ Management review and approval ■ Backup verification ■ Training and awareness ■ Disaster recovery (DR) and business continuity (BC)

A closer look at these areas would show account management metrics to likely focus on user activities that might take place in the context of a help

desk on things like “mean time to resolution,” “average response time,” and “number of support emails” as they relate to service tickets. Account management metrics might also focus on details related to user accounts in a system, like last log-on time, status of account, and time of last password change.

Management review and approval metrics typically focus on products and processes, including the review and approval process itself. Things like “time to resolve defects,” “number of defects identified,” “defect detection effectiveness,” “average cost per defect,” “deliverables,” and “process used by reviewers” might be metrics included as part of management review and approval, along with many other metrics.

Backup verification refers to the process of verifying that backed up data is valid and accessible. This typically involves routine restores of subsets of data to confirm the availability and integrity of all information. It should also include a more thorough exercise from time to time that assumes a worst-case scenario, where significant amounts of data have been lost, that requires the restoration of that data from multiple sources. Related metrics might include “number of backups verified,” “time between backup verification exercises,” “number of verified files or total amount of data verified.”

Training and awareness metrics examine things like “number of employees who completed training” as well as results from phishing-related exercises, like “number of times phished” and “phishing emails reported.”

Finally, DR and BC metrics include things like Recovery Time Objective (RTO) and Recovery Point Objectives (RPO), and they might also include

others, like “total plans that cover each critical process,” “actual time required to restore a process,” and “time between plan updates,” to name a few examples.

A comparison between KPIs and KRIs can be found in [Table 6-20](#).

Key Performance Indicators (KPIs)	Key Risk Indicators (KRIs)
<ul style="list-style-type: none">■ Backward looking metrics ■ Metrics that indicate the achievement of performance targets■ Provide insights about risk events that have already affected the organization	<ul style="list-style-type: none">■ Forward looking metrics ■ Metrics that indicate the level of exposure to operational risk■ Help to better monitor potential future shifts in risk conditions or new emerging risks

Table 6-20: **KPI vs. KRI**

6.4 Analyze test output and generate report

6.4.1 Test Output

CORE CONCEPTS

- **Results of security assessments and testing should include steps related to: remediation, exception handling, and ethical disclosure.**

As part of security assessments and testing, it is important to note what steps will be taken to address issues and vulnerabilities identified during the assessment and testing. It is equally important to note what steps might *not* be taken and why this is the case. Finally, as is becoming more the rule, it's

important that newly discovered vulnerabilities be disclosed and shared with others. These activities are summarized in [Table 6-21](#), and details from each activity should be incorporated into a report.

Remediation	Based upon assessments and testing, remediation steps for all identified vulnerabilities should be documented.
Exception handling	If an identified vulnerability will not be remediated, this should be documented too, including the reason why that's the case. For example, perhaps remediation would cost too much relative to the value of the asset at risk or the chance of the risk being realized. Regardless of the reason, exceptions should be carefully considered and noted.
Ethical disclosure	Some identified vulnerabilities might be new discoveries and point to significant flaws and weaknesses in widely used software and hardware. In this context and for the sake of other users, customers, and vendors, it's very important that newly discovered vulnerabilities be shared to the extent necessary to protect anybody who may be exposed to the same vulnerabilities.

Table 6-21: Test Output



6.5 Conduct or facilitate security audits

6.5.1 Audit Process

CORE CONCEPTS

- Audit approaches include: internal, external, and third party.
- Internal audits involve employees focused on organizational processes.
- External audits involve employees focusing on vendor processes.
- Third-party audits involve independent auditors focusing on vendor processes.
- Audit plans typically include: defining the audit objective, defining the audit scope, conducting the audit, and refining the audit process.

Understand components of an audit plan

Assessments and third-party audits are an integral part of operations and the security assessment and testing strategy of an organization. Audits consist of internal, external, and third-party efforts, and oftentimes a hybrid approach is utilized. The security function needs to support the audit process. Regardless of the type of audit, most audit plans include the following steps:

- Define the audit objective
- Define the audit scope
- Conduct the audit
- Refine the audit process

To the last point, a typical audit process might include:

- Determining audit goals
- Involving the right business unit leader(s)
- Determining the audit scope
- Choosing the audit team
- Planning the audit
- Conducting the audit
- Documenting the audit results
- Communicating the results

Expanding upon the types of audits, internal audits involve an organization's own employees examining the organization's own systems; external audits could be a company's employees examining a vendor's systems or external auditors, for example, an independent audit firm's auditor looking in and providing an independent assessment of a company's controls. Third-party audits involve an independent auditor examining the operations and governance of a service provider. The service provider pays the independent auditor to conduct the audit and produce a report. The service provider can then provide this report to their customers. As a reminder, the different audit approaches were listed in [Table 6-2](#).

6.5.2 System and Organization Controls (SOC) Reports

CORE CONCEPTS

- Audit standards have matured over the years from SAS70 → SSAE 16 → SSAE 18.
- ISAE 3402 is the international standard in assurance engagements and is quite similar to SSAE 16/18 standards, with slight variations.
- SOC audits are used to build trust between service organizations and their customers.
- SOC 1 reports are quite basic and focus on financial reporting risks.
- SOC 2 reports are much more involved and focus on the controls related to the five trust principles: security, availability, confidentiality, processing integrity, and privacy.
- SOC 3 reports are stripped down versions of SOC 2 reports—typically used for marketing purposes.
- Type 1 reports focus on a point in time (SOC 1 and SOC 2).
- Type 2 reports focus on a period of time, covering design, and operating effectiveness (SOC 1 and SOC 2).

Over the years, audit standards have evolved. Years ago, SAS 70 was the gold standard. It was superseded by SSAE 16, which in turn has been superseded by SSAE 18. In each case, the standard outlines how to conduct third-party audits. In the United States, the American Institute of Certified Public Accountants (AICPA) is the governing body that oversees and refines these standards that essentially say, “Anyone who is going to conduct audits should ensure that the following details are included.” The SSAE 18 standard defines three types of System and Organization Controls (SOC) reports: (the SOC acronym used to stand for Service Organization Controls, but it has since been changed). An SOC audit is a method for building trust between a service organization and its customers. The three types of SOC report are:

SOC 1 reports focus on **financial reporting risks**.

SOC 2 reports focus on what are known as the **five trust principles: security, availability, confidentiality, processing integrity, and privacy**. A SOC 2 report will cover controls related to security, availability, and confidentiality. Coverage of controls related to processing integrity and privacy are optional, and it therefore may or may not be included in the report. SOC 2 reports can be quite detailed documents that contain information about an organization’s controls and how they operate as well as details about an organization’s systems. In fact, SOC 2 reports often contain a fair amount of confidential information about an organization, and they should be protected from accidental disclosure or disclosure to unauthorized people. However, certain information contained within a SOC 2 report would be valuable for new prospective customers to know. This is where SOC 3 reports can be very helpful.

An **SOC 3** report is essentially a stripped down and sanitized version of a SOC 2 report. It’s a **marketing tool** that potential customers or interested parties can read to gain a basic understanding of a

service provider's controls.

Understand the difference between SOC 1, SOC 2, and SOC 3 reports and what each report entails

As security professionals, the most meaningful among available SOC reports are SOC 2.

It's important to also understand that two types of SOC 1 and SOC 2 reports exist. These are known as Type 1 and Type 2, also known as Type I and Type II.

A **Type 1** report focuses on the **design of controls at a point in time**. To conduct a Type 1 audit, an auditor will come into an organization and focus on paperwork (policies, procedures, baselines, and so on). The auditor is essentially evaluating whether a process is properly designed on the day they looked at it. Do policies and procedures exist? Are they documented? Do they contain expected information?

This examination is done from the perspective of a point in time, and just speaks to whether a control appears to be appropriately designed at that point in time. A Type 1 audit does not provide any indication as to whether controls are operating effectively.

Understand the difference between Type 1 and Type 2 SOC reports and which is more comprehensive

A **Type 2** report examines not only the design of a control but more importantly the **operating effectiveness over a period of time**, typically a year. A Type 2 report covers everything in a Type 1 report and then goes much deeper to confirm that controls are operating effectively. Looking at change management, for example, the auditor would confirm that a change management policy exists as well as associated procedures—that the controls are properly designed. Then the auditor would examine a “population” of changes—perhaps all the changes that occurred during the past year. From that population, the auditor would choose a subset of changes and examine them closely. Did the control operate effectively? Was testing, including regression, performed? Was the change management review board involved? Did the appropriate stakeholders approve the change? Were the changes documented? The auditor is going to dig much deeper and confirm the operating effectiveness related to all samples examined, and again, this operating effectiveness will cover a period of time.

From the types of reports defined, it should be clear that **SOC 2, Type 2 are the most desirable reports for security professionals**, as they report on the operating effectiveness of the security controls at a service provider over a period of time.

A Type 1 report would usually be used during the first year that a service provider begins having a third-party audit conducted. If a company is just ramping up and undergoing audits, they're likely to have issues with their controls that need to be rectified, especially for the sake of long-term customer perceptions and credibility. By pursuing a Type 1 audit first, the auditor can identify gaps, missing controls, and other problems, with the expectation that the organization will reconcile everything

during the coming months and that a Type 2 audit will be conducted the next year. So, SOC 2, Type 1 reports typically are used in the first year and SOC 2, Type 2 reports then become the norm in order to show operational control continuity and compliance. The various SOC report types are depicted in Figure 6-6.

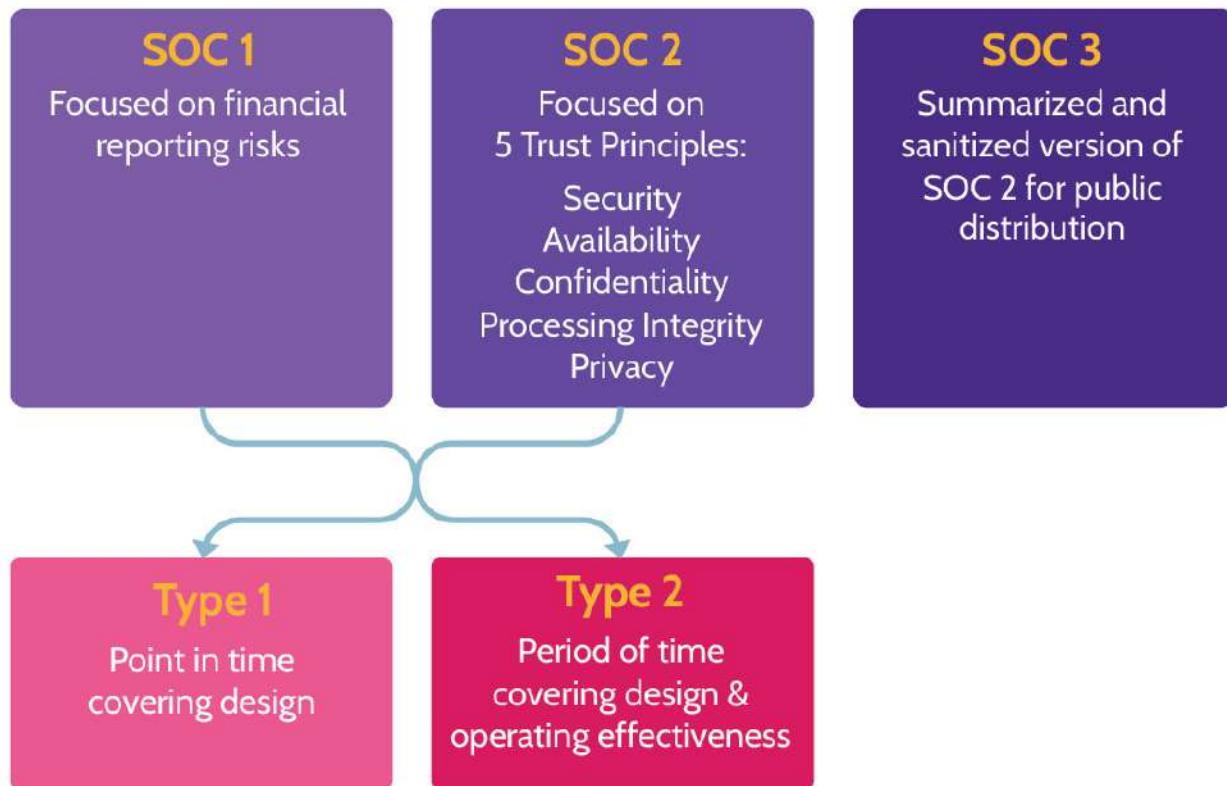


Figure 6-6: SOC Report Types

6.5.3 Audit Roles and Responsibilities

CORE CONCEPTS

- **Audit roles include: executive (senior) management, audit committee, security officer, compliance manager, internal auditors, and external auditors.**
- **Audit responsibilities vary based upon the audit role.**

It's important that executive (senior) management understand that the tone must be set from the top, and this applies to assurance as well. Yes, auditors are doing the work, but executive management must clearly articulate that assurance is important and that the process is supported and a priority for the organization.

At the same time, the Audit Committee, which is made up of key members of the Board as well as senior stakeholders from across the organization, should provide oversight and direction to the audit program. The Chief Security Officer's (CSO) or Chief Information Security Officer's (CISO) role is to advise on security-related matters. Compliance managers are usually responsible for an audit function, scheduling audits as necessary or required, ensuring that auditors are hired and trained, among other things. Internal auditors work for the organization they audit and provide assurance that corporate internal controls are operating effectively. External auditors work outside the organization—typically for an independent organization—and come inside to examine an organization's controls. A summary of these roles can be found in [Table 6-22](#).

Understand audit roles and responsibilities

Executive (Senior) Management	Sets the proper tone from the top, promotes the audit process, and provides support where needed
Audit Committee	Composed of members of the Board/senior stakeholders to provide oversight of the audit program
Security Officer	Advise on security related risks to be evaluated in the audit program
Compliance Manager	Ensure corporate compliance with applicable laws and regulations, professional standards, and company policy
Internal Auditors	Company employees who provide assurance that corporate internal controls are operating effectively
External Auditors	Provide an unbiased and independent audit report as they are independent of the entity being audited

Table 6-22: Audit Roles



MINDMAP REVIEW VIDEOS

Security Assessment and Testing												
Testing a System			Testing Techniques				Testers / Assessors			Metrics		
Unit	Methods & Tools	Runtime	Access to Code	Techniques	Operational	Internal	External	Third-Party	Roles	Focus	KPIs	KRIs
Validation	Manual	Boundary Value Analysis	Positive	Boundary Value Analysis	SOC 1	Executive Management	Type 1	Type 2	Audit Committee	Focus	KPIs	KRIs
Verification	Automated	Equivalence Partitioning	Negative	Equivalence Partitioning	SOC 2	Security Officer			Security Officer			
Rigour	Static	Real User Monitoring	Misuse	Real User Monitoring	SOC 3	Compliance Manager			Compliance Manager			
	Dynamic	Synthetic Performance Monitoring		Synthetic Performance Monitoring		Internal Auditors			Internal Auditors			
	Fuzz	Regression Testing		Regression Testing		External Auditors			External Auditors			
	White											
	Black											
	Black											

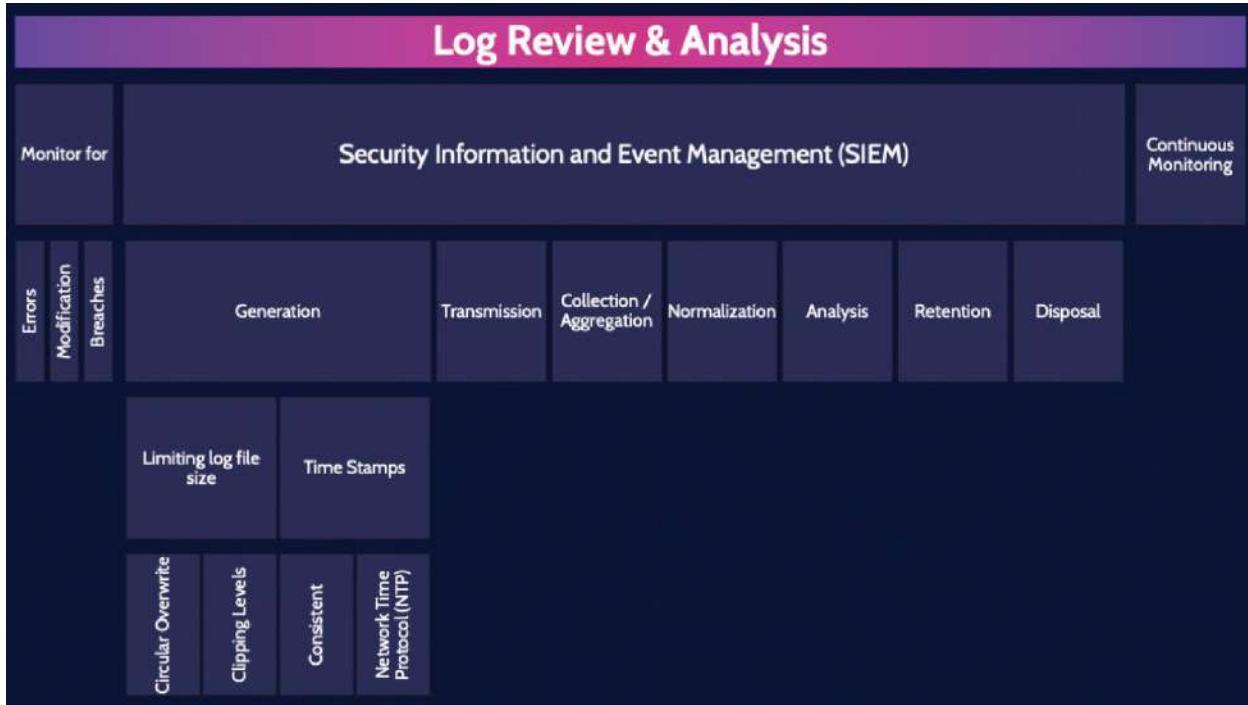
Security Assessment and Testing

dcgo.ca/CISSPmm6-1

Identifying Vulnerabilities							
Vulnerability Assessment	Penetration Testing	Process		Testing Techniques			
Reconnaissance		Enumeration		Vulnerability Analysis		Execution	
Document Findings							
	Perspective			Approach			
Internal		External		Blind			
	Double-blind			Zero (black)			
				Partial (gray)			
				Full (white)			
	Credentialed / Authenticated			Uncredentialed / Unauthenticated			
	Banner grabbing & Fingerprinting						
	CVE						
	CVSS						
	Interpreting & understanding results						
	False positive vs. False negative						

Identifying Vulnerabilities

dcgo.ca/CISSPmm6-2

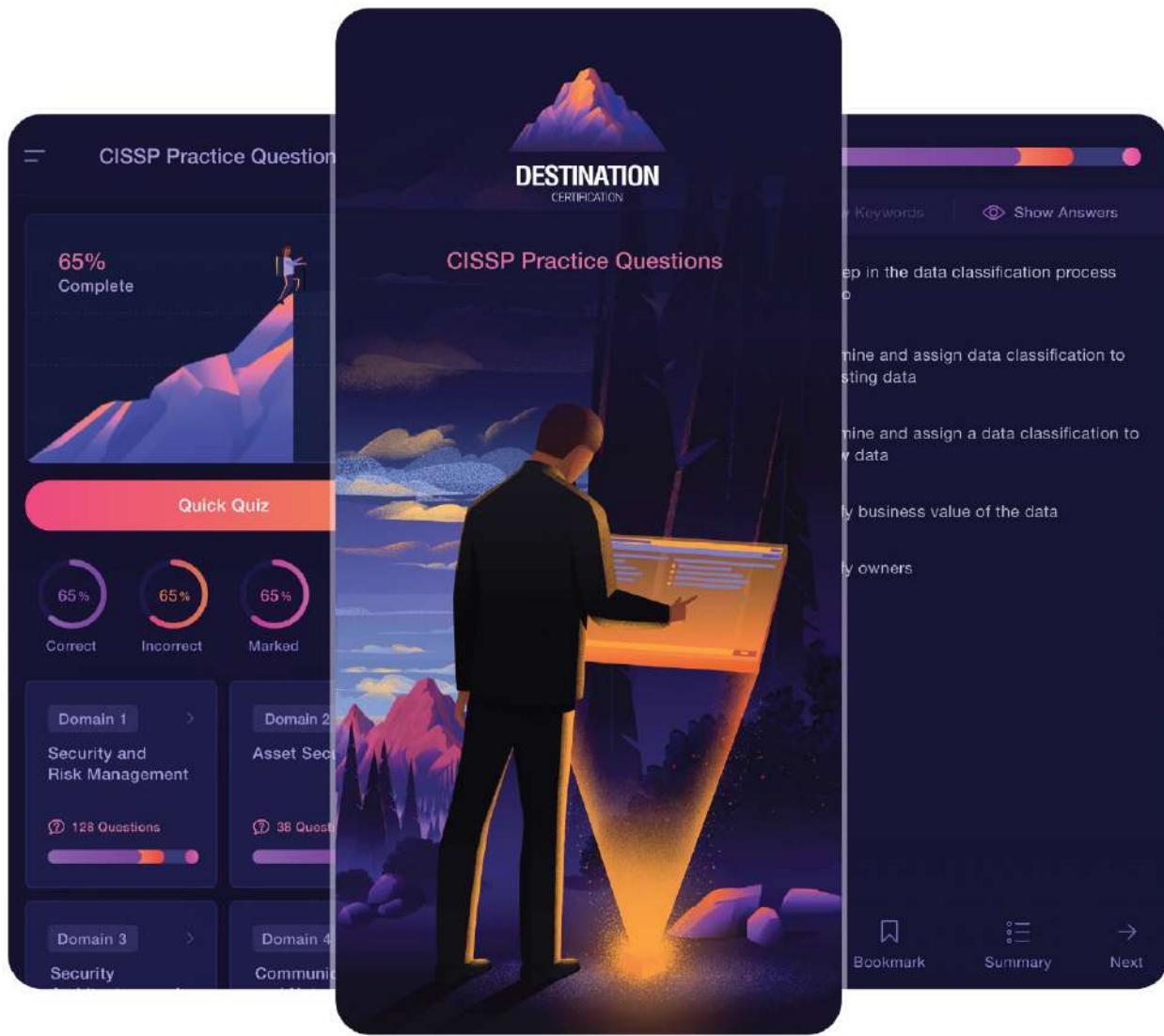


Log Review and Analysis

dcgo.ca/CISSPmm6-3

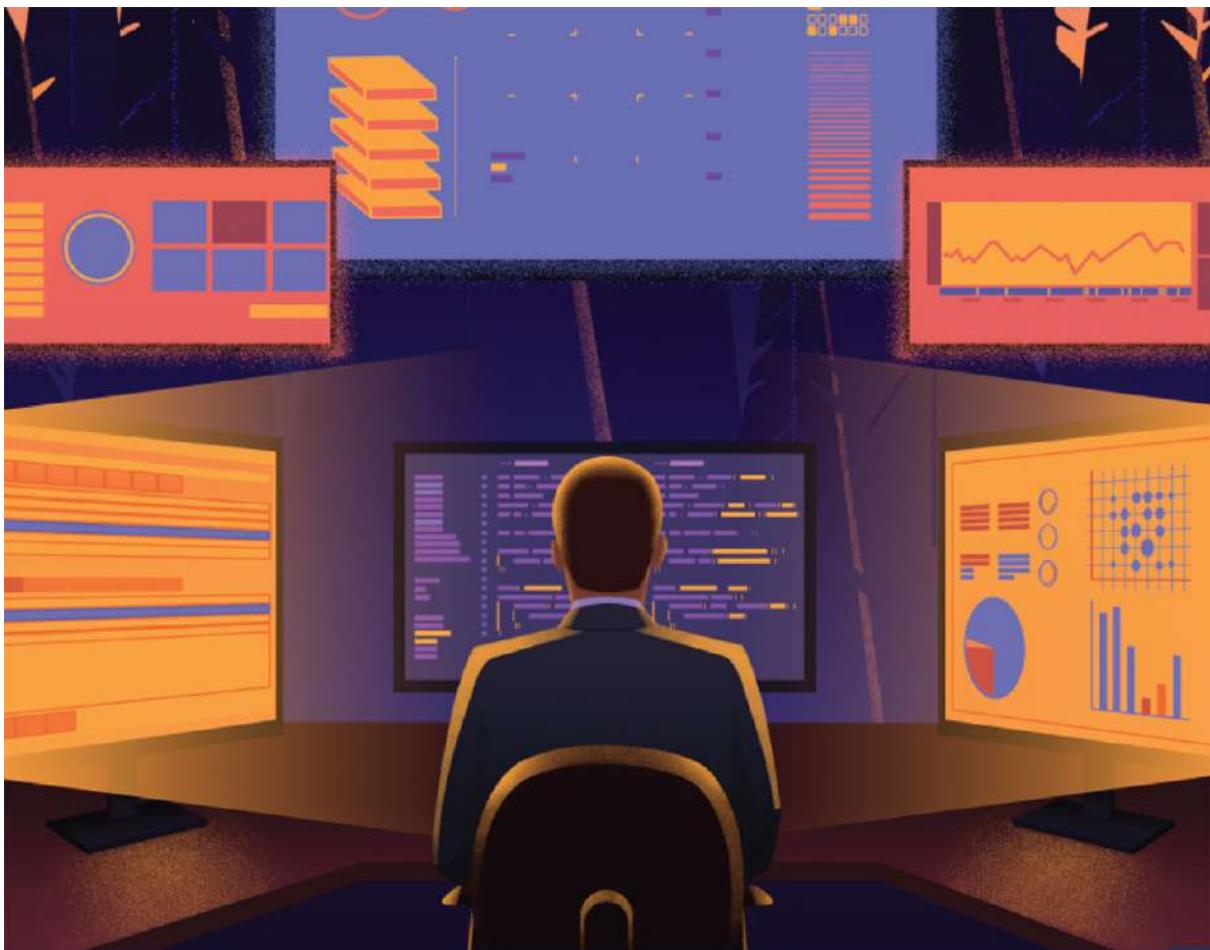


CISSP PRACTICE QUESTION APP



Download the Destination CISSP Practice Question app for Domain 6 practice questions

dcgo.ca/PracQues



DOMAIN 7

Security Operations

DOMAIN 7

SECURITY OPERATIONS

Domain 7 is where the rubber meets the road and focuses on integrating security within organizational operations. Oftentimes, this integration is in the form of processes that are put in place and configured to ensure that systems remain secure. Security operations are the day-to-day activities that the security function performs to support the entire organization in achieving its goals and objectives.

7.1 Understand and comply with investigations

7.1.1 Securing the Scene

CORE CONCEPTS

- **Securing the scene is an essential and critical part of every investigation.**
- **Securing the scene might include any/all of the following: sealing off access to the area where a crime may have been committed, taking photographs, documenting the location of evidence, and avoiding touching anything—computers, mobile devices, thumb drives, hard drives, and so on—that may have been used as part of the crime.**

One facet of security operations is investigations. Imagine you're a very experienced and capable investigator. To conduct a proper investigation, evidence needs to be

secured in the appropriate manner. Security professionals may need to support the entire investigation process and do it correctly. **The bottom line: investigators need to be able to conduct reliable investigations that will stand up to scrutiny and cross-examination.**

As an experienced and capable investigator, imagine that your organization has been breached and you've managed to track the hacker back to their criminal lair—their parents' basement. As you descend the stairs, you discover the scene from where the crime was perpetrated, and you quickly make note of the computer that was used. Are you going to walk up to it and start moving the mouse around or typing on the keyboard to start gaining insights about what this person did? Of course not, because experience has taught you that doing so might contaminate the scene. In fact, the first thing you'll do is **secure the scene** to prevent contamination from taking place. That might involve blocking off the area where the crime took place, taking photographs, documenting the location of evidence in the scene, taking snapshots of the system, and so on. With forensic computer investigations, the exact same concept applies. Otherwise, **once evidence has been contaminated, it can't be decontaminated.** *It's imperative to employ the right process and approach from the beginning.*

7.1.2 Evidence Collection and Handling

CORE CONCEPTS

- The forensic investigation process should include, among other things, the identification and securing of a crime scene, proper collection of evidence that preserves its integrity and the chain of custody, examination of all evidence, further analysis of the most compelling evidence, and final reporting.
- Evidence collection should be guided by early establishment and maintenance of the chain of custody.
- The chain of custody focuses on who handled what evidence, when, and where, and its primary focus is control, which implies integrity.
- Sources of information/evidence include: oral/written statements, written documents, computer systems, visual/audio recordings, photographs, surveillance footage, and so on.
- MOM = motive, opportunity, means.

Forensic Investigation Process

Understand the steps of the forensic investigation process and what happens at each step

The need to perform digital forensics exists within any organization that uses computer systems, networks, and the multitude of electronic devices currently available in the marketplace. Whether in response to a crime or incident, a breach of organizational policy, troubleshooting system or network issues, or a number of other reasons, digital

forensic methodologies can assist in finding answers, solving problems, and, in some cases, successfully prosecuting crimes.

As the term *methodologies* implies, digital forensic processes can vary; but, certain digital forensic science practices and standards are typically consistent, regardless of context. First among them relates to identifying and securing the scene, which is focused on protecting potential evidence from being touched, removed, or otherwise contaminated until it can be properly examined. This step also marks the beginning of the chain of custody, which is critical where a crime may have been committed and for the sake of all evidence that may ultimately be admissible as part of a trial. After a scene is identified and secured properly, the formal collection of evidence can take place. Whether dealing with physical or digital evidence, proper care must be taken to protect the integrity of whatever is collected. Forensic policies, standards, and procedures can aid with the collection process to ensure the integrity of the evidence collected as well as establish the chain of custody.

Once collected, evidence and data can be examined and analyzed via automated and manual means to determine what might be of consequential interest for the sake of building a case, identifying a culprit, or otherwise moving further along with an investigation.

Finally, as noted further below, results of the analysis should be compiled in a report. The report should describe every facet of the investigative process, from beginning to end, as well as action items to be completed, recommendations for improvement, and anything else that may prove valuable. Additionally, because multiple audiences may exist, the report may need to be compiled in different formats or with varying levels of detail for the sake of a given audience.

Sources of Information and Evidence

Sources of information and evidence as part of a computer security investigation often include oral and written statements, documents, audio/visual records, and of course, computer systems. For purposes of Domain 7, the primary focus will be computer systems, networks, and network devices. Possible sources of information are noted in [Table 7-1](#).

Oral/written statements	Statements given to police, investigators, or as testimony in court by people who witness a crime or who may have information deemed pertinent to an investigation.
Written documents	Handwritten, typed, or printed documents such as checks, letters, wills, receipts, or contracts, to name a few, that may be relevant for the sake of an investigation.
Computer systems	In the context of an investigation, a computer system could include the unit that houses the CPU, motherboard, and other system-related components that might store data in a non-volatile manner, as well as the

	storage devices—SSD, HDD (external/internal), USB device, and so on—and any other peripheral that may have been connected to a computer while a crime was committed.
Visual/audio	As part of a computer security investigation, visual and audio evidence could include photographs, video and taped recordings, surveillance footage from security cameras, and so on.

Table 7-1: Sources of Information and Evidence

In addition to sources of information, you also need to take numerous types of evidence into account. Those have been noted in [Table 7-2](#).

Real evidence	Real evidence is tangible physical objects (e.g., hard drives, SSDs, USB drives)—not the data on them. Real evidence can be physically held, touched, and inspected, and this type of evidence is often very important in a case. It is often used to prove or disprove a factual issue in a trial.
Direct evidence	Direct evidence speaks for itself and requires no inference (e.g., eyewitness accounts, confessions, a smoking gun). Direct evidence directly proves a fact being discussed. An example of direct evidence is video footage showing a defendant breaking into the computer storage area and walking out with two laptops.
Circumstantial evidence	Also referred to as indirect evidence, circumstantial evidence suggests a fact by implication or inference and can prove an intermediate fact. An

	example of circumstantial evidence is a witness testifying that the defendant was near the computer storage area after it had been broken into.
Corroborative evidence	Corroborative evidence supports facts or elements of the case, not a fact on its own, but supports other facts . Corroborating evidence can be very powerful, as it serves to uphold and confirm testimony of witnesses and other forms of evidence.
Hearsay evidence	Hearsay evidence is testimony from witnesses who were not present . No firsthand proof of accuracy or reliability exists, but the content is being offered to prove the truth of the matter at hand. Hearsay evidence is usually inadmissible in a court, unless an exception to hearsay rules is made.
Best evidence rule	The best evidence rule essentially states that original evidence rather than a copy or duplicate of the evidence should be entered as evidence.
Secondary evidence	Secondary evidence is a reproduction of, or substitute for, an original document or item of proof (e.g., print out of log files). In cases where original evidence no longer exists, a court may allow secondary evidence to be presented in a trial.

Table 7-2: Evidence Types

Motive Opportunity Means (MOM)

Another concept often employed while conducting investigations is what's known as MOM. MOM stands for motive, opportunity, and means, and it serves as a guide

when conducting an investigation. In other words, what might have motivated the suspect? Did the suspect have the opportunity to perpetrate the crime? Did the suspect have the means?

7.1.3 Locard's Exchange Principle

CORE CONCEPTS

- **Locard's exchange principle: with every crime, something is taken and left behind**

Understand what is meant by Locard's exchange principle

A simple yet effective method for identifying where to look for evidence is the notion that, whenever a crime is committed, something is taken and something is left behind. This explains why pictures are taken, carpets are vacuumed, fingerprints lifted, and crime scenes are otherwise meticulously examined. A nineteenth-century French criminologist, Dr. Edmund Locard, formulated the foundation of forensic science based upon this fact. In simple terms, this means that every time two objects interact, some type of transfer occurs—something is taken and something is left behind.

7.1.4 Digital/Computer Forensics

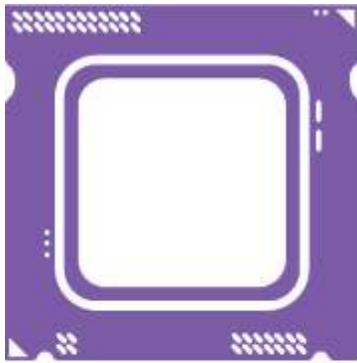
CORE CONCEPTS

- **Digital forensics is the scientific examination and analysis of data.**
- **Live evidence is data that is stored in a running system in places like random access memory (RAM), cache, buffers, and so on.**
- **Forensic copies refer to identical, bit-for-bit copies of a digital media source, like a hard drive.**
- **Digital forensics tools, tactics, and procedures facilitate proper and immediate response to live systems.**
- **Artifacts are remnants of breach or attempted breach and can act like breadcrumbs that point to the path followed or activities pursued by an attacker.**

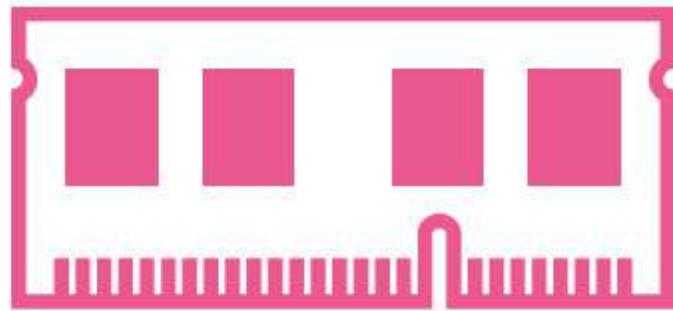
Understand the implications and challenges of working with live evidence

With computer security, the primary focus is computer, or digital, forensics. **Digital forensics** is the scientific examination and analysis of data from storage media in such a way that the information can be used as part of an investigation to identify the culprit or the root cause of an incident.

Live Evidence



CPU Cache



RAM

With digital forensics, one of the primary considerations is what's referred to as live evidence. Live evidence is data that is stored in a running system in places like random access memory (RAM), CPU, cache, buffers, and so on. If the keyboard is tapped, if the mouse is moved, and certainly if the plug is pulled or the system is powered off, the live evidence changes or disappears completely. ***Examining a live system changes the state of the evidence.*** This fact makes it immediately clear that examination of a live system requires expert knowledge and specialized tools to extract live evidence and minimize contamination. As noted, live evidence is often stored in locations like RAM, cache, and the CPU, and if power to the system is disrupted, the live evidence is gone.

Bit for Bit copies



Forensic Copies

Another major source of digital evidence on a computer system is the hard drive. For example, imagine you're investigating a crime and have discovered a laptop that was likely used to commit the crime. The laptop is already turned off, so there are no concerns about contaminating live evidence. However, the hard drive might contain evidence, so you remove it to conduct forensic analysis.

*Whenever a forensic investigation of a hard drive is conducted, two identical **bit-for-bit copies** of the original hard drive should be created first.* Then, the original hard drive should be placed in an evidence bag, the bag sealed, and the drive never touched again. Similarly, the first copy

of the hard drive should be treated the same. The second copy of the drive is the working copy.

Understand the importance of creating bit-for-bit copies of a hard drive

What does “bit-for-bit copy” mean? It simply means it is an exact copy, down to every bit on the original drive, and specialized tools are required to create bit-for-bit copies. After the bit-for-bit copies are created, the original drive and the two copies should be hashed. If the hash values match, the copies are exact, bit-for-bit copies. That ensures integrity is maintained.

Why is it harder to do forensic analysis of mobile devices?

Understand why forensic analysis of computers and mobile devices can be difficult

- Manufacturers **frequently change** operating system structure, file structure, services, and connectors ■
- No **single method or tool** can extract all the data ■
- Hibernation and suspension of applications ■
- Extensive new training required for examiners

Reporting and Documentation

The final phase of the forensic investigation process is reporting and documentation, though documentation should be an integral component of every step in the process. At this point, all evidence has been examined, and the most relevant evidence should be documented for the sake of use by all relevant stakeholders, including:

- Prosecution/Defense
- Judge/Jury
- Regulators
- Investors
- Insurers

Artifacts

Understand the importance and potential relevance of artifacts to an investigation

Forensic artifacts are remnants of a breach or attempted breach of a system or network, and they may or may not be relevant to an investigation or response. They're breadcrumbs that can potentially lead back to an intruder or at least identify their actions and the path they followed while in the system or network. Artifacts can be found in numerous places, including:

- Computer systems
- Web browsers
- Mobile devices
- Hard drives
- Flash drives

Examples of artifacts include IP addresses, hashes, file name/type, registry keys (Windows), URLs, operating system information, as well as logged information, like account updates, profile changes, file changes, and so on, that point to malicious behavior. Undoubtedly with so many potential

sources of artifacts available, identifying relevant artifacts can be akin to finding a needle in a haystack—a very large haystack—and the forensic investigator must be very skilled and careful in evaluating what is most pertinent and therefore most valuable for the sake of the investigation. Artifacts that support or refute a hypothesis related to an investigation or response can be used as evidence.

7.1.5 Chain of Custody

CORE CONCEPTS

- The primary focus of the chain of custody is control of evidence to maintain integrity for the sake of presentation in court

Chain of Custody

Understand the “chain of custody” and its importance

One very important aspect of evidence collection is that the **chain of custody** must be established immediately and maintained. The *chain of custody* is ultimately focused on having **control** of the evidence: who collected and handled what evidence, when, and where. Crime scenes should always be thoroughly documented via photos and diagrams. Additionally, evidence should be collected in a manner that protects it from tampering, contamination, or other things such as corruption or deterioration. It's

imperative that evidence be controlled from the moment of collection to the moment it might be presented in court, which could potentially be years later. Regardless of the time frame, if the chain of custody is maintained, evidence will have a better opportunity to be admissible.

A useful way to think about establishing the chain of custody is to tag, bag, and carry the evidence as depicted in [Figure 7-1](#). Tag the evidence to clearly note where the evidence was collected, on what date, and by whom. Bag the evidence—carefully store the evidence to minimize contamination. Carry the evidence back to a secured evidence storage location, such as an evidence locker.



Figure 7-1: **Chain of Custody**



7.1.6 Five Rules of Evidence

CORE CONCEPTS

- **The five rules of evidence state that evidence should be: authentic, accurate, complete, convincing or reliable, and admissible.**
- **To best ensure that the five rules are achieved, evidence chain of custody must be maintained.**

For any evidence to stand the best chance of surviving legal and other scrutiny, it should exhibit five characteristics, also known as the “five rules of evidence,” described in [Table 7-3](#).

Understand the five rules of evidence and their meaning

Authentic	Evidence is not fabricated or planted and can be proven so through crime scene photos or things like bit-for-bit copies of hard drives. This points back to securing the scene, to ensure the best chance of preserving all critical pieces of evidence for the sake of an investigation and any legal proceedings.
Accurate	Evidence has not been changed or modified—it has integrity .
Complete	Evidence must be complete , and all parts presented. In other words, <i>all</i> pieces of evidence must be available and shared, whether they support or fail to support the case.
Convincing or Reliable	Evidence must be conveyed in a manner that allows anybody to understand what is being presented. Evidence must display a high degree of veracity—it must demonstrate a high degree of truth. Additionally, nontechnical people, including judges and juries, must be able to understand what is being presented.
Admissible	Evidence is accepted as part of a case and allowed into the court proceedings. Chain of custody can help, but it does not guarantee admissibility.

**Table 7-3: Five Rules of Evidence
Investigative Techniques**

Several investigative techniques can be used when

conducting analysis.

One of them is media analysis. Media might include the analysis of things like hard drives, flash drives, tapes, CDs, USB drives, or anything similar. With media analysis, oftentimes the search is for what's not there as much as what is there. For example, when examining a hard drive, if someone has deleted a file, is that file gone from the hard drive? Oftentimes the pointer to the file has simply been deleted, but the file is still there. Media analysis examines the bits on a hard drive that may no longer have pointers, but the data is still there.

Software analysis focuses on an application, especially malware. With malware, the goal is to determine exactly how it works and what it is trying to do. An important facet of this relates to attribution and trying to determine who or where the software was created. Oftentimes, the source code can offer clues and pointers that help to pinpoint this information.

Network analysis attempts to understand how a network might have been penetrated, how the network was traversed, and what systems may have been breached. Typically, system log files provide the best source of information for network analysis.

7.1.7 Types of Investigations

Table 7-4 provides a summary of types of investigations, relating to an incident.

	Overview	Who drives the investigation?
Criminal	Deal with crimes and oftentimes with accompanying legal punishment . Convictions often lead to time in jail as well as a criminal record. These are conducted by law enforcement at the local, state, and federal levels, depending upon the nature and severity of the crime, and punishment can potentially be very harsh.	Primarily law enforcement with support from the organization
Civil	Deal with disputes between individuals or organizations , and whichever party is found guilty usually pays a fine or other monetary penalty as well as related court costs.	Organizations, individuals and their attorneys
Regulatory	Deals with violations of regulated activities	Associated regulatory body
Administrative	Focus on internal violations of organizational policies and incidents identified by an	The organization

	organization. Perhaps employee misconduct was involved, or policies or procedures were broken, or a hacker absconded with some credit card numbers. Unless it's determined that criminal activity resulted, administrative investigations are opened and closed by the organization itself.
--	---

Table 7-4: Types of Investigations

In the case of criminal and civil investigations in the context of an organization, the investigation might start internally. Once it becomes clear that criminal activity has taken place, law enforcement should be contacted, at which point the investigation would be handed over to them, and they would drive the investigation forward from there.



7.2 Conduct logging and monitoring activities

7.2.1 Security Information and Event Management (SIEM)

CORE CONCEPTS

- Security Information and Event Management (SIEM) systems ingest logs from multiple sources, compile and analyze log entries, and report relevant information.

Understand at a high level what a SIEM system is and its capabilities

The topic of logging and monitoring was first introduced in Domain 6, and now we're going to look at it more closely in the context of security information and event management

(SIEM) systems. SIEM systems ingest logs from disparate devices throughout an organization, they aggregate and correlate the log entries and look for interesting activity. Relevant findings are reported, so additional action can be taken. At a high level, [Figure 7-2](#) shows how a SIEM system operates.

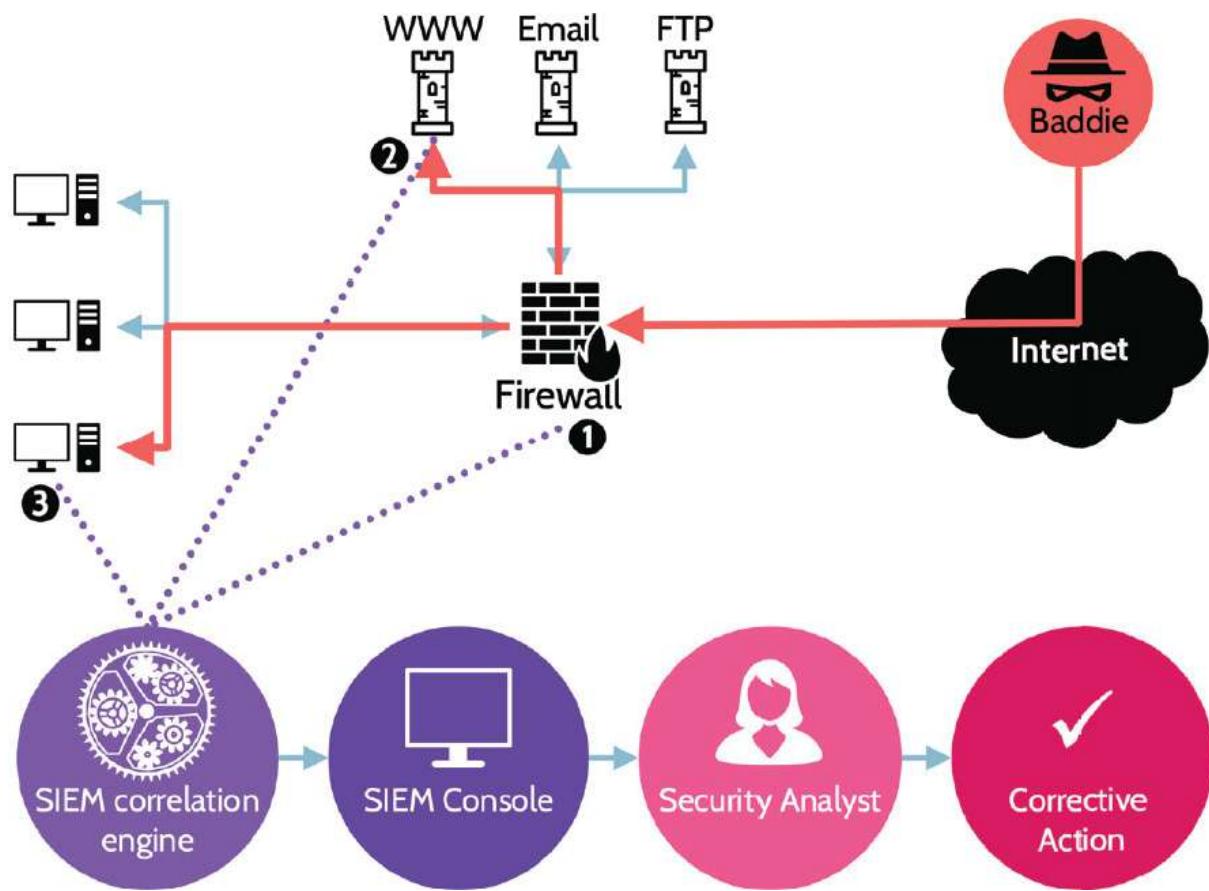


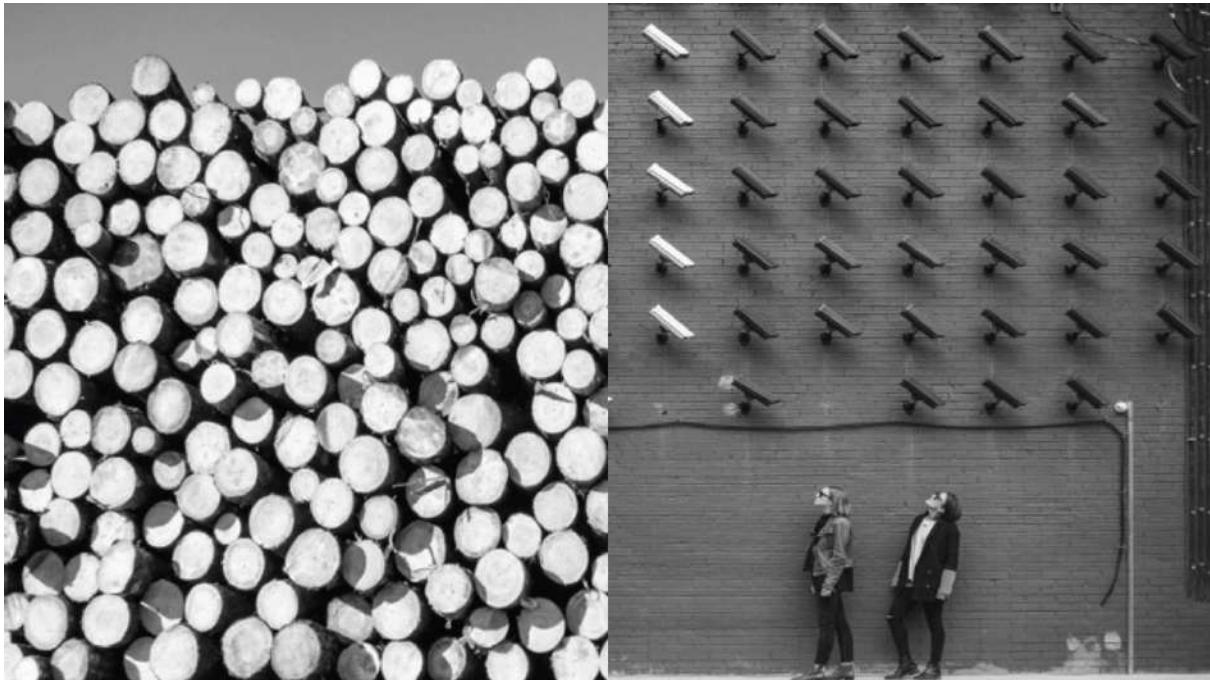
Figure 7-2: **SIEM Operation**

Let's examine a simple example of how this might work.

Imagine an attacker who starts poking at your network to determine what's where and what can be accessed. As this

activity takes place, log events are being generated. A log event is simply a record of any event of interest. Most events that a SIEM ingests are meaningless, even some of the events being generated by the hacker, because the hacker is simply poking around. Now, let's further imagine that the hacker finds an entry point and successfully gains access to a web server, and from the web server they're able to locate a back channel that leads to internal systems. These additional activities, like before, continue to generate events, but now—when correlated to earlier events—they take on new meaning and, very likely, generate some type of alert that a security analyst would be tasked with examining further. Very quickly, the analyst is going to try and determine if this is a false-positive or if something malicious is really taking place, and this highlights an important point about SIEM systems in general and their installation and operation.

SIEM systems are much more than just technology. Yes, the technology is important and complicated and very hard to implement, but it's more than these facts. SIEM systems also require trained personnel. Let's look briefly at a real-life example to explain.



Company A (generic name used to protect the innocent) suffered two major breaches. After the first breach, they installed several expensive and sophisticated systems, which helped detect the second breach. In fact, two different teams—one in the United States and one in India—detected the breach. But there was a breakdown, and despite the technology that alerted about the breach, the alert was not properly escalated and was ultimately ignored. So, the technology was in place, it detected the hackers, and it even alerted the security team; then everything fell apart with the process of escalation and dealing with the situation. Again, this example highlights the fact that SIEM systems are complex and require expertise to install and tune properly, and they require a properly trained team that understands how to read and interpret what they're seeing

as well as what escalation procedures to follow when a legitimate alert is raised. SIEM systems represent technology, process, and people, and each is relevant to overall effectiveness.

Figure 7-3 depicts another example.

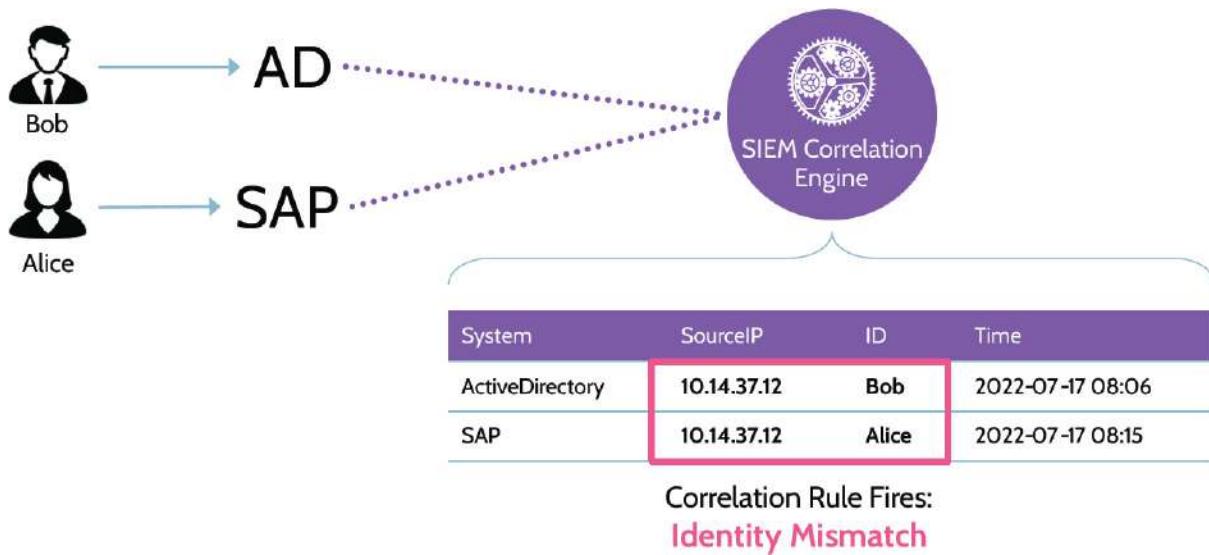


Figure 7-3: **SIEM Example**

Imagine two users, Bob and Alice, both working from home. Bob's an active user with a valid account, and he logs into Active Directory (AD). When Bob logs into Active Directory, an event is logged and captured by the SIEM system. Several minutes later, Alice logs into SAP. Alice is also a legitimate user with a valid account, and the event is similarly logged to the SIEM system. However, this time the SIEM system generates an alert because the IP address associated with Alice is the same as the IP address

associated with Bob, which wouldn't make sense because they both live at different homes. This is not normal and may indicate a problem that needs to be further evaluated by an analyst. For example, an attacker may have compromised legitimate user accounts and is using those to access company systems. At the very least, this looks suspicious. Without the log events being correlated in the context of the SIEM system, this anomaly wouldn't have been identified because events logged in Active Directory and events logged in the SAP system are separate. However, once ingested and aggregated by the SIEM system, correlation took place, and the once disparate events now tell a different story. This functionality is one of the major advantages of SIEM systems.

Both examples point to the real power of SIEM systems. With traditional logging, where log files are only captured and maintained on individual systems, it would take enormous effort to access and analyze all the captured data and determine if anything out of the ordinary or malicious is taking place. It's like trying to find the proverbial needle in the haystack. With a SIEM system, significant intelligence is incorporated into their functionality, which allows significant amounts of logged events and analysis and correlation of the same to take place very quickly. And with the proper tuning, SIEM systems can identify and alert on even the least apparent telltale sign.



Figure 7-4: SIEM Capabilities

Let's look a bit deeper at SIEM system capabilities as also highlighted in [Figure 7-4](#).

- SIEM systems allow for the **aggregation** of logged events from multiple systems. In other words, events logged in systems located throughout an organization's network can all be brought under one umbrella—the SIEM system.
- Once aggregated, logs typically need to be **normalized**, because different systems log events using different formats. For example, one system might log events using a twelve-hour clock, while another system might use a twenty-four-hour clock, or the dates might be in month/day/year format on one system and day/month/format on another. Events should be deduplicated, or simply deduped. This means that duplicate events are eliminated. Normalization helps clean up data, put it in the same format, and eliminate redundancy, so it can be easily analyzed and suspicious activity flagged,

based upon rules that have been programmed into the system.

- After data has been normalized, **correlation** seeks to line up events and determine which of them alone and in combination might be important and indicative of problems.
- **Secure storage** is the concept where the SIEM system keeps a copy of all logged events from each device. The system is designed to store data for long periods of time, and ideally those log files are read-only, to prevent tampering or deletion.
- **Analysis** and **reporting** refer to the SIEM system rules that have been put in place, looking at data related to the same, and taking action when appropriate. As noted earlier, event log data can be collected from virtually every system within an organization. However, SIEM systems should be used to capture relevant logs, based on organizational risk, required budget, and regulatory obligations.

Example Sources of Event Data

As noted above, SIEM systems allow log data from multiple sources, such as those noted below, to be captured in one location.

- Security appliances ■ Network devices ■ DLP
- Data activity ■ Applications ■ Operating systems ■ Servers ■ IPS/IDS

Threat Intelligence

The term “threat intelligence” is an umbrella term encompassing threat research and analysis and emerging threat trends. It is an important element of any organization’s digital security strategy that equips security professionals to proactively anticipate, recognize, and respond to threats. Many SIEM solutions offer threat intelligence subscriptions that add additional capabilities, strength, and value to already robust systems. However, actionable threat intelligence can also be gleaned from documents like vendor trend reports, public sector team reports (like US-CERT), related information sharing and analysis centers (ISACs), and more.

User and Entity Behavior Analytics (UEBA)

UEBA (also known as UBA) analysis engines are typically included with SIEM solutions or may be added via subscription. As the name implies, UEBA focuses on the analysis of user and entity behavior. At its core, UEBA monitors the behavior and patterns of users and entities, logs and correlates the underlying data, analyzes the data, and triggers alerts when necessary. The analytics

component of a UEBA solution is based on machine learning, which allows a baseline for each user and entity to be created. If future behavior deviates from what is considered normal, an alert can be fired, and potential further action can be taken based upon the perceived or real threat. UEBA solutions can be used to address insider threats, hacked privileged accounts, or brute-force attacks, to name a few examples, and the sophisticated manner through which UEBA can detect behavioral shifts and anomalies and alert a security team before a breach occurs or progresses too far can potentially prove invaluable to an organization.

7.2.2 Continuous Monitoring

CORE CONCEPTS

- After a SIEM is set up, configured, tuned, and running, it must be routinely updated and continuously monitored to function most effectively.
- Effective continuous monitoring encompasses technology, processes, and people.

Understand the concept of continuous monitoring and the value it provides to an organization

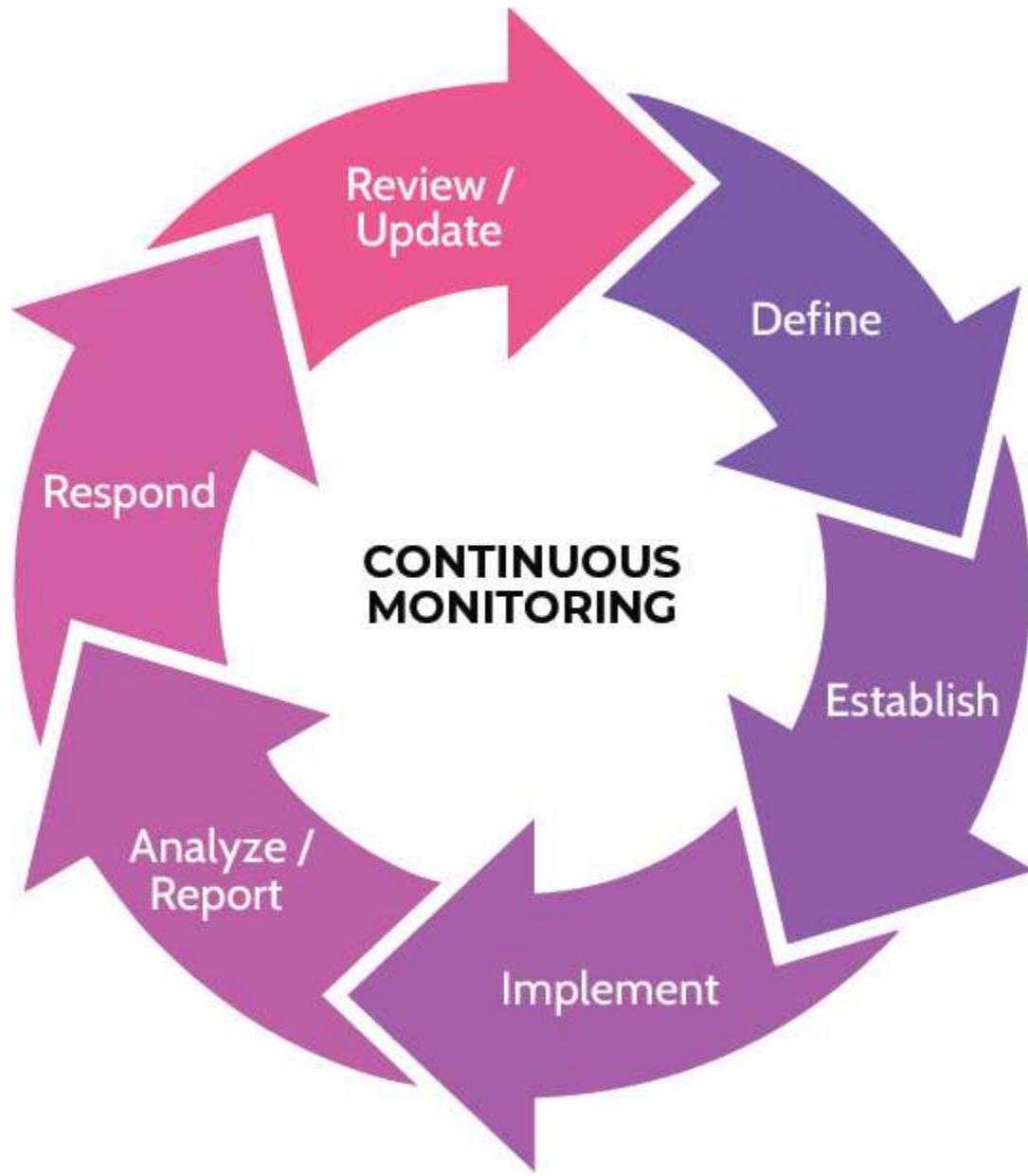


Figure 7-5: **Continuous Monitoring Steps**

Setting up a SIEM system can be a very long and arduous process, sometimes taking months or even longer, depending on the complexity of the environment and needs of the organization. However, once the system has been configured and is running, the work is not complete.

In fact, the SIEM system must be updated and monitored continuously, because ■ The threat environment is constantly changing, ■ New vulnerabilities are constantly emerging, ■ Assets in the organization are changing, ■ New monitoring rules need to be configured and programmed, ■ The balance between false-positives and false-negatives must be closely monitored and responded to accordingly.

Understand what continuous monitoring encompasses

Finally, and perhaps most importantly, focus should be not only on the technology and related processes, but also on the people utilizing the technology and the escalation processes, so that timely and proper responses can take place before a breach can cause significant damage to the organization. The full continuous monitoring life cycle has been provided in [Figure 7-5](#).

In addition to the topics and subject matter covered here in 7.2, the Official CISSP Certification Exam Outline includes other topics in this section that are covered elsewhere in the book.

Please refer to Domain 4 for details on intrusion detection and prevention, and egress (and ingress) monitoring.

Please refer to Domain 6 for details on log management.

7.2.3 Security Orchestration, Automation, and Response (SOAR)

SOAR is a collection of compatible technologies that take input and data from disparate sources—such as other devices, email, Security Information Event Management (SIEM) systems, user submissions, and manual input—and apply rules and workflows aligned to organizational processes and procedures.

SOAR focuses on three key areas:

- Threat and vulnerability management ■ Incident response ■ Security operations automation SOAR tools can be integrated with other technologies and automated to orchestrate a desired outcome and better visibility. SOAR tools typically include incident and threat intelligence management features, reporting, and data analytics. Additionally, through machine-learning and machine-powered assistance, SOC activities and threat detection and response by analysts can be efficiently and consistently improved.

7.3 Perform configuration management (CM) (e.g., provisioning, baselining, automation)

7.3.1 Asset Inventory

CORE CONCEPTS

- Provisioning relates to the deployment of assets—hardware, software, devices, and so on—with in an organization.
- Part of provisioning should include maintaining and updating a related asset inventory database anytime an asset is added or removed.
- Assets should be managed as part of an overall asset management life cycle.



Figure 7-6: **Asset Management**

The full asset management life cycle can be seen in [Figure 7-6](#).

Items can be requested to be procured after careful planning takes place about what's needed in the

environment. However, they also need to be securely provisioned. That refers to how things like hardware, software, devices, and so on are provisioned or deployed. Secure provisioning is about the deployment process. For example, when a firewall is purchased, it's typically configured with default settings by the vendor. These default settings usually include default admin and user accounts and passwords as well as other configuration settings. It's not a best practice to deploy any device with default settings. In fact, a process already discussed—system hardening—should be used to ensure that the new system is properly secured and deployed according to policy and baselines established by the organization.

Additionally, whenever deploying a new asset—whether hardware or software—an associated asset inventory database should be updated. Assets represent an organization's attack surface. In other words, each asset is something of value that an attacker might target. Without a current asset inventory, including details about the asset owner, the assets have a very low chance of being patched, configured, scanned routinely, or otherwise kept up to date. It's easy to see that secure provisioning ties into the concept of asset management and the asset life cycle very closely.

7.3.2 Configuration Management

CORE CONCEPTS

- Configuration management is an integral part of secure provisioning.
- Configuration management relates to the proper configuration of a device at the time of deployment.
- Policies, standards, baselines, and procedures inform configuration management.
- Hardening should be considered as part of the configuration management process.
- Automated provisioning tools can help ensure consistency with the configuration and deployment process and save time.

Understand the value and key benefits of configuration management

A significant aspect of secure provisioning is configuration management. As the term implies, configuration management focuses on the proper configuration when a device is first deployed, and this is where tools like baselines, policies, and standards are utilized. Hardening—ensuring that only the necessary services and features of a given hardware device or piece of software are available—should be part of the configuration management process. Additionally—especially in large organizations—automated tools can be used for provisioning purposes. The use of

automation can help ensure consistency with the provisioning process as well as save time.

Device configurations should be documented and reviewed on a periodic basis, and things like credentialed vulnerability scans can be used as a way to review hardware and software configurations. In fact, entire software suites exist for purposes of examining configurations of almost any type of system in an organization.

- Identify assets to keep under control
 - Configure assets
 - Document configuration
 - Verify configuration
- 7.4 Apply foundational security operations concepts**

7.4.1 Foundational Security Operations Concepts

CORE CONCEPTS

- Implementation of foundational security operations concepts can significantly improve security within an organization.
- Foundational security operations concepts include: need to know/least privilege, separation of duties (SoD) and responsibilities, privileged account management (PAM), job rotation, and service level agreements (SLA).

Understand foundational security operations concepts

Privileged account management: Privileged accounts refers to system accounts that typically have significant power. System administrator accounts are one example of a privileged account, and they often give what's known as "root" or "admin" access to a system. In other words, the account can access and control every part of the system. In the wrong hands this type of control can obviously lead to significant problems. Therefore, privileged accounts should be very carefully managed. For one, access to them should be restricted as much as possible. For another, within an IT department, for example, personnel should have regular user accounts as well as accounts with increased privileges that are only used when needed. Additionally, privileged accounts should require additional authentication, like multifactor authentication. This should be the rule, not the exception. Finally, the use of privileged accounts should always be accompanied by increased logging and monitoring.

Need to know and least privilege were discussed in [5.1.1](#), and [Table 7-5](#) provides a summary of these terms.

Need to Know	Least Privilege
Restricting a user's KNOWLEDGE (access to data) to only the data	Restricting a user's ACTIONS/PRIVILEGES to only those

required for them to perform their role	required for them to perform their role
---	---

Table 7-5: Need to Know and Least Privilege

Job rotation: Another excellent fraud detection and deterrent technique is what's known as job rotation. When an organization employs job rotation, they're essentially telling employees that from time to time, another employee will assume their duties and they'll assume the duties of somebody else. Not only does this help prevent fraud and the perpetration of crimes, but it can also help ensure accurate processes as well as cross-training of employees to avoid a single point of failure.

Service Level Agreements (SLAs): SLAs are legal contracts and are part of the overall contract between a customer and vendor. They contain terms denoting related time frames against performance of specific operations that have been agreed upon. For example, an SLA could require a vendor to respond to a certain type of incident within one hour.

7.5 Apply resource protection techniques

7.5.1 Protecting Media

CORE CONCEPTS

- Media management should consider all types of media as well as short- and long-term needs.
- Mean Time Between Failure (MTBF) is an important criterion to consider when evaluating storage media, especially where valuable or sensitive information is concerned.
- Media protection techniques considers media and media management needs and incorporates several tools for the sake of protection of media.

Within any organization, data is one of the most important assets, and it follows that protection of data is critical for the ongoing success of the organization. This fact, therefore, presents some unique needs as well as some challenges. For one thing, data may be stored on a variety of media. For another, data may need to be kept for very long periods of time.

Media

Depending upon how data is being used, storage requirements, portability, and other factors, data storage media might include any of the following:

- Paper
- Microforms (microfilm and microfiche)
- Magnetic (HD, disks, and tapes)
- Flash memory (SSD and memory cards)
- Optical (CD and DVD)

The **Mean Time Between Failure (MTBF)** of media can help determine the best media to use for a given need, but no media is going to reliably last for significantly long periods of time. To best retain and protect data for very long periods of time, **processes** must be put in

place that constantly move the data to new media, and **file formats should be updated in order to maintain compatibility** with applications that can manage the data. Additionally, and most importantly, the **protection of the data must be updated to reflect current cryptography standards** versus standards that may have been employed when the data was first created.

Media Management

When employing a media management process, some key considerations should include the items listed below, and they should try to look into the future as far as possible.

- Confidentiality ■ Access speeds ■ Portability ■ Durability ■ Media format ■ Data format For example, looking at confidentiality, perhaps a certain cryptographic algorithm would more than suffice to protect the data today, but will it suffice in ten years? If data needs to be protected that long, or longer, perhaps the strongest algorithm available should be used instead.

Media Protection Techniques

Associated with media management is protection of the media itself, which typically involves policies and procedures, access control mechanisms, labeling and marking, storage, transport, sanitization, use, and end of life.

Which elements are involved and to what degree they are used points back to the value of the data stored on the media, relative to the goals and objectives and associated risk management process of the organization.

Hardware and Software Asset Management

Closely aligned with the information above is the management of hardware and software assets. To best manage either asset class, an inventory of all assets must exist, and it must be maintained. An owner should be assigned to each asset, with each owner accountable for protecting that asset, including things such as patching (software and firmware), maintaining proper licensing, and determining the most appropriate and secure configuration before deployment and on an ongoing basis. To summarize, the items noted below must be present.

- Asset management life cycle ■ Inventories ■ Patching ■
- Software licensing ■ Secure configuration



7.6 Conduct incident management

7.6.1 Incident Response Process

CORE CONCEPTS

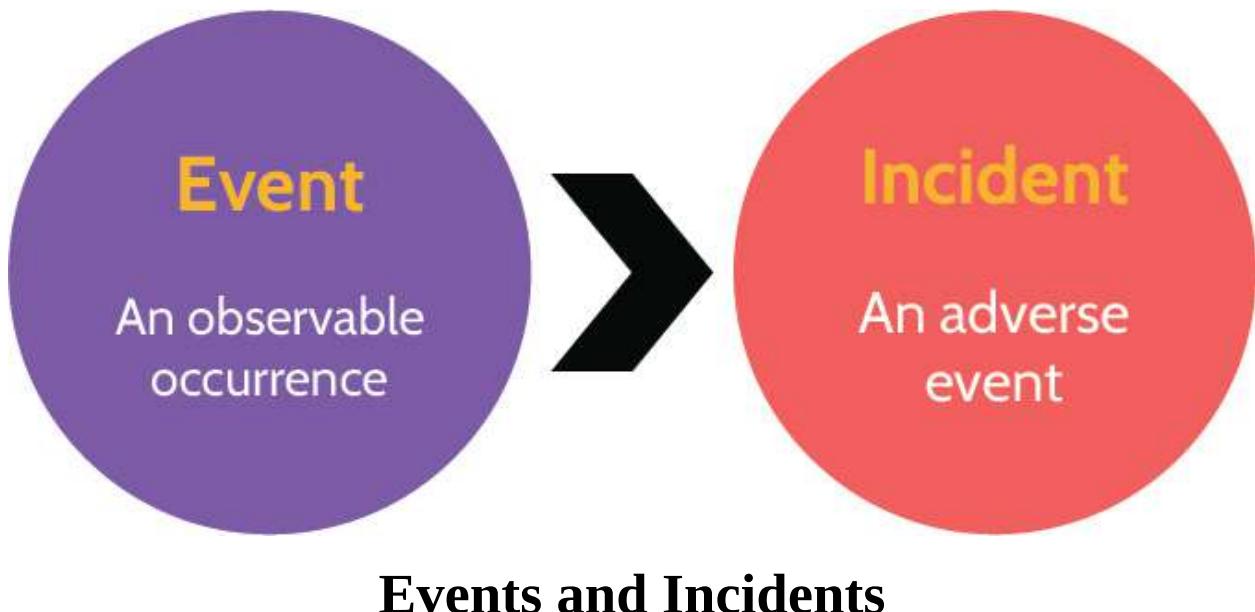
- **Incident response is the process used to detect and respond to incidents.**
- **Event = observable occurrence of something.**
- **Incident = an adverse event.**
- **Not all events are incidents.**
- **Incident response process includes: preparation, detection, response, mitigation, reporting, recovery, remediation, and lessons learned.**

Incident response is the process used to detect and respond to incidents and to reduce the impact when incidents

occur. Incident response attempts to keep a business operating or to restore operations as quickly as possible in the wake of an incident. Furthermore, for the sake of ongoing operations, incident response seeks to learn from mistakes and strengthen the organization as a result.

Goals of Incident Response

- Provide an effective and efficient response to **reduce impact** to the organization ■ Maintain or restore **business continuity**
- **Defend** against future attacks



To know when an incident response process should be initiated, an important distinction needs to be made. Namely, what distinguishes an ***incident*** from an ***event***. Events take place on a continual basis, and the vast majority of these are insignificant; however, events that lead to some type of adversity can be deemed incidents, which should then trigger an organization's incident response process.

Detection Examples

Organizations must have tools in place that can help detect and identify incidents. Most organizations use one or more of the tools noted below for detection purposes, and some of them—like IPS/IDS, DLP, and SIEM—require quite a bit of configuration and tuning to work optimally. Thus, a combination of automated and manual tools is usually the best and most effective approach.

- IPS/IDS
- DLP
- Anti-malware ■ SIEM
- Administrative review ■ Motion sensors ■ Cameras ■ Guards

Examples of Incidents

Incidents can take on many shapes and forms, and the examples below show a good mix of what might be called an incident. It's not important to memorize this list, as it's certainly not exhaustive; it simply serves to

illustrate what types of incidents might be detected that require a formal response via the incident response process.

- Malware ■ Hacker attack ■ Insider attack ■ Employee error ■ System error ■ Data corruption ■ Workplace injury

Process Steps

The incident response process is outlined in [Figure 7-7](#) and then further defined in [Table 7-6](#).

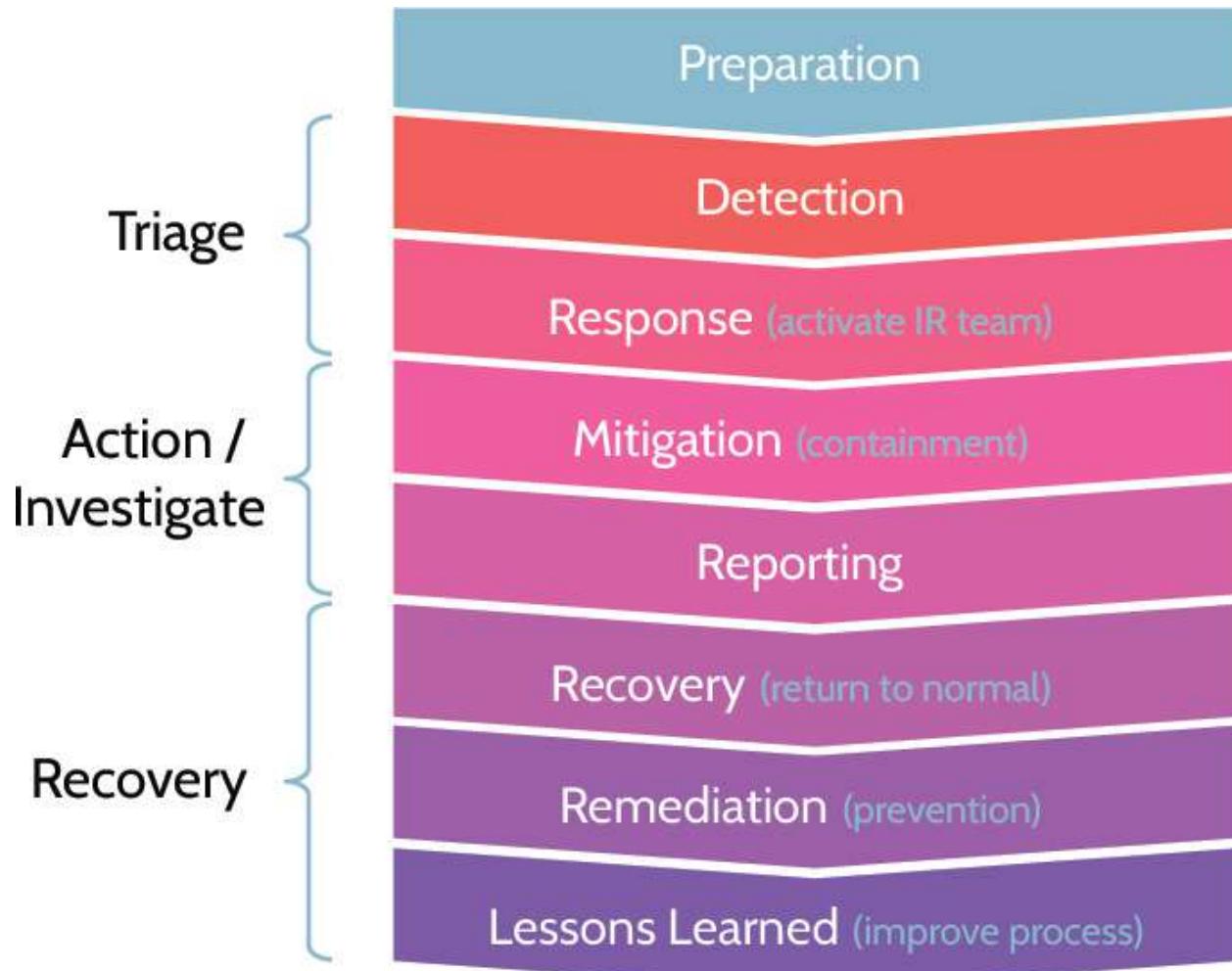


Figure 7-7: Incident Response Process

Preparation	Preparation is critical, as this will help anticipate all the steps to follow. Preparation would include things like developing the IR process, assigning IR team members, and everything related to what happens when an incident is identified.
Detection	Pointing back to the distinction between an event and an incident, <i>the goal of detection is to identify an adverse event—an incident—and to begin dealing with it.</i>
Response (IR Team)	After an incident has been identified, the IR Team should be activated. Among the first steps taken by the IR Team will be an impact assessment to determine how big of a deal is the incident, how long might the impact be experienced, who else might need to be involved, and so on.

Mitigation (containment)	In addition to conducting an impact assessment, the IR Team will attempt to minimize—to contain—damage or impact from the incident. The IR Team’s job at this point is not to fix the problem; it’s simply to try and prevent further damage from taking place. For example, if a fire has broken out, the IR Team will focus on ensuring human safety, extinguishing the fire, and confirming that no hot spots exist that could flare up again.
Reporting	Reporting occurs throughout the incident response process. Once an incident is mitigated, formal reporting takes place, because there are often numerous stakeholders that need to understand what has happened. Especially in situations where a major incident has taken place, these stakeholders could include senior management, Board members, HR, Legal, IT, customers, vendors, PR, and even outside media outlets. As there are often numerous stakeholders seeking updates, this can quickly distract the IR Team from focusing on their role and responding to the incident, so it’s important that one person be designated as the point person for purposes of reporting. Taking this approach helps keep the message consistent, and it allows those most directly involved with responding to the incident to stay focused on their jobs.
Recovery (return to normal)	At this point, the goal is to start returning things to normal, to getting back to business as usual. For example, looking back at the fire example, this is when cleanup of water and debris would take place, walls and ceilings replaced, systems put back in place, and so on.
Remediation (prevention)	In parallel with recovery, remediation is also taking place. The goal of remediation is to implement fixes and improvements to systems and processes to prevent a similar incident from occurring again.
Lessons Learned (improve process)	<p>Lessons learned steps back and takes a more all-encompassing view of the situation related to an incident and asks questions like:</p> <ul style="list-style-type: none"> ■ “What additional processes can be put in place?” ■ “How did our organization get here?” ■ “What can be done differently to improve how we run and protect our organization?” <p>The findings from the lessons learned are used to further improve systems and processes to try to prevent future incidents from occurring.</p>

Table 7-6: Definitions of Incident Response Steps

7.7 Operate and maintain detective and preventive measures

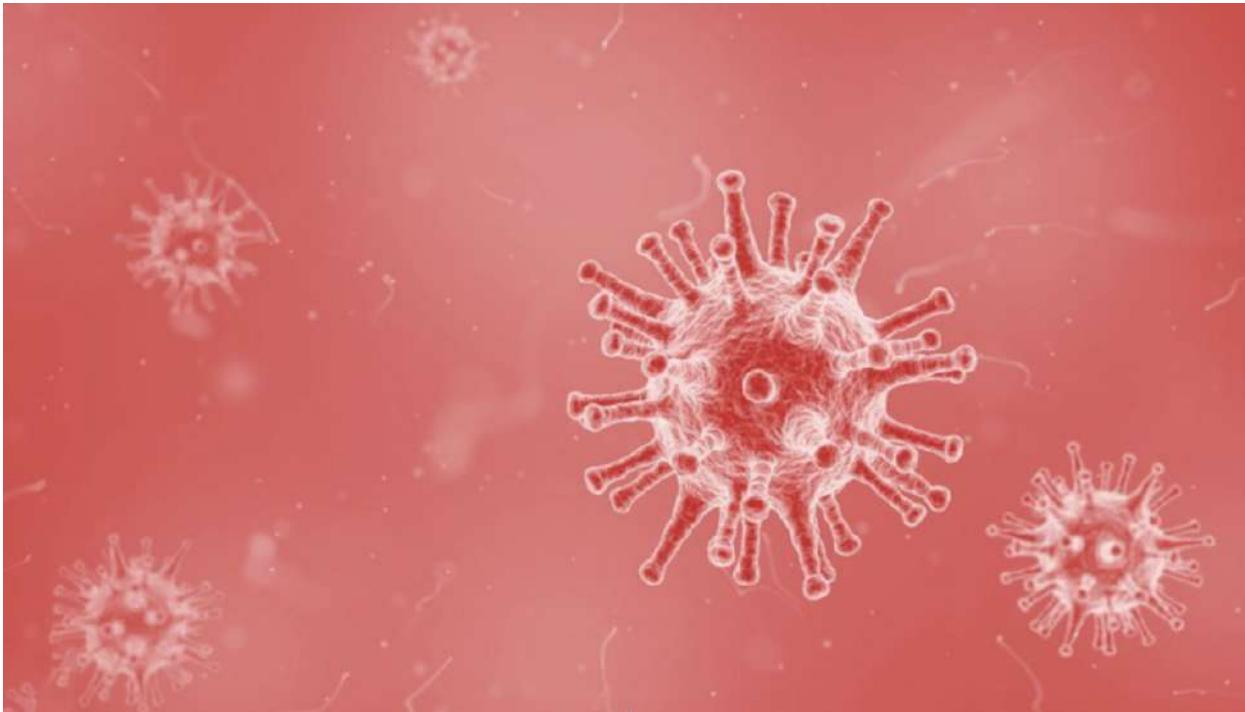
7.7.1 Malware

CORE CONCEPTS

- **Malware** is malicious software that negatively impacts a system.
- **Many types of malware exist.**
- **Virus** = malware that has to be triggered by a user.
- **Worm** = malware that can self-propagate and spread through a network on its own.
- **Logic bomb** = malware that executes based on logic embedded in the code.
- **Trojan horse** = malware that looks harmless but contains malicious code.
- **Polymorphic** = malware that can change multiple aspects of itself to evade detection.
- **Ransomware** = malware that encrypts a system or network of systems and then demands a ransom payment to gain access to the decryption key.
- **Rootkit** = malware that attempts to mask its presence on a system and is a collection of malware tools that can be used by an attacker to meet multiple needs.
- **Zero-day** = malware that has never been seen in the wild before, and therefore no detection or antivirus signatures exist for it.

Understand different types of malware and characteristics of each

Malware is malicious software that negatively impacts a system, and a number of different types of malware exist. It's very common that malware demonstrates multiple different personality traits. For example, a piece of malware can be considered a worm and it can be polymorphic. [Table 7-7](#) describes different types of malware and their traits in more detail.



Types of Malware

Virus	The defining characteristic of a virus is it's a piece of malware that has to be triggered in some way by the user.
Worm	A worm is a piece of malware that is able to self-propagate and spread through a network or a series of systems on its own, by exploiting a vulnerability in those systems. Essentially, a worm can be much more dangerous than a virus.
Companion	Companion malware is a type of malware that attaches itself to legitimate programs and runs simultaneously with them. It creates a different infected file with a similar name, leading the user to execute the infected file unintentionally. For example, creating files with similar names, but different extensions, may cause the execution of the infected file before the intended application..
Macro	A macro is something often found in Microsoft Office products, like Excel, and is created using a very simple programming language to automate tasks. Because a programming language is involved, macros can be programmed to be malicious and harmful, and running macros automatically when opening an Excel or similar file—especially if the file originated elsewhere—is discouraged.

Multipartite	Multipartite means the malware spreads in different ways . Stuxnet is a perfect example of multipartite malware. Stuxnet was first introduced to and infected a system via the system's USB port. Once the system was compromised, it spread harmlessly to other systems until reaching the target systems—Iran's nuclear centrifuges—at which point it sped the centrifuges to the point of failure while indicating to plant workers that everything was working normally.
Polymorphic	This term is best understood by looking at its parts: <i>poly</i> and <i>morphic</i> . <i>Poly</i> means “many”; <i>morphic</i> means “having a form or shape.” Put together, the word <i>polymorphic</i> means “many forms or shapes.” Specifically, every time it replicates across a network, polymorphic malware can change aspects of itself , like file name, file size, code structure, and so on, in order to evade detection .
Trojan	A Trojan horse is malware that looks harmless , or even desirable, but contains malicious code . Trojans are often found in software that is easily and freely downloadable from the internet—audio and video codec files often harbor this type of malware.
Botnet	A botnet is many infected systems that have been harnessed together and act in unison . They're typically managed via some type of command and control structure created by the attacker. Botnets can be used for cryptocurrency mining, DDoS attacks, sending spam, and so on.
Boot sector infector	Boot sector infectors are pieces of malware that are able to install themselves in the boot sector of a hard drive . That allows it to run upon system boot up. Boot sector infectors are extraordinarily difficult to detect, and once installed, they're very difficult to remove. Most operating systems can't read the boot sector by default, so boot sector malware often stays very well hidden, almost as if invisible.
Hoaxes/Pranks	Hoaxes and pranks are not actually software. Rather, they're typically social engineering —via email or other means—that intends harm (hoaxes) or just a good laugh (pranks). For example, a hoax could be something like telling a person they can speed up their system by going to a command prompt and typing DEL *.* In fact, following those instructions could potentially lead to the deletion of everything on a system. A prank could be something like editing the display settings of a system and flipping the screen upside down.
Logic bomb	A logic bomb is a bit of code that, based on some logic, will execute . A classic example of this is a situation that involved a system administrator at a small/medium-size organization. He was the only IT person and had worked for the organization for years. Over time, he grew disgruntled and wrote a little logic bomb and planted it in a system. The logic bomb did one simple thing: it queried the HR database every morning and asked one

	<p>simple question, “Am I still an employee?” One day, when the IT person was no longer an employee, the logic bomb continued to execute and deleted every bit of data from the systems. Think scorched earth—backups, files, email—everything was wiped out. As a result, the organization was seriously impacted. The former IT guy went to jail.</p>
Stealth	<p>Stealth malware is malware that uses various active techniques to avoid detection. Stealth malware will attempt to actively disable the security capabilities of the system (e.g., disable the antivirus software). Once a computer is infected with stealth malware, detection and removal can be a complex process.</p>
Ransomware	<p>Ransomware is gaining in popularity very rapidly. It is a type of malware that typically encrypts a system or a network of systems, effectively locking users out, and then demands a ransom payment (usually in the form of a digital currency, like Bitcoin) to gain access to the decryption key.</p> <p>Many ransomware attacks begin with the attacker gaining access to the target network and exfiltrating a significant amount of data from the organization. Once this is done, system and network encryption takes place. By threatening to release the exfiltrated data to the public and causing the organization reputational and other harm, the attacker increases the likelihood of receiving the ransom payment.</p>
Rootkit	<p>Similar to stealth malware, rootkit malware attempts to mask its presence on a system. As the name implies, a rootkit typically includes a collection of malware tools that an attacker can utilize according to specific goals. For example, one tool in a rootkit might be used to steal passwords, while another tool might be used to plant a backdoor or mask its presence by deleting critical system files and replacing them with fake ones.</p>
Data diddler	<p>A data diddler is a piece of malware that makes very small changes over a long period of time to evade detection. One example of this is known as a salami attack. How do you slice salami? Very thinly. A salami attack is specific to financial systems, with regards to calculating interest and taxes. Oftentimes these numbers are never calculated perfectly, thus allowing a salami attack to shave off fractions of amounts and put them in another account. Over time and with enough transactions, the shaved amounts can add up to quite a significant sum of money.</p>
Zero day	<p>A zero day is any type of malware that's never been seen in the wild before. The vendor of the impacted product is unaware, as are security companies that create antimalware software intended to protect systems. Certainly, customers are completely unaware of a zero day. The only person aware of a zero day is the creator of the malware; they're using it for the first time. Because this is day zero of the malware being “in the wild,” no</p>

detection signatures yet exist, and this fact makes zero day malware potentially very dangerous.

Table 7-7: Malware Types

Third-Party Provided Security Services

As cloud technology and services have become more common, so too have phrases like “third-party provided security services.” Third-party provided security services simply refer to the menu of security-related services that can be contracted. So, in addition to the many security-related practices and controls discussed in this section, SIEM, auditing, penetration testing, antivirus and malware, and forensic services can be provided by third-party providers.

7.7.2 Anti-malware

CORE CONCEPTS

- Anti-malware software designed to prevent malware from being triggered.
- Policy and user training and awareness can help prevent malware outbreaks.
- Signature-based anti-malware software can also help prevent malware outbreaks.
- Heuristic anti-malware software looks at the behavior of code or a file to help identify malware.

What is one of the most effective ways to prevent malware outbreaks?

Anti-malware, as the name suggests, is training or software designed to prevent malware from being triggered. One of the best anti-malware solutions is effective policy and providing user training and awareness to staff members. Considering that a virus requires human interaction—typically the click of a mouse on a link or opening of an attachment—to trigger it, it follows that providing a basic understanding of security and steps to follow can help prevent virus outbreaks.

Understand the difference between signature-based and heuristic malware detection software

More technical methods of detecting malware include using **signature-based antimalware** systems. These systems contain what are known as definition files—files that include signature characteristics of currently known malware—and scan systems using this information to detect suspicious and compromised files. These systems are **unable to detect zero-day malware**, and they're only accurate based upon accurate definition files, so they constantly need to be updated.

Heuristic systems, on the other hand, do not look for malware based on a particular pattern or signature. Rather, they look at the underlying code or behavior of a file. For instance, if a heuristic system identified an executable file as suspicious, it might run the file in what's known as a sandbox, to see how the file behaves. Heuristic systems generally work one of two ways: 1. **Static** code scanning techniques: the scanner scans code in files, similar to white box testing 2. **Dynamic** techniques: the scanner runs executable files in a sandbox to observe their behavior.

Either way, by examining code or running in a sandbox, heuristic systems are attempting to determine what the code does or how the file behaves. If it detects something suspicious, it will block the file. Most malware is designed to know when it's being analyzed, especially in the context of a sandbox, and will behave normally (malicious operation will remain dormant). Then, once on a real system, it behaves maliciously. Also, heuristic systems tend to generate high numbers of false-positives; in other words, perfectly legitimate files and application programs are flagged as suspicious, which slows down productivity of staff members. The major advantage of heuristic scanners is they **can potentially detect zero day malware**.

Activity monitors monitor running systems to see what processes are active. If something suspicious is detected, an alert will be raised. Malware often installs itself on a system and runs in the background. Activity monitors are designed to detect the malware and send an alert.

Change detection (also known as file integrity monitoring), commonly found in Linux-based systems, focuses on modification of key system files. Change detection systems first create hashes of key operating system files and store the hashes in a secure location. Then, as frequently as necessary—every hour, every day, whatever interval makes sense—the tool will rehash the files and compare the new hashes against the stored hashes. If they match, everything is good; if they don't, something might be wrong and an alert is generated. Similar to other types of systems, especially signature-based antimalware systems, change detection systems require continual updating in order to be most effective.

Machine Learning and Artificial Intelligence (AI)-Based Tools

To understand what is meant by machine learning (ML) and artificial intelligence (AI)-based tools, it is important to have an understanding of each term. Of the two, artificial intelligence has been in use longer, and over the years its meaning has changed to reflect advances in technology and the goals trying to be achieved. In today's context, AI development has focused more on the use of human intelligence as a model and not as a goal unto itself.

With this understanding, machine learning (ML), which is often viewed as a subset of artificial intelligence, can be used for purposes of business growth, improving customer selection and

satisfaction, and optimizing processes, logistics, speed of delivery, and quality, among others. In its simplest form, ML employs the strength of artificial intelligence, where a computer is used to learn from data inputs (the past) and make predictions (the future). In reality, ML systems are often networked computers that utilize very powerful processors to run complex algorithms in order to derive meaningful and actionable insights for the sake of future endeavors based upon data inputs generated by past events.

Together, ML and AI-based tools can:

- Empower systems to use data to learn and improve without being explicitly programmed ■ Make predictions through the use of mathematical models to analyze patterns **ML/AI Security Application**

Based upon the above information, in the context of security, ML/AI is specifically being leveraged to provide: ■ Threat detection and classification ■ Network risk scoring ■ Automation of routine security tasks and optimization of human analysis ■ Response to cybercrime: ■ Unauthorized access ■ Evasive malware ■ Spear phishing In addition to the topics and subject matter covered here in [section 7.7](#), the Official CISSP Certification Exam Outline includes other topics in this section that are covered elsewhere in the book. Please refer to Domain 4 for details on the following topics: ■ Firewalls (e.g., next generation, web application, network) ■ Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS) ■ Whitelisting/blacklisting ■ Sandboxing ■ Honeypots/honeynets [7.8 Implement and support patch and vulnerability management](#)

7.8.1 Patch Management

CORE CONCEPTS

- **Patch management helps create a secure environment by fixing—patching—security flaws and vulnerabilities in systems.**
- **Patching only secures a system against known vulnerabilities.**
- **Change management should be part of a patch management program.**
- **Determining patch levels can be done via: agent, agentless, passive methods.**
- **Deploying patches can be done manually or automatically.**

Patch management is a *proactive process to create a consistently configured environment that is secure against known vulnerabilities*. Patches fix security flaws and vulnerabilities in systems. Patches can also improve performance and add functionality.

Understand why timely and consistent application of patches is beneficial and important

Whenever deploying patches, it's important to do so in a manner that leaves the operating environment consistently configured. Patches should be deployed to the entire environment and verified that they were deployed properly and everything is consistently configured. Many patch management systems actually include the capability of knowing when new patches are available and specifically which ones need to be installed. For example, most Windows users are very familiar with the persistent notifications when a patch is available for installation. A number of other systems leave patching up to the system owner and do not indicate that a patch is available or needs to be installed. One other important aspect of patch management is the need for threat intelligence capabilities. Knowledge of new threats is imperative, along with up-to-date system inventories, including patching needs. Threat intelligence can be developed internally, through hardware and software vendor news feeds, email lists, and so on. The point is that patch management should be as proactive as possible to remain as secure as possible.

Once the need for an available patch has been identified, a change management process should be employed as part of the decision to move forward and install the patch. The full patch management life cycle can be seen in [Figure 7-8](#).



Figure 7-8: Patch Management Life Cycle

Determining Patch Levels

Understand the methods used to determine patch levels and ramifications/challenges of each method

As alluded to, it's important to be able to determine patch levels of systems, and several ways to do this exist. One way is agent-based. An agent is a small program installed on a host, and it monitors the host for patch needs. The agent knows what software/patches are installed on the host, and it routinely

compares them to a master database to see if any needs exist. If patches are needed, the agent typically automatically initiates an update process.

Agentless, as the name implies, means no agent is installed on the host system. Rather, monitoring software or a patch scanning system will routinely connect to the host and check patch levels.

Finally, there's what's known as passive detection. With this approach, a look back at vulnerability analysis is required—specifically with how operating system and software versions are detected on a system. This is done through fingerprinting, where activity on a system—network packets particularly—can identify system and software versions. Based upon this approach and data gathered, it can be possible to determine the patch level of systems and then respond accordingly.

Three methods for determining patch levels are summarized in [Table 7-8](#).

Agent	Agentless	Passive
Update software (agent) installed on devices	Remotely connect to each device	Monitor traffic to infer patch levels

Table 7-8: Patch Levels of Systems

Deploying Patches

Understand patch deployment methods and why it might make sense to use a manual method

Once the need is identified, patches can be deployed via manual or automated means. With a manual approach, somebody actually logs into the target system and installs the software. With an automated approach, software is used to roll out the patches. A good example of the latter is Microsoft's Windows Server Update Services (WSUS), which helps maintain and update patches on Windows computers. With regards to high-value, high-priority production systems, automated patching shouldn't be used because patching sometimes breaks things. So, manual patching of these types of systems is often the best approach, as it allows for much better control and response if an issue arises. For rank-and-file systems, automated patching is the best approach and can help an organization maintain a consistently configured environment that was mentioned earlier.

Manual versus automated patching are summarized in [Table 7-9](#).

Manual	Automated
--------	-----------

Somebody logs into the system and installs the software	Patching software automatically rolls out the software updates
---	--

Table 7-9: Patch Deployment Methods

7.9 Understand and participate in change management processes

7.9.1 Change Management

CORE CONCEPTS

- Change management ensures that changes are made in a deliberate manner.
- The change management process includes multiple steps that build upon each other.

Change management is important in the context of security operations. In essence, ***change management ensures that costs and benefits of changes are analyzed and changes are made in a controlled manner to reduce risks.*** The actual change management process is described below.

Change Management Process

Understand the change management process and what happens at each step

The first step is known as a change *request*. A change request can come from any part of an organization and pertain to just about any topic. A business owner might want new functionality. Someone in IT identified a misconfiguration that requires a change. A threat management application identified that a system is vulnerable and needs to be patched. Changes can come from everywhere. When a change request is made, the *impact of the change must be assessed*. How big of a deal is this change?

If a vulnerable system requires a patch, should the change management process take the typical amount of time?

No, this is when emergency change management can be utilized. Related to the severity of the proposed change, the size of the change must be considered. If the change is minor and relates to a low-value system of secondary importance, how many levels of review and approval should be involved?

Common sense would dictate not many levels, but if this is a multimillion-dollar change that's going to affect multiple stakeholders and customers, how many levels of review and approval should now be involved? Probably quite a few.

After a change request has been assessed, *approval* should follow. Approval can take place at multiple stages, but if a significant and potentially costly change is being considered, it should definitely be approved first. Approval also takes place in the context of starting and designing a change as well as prior to implementation. Who approves change can vary significantly, but the owner of the system and potentially other relevant stakeholders should definitely be part of the approval process. This fact explains why **Change Advisory Boards (CAB)** are often utilized as part of the change management process, because they include key stakeholders from throughout an organization.

Once approved, the change should be *built and tested*, and key people should be *notified* before *implementing* the change. Testing might include things like regression testing, where changes can be tested to ensure that everything still works, including the new functionality. Other types of validation testing can also be conducted. Once a change has been built and tested, stakeholders notified, and the change implemented, management and other relevant parties should once again be notified to *validate* the change.

Finally, and perhaps most important, it's critical to update documentation and *versions and baselines*. In fact, documentation should take place throughout the process. The discipline and rigor that a company places upon change management is directly proportional to how well a company operates. In other words, if a company has terrible or nonexistent change management, their environment is most likely a mess, pointing toward disaster. It's undoubtedly reactive, and employees are probably constantly fighting fires. Without enough or proper change management, the environment is likely chaos. Contrarily, with too much change management, every change often requires too much time and too many layers of approval. In this context, people avoid and go around the process, which can lead to chaos as well. Proper change management strikes a balance between too much and too little oversight. All required change management steps have been depicted in [Figure 7-9](#) and described in [Table 7-10](#).



Figure 7-9: Change Management Steps

Change request

A change request can come from any part of an organization and pertain to almost any topic. Organizations typically use some type of change management software that includes a request portal, among other tools that help manage and track the overall process.

Assess Impact	After a change request is made, however small the request might be, the impact of the potential change must be assessed. Additionally, the size of the change should be included in this assessment.
Approval	Based upon the requested change and related impact assessment, common sense plays a big part in the approval process. If the requested change relates to a critical need, perhaps emergency change management protocols can be utilized. If the requested change relates to noncritical, lower-value items, the levels of review and approval should probably be minimal. However, if the change is significantly costly and going to impact multiple stakeholders, the levels of review and approval should likely be high.
Build and Test	After approval, any change should be developed and tested, ideally in a test environment. Testing should be thorough and include all steps necessary to ensure proper functionality of the change as well as viability of existing functionality.
Notification	Prior to implementing any change, key stakeholders should be notified.
Implement	After testing and notification of stakeholders, the change should be implemented.
Validation	Once implemented, senior management and stakeholders should again be notified to validate the change.
Version and Baseline	Documentation should take place at each of the steps noted, but at this point, it's critical to make sure all documentation is complete and to identify the version and baseline related to a given change. This last step is especially helpful, as maintaining discipline in this area can help an organization operate most effectively and efficiently in a proactive manner.

Table 7-10: **Detailed Description of Change Management Steps**

7.10 Implement recovery strategies

7.10.1 Failure Modes

CORE CONCEPTS

- Failure modes refers to what happens in an environment when something—component in a system, an entire system, facility—fails.
- Three failure modes: fail-soft (fail-open), fail-secure (fail-closed), fail-safe.
- Fail-safe prioritizes the safety of people.

Understand the three types of failure modes and what each one prioritizes

Within any environment, consideration must be given to when things fail. Failure refers to components within a system, for example, disk drives or power supplies, as well as entire systems, like a firewall, or facilities, like an office or warehouse. Several failure modes exist, and they are explained in [Table 7-11](#).

Fail-soft (Fail-open)	Fail into a state of less security, for example, a firewall allowing all traffic through if it fails. This could cause significant security problems.
Fail-secure (Fail-closed)	Fail into a state of the same or greater security, for example, a firewall blocking all traffic if it fails. Though this might block legitimate users, it would also maintain security by blocking all attackers.
Fail-safe	Fail into a state that prioritizes the safety of people , for example, the doors to a secure facility unlock automatically in the event of a fire alarm going off.

Table 7-11: Failure Modes



7.10.2 Backup Storage Strategies

CORE CONCEPTS

- **Backup strategies** are driven by organizational goals and objectives and typically focus on backup and restore time as well as storage needs.
- Archive bit is technical detail—metadata—that indicates the status of a backup relative to a given backup strategy; 0 = no changes to file or no backup required, and 1 = file has been modified or backup required.
- **Backup strategies:** incremental, differential, full, mirror
- **Incremental backup:** Changes since last incremental backup
- **Differential backup:** Changes since last *full* backup
- **Full:** All data
- **Mirror backup** = an exact copy of a data set is created, and no compression is used.
- **Backup rotations** refers to different types of tape backup strategies, like first in, first out (FIFO), grandfather-father-son, and Tower of Hanoi, to

name a few. Each strategy dictates when a tape is used for backup, how long a tape is retained, and restoration requirements.

- **Reasons for rotating backups include when a given tape is used, how long a tape is retained, and restoration requirements.**
- **A backup checksum is also known as a cyclic redundancy check or CRC. A CRC ensures the integrity of data through a bit of math and can be used anywhere data resides.**

Understand different backup strategies and how a backup strategy fits into an overall recovery strategy

Recovery strategies focus on several key things, including bringing systems back online quickly, minimizing data loss, and architecting systems in a manner that precludes system downtime if a system failure occurs.

An important component of recovery strategies is backup strategies, which focus on backing up data in such a manner that if the primary data store is lost, corrupted, stolen, or otherwise compromised, the data can be recovered and restored. Before diving in a bit deeper on this topic, it's important to understand a technical detail known as an *archive bit*. Within virtually every operating and associated file system, each file will have an associated bit. The bit is simply metadata, which is data about data.

Archive Bit

An archive bit, as depicted in [Figure 7-10](#), can be set to zero (0) or one (1), and the setting indicates if the file needs to be backed up or not. If the archive bit is set to 0, the file does not need to be backed up; if the archive bit is set to 1, the file needs to be backed up. When a backup is performed, the archive bit is usually set to 0. After the backup, if any of those files are modified or if new files are created, the archive bit is set to 1, indicating that the file needs to be backed up.

0

No changes to file
No backup required

1

File has been modified
Backup required

Figure 7-10: Archive Bit

This is important to understand because different backup strategies deal with the archive bit differently. Incremental and differential backup strategies do not treat the archive bit in the same manner.



Figure 7-11: Incremental vs. Differential

Figure 7-11, each diagram shows the days of the week on the horizontal line, because backup strategies are typically based on a seven-day calendar week. The varying height bars above each day represent the amount of data being backed up during each backup cycle.

Incremental Backup

An Incremental backup starts by performing a full backup, represented by the bar above Sunday. A **full backup** means that every file, regardless of the archive bit setting, is backed up, and the archive bit on each file is then set to 0. On Monday, the typical beginning of a work week, people show up and start creating or modifying files. The archive bit associated with every new or modified file is set to 1. On Monday evening, when the incremental backup is run, all of the files with an archive bit set to 1 are backed up, and the archive bit is reset to 0. On Tuesday, people once again come to work and create or

modify files, and all of the files that have been created or changed since Monday are backed up during Tuesday's incremental backup. As the diagram depicts, incremental backups typically lead to a small amount of data being backed up each time. **Only the changes—new or modified files—since the last backup are backed up each evening.**

Differential Backup

A differential starts the same way, by performing a full backup, and the archive bit on each file is set to 0. On Monday, as before, people show up to work and create or modify files, and the archive bit on each new or modified file is set to 1. Here's where the difference exists: On Monday evening, when the differential backup is performed, the archive bit of backed up files is not reset to 0. All of the files whose archive bit is set to 1 are backed up, and the archive bit on each file remains set at 1. On Tuesday, people again create and modify files, and the archive bit on those files is set 1. Now, during Tuesday's differential backup, all of the new and changed files from Monday and Tuesday are backed up, and again, the archive bit on every backed up file stays set to 1. On Wednesday, Thursday, Friday, and Saturday, the same thing happens. As the graph depicts, differential backups result in more and more files being backed up each evening. In other words, **all of the changes since the last full backup** (on Sunday) are backed up each evening.

As one might imagine, each backup strategy comes with pros and cons.

One advantage of incremental backups is they're typically very fast, especially toward the end of the week when much less data is being backed up. However, a big disadvantage of incremental backups relates to recovery. Imagine a system has failed on a Saturday night, and we need to restore all of the data. First, Sunday's full backup would need to be restored, and then each incremental backup prior to Saturday (or including Saturday, if Saturday's backup took place prior to the failure) would need to be restored. This obviously represents quite a few backup tapes as well as time to recover.

Contrary to a longer recovery time with incremental backups, recovery with differential is much quicker. At most, only two backup tapes would ever need to be restored—the full backup and the most recent differential. This is a significant advantage relative to incremental backup. However, differential backup takes more time to perform, especially at the end of the week, and much more data storage is used.

Which strategy should be utilized? As with many decisions, the goals and objectives of the organization should drive this decision. What's the business trying to achieve? What are they willing to pay for? Similar questions will ultimately help drive the response to this question.

Mirror Backup

Another type of backup is known as a **mirror backup**, where **an exact copy of a data set is created, and no compression is used**. Mirror backups can often be created and restored quickly, but at the cost of significant amounts of data storage. Of the three types of backups mentioned, mirror is the fastest to backup and restore, but it requires a tremendous amount of data storage; incremental backups require the least amount of data storage. [Table 7-12](#) contains a comparison of the prementioned back strategies.

Type	Data Backed Up	Backup Time	Restore Time	Storage
Mirror	Exact copy with no compression	Fastest	Fastest	High
Full	ALL data	Slowest	Fast	High
Differential	Full and then <i>changes since last full backup</i>	Moderate	Moderate	Moderate
Incremental	Full and then <i>changes since last backup</i>	Fast	Slow	Lowest

Table 7-12: Summary of Backup Strategies

Backup Storage Strategies (e.g., cloud storage, onsite, offsite)

Understand backup storage strategies, including reasons for tape rotation

Questions are often raised about which locations are optimal for backup storage. **Onsite backups involve storing backup data in the same place as the original data.** They are local backups. **Offsite backups involve storing the data in a separate location from the original data.** A good choice would be to store the backup media at a geographically remote location. Geographically remote means the backup location is far enough away from the primary location so a natural disaster, political uprising, or other significant event at the primary location won't impact the backup location. **Cloud backups involve storing data in the cloud.** Not only is cloud storage in a separate location, but it provides high availability, often at a relatively low cost. Other storage strategies are outlined below:

- **Electronic vaulting:** This term usually implies some type of automated tape management system, like a tape jukebox. These systems contain numerous backup tapes, which are managed automatically by a robotic arm and backup scheduling software.

■ **Tape rotation:** This refers to different types of tape backup strategies. Each strategy dictates when a tape is used for backup, how long a tape is retained, and restoration requirements.

Cyclic Redundancy Check (CRC)

Understand the purpose behind the use of checksums/CRC

In the context of backups and moving data around a network, a term known as cyclic redundancy check, or CRC, is often used. A CRC is a bit of math that is used to ensure the integrity of data, and CRC can be used anywhere data resides—hard drive, when moving data across a network, and even in RAM. Among other uses, a CRC can be used to validate backups in order to ensure the integrity of the backed up data. *A CRC is a method of detecting accidental changes to data at rest or in motion.*

7.10.3 Spare Parts

CORE CONCEPTS

- Spare parts strategies include: cold spare, warm spare, hot spare.
- Which strategy is used is often dependent upon organizational goals and objectives.
- Key considerations include: cost, criticality.

Oftentimes people take for granted that systems in a data center “just work” all the time. But what happens when those systems stop working due to a part failing? Business stops. So, with any critical systems, it’s important to have spares of critical components for those systems on hand. Critical components include things like power supplies, cooling fans, hard drives, and so on. Then, if the primary part in a system fails, a spare part can be installed and up and running very quickly. Three types of spare parts exist, and the name points to the location. [Table 7-13](#) contains a summary of all spare types.

Cold Spare	Warm Spare	Hot Spare
<ul style="list-style-type: none">■ Parts sitting on a shelf, perhaps in a storage room.■ If the primary part fails, the system will go offline until the needed spare part can be retrieved and installed.	<ul style="list-style-type: none">■ Installed in a computer system, but not powered and available.■ If the primary part fails, the system will go offline, but getting it back online quickly is often a simple matter of switching over to the spare part.	<ul style="list-style-type: none">■ Installed in a system, powered, and available.■ If the primary part fails, the spare part instantly takes over the primary part’s function.■ Hot spares are expensive, as are the systems within

■ Cold spares result in downtime, but they're also the least expensive.	■ Warm spares lead to less downtime, but this benefit costs a bit more.	which they're used, but they also provide significant benefits to an organization.
---	---	--

Table 7-13: Spare Types

7.10.4 Redundant Array of Independent Disks (RAID)

CORE CONCEPTS

- **Redundant array of independent disks (RAID)** refers to multiple drives being used in unison in a system to achieve greater speed or availability.
- **RAID 0**—also known as striping—provides significant speed data writing and reading advantages.
- **RAID 1**—also known as mirroring—utilizes redundancy to provide reliable availability of data.
- **RAID 10**—mirroring and striping—requires a minimum of four hard drives and provides benefits of striping (speed) and mirroring (availability) in one solution; this type of RAID is typically one of the most expensive.
- **RAID 5**—parity protection—requires a minimum of three hard drives and provides a cost-effective balance between RAID 0 and RAID 1; RAID 5 utilizes a parity bit, computed from an XOR operation, for purposes of storing and restoring data.

Understand the primary types of RAID and pros and cons of each

Redundant array of independent discs, better known as RAID, is the concept that instead of one hard drive being utilized in a system, multiple drives are used in unison to achieve greater speed or availability. In this context, hard drives can be grouped in a number of different ways—called RAID levels—to achieve these goals. Three of the best-known RAID levels are RAID 0, RAID 1, and RAID 5. RAID 0 is also known as striping, RAID 1 as mirroring, and RAID 5 as parity protection. We are going to focus on the most well-known RAID levels.

RAID 0—Striping

RAID 0 (shown in [Figure 7-12](#)), also known as striping, works as follows. Imagine a file is sent to the RAID controller on a system. The RAID controller is simply a piece of hardware (sometimes it's software) that manages the storage of data on connected hard drives. Depending upon the type of RAID

configured, the controller will make the appropriate decision about how to best store the file. With RAID 0, the file is split into two pieces, and one piece is saved on one hard drive and the other on the other hard drive. RAID 0—striping—is all about speed of writing and reading data, because of the power of the RAID controller. If the file is split and written to two drives at the same time, write speed is effectively doubled. Similarly, when reading the file back from two drives at the same time, read speed is also doubled. RAID 0 is all about speed.

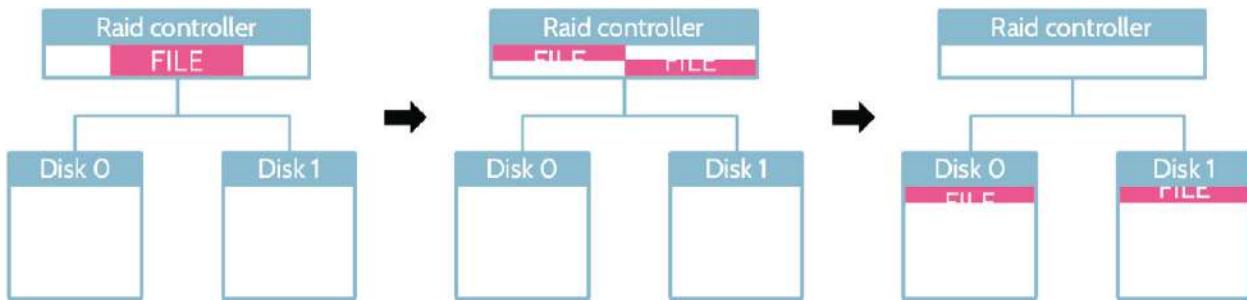


Figure 7-12: RAID 0

RAID 1—Mirroring

RAID 1 (shown in [Figure 7-13](#)), also known as mirroring, is set up exactly the same way as RAID 0. However, in this case, instead of splitting a file and putting a portion of each on separate drives, the file is written to each drive. In other words, the same file now resides on two different drives; thus the term mirroring. If one drive fails, the file can easily be recovered because it was written to two locations. RAID 1 is all about availability through redundancy.

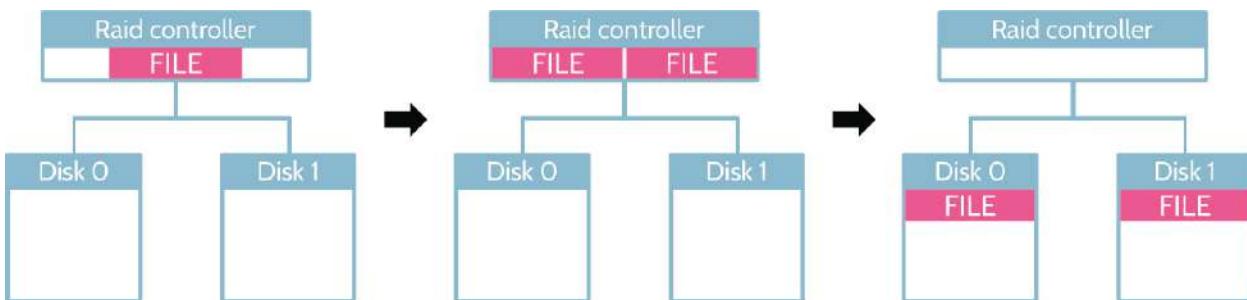


Figure 7-13: RAID 1

RAID 10—Mirroring and Striping

RAID 10 or RAID 1+0 or 0+1 (shown in [Figure 7-14](#)) is, in a sense, the best of both RAID 0 and RAID 1. With each of the latter approaches, a minimum of two hard drives is required. With RAID 10, a minimum of four drives is required, because you’re treating data the way it is treated using RAID 0 and RAID 1. When a file is sent to the RAID controller, two copies are made, and then each copy is split, resulting in four chunks of data. RAID 10 offers the advantage of significant speed and redundancy, but these advantages come with the associated need for and cost of more hard drives.

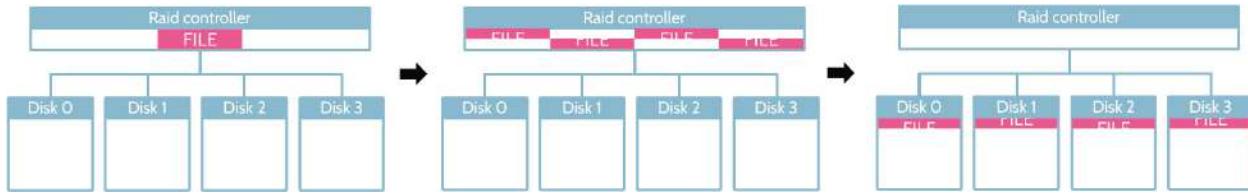


Figure 7-14: RAID 10

RAID 5—Parity Protection

RAID 5 (shown in Figure 7-15) attempts to strike a balance between RAID 0 and RAID 1 and provide availability in an efficient and cost-effective manner. Unlike RAID 0 and 1, where a minimum of two hard drives is required to work, RAID 5 requires a minimum of three hard drives. As before, a file is sent to the RAID controller, and the file is split into two parts, and one more piece of data—the **parity data**—is computed, using a binary mathematics operation known as XOR, or *exclusive or*. With the parity data in place, if either part of the file was lost, it could be reconstructed using the remaining part and the data contained in the parity data. So three pieces of data exist—the two file chunks and the parity data—and each is written to a separate hard drive. Additionally, to further provide protection against loss, the parity data is typically not written to only one drive. A round robin methodology is employed that varies the location of file chunks and parity data. RAID 5 offers very quick read/write speeds as well as redundancy.

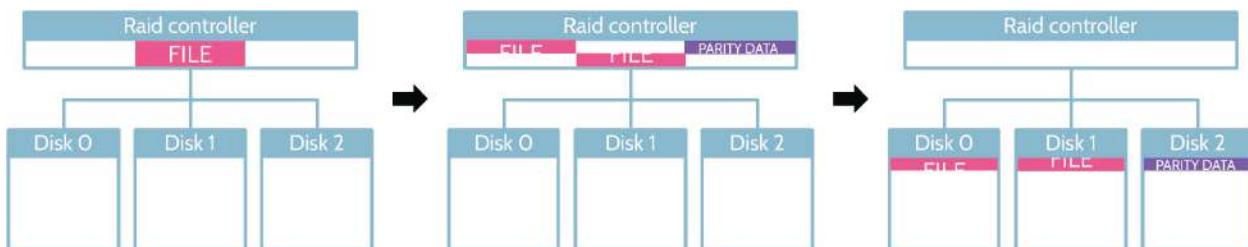


Figure 7-15: RAID 5

A comparison between the different RAID types is provided in Table 7-14.

RAID	Data Redundancy	Read/Write Performance	Min. # of Drives	
0 Striping	✗	Highest	2	<ul style="list-style-type: none"> ■ All about increasing read/write performance ■ No data resiliency
1	✓	Moderate	2	<ul style="list-style-type: none"> ■ Great data resiliency ■ No

Mirroring				performance improvement
1+0 Striping + Mirroring	✓	Highest	4	<ul style="list-style-type: none"> ■ Performance and resiliency ■ But requires a lot of hard drives
5 Parity Protection	✓	High	3	<ul style="list-style-type: none"> ■ Good balance of performance and resiliency and requires less hard drives than RAID 10

Table 7-14: RAID Type Comparison

7.10.5 Clustering and Redundancy

CORE CONCEPTS

- **Clustering** = a group of systems working together to handle a load.
- **Redundancy** = typically a primary system and secondary system, with the secondary system in standby mode and ready to take over if something goes wrong with the primary system.
- **Clustering and redundancy both include high availability as a by-product of their configuration.**

Understand the difference between clustering and redundancy and a primary by-product of each approach

Clustering and redundancy are two other important terms pertaining to recovery and protection of systems and data. Both concepts point to high availability (HA) of systems.

Clustering refers to a group of systems working together to handle a load. This is often seen in the context of web servers that support a website. Typically, incoming traffic will be managed by a load balancer that distributes requests to multiple web servers, the cluster. With clustering, if one system goes down, the amount of overall performance for the cluster drops by an equivalent amount.

Redundancy also involves a group of systems, but unlike a cluster, where all the members work together, redundancy typically involves a primary system and a secondary system. The primary system does all the work, while the secondary system is in standby mode. If the primary system fails, activity can fail over to the secondary. One or more secondary systems can exist, but there is always only one primary. With redundancy, if the primary system goes down, there is no loss in performance, because the secondary system takes over, and typically any secondary system will be configured exactly the same as the primary system.

Clustering and redundancy both include high availability (HA) as a by-product, which can help ensure ongoing operations in the face of planned/unplanned system outages, failure of components, or other disruptions to operations. Clustering is about systems working together, and redundancy is about one primary and one or more secondary systems.

Clustering and redundancy are summarized in [Table 7-15](#).

Clustering	Redundancy
A cluster of multiple systems work together to support a workload	A single primary system supporting the entire workload and one or more secondary systems that will take over if the primary one fails

Table 7-15: Clustering and Redundancy

7.10.6 Recovery Site Strategies

CORE CONCEPTS

- Recovery site strategies consider multiple elements of an organization—people, data, infrastructure, and cost, to name a few examples—as well as factors like availability and location.
- Geographically remote and geographic disparity refer to where a recovery site is located relative to the primary site.
- Internal recovery sites are owned by the organization; external recovery sites are owned by a service provider.
- Multiple processing sites are more than one site where key business functionality is performed.
- System resilience, high availability, quality of service (QOS), and fault tolerance are achieved using recovery site strategies as well as things like clustering, redundancy, replication, spare parts, and RAID.

Recovery to this point has been more focused on systems or components of systems. Recovery of sites is equally important. For example, what happens if an entire data center goes down?

Understand the time required to bring each type of recovery site online

Different recovery strategies exist, and elements like time to recover and money are important components of each. Something called a cold site is relatively inexpensive, but it takes a relatively long time to bring online. A mirrored or redundant site, on the other hand, can be brought online very quickly, but this type of site is also extremely expensive. [Figure 7-16](#) illustrates this fact, and each type of strategy will be discussed in a bit more detail below.

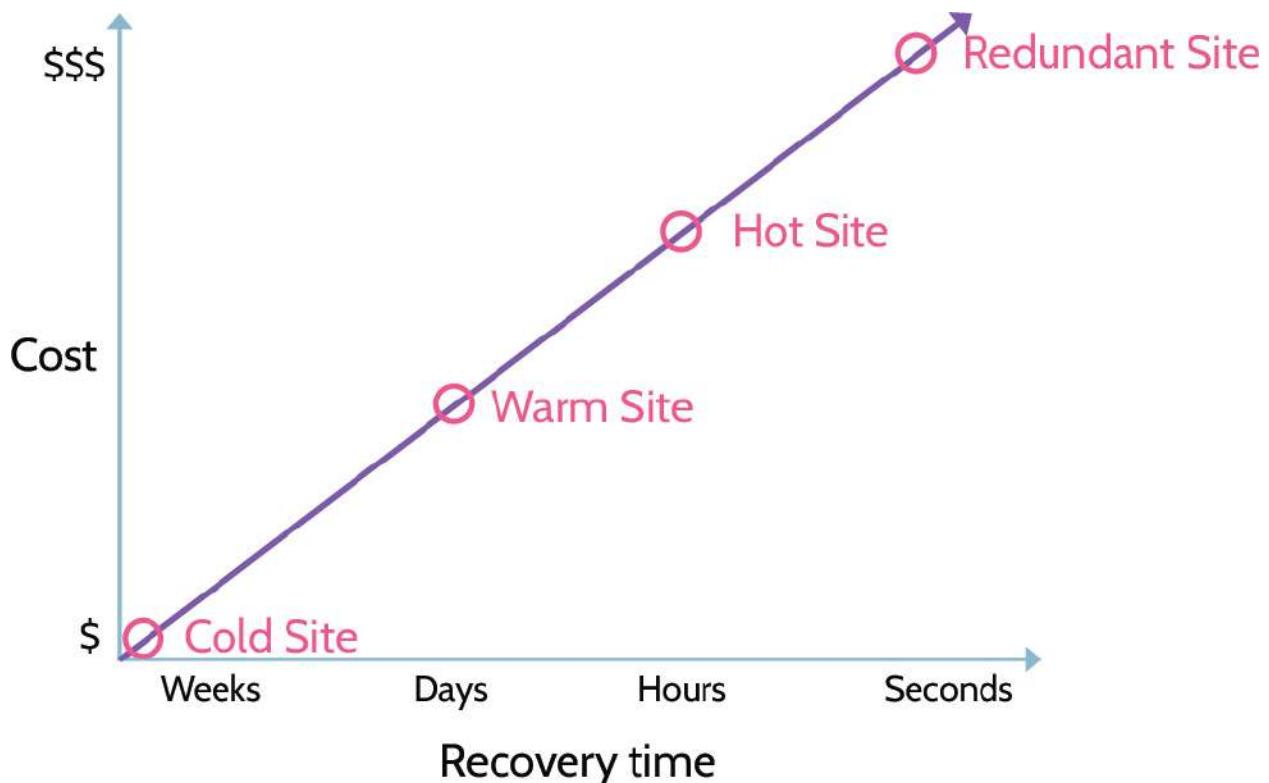


Figure 7-16: Recovery Site Types

Before looking more closely at different types of recovery sites, however, it's important to examine components related to each strategy.

Infrastructure and HVAC refer to the shell of a building and the heating, ventilation, and cooling equipment—pretty much the basic items that make up any building. It may also include equipment racks and basic cabling, but no computers or networking hardware. Racks and basic cabling are cheap; computer and networking hardware can be very expensive, potentially costing millions of dollars. Of course, data is the component that runs on the systems, and people are the component that run and manage the data center on a daily basis. As can be seen, a bare shell is much less expensive than a site that contains everything to run an organization. Recovery time, as noted above, can vary and is essentially the length of time to get back up and running.

The cheapest recovery option—**cold site**—is the least expensive for a reason. With this strategy, you essentially get a shell of a building that needs to be populated with equipment, data, and people. It usually takes weeks to bring a cold site online. The biggest advantage of a cold site is cost—it's cheap.

A **warm site** is better than a cold site, because in addition to the shell of a building, basic equipment is installed—racks are in place, cables are run, and so on. Servers, network, and other equipment as well as data and people are the missing components. Unlike a cold site, a warm site can be brought online in a matter of days.

Hot sites, compared to cold and warm sites, are significantly more expensive, because everything is ready to go except for the data and people. Servers, networking equipment, and so on are all in place, only waiting for data to be restored and people to operate the site. Hot sites can usually be brought online in a matter of hours.

A form of hot site is known as a **mobile site**. A mobile site is a hot site on wheels, and many companies use them in anticipation of emergency or severe business disruption. Essentially, a mini data center is built inside something like a shipping container, and then if something happens, the shipping container can be loaded onto a truck and moved where needed. The federal government uses these when hurricanes or other natural disasters strike. Similar to a stationary hot site, a mobile site can also be brought online quickly—in days, or even hours—with the time to move to location often being the biggest determining factor.

Finally, a **redundant site** is extremely expensive, because the basic infrastructure and equipment, expensive equipment, data, and people are in place and ready to go. A redundant site can be architected in such a manner that the primary site can automatically fail over to the redundant site if required. Thus, a redundant site can be online and running instantaneously or in a matter of seconds. Of course, this benefit comes at very high cost, and it is usually as much as the cost of the primary site.

The recovery sites are summarized in [Table 7-16](#).

	Cold	Warm	Hot	Mobile	Redundant
People					✓
Data					✓
Computer Hardware			✓	✓	✓
Basic Equipment		✓	✓	✓	✓
Infrastructure/HVAC	✓	✓	✓	✓	✓

Cost	\$	\$\$	\$\$\$	\$\$\$\$	\$\$\$\$\$
Recovery	Weeks	Days	Hours	Days/Hours	Instant/Seconds

Table 7-16: Recovery Site Type Comparison

Understand the importance of geographic disparity to recovery site strategies



Geographically Remote/Geographic Disparity

Another topic of relevance is the notion of a recovery site being geographically remote. In other words, if an organization's primary site is on the East Coast of North America, perhaps the recovery site should be somewhere in the Midwest or West Coast of North America. The idea is that you don't want whatever may have brought a primary site down to also impact the recovery site.

Internal versus External Recovery Sites

The words internal and external refer to who owns the recovery site. Internal sites are owned by the organization; external sites are owned by a third party. A good example of an external site provider is

Sungard. Companies like Sungard offer organizations a wide portfolio of colocation, data, and recovery centers around the globe. However, as the use of cloud services has continued to expand, many organizations use it as a significant part of their backup and recovery strategy.

Internal versus external recovery sites are summarized in [Table 7-17](#).

Internal Recovery Site	External Recovery Site
Owned and operated by the organization	Provided by a service provider

Table 7-17: Internal vs. External Recovery Sites

Reciprocal Agreements

Another important concept is known as a reciprocal agreement, where two companies agree to support each other if either company suffers an outage. Each company essentially says to the other, “If you experience downtime, you can recover your key systems in our data center.” In reality, reciprocal agreements are quite rare, especially in the context of private enterprise.

Resource Capacity Agreements

Resource capacity agreements are agreements that organizations make with vendors to ensure that they can secure the resources they need during a disaster. These are critical for business continuity.

Multiple Processing Sites

Multiple, or multiprocessing, sites are more than one site where key business functionality is performed. A great example of where this is employed is in the credit card processing industry, where transactions are processed simultaneously at multiple sites located in different parts of the country. A primary record of the transaction is maintained, and the secondary records are there in case something happens at that site where the primary record is located. The premise of multiprocessing sites is geographically dispersed redundant processing, and this functionality is architected from the beginning, which makes it quite expensive but also very effective and reliable.

Disaster Recovery Solutions Summary

With the recovery site landscape described above, how does an organization decide which type makes the most sense? When considering this question, every organization should consider a couple of variables to help determine the answer. The two variables—RPO and RTO (shown in [Figure 7-17](#))—point to data and time as relates to recovering from a disaster.

RPO stands for recovery point objective, and it refers to how much data an organization could afford to lose. The answer to this question specifically drives data backup and recovery strategies.