Chapter 9
Implementing Controls to Protect Assets

CompTIA Security+ objectives covered in this chapter:

1.2 Compare and contrast types of attacks.

- Social engineering (Tailgating)

1.6 Explain the impact associated with types of vulnerabilities.

- Vulnerable business processes, System sprawl/undocumented assets, Architecture/ design weaknesses

2.1 Install and configure network components, both hardware- and software-based, to support organizational security.

- Load balancer (Scheduling [Affinity, Round-robin], Active-passive, Active-active, Virtual IPs)

2.2 Given a scenario, use appropriate software tools to assess the security posture of an organization.

- Backup utilities

2.3 Given a scenario, troubleshoot common security issues.

- Asset management

3.1 Explain use cases and purpose for frameworks, best practices and secure configuration guides.

- Defense-in-depth/layered security (Vendor diversity, Control diversity [Administrative, Technical], User training)

3.8 Explain how resiliency and automation strategies reduce risk.

- Distributive allocation, Redundancy, Fault tolerance, High availability, RAID

3.9 Explain the importance of physical security controls.

- Lighting, Signs, Fencing/gate/cage, Security guards, Alarms, Safe, Secure cabinets/ enclosures, Protected distribution/Protected cabling, Airgap, Mantrap, Faraday cage, Lock types, Biometrics,

Barricades/bollards, Tokens/cards, Environmental controls (HVAC, Hot and cold aisles, Fire suppression), Cable locks, Cameras, Motion detection, Logs, Infrared detection, Key management

4.3 Given a scenario, implement identity and access management controls.

- Physical access control (Proximity cards, Smart cards)

5.2 Summarize business impact analysis concepts.

- RTO/RPO, MTBF, MTTR, Mission-essential functions, Identification of critical systems, Single point of failure, Impact (Life, Property, Safety, Finance, Reputation), Privacy impact assessment, Privacy threshold assessment

5.6 Explain disaster recovery and continuity of operation concepts.

- Recovery sites (Hot site, Warm site, Cold site), Order of restoration, Backup concepts (Differential, Incremental, Snapshots, Full), Geographic considerations (Off-site backups, Distance, Location selection, Legal implications, Data sovereignty), Continuity of operation planning (Exercises/tabletop, After-action reports, Failover, Alternate processing sites, Alternate business practices)

**

You can't eliminate risk to an organization's assets. However, you can reduce the impact of many threats by implementing security controls. It's common to implement several controls using a defense-in-depth strategy. Physical security controls help protect access to secure areas. Redundancy and fault-tolerance strategies help eliminate single points of failure for critical systems. Backups ensure that data remains available even after data is lost. More in-depth business continuity strategies help ensure mission-critical functions continue to operate even if a disaster destroys a primary business location. This chapter covers these concepts.

Implementing Defense in Depth

Defense in depth (also known as layered security) refers to the security practice of implementing several layers of protection. You can't simply take a single action, such as implementing a firewall or installing antivirus software, and consider yourself protected. You must implement security at several different layers. This way, if one layer fails, you still have additional layers to

protect you.

If you drive your car to a local Walmart, put a five-dollar bill on the dash, and leave the keys in the car and the car running, there is a very good chance the car won't be there when you come out of the store. On the other hand, if you ensure nothing of value is visible from the windows, the car is locked, it has an alarm system, and it has stickers on the windows advertising the alarm system, it's less likely that someone will steal it. Not impossible, but less likely.

You've probably heard this as "there is no silver bullet." If you want to kill a werewolf, you can load your gun with a single silver bullet and it will find its mark. The truth is that there is no such thing as a silver bullet. (Of course, there is no such thing as a werewolf either.)

Applied to computers, it's important to implement security at every step, every phase, and every layer. Information technology (IT) professionals can never rest on their laurels with the thought they have done enough and no longer need to worry about security.

Control diversity is the use of different security control types, such as technical controls, administrative controls, and physical controls. For example, technical security controls such as firewalls, intrusion detection systems (IDSs), and proxy servers help protect a network. Physical security controls can provide extra protection for the server room or other areas where these

devices are located. Administrative controls such as vulnerability assessments and penetration tests can help verify that these controls are working as expected.

Vendor diversity is the practice of implementing security controls from different vendors to increase security. As an example, Chapter 3, "Exploring Network Technologies and Tools," describes a demilitarized zone (DMZ). Many DMZs use two firewalls and vendor diversity dictates the use of firewalls from different vendors. For example, one firewall could be a Cisco firewall and the other one could be a Check Point firewall. If a vulnerability is discovered in one of these firewalls, an attacker might be able to exploit it. However, it's unlikely that both firewalls would develop a vulnerability at the same time.

User training also helps provide defense in depth. If users engage in risky behaviors, such as downloading and installing files from unknown sources or responding to phishing emails, they can give attackers a path into an organization's network. However, providing regular training to users on common threats, and emerging threats, helps them avoid these types of attacks.

Remember this

Layered security, or defense-in-depth practices, uses control diversity, implementing administrative, technical, and physical security controls. Vendor diversity utilizes controls from different vendors. User training informs users of threats, helping them avoid common attacks.

Comparing Physical Security Controls

A physical security control is something you can physically touch, such as a hardware lock, a fence, an identification badge, and a security camera. Physical security access controls attempt to control entry and exits, and organizations commonly implement different controls at different boundaries, such as the following:

- **Perimeter.** Military bases and many other organizations erect a fence around the entire perimeter of their land. They often post security guards at gates to control access. In some cases, organizations install barricades to block vehicles.
- **Buildings.** Buildings commonly have additional controls for both safety and security. For example, guards and locked doors restrict entry so only authorized personnel enter. Many buildings include lighting and video cameras to monitor the entrances and exits.
- **Secure work areas.** Some companies restrict access to specific work areas when employees perform classified or restricted access tasks. In some cases, an organization restricts access to all internal work areas. In other words, visitors can enter the lobby of a building, but they are not able to enter internal work areas without an escort.
- **Server and network rooms.** Servers and network devices such as routers and switches are normally stored in areas where only the appropriate IT personnel can access them. These spaces may be designated as server rooms or wiring closets. It's common for an organization to provide additional physical security for these rooms to prevent attackers from accessing the equipment. For example, locking a wiring closet prevents an attacker from installing illicit monitoring hardware, such as a protocol analyzer, to capture network traffic.
- **Hardware.** Additional physical security controls protect individual systems. For example, server rooms often have locking cabinets to protect servers and other equipment installed in the equipment bays. Cable locks protect laptop computers, and smaller devices can be stored in safes.
- **Airgap.** An ***airgap*** is a physical security control that ensures that a computer or network is physically isolated from another computer or network. As an example, you can isolate a computer from a network by ensuring that it is not connected to any other system in the network. This lack of connectivity provides an airgap. This is often done to separate classified networks from unclassified networks.

Using Signs

A simple physical security control is a sign. For example, an

“Authorized Personnel Only” sign will deter many people from entering a restricted area. Similarly, “No Trespassing” signs let people know they shouldn’t enter. Of course, these signs won’t deter everyone, so an organization typically uses additional physical security measures.

Comparing Door Lock Types

It’s common to secure access to controlled areas of a building with door locks, and there are many different lock types. A door access system is one that only opens after some access control mechanism is used. Some common door access systems are cipher locks, proximity cards, and biometrics.

When implementing door access systems, it’s important to limit the number of entry and exit points. As an example, if a data center has only one entrance and exit, it is much easier to monitor this single access point. You can control it with door locks, video surveillance, and guards. On the other hand, if the data center has two entry/exit points, you need another set of controls to control access in both places.

Another important consideration with door access systems is related to personnel safety and fire. In the event of a fire, door access systems should allow personnel to exit the building without any form of authentication.

Remember this

In the event of a fire, door access systems should allow personnel to exit the building without any form of authentication. Access points to data centers and server rooms should be limited to a single entrance and exit whenever possible.

Securing Door Access with Cipher Locks

Cipher locks often have four or five buttons labeled with numbers. Employees press the numbers in a certain order to unlock the door. For example, the cipher code could be 1, 3, 2, 4. Users enter the code in the correct order to gain access. Cipher locks can be electronic or manual. An electronic cipher lock automatically unlocks the door after you enter the correct code into the keypad. A manual cipher lock requires the user to turn a handle after entering the code.

To add complexity and reduce brute force attacks, many manual cipher

locks include a code that requires two numbers entered at the same time. Instead of just 1, 3, 2, 4, the code could be 1/3 (entered at the same time), then 2, 4, 5.

One challenge with cipher locks is that they don't identify the users. Further, uneducated users can give out the cipher code to unauthorized individuals without understanding the risks. Shoulder surfers might attempt to discover the code by watching users as they enter. Security awareness training can help reduce these risks.

Securing Door Access with Cards

It's also possible to secure access to areas with proximity cards or smart cards. ***Proximity cards*** are small credit card-sized cards that activate when they are in close proximity to a card reader. Many organizations use these for access points, such as the entry to a building or the entry to a controlled area within a building. The door uses an electronic lock that only unlocks when the user passes the proximity card in front of a card reader.

Similarly, it's possible to use smart cards or physical tokens (described in Chapter 2, "Understanding Identity and Access Management") for door access. In some scenarios, the smart cards include proximity card electronics. In other scenarios, users must insert the smart card into a smart card reader to gain access.

You've probably seen proximity card readers implemented with credit card readers. Many self-serve gasoline stations and fast-food restaurants use them. Instead of swiping your credit card through a magnetic reader, you simply pass it in front of the reader (in close proximity to the reader), and the reader extracts your credit card's information.

These are becoming popular elsewhere, too. For example, if you stay at a Walt Disney World property, they can issue you a bracelet that includes the functionality of a proximity card. To enter your hotel room, you wave your bracelet in front of the door. If you want to buy food or souvenirs or pay for almost anything, you can simply wave your bracelet in front of a card reader to complete your purchase.

The card (and bracelet) doesn't require its own power source. Instead, the electronics in the card include a capacitor and a coil that can accept a charge from the proximity card reader. When you pass the card close to the reader, the reader excites the coil and stores a charge in the capacitor. Once charged, the card transmits the information to the reader using a radio

frequency. When used with door access systems, the proximity card can send just a simple signal to unlock the door. Some systems include details on the user and record when the user enters or exits the area. When used this way, it's common to combine the proximity card reader with a keypad requiring the user to enter a personal identification number (PIN). This identifies and authenticates the user with multifactor authentication. The user has something (the proximity card) and knows something (a PIN).

Many organizations use proximity cards with turnstiles to provide access for a single person at a time. These are the same type of turnstiles used as entry gates in subways, stadiums, and amusement parks.

Remember this

Proximity cards are credit card-sized access cards. Users pass the card near a proximity card reader and the card reader then reads data on the card. Some access control points use proximity cards with PINs for authentication.

Securing Door Access with Biometrics

It's also possible to use biometric methods as an access control system. One of the benefits is that some biometric methods provide both identification and authentication. When connected to a back-end database, these systems can easily record the activity, such as who entered the area and when.

For example, you can install a retina scanner at the entrance to a secure server room. When individuals want to enter, the biometric scanner identifies and authenticates them. It's important to ensure you use an accurate biometric system and configure it to use a low false acceptance rate, as described in Chapter 2. Otherwise, it might falsely identify unauthorized individuals and grant them access.

Remember this

Door access systems include cipher locks, proximity cards, and biometrics. Cipher locks do not identify users. Proximity cards can identify and authenticate users when combined with a PIN. Biometrics can also identify and authenticate users.

Tailgating

Chapter 6, "Comparing Threats, Vulnerabilities, and Common Attacks," discusses several types of social engineering attacks and tailgating is another one. ***Tailgating*** (also called piggybacking) occurs when one user follows closely behind another user without using credentials. For example, if Lisa opens a door with her proximity card and Bart follows closely behind her without using a proximity card, Bart is tailgating. If authorized users routinely do this, it indicates the environment is susceptible to a social engineering attack where an unauthorized user follows closely behind an authorized user.

As an example, an organization hired a security company to perform a vulnerability assessment. The company sent one of its top security professionals (who happened to be an attractive woman) to see if she could get into the building. She saw that employees were using proximity cards to get into the building, but she didn't have one. Instead, she loaded herself up with a book bag and a laptop—ensuring her hands weren't free. She timed her approach carefully and followed closely behind an employee with a proximity card. She flashed a friendly smile, and sure enough, the employee held the door open for her.

Most of us learn to be polite and courteous and social engineers take advantage of this. It's polite to hold a door open for people who have their hands full. In contrast, it's rude to slam the door in the face of someone following behind us. However, most users don't want to help criminals. Security awareness programs and training help users understand how criminals use tactics such as tailgating. Educated users are less likely to be tricked, even by a friendly smile from an attractive woman.

High-traffic areas are most susceptible to tailgating attacks. Security guards can be an effective preventive measure at access points, but they need to be vigilant to ensure that tailgating does not occur. The best solution is a mantrap.

Preventing Tailgating with Mantraps

A ***mantrap*** is a physical security mechanism designed to control access to a secure area through a buffer zone. Personnel use something like a

proximity card to gain access, and the mantrap allows one person, and only one person, to pass through. Because they only allow one person through at a time, mantraps prevent tailgating. Mantraps get their name due to their ability to lock a person between two areas, such as an open access area and a secure access area, but not all of them are that sophisticated.

An example of a simple mantrap is a turnstile similar to what you see in many public transport systems. Even if you've never ridden the subway in one of many U.S. cities or the Tube in London, you've probably seen turnstiles in movies such as *While You Were Sleeping*. When customers present a token, the turnstile unlocks and allows a single person through at a time. Similarly, users unlock the turnstile mantrap with something like a proximity card.

A sophisticated mantrap is a room, or even a building, that creates a large buffer area between the secure area and the unsecured area. Access through the entry door and the exit door is tightly controlled, either with guards or with an access card such as a proximity card.

It's also possible to require identification and authentication before allowing passage through a mantrap. For example, a retina scanner can identify individuals and restrict access to only authorized individuals. Similarly, some card reader systems support the use of unique PINs assigned to the user. Users present their card and enter their PIN to gain access before the mantrap opens.

Remember this

Tailgating is a social engineering tactic that occurs when one user follows closely behind another user without using credentials. Mantraps allow only a single person to pass at a time. Sophisticated mantraps can identify and authenticate individuals before allowing access.

Increasing Physical Security with Guards

Many organizations use security guards to control access to buildings and secure spaces. If employees have ID badges, guards can check these

badges prior to granting the employees access. Even if ID badges aren't used, guards can still verify people's identity using other identification. Similarly, the security guards can restrict access by checking people's identity against a preapproved access control list. In some cases, guards record all access in an access log. Security guards can also take a less-active role to deter security incidents. For example,
a security guard can deter tailgating incidents by observing personnel when they use their proximity card to gain access to a secure area.

Monitoring Areas with Cameras

Organizations are increasingly using security cameras in the workplace and surrounding areas for video surveillance. This includes areas outside of a building, such as a parking lot, and all building entrances and exits. Additionally, many organizations use cameras to monitor internal entrances of high-security areas, such as the entrance of a data center or server room.

Cameras are connected to a closed-circuit television (CCTV) system, which transmits signals from video cameras to monitors that are similar to TVs. In addition to providing security, CCTV can also enhance safety by deterring threats.

Organizations often use video cameras within a work environment to protect employees and enhance security in the workplace. In addition to live monitoring, most systems include a recording element, and they can verify if someone is stealing the company's assets. By recording activity, videos can be played back later for investigation and even prosecution.

Video surveillance provides the most reliable proof of a person's location and activity. Access logs provide a record, but it's possible to circumvent the security of an access log. For example, if Bart used your proximity card to gain access to a secure space, the log will indicate you entered, not Bart. In contrast, if the video shows that Bart entered the room at a certain time of day, it's not easy for Bart to refute the video.

Remember this

Video surveillance provides reliable proof of a person's location and activity. It can identify who enters and exits secure areas and can record theft of assets.

When using video surveillance in a work environment, it's important to

respect privacy and to be aware of privacy laws. Some things to consider are:

- **Only record activity in public areas.** People have a reasonable expectation of privacy in certain areas, such as locker rooms and restrooms, and it is often illegal to record activity in these areas.
- **Notify employees of the surveillance.** If employees aren't notified of the surveillance, legal issues related to the video surveillance can arise. This is especially true if the recordings are used when taking legal and/or disciplinary actions against an employee.
- **Do not record audio.** Recording audio is illegal in many jurisdictions, without the express consent of all parties being recorded. Many companies won't even sell surveillance cameras that record audio.

Fencing, Lighting, and Alarms

Fences provide a barrier around a property and deter people from entering. When using a fence, it's common to control access to the area via specific gates. Guards often monitor these gates and ensure only authorized individuals can enter. When additional security is required, organizations sometimes configure dual gates, allowing access into one area where credentials are checked before allowing full access. This effectively creates a cage preventing full access, but also prevents unauthorized individuals from escaping.

Installing lights at all the entrances to a building can deter attackers from trying to break in. Similarly, lighting at the entrances of any internal restricted areas can deter people from trying to enter. Many organizations use a combination of automation, light dimmers, and motion sensors to save on electricity costs without sacrificing security. The lights automatically turn on at dusk, but in a low, dimmed mode. When the motion sensors detect any movement, the lights turn on at full capacity. They automatically turn off at dawn.

It's important to protect the lights. For example, if an attacker can remove the light bulbs, it defeats the control. Either place the lights high enough so that they can't be easily reached, or protect them with a metal cage.

Alarms provide an additional physical security protection. This includes

alarms that detect fire and alarms that detect unauthorized access. Fire alarms detect smoke and/or heat and trigger fire suppression systems. Burglary prevention systems monitor entry points such as doors and windows, detecting when someone opens them.

You can also combine motion detection systems with burglary prevention systems. They detect movement within monitored areas and trigger alarms. Obviously, you wouldn't have motion detection systems turned on all the time. Instead, you'd turn them on when people will not be working in the area, such as during nights or weekends.

You might have noticed that fencing, lighting, and alarms can all be combined with motion detection. At the most basic level, motion detection methods detect moving objects. Many motion detectors use microwave technologies to detect movement. This is like the technology used in some police radar speed guns.

A more advanced method is infrared detection. Infrared detectors sense infrared radiation, sometimes called infrared light, which effectively sees a difference between objects of different temperatures. As an example, a person is much warmer than objects in a room and easily stands out using an infrared detector. This can help eliminate false alarms by sensing more than just motion, but motion from objects of different temperatures.

Remember this

Fencing, lighting, and alarms all provide physical security. They are often used together to provided layered security. Motion detection methods are also used with these methods to increase their effectiveness. Infrared detectors detect movement by objects of different temperatures.

Securing Access with Barricades

In some situations, fencing isn't enough to deter potential attackers. To augment fences and other physical security measures, organizations erect stronger barricades. As an example, military bases often erect strong, zigzag barricades that require vehicles to slow down to navigate through them. This prevents attackers from trying to ram through the gates.

Businesses and organizations need to present an inviting appearance, so they can't use such drastic barricades. However, they often use ***bollards***, which

are short vertical posts, composed of reinforced concrete and/or steel. They often place the bollards in front of entrances about three or four feet apart. They typically paint them with colors that match their store. You've probably walked through a set of bollards multiple times without giving them a second thought. However, thieves who are contemplating driving a car or truck through the entrance see them.

Many thieves have driven vehicles right through the front of buildings, and then proceeded to steal everything in sight. Depending on the strength of the walls, criminals might even be able to drive through a wall with a truck. Strategically placed bollards will prevent these types of attacks.

Remember this

Barricades provide stronger barriers than fences and attempt to deter attackers. Bollards are effective barricades that can block vehicles.

Using Hardware Locks

You can implement simple physical security measures to prevent access to secure areas. For example, you can use hardware locks—similar to what you use to secure your home—to secure buildings as well as rooms within buildings. Companies that don't have the resources to employ advanced security systems often use these types of hardware locks.

Instead of allowing free access to wiring closets or small server rooms, small organizations use these types of locks to restrict access. Although these locks aren't as sophisticated as the ones used by large organizations, they are much better than leaving the rooms open and the equipment exposed.

Key management is an important concept to consider when using hardware locks. Proper key management ensures that only authorized personnel can access the physical keys. This might be done by locking keys within a safe or locking cabinet.

Securing Mobile Computers with Cable Locks

Cable locks are a great theft deterrent for mobile computers, and even many desktop computers at work. Computer cable locks work similar to how a bicycle cable lock works. However, instead of securing a bicycle to a bike rack or post, a computer cable lock secures a computer to a piece of

furniture.

The user wraps the cable around a desk, table, or something heavy, and then plugs it into an opening in the laptop specifically created for this purpose. Most cable locks have a four-digit combo. If you (or anyone) remove the cable lock without the combo, it will likely destroy the laptop.

Another common use of cable locks is for computers in unsupervised labs. For example, you can secure laptop or desktop computers with cable locks in a training lab. This allows you to leave the room open so that students can use the equipment, but the cable locks prevent thieves from stealing the equipment.

Securing Servers with Locking Cabinets

Larger companies often have large server rooms with advanced security to restrict access. Additionally, within the server room, administrators use locking cabinets or enclosures to secure equipment mounted within the bays. An equipment bay is about the size of a large refrigerator and can hold servers, routers, and other IT equipment. These bays have doors in the back and many have doors in the front, too. Administrators lock these doors to prevent unauthorized personnel from accessing the equipment.

Offices often have file cabinets that lock, too, so it's important to pay attention to the context when referring to locking cabinets. For example, if you want to secure equipment within a server room, a locking cabinet is one of many physical security controls you can use. If you want to secure unattended smartphones in an office space, you can also use a locking cabinet, but this is an office file cabinet that locks.

Remember this

Cable locks are effective threat deterrents for small equipment such as laptops and some workstations. When used properly, they prevent losses due to theft of small equipment. Locking cabinets in server rooms provide an added physical security measure. A locked cabinet prevents unauthorized access to equipment mounted in server bays.

Securing Small Devices with a Safe

Locking file cabinets or safes used in many offices help prevent the theft of smaller devices. For example, you can store smaller devices such as external USB drives or USB flash drives in an office safe or locking cabinet when they aren't in use. Depending on the size of the office safe and office cabinet, you might also be able to secure laptops within them.

Asset Management

Asset management is the process of tracking valuable assets throughout their life cycles. For example, organizations commonly implement processes to track hardware such as servers, desktop computers, laptop computers, routers, and switches. An effective asset management system can help reduce several vulnerabilities:

- **Architecture and design weaknesses.** Asset management helps reduce architecture and design weaknesses by ensuring that purchases go through an approval process. The approval process does more than just compare costs. It also evaluates the purchase to ensure it fits in the overall network architecture. Unapproved assets often weaken security by adding in additional resources that aren't managed.
- **System sprawl and undocumented assets.** ***System sprawl*** occurs when an organization has more systems than it needs, and systems it owns are underutilized. Asset management begins before the hardware is purchased and helps prevent system sprawl by evaluating the purchase. Additionally, after the purchase is completed, asset management processes ensure hardware is added into the asset management tracking system. This ensures that the assets are managed and tracked from cradle to grave.

Many organizations use automated methods for inventory control. For example, radio- frequency identification (RFID) methods can track the movement of devices. These are the same types of devices used in stores to prevent shoplifting. If someone exits without paying, the RFID device transmits when the shoplifter gets close to the exit door and sounds an alarm. Organizations won't necessarily have an alarm, but they can track the movement of devices.

Mobile devices are easy to lose track of, so organizations often use asset-tracking methods to reduce losses. For example, when a user is issued a

mobile device, asset-tracking methods record it. Similarly, if the user leaves the company, asset-tracking methods ensure the user returns the device.

Implementing Environmental Controls

Although environmental controls might not seem security related, they directly contribute to the availability of systems. This includes ensuring temperature and humidity controls are operating properly, fire suppression systems are in place, and proper procedures are used when running cables.

Heating, Ventilation, and Air Conditioning

Heating, ventilation, and air conditioning (***HVAC***) systems are important physical security controls that enhance the availability of systems. Quite simply, computers and other electronic equipment can't handle drastic changes in temperatures, especially hot temperatures. If systems overheat, the chips can actually burn themselves out.

The cooling capacity of HVAC systems is measured as tonnage. This has nothing to do with weight, but instead refers to cooling capacity. One ton of cooling equals 12,000 British thermal units per hour (Btu/hour), and typical home HVAC systems are three-ton units. Higher-tonnage HVAC systems can cool larger areas or areas with equipment generating more heat.

The amount of air conditioning needed to cool a massive data center is much greater than you need to cool your home, primarily because of all the heat generated by the equipment. If your home air conditioner fails in the middle of summer, you might be a little uncomfortable for a while, but if the data center HVAC system fails, it can result in loss of availability and a substantial loss of money.

I worked in several environments where we had a policy of shutting down all electronics when the room temperature reached a certain threshold. When we didn't follow the policy, the systems often developed problems due to the heat and ended up out of commission for a lot longer than the AC.

Most servers aren't in cases like a typical desktop computer. Instead, they are housed in rack-mountable cases. These rack-mountable servers are installed in equipment cabinets (also called racks or bays) about the size of tall refrigerators. A large data center will have multiple cabinets lined up beside each other in multiple rows.

These cabinets usually have locking doors in the front and rear for physical security. The doors are perforated with cold air coming in the front, passing over and through the servers to keep them cool, and warmer air exiting out the rear. Additionally, a server room has raised flooring with air conditioning pumping through the space under the raised floor.

Remember this

Higher-tonnage HVAC systems provide more cooling capacity. This keeps server rooms at lower operating temperatures and results in fewer failures.

Hot and Cold Aisles

Hot and cold aisles help regulate the cooling in data centers with multiple rows of cabinets. The back of all the cabinets in one row faces the back of all the cabinets in an adjacent row. Because the hot air exits out the back of the cabinet, the aisle with the backs facing each other is the hot aisle.

Similarly, the front of the cabinets in one row is facing the front of the cabinets in the adjacent row. Cool air is pumped through the floor to this cool aisle using perforated floor tiles in the raised flooring. This is the cold aisle. In some designs, cool air is also pumped through the base of the cabinets. This depends on the design of the cabinets and the needs of the equipment. Consider what happens if all the cabinets had their front facing the same way without a hot/cold aisle design. The hot air pumping out the back of one row of cabinets would be sent to the front of the cabinets behind them. The front row would have very cold air coming in the front, but other rows would have warmer air coming in the front.

Of course, an HVAC also includes a thermostat as a temperature control and additional humidity controls. The thermostat ensures that the air temperature is controlled and maintained. Similarly, humidity controls ensure that the humidity is controlled. High humidity can cause condensation on the equipment, which causes water damage. Low humidity allows a higher incidence of electrostatic discharge (ESD).

HVAC and Fire

HVAC systems are often integrated with fire alarm systems to help prevent a fire from spreading. One of the core elements of a fire is oxygen. If

the HVAC system continues to operate normally while a fire is active, it continues to pump oxygen, which feeds the fire. When the HVAC system is integrated with the fire alarm system, it controls the airflow to help prevent the rapid spread of the fire. Many current HVAC systems have dampers that can control airflow to specific areas of a building. Other HVAC systems automatically turn off when fire suppression systems detect a fire.

Remember this

HVAC systems increase availability by controlling temperature and humidity. Temperature controls help ensure a relatively constant temperature. Humidity controls reduce the potential for damage from electrostatic discharge and damage from condensation. HVAC systems should be integrated with the fire alarm systems and either have dampers or the ability to be turned off in the event of a fire.

Fire Suppression

You can fight fires with individual fire extinguishers, with fixed systems, or both. Most organizations included fixed systems to control fires and place portable fire extinguishers in different areas around the organization. A fixed system can detect a fire and automatically activate to extinguish the fire. Individuals use portable fire extinguishers to extinguish or suppress small fires.

The different components of a fire are heat, oxygen, fuel, and a chain reaction creating the fire. Fire suppression methods attempt to remove or disrupt one of these elements to extinguish a fire. You can extinguish a fire using one of these methods:

- **Remove the heat.** Fire extinguishers commonly use chemical agents or water to remove the heat. However, water should never be used on an electrical fire.
- **Remove the oxygen.** Many methods use a gas, such as carbon dioxide (CO_2) to displace the oxygen. This is a common method of fighting electrical fires because CO_2 and similar gasses are harmless to electrical equipment.
- **Remove the fuel.** Fire-suppression methods don't typically fight a fire this way, but of course, the fire will go out once all the material is burned.

- **Disrupt the chain reaction.** Some chemicals can disrupt the chain reaction of fires to stop them.

When implementing any fire suppression system, it's important to consider the safety of personnel. As an example, if a fire suppression system uses a gas such as carbon dioxide (CO_2) to displace the oxygen, it's important to ensure that personnel can get out before the oxygen is displaced.

Similarly, consider an exit door secured with a proximity card. Normally, employees open the door with the proximity card and the system records their exit. What happens if a fire starts and power to the building is lost? The proximity card reader won't work, and if the door can't open, employees will be trapped. It's important to ensure that an alternative allows personnel to exit even if the proximity card reader loses power. Of course, this might introduce a vulnerability to consider. You don't want an attacker to access a secure data center just by removing power to the proximity reader.

Environmental Monitoring

Environmental monitoring includes temperature and humidity controls. From a very basic perspective, an HVAC system monitors the current temperature and humidity and makes adjustments as necessary to keep the temperature and humidity constant.

Large-scale data centers often have sophisticated logging capabilities for environmental monitoring. The HVAC system still attempts to keep the temperature and humidity constant. However, the logs record the actual temperature and humidity at different times during the day. This allows administrators to review the performance of the HVAC system, to see if it is able to keep up with the demands within the data center.

Shielding

Shielding helps prevent electromagnetic interference (EMI) and radio frequency interference (RFI) from interfering with normal signal transmissions. It also protects against unwanted emissions and helps prevent an attacker from capturing network traffic.

Although you might see EMI and RFI in the same category as

EMI/RFI, they are different. EMI comes from different types of motors, power lines, and even fluorescent lights. RFI comes from radio frequency (RF) sources such as AM or FM transmitters. However, shielding used to block interference from both EMI and RFI sources is often referred to as simply EMI shielding.

Attackers often use different types of eavesdropping methods to capture network traffic. If the data is emanating outside of the wire or outside of an enclosure, attackers may be able to capture and read the data. EMI shielding fulfills the dual purpose of keeping interference out and preventing attackers from capturing network traffic.

Protected Cabling

Twisted-pair cable, such as CAT5e and CAT6 cable, comes in both shielded twisted-pair (STP) and unshielded twisted-pair (UTP) versions. The shielding helps prevent an attacker from capturing network traffic and helps block interference from corrupting the data.

When data travels along a copper wire (such as twisted-pair), it creates an induction field around the wire. If you have the right tools, you can simply place the tool around the wire and capture the signal. The shielding in STP cable blocks this. Fiber-optic cable is not susceptible to this type of attack. Signals travel along a fiber-optic cable as light pulses, and they do not create an induction field.

Protected Distribution of Cabling

Physical security includes planning where you route cables and how you route them. Skilled network administrators can cut a twisted-pair cable, attach an RJ-45 connector to each end, and connect them back together with an adapter in less than 5 minutes. Experienced fiber-optic cable technicians can do the same thing with a fiber-optic cable within 10 minutes.

If an attacker did this, he could connect the cut cable with a hub, and then capture all the traffic going through the hub with a protocol analyzer. This represents a significant risk.

One method of reducing this risk is to run cables through cable troughs or wiring ducts. A cable trough is a long metal container, typically about 4 inches wide by 4 inches high. If you run data cables through the cable trough, they aren't as accessible to potential attackers. In contrast, many organizations

simply run the cable through a false ceiling or a raised floor.

In addition to considering physical security, it's important to keep the cables away from EMI sources. As an example, if technicians run cables over or through fluorescent lighting fixtures, the EMI from the lights can disrupt the signals on the cables. The result is intermittent connectivity for users.

Faraday Cage

A ***Faraday cage*** is typically a room that prevents signals from emanating beyond the room. It includes electrical features that cause RF signals that reach the boundary of the room to be reflected back, preventing signal emanation outside the Faraday cage. A Faraday cage can also be a small enclosure.

In addition to preventing signals from emanating outside the room, a Faraday cage also provides shielding to prevent outside interference such as EMI and RFI from entering the room. At a very basic level, some elevators act as a Faraday cage (though I seriously doubt the designers were striving to do so). You might have stepped into an elevator and found that your cell phone stopped receiving and transmitting signals. The metal shielding around the elevator prevents signals from emanating out or signals such as the cell phone tower signal from entering
the elevator.

On a smaller scale, electrical devices such as computers include shielding to prevent signals from emanating out and block interference from getting in.

Remember this

EMI shielding prevents outside interference sources from corrupting data and prevents data from emanating outside the cable. Cable troughs protect cables distributed throughout a building in metal containers. A Faraday cage prevents signals from emanating beyond the cage.

Adding Redundancy and Fault

Tolerance

One of the constants with computers, subsystems, and networks is that they will fail. It's one of the few things you can count on. It's not a matter of *if* they will fail, but *when*. However, by adding redundancy into your systems and networks, you can increase the reliability of your systems even when they fail. By increasing reliability, you increase one of the core security goals: availability.

Redundancy adds duplication to critical system components and networks and provides ***fault tolerance***. If a critical component has a fault, the duplication provided by the redundancy allows the service to continue as if a fault never occurred. In other words, a system with fault tolerance can suffer a fault, but it can tolerate it and continue to operate. Organizations often add redundancies to eliminate single points of failure.

You can add redundancies at multiple levels:

- Disk redundancies using RAID
- Server redundancies by adding failover clusters
- Power redundancies by adding generators or an UPS
- Site redundancies by adding hot, cold, or warm sites

Single Point of Failure

A ***single point of failure*** is a component within a system that can cause the entire system to fail if the component fails. When designing redundancies, an organization will examine different components to determine if they are a single point of failure. If so, they take steps to provide a redundancy or fault-tolerance capability. The goal is to increase reliability and availability of the systems.

Some examples of single points of failure include:

- **Disk.** If a server uses a single drive, the system will crash if the single drive fails. Redundant array of inexpensive disks (RAID) provides fault tolerance for hard drives and is a relatively inexpensive method of adding fault tolerance to a system.
- **Server.** If a server provides a critical service and its failure halts the service, it is a single point of failure. Failover clusters (discussed later in this chapter) provide fault tolerance for critical servers.

- **Power.** If an organization only has one source of power for critical systems, the power is a single point of failure. However, elements such as uninterruptible power supplies (UPSs) and power generators provide fault tolerance for power outages.

Although IT personnel recognize the risks with single points of failure, they often overlook them until a disaster occurs. However, tools such as business continuity plans (covered later in this chapter) help an organization identify critical services and address single points of failure.

Remember this

A single point of failure is any component whose failure results in the failure of an entire system. Elements such as RAID, failover clustering, UPSs, and generators remove many single points of failure. RAID is an inexpensive method used to add fault tolerance and increase availability.

Disk Redundancies

Any system has four primary resources: processor, memory, disk, and the network interface. Of these, the disk is the slowest and most susceptible to failure. Because of this, administrators often upgrade disk subsystems to improve their performance and redundancy.

Redundant array of inexpensive disks (***RAID***) subsystems provide fault tolerance for disks and increase the system availability. Even if a disk fails, most RAID subsystems can tolerate the failure and the system will continue to operate. RAID systems are becoming much more affordable as the price of drives steadily falls and disk capacity steadily increases. While it's expected that you are familiar with RAID subsystems, the following sections provide a short summary to remind you of the important details.

RAID-0

RAID-0 (striping) is somewhat of a misnomer because it doesn't provide any redundancy or fault tolerance. It includes two or more physical disks. Files stored on a RAID-0 array are spread across each of the disks.

The benefit of a RAID-0 is increased read and write performance. Because a file is spread across multiple physical disks, the different parts of

the file can be read from or written to each of the disks at the same time. If you have three 500 GB drives used in a RAID-0, you have 1,500 GB (1.5 TB) of storage space.

RAID-1

RAID-1 (mirroring) uses two disks. Data written to one disk is also written to the other disk. If one of the disks fails, the other disk still has all the data, so the system can continue to operate without any data loss. With this in mind, if you mirror all the drives in a system, you can actually lose half of the drives and continue to operate.

You can add an additional disk controller to a RAID-1 configuration to remove the disk controller as a single point of failure. In other words, each of the disks also has its own disk controller. Adding a second disk controller to a mirror is called disk duplexing.

If you have two 500 GB drives used in a RAID-1, you have 500 GB of storage space. The other 500 GB of storage space is dedicated to the fault-tolerant, mirrored volume.

RAID-2, RAID 3, and RAID-4 are rarely used.

RAID-5 and RAID-6

A RAID-5 is three or more disks that are striped together similar to RAID-0. However, the equivalent of one drive includes parity information. This parity information is striped across each of the drives in a RAID-5 and is used for fault tolerance. If one of the drives fails, the system can read the information on the remaining drives and determine what the actual data should be. If two of the drives fail in a RAID-5, the data is lost.

RAID-6 is an extension of RAID-5, and it includes an additional parity block. A huge benefit is that the RAID-6 disk subsystem will continue to operate even if two disk drives fail. RAID-6 requires a minimum of four disks.

Remember this

RAID subsystems, such as RAID-1, RAID-5, and RAID-6, provide fault tolerance and increased data availability. RAID-5 can survive the failure of one disk. RAID-6 can survive the failure of two disks.

RAID-10

A RAID-10 configuration combines the features of mirroring (RAID-1) and striping (RAID-0). RAID-10 is sometimes called RAID 1+0. A variation is RAID-01 or RAID 0+1 that also combines the features of mirroring and striping but implements the drives a little differently.

The minimum number of drives in a RAID-10 is four. When adding more drives, you add two (or multiples of two such as four, six, and so on). If you have four 500 GB drives used in a RAID-10, you have 1 TB of usable storage.

Server Redundancy and High Availability

High availability refers to a system or service that needs to remain operational with almost zero downtime. Utilizing different redundancy and fault-tolerance methods, it's possible to achieve 99.999 percent uptime, commonly called five nines. This equates to less than 6 minutes of downtime a year: 60 minutes × 24 hours × 365 days ×.00001 = 5.256 minutes. Failover clusters are a key component used to achieve five nines.

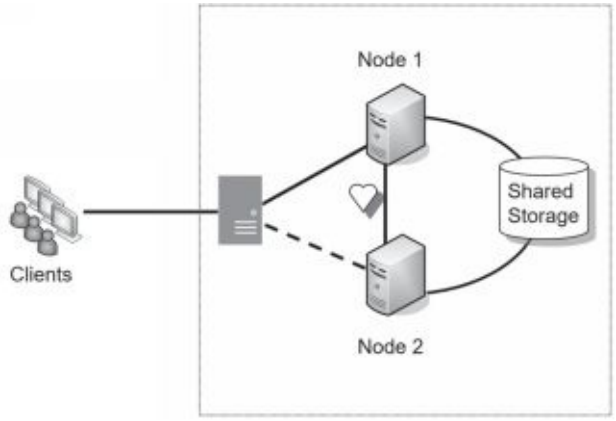
Although five nines is achievable, it's expensive. However, if the potential cost of an outage is high, the high cost of the redundant technologies is justified. For example, some web sites generate a significant amount of revenue, and every minute a web site is unavailable represents lost money. High-capacity failover clusters ensure the service is always available even if a server fails.

Distributive allocation is another option to provide both high availability and scalability, though it is typically used primarily in scientific applications. In a distributed application model, multiple computers (often called nodes) are configured to work together to solve complex problems. These computers are configured within a local network. A central processor divides the complex problem into smaller tasks. It then coordinates tasking of the individual nodes and collecting the results. If any single nodes fail, the central processor doesn't task it anymore, but overall processing continues, providing high availability. This also provides high scalability because it is relatively easy to add additional nodes and task them when

they come online.

Failover Clusters for High Availability

The primary purpose of a failover cluster is to provide high availability for a service offered by a server. Failover clusters use two or more servers in a cluster configuration, and the servers are referred to as nodes. At least one server or node is active and at least one is inactive. If an active node fails, the inactive node can take over the load without interruption to clients.



Consider Figure 9.1, which shows a two-node active-passive failover cluster. Both nodes are individual servers, and they both have access to external data storage used by the active server. Additionally, the two nodes have a monitoring connection to each other used to check the health or heartbeat of each other.

Figure 9.1: Failover cluster

Imagine that Node 1 is the active node. When any of the clients connect, the cluster software (installed on both nodes) ensures that the clients connect to the active node. If Node 1 fails, Node 2 senses the failure through the heartbeat connection and configures itself as the active node. Because both nodes have access to the shared storage, there is no loss of data for the client. Clients may notice a momentary hiccup or pause, but the service continues.

You might notice that the shared storage in Figure 9.1 represents a single point of failure. It's not uncommon for this to be a robust hardware RAID-10. This ensures that even if a hard drive in the shared storage fails, the service will continue. Additionally, if both nodes are plugged into the same power grid, the power represents a single point of failure. They can each be protected with a separate UPS, and use a separate power grid.

It's also possible to configure the cluster as an active-active cluster. Instead of one server being passive, the cluster balances the load between

both servers.

Cluster configurations can include many more nodes than just two. However, nodes need to have close to identical hardware and are often quite expensive, but if a company truly needs to achieve 99.999 percent uptime, it's worth the expense.

Load Balancers for High Availability

A ***load balancer*** can optimize and distribute data loads across multiple computers or multiple networks. For example, if an organization hosts a popular web site, it can use multiple servers hosting the same web site in a web farm. Load-balancing software distributes traffic equally among all the servers in the web farm, typically located in a DMZ.

The term load balancer makes it sound like it's a piece of hardware, but a load balancer can be hardware or software. A hardware-based load balancer accepts traffic and directs it to servers based on factors such as processor utilization and the number of current connections to the server. A software-based load balancer uses software running on each of the servers in the load-balanced cluster to balance the load. Load balancing primarily provides scalability, but it also contributes to high availability. Scalability refers to the ability of a service to serve more clients without any decrease in performance. Availability ensures that systems are up and operational when needed. By spreading the load among multiple systems, it ensures that individual systems are not overloaded, increasing overall availability.

Consider a web server that can serve 100 clients per minute, but if more than 100 clients connect at a time, performance degrades. You need to either scale up or scale out to serve more clients. You scale the server up by adding additional resources, such as processors and memory, and you scale out by adding additional servers in a load balancer.

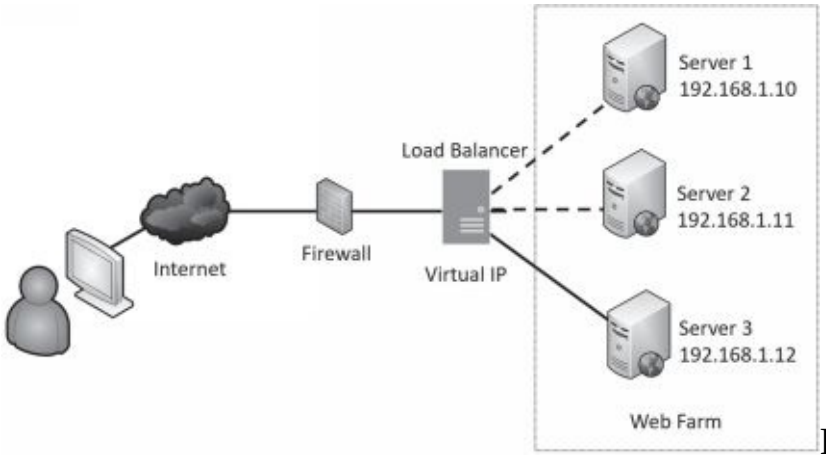


Figure 9.2 shows an example of a load balancer with multiple web servers configured in a web farm. Each web server includes the same web application. A load balancer uses a scheduling technique to determine where to send new requests. Some load balancers simply send new requests to the servers in a ***round-robin*** fashion. The load balancer sends the first request to Server 1, the second request to Server 2, and so on. Other load balancers automatically detect the load on individual servers and send new clients to the least used server.

Figure 9.2: Load balancing

Some load balancers use source address ***affinity*** to direct the requests. Source affinity sends requests to the same server based on the requestor's IP address. As an example, imagine that Homer sends a request to retrieve a web page. The load balancer records his IP address and sends his request to Server 3. When he sends another request, the load balancer identifies his IP address and sends his request to Server 3 again. Source affinity effectively sticks users to a specific server for the duration of their sessions.

A software-based load balancer uses a virtual IP. For example, imagine the IP address of the web site is 72.52.206.134. This IP address isn't assigned to a specific server. Instead, clients send requests to this IP address and the load-balancing software redirects the request to one of the three servers in the web farm using their private IP addresses. In this scenario, the actual IP address is referred to as a virtual IP.

An added benefit of many load balancers is that they can detect when a server fails. If a server stops responding, the load-balancing software no longer sends clients to this server. This contributes to overall high availability for the load balancer.

Remember this

Failover clusters are one method of server redundancy and they provide high availability for servers. They can remove a server as a single point of failure. Load balancing increases the overall processing power of a service by sharing the load among multiple servers. Configurations can be active-passive, or active-active. Scheduling methods include round-robin and source IP address affinity. Source IP address affinity scheduling ensures clients are redirected to the same server for an entire session.

Clustering Versus Load Balancing

It's worth mentioning that CompTIA has grouped both clustering and load balancing into the same category of load balancing in the objectives. Many IT professionals do the same thing, though technically they are different concepts. In general, failover clusters are commonly used for applications such as database applications. Load balancers are often used for services, such as web servers in a web farm.

Power Redundancies

Power is a critical utility to consider when reviewing redundancies. For mission-critical systems, you can use uninterruptible power supplies and generators to provide both fault tolerance and high availability. An UPS provides fault tolerance for power and can protect against power fluctuations. It provides short-term power. Generators provide long-term power in extended outages.

Protecting Data with Backups

Backups are copies of data created to ensure that if the original data is lost or corrupted, it can be restored. Maybe I should restate that. Backups are copies of data created to ensure that *when* the original data is lost or corrupted, it can be restored. The truth is, if you work with computers long enough, you will lose data. The difference between a major catastrophe and a minor inconvenience is the existence of a usable backup.

A Backup Horror Story (Sidebar)

A friend of mine was a consultant for small businesses and was once hired to help a small business owner recover some lost data. The owner had been growing his business for about five years and had just about everything related to his business (client lists, billing information, proposals, agreements, and more) on one system. The system crashed.

The consultant tried to restore information from the disk but wasn't successful. The business owner panicked, knowing he simply needed the information. If he couldn't get the data back, his business might fail.

Although it's expensive, it is possible to have a clean-room facility take a hard drive apart and read the data at the bit level to restore at least some of the data. At this point, the owner was willing to try anything, so he paid the high price and they sent the disk to a recovery facility. Unfortunately, the disk suffered a catastrophic failure, and they weren't able to retrieve any meaningful data even in the clean room.

My friend visited the owner to relay the bad news. He said that when he left, the owner had his head in his hands and was literally crying. The business he had built for five years was close to ruins without much chance for recovery.

The worst part of this story is that it's repeated over and over with many different people in many different environments. Too many people don't recognize the importance of backups until they've lost their data. Unfortunately, by then, it's too late.

It's important to realize that redundancy and backups are not the same thing. Protecting data with a RAID-1 or RAID-10 does not negate the need for backups. If a fire destroys a server, it also destroys the data on the RAID. Without a backup, all of the data is gone. Forever.

Comparing Backup Types

Backup utilities support several different types of backups. Even though third-party backup programs can be quite sophisticated in what they do and how they do it, you should have a solid understanding of the basics.

The most common media used for backups is tape. Tapes store more data and are cheaper than other media, though some organizations use hard drives for backups. However, the type of media doesn't affect the backup type.

The following backup types are commonly used:

- **Full backup.** A full (or normal backup) backs up all the selected data.
- **Differential backup.** This backs up all the data that has changed or is different since the last full backup.
- **Incremental backup.** This backs up all the data that has changed since the last full or incremental backup.
- **Snapshots.** A snapshot backup captures the data at a point in time. It is sometimes referred to as an image backup.

Full Backups

A ***full backup*** backs up all data specified in the backup. For example, you could have several folders on the D: drive. If you specify these folders in the backup program, the backup program backs up all the data in these folders.

Although it's possible to do a full backup on a daily basis, it's rare to do so in most production environments. This is because of two limiting factors:

- **Time.** A full backup can take several hours to complete and can interfere with operations. However, administrators don't always have unlimited time to do backups and other system maintenance. For example, if a system is online 24/7, administrators might need to limit the amount of time for full backups to early Sunday morning to minimize the impact on users.
- **Money.** Backups need to be stored on some type of media, such as tape or hard drives. Performing full backups every day requires more media, and the cost can be prohibitive. Instead, organizations often combine full backups with differential or incremental backups.

However, every backup strategy must start with a full backup.

Restoring a Full Backup

A full backup is the easiest and quickest to restore. You only need to restore the single full backup and you're done. If you store backups on tapes, you only need to restore a single tape. However, most organizations need to balance time and money and use either a full/differential or a full/incremental backup strategy.

Differential Backups

A ***differential backup*** strategy starts with a full backup. After the full backup, differential backups back up data that has changed or is different since the last full backup.

For example, a full/differential strategy could start with a full backup on Sunday night. On Monday night, a differential backup would back up all files that changed since the last full backup on Sunday. On Tuesday night, the differential backup would again back up all the files that changed since the last full backup. This repeats until Sunday, when another full backup starts the process again. As the week progresses, the differential backup steadily grows in size.

Order of Restoration for a Full/Differential Backup Set

Assume for a moment that each of the backups was stored on different tapes. If the system crashed on Wednesday morning, how many tapes would you need to recover the data?

The answer is two. You would first recover the full backup from Sunday. Because the differential backup on Tuesday night includes all the files that changed after the last full backup, you would restore that tape to restore all the changes up to Tuesday night.

Incremental Backups

An ***incremental backup*** strategy also starts with a full backup. After the full backup, incremental backups then back up data that has changed since the last backup. This includes either the last full backup, or the last incremental backup.

As an example, a full/incremental strategy could start with a full backup on Sunday night. On Monday night, an incremental backup would back up all

the files that changed since the last full backup. On Tuesday night, the incremental backup would back up all the files that changed since the incremental backup on Monday night. Similarly, the Wednesday night backup would back up all files that changed since the last incremental backup on Tuesday night. This repeats until Sunday when another full backup starts the process again. As the week progresses, the incremental backups stay about the same size.

Order of Restoration for a Full/Incremental Backup Set

Assume for a moment that each of the backups were stored on different tapes. If the system crashed on Thursday morning, how many tapes would you need to recover the data?

The answer is four. You would first need to recover the full backup from Sunday. Because the incremental backups would be backing up different data each day of the week, each of the incremental backups must be restored—and must be restored in chronological order.

Sometimes, people mistakenly think the last incremental backup would have all the relevant data. Although it might have some relevant data, it doesn't have everything.

As an example, imagine you worked on a single project file each day of the week, and the system crashed on Thursday morning. In this scenario, the last incremental backup would hold the most recent copy of this file. However, what if you compiled a report every Monday but didn't touch it again until the following Monday? Only the incremental backup from Monday would include the most recent copy. An incremental backup from Wednesday night or another day of the week wouldn't include the report.

Remember this

If you have unlimited time and money, the full backup alone provides the fastest recovery time. Full/incremental strategies reduce the amount of time needed to perform backups. Full/differential strategies reduce the amount of time needed to restore backups.

Choosing Full/Incremental or Full/Differential

A logical question is, “Why are there so many choices for backups?” The answer is that different organizations have different needs. For example, imagine two organizations perform daily backups to minimize losses. They each do a full backup on Sunday, but are now trying to determine if they should use a full/incremental or a full/differential strategy.

The first organization doesn’t have much time to perform maintenance throughout the week. In this case, the backup administrator needs to minimize the amount of time required to complete backups during the week. An incremental backup only backs up the data that has changed since the last backup. In other words, it includes changes only from a single day. In contrast, a differential backup includes all the changes since the last full backup. Backing up the changes from a single day takes less time than backing up changes from multiple days, so a full/ incremental backup is the best choice.

In the second organization, recovery of failed systems is more important. If a failure requires restoring data, they want to minimize the amount of time needed to restore the data. A full/ differential is the best choice in this situation because it only requires the restoration of two backups, the full and the most recent differential backup. In contrast, a full/incremental can require the restoration of several different backups, depending on when the failure occurs.

Snapshot Backup

A snapshot backup captures the data at a moment in time. It is commonly used with virtual machines and sometimes referred to as a checkpoint. Chapter 1, “Mastering Security Basics,” discusses virtual machines (VMs) and administrators often take a snapshot of a VM before a risky operation such as an update. If the update causes problems, it’s relatively easy to revert the VM to the state it was in before the update.

Testing Backups

I’ve heard many horror stories in which personnel are regularly performing backups thinking all is well. Ultimately, something happens and they need to restore some data. Unfortunately, they discover that none of the backups hold valid data. People have been going through the motions, but

something in the process is flawed.

The only way to validate a backup is to perform a test restore. Performing a test restore is nothing more than restoring the data from a backup and verifying its integrity. If you want to verify that you can restore the entire backup, you perform a full restore of the backup. If you want to verify that you can restore individual files, you perform a test restore of individual files. It's common to restore data to a different location other than the original source location, but in such a way that you can validate the data.

As a simple example, an administrator can retrieve a random backup and attempt to restore it. There are two possible outcomes of this test, and both are good:

- **The test succeeds.** Excellent! You know that the backup process works. You don't necessarily know that every backup tape is valid, but at least you know that the process is sound and at least some of your backups work.
- **The test fails.** Excellent! You know there's a problem that you can fix before a crisis. If you discovered the problem after you actually lost data, it wouldn't help you restore the data.

An additional benefit of performing regular test restores is that it allows administrators to become familiar with the process. The first time they do a restore shouldn't be in the middle of a crisis with several high-level managers peering over their shoulders.

Protecting Backups

If data is important enough to be backed up, it's important enough to protect. Backup media should be protected at the same level as the data that it holds. In other words, if proprietary data enjoys the highest level of protection within an organization, then backups of this data should also have the highest level of protection.

Protecting backups includes:

- **Storage.** This includes using clear labeling to identify the data and physical security protection to prevent others from easily accessing it while it's stored.
- **Transfer.** Data should be protected any time it is transferred from one location to another. This is especially true when transferring a copy of the backup to a separate geographical location.

- **Destruction.** When the backups are no longer needed, they should be destroyed. This can be accomplished by degaussing the media, shredding or burning the media, or scrubbing the media by repeatedly writing varying patterns of 1s and 0s onto the media.

Backups and Geographic Considerations

Organizations typically create a backup policy to answer critical questions related to backups. The backup policy is a written document and will often identify issues such as what data to back up, how often to back up the data, how to test the backups, and how long to retain the backups.

Additionally, it's important to address special geographic considerations, such as the following:

- **Off-site backups.** A copy of a backup should be stored in a separate geographic location. This protects against a disaster such as a fire or flood. Even if a disaster destroys the site, the organization will still have another copy of the critical data.
- **Distance.** Many organizations have specific requirements related to the distance between the main site and the off-site location. In some scenarios, the goal is to have the off-site location relatively close so that backups can be easily retrieved. However, in other scenarios, the off-site location must be far away, such as 25 miles or further away.
- **Location selection.** The location is often dependent on environmental issues. As an example, consider an organization located in California near the San Andreas fault. The off-site backup location should be far enough away that an earthquake at the primary location doesn't affect the off-site location.
- **Legal implications.** The legal implications related to backups depends on the data stored in the backups. For example, if the backups include Personally Identifiable Information (PII) or Protected Health Information (PHI), the backups need to be protected according to governing laws.
- **Data sovereignty.** ***Data sovereignty*** refers to the legal

implications when data is stored off-site. If the backups are stored in a different country, they are subject to the laws of that country. This can be a concern if the backups are stored in a cloud location, and the cloud servers are in a different country. For example, imagine that an organization is located in the United States. It routinely does backups and stores them with a cloud provider. The cloud provider has some servers in the United States, some in Canada, and some in Mexico. If the organization's backups are stored in the other countries, it can be subject to additional laws and regulations.

Remember this

Test restores are the best way to test the integrity of a company's backup data. Backup media should be protected with the same level of protection as the data on the backup. Geographic considerations for backups include storing backups off-site, choosing the best location, and considering legal implications and data sovereignty.

Comparing Business Continuity Elements

Business continuity planning helps an organization predict and plan for potential outages of critical services or functions. The goal is to ensure that critical business operations continue and the organization can survive the outage. Organizations often create a business continuity plan (BCP). This plan includes disaster recovery elements that provide the steps used to return critical functions to operation after an outage.

Disasters and outages can come from many sources, including:

- Fires
- Attacks
- Power outages
- Data loss from any cause
- Hardware and software failures
- Natural disasters, such as hurricanes, floods, tornadoes, and earthquakes

Addressing all of these possible sources takes a lot of time and effort.

The goal is to predict the relevant disasters, their impact, and then develop recovery strategies to mitigate them. One of the first things an organization completes is a business impact analysis.

Business Impact Analysis Concepts

A ***business impact analysis (BIA)*** is an important part of a BCP. It helps an organization identify critical systems and components that are essential to the organization's success. These critical systems support mission-essential functions. The BIA also helps identify vulnerable business processes. These are processes that support mission-essential functions.

As an example, imagine an organization has an online e-commerce business. Some basic mission-essential functions might include serving web pages, providing a shopping cart path, accepting purchases, sending email confirmations, and shipping purchases to customers. The shopping cart path alone is a business process and because it is essential to the mission of e-commerce sales, management will likely consider it a vulnerable business process to protect. The customer needs to be able to view products, select a product, enter customer information, enter credit card data, and complete the purchase. Some critical systems that support the web site are web servers and a back-end database application hosted on one or more database servers.

If critical systems and components fail and cannot be restored quickly, mission-essential functions cannot be completed. If this lasts too long, it's very possible that the organization will not survive the disaster.

For example, if a disaster such as a hurricane hit, which services must the organization restore to stay in business? Imagine a financial institution. It might decide that customers must have uninterrupted access to account data through an online site. If customers can't access their funds online, they might lose faith with the company and leave in droves.

However, the company might decide to implement alternate business practices in other elements of the business. For example, management might decide that accepting and processing loan applications is not important enough to continue during a disaster. Loan processing is still important to the company's bottom line, but a delay will not seriously affect its ability to stay in business. In this scenario, continuous online access is a mission-essential function, but processing loan applications during a disaster is not mission-essential.

The time to make these decisions is not during a crisis. Instead, the organization completes a BIA in advance. The BIA involves collecting information from throughout the organization and documenting the results. This documentation identifies core business or mission requirements. The BIA does not recommend solutions. However, it provides management with valuable information so that they can focus on critical business functions. It helps them address some of the following questions:

- What are the critical systems and functions?
- Are there any dependencies related to these critical systems and functions?
- What is the maximum downtime limit of these critical systems and functions?
- What scenarios are most likely to impact these critical systems and functions?
- What is the potential loss from these scenarios?

As an example, imagine an organization earns an average of $5,000 an hour through online sales. In this scenario, management might consider online sales to be a mission-essential function and all systems that support online sales are critical systems. This includes web servers and back-end database servers. These servers depend on the network infrastructure connecting them, Internet access, and access to payment gateways for credit card charges.

After analysis, they might determine that the maximum allowable outage for online sales is five hours. Identifying the maximum downtime limit is extremely important. It drives decisions related to recovery objectives and helps an organization identify various contingency plans and policies.

Impact

The BIA evaluates various scenarios, such as fires, attacks, power outages, data loss, hardware and software failures, and natural disasters. Additionally, the BIA attempts to identify the impact from these scenarios.

When evaluating the impact, a BIA looks at multiple items. For example, it might attempt to answer the following questions related to any of the scenarios:

- Will a disaster result in loss of life? Is there a way to minimize the risk to personnel?
- Will a disaster result in loss of property?

- Will a disaster reduce safety for personnel or property?
- What are the potential financial losses to the organization?
- What are the potential losses to the organization's reputation?

For example, a database server might host customer data, including credit card information. If an attacker was able to access this customer data, the cost to the organization might exceed millions of dollars.

You might remember the attack on retail giant Target during November and December 2013. Attackers accessed customer data on more than 110 million customers, resulting in significant losses for Target. Estimates of the total cost of the incident have ranged from $600 million to over $1 billion. This includes loss of sales—Target suffered a 46 percent drop in profits during the last quarter of 2013, compared with the previous year. Customers were afraid to use their credit cards at Target and simply stayed away. It also includes the cost to repair their image, the cost of purchasing credit monitoring for affected customers, fines from the payment-card industry, and an untold number of lawsuits. Target reportedly has $100 million in cyber insurance that helped them pay claims related to the data breach.

Remember this

The BIA identifies mission-essential functions and critical systems that are essential to the organization's success. It also identifies maximum downtime limits for these systems and components, various scenarios that can impact these systems and components, and the potential losses from an incident.

Privacy Impact and Threshold Assessments

Two tools that organizations can use when completing a BIA are a privacy threshold assessment and a privacy impact assessment. National Institute of Standards and Technology (NIST) Special Publication (SP) 800-122, "Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)," covers these in more depth, but refers to a privacy threshold assessment as a privacy threshold analysis.

The primary purpose of the ***privacy threshold assessment*** is to help the organization identify PII within a system. Typically, the threshold assessment is completed by the system owner or data owner by answering a simple questionnaire.

If the system holds PII, then the next step is to conduct a ***privacy impact assessment***. The impact assessment attempts to identify potential risks related to the PII by reviewing how the information is handled. The goal is to ensure that the system is complying with applicable laws, regulations, and guidelines. The impact assessment provides a proactive method of addressing potential risks related to PII throughout the life cycle of a computing system.

Remember this

A privacy threshold assessment is typically a simple questionnaire completed by system or data owners. It helps identify if a system processes data that exceeds the threshold for PII. If the system processes PII, a privacy impact assessment helps identify and reduce risks related to potential loss of the PII.

Recovery Time Objective

The recovery time objective (***RTO***) identifies the maximum amount of time it can take to restore a system after an outage. Many BIAs identify the maximum acceptable outage or maximum tolerable outage time for mission-essential functions and critical systems. If an outage lasts longer than this maximum time, the impact is unacceptable to the organization.

For example, imagine an organization that sells products via a web site generates $10,000 in revenue an hour. It might decide that the maximum acceptable outage for the web server is five minutes. This results in an RTO of five minutes, indicating any outage must be limited to less than five minutes. This RTO of five minutes only applies to the mission-essential function of online sales and the critical systems supporting it.

Imagine that the organization has a database server only used by internal employees, not online sales. Although the database server may be valuable, it is not critical. Management might decide they can accept an outage for as long as 24 hours, resulting in an RTO of less than 24 hours.

Recovery Point Objective

A recovery point objective (***RPO***) identifies a point in time where data loss is acceptable. As an example, a server may host archived data that has very few changes on a weekly basis. Management might decide that some

data loss is acceptable, but they always want to be able to recover data from at least the previous week. In this case, the RPO is one week.

With an RPO of one week, administrators would ensure that they have at least weekly backups. In the event of a failure, they will be able to restore recent backups and meet the RPO.

In some cases, the RPO is up to the minute of the failure. For example, any data loss from an online database recording customer transactions might be unacceptable. In this case, the organization can use a variety of techniques to ensure administrators can restore data up to the moment of failure.

Remember this

The recovery time objective (RTO) identifies the maximum amount of time it should take to restore a system after an outage. It is derived from the maximum allowable outage time identified in the BIA. The recovery point objective (RPO) refers to the amount of data you can afford to lose.

Comparing MTBF and MTTR

When working with a BIA, experts often attempt to predict the possibility of a failure. For example, what is the likelihood that a hard disk within a RAID configuration will fail? The following two terms are often used to predict potential failures:

- **Mean time between failures (MTBF).** The mean time between failures (***MTBF***) provides a measure of a system's reliability and is usually represented in hours. More specifically, the MTBF identifies the average (the arithmetic mean) time between failures. Higher MTBF numbers indicate a higher reliability of a product or system. Administrators and security experts attempt to identify the MTBF for critical systems with a goal of predicting potential outages.
- **Mean time to recover (MTTR).** The mean time to recover (***MTTR***) identifies the average (the arithmetic mean) time it takes to restore a failed system. In some cases, people interpret MTTR as the mean time to repair, and both mean essentially the same thing. Organizations that have maintenance contracts often specify the MTTR as a part of the contract. The supplier agrees that it will, on average, restore a failed system within the MTTR time. The MTTR

does not provide a guarantee that it will restore the system within the MTTR every time. Sometimes, it might take a little longer and sometimes it might be a little quicker, with the average defined by the MTTR.

Continuity of Operations Planning

Continuity of operations planning focuses on restoring mission-essential functions at a recovery site after a critical outage. For example, if a hurricane or other disaster prevents the company from operating in the primary location, the organization can continue to operate the mission-essential functions at an alternate location that management previously identified as a recovery site. Failover is the process of moving mission-essential functions to the alternate site.

Recovery Sites

A ***recovery site*** is an alternate processing site that an organization can use after a disaster. The three primary types of recovery sites are hot sites, cold sites, and warm sites. These alternate locations could be office space within a building, an entire building, or even a group of buildings. Two other types of recovery sites are mobile sites and mirrored sites. The following sections provide more details on these sites.

Hot Site

A ***hot site*** would be up and operational 24 hours a day, seven days a week and would be able to take over functionality from the primary site quickly after a primary site failure. It would include all the equipment, software, and communication capabilities of the primary site, and all the data would be up to date. In many cases, copies of backup tapes are stored at the hot site as the off-site location.

In many cases, a hot site is another active business location that has the capability to assume operations during a disaster. For example, a financial institution could have locations in two separate cities. The second location provides noncritical support services, but also includes all the resources necessary to assume the functions of the first location.

Some definitions of hot sites indicate they can take over

instantaneously, though this isn't consistent. In most cases, it takes a little bit of time to transfer operations to the hot site, and this can take anywhere from a few minutes to an hour.

Clearly, a hot site is the most effective disaster recovery solution for high-availability requirements. If an organization must keep critical systems with high-availability requirements, the hot site is the best choice. However, a hot site is the most expensive to maintain and keep up to date.

Remember this

A hot site includes personnel, equipment, software, and communication capabilities of the primary site with all the data up to date. A hot site provides the shortest recovery time compared with warm and cold sites. It is the most effective disaster recovery solution, but it is also the most expensive to maintain.

Cold Site

A ***cold site*** requires power and connectivity but not much else. Generally, if it has a roof, electricity, running water, and Internet access, you're good to go. The organization brings all the equipment, software, and data to the site when it activates it.

I often take my dogs for a walk at a local army base and occasionally see soldiers activate an extreme example of a cold site. On most weekends, the fields are empty. Other weekends, soldiers have transformed one or more fields into complete operational sites with tents, antennas, cables, generators, and porta-potties.

Because the army has several buildings on the base, they don't need to operate in the middle of fields, but what they're really doing is testing their ability to stand up a cold site wherever they want. If they can do it in the field, they can do it in the middle of a desert, or anywhere else they need to.

A cold site is the cheapest to maintain, but it is also the most difficult to test.

Warm Site

You can think of a ***warm site*** as the Goldilocks solution—not too hot and not too cold, but just right. Hot sites are generally too expensive for most organizations, and cold sites sometimes take too long to configure for

full operation. However, the warm site provides a compromise that an organization can tailor to meet its needs.

For example, an organization can place all the necessary hardware at the warm site location but not include up-to-date data. If a disaster occurs, the organization can copy the data to the warm site and take over operations. This is only one example, but there are many different possibilities of warm site configurations.

Site Variations

Although hot, cold, and warm sites are the most common, you might also come across two additional alternate site types: mobile and mirrored.

Amobilesiteisaself-containedtransportableunitwithalltheequipmentneededforspecific requirements. For example, you can outfit a semitrailer with everything needed for operations, including a satellite dish for connectivity. Trucks, trains, or ships haul it to its destination and it only needs power to start operating.

Mirrored sites are identical to the primary location and provide 100 percent availability. They use real-time transfers to send modifications from the primary location to the mirrored site. Although a hot site can be up and operational within an hour, the mirrored site is always up and operational.

Order of Restoration

After the disaster has passed, you will want to return all the functions to the primary site. As a best practice, organizations return the least critical functions to the primary site first. Remember, the critical functions are operational at the alternate site and can stay there as long as necessary. If a site has just gone through a disaster, it's very likely that there are still some unknown problems. By moving the least critical functions first, undiscovered problems will appear and can be resolved without significantly affecting mission-essential functions.

Remember this

A cold site will have power and connectivity needed for a recovery site, but little else. Cold sites are the least expensive and the hardest to test. A warm site is a compromise between a hot site and a

cold site. Mobile sites do not have dedicated locations, but can provide temporary support during a disaster.

Disaster Recovery

Disaster recovery is a part of an overall business continuity plan. Often, the organization will use the business impact analysis to identify the critical systems and components and then develop disaster recovery strategies and disaster recovery plans (DRPs) to address the systems hosting these functions.

In some cases, an organization will have multiple DRPs within a BCP, and in other cases, the organization will have a single DRP. For example, it's possible to have individual DRPs that identify the steps to recover individual critical servers and other DRPs that detail the recovery steps after different types of disasters such as hurricanes or tornadoes. A smaller organization might have a single DRP that simply identifies all the steps used to respond to any disruption.

A DRP or a BCP will include a hierarchical list of critical systems. This list identifies what systems to restore after a disaster and in what order. For example, should a server hosting an online web site be restored first, or a server hosting an internal application? The answer is dependent on how the organization values and uses these servers. In some cases, systems have interdependencies requiring systems to be restored in a certain order.

If the DRP doesn't prioritize the systems, individuals restoring the systems will use their own judgment, which might not meet the overall needs of the organization. For example, Nicky New Guy might not realize that a web server is generating $5,000 an hour in revenue but does know that he's responsible for keeping a generic file server operational. Without an ordered list of critical systems, he might spend his time restoring the file server and not the web server.

This hierarchical list is valuable when using alternate sites such as warm or cold sites, too. When the organization needs to move operations to an alternate site, the organization will want the most important systems and functions restored first.

Similarly, the DRP often prioritizes the services to restore after an outage. As a rule, critical business functions and security services are restored first. Support services are restored last.

The different phases of a disaster recovery process typically include the following steps:

- **Activate the disaster recovery plan.** Some disasters, such as earthquakes or tornadoes, occur without much warning, and a disaster recovery plan is activated after the disaster. Other disasters, such as hurricanes, provide a warning, and the plan is activated when the disaster is imminent.
- **Implement contingencies.** If the recovery plan requires implementation of an alternate site, critical functions are moved to these sites. If the disaster destroyed on-site backups, this step retrieves the off-site backups from the off-site location.
- **Recover critical systems.** After the disaster has passed, the organization begins recovering critical systems. The DRP documents which systems to recover and includes detailed steps on how to recover them. This also includes reviewing change management documentation to ensure that recovered systems include approved changes.
- **Test recovered systems.** Before bringing systems online, administrators test and verify them. This may include comparing the restored system with a performance baseline to verify functionality.
- **After-action report.** The final phase of disaster recovery includes a review of the disaster, sometimes called an after-action review. This often includes a lessons learned review to identify what went right and what went wrong. After reviewing the after-action report, the organization often updates the plan to incorporate any lessons learned.

Remember this

A disaster recovery plan (DRP) includes a hierarchical list of critical systems and often prioritizes services to restore after an outage. Testing validates the plan. The final phase of disaster recovery includes a review to identify any lessons learned and may include an update of the plan.

Testing Plans with Exercises

Business continuity plans and disaster recovery plans include testing.

Testing validates that the plan works as desired and will often include testing redundancies and backups. There are several different types of testing used with BCPs and DRPs.

NIST SP 800-34, "Guide to Test, Training, and Exercise Programs for IT Plans and Capabilities," provides detailed guidance on testing BCP and DRP plans. SP 800-34 identifies two primary types of exercises: tabletop exercises and functional exercises.

A ***tabletop exercise*** (also called a desktop exercise or a structured walk-through) is discussion-based. A coordinator gathers participants in a classroom or conference room, and leads them through one or more scenarios. As the coordinator introduces each stage of an incident, the participants identify what they'll do based on the plan. This generates discussion about team members' roles and responsibilities and the decision-making process during an incident. Ideally, this validates that the plan is valid. However, it sometimes reveals flaws. The BCP coordinator ensures the plans are rewritten if necessary.

Functional exercises provide personnel with an opportunity to test the plans in a simulated operational environment. There is a wide range of functional exercises, from simple simulations to full-blown tests. In a simulation, the participants go through the steps in a controlled manner without affecting the actual system. For example, a simulation can start by indicating that a server failed. Participants then follow the steps to rebuild the server on a test system. A full- blown test goes through all the steps of the plan. In addition to verifying that the test works, this also shows the amount of time it will take to execute the plan.

Some of the common elements of testing include:

- **Backups.** Backups are tested by **restoring** the data from the backup, as discussed in the "Testing Backups" section earlier in this chapter.
- **Server restoration.** A simple disaster recovery exercise rebuilds a server. Participants follow the steps to **rebuild** a server using a test system without touching the live system.
- **Server redundancy.** If a server is within a failover cluster, you can test the cluster by taking a primary node offline. **Another** node within the cluster should automatically assume the role of this offline node.
- **Alternate sites.** You can test an alternate site (hot, cold, or warm)

by moving some of the functionality to the alternate site and ensuring the alternate site works as desired. It's also possible to test individual elements of an alternate site, such as Internet connectivity, or the ability to obtain and restore backup media.

Remember this

You can validate business continuity plans through testing. Tabletop exercises are discussion-based only and are typically performed in a classroom or conference setting. Functional exercises are hands-on exercises.

Chapter 9 Exam Topic Review

When preparing for the exam, make sure you understand these key concepts covered in this chapter.

Implementing Defense in Depth

- Layered security (or defense in depth) employs multiple layers of security to protect against threats. Personnel constantly monitor, update, add to, and improve existing security controls.
- Control diversity is the use of different security control types, such as technical controls, administrative controls, and physical controls.
- Vendor diversity is the practice of implementing security controls from different vendors to increase security.

Comparing Physical Security Controls

- Physical security controls are controls you can physically touch. They often control entry and exit points, and include various types of locks.
- An airgap is a physical security control that ensures that a computer or network is physically isolated from another computer or network.
- Controlled areas such as data centers and server rooms should only have a single entrance and exit point. Door lock types include cipher locks, proximity cards, and biometrics.
- A proximity card can electronically unlock a door and helps prevent unauthorized personnel from entering a secure area. By themselves, proximity cards do not identify and authenticate users. Some systems combine proximity cards with PINs for identification and authentication.
- Tailgating occurs when one user follows closely behind another user without using credentials. A mantrap can prevent tailgating.
- Security guards are a preventive physical security control and they can prevent unauthorized personnel from entering a secure area.

A benefit of guards is that they can recognize people and compare an individual's picture ID for people they don't recognize.

- Cameras and closed-circuit television (CCTV) systems provide video surveillance. They provide reliable proof of a person's identity and activity.
- Fencing, lighting, and alarms are commonly implemented with motion detection systems for physical security. Infrared motion detection systems detect human activity based on the temperature.
- Barricades provide stronger physical security than fences and attempt to deter attackers. Bollards are effective barricades that allow people through, but block vehicles.
- Cable locks secure mobile computers such as laptop computers in a training lab. Server bays include locking cabinets or enclosures within a server room. Small devices can be stored in safes or locking office cabinets to prevent the theft of unused resources.
- Asset management processes protect against vulnerabilities related to architecture and design weaknesses, system sprawl, and undocumented assets.
- Heating, ventilation, and air conditioning (HVAC) systems control airflow for data centers and server rooms. Temperature controls protect systems from damage due to overheating.
- Hot and cold aisles provide more efficient cooling of systems within a data center.
- EMI shielding prevents problems from EMI sources such as fluorescent lighting fixtures. It also prevents data loss in twisted-pair cables. A Faraday cage prevents signals from emanating beyond a room or enclosure.

Adding Redundancy and Fault Tolerance

- A single point of failure is any component that can cause the entire system to fail if it fails.
- RAID disk subsystems provide fault tolerance and increase availability. RAID-1 (mirroring) uses two disks. RAID-5 uses three or more disks and can survive the failure of one disk. RAID-6 and RAID-10 use four or more disks and can survive the failure of two

disks.

- Load balancers spread the processing load over multiple servers. In an
- active- active configuration, all servers are actively processing requests. In an active-passive configuration, at least one server is not active, but is instead monitoring activity ready to take over for a failed server. Software-based load balancers use a virtual IP.
- Affinity scheduling sends client requests to the same server based on the client's IP address. This is useful when clients need to access the same server for an entire online session. Round-robin scheduling sends requests to servers using a predefined order.

Protecting Data with Backups

- Backup strategies include full, full/differential, full/incremental, and snapshot strategies. A full backup strategy alone allows the quickest recovery time.
- Full/incremental backup strategies minimize the amount of time needed to perform daily backups.
- Test restores verify the integrity of backups. A test restore of a full backup verifies a backup can be restored in its entirety.
- Backups should be labeled to identify the contents. A copy of backups should be kept off-site.
- It's important to consider the distance between the main site and the off-site location.
- The data contained in the backups can have legal implications. If it includes Personally Identifiable Information (PII) or Protected Health Information (PHI), it must be protected according to governing laws.
- The location of the data backups affects the data sovereignty. If backups are stored in a different country, the data on the backups is now subject to the laws and regulations of that country.

Comparing Business Continuity Elements

- A business impact analysis (BIA) is part of a business continuity plan (BCP) and it identifies mission-essential functions, critical

systems, and vulnerable business processes that are essential to the organization's success.

- The BIA identifies maximum downtimes for these systems and components. It considers various scenarios that can affect these systems and components, and the impact to life, property, safety, finance, and reputation from an incident.
- A privacy threshold assessment identifies if a system processes data that exceeds the threshold for PII. If the system processes PII, a privacy impact assessment helps identify and reduce risks related to potential loss of the PII.
- A recovery time objective (RTO) identifies the maximum amount of time it should take to restore a system after an outage. The recovery point objective (RPO) refers to the amount of data you can afford to lose.
- Mean time between failures (MTBF) identifies the average (the arithmetic mean) time between failures. The mean time to recover (MTTR) identifies the average (the arithmetic mean) time it takes to restore a failed system.
- Continuity of operations planning identifies alternate processing sites and alternate business practices. Recovery sites provide alternate locations for business functions after a major disaster.
- A hot site includes everything needed to be operational within 60 minutes. It is the most effective recovery solution and the most expensive. A cold site has power and connectivity requirements and little else. It is the least expensive to maintain. Warm sites are a compromise between hot sites and cold sites.
- Periodic testing validates continuity of operations plans. Exercises validate the steps to restore individual systems, activate alternate sites, and document other actions within a plan. Tabletop exercises are discussion-based only. Functional exercises are hands-on exercises.

Online References

- Do you know how to answer performance-based questions? Check out the online extras at *http://gcgapremium.com/501-extras.*

Chapter 9 Practice Questions

1. After a recent attack on your organization's network, the CTO is insisting that the DMZ uses two firewalls and they are purchased from different companies. Which of the following BEST describes this practice?

   A. Single-layer security
   B. Vendor diversity
   C. Control diversity
   D. Redundancy

2. Management within your organization wants to create a small network used by executives only. They want to ensure that this network is completely isolated from the main network. Which of the following choices BEST meets this need?

   A. Airgap
   B. Mantrap
   C. Control diversity
   D. Infrared motion detectors

3. A security professional has reported an increase in the number of tailgating violations into a secure data center. Which of the following can prevent this?

   A. CCTV
   B. Mantrap
   C. Proximity card
   D. Cipher lock

4. Lisa is the new chief technology officer (CTO) at your organization. She wants to ensure that critical business systems are protected from isolated outages. Which of the following would let her know how often these systems will experience outages?

   A. MTTR
   B. MTBF
   C. RTO
   D. RPO

5. Thieves recently rammed a truck through the entrance of your company's main building. During the chaos, their partners proceeded to

steal a significant amount of IT equipment. Which of the following choices can you use to prevent this from happening again?

A. Bollards
B. Guards
C. CCTV
D. Mantrap

6. You are a technician at a small organization. You need to add fault-tolerance capabilities within the business to increase the availability of data. However, you need to keep costs as low as possible. Which of the following is the BEST choice to meet these needs?

A. Alternate processing site
B. RAID-10
C. Backups
D. Faraday cage

7. Flancrest Enterprises recently set up a web site utilizing several web servers in a web farm. The web farm spreads the load among the different web servers. Visitor IP addresses are used to ensure that clients always return to the same server during a web session. Which of the following BEST describes this configuration?

A. Affinity
B. Round-robin
C. Virtual IP
D. Active-passive

8. Your organization is planning to deploy a new e-commerce web site. Management anticipates heavy processing requirements for a back-end application. The current design will use one web server and multiple application servers. Which of the following BEST describes the application servers?

A. Load balancing
B. Clustering
C. RAID
D. Affinity scheduling

9. Flancrest Enterprises recently set up a web site utilizing several web servers in a web farm. The web farm spreads the load among the different web servers by sending the first request to one server, the next request to the second server, and so on. Which of the following BEST

describes this configuration?

A. Affinity
B. Round-robin
C. Airgap
D. Mantrap

10. Flancrest Enterprises recently set up a web site utilizing several web servers in a web farm. The web servers access a back-end database. The database is hosted by a database application configured on two database servers. Web servers can access either of the database servers. Which of the following BEST describes the configuration of the database servers?

A. Active-passive
B. Round-robin
C. Affinity
D. Active-active

11. Your organization has decided to increase the amount of customer data it maintains and use it for targeted sales. However, management is concerned that they will need to comply with existing laws related to PII. Which of the following should be completed to determine if the customer data is PII?

A. Privacy threshold assessment
B. Privacy impact assessment
C. Tabletop exercise
D. Affinity scheduling

12. Your backup policy for a database server dictates that the amount of time needed to perform backups should be minimized. Which of the following backup plans would BEST meet this need?

A. Full backups on Sunday and full backups on the other six days of the week
B. Full backups on Sunday and differential backups on the other six days of the week
C. Full backups on Sunday and incremental backups on the other six days of the week
D. Differential backups on Sunday and incremental backups on the other six days of the week

13. You are helping implement your company's business continuity plan. For one system, the plan requires an RTO of five hours and an RPO

of one day. Which of the following would meet this requirement?

A. Ensure the system can be restored within five hours and ensure it does not lose more than one day of data.

B. Ensure the system can be restored within one day and ensure it does not lose more than five hours of data.

C. Ensure the system can be restored between five hours and one day after an outage.

D. Ensure critical systems can be restored within five hours and noncritical systems can be restored within one day.

14. A security analyst is creating a document that includes the expected monetary loss from a major outage. She is calculating the potential impact on life, property, finances, and the organization's reputation. Which of the following documents is she MOST likely creating?

A. BCP

B. BIA

C. MTBF

D. RPO

15. A security expert at your organization is leading an on-site meeting with key disaster recovery personnel. The purpose of the meeting is to perform a test. Which of the following BEST describes this test?

A. Functional exercise

B. Full-blown test

C. Tabletop exercise

D. Simulation to perform steps of a plan

Chapter 9 Practice Question Answers

1. **B.** The chief technology officer (CTO) is recommending vendor diversity for the demilitarized zone (DMZ). Firewalls from different companies (vendors) provide vendor diversity. This also provides defense in depth or layered security, but not single-layer security. Control diversity is the use of different controls such as technical, administrative, and physical. Redundancy is the use of duplicate components for fault tolerance, but the two firewalls work together in the DMZ.

2. **A.** An airgap ensures that a computer or network is physically isolated from another computer or network. A mantrap helps prevent unauthorized entry and is useful for preventing tailgating. Control diversity is the use of different controls such as technical, administrative, and physical, but it doesn't necessarily isolate networks. Infrared motion detectors sense motion from infrared light, but they don't isolate networks.

3. **B.** A mantrap is highly effective at preventing unauthorized entry and can also be used to prevent tailgating. CCTV uses cameras for video surveillance and it can record unauthorized entry, but it can't prevent it. A proximity card is useful as an access control mechanism, but it won't prevent tailgating, so it isn't as useful as a mantrap. A cipher lock is a door access control, but it can't prevent tailgating.

4. **B.** The mean time between failures (MTBF) provides a measure of a system's reliability and would provide an estimate of how often the systems will experience outages. The mean time to recover (MTTR) refers to the time it takes to restore a system, not the time between failures. The recovery time objective (RTO) identifies the maximum amount of time it can take to restore a system after an outage. The recovery point objective (RPO) identifies a point in time where data loss is acceptable.

5. **A.** Bollards are effective barricades that can block vehicles. Guards can restrict access for personnel, but they cannot stop trucks from ramming through a building. Closed-circuit television (CCTV) or a

similar video surveillance system can monitor the entrance, but it won't stop the attack. Mantraps prevent tailgating, but they most likely won't stop a truck.

6. **B.** A redundant array of inexpensive disks 10 (RAID-10) subsystem provides fault tolerance for disks and increases data availability. An alternate processing site might be used for a mission-essential function, but it is expensive and does much more than increase the availability of data. Backups help ensure data availability, but they do not help with fault tolerance. A Faraday cage is a room or enclosure that prevents signals from emanating beyond the room.

7. **A.** Source address IP affinity scheduling allows a load balancer to direct client requests to the same server during a web session. Round-robin scheduling simply sends each request to the next server. Load balancers can use a virtual IP, but this refers to the IP address of the web server, not the IP address of a visitor. An active-passive configuration has at least one server that is not actively serving clients, but the scenario doesn't indicate any of the servers are in a passive mode.

8. **A.** The design is using load balancing to spread the load across multiple application servers. The scenario indicates the goal is to use multiple servers because of heavy processing requirements, and this is exactly what load balancing does. Clustering is typically used to provide high availability by failing over to another server if one server fails. RAID provides fault tolerance for disk drives, not servers. Affinity scheduling helps ensure clients go to the same server during a session, but this isn't relevant to this scenario.

9. **B.** A round-robin scheduling scheme allows a load balancer to send requests to servers one after another. Affinity scheduling directs user requests to a specific server based on the user's IP address to ensure that the user accesses the same server during a web session. An airgap ensures that computing systems are physically separated from each other and is unrelated to this question. A mantrap prevents unauthorized entry using the social engineering tactic of tailgating.

10. **D.** The database servers are in an active-active load-balancing configuration because web servers can query both database servers. In an active-passive configuration, only one of the database servers would be

answering queries at any given time. Round-robin and affinity are two methods of scheduling the load balancing in an active-active configuration.

11. **A.** A privacy threshold assessment helps an organization identify Personally Identifiable Information (PII) within a system, and in this scenario, it would help the organization determine if the customer data is PII. A privacy impact assessment is done after you have verified that the system is processing PII, not to determine if the data is PII. A tabletop exercise is a discussion-based exercise used to talk through a continuity of operations plan. Affinity scheduling is a load-balancing scheduling scheme using the client's IP address and is unrelated to PII.

12. **C.** A full/incremental backup strategy is the best option with one full backup on one day and incremental backups on the other days. The incremental backups will take a relatively short time compared with the other methods. A full backup every day would require the most time every day. Differential backups become steadily larger as the week progresses and take more time to back up than incremental backups. Backups must start with a full backup, so a differential/incremental backup strategy is not possible.

13. **A.** The recovery time objective (RTO) identifies the maximum amount of time it should take to restore a system after an outage. The recovery point objective (RPO) refers to the amount of data you can afford to lose. RTO only refers to time, not data. RPO refers to data recovery points, not time to restore a system.

14. **B.** A business impact analysis (BIA) includes information on potential monetary losses along with the impact on life, property, and the organization's reputation. It is the most likely document of those listed that would include this information. A business continuity plan (BCP) includes a BIA, but the BIA is more likely to include this information than the BCP is. The mean time between failures (MTBF) provides a measure of a system's reliability. The recovery point objective (RPO) refers to the amount of data you can afford to lose, but it does not include monetary losses.

15. **C.** A tabletop exercise is discussion-based and is typically performed in a classroom or conference room setting. Because this is a

meeting that includes disaster recovery personnel, it is a tabletop exercise. Functional exercises are hands-on exercises and include simulations and full-blown tests.