**Lecture: Account & Identity Management**

**Content Author:** Dr. Michael J. Simko

A picture containing text, table, shop, worktable

Description automatically generatedThis lecture will dive into the account management lifecycle and the concepts of Authentication, Authorization and Accounting (AAA). Several stages of creating, managing and terminating accounts are used to provide effective means of ensuring AAA. As a defensive cybersecurity architect, you should not only be aware of the technical aspect of creating and managing an account, but also the importance of each stage, often taken for granted throughout the process. Options for account management have continued to evolve over time.

**Organizational Account Management**

Every organization has employees which typically require access to company owned assets, data and internal network. User accounts provide basic computer functionality while administrative accounts provide the ability to manage user and system accounts. With technology constantly evolving, it is safe to assume that nearly every modern day organization functions with the use of technology and internet access. No matter what position, employees generally require some type of computer access, whether it is to perform their job duties or clock their time. There are several best practices that organizations should implement to effectively manage the account lifecycle from start to end. Effectively managing company owned accounts will reduce insider threat risk and disgruntled employee retaliation to a minimum level.

**Account Provisioning**

Account Provisioning involves configuring and maintaining access to specific technology devices. Rouse (2010) shared several examples of account provisioning to include:

* Discretionary Account Provisioning - The administrator determines what access the user should have.
* Self-Service Account Provisioning - Users assist in provisioning to lighten the administrator's overhead.
* Workflow-Based Account Provisioning - Designated approvers must validate the requirement for access, prior to granting access.
* Automated Account Provisioning - All accounts are added uniformly, utilizing a centralized management system or application.

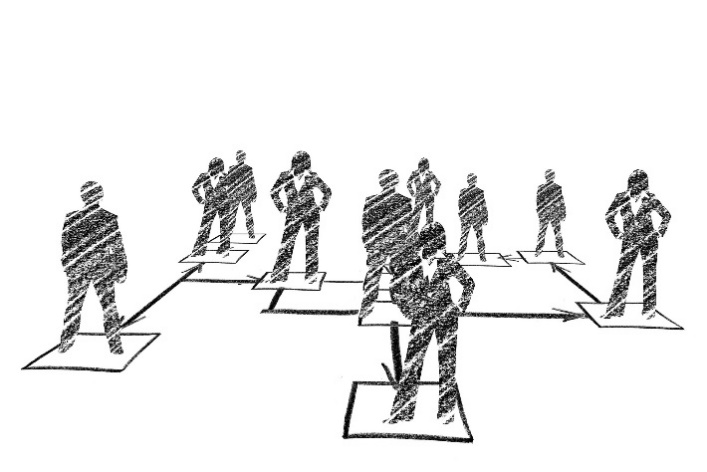
The majority of organizations utilize a Workflow-based account provisioning process. Account requests are initiated through a specific section within the Information Technology Department, while designated approvers (typically involving the employee's supervisor, the information owner and a cybersecurity specialist) validate the requirement of access and read, write or execute capabilities.

A picture containing text

Description automatically generated**Account Request documentation (logging)**

Written documentation or specific paperwork that outlines procedures and prerequisites for requesting an account must be implemented. This paperwork should also include a user access agreement to ensure the employee understands and acknowledges company technology usage policy. The majority of organizations utilize such paperwork for new employees to initiate new account requests or modifications to current accounts. As an administrator, you must emphasize due care in ensuring the account requester does indeed possess the requirement to obtain the account, as well as a need to know. It is best practice to maintain either hard or soft copy of account requests for at least up to three years.

**Account Creation & Staging**

Creating an account requires several steps as briefly discussed in account provisioning. Establishing organizational units (OUs) or a folder hierarchy to manage accounts is a necessity. For example, a staging folder should be established to store new accounts in until appropriate access is determined. Different folders typically have different rights attached to them and various folders may help establish role based access. Role based access is specific to the position requirements of an individual. For example, a budget employee within the organization will require different rights to software applications than a human resources employee. Role based access can differentiate and manage several users under a set of permissions pertaining to a specific folder or OU.

To summarize, account staging is an effective method to manage new user accounts, prior to moving to the appropriate organizational unit. Role based access control will reduce the amount of time required to configure individual permissions, ultimately reducing administrator overhead.

**Account management**

Logo

Description automatically generated with medium confidenceLogo

Description automatically generated with medium confidenceAdministrators should configure both security requirements and account policies for each account. Managing and reviewing accounts frequently is required to determine if account permissions must be modified over time and also to prevent privilege creep. Privilege creep is when an employee begins a position with specific privileges or permissions and is promoted or hired into another position within the organization (without permissions adjusted appropriately). As the employee switches positions, additional permissions are granted without revoking non-necessary privileges, resulting in stackable permissions or privilege creep. Consistently reviewing employee privilege requirements and role based access will reduce the risk of privilege creep.

Account policies ensure best defensive security practice. Configuring the requirement for specific account password requirements such as length, complexity, maximum attempts prior to account lockout and password change requirements will help lower the risk of password cracking or guessing. The administrator should ensure account logging is enabling to detect any fraudulent activity. Mandatory vacations force the employee to take time off from work, so administrators can review account logs in the event of any suspicion. Job rotation moves employees to different roles to detect any differences in job function. This ensures one person is not solely relied upon for critical functions resulting in a single point of failure.

**Account termination/deletion**

As employees move on to new opportunities or are terminated; revoking account access immediately is a best practice. The organization should already have an exit interview process established with several required steps for employees leaving the organization. Account deactivation should be part of this process. Upon deactivating an account, the account should be staged in a folder or OU specifically for transitioned employees. This will help the organization manage which accounts have been deactivated, which employees have left the organization and provide the ability to maintain any account specific information required based off the retention policy of the organization. It is common for the disabled accounts to be deleted after a certain period of time or as indicated by organization's retention policy.

**Single Sign-On (SSO)**

Graphical user interface, diagram

Description automatically generatedYou should be familiar with the term single sign-on. Many organizations are utilizing SSO to simplify account access for multiple platforms or services. For example, at Walsh you use your credentials to access multiple platforms to include but not limited to the virtual lab environment, Email and Moodle.

**Identify as a Service (IDaaS)**

It is not uncommon these days for organizations to outsource identity and account management to a third party. IDaaS offers a solution in which infrastructure, software and accounts are management typically on a cloud by the third party. This service alleviates a lot of stress and labor of the account management lifecycle for the organization, as the third party assumes the responsibility.

**Tying it all together:**

AAA ensures employees are validated to prove their identity is accurate, authenticated to ensure their credentials (username and password) are correct and granted access dependent upon what level of access is required. As we learned, the account provisioning ensures each employee possess appropriate access. The creation and staging of the account establishes credentials used for authentication. The permissions configured by the administrative are used for authorization to specific functions or data. Lastly, the accounting phase ensures all actions from the identification phase to the authorization phase have been maintained in an audit trail or log.

Do you know these terms? If not, reference above.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| AAA | Discretionary Account Provisioning | Self-Service Account Provisioning | Workflow-Based Account Provisioning | Automated Account Provisioning |
| Logging | OUs | Privilege creep | SSO | IDaaS |

Several tools are available for account management both within specific operating systems and as additional services or tools. While Authentication, Authorization and Accounting (AAA) services and tools may be implemented through means of servers, services or network protocols, there are many options available for implementing AAA. Below, I describe a few commonly implemented account management tools and AAA solutions.

|  |  |
| --- | --- |
| Logo, company name  Description automatically generated | **Microsoft Windows Active Directory**  Active Directory (AD) is a directory service, commonly used on Microsoft Windows based platforms. A Microsoft Windows domain controller typically runs active directory services with the ability to authenticate and authorize users within the system or domain. AD also provides means for establishing various types of accounts, permissions and security policies. This is where you would not only create accounts, but also manage both accounts and system permissions for a company utilizing a Windows based infrastructure and configuration. |
| Logo  Description automatically generated | **Ubuntu Account Management**  Like Windows, Linux based operating systems offer their own account management features. Ubuntu has the option of configuring and managing accounts both through a graphical user interface (GUI) and command line. See this week's additional resources for a comparative analysis of adding and managing users in Ubuntu, through the GUI and command line. |
| A blue and white logo  Description automatically generated with low confidence | **Remote Authentication Dial-In User Service (RADIUS)**  RADIUS is a networking and AAA protocol for managing access on a network. RADIUS uses two functions known as access-request and accounting-request. A RADIUS server will receive access requests and determine whether they should be authorized or not. The RADIUS protocol furthermore provides authorization once authentication has been completed. There are free open source tools available such as FreeRADIUS. FreeRADIUS is an Internet Engineering Task Force (IETF) protocol for AAA which is capable of implementing RADIUS, the Dynamic Host Configuration Protocol (DHCP), Bidirectional Forwarding Detection (BFD) and the Address Resolution Protocol (ARP). This week's additional resources provides the FreeRADIUS website for downloads, tutorials and capabilities. |
| Diagram  Description automatically generated | **Diameter**  Diameter is RADIUS's replacement. Diameter provides additional capabilities that RADIUS does not, such as Stream Control Transmission Protocol (SCTP), acknowledgements at the application level layer and capability negation. Although Diameter does not possess encryption, it can be used in conjunction with IPSEC or TLS for transport level security. |
| Graphical user interface, application  Description automatically generated | **Terminal Access Controller Access-Control System Plus (TACACS+)**  TACACS+ is a replacement for its predecessor, TACACS. TACACS+ provides AAA services but unfortunately is not compatible with XTACACS or the original TACACS. TACACS+ is designed by Cisco, providing authorization for individual commands and the ability to encrypt all content within each packet of data. Network administrators commonly use TACACS+ (for authentication and authorization) to make switching and routing configurations or modifications on an enterprise network. |
| A picture containing text, clipart  Description automatically generated | **Kerberos**  Kerberos, developed by the Massachusetts Institute of Technology (MIT), uses symmetric cryptography while providing strong authentication for both server and client applications. Kerberos is available for free from MIT to help ensure communication confidentiality and integrity. More on the Kerberos Network Authentication Protocol is available within this week's additional resources. |

## How does Kerberos work?

According to Petters (2018), there are eight basic steps within the Kerberos authentication process:

1. A client requests an authentication ticket or TGT from the Key Distribution Center (KDC).
2. The KDC validates credentials, followed by sending an encrypted TGT and session key back.
3. The ticket granting service (TGS) secret key encrypts the TGT.
4. The client stores the TGT until it expires. When it expires the local session manager requests another TGT.

When a client requests access to a network resource or service, this happens:

1. The client sends the TGT to the TGS with the service principal name (SPN) of the resource or service the client wishes to access.
2. The KDC verifies the TGT of the specific user and whether or not they are able to access the service.
3. The TGS sends a valid session key for the service back to the client.
4. The client forwards the session key to the service to validate the user does indeed have access, and the service grants access.

# Lecture: Securing Servers

**Content Author:** Dr. Michael J. Simko

Servers provide the critical applications, services and software required for modern day organizations to function. Whether Microsoft Windows or Linux based, server security consists of:

* Specific configurations
* Security settings and
* A Standardized methodology to system hardening.

We harden servers to ensure our best defense against hackers finding an entry point into our networks, stealing proprietary or company owned data or sabotaging organizational services and the company's ability to generate revenue.

Shape

Description automatically generated with low confidence

This lecture will address common server hardening best practices, while this week's lab will provide you an opportunity to implement many of these concepts with a hands on approach.

So, how would you build servers using basic hardening techniques?

Icon

Description automatically generated

## Access Control: Physical and Technical Controls

The first aspect to take into consideration when securing servers is location. How will your infrastructure, servers and hardware be physically protected? Will the servers be located in server racks that are lockable? Will USB ports be disabled to prevent malware from being spread onto the system? Will the data center have sufficient access control (proximity/magnetic strip card readers for door access, biometrics and authorized access criteria/need to know, etc.)? Several factors must be considered to limit physical access to servers, in addition to technical or logical controls.

Logo, icon

Description automatically generated

## Group Policy Objects

It may be difficult to implement multiple security technical implementation guides or security settings on each and every server within your cluster or server farm. Utilizing and configuring group policy objects (GPOs) with a bulk of security settings can ease the overall process. You can apply a GPO to a container with multiple servers in it. All of the servers will be stored under the same set of rules or security requirements defined by the GPO.

Application

Description automatically generated with medium confidence

## Encryption

Encrypting server hard drives and data the server may process in transit and at rest is essential. This will make it more difficult to tamper with server hard drives in the event they are stolen. Encrypting sensitive data and files located on the systems will ensure data confidentiality and integrity.

Icon

Description automatically generated

## Antivirus/Anti-malware

Utilizing live system monitoring tools such as antivirus and anti-malware suites can quickly identify malware before it is too late. Many times, it is an organizations best defense due to these software suites being capable of not only identifying potential threats, but also isolating them to prevent further harm to systems within the enterprise. These software based solutions require frequent updates to signatures in order to adequately implement current and best defensive measures.

Text

Description automatically generated with medium confidence

## Patch Management and Firmware

Frequent patching and firmware updates help mitigate constant evolving vulnerabilities. Throughout the lifecycle of software, vulnerabilities are often identified. Sometimes fixing one vulnerability opens up an avenue for another vulnerabilities. Configuring Windows Server Update Services (WSUS) or a System Center Configuration Manager (SCCM) for updates is highly recommended. Manual rather than automated patching may not be as effective due to human error and possibility of missing fine details. Verifying the latest versions of firmware can also hinder potential vulnerabilities from being exploited.

Logo

Description automatically generated

## Audit Trails

System and event logging may not only serve as a detective control for identifying fraudulent activity, but also provides an ample amount of evidence pertaining to breach attempts on your servers. The security logs are available on each individual server, while a Syslog server can store logs for a wide range of devices. Logs should be reviewed on a continuous basis.

A picture containing chart

Description automatically generated

## Port Filtering

Check all ports and services that are enabled on your server by default. Ensure that unnecessary services and ports are not enabled. For example, you will commonly not require FTP ports 20/21 if you are running a server providing web browsing capabilities.

Icon

Description automatically generated

## Unnecessary Services

As stated with port filtering, negate any unnecessary services. The less services you have running that could potentially be used to breach your system, the better. Frequently audit required services to ensure you are staying on top of the evolving requirements of your organization.

Whiteboard

Description automatically generated

## Network Sharing

Review folders and files that are available or shared on the network. If your server is acting as a file share, ensure permissions (read, write and execute) are configured appropriately for each file and folder.

Icon

Description automatically generated

## Script Management

If you or any system administrators will be using scripts, you need to ensure the scripts are not performing actions that could create an avenue of attack without your knowledge. Customizing and writing your own scripts is okay, but ensure the scripts are truly meeting your expectations or desired function without exposing your server.

Icon

Description automatically generated

## Host based IDS/IPS

Similar to antivirus and anti-malware, implementing an automated and layered defense can strengthen your system even more. While antivirus and anti-malware commonly protects you from software based attacks; implementing host based intrusion detection/prevention systems can counter network based attack.

Logo, icon

Description automatically generated

## File Integrity

Running integrity checks on files will ensure your data has not been altered in any unauthorized way. It is important to audit file integrity frequently to identify any fraudulent activity or insider threat.

Icon

Description automatically generated

## IP Filtering

Limiting access to only essential hosts or clients requiring access to your server will lessen the likelihood of a person with malicious intent easily accessing the system. IP Filtering can be enforced to ensure only known and trusted IPs are able to access the server.

A picture containing icon

Description automatically generated

## Default Permissions/Accounts

When building servers, standard usernames and passwords are common. For example, in a Windows environment, the administrator account will use the username "administrator" and the password "password".

Icon

Description automatically generated

## Guest Accounts

When building servers, standard usernames and passwords are common. For example, in a Windows environment, the administrator account will use the username "administrator" and the associated password.

Logo

Description automatically generated

## Vulnerability Management

We talked about the importance of updating servers, but how do we know we are truly aware of all potential vulnerabilities pertaining to our server? Using a vulnerability scanner can identify additional vulnerabilities that may not show up as a required or optional update for the server. Frequently conducting vulnerability scans is a best practice to stay current on potential threats that arise pertaining to your server.

Logo

Description automatically generated

## Known Vulnerabilities and Advanced Persist Threats (APTs)

Staying informed of current vulnerabilities and threats successfully exploiting organizations globally is a must. If multiple organizations are being exploited by the same vulnerability or weakness; there may be a chance your servers, systems and organization can fall victim to the same threat. Staying current on these threats and investigating whether or not the attack has the possibility of harming your systems will keep you and your organization less likely to be victimized.

Logo

Description automatically generated

## Policy/Procedures

So, we have discussed all of these great methods for securing servers but what good are they if there is not a standardized baseline, policy or procedure for configuring future servers? Enforcing a set of standardized procedures that clearly articulates server security requirements will ensure the process is uniform across the organization. We would not want to waste valuable time hardening a handful of servers, if there are other unhardened servers that offer an avenue of attack or exploit.

In this lecture, we are going to address many variations of implementing security within Linux based servers.

In the previous lecture on server hardening, we addressed some basic best practices for establishing baseline security on all servers. The assumption that Linux servers are naturally more hardened than Windows servers is not accurate. All software is equally vulnerable to various threats overtime and specific configurations must be incorporated into system configuration in order to provide sufficient due care.

Let's talk about a few specific best practices for hardening Linux boxes!

## Locking Down Cronjobs

Cron is a utility in Linux that has the ability to determine who can and who cannot perform certain jobs.

**How to Determine Who Can and Who Cannot Perform Certain Jobs?**

|  |  |
| --- | --- |
| Manage what users you would like run jobs | under the /etc/cron.allow file |
| Determine who should not have the ability to run cron | through the /etc/cron.deny file |
| If you do not want any users to use cron | include All in the cron.deny file |

A person looking at a computer

Description automatically generated with low confidence

## Ensure Systems are Kept Updated

You should have caught on by now, it is imperative to keep systems updated. Keeping systems updated can prevent exploitation of known vulnerabilities. Configuring routine updates will ease the requirement of having to remember manually updating frequently.

## Audit Network Services

As with Windows servers, unnecessary open ports and services provide an avenue for attack. Auditing unneeded ports and services with the netstat command will provide you a list of ports that are currently open, while the chkconfig command will provide you the opportunity to disable any that may not be required for the system in its current state.

## Review Installed Packages and Services

Multiple installed services provide numerous vectors for attack. Disabling or removing any unneeded installed packages or services can minimize the field of opportunity for nefarious actors. The chkconfig command will also provide the ability to audit services currently active on runlevel 3. Runelevel 3, commonly used by Linux when booting up, is a Linux kernel level that supports multiple user mode through command line.

**What Commands 'to Type' to Audit Currently Active Services?**

|  |  |
| --- | --- |
| Find running services | /sbin/chkconfig -list |grep '3:on' |
| Turn off a service | checkconfig serviceName off |
| Remove a packet | sudo apt-get remove package-name |

## Physical Security of Systems

Disabling USB boot or USB Flash drives altogether is a best physical security practice for all systems. Doing so can prevent injection of malware onto a system and theft of data. In Linux, creating a file with /etc/modprobe.d/no-usb and adding the line install usb-storage /bin/true will prevent USB storage devices from being detected. It is also a good idea to disable the CTRL ALT DELETE reboot command. This will negate accidents from happening and hinder an insider threat from attempting to boot off any type of media.

## Secure Shell

Using protocols that include encryption during communication with a server is necessary. Older protocols such as Telnet and rlogin are not recommended for use.

It is recommended to:

* + Never login specifically as "root"
  + Utilize the sudo command for executing commands
  + Change the default port (22) to a higher level port. Doing so will make it more difficult for an attacker to accomplish malicious intent.
  + Implement a SSH Banner to users, prior to SSH authentication. This will provide a message to users with appropriate warnings and messages.

A person working on a computer

Description automatically generated with medium confidence

## SELinux

Security-Enhanced Linux is a built in security for Linux based systems. The status of security-enhanced Linux can be viewed using the command: system-config-selinux, getenforce or sestatus commands. To enable SELinux when it is disabled type: setenforce enforcing. There are three modes of SELinux operation.

**What are the 3 Modes of SELinux Operation?**

**Permissive** - SELinux uses warning and log actions without actually enabling the system security policy. Permissive mode can help troubleshoot SELinux issues.

**Enforcing** - The SELinux security is both enforced and enabled on a specific system.

**Disabled** - Turns off security-enhanced Linux.

## Multiple Disk Partitions

Separating data and files into multiple partitions is a great method to implement defense in depth. Having multiple partitions will negate the possibility of a single point of failure. It is best to install third party applications on separate file structures under /opt.

## Password Management

Password policy is necessary to prevent exploitation of weak passwords or password policies. You should audit your system for any accounts that have blank passwords and require a complex password. For password management:

* + Use the pam\_cracklib module (vt /etc/pam.d/system-auth) in pluggable authentication modules to Require strong passwords for users.
  + **Set expiration criteria** for passwords.

**What Commands 'to Type' Regarding Expiration Criteria?**

|  |  |
| --- | --- |
| Check a user's password age | chage -l username |
| Change min days | chage -m |
| Change max days | chage -M |
| Set a warning for said amount of days | chage -W |

* + To **Prevent users from using old passwords**, see the /etc/security/opasswd file with the PAM module. You may set how many previous passwords to allow/disallow. For example, you may disallow the previous three passwords when a user attempts to change their password.

## Account Management

Manual account management helps an administrator stay in control. There are a few basics commands for locking and unlocking accounts.

**What 'to Type' for Locking and Unlocking Accounts?**

|  |  |
| --- | --- |
| Lock an account | password -l accountName |
| Unlock an account | passwd -u accountName |

## User Activity Surveillance

Psacct & acct packages monitor activities on a Linux system. There are several options available for user monitoring.

**What Commands 'to Type' for Different Options of User Monitoring?**

|  |  |
| --- | --- |
| Print data regarding user logins and logouts | ac |
| View commands executed by a user | lastcomm |
| Turn accounting on or off | accton |
| Summarize previous commands that were executed | sa |
| Generate a list of recently logged in users | last and lastb |

## Audit Log Review

While many logs exist in Linux systems, the most common logs that should be reviewed regularly are listed below.

**What are Most Common Linux Logs?**

|  |  |
| --- | --- |
| /var/log/message | The log where current activity and system data is kept |
| /var/log/auth.log | Authentication logs |
| /var/log/kern.log | Kernel logs (provides several Linux related logs) |
| /var/log/yum.log | Yum logs (contains install logs of yum packages) |
| /var/log/utmp | Records of login |
| /var/log/cron.log | Cron job logs (contain who ran a command, what command was ran, and a date/time stamp of when it happened) |
| /var/log/secure | Authentication logs |
| /var/log/maillog | Mail server logs |
| /var/log/mysqld.log | MySQL Database logs |
| /var/log/boot.log | Boot logs |

**Chapter 11**

Secure Application Development,

Deployment, and Automation Concepts

In this chapter, you will

• Learn to implement secure application development

• Understand secure development concepts

• Explore the addition of security to automated/agile development processes

Software development is a complex process with many different issues, from

design, to coding, to testing and deployment, that need to be considered and

managed to achieve the desired goals of secure software. Developing and using a

security-inclusive application development process is essential. Expanding that

process to include delivery and post-deployment issues is essential, and one of

the key tools to achieve these objectives is the use of automation and scripting.

This chapter covers these issues with respect to the Security+ Version 6 exam

objectives.

Certification Objective This chapter covers CompTIA Security+ exam

objective 2.3: Summarize secure application development, deployment, and

automation concepts.

Environment

Most organizations have multiple, separate computing environments designed to

provide isolation between the functions of development, test, staging, and

production. The primary purpose of having these separate environments is to

prevent security incidents arising from untested code ending up in the production

environment. The hardware of these environments is segregated, and access

control lists are used to prevent users from accessing more than one environment

at a time. Moving code between environments requires a special account that can

access both, minimizing issues of cross-contamination.

Development

The development environment is sized, configured, and set up for developing

applications and systems. Unlike production hardware, the development

hardware does not have to be scalable, and it probably does not need to be as

responsive for given transactions. The development platform does need to use

the same OS type and version as used in the production environment, for

developing on Windows and deploying to Linux is fraught with difficulties that

can be avoided by matching the environments in terms of OS type and version.

After code is successfully developed, it is moved to a test system.

Test

The test environment fairly closely mimics the production environment—same

versions of software, down to patch levels, same sets of permissions, same file

structures, and so forth. The purpose of the test environment is to test a system

fully prior to deploying it into production to ensure that it is bug-free and will

not disrupt the production environment. The test environment may not scale like

production, but from the viewpoint of the software/hardware footprint, it will

look exactly like production. This is important to ensure that system-specific

settings are tested in an environment identical to that in which they will be run.

Staging

The staging environment is an optional environment, but it is commonly used

when an organization has multiple production environments. After passing

testing, the system moves into staging, from where it can be deployed to the

different production systems. The primary purpose of staging is to serve as a

sandbox after testing, so the test system can test the next set while the current set

is deployed across the enterprise. One method of deployment is a staged

deployment, where software is deployed to part of the enterprise and then a

pause occurs to watch for unseen problems. If none occur, the deployment

continues, stage by stage, until all of the production systems are changed. By

moving software in this manner, you never lose the old production system until

the end of the move, giving you time to monitor and catch any unforeseen

problems. This also prevents the total loss of production to a failed update.

Production

The production environment is where the systems work with real data, doing the

business that the system is intended to perform. This is an environment where,

by design, very few changes occur, and those that do must first be approved and

tested via the system’s change management process.

EXAM TIP Understand the structure and purpose of the different

environments so that when given a scenario and asked to identify which

environment is appropriate, you can pick the best answer: development, test,

staging, or production.

Quality Assurance (QA)

Quality assurance (QA) is a common step in any manufacturing process, and

software is no exception. Ensuring that quality is in a product is a process issue,

not an inspection issue. Yes, testing is still needed, but the current state of the art

is to drive security and quality issues via the actual software build process, not to

have a series of inspectors after it is built. This being said, there is still a role for

people who focus on quality and security issues in maintaining the bug register

(a listing of all bugs) and helping the correct people on the team get the correct

information with respect to building secure software.

Provisioning and Deprovisioning

Provisioning is the process of assigning permissions or authorities to objects.

Users can be provisioned into groups, and computer processes or threads can be

provisioned to higher levels of authority when executing. Deprovisioning is the

removal of permissions or authorities. In secure coding, the practice is to

provision a thread to an elevated execution permission level (for example, root)

only during the time that the administrative permissions are needed. After those

steps have passed, the thread can be deprovisioned back to a lower access level.

This combination lowers the period of time an application is at an increased

level of authority, thus reducing the risk exposure should the program get

hijacked or hacked.

Integrity Measurement

Integrity is defined in the security field as a determination that data has no

unauthorized changes. In a software development and deployment environment,

this is a very important issue because even little changes can cause huge issues

and can be difficult to detect. Maintaining control over a codebase means that

two things are happening. First, you have control over the copies in such a way

that people are only working on a legitimate codebase. Nothing ruins a day faster

than learning your all-day programming session was performed on a set of code

that’s not being updated—in other words, you worked on the wrong copy. While

not as disastrous as painting the wrong house (you don’t have to paint it back),

the work is in essence lost. When code is undergoing constant change from

multiple authors, this is not as simple as it seems, and some form of version

control is required. Second, you maintain a log of the changes and a method of

identifying the versions. The version control system you use should keep track

of the versions, but to clearly identify a set of code requires a different tool. A

hash algorithm creates a unique hash value for each unique item it operates on,

and codebases are digital objects. Maintaining a directory of hash values that

denote the different versions of the codebase is how integrity controls are

annotated within the code. If you have a version of the code, you can hash it and

look up in the version table to see which version you have. This is superior to

labeling in the code with metadata because the labels can be changed, but the

hash is tied to the code. When code is released for deployment, it is typically

signed digitally, and again the hash values ensure users that the code has not

been changed.

Secure Coding Techniques

Software security begins with code that is secure and free of vulnerabilities.

Unfortunately, all code has weaknesses and vulnerabilities, so instantiating the

code in a manner that has effective defenses to prevent the exploitation of

vulnerabilities can help maintain a desired level of security. Proper handling of

configurations, errors and exceptions, and inputs can assist in the creation of a

secure application. Testing the application throughout the system lifecycle can

determine the actual security risk profile of a system.

There are numerous individual elements in the secure development lifecycle

(SDL) that can assist a team in developing secure code. Correct SDL processes,

such as input validation, proper error and exception handling, and cross-site

scripting and cross-site request forgery mitigations, can improve the security of

code. Process elements such as security testing, fuzzing, and patch management

also help to ensure applications meet a desired risk profile.

Normalization

Normalization is an initial step in the input validation process. Specifically, it is

the process of creating the canonical form, or simplest form, of a string before

processing. Strings can be encoded using Unicode and other encoding methods.

This makes byte-by-byte comparisons meaningless when trying to screen user

input of strings. Checking to see whether the string is “rose” can be difficult

when “A Rose is a rose is a r%6fse.” The process of normalization converts all

of these to “rose,” where they can then be screened as valid input.

Different libraries exist to assist developers in performing this part of input

validation. Developers should always normalize their inputs prior to validation

steps to remove Unicode and other encoding issues. Per the Unicode standard,

“When implementations keep strings in a normalized form, they can be assured

that equivalent strings have a unique binary representation.”

A Rose is a rose is a r%6fse

Canonical form refers to simplest form and, because of the many encoding

schemes in use, can be a complex issue. Characters can be encoded in ASCII,

Unicode, hex, UTF-8, and even combinations of these. So, if the attacker

desires to obfuscate a response, then several things can happen.

By URL-encoding URL strings, it may be possible to circumvent filter

security systems and intrusion detection systems. For example, the URL

http://www.myweb.com/cgi?file=/etc/passwd

can become the following:

Double-encoding can complicate the matter even further. Thus, the round 1

decoding

scripts/..%255c../winnt

becomes the following:

And the round 2 decoding

scripts/..%5c../winnt

becomes the following:

scripts/..\../winnt

The bottom line is simple: know that encoding can be used and plan for it

when designing input-verification mechanisms. Expect encoded transmissions

to be used to attempt to bypass security mechanisms.

Stored Procedures

Stored procedures are methods of interfacing with database engines. Stored

procedures are precompiled scripted methods of data access that offer many

advantages. First is speed. Because they are precompiled, they can run much

more efficiently in the production environment. But because they are scripted in

advance, they offer much less flexibility than other methods such as using

parameterized queries or building and executing SQL statements on the fly in an

application program.

EXAM TIP A stored procedure is a group of one or more statements stored

within a database. Stored procedures are used in programming languages such as

SQL, Java, C++, and C.

Obfuscation/Camouflage

Obfuscation or camouflage is the hiding of obvious meaning from observation.

While obscurity is not considered adequate security under most circumstances,

adding obfuscation or camouflage to a system to make it harder for an attacker to

understand and exploit is a good thing. Numbering your e-mail servers email1,

email2, email3, . . . tells an attacker what namespace to explore. Removing or

hiding these hints makes the work harder and offers another layer of protection.

This works well for data names and other exposed elements that have to be

exposed to the outside. Where this does not work well is in the construction of

code. Obfuscated code, or code that is hard or even nearly impossible to read, is

a ticking time bomb. The day will come when someone will need to read the

code, figure out how it works so it can be modified, or determine why it is not

working. If programmers have issues reading and understanding the code,

including how it functions and what it is supposed to do, how can they

contribute to its maintenance?

Code Reuse and Dead Code

Modern software development includes the extensive reuse of components.

From component libraries to common functions across multiple components,

there is significant opportunity to reduce development costs through reuse. This

can also simplify a system through the reuse of known elements. The downside

of massive reuse is associated with a monoculture environment, which is where

a failure has a larger footprint because of all the places it is involved with.

During the design phase, decisions should be made as to the appropriate level

of reuse. For some complex functions, such as in cryptography, reuse is the

preferred path. In other cases, where the lineage of a component cannot be

established, the risk of use may outweigh the benefit. Additionally, the inclusion

of previous code, sometimes referred to as legacy code, can reduce development

efforts and risk.

EXAM TIP The use of legacy code in current projects does not exempt that

code from security reviews. All code should receive the same scrutiny, especially

legacy code that may have been developed prior to the adoption of software

development lifecycle (SDLC) processes.

Dead code is code that, while it may be executed, obtains results that are

never used elsewhere in the program. There are compiler options that can

remove dead code, called dead code elimination, but these must be used with

care. Assume you have a section of code that you put in specifically to set a

secret value to all zeros. The logic is as follows: generate a secret key, use the

secret key, set the secret key to zero. You set the secret key to zero to remove the

key from memory and keep it from being stolen. But along comes the dead code

removal routine. It sees you set the value of secretkey == 0, but then you never

use it again. So, the compiler, in optimizing your code, removes your protection

step.

Server-Side vs. Client-Side Execution and Validation

In a modern client/server environment, data can be checked for compliance with

input/output requirements either on the server or on the client. There are

advantages to verifying data elements on a client before sending them to the

server—namely, efficiency. Doing checks on the client saves a round trip, and its

delays, before the user is alerted to a problem. This can improve the usability of

software interfaces.

The client is not a suitable place to perform any critical value checks or

security checks. The reasons for this are twofold. First, the client can change

anything after the check. Second, the data can be altered while in transit or at an

intermediary proxy. For all checks that are essential, either for business reasons

or for security, the verification steps should be performed on the server side,

where the data is free from unauthorized alterations. Input validation checks can

be safely performed only on the server side.

EXAM TIP All input validation should be performed on the server side of the

client–server relationship, where it is free from outside influence and change.

Memory Management

Memory management encompasses the actions used to control and coordinate

computer memory, assigning memory to variables, and reclaiming it when it is

no longer being used. Errors in memory management can result in a program

that has a memory leak, and it can grow over time, consuming more and more

resources. The routine to clean up memory that has been allocated in a program

but is no longer needed is called garbage collection. In the C programming

language and C++, where there is no automatic garbage collector, the

programmer must allocate and free memory explicitly. One of the advantages of

newer programming languages such as Java, C#, Python, and Ruby is that they

provide automatic memory management with garbage collection. This may not

be as efficient as specifically coding in C, but it is significantly less error prone.

Use of Third-Party Libraries and Software

Development Kits (SDKs)

Programming today is to a great extent an exercise in using third-party libraries

and software development kits (SDKs). This is because once code has been

debugged and proven to work, rewriting it is generally not a valuable use of

time. Also, some fairly complex routines, such as encryption, have vetted,

proven library sets that remove a lot of risk from programming these functions.

EXAM TIP Software developers use packaged sets of software programs and

tools called SDKs to create apps for specific vender platforms.

Data Exposure

Data exposure is the loss of control over data from a system during operations.

Data must be protected during storage, during communication, and even at times

during use. It is up to the programming team to chart the flow of data through a

system and ensure it is protected from exposure throughout the process. Data can

be lost to unauthorized parties (a failure of confidentiality) and, equally

dangerous, can be changed by an unauthorized party (a failure of integrity).

EXAM TIP The list of elements under secure coding techniques is long and

specific in the CompTIA S+ exam objectives. It is important to understand the

differences so you can recognize which one best fits the context of the question.

Open Web Application Security Project

(OWASP)

The Open Web Application Security Project (OWASP) is a nonprofit foundation

dedicated to improving web-based application software security. Best known for

its top ten list of software vulnerabilities associated with website applications,

OWASP also has a multitude of useful guidelines on its website,

www.owasp.org. OWASP is a resource that should be actively used by web

application programmers to prevent vulnerabilities that are common in web

applications. This site has tons of resources to assist developers in producing

better and more secure apps.

Software Diversity

Software is not a single product. There are many different forms, and these can

be characterized by a wide range of differentiators. Software can be categorized

by elements such as platform (PC, server, mobile, IoT device, cloud),

programming language, interfaces (web, API, messaging, direct connections),

purpose, and a whole host of other factors. One can say that each project in the

end is unique. But, the fact that someone can point to why their software is

different or special does not diminish the fact that it is a series of instructions for

a computer to operate on, and that based on design decisions, coding decisions,

and environment decisions, it can and will have vulnerabilities that could enable

an attacker to do things that are not desired outcomes. Hence, all software needs

security. Having a proper security process as part of the development process is

important to reduce vulnerabilities and manage security issues as they are

uncovered.

Another key aspect of software diversity is the issue of monoculture

avoidance. As many systems in an enterprise have common elements, such as

the operating system, key libraries, and so on, there exists the possibility for

common vulnerabilities to affect many components. A consequence of software

systems sharing common vulnerabilities is an increased susceptibility to

malware and other attacks with common methods. The primary method of

beating this systemic risk is through software diversity, having different

components with different software elements.

Compilers

Compilers take computer programs written in one language and convert them to

a set of codes that can run on a specific set of hardware. Modern compilers can

take high-level, platform-agnostic code and convert it to machine language code

that actually can run on a given platform. In the process of doing this

transformation, compilers can manage various aspects of a program, such as

memory, code efficiency, and more.

Binaries

Ultimately in the end, all digital computer systems are binary machines. Binary

machines operate in one of two states: on (1) or off (0). Grouping these signals

(the 1s and 0s) together into words and larger memory and processing structures

is what makes computers capable of doing their work. But one interesting aspect

of all of this is reproducibility. Two identical computers can run the exact same

thing, and the signals and the memory structures will all be identical, because

that is how computers work. This leads to another form of important diversity:

randomization. Although all computer memory is a collection of 1s and 0s, how

these signals are arranged has implications. Having two machines, or more, with

completely identical memory layouts again provides attackers a reproducible

target. This has led to defenses that include randomizing memory layouts, where

the pattern is specific to each boot of the machine and is only known to the

machine.

Binary diversity is the creation of identically functioning binary images, but

with different specific instantiations. Different locations for memory variables,

different pointer offsets, and different layouts in computer memory can all be

done today and yet completely preserve functionality. This type of defense

makes it difficult for an attacker to bypass controls and inject something directly

into memory.

NOTE Taking binary diversity to the extreme, one can run a set of variants

simultaneously in a multivariant execution environment (MVEE). The system

then unifies inputs/outputs and monitors the operation, enabling detection of

when variants diverge in behavior. This indicates abnormal behavior and enables

the system to react and recover from a bad result stream.

Automation/Scripting

Automation through scripting and other programmable means has great utility in

software development. The use of these technology-backed methods has led to a

field of development known as DevOps. DevOps is a combination of

development and operations—in other words, a blending of tasks performed by a

company’s application development and systems operations teams. DevOps

emphasizes communication and collaboration between product management,

software development, and operations professionals to facilitate continuous

development, continuous integration, continuous delivery, and continuous

monitoring processes. DevOps can be considered the anti-waterfall model

because rather than going from phase to phase, in DevOps, as small changes are

ready to advance, they advance. This leads to many small incremental changes

but less time between updates and less time to fix or change things. Secure

DevOps is the addition of security steps to the DevOps process. Just as you can

add security steps to the waterfall model, or any other software development

model, you can add them to DevOps as well, resulting in a secure DevOps

outcome.

Automated Courses of Action

One of the key elements of DevOps is automation. DevOps relies on automation

for much of its efficiencies. Security automation can do the same for security

that automation has in DevOps. Automating routines and extensive processes

allows fewer resources to cover more of the environment in a more effective and

efficient manner. Automation removes manual labor costs, especially for skilled

cybersecurity personnel. Rather than replacing the personnel with scripts, the use

of automation allows the personnel to spend their time doing value-added

analysis work.

EXAM TIP The implications of continuous

monitoring/validation/integration/delivery/deployment will likely depend on the

details of the question, the context, and the specific question being asked. To

determine which aspect is correct requires careful examination of the context of

the question. Learn the differences, not just the context of “continuous.”

Continuous Monitoring

Continuous monitoring is the term used to describe the technologies and

processes employed to enable rapid detection of compliance issues and security

risks. More than just a buzzword, continuous monitoring is one of the most

important tools available for risk management. Automation and scripts are

commonly used as part of a continuous monitoring framework, as they can

provide 24/7/365 monitoring of processes and conditions, feeding alerts into the

organization’s monitoring system for review and action by security personnel.

Continuous Validation

Continuous validation is the extension of testing to support the continuous

process of software development that occurs in DevOps. As code is changed in

the DevOps process, the new code must be tested with the existing codebase to

ensure functionality and stability. Making this process part of the continuous

development process is essential to keeping development on a timely trajectory.

Continuous Integration

Continuous integration is the DevOps manner of continually updating and

improving the production codebase. By using high levels of automation and

safety nets of automated back-out routines, continuous integration allows for

testing and updating even minor changes without a lot of overhead. This means

that rather than several large updates, with many integrated and many potentially

cross-purpose update elements, all squeezed into a single big package, a whole

series of smaller single-purpose integrations is run. Thus, when testing, you have

isolated the changes to a small manageable number, without the significance of

multiple potential interactions. This reduces interaction errors and other types of

errors that are time-consuming to chase down.

Continuous Delivery

Continuous delivery is a natural extension of continuous integration so you can

quickly release new changes to production in a sustainable way. Continuous

delivery relies on automated testing and is an automated release process that

enables the delivery of updates when they are complete, at any point of time, as

opposed to a fixed release schedule. When code is ready to be released to

production, continuous delivery is the automation of that step, but still under

specific operator control.

Continuous Deployment

Continuous deployment is continuous delivery on autopilot. It goes one step

further than continuous delivery in that the release is automatic. With this

practice, every change that passes all stages of your production pipeline is

released to production. There’s no human intervention, and when all gates are

met (that is, there are no failed tests), continuous deployment automatically

sends the code to production.

EXAM TIP Continuous deployment goes one step further than continuous

delivery—every change that passes all stages of your production pipeline is

automatically released to customers.

Elasticity

Elasticity is the characteristic that something is capable of change without

breaking. One of the strengths of cloud computing is its elasticity. One can add

or subtract resources to or from a cloud environment almost automatically

without issue. Elasticity in software works in the same fashion—how resilient

the software is to changes in its environment while remaining secure. For

software to be elastic, it needs to be able to run under a variety of different

conditions. Legacy software that runs in a single thread, although easier to write,

is not elastic. When single-threaded software gets employed in an environment

of VMs, multiple processors, and cloud environments, its performance is limited

to a single thread. Multithreaded software can scale and adapt better, but this

also increases the complexity, bringing in issues such as race conditions. For

scalability to be stable and sustainable, the software needs to be elastic.

Scalability

Scalability is the characteristic of a software system to process higher workloads

on its current resources (scale up) or on additional resources (scale out) without

interruption. Scalability is important in web systems, databases, application

engines, and cloud systems. Workloads can vary, and cloud/container systems

can add processing and storage, but the software must be capable of addressing

the changes in an environment. While this seems obvious, the devil is in the

details. Timing loops can affect the ability of software to run on faster hardware,

as the system can only run as fast as its slowest link. Scaling out to multiple

machines brings in issues of synchronization and coordination. All of these

issues can be solved, but this has to happen during design and development, not

after delivery.

Version Control

Programs are developed, released, and used, and then changes are needed, either

to alter functionality, to fix errors, or to improve performance. This leads to

multiple versions of programs. Version control is as simple as tracking which

version of a program is being worked on, whether in dev, test, or production.

Versioning tends to use the first whole number to indicate major releases and

uses numbers after a decimal point to indicate minor changes.

Having the availability of multiple versions brings into focus the issue of

change management. How does a firm manage which versions are currently

being used, and how do they coordinate changes as they are released by a

manufacturer? In traditional software publishing, a new version required a new

install and fairly significant testing because the level of change could be drastic

and call into question issues of compatibility, functionality, and even correctness.

DevOps turned the tables on this equation by introducing the idea that

developers and production work together and create in essence a series of microreleases so that any real problems are associated with single changes and not

bogged down by interactions between multiple module changes.

Whether you are in traditional software publishing or operating in the

DevOps world, you still need a change management process that ensures all

changes in production are authorized, properly tested, and rolled back if they

fail, and you must maintain current and accurate documentation.

Chapter Review

In this chapter, you became acquainted with secure application development,

deployment, and automation concepts. The chapter opened with a discussion of

the environment, including development, test, staging, production, and quality

assurance elements. It then examined provisioning and deprovisioning, and

integrity management.

The chapter then moved on to secure coding techniques. These include

normalization, stored procedures, obfuscation/camouflage, and code reuse and

dead code. This section included a discussion of server-side versus client-side

validation elements and it closed with memory management, the use of thirdparty libraries and software development kits, and data exposure.

The Open Web Application Security Project (OWASP) and software diversity

were discussed next. Under software diversity, issues of compiler and binary

diversity are covered. The chapter then moved on to the issues from the world of

DevOps or automation and scripting as applied to software development. The

topics in this section included automated courses of action, continuous

monitoring, continuous validation, continuous integration, continuous delivery,

and continuous deployment.

The chapter closed with an examination of elasticity and scalability followed

by version control.

Questions

To help you prepare further for the exam, and to test your level of preparedness,

answer the following questions and then check your answers against the correct

answers at the end of the chapter.

1. To develop secure software that prevents attackers from directly injecting

attacks into computer memory and manipulating the application’s process,

one should employ which method?

A. Elasticity

B. Dead code

C. Normalization

D. Software diversity

2. Problems in which phase will specifically stop continuous deployment but

not necessarily continuous delivery?

A. Continuous integration

B. Continuous monitoring

C. Continuous validation

D. Continuous development

3. Why is memory management important in software development?

A. A program can grow and consume other program spaces.

B. Memory is expensive.

C. Memory can be a speed issue.

D. None of the above.

4. When a program is installed and needs permissions, what is this called?

A. Staging

B. Provisioning

C. Continuous integration

D. Version control

5. Which of the following statements concerning elasticity and scalability are

true?

A. Scalability requires elasticity.

B. Elasticity involves enabling software to use more processors to do

more work.

C. Elasticity means being prepared to take advantage of scalability.

D. All of the above.

6. To protect software from reverse engineering by attackers, developers can

use which of the following?

A. Dead code

B. Obfuscation

C. Binary diversity

D. Stored procedures

7. To manage various releases of software over time, the organization uses

which of the following?

A. Staging environment

B. Provisioning and deprovisioning steps

C. Version control

D. Continuous integration

8. Which of the following environments is used to test compatibility against

multiple target environments?

A. Production

B. Test

C. Quality assurance

D. Staging

9. The fact that there are multiple methods of representing an object in a

computer system can lead to issues when logical comparisons are needed.

What can be used to ensure accuracy of comparison elements?

A. Normalization

B. Stored procedures

C. Third-party libraries

D. Third-party software development kits

10. What is the only sure method of ensuring input is valid before use on a

server?

A. Use of third-party libraries and software development kits

B. Server-side validation

C. Stored procedures

D. Client-side validation

Answers

1. D. Software diversity in the form of diverse binaries will prevent direct

memory attacks against known software structures.

2. C. Continuous validation is required to ensure error-free software, and

errors will stop continuous deployment.

3. A. Memory management failures can lead to a program growing in size

when executing. This can result in either its own failure or the diminishing

of memory resources for other programs.

4. B. Provisioning is the assignment of permissions or authorities to objects.

5. D. All of the above is the correct answer. Scalability requires elasticity to

scale, elasticity involves enabling software to use more processors to do

more work, and elasticity means developing software that is prepared to

take advantage of scalability.

6. B. Obfuscation is the technique of hiding properties to prevent examination.

Making code hard to decompile and not storing any specific clues in the

source code can make reverse engineering a challenge.

7. C. Version control comprises the processes and procedures employed to

manage different releases of software over time.

8. D. The staging environment can be used to manage software releases

against different targets to ensure compatibility.

9. A. Normalization is the process of reducing items to a canonical form

before comparisons to ensure appropriate logical matching.

10. B. Server-side validation is the only sure validation method for inputs to the

application.

CHAPTER 12

Authentication and Authorization

In this chapter, you will

• Learn how to identify and implement authentication methods, factors, and

attributes

• Learn about authorization design concepts and requirements

One of the core tenets of computer security is the concept that all actions will be

controlled via a system of approvals; for example, only authorized parties can

perform the actions of accessing a resource, operating on a resource, and storing

an item. Identity and access management systems are the mechanisms by which

this is accomplished. This chapter examines the foundational elements behind

authentication systems.

Certification Objective This chapter covers CompTIA Security+ exam

objective 2.4: Summarize authentication and authorization design concepts.

Authentication Methods

Authentication is the process of verifying an identity previously established in a

computer system. There are a variety of methods of performing this function,

each with its advantages and disadvantages, as detailed in the following sections.

Directory Services

A directory is a data storage mechanism similar to a database, but it has several

distinct differences designed to provide efficient data-retrieval services

compared to standard database mechanisms. A directory is designed and

optimized for reading data, offering very fast search and retrieval operations.

The types of information stored in a directory tend to be descriptive attribute

data. A directory offers a static view of data that can be changed without a

complex update transaction. The data is hierarchically described in a treelike

structure, and a network interface for reading is typical. Common uses of

directories include e-mail address lists, domain server data, and resource maps of

network resources. The Lightweight Directory Access Protocol (LDAP) is

commonly used to handle user authentication and authorization and to control

access to Active Directory (AD) objects.

To enable interoperability, X.500 was created as a standard for directory

services. The primary method for accessing an X.500 directory is through the

Directory Access Protocol (DAP), a heavyweight protocol that is difficult to

implement completely, especially on PCs and more constrained platforms. This

led to LDAP, which contains the most commonly used functionality. LDAP can

interface with X.500 services and, most importantly, can be used over TCP with

significantly less computing resources than a full X.500 implementation. LDAP

offers all of the functionality most directories need and is easier and more

economical to implement; hence, LDAP has become the Internet standard for

directory services. LDAP standards are governed by two separate entities,

depending on use: the International Telecommunication Union (ITU) governs the

X.500 standard, and LDAP is governed for Internet use by the Internet

Engineering Task Force (IETF). Many RFCs apply to LDAP functionality, but

some of the most important are RFCs 4510 through 4519.

When integrating with cloud-based systems, you might find managing

credentials across the two different domains challenging. Different vendors have

created directory-based technologies to address this, such as AWS Directory

Service for Microsoft Active Directory, also known as AWS Managed Microsoft

AD. This service enables your directory-aware workloads and AWS resources to

use a managed Active Directory in the AWS Cloud. Because AWS Managed

Microsoft AD is built on the actual Microsoft Active Directory, you can use

standard Active Directory administration tools and take advantage of built-in

Active Directory features, such as Group Policy and single sign-on (SSO)

features.

EXAM TIP A client starts an LDAP session by connecting to an LDAP server,

called a Directory System Agent (DSA), which is by default on TCP and UDP

port 389 or on port 636 for LDAPS (LDAP over SSL).

Federation

Federation, or identity federation, defines policies, protocols, and practices to

manage identities across systems and organizations. Federation’s ultimate goal is

to allow users to seamlessly access data or systems across domains. Federation is

enabled through the use of industry standards such as Security Assertion Markup

Language (SAML), discussed in Chapter 24, “Implement Authentication and

Authorization.”

EXAM TIP Federated identity access management systems allow users to

authenticate and access resources across multiple enterprises using a single

credential. But don’t confuse this with single sign-on (SSO), which allows users

access to multiple resources within a single organization or enterprise.

Attestation

Attestation is the supplying of proof or evidence of some fact. In the case of

authentication, attestation can be done by a service that checks the credentials

supplied, and if they are correct and match the required values, the service can

attest that the entry is valid or correct. Attestation is used throughout

cybersecurity whenever a third party or entity verifies an object as valid or an

item as correct in value.

Technologies

There are multiple ways to perform authentication, and multiple technologies

can be employed to assist in the effort.

Time-based One-Time Password (TOTP)

The Time-based One-Time Password (TOTP) algorithm is a specific

implementation of an HOTP (discussed next) that uses a secret key with a

current timestamp to generate a one-time password (OTP). It is described in

RFC 6238 (May 2011).

HMAC-based One-Time Password (HOTP)

HMAC-based One-Time Password (HOTP) is an algorithm that can be used to

authenticate a user in a system by using an authentication server. (HMAC stands

for hash-based message authentication code.) It is defined in RFC 4226

(December 2005).

EXAM TIP HOTP passwords can remain valid and active for an unknown

time period. TOTP passwords are considered more secure because they are valid

for short amounts of time and change often.

Short Message Service (SMS)

The use of Short Message Service (SMS), or text messaging, in a cell phone

provides a second authentication factor that is sent to a preidentified number.

The message that is sent provides a code that the user enters into the system.

This code typically has an expiration time, as shown in Figure 12-1. This is a

way of verifying that the first credential, usually a password, was entered by the

person expected—assuming they have control over the cell phone. This is a

practical example of multifactor authentication, which is discussed later in this

chapter.

Figure 12-1 Sample SMS verification codes

Token Key

Token keys are physical devices that carry a digital token used to identify the

user. This is a “something you have” element in a multifactor authentication

scheme. The format of the actual token can vary from a smart card, to a

keyboard fob, to a USB device. Proximity cards used in physical access systems

are token-carrying devices.

As in all “something you have” elements, tokens are a proof of possession

type of event, and to prevent their use if lost, they are backstopped with a PIN

code. Different tokens can carry different forms of keys. The keys can be

dynamic, changing over time, or static. Dynamic tokens add security in that the

value changes over time and cannot be captured and replayed. An example of a

commercial token is shown in Figure 12-2.

Figure 12-2 Token authenticator from Blizzard Entertainment

Static Codes

Static codes are just that—codes that do not change, or are static in nature. There

are many use cases where these are essential, such as devices without user

intervention. Devices that do not have user intervention are widely deployed in

many systems. An example would be a smart electric meter, a device that needs

to communicate with other systems and authenticate its identity. The use of static

codes has a weakness in that, if compromised, the keys are no longer valid. The

standard is to use cryptographic protection of all transmission of static codes,

making the code unreadable even if the communication channel data is copied.

Authentication Applications

Need a second factor for authentication? We have an app for that. And this is not

just a joke, but an increasingly common method of authentication that works by

verifying that a user has a given mobile device in their possession. An

authentication application functions by accepting user input, and if the user

input is correct, it can pass the appropriate credentials to the system requesting

authentication. This can be in the form of either a stored digital value or a onetime code in response to a challenge. Authentication applications exist for a

variety of platforms—from Android to iOS, Linux, and Windows—and there are

multiple vendors for each platform. The use of the application on the device is a

second factor of authentication and is part of a multifactor authentication

scheme.

Push Notifications

Push notification authentication supports user authentication by pushing a

notification directly to an application on the user’s device. The user receives the

alert that an authentication attempt is taking place, and they can approve or deny

the access via the user interface on the application. The push notification itself is

not a secret; it is merely a means by which the user can authenticate and approve

access. This is an out-of-band communication and demonstrates a second

communication channel, thus making account hacking significantly more

challenging.

Phone Call

Another form of authenticating a user has an interaction with the system via a

phone call. The authentication phone call is delivered from the authentication

system to a specified phone number, which then can verify that the user is in

possession of the actual mobile device.

EXAM TIP Tokens represent something you have with respect to

authentication as well as devices that can store more information than a user can

memorize, which makes them very valuable for access control. The details in the

scenario preceding a question will provide the necessary criteria to pick the best

token method for the question.

Smart Card Authentication

A smart card (also known as an integrated circuit card [ICC] or chip card) is a

credit card–sized card with embedded integrated circuits that is used to provide

identification security authentication. Smart cards can increase physical security

because they can carry long cryptographic tokens that are too long to remember

and too large a space to guess. Also, because of the manner in which smart cards

are employed and used, copying the number is not a practical option. Smart

cards can be used in a variety of situations where you want to combine

something you know (a PIN or password) with something you have (and can’t be

duplicated, such as a smart card). Many standard corporate-type laptops come

with smart card readers installed, and their use is integrated into the Windows

user access system.

Biometrics

Biometric factors are measurements of certain biological factors to identify one

specific person from others. These factors are based on parts of the human body

that are unique. The most well-known of these unique biological factors is the

fingerprint. Fingerprint readers have been available for several years in laptops

and other mobile devices, on keyboards, and as stand-alone USB devices.

However, many other biological factors can be used, such as the retina or iris

of the eye, the geometry of the hand, and the geometry of the face. When these

are used for authentication, there is a two-part process: enrollment and then

authentication. During enrollment, a computer takes the image of the biological

factor and translates it to a numeric value, called a template. When the user

attempts to authenticate, the biometric feature is scanned by the reader, and the

computer computes a value in the same fashion as the template and then

compares the numeric value being read to the one stored in the database. If they

match, access is allowed. Since these physical factors are unique, theoretically

only the actual authorized person would be allowed access.

In the real world, however, the theory behind biometrics breaks down. Tokens

that have a digital code work very well because everything remains in the digital

realm. A computer checks your code, such as 123, against the database; if the

computer finds 123 and that number has access, the computer opens the door.

Biometrics, however, take an analog signal, such as a fingerprint or a face, and

attempt to digitize it, and it is then matched against the digits in the database.

The problem with an analog signal is that it might not encode the exact same

way twice. For example, if you came to work with a bandage on your chin,

would the face-based biometrics grant you access or deny it? Because of this, the

templates are more complex in a manner where there can be a probability of

match, or closeness measurement.

Fingerprint

A fingerprint scanner measures the unique pattern of a person’s fingerprint and

translates that pattern into a numerical value, or template, as discussed in the

previous section. Fingerprint readers can be enhanced to ensure that the pattern

is a live pattern—one with circulating blood or other detectable biological

activity—to prevent simple spoofing with a Play-Doh mold of the print.

Fingerprint scanners are cheap to produce and have widespread use in mobile

devices. One of the challenges of fingerprint scanners is that they don’t function

if the user is wearing gloves (for example, medical gloves) or has worn off their

fingerprints through manual labor, as many involved in the sheetrock trade do

through normal work.

Retina

A retinal scanner examines blood vessel patterns in the back of the eye. Believed

to be unique and unchanging, the retina is a readily detectable biometric. Retinal

scanning does suffer from lack of user acceptance, as it involves a laser scanning

the inside of the user’s eyeball, which raises some psychological issues for some

users who are wary of letting a laser scan the inside of their eye. This detection

requires the user to be right in front of the device for it to work. It is also more

expensive because of the precision of the detector and the involvement of lasers

and users’ vision.

Iris

An iris scanner works in a manner similar to a retinal scanner in that it uses an

image of a unique biological measurement (in this case, the pigmentation

associated with the iris of the eye). This can be photographed and measured from

a distance, removing the psychological impediment of placing one’s eye up close

to a scanner. The downside to being able to capture an iris scan at a distance is

that it’s easy to do without a person’s knowledge, and even to construct contact

lenses that mimic a pattern. There are also some other issues associated with

medical conditions such as pregnancy and some diseases that can be detected by

changes in a person’s iris and, if revealed, would be a privacy violation.

Facial

Facial recognition was mostly the stuff of sci-fi until it was integrated into

various mobile phones. A sensor that recognizes when you move the phone into

a position to see your face, coupled with a state of not being logged in, turns on

the forward-facing camera, causing the system to look for its enrolled owner.

This system has proven to have fairly high discrimination and works fairly well,

with only one drawback: another person can move the phone in front of the

registered user and it can unlock. In essence, another user can activate the

unlocking mechanism when the user is unaware. The other minor drawback is

that for certain transactions, such as positive identification for financial

transactions, the position of the phone on an NFC location, together with the

user’s face needing to be in a certain orientation with respect to the phone, leads

to some awkward positions. In other words, having to put your face in a proper

position on the phone to identify you while holding it against the counter-height

NFC credit card reader can be awkward.

Voice

Voice recognition is the use of unique tonal qualities and speech patterns to

identify a person. Long the subject of sci-fi movies, this biometric has been one

of the hardest to develop into a reliable mechanism, primarily because of

problems with false acceptance and rejection rates, which will be discussed a bit

later in the chapter.

Vein

A different biometric is the use of blood vein patterns to identify a user. Humans

have a common vascular system, but the individual elements can vary in size and

microstructure, and these fine-grained patterns are believed to be unique.

Sensors can measure these patterns and use them to identify a user. Three

common vascular pattern locations are used: palms, fingers, and the veins in the

retina. This measurement is done via spectral analysis of the tissue, using

frequencies that detect the hemoglobin in the blood. These are noninvasive

measurements, but they do require close proximity to the user’s item under

measurement.

Gait Analysis

Gait analysis is the measurement of the pattern expressed by a person as they

walk. An analysis of the gait, its length, the speed, and the rate of movement of

specific points provides a unique signature that can be recorded and compared to

previous samples. Even when not used for authentication, as a previous sample

is required, gait analysis can be used to identify a suspect in a group of others,

enabling the tracking of individuals in a crowd. From an access control

perspective, in high-security situations, a camera can record the gait of incoming

personnel and compare it to known values, providing a remote and early

additional factor in determining identity.

Efficacy Rates

Biometric measurements have a level of uncertainty, and thus the efficacy of

biometric solutions has been an issue since they were first developed. As each

generation of sensor improved the accuracy of the measurements, the errors have

been reduced to what is now a manageable level. For biometrics to be effective,

they must have both low false positive rates and low false negative rates. The

terms false acceptance rate (FAR) and false rejection rate (FRR) describe the

chance that an incorrect user will be falsely accepted or a correct user will be

falsely rejected, respectively, as covered in detail in the next sections. These two

measures are different, and while a low false rejection rate is important for

usability, a low false acceptance rate is more important from a security

perspective. Users having to repeat trying to authenticate is bad, but having

authentication occur for unauthorized users is worse.

The FIDO Alliance, a leading authentication standards and certification body,

has specifications for error rates. FRR should be below 3 percent (no more than

three errors in 100 attempts) and FAR should be below 0.01 percent (no more

than one error in 10,000 attempts). As in all defense-in-depth scenarios, the

backstop is a lockout function where devices will lock after a certain number of

failed attempts. This makes the FAR more secure than just the simple

percentage.

False Acceptance

The false acceptance rate (FAR) determines what level of false positives is

allowed in the system. A false acceptance (or false positive) is demonstrated by

the grayed-out area in Figure 12-3. In this area, the two curves overlap, and the

decision has been made that at this threshold (or better) an accept signal will be

given. Thus, if you are not a match, but your measured value falls on the upper

end of the nonmatch curve (in the gray area), you will be considered a match,

and hence become a false positive. The false acceptance rate is expressed as the

probability that the system incorrectly identified a match between the biometric

input and the stored template value.

Figure 12-3 False acceptance rate

When selecting the threshold value, the designer must be cognizant of two

factors. One is the rejection of a legitimate biometric—the area on the match

curve below the threshold value. The second is the acceptance of false positives.

The more the curves overlap, the larger the problem, because once a threshold is

chosen, that number defines the FAR. Setting the threshold higher will decrease

false positives but increase false negatives or rejections. This would increase the

false rejection rate, discussed in the next section.

False Rejection

The false rejection rate (FRR) determines what level of false negatives, or

rejections, are going to be allowed in the system. A false rejection is

demonstrated by the grayed-out area in Figure 12-4. In this section, the curves

overlap, and the decision has been made that at this threshold (or lower), a reject

signal will be given. Thus, if you are on the lower end of the match curve (in the

gray area), you will be rejected, even if you should be a match. The false

rejection rate is expressed as the probability that the system incorrectly rejected a

legitimate match between the biometric input and the stored template value.

Figure 12-4 False rejection rate

When comparing the FAR and the FRR, one realizes that, in most cases,

whenever the curves overlap, they are related. This brings up the issue of the

crossover error rate. Both the FAR and the FRR are set by choosing the threshold

value. This is done when the system is set up and reflects the choice of which

error rate is more important. If you want to make it harder for a false positive to

occur, you will cause many failed authorizations of legitimate users because they

will be seen by the system as on the other curve. If you want to make sure all

legitimate users do not experience trouble during scans, then some unauthorized

users will get accepted (false positives) because they will be interpreted by the

system as being on the wrong curve based on where the threshold is set.

Crossover Error Rate

The crossover error rate (CER) is where both accept and reject error rates are

equal. This is the desired state for the most efficient operation, and it can be

managed by manipulating the threshold value used for matching. In practice, the

values may not be exactly the same, but they will typically be close to each

other. Figure 12-5 demonstrates the relationship between the FAR, FRR, and

CER.

Figure 12-5 FRR, FAR, and CER compared

EXAM TIP Remember that the crossover error rate (CER) is the percentage at

which the false acceptance rate (FAR) and false rejection rate (FRR) are equal.

Multifactor Authentication (MFA) Factors

and Attributes

Multifactor authentication (or multiple-factor authentication) is simply the

combination of two or more types of authentication. Five broad categories of

authentication can be used: what you are (for example, biometrics), what you

have (for instance, tokens), what you know (passwords and other information),

somewhere you are (location), and something you do (physical performance).

Two-factor authentication combines any two of these before granting access. An

example would be a card reader that then turns on a fingerprint scanner—if your

fingerprint matches the one on file for the card, you are granted access. Threefactor authentication would combine three types, such as a smart card reader that

asks for a PIN before enabling a retina scanner. If all three, card (physical), PIN

(knowledge), and scan (biometric), correspond to a valid user in the computer

database, access is granted.

EXAM TIP Two-factor authentication combines any two methods, matching

items such as a token with a biometric. Three-factor authentication combines any

three, such as a passcode, biometric, and a token.

Multifactor authentication methods greatly enhance security by making it

very difficult for an attacker to obtain all the correct materials for authentication.

They also protect against the risk of stolen tokens, as the attacker must have the

correct biometric, password, or both. More important, multifactor authentication

enhances the security of biometric systems by protecting against a spoofed

biometric. Changing the token makes the biometric useless unless the attacker

can steal the new token. It also reduces false positives by trying to match the

supplied biometric with the one that is associated with the supplied token. This

prevents the computer from seeking a match using the entire database of

biometrics. Using multiple factors is one of the best ways to ensure proper

authentication and access control.

Factors

Factors are the specific elements that comprise an item of proof. These items

can be grouped into three classes: something you know (passwords), something

you have (token), and something you are (biometrics). Each of these has

advantages and disadvantages, as discussed in the following sections.

Something You Know

The most common example of something you know is a password. One of the

challenges with using something you know as an authentication factor is that it

can be “shared” (or duplicated) without you knowing it. Another concern with

using something you know is that because of the vast number of different

elements a typical user has to remember, they do things to assist with memory,

such as repeating passwords, making slight changes to a password, such as

incrementing the number from password1 to password2, and writing them down.

These are all common methods used to deal with password sprawl, yet they each

introduce new vulnerabilities.

Another form of authentication via something you know is identity-driven

authentication. In identity-driven authentication, you contact someone to get

access, and they respond with a series of challenge questions. Sometimes the

questions are based on previously submitted information, and sometimes the

questions are based on publicly known information, such as previous addresses,

phone numbers, cars purchased/licensed, and so on. Again, the proper

respondent will know these answers, whereas an imposter will not. These tests

are timed, and if the respondent takes too long (for example, taking the time to

perform a lookup), they will fail.

Something You Have

Something you have specifically refers to security tokens and other items that a

user can possess physically. One of the challenges with using something you

have as an authentication factor is that you have to have it with you whenever

you wish to be authenticated, and this can cause issues. It also relies on

interfaces that may not be available for some systems, such as mobile devices,

although interfaces, such as one-time password (OTP) generators, are device

independent. OTP generators create new passwords on demand, against a

sequence that is known only to the OTP generator and the OTP element on the

system accepting the password.

One of the challenges of something you have is the concept of “something

you lost,” such as something you left in a briefcase, at home, and so on. Just as

leaving your key ring with your office key can force a return trip back home to

get it, so can leaving a dongle or other security element that is “something you

have” in nature. And if something you have becomes something you had stolen,

the implications are fairly clear—you don’t have access and you have to reidentify yourself to get access again.

Something You Are

Something you are specifically refers to biometrics. One of the challenges with

using “something you are” artifacts as authentication factors is that typically

they are hard to change; once assigned, they inevitably become immutable, as

you can change fingers, but only a limited number of times, and then you run out

of changes. Another challenge with biometrics is that cultural or other issues

associated with measuring things on a person may exist. For example, people in

some cultures object to having their pictures taken. Another example is that

physical laborers in some industries tend to lack scannable fingerprints because

they are worn down. Some biometrics are not usable in certain environments; for

instance, in the case of medical workers, or workers in clean-room

environments, their personal protective gear inhibits the use of fingerprint

readers and potentially other biometric devices.

Attributes

Attributes are collections of artifacts, like the factors previously presented, but

rather than focus on the authentication item, they focus on elements associated

with the user. Common attributes include the user’s location, their ability to

perform a task, or something about the user themselves. These attributes are

discussed in the following sections.

Somewhere You Are

One of the more discriminant authentication factors is your location, or

somewhere you are. When a mobile device is used, GPS can identify where the

device is currently located. When you are logged on to a local, wired desktop

connection, it shows you are in the building. Both of these can be compared to

records to see if you are really there or should be there. If you are badged into

your building, and at your desk on a wired PC, then a second connection with a

different location would be suspect, as you can only be in one place at a time.

With geofencing (see Chapter 21, “Secure Mobile Solutions,” for details),

location becomes a big thing for marketing services pushing content to devices

when in specific locations. Location services on mobile devices, coupled with

geofencing, can alert others when you are in a specific area—not specifically

authentication, but leading toward it.

Something You Can Do

Something you can do specifically refers to a physical action that you perform

uniquely. An example of this is a signature; the movement of the pen and the

two-dimensional output are difficult for others to reproduce. This makes it useful

for authentication, but challenges exist in capturing the data, as signature pads

are not common peripherals on machines. Gait analysis, presented earlier, is

another example of this attribute. Something you can do is one of the harder

artifacts to capture without specialized hardware, making it less ubiquitous as a

method of authentication.

Something You Exhibit

Something you exhibit is a special case of a biometric. An example would be a

brainwave response to seeing a picture. Another example would be the results of

a lie detector test. The concept is to present a trigger and measure a response that

cannot be faked. As sensors improve, tracking eye movement and sensing other

aspects of responses will become forms that can be used to assist in

authentication.

Someone You Know

Just as passwords relate to possession of knowledge, someone you know relates

to a specific memory, but in this case an individual. This is the classic “having

someone vouch for you” attribute. Electronically, this can be done via a chain of

trust model, and it was commonly used in the past as a result of people signing

each other’s keys, indicating trust.

EXAM TIP Be able to differentiate between the three factors for authentication

(something you know, have, and are) as well as the four attributes (somewhere

you are, something you can do and exhibit, and someone you know). These are

easily tested on the exam. Be sure you recognize examples for each factor to

match to a scenario-type question.

Authentication, Authorization, and

Accounting (AAA)

Authentication is the process of verifying an identity previously established in a

computer system. There are a variety of methods for performing this function,

each with its advantages and disadvantages. Authentication methods and their

advantages and disadvantages were described throughout the chapter.

Authorization is the process of permitting or denying access to a specific

resource. Once identity is confirmed via authentication, specific actions can be

authorized or denied. Many types of authorization schemes are used, but the

purpose is the same: determine whether a given user who has been identified has

permissions for a particular object or resource being requested. This

functionality is frequently part of the operating system and is transparent to

users.

Accounting is the process of ascribing resource usage by account for the

purpose of tracking resource utilization. This is a basic accounting function that

is still used by some enterprises.

The separation of tasks, from identification to authentication to authorization,

has several advantages. Many methods can be used to perform each task, and on

many systems several methods are concurrently present for each task. Separation

of these tasks into individual elements allows combinations of implementations

to work together. Any system or resource, be it hardware (router or workstation)

or a software component (database system), that requires authorization can use

its own authorization method once authentication has occurred. This makes for

efficient and consistent application of these principles.

EXAM TIP Authentication is the process of validating an identity.

Authorization is the process of permitting or denying access to resources.

Accounting is the process of keeping track of the resources a user accesses.

Together, they make up the AAA framework for identity access security.

Cloud vs. On-premises Requirements

Authentication to cloud versus on-premises requirements is basically a revisiting

of the identity and authentication problem all over again. When establishing

either a cloud or on-premises system, you use identity and authentication as the

foundation of your security effort. Whether you use an Active Directory

methodology or other system to manage identities on premises, when you’re

establishing a cloud-based system, the options need to be completely reviewed

and appropriate choices made based on the use of the cloud in the enterprise.

Simple methods include a completely new independent system, although this

increases costs and reduces usability when the number of users grows. Solutions

such as federated authentication and single sign-on exist, and the proper

determination of authentication processes should rest on data criticality and who

needs access.

Chapter Review

In this chapter, you became acquainted with design concepts associated with

authentication and authorization. The chapter opened with an examination of the

authentication methods, including directory services, federation, attestation, and

technologies. The technologies we examined were time-based one-time

passwords, HMAC-based one-time passwords, Short Message Service, token

keys, static codes, authentication applications, push notifications, and phone

calls. The initial section closed with a discussion of smart card–based

authentication.

The discussion of biometrics included both factors and usage. You learned

about several different biometric technologies: fingerprint scanning, retinal

scanning, iris scanning, facial recognition, voice recognition, vein patterns, and

gait analysis. The methods and analytics covered included efficacy rates: false

acceptance rate, false rejection rate, and crossover error rate.

The chapter continued with an examination of multifactor authentication

factors and attributes. The three factors presented were something you know,

something you have, and something you are. Four attributes followed: something

you are, something you can do, something you exhibit, and someone you know.

The chapter concluded with an examination of authentication, authorization, and

accounting as well as cloud versus on-premises requirements.

Questions

To help you prepare further for the CompTIA Security+ exam, and to test your

level of preparedness, answer the following questions and then check your

answers against the correct answers at the end of the chapter.

1. During a visit to a hosting center where your organization keeps some

offsite servers, you see a door with an odd-looking panel next to it. You see

people approaching the panel and placing their eyes into a hooded viewer.

A few seconds after they’ve done this, the door unlocks. What type of

biometric scanner might this be?

A. Voice recognition scanner

B. Retinal scanner

C. Fingerprint scanner

D. Facial recognition scanner

2. You’ve spent the last week tweaking a fingerprint-scanning solution for

your organization. Despite your best efforts, roughly 1 in 50 attempts will

fail, even if the user is using the correct finger and their fingerprint is in the

system. Your supervisor says 1 in 50 is “good enough” and tells you to

move on to the next project. Your supervisor just defined which of the

following for your fingerprint scanning system?

A. False rejection rate

B. False acceptance rate

C. Critical threshold

D. Failure acceptance criteria

3. Which of the following algorithms uses a secret key with a current

timestamp to generate a one-time password?

A. Hash-based Message Authentication Code

B. Date-Hashed Message Authorization Password

C. Time-based One-Time Password

D. Single sign-on

4. With regard to authentication, an access token falls into which factor

category?

A. Something you are

B. Something you have

C. Something you know

D. Something you see

5. Which of the following is not a common form of hardware token?

A. Proximity card

B. Common access card

C. USB token

D. Iris scan

6. While depositing cash from a charity fundraiser at a local bank, you notice

bank employees are holding up cards next to a panel near a door. A light on

the panel turns green and the employees are able to open the door. The light

on the panel is normally red. What type of electronic door control is this

bank using?

A. Iris scanner

B. Hardware token

C. Proximity card

D. Symmetric key token

7. Your colleague is telling you a story she heard about a way to trick

fingerprint scanners using gummy bears. She heard that if you press a

gummy bear against an authorized user’s finger, you can then use that

gummy bear as their fingerprint to fool a fingerprint scanner. If this works,

the result is an example of which of the following?

A. False negative

B. False positive

C. Crossover positive

D. Crossover negative

8. To ensure customers entering credentials in your website are valid and not

someone with stolen credentials, your team is tasked with designing

multifactor authentication. Which of the following would not be a good

choice?

A. Static code

B. Phone call

C. Authentication application

D. Short Message Service

9. When you’re designing and tweaking biometric systems, the point where

both the accept and reject error rates are equal is known as which of the

following?

A. Crossover acceptance rate

B. Accept-reject overlap rate

C. Crossover error rate

D. Overlap acceptance rate

10. Which of the following is not a term used in multifactor authentication?

A. Someone you know

B. Somewhere you are

C. Something you have

D. Something you see

Answers

1. B. This is most likely a retinal scanner. Retinal scanners examine blood

vessel patterns in the back of the eye. Retinal scanning must be done at

short distances; the user has to be right at the device for it to work.

2. A. Your supervisor just defined the false rejection rate (FRR) for your

system. The FRR is the level of false negatives, or rejections, that are going

to be allowed in the system. In this case, your supervisor is willing to accept

one false rejection for every 50 attempts.

3. C. The Time-based One-Time Password (TOTP) algorithm is a specific

implementation of an HOTP that uses a secret key with a current timestamp

to generate a one-time password. Note that timestamp is the key clue in the

question.

4. B. An access token is a physical object that identifies specific access rights,

and in authentication it falls into the “something you have” factor category.

5. D. An iris scan would be considered a biometric technique and is not a

hardware token. A hardware token is a physical item the user must be in

possession of to access their account or certain resources.

6. C. The bank employees are using proximity cards, which are contactless

access cards that provide information to the electronic door control system.

Proximity cards just need to be close enough to the scanner to work—they

do not need to actually touch the scanner.

7. B. This is an example of a false positive. A false positive occurs when a

biometric is scanned and allows access to someone who is not authorized.

8. A. Static codes can be captured and replayed and are not well suited for

systems with active users.

9. C. The crossover error rate (CER) is the rate where both accept and reject

error rates are equal. This is the desired state for the most efficient

operation of a biometric system, and it can be managed by manipulating the

threshold value used for matching.

10. D. Something you see is neither a factor (something you know, something

you have, or something you are) nor an attribute (somewhere you are,

something you can do, something you exhibit, or someone you know).

CHAPTER 13

Cybersecurity Resilience

In this chapter, you will

• Examine the elements that create redundancy

• Understand the types of backups and the roles they play in resilience

Systems are designed to operate for a purpose, and we use the term risk to

describe the outcomes when issues degrade performance from an optimal state.

For a variety of reasons, expecting a system to run flawlessly, at all times, is

unreasonable. We can put defenses in place to mitigate the issues that occur

when a system is degraded, but it still raises the question of how one gets back to

full performance. This is where resiliency comes in. A resilient system is one

that can return to proper operating conditions after having something go wrong.

And in today’s increasingly hostile environment, this is an important security

measure.

Certification Objective This chapter covers CompTIA Security+ exam

objective 2.5: Given a scenario, implement cybersecurity resilience.

Redundancy

Redundancy is the use of multiple, independent elements to perform a critical

function, so that if one element fails, there is another that can take over the work.

When developing a resiliency strategy for ensuring that an organization has what

it needs to keep operating, even if hardware or software fails or if security is

breached, you should consider other measures involving redundancy and spare

parts. Some common applications of redundancy include the use of redundant

servers, redundant connections, and redundant ISPs. The need for redundant

servers and connections may be fairly obvious, but redundant ISPs may not be

so, at least initially. Many ISPs already have multiple accesses to the Internet on

their own, but by having additional ISP connections, an organization can reduce

the chance that an interruption of one ISP will negatively impact the

organization. Ensuring uninterrupted access to the Internet by employees or

access to the organization’s e-commerce site for customers is becoming

increasingly important.

Many organizations don’t see the need for maintaining a supply of spare

parts. After all, with the price of storage dropping and the speed of processors

increasing, why replace a broken part with older technology? However, a ready

supply of spare parts can ease the process of bringing the system back online.

Replacing hardware and software with newer versions can sometimes lead to

problems with compatibility. An older version of some piece of critical software

may not work with newer hardware, which may be more capable in a variety of

ways. Having critical hardware (or software) spares for critical functions in the

organization can greatly facilitate maintaining business continuity in the event of

software or hardware failures.

EXAM TIP Redundancy is an important factor in both security and reliability.

Make sure you understand the many different areas that can benefit from

redundant components.

Geographic Dispersal

An important element to factor into the cost of the backup strategy is the expense

of storing the backups. A simple strategy might be to store all your backups

together for quick and easy recovery actions. This is not, however, a good idea.

Suppose the catastrophic event that necessitated the restoration of backed-up

data was a fire that destroyed the computer system the data was processed on. In

this case, any backups that were stored in the same facility might also be lost in

the same fire.

The solution is to keep copies of backups in separate locations. The most

recent copy can be stored locally, as it is the most likely to be needed, while

other copies can be kept at other locations. Depending on the level of security

your organization desires, the storage facility itself could be reinforced against

possible threats in your area (such as tornados or floods). A more recent advance

is online backup services. A number of third-party companies offer high-speed

connections for storing data in a separate facility. Transmitting the backup data

via network connections alleviates some other issues with physical movement of

more traditional storage media, such as the care during transportation (tapes do

not fare well in direct sunlight, for example) or the time that it takes to transport

the tapes.

Disk

Disks are the primary storage mechanism in a system, whether composed of

physical hard drives with spinning platters or solid-state memory devices. The

term disk refers to the spinning platter historically, but more and more storage is

being handled by solid-state memory. Also, the logical construct of a disk can be

mapped across multiple physical storage elements.

Redundant Array of Inexpensive Disks (RAID) Levels

A common approach to increasing reliability in disk storage is employing a

redundant array of inexpensive disks (RAID). RAID takes data that is normally

stored on a single disk and spreads it out among several others. If any single disk

is lost, the data can be recovered from the other disks where the data also

resides. With the price of disk storage decreasing, this approach has become

increasingly popular to the point that many individual users even have RAID

arrays for their home systems. RAID can also increase the speed of data

recovery as multiple drives can be busy retrieving requested data at the same

time instead of relying on just one disk to do the work.

Several different RAID approaches can be considered:

• RAID 0 (striped disks) simply spreads the data that would be kept on the

one disk across several disks. This decreases the time it takes to retrieve

data because the data is read from multiple drives at the same time, but it

does not improve reliability because the loss of any single drive will result

in the loss of all the data (since portions of files are spread out among the

different disks). With RAID 0, the data is split across all the drives with

no redundancy offered.

• RAID 1 (mirrored disks) is the opposite of RAID 0. RAID 1 copies the

data from one disk onto two or more disks. If any one disk is lost, the data

is not lost since it is also copied onto the other disk(s). This method can be

used to improve reliability and retrieval speed, but it is relatively

expensive when compared to other RAID techniques.

• RAID 2 (bit-level error-correcting code) is not typically used, as it stripes

data across the drives at the bit level as opposed to the block level. It is

designed to be able to recover the loss of any single disk through the use

of error-correcting techniques.

• RAID 3 (byte-striped with error check) spreads the data across multiple

disks at the byte level with one disk dedicated to parity bits. This

technique is not commonly implemented because input/output operations

can’t be overlapped due to the need for all to access the same disk (the

disk with the parity bits).

• RAID 4 (dedicated parity drive) stripes data across several disks but in

larger stripes than in RAID 3, and it uses a single drive for parity-based

error checking. RAID 4 has the disadvantage of not improving data

retrieval speeds since all retrievals still need to access the single parity

drive.

• RAID 5 (block-striped with error check) is a commonly used method that

stripes the data at the block level and spreads the parity data across the

drives. This provides both reliability and increased speed performance.

This form requires a minimum of three drives.

RAID 0 through 5 are the original techniques, with RAID 5 being the most

common method used, as it provides both the reliability and speed

improvements. Additional methods have been implemented, such as duplicating

the parity data across the disks (RAID 6) and a stripe of mirrors (RAID 10).

Some levels can be combined to produce a two-digit RAID level. RAID 10,

then, is a combination of levels 1 (mirroring) and 0 (striping), which is why it is

also sometimes identified as RAID 1 + 0. Mirroring is writing data to two or

more hard disk drives (HDDs) at the same time—if one disk fails, the mirror

image preserves the data from the failed disk. Striping breaks data into “chunks”

that are written in succession to different disks.

EXAM TIP Knowledge of the basic RAID structures by number designation is

a testable element and should be memorized for the exam. RAID 0 and RAID 1

both require a two-drive minimum. Both RAID 3 and RAID 5 have a three-drive

minimum. RAID 10 (also called 1+0) requires four drives at a minimum.

Multipath

Between the storage systems and the server/computer is an I/O interface. This

I/O interface converts the information from the computer to a form that works

for the specific storage system. There are different interfaces for different types

of storage systems (for example, RAID, SCSI, Fiber Channel, and SATA), each

designed to deal with the necessary data transfers. When a storage element is

connected by multiple adapters, this provides redundancy in the event of a

problem with one of the adapters. This is referred to as a multipath connection

and is commonly employed in high-reliability servers and critical systems.

Figure 13-1 shows a server with two host bus adapters (HBAs), along with two

storage area network (SAN) switches and two RAID controllers. This provides

two independent paths from server to data.

Figure 13-1 Multipath configuration of a RAID device to a server

Network

A network is the infrastructure element that connects all the IT components in

the enterprise. A network can serve as a point of failure, or it can be a system of

redundant connections that can be resilient under various traffic loads and

connectivity conditions. Having a properly architected network that has multiple

independent pathways and infrastructure elements designed to increase

redundancy is important. Two major elements to consider are load balancers and

network interface card (NIC) teaming to remove some of the common modes of

network-related traffic failures.

Load Balancers

Certain systems, such as servers, are more critical to business operations and

should therefore be the object of fault-tolerance measures. A common technique

used in fault tolerance is load balancing through the use of a load balancer,

which moves loads across a set of resources in an effort not to overload

individual servers. This technique is designed to distribute the processing load

over two or more systems. It is used to help improve resource utilization and

throughput but also has the added advantage of increasing the fault tolerance of

the overall system since a critical process may be split across several systems.

Should any one system fail, the others can pick up the processing it was

handling. While there may be an impact to overall throughput, the operation

does not go down entirely. Load balancing is often utilized for systems handling

websites, high-bandwidth file transfers, and large Internet Relay Chat (IRC)

networks. Load balancing works by a series of health checks that tell the load

balancer which machines are operating and by a scheduling mechanism that

spreads the work evenly. Load balancing is best for stateless systems, as

subsequent requests can be handled by any server, not just the one that processed

the previous request.

Network Interface Card (NIC) Teaming

If a server has multiple network interface cards (NICs) connecting it to a switch

or router, it will have multiple addresses, one for each NIC. NIC teaming is an

alternative means of connecting used by servers that have multiple network

interface cards and wish to enjoy the benefits of load balancing, fault tolerance,

and failover without requiring added infrastructure to do it. When NIC teaming

is used, the OS combines the NICs into a virtual NIC from the OS perspective. If

one or more of the connections have traffic issues or connectivity issues, the

other NICs can carry the load. Using NIC teaming allows your server to have

redundancy and increased bandwidth, even in the event any of your physical

adapters or cabling fails.

EXAM TIP NIC teaming groups multiple NICs together to form a logical

network device called a bond. This provides for load balancing and fault

tolerance.

Power

Power is required for all machines to operate, and having a reliable and resilient

source of electrical power is critical for continued operations of enterprise

computing. Servers and networking equipment are always on, and even

occasional outages due to equipment failures need to be planned for and

managed to provide an appropriate level of service. In a modern enterprise,

equipment such as uninterruptible power supplies, generators, dual supplies, and

managed power distribution all support having the proper levels of electricity

available to the networking equipment all the time.

Uninterruptible Power Supply (UPS)

Uninterruptible power supplies (UPSs) are power supply systems that can

function using a temporary battery backup in the event of a power failure. UPSs

typically do not have sufficient battery backup capability to last for long, but

they are designed to keep equipment running while backup power, such as from

a generator, is connected. In an enterprise system, most UPSs are designed and

rated for typically 20 minutes of runtime. This is enough time for the backup

generators to start, or in the event that power cannot be restored, for the servers

to transfer to a secondary site as part of a continuity of operations plan and then

to gracefully shut down.

Generator

Backup generators are used to provide power when normal sources of electricity

are lost. The power for these devices is either natural gas or diesel, and they

produce sufficient electrical power to cover the desired services during a power

outage. Generators come with a host of requirements, including maintenance and

testing, and they require significant electrical architecture work to isolate the

desired circuits. The objective typically isn’t to provide power to everything that

the normal power source supplies, as the scale of generation can be costly in

some instances. The circuits energized by the backup generator are separate

circuits that provide power to the desired components. Sizing of the backup

generator is done with respect to the load, and because of the physical

infrastructure, it is not easy or cost efficient to continuously resize the backup

power. The other issue is that when long-term use happens in the case of diesel

generators, a resupply of fuel needs to be managed. As these systems are

typically used during natural disasters, having contracts in place that function

during the disaster are important for refueling and maintenance operations.

Dual Supply

Individual pieces of equipment have power supplies in them to convert the line

power in the facility to the voltages and currents used by the devices. The

individual power supplies are one of the weakest links in a piece of equipment,

as they tend to fail at a much higher rate than the lower-voltage equipment that

they supply power to; hence, a plan needs to be in place for these when they fail.

In cases where a minor outage is okay, having spare power supplies that can be

replaced offline works.

For servers and other pieces of critical infrastructure, however, having a

redundant, dual-supply system is essential. A dual supply is a system where two

independent power supply units, either capable of handling the load, are used. In

the event that either supply is lost, the other continues to carry the load.

Typically, these devices are also made to be hot swappable, so in the event of a

failure, the bad supply can be replaced without powering down the unit.

Managed Power Distribution Units (PDUs)

A managed power distribution unit (PDU) is a device designed to handle the

electrical power for server racks. A fully populated server rack can use as much

as 30kVA, or 10 times the amount of electricity for a typical home. This is why

server rooms need special HVAC to handle the heat distribution, and they use

managed power distribution units to efficiently handle the electrical side. A PDU

can take three-phase 440/240VAC power in and convert it to either single-phase

110VAC or 48VDC power. The objective of a PDU is to efficiently convert the

power, and manage the heat from the conversion, while producing a power flow

that is conditioned from spikes and over/under voltage conditions. Most PDUs

offer extensive monitoring capability, so entire racks of servers can have their

power remotely monitored for conditions that might cause issues.

EXAM TIP Given a scenario, you should understand how each device is used

to manage power and provide cybersecurity resilience.

Replication

Replication is a simple form of redundancy—that is, having another copy of

something should something happen to the original. Dual power supplies

replicate the power. Having a redundant array of disks to store data is another

form of replication, as are backups and having offsite alternate operations for

business continuity purposes. In situations where having something specific is

essential, hearing someone say “two is one, and one is none” is a commonplace.

Common ways of seeing replication in everyday enterprise operations include

the use of storage area networks and virtual machine technologies.

Storage Area Network (SAN)

A storage area network (SAN) is a dedicated network that connects compute

elements to storage elements. This network can be optimized for the types of

data storage needed, in terms of size and data rates, in terms of format, and in

terms of access criteria. Having the old-school model of the data stored on disks

attached directly to a machine represents a failure mode when the machine fails.

It also has issues when scaling to large quantities such as enterprise databases

with multiple users. The SAN resolves this point of failure by making the data

storage independent of any individual computer and can even interface to

multiple redundant storage systems to allow redundancy on the side of data

storage as well.

VM

Virtual machine (VM) technologies can enable replication of processing units

that can be manipulated between different computers. Having a website with

multiple identical servers to handle the load still has the issues associated with

failures of individual servers and the rebuilding of those server components on

the software side. VM technology resolves that by allowing multiple copies of a

specific instance to be used on different hardware and with centralized

monitoring and management. Need an extra web server because of the current

system load? Just start another VM. It takes seconds as opposed to provisioning

and building a server—a process measured in hours or days. VMs have

revolutionized corporate computing operations because they allow

administrators to manage the compute side easily by pointing and clicking to add

or remove capacity and processing power using known-good images. Should a

particular instance be modified by an unauthorized actor, a known-good

replacement image can be established quickly, returning capacity to the

enterprise. Also, the proper deployment of VMs and server technologies can

provide hardware independence for specific operating images, enabling efficient

use of server resources.

On-premises vs. Cloud

When you’re examining redundancy, one factor to consider is location. Is the

work going to happen on the premises or is it being done in the cloud? By itself,

this does not provide redundancy, but once the determination of location is

made, then factors can be employed to ensure appropriate levels of redundancy

based on risk.

Backup Types

A key element in business continuity/disaster recovery (BC/DR) plans is the

availability of backups. This is true not only because of the possibility of a

disaster but also because hardware and storage media will periodically fail,

resulting in loss or corruption of critical data. An organization might also find

backups critical when security measures have failed and an individual has gained

access to important information that may have become corrupted or at the very

least can’t be trusted. Data backup is thus a critical element in these plans, as

well as in normal operation. There are several factors to consider in an

organization’s data backup strategy:

• How frequently should backups be conducted?

• How extensive do the backups need to be?

• What is the process for conducting backups?

• Who is responsible for ensuring backups are created?

• Where will the backups be stored?

• How long will backups be kept?

• How many copies will be maintained?

Keep in mind that the purpose of a backup is to provide valid, uncorrupted

data in the event of corruption or loss of the original file or the media where the

data was stored. Depending on the type of organization, legal requirements for

maintaining backups can also affect how it is accomplished.

There are four main forms of backups: full, incremental, differential, and

snapshot. Each of these has advantages and disadvantages in terms of time to

back up and restore as well as complexity. These backup types are described in

the upcoming sections.

Understanding the purpose of the archive bit is important when you read

about the backup types. The archive bit is used to indicate whether a file has (1)

or has not (0) changed since the last backup. The bit is set (changed to a 1) if the

file is modified, or in some cases, if the file is copied, the new copy of the file

has its archive bit set. The bit is reset (changed to a 0) when the file is backed up.

The archive bit can be used to determine which files need to be backed up when

using methods such as the differential backup method.

EXAM TIP When learning about the following backup types, be sure to pay

attention to the details concerning how many backups are needed for a restore.

Here’s a typical exam question: “With this type of backup (differential or

incremental) and a seven-day backup scheme, how many backup tapes are

needed for a restore?” Note that this is not a simple case of memorization

because you need the details from the scenario to answer the question. Also, you

need to know the “order of restoration” of the backups.

Full

The easiest type of backup to understand is the full backup. In a full backup, all

files and software are copied onto the storage media. Restoration from a full

backup is similarly straightforward—you must copy all the files back onto the

system. This process can take a considerable amount of time. Consider the size

of even the average home PC today, for which storage is measured in tens and

hundreds of gigabytes. Copying this amount of data takes time. In a full backup,

the archive bit is cleared.

EXAM TIP A full backup copies all data and clears/resets the archive bit. This

process takes considerable time to complete but allows for a complete restore

with one tape.

Incremental

The incremental backup is a variation on a differential backup, with the

difference being that instead of copying all files that have changed since the last

full backup, the incremental backup backs up only files that have changed since

the last full or incremental backup occurred, thus requiring fewer files to be

backed up. With incremental backups, even less information will be stored in

each backup. Just as in the case of the differential backup, the incremental

backup relies on the occasional full backup being accomplished. After that, you

back up only files that have changed since the last backup of any sort was

conducted. To restore a system using this type of backup method requires quite a

bit more work. You first need to go back to the last full backup and reload the

system with this data. Then you have to update the system with every

incremental backup that has occurred since the full backup. The advantage of

this type of backup is that it requires less storage and time to accomplish. The

disadvantage is that the restoration process is more involved. Assuming that you

don’t frequently have to conduct a complete restoration of your system,

however, the incremental backup is a valid technique. An incremental backup

will clear the archive bit.

EXAM TIP To perform a restore from incremental backup, you need the last

full backup and every incremental tape since the last full backup.

Snapshot

A snapshot is a copy of a virtual machine at a specific point in time. A snapshot

is created by copying the files that store the virtual machine. One of the

advantages of a virtual machine over a physical machine is the ease with which

the virtual machine can be backed up and restored—the ability to revert to an

earlier snapshot is as easy as clicking a button and waiting for the machine to be

restored via a change of the files.

Differential

In a differential backup, only the files that have changed since the last full

backup was completed are backed up. This also implies that periodically a full

backup needs to be accomplished. The frequency of the full backup versus the

interim differential backups depends on your organization and needs to be part of

your defined strategy. Restoration from a differential backup requires two steps:

the last full backup first needs to be loaded and then the last differential backup

performed can be applied to update the files that have been changed since the

full backup was conducted. Again, this is not a difficult process, but it does take

some time. The amount of time to accomplish the periodic differential backup,

however, is much less than that for a full backup, and this is one of the

advantages of this method. Obviously, if a lot of time has passed between

differential backups, or if most files in your environment change frequently, then

the differential backup does not differ much from a full backup. It should also be

obvious that to accomplish the differential backup, the system has to have a

method to determine which files have been changed since some given point in

time. The archive bit is not cleared in a differential backup since the key for a

differential is to back up all files that have changed since the last full backup.

EXAM TIP To perform a restore from differential backup, you need the last

full backup and the most recent differential backup tape.

The amount of data that will be backed up, and the time it takes to accomplish

this, has a direct bearing on the type of backup that should be performed. The

following table outlines the three basic types of backups that can be conducted,

the amount of space required for each, and the ease of restoration using each

strategy.

There are times when each of these methods makes sense. If you have a large

amount of data, but most is static (changes slowly if ever), then the small

changes are best captured with differentials. If the whole data structure is

changing, then full backups make more sense. Understanding the data is part of

the key to understanding the correct mechanisms for backup and restore.

Tape

Tape drives are an older form of data storage mechanism, and they are

characterized by their sequential read/write access. A disk allows you to directly

access specific elements randomly, whereas a tape system stores everything in

one long structure, requiring you to physical move the tape if you wish to access

an element halfway through the storage. For general-purpose storage, this

sequential access mechanism tends to create significant performance issues. But

for backups and restores, these operations are sequential in nature, and thus tape

is still well suited for this type of operation. For bulk storage of backups, tape is

still a viable alternative in terms of cost and performance.

Disk

The term disk refers to either a physical hard drive with spinning platters or a

solid-state memory device. Backing up a disk is a common operation for a single

computer because most computers have very few disks, and this is a logical

structure to maintain and restore. For client-based PCs, a disk backup can make

sense, and many systems can perform a full, incremental, snapshot, or

differential backup of a disk.

Copy

Copying is the simplest form of backup for a file or set of files. Users can use

this option with ease, as the scope of their data backup requirement is typically

small (for example, saving a copy of a critical document or an important

picture). However, this method breaks down when the scope expands to larger

and larger sets of data, and for large-scale backups, one of the previous methods

is more efficient both for backing up and restoring. One of the advantages of

having users make copies of critical documents is the ability to do a quick

restore in the event of an overwrite error.

Network Attached Storage (NAS)

Network attached storage (NAS) is the use of a network connection to attach

external storage to a machine. This is a simple method of extending storage, and

the connection can be managed over either a USB connection or the Ethernet

network connection. In either case, NAS is a simple extension of data storage to

an external system, and typically these devices do not transfer data fast enough

for regular operations. However, they do work well as an external site for databackup-and-recover solutions on a smaller, single-machine scale.

Storage Area Network (SAN)

As mentioned previously, a storage area network (SAN) is a dedicated network

that connects compute elements to storage elements. This network can be

optimized for the types of data storage needed, in terms of size and data rates, in

terms of format, and in terms of access criteria. Using a SAN as part of a backup

solution is a good example of using technology to solve complex problems.

Multiple different servers across the enterprise can connect via a SAN to a

backup array, enabling efficient and effective backups in a manageable and

flexible manner.

EXAM TIP NAS is a single storage device that serves files over the network to

a machine. It’s a simple form of external storage. A SAN, on the other hand, is a

network of multiple devices designed to manage large and complex sets of data

in real time at processor speed.

Cloud

Just as NAS and SANs can be used as locations to store data backups, so can the

cloud. Numerous cloud-based backup security vendors and products place the

data storage of a backup in the cloud. The advantages are all of the cloud

advantages: offsite, can have multiple redundant copies, and available via the

Web for recovery. The disadvantages are the same: the backup is on another box

and it is protected only by the legal agreement between the user and the backup

vendor. Also, these contracts tend to favor the backup vendor, not the client. So

while the cloud can result in less on-premises administration of data, it can

increase security concerns because someone else is protecting the data, under the

guidelines of a contractual document that may or may not reflect current risk

postures.

It is important to realize that cloud storage has invaded the desktop of many

users. The wide range of basic cloud sync providers includes Dropbox, Box,

Microsoft OneDrive, Google Drive, and iCloud, as well as many lesser-known

entities. Understanding the risk associated with data in these situations matters in

a corporate environment because what might be convenient or seem like a good

idea from a user perspective might put data at risk of disclosure.

Image

An image-based backup is a specific structure of the backup file to match that of

the system being backed up. This may take more time and space, but it is also

guaranteed not to miss anything because it backs up everything, including the

deleted data and free space. For critical systems, this provides a complete

capture of the system as it was at the time of backup, including all nonpersistent

data associated with the OS. Image backups can provide extra levels of

assurance when certain types of failures (due to a malware attack, for example)

leave a system unusable.

Online vs. Offline

Online backups are those that are stored in a location that is accessible via the

Internet. This provides flexibility in recovery by making the backup data

available anywhere there is a network connection. Offline backups are those

stored on an offline system that is not accessible via the Internet. Online backups

have the advantage of providing geographic separation of the backups from the

original system.

Offsite Storage

Offsite backups are ones stored in a location separate from the system being

backed up. This can be important in regard to problems that affect an area larger

than a single room. A building fire, a hurricane, a tornado—these are all disasters

that occur frequently and typically affect more than just a single room or

building. Having backups offsite alleviates the risk of losing the backups to the

same problem. In today’s high-speed network world with cloud services, storing

backups in the cloud is an option that can resolve many of the risks and issues

associated with backup availability.

Distance Considerations

The distance associated with an offsite backup is a logistics problem. If you need

to restore a system and the backup is stored hours away by car, that increases the

recovery time. The delay resulting from physical movement of backup tapes has

been alleviated in many systems through networks that move the data at the

speed of the network. Distance is also critical when examining the reach of a

disaster. It is important that the offsite location is far enough away that it is not

affected by the same incident. This includes the physical location of a cloud

storage provider’s servers. If your business is in Puerto Rico and so is your cloud

provider’s servers, for example, Hurricane Maria likely made your data

unavailable for a long time.

Nonpersistence

Nonpersistence refers to system items that are not permanent and can change.

An example of something that is nonpersistent is the registry in Microsoft

Windows, which is a dynamic list of configuration criteria. Nonpersistence needs

to be appropriately managed, and systems that have this characteristic typically

have mechanisms built in to manage this diversity. For VMs, where the current

state of the system is continually changing, and thus the image is changing, we

have snapshots. Snapshots provide a copy of the system at a point in time that

you can then persist to use as a recovery point or backup.

EXAM TIP In the event of a failure in a nonpersistent system, all data is lost.

The resilience and recovery of those conditions must occur from external

sources. Think memory when you turn off your computer; it must be reloaded

when you restart.

Revert to Known State

Things eventually go wrong, and when something goes wrong, you want to

recover to a known point. Having the ability to recover to a known state is

referred to as reverting to a known state. Modern OSs are a prime example of

nonpersistence; they are regularly changing with new data, new software, new

configurations, new drivers, and so on. While data backups can bring back the

data elements of a system, bringing back the configuration of a system, including

driver files and patches, is more complicated. How do you recover a system after

a patch goes awry, or a new driver brings a level of instability? Many OSs have

the ability to roll back to a previous known configuration: both servers and

desktops can be rolled back, restoring the system to a previous point in time

while leaving the files intact—back to a condition where the OS previously

worked properly.

Last Known-Good Configuration

When you have a system without persistence, you need a means to recover to a

known-good state. On a boot failure, Microsoft Windows can give you an option

to revert to the last known-good configuration, which is a means of reverting to a

known state. In Windows 7, this was a direct menu option. In Windows 10, this

option is buried under the Windows Recovery system. The methods of accessing

it vary based on the type of issue, and whether or not you can get into Windows

itself. If Windows fails on three subsequent boots in sequence, it will present you

with recovery options rather than trying to boot again.

Live Boot Media

One means of beginning with a known configuration and a known state is to

boot to live boot media, which is a bootable flash drive or DVD source that

contains a complete bootable image of the OS. Using this as a means of starting

in a known state is common in digital forensics investigations.

High Availability

One of the objectives of security is the availability of data and processing power

when an authorized user desires it. High availability refers to the ability to

maintain the availability of data and operational processing (services) despite a

disrupting event. Generally this requires redundant systems, both in terms of

power and processing, so that should one system fail, the other can take over

operations without any break in service. High availability is more than data

redundancy; it requires that both data and services be available.

EXAM TIP Fault tolerance and high availability are similar in their goals, yet

they are separate in application. High availability refers to maintaining both data

and services in an operational state, even when a disrupting event occurs. Fault

tolerance is a design objective to achieve high availability should a fault occur.

Scalability

Scalability is a design element that enables a system to accommodate larger

workloads by adding resources either making hardware stronger (scaling up) or

adding additional nodes (scaling out). This term is commonly used in server

farms and database clusters, as these both can have scale issues with respect to

workload. Both elasticity and scalability have an effect on system availability

and throughput, which can be significant security- and risk-related issues.

EXAM TIP Elasticity and scalability seem to be the same thing, but they are

different. Elasticity is related to dynamically scaling a system with workload

(scaling out), whereas scalability is a design element that enables a system both

to scale up to more capable hardware and to scale out to more instances.

Restoration Order

Data restoration operations are designed to take an alternative copy of the data

and put it back into a working system. If you back up a database and then later

need to use the backup to restore the database, this is data restoration. But the

order of restoration can make a difference. If you have a large database that takes

days to back up and restore, then having a backup solution that allows you to

restore the selected parts most needed faster can be a lifesaver. This is not just a

technology issue; it requires planning and coordination because the most

important data needs to be identified and then backed up in a manner that

facilitates its quick restore. Developing a restoration plan, along with an order of

what needs to be restored first, second, and so on, is important because this will

drive certain operations when backing up the data in the first place.

Diversity

Most failures come from a series of common causes, either in the environment or

the equipment. If you have a bunch of identical equipment, the advantage is you

can have spares for the commonly known issues. The disadvantage is that these

commonly known problems tend to affect all of the systems. Having a

monoculture of all OSs being identical adds efficiency to patching, but it also

adds risk in common failure modes across the entire enterprise. Having diversity

in technologies, vendors, processes, and controls can assist in resiliency through

differences in failure modes. The virus that hurts one OS typically has no effect

on another. Building diversity into systems to allow parallel operations using

different technologies, vendors, processes, and controls can provide a means to

continue operation even during times of systems failures.

Technologies

The security industry has multiple technologies that can be employed across the

enterprise in an effort to mitigate security risk. Employing the concept of defense

in depth, it is best practice not to use a single technology, but to use several

different technologies in an overlapping fashion, forcing an attacker to bypass

them all to achieve their objective. Having firewalls, ACLs, bastion hosts in a

screened subnet (DMZ), and network monitoring is an example of multiple

technologies designed to detect unauthorized network activity. Having a diverse

set of these elements improves the chances of catching an attacker, even when

they can beat one or two control elements.

Vendors

Different vendors approach security problems with different methodologies,

different toolsets, different policies and procedures, and different technologies.

Adversaries have developed methods of beating different vendors, but if multiple

vendors are brought into play, this makes it all that much more difficult for an

adversary to bypass all of the employed options. Having diversity in the vendors

used for security prevents vendor-specific forms of single points of failure and

creates a more robust set of defensive capabilities.

Crypto

For cryptographic solutions to work, both sides must agree on algorithms, keys,

and other parameters, but diversity can still exist in this environment. A prime

example is in the TLS cipher suite, a set of different crypto protocols,

preassigned to facilitate flexibility in establishing a connection. When you

establish a TLS-enabled connection, the server and client both negotiate a

selection of protocol parameters from the preassigned list, enabling a secure

connection to be established. The same server with a different host, doing the

same exercise, can end up with different crypto choices, but in the end it’s still a

secure connection. Having multiple options configured and available enables the

removal of one if something affects it, while still providing a means of

connecting via alternative options.

Controls

Defense in depth is a security principle where multiple layers of different

security mechanisms are used to ensure catching a risk. This is the use of

diversity in controls. Modern networks employ not just a firewall but also a

screened subnet (DMZ), bastion hosts, and ACLs, all working together in a

coordinated fashion to make bypassing the entire suite of controls nearly

impossible.

EXAM TIP Diversity is about having multiple different sets of controls to

provide for risk mitigation. Diversity should be practiced in all aspects and used

to enhance security. A performance-based question that considers diversity

should be examined in light of which element is most efficient to manipulate—

technologies, vendors, crypto, or controls—and the answer is found in the

specifics of the question.

Chapter Review

In this chapter, you became acquainted with the aspects of cybersecurity

resilience. The chapter opened with an examination of elements that lead to

redundancy, such as geographic dispersal; disks, including RAID and multipath

solutions; network redundancy from load balancers and NIC teaming; and power

supply concerns, including UPSs, generators, dual supplies, and PDUs. The

issues around replication using both SAN and VMs were covered, as was a

comparison of on-premises and the cloud.

The topic of backups was covered, including the backup methodologies of

full, incremental, snapshot, and differential. Backup technologies of tape, disk,

copy, NAS, SAN, cloud, and image were presented as well. An examination of

online versus offline backup locations was provided, including offsite storage

and distance considerations.

The issues associated with nonpersistence, including reverting to a known

state, last known-good configuration, and live boot media, were covered. An

examination of high availability and scalability as well as restoration order

followed. The chapter concluded with a consideration of diversity and how

diversity in technology, vendors, crypto, and controls can be used to mitigate

risk.

Questions

To help you prepare further for the exam, and to test your level of preparedness,

answer the following questions and then check your answers against the correct

answers at the end of the chapter.

1. Which backup strategy includes only the files and software that have

changed since the last full backup?

A. Incremental

B. Full

C. Snapshot

D. Differential

2. Which backup strategy focuses on copies of virtual machines?

A. Incremental

B. Full

C. Snapshot

D. Differential

3. When discussing location for storage of backups, which of the following

statements are true? (Choose all that apply.)

A. The most recent copy should be stored offsite, as it is the one that is

most current and is thus the most valuable.

B. Offsite storage is generally not necessary, except in cases where the

possibility of a break-in at the main facility is high.

C. Offsite storage is a good idea so that you don’t lose your backup to the

same event that caused you to lose your operational data and thus need

the backup.

D. The most recent copy can be stored locally, as it is the most likely to be

needed, while other copies can be kept at other locations.

4. To deal with nonpersistence in a system, which of the following items offer

risk mitigation? (Choose all that apply.)

A. Image backups

B. Cloud

C. Last known-good configuration

D. Revert to a known state

5. To have easily available quick backup of critical user documents, which of

the following is recommended for backing these items up?

A. Differential

B. Snapshot

C. Copy

D. NAS

6. You have offices in six locations across town and wish to utilize a common

backup restore methodology. Which would be the most efficient solution

for your small offices?

A. SAN

B. NAS

C. Cloud

D. Offline

7. Which of the following statements is true about redundancy?

A. It prevents failures.

B. It is complicated and expensive to do.

C. It applies only to hardware.

D. It can be done across many systems.

8. What distinguishes high availability systems?

A. The ability to change with respect to usage conditions

B. The ability to process, even in times of disruption

C. Automated backups and recovery functions

D. The use of diversity to mitigate single threats

9. The continual changing of information in a system is referred to as what?

A. Nonpersistence

B. Snapshots

C. Differentials

D. Images

10. A PDU provides management of what in an enterprise?

A. Redundant backup processing

B. Power distribution to servers

C. Improved network connection to data storage

D. Load balancing

Answers

1. D. In a differential backup, only the files and software that have changed

since the last full backup was completed are backed up. The incremental

backup is a variation on a differential backup, with the difference being that

instead of copying all files that have changed since the last full backup, the

incremental backup backs up only files that have changed since the last full

or incremental backup occurred, thus requiring fewer files to be backed up.

In a full backup, all files and software are copied onto the storage media.

Snapshots refer to copies of virtual machines.

2. C. Snapshots refer to copies of virtual machines. The incremental backup is

a variation on a differential backup, with the difference being that instead of

copying all files that have changed since the last full backup, the

incremental backup backs up only files that have changed since the last full

or incremental backup occurred, thus requiring fewer files to be backed up.

In a full backup, all files and software are copied onto the storage media. In

a differential backup, only the files and software that have changed since

the last full backup was completed are backed up.

3. C and D. Offsite storage is a good idea so that you don’t lose your backup

to the same event that caused you to lose your operational data and thus

need the backup. Additionally, the most recent copy can be stored locally,

as it is the most likely to be needed, while other copies can be kept at other

locations.

4. A, C, and D. Image backups capture the nonpersistence of the OS. Also,

reverting to a known state and using the last known-good configuration

both can resolve nonpersistence issues. Cloud (answer B) is not a direct

answer, as by itself, the cloud does not offer persistence to a nonpersistent

system. An image backup has everything, so restoring from it can resolve a

persistence problem. For the cloud to be involved, it would be as a

secondary item (that is, a place to store an image backup), but then it is not

actually directly involved.

5. C. User-managed copies on external media of critical documents can make

it very easy for the end user to manage recovery in a quick manner.

6. C. Cloud backup solutions can be ideal for small offices, and with the

different offices, centralized administration can be added.

7. D. A wide range of options are associated with creating redundant systems

—some as simple as configuration elements and system choices.

8. B. High availability systems continue to process data even when disrupting

events occur.

9. A. Nonpersistence refers to system items such as memory and registry

elements that are not permanent and can change over time, even while

running.

10. B. Power distribution units provide a centralized means of managing and

monitoring the power delivered to servers in a rack.

CHAPTER 14

Embedded and Specialized Systems

In this chapter, you will

• Explore the security implications of embedded systems

• Explore the security implications of smart devices/IoT

• Explore the security implications of SCADA systems

Cybersecurity is not just limited to IT systems in the enterprise. A significant

number of embedded systems and specialized systems produce and consume

data to achieve functionality. These systems require cybersecurity as well if their

functionality is to be protected from adverse risk. This chapter covers the unique

nature of these systems and how that relates to providing protection for them.

Certification Objective This chapter covers CompTIA Security+ exam

objective 2.6: Explain the security implications of embedded and specialized

systems.

Embedded Systems

Embedded systems is the name given to computers that are included as an

integral part of a larger system, typically hardwired in. From computer

peripherals like printers, to household devices like smart TVs and thermostats, to

the car you drive, embedded systems are everywhere. Embedded systems can be

as simple as a microcontroller with fully integrated interfaces (a system on a

chip) or as complex as the dozens of interconnected embedded systems in a

modern automobile. Embedded systems are designed with a single control

purpose in mind and have virtually no additional functionality, but this does not

mean that they are free of risk or security concerns. The vast majority of security

exploits involve getting a device or system to do something it is capable of

doing, and technically designed to do, even if the resulting functionality was

never an intended use of the device or system.

The designers of embedded systems typically are focused on minimizing

costs, with security seldom seriously considered as part of either the design or

the implementation. Because most embedded systems operate as isolated

systems, the risks have not been significant. However, as capabilities have

increased, and these devices have become networked together, the risks have

increased significantly. For example, smart printers have been hacked as a way

into enterprises, and as a way to hide from defenders. Also, when nextgeneration automobiles begin to talk to each other, passing traffic and other

information between them, and begin to have navigation and other inputs

beamed into systems, the risks will increase and security will become an issue.

This has already been seen in the airline industry, where the separation of inflight Wi-Fi, in-flight entertainment, and cockpit digital flight control networks

has become a security issue.

Raspberry Pi

The Raspberry Pi is a highly successful, low-cost, single-board computer.

Millions of these devices have found their way into a wide range of applications

—from use by hobbyists to prototype engineers, and even as production

elements in some cases. This low-cost (less than $50) highly capable computing

device offers a lot of features and connectivity. A quad-core ARM processor, 8

GB of RAM, connectivity via Ethernet, Bluetooth, USB, 2.4GHz and 5GHz WiFi, and host of connectivity options for displays, I/O, and storage all make this a

versatile platform. Securing a Raspberry Pi is similar to securing any other

system. One has to consider the environment in which it will be deployed, how it

is connected to other users, and what data and sensitive information is involved.

Also, remember that this is, in most cases, a full Linux environment that requires

permissions and other basic security elements.

A Raspberry Pi running a science fair project with no connectivity to the Web

has a completely different scope than one that is connected to the Internet and

being used to log sensitive data necessary for production in an enterprise setting.

Determining the risk profiles and addressing them as appropriate is still a needed

and important task.

Field Programmable Gate Arrays (FPGAs)

Field programmable gate arrays (FPGAs) are electronic circuits that are

programmed to perform a specific function. These semiconductor devices are

based around a matrix of configurable logic blocks (CLBs) that are connected

via programmable interconnects, and in essence the logic is programmed before

use. FPGAs are designed to be reprogrammed to the desired functionality

requirements after manufacturing, and they can typically be reprogrammed as

designs of the functionality evolve. Although not as fast as application-specific

integrated circuits (ASICs), which are custom manufactured for specific design

tasks, the programmability and reprogrammability capabilities of FPGAs

provide significant design flexibility. FPGAs and ASICs are found in a lot of

custom devices, where a full-blown computer with an operating system (OS) and

all that it entails is not needed.

Arduino

The Arduino is a single-board microcontroller, not a full-fledged computer like

the Raspberry Pi. The Arduino is simpler, designed to provide computer control

to hardware projects without the overhead of a full computer, OS, and so on.

While a Raspberry Pi is designed as a computer, the Arduino is designed as a

controller, specifically for interfacing with sensors and devices. The Arduino can

respond to sensor levels and actuate devices based on programming that is

loaded onto the device. This coding works when power is applied; if power is

lost, once it is restored, the device can begin functioning again—unlike a

computer, which would have to reboot and start over. Expansion of the Arduino

platform is done via a series of boards called shields that can add specific

functionality in networking, display, data collection, and so on.

EXAM TIP Understand static environments—systems in which the hardware,

OS, applications, and networks are configured for a specific function or purpose.

These systems are designed to remain unaltered through their lifecycle, rarely

requiring updates.

Supervisory Control and Data Acquisition

(SCADA) / Industrial Control System (ICS)

SCADA is an acronym for supervisory control and data acquisition, a system

designed to control automated systems in cyber-physical environments. SCADA

systems have their own smart components, each of which is an example of an

embedded system. Together they form a SCADA system, which can control

manufacturing plants, traffic lights, refineries, energy networks, water plants,

building automation and environmental controls, and a host of other systems. A

SCADA system is also known as a distributed control system (DCS) or an

industrial control system (ICS), depending on the industry and the configuration.

Where computers control a physical process directly, a SCADA system likely is

involved.

Most SCADA systems involve multiple components networked together to

achieve a set of functional objectives. These systems frequently include a

human–machine interface (HMI), where an operator can exert a form of

directive control over the operation of the system under control. SCADA

systems historically have been isolated from other systems, but the isolation is

decreasing as these systems are being connected across traditional networks to

improve business functionality. Many older SCADA systems were air gapped

from the corporate network; that is, they shared no direct network connections.

This meant that data flows in and out were handled manually and took time to

accomplish. Modern systems removed this constraint and added direct network

connections between the SCADA networks and the enterprise IT networks.

These connections increase the attack surface and the risk to the system; the

more they resemble an IT networked system, the greater the need for security

functions.

SCADA systems have been drawn into the security spotlight with the Stuxnet

attack on Iranian nuclear facilities, initially reported in 2010. Stuxnet is malware

designed to attack a specific SCADA system and cause failures, resulting in

plant equipment damage. This attack was complex and well designed, crippling

nuclear fuel processing in Iran for a significant period of time. This attack raised

awareness of the risks associated with SCADA systems, whether connected to

the Internet or not (Stuxnet crossed an air gap to hit its target).

Facilities

SCADA systems find many uses in facilities, ranging from the building

automation systems of the HVAC system, to pumps for water pressure,

escalators and elevators, and fire alarms—the lists just keep going on. Many of

these systems are independent systems where data is collected from sensors and

used for a specific purpose (elevator scheduling based on buttons on floors, for

instance). Others, such as building access controls, locked/secured doors, and

fire alarm systems, may be interconnected to ensure safety under specific

conditions and security under other conditions. Some of these systems are

connected via the Internet for remote monitoring or control. As for all systems,

understanding how to get access to a system and securing those access points is

key for securing this type of SCADA employment.

Industrial

Industrial facilities have some of the same needs as other facilities—the

computer control of various processes, such as security, environmental

monitoring, fire alarms, and more. The key element is to understand that

virtually any facility has a data collection–data response system, whether it’s a

simple HVAC or thermostat system or a more complex system such as a

surveillance or fire alarm system. Each of these systems can stand alone, be

partially integrated with other systems, fully integrated with others, or connected

to the Internet; the combinations are almost endless and are tailored to meet the

requirements of the facility.

Manufacturing

Manufacturing systems add another layer of computer-controlled processes to

the industrial/facility mix—those of the actual manufacturing process itself.

Manufacturing equipment is commonly computer controlled, using devices such

as programmable logic controllers (PLCs), to execute a process-specific set of

instructions based on sensor readings and actuator settings. These systems can be

differentiated by a wide range of specific attributes, but the term SCADA is

commonly used to cover them.

These systems may be connected to the Internet or have outside access for

third-party vendors. Because the SCADA systems that are running your

manufacturing are typically very critical to your enterprise, these systems require

protection from attackers. The standard practice for this is one of strict network

segmentation.

Energy

Energy systems range from electrical to chemical, petroleum, pipelines, nuclear,

solar, hydrothermal, and more. Each of these systems has multiple systems under

computer control, typically using the same types of SCADA components as

other categories already discussed. In the case of energy distribution, such as

pipelines and electricity, a further complication is the distributed nature of these

elements, where they are geographically spread out (in many cases in our

communities). This distribution of components “outside the corporate walls”

adds a unique physical security aspect to these systems.

Logistics

Logistics systems are the systems that move material from point A to point B.

These systems can involve sea, surface (roads and rail), and air transport. There

are two basic elements that will be under control: the transport system itself and

the material being moved.

EXAM TIP When examining SCADA systems, you have three things to worry

about: the value of the information being protected, physical access to the

system, and logical (typically network) access to the data. When examining the

question, parse the question for the specific detail that matters.

Internet of Things (IoT)

The Internet of Things (IoT) is a term used to describe a wide range of devices

that connect directly via the Internet to create a distributed sensor and processing

system to achieve a specific function. As opposed to general-purpose devices,

like computers and networking equipment, IoT devices are purpose built; they

are designed to do a specific task. All these devices have a couple similarities.

They all have a network interface because connectivity is their purpose as a

smart device or a member of the Internet of Things club. On that network

interface is some form of compute platform. With complete computer

functionality now included on the system on a chip platform (covered in a later

section), these tiny devices can have a complete working computer for just a few

dollars in cost. The use of a Linux-type kernel as the core engine makes

programming easier, as the base of programmers is very large. These devices

also can be mass produced and at relatively low cost. The scaling of the software

development over literally millions of units makes costs scalable. Functionality

is king, meaning that security or anything that might impact new expanded

functionality has to take a backseat.

Sensors

Sensors are devices that measure some physical item and return data that can be

used by a system. Sensors come in an endless array of sizes, shapes, and

physical constraints. Sensors can be used to measure temperatures, pressures,

voltages, positions, humidity levels—the list goes on. Sensors can return the data

as either a digital or analog signal. Analog sensors require an analog-to-digital

conversion before the data can be used by a computer, although many interface

boards do this translation automatically. When designing a system, you need to

determine what needs to be measured, over what range, and at what precision, as

well as environmental and other conditions; these factors all shape the

specification for a sensor and determine the cost.

Smart Devices

Smart devices and devices that comprise the IoT have taken the world’s markets

by storm. From key fobs that can track the location of items via GPS, to cameras

that can provide surveillance, to connected household appliances, TVs,

dishwashers, refrigerators, crock pots, washers, and dryers—anything with a

microcontroller now seems to be connected to the Web so that it can be

controlled remotely. Artificial intelligence (AI) has also entered into the mix,

enabling even greater functionality, embodied in products such as Amazon Echo,

Google Home, Microsoft Cortana, and Apple Siri. Computer-controlled light

switches, LED light bulbs, thermostats, and baby monitors—the smart home has

become a reality, connecting everything to the Internet. You can carry a key fob

that your front door recognizes, unlocking itself before you get to it. Of course,

the security camera sees you first and alerts the system that someone is coming

up the driveway. The only thing that can be said with confidence about this

revolution is someone will figure out a how and a why to connect virtually

anything to the network.

Wearables

Wearable technologies include everything from biometric sensors measuring

heart rate, to step counters measuring how far one walks, to smart watches that

combine all these functions and more. By measuring biometric signals, such as

pulse rate, and body movements, it is possible to measure fitness and even sleep.

These wearable devices are built using very small computers that run a real-time

operating system, usually built from a stripped-down Linux kernel. As with all

information-containing devices, how does one protect the data? As wearables

learn more and more of your personal data, they become a source of interest for

hackers. Protecting the data is the security objective for these devices.

Things you can do to start protecting personal data include checking the

default settings, checking the privacy settings, turning off location tracking,

reading the privacy policies, and, where possible, using a passcode to protect

your personal information (PI).

Facility Automation

Low-cost sensors in an IoT package offer several advantages, including but not

limited to, network delivery of data, significant data collection capability, and

cost advantages with scale. In large facilities, this means security systems,

HVACs, fire sensors, and so on can provide large-scale coverage, enabling

automation of data collection that used to be done manually via a person walking

around. Automation is more than just remote operation; apps such as IFTTT (If

This Then That) systems can respond to changing conditions and use multiple

indicators, including dates and times. Automation can improve risk because it

removes errors and improves speed of response.

Weak Defaults

Whenever items are manufactured or produced in large quantities, specific

specializations such as default credentials are a challenge. The typical process is

to have default credentials on a device and then expect the user to change them.

This expectation of a user changing credentials commonly results in poor

security. Weak defaults are a condition where default conditions are generally

known, including admin account and password, leaving the system completely

vulnerable to an attacker. But even if the password is changed, in cases where

large numbers of devices are deployed, is it reasonable to expect they all got the

default credentials changed to unique passwords? One of the challenges of IoT

deployment and security is managing literally thousands or millions of devices—

and the credentials.

EXAM TIP The Internet of Things is all about connectivity of low-cost

(relative) items at scale. Deployments of hundreds, thousands, and even millions

of devices have been done, and the data can provide great insights that can only

be seen with data at scale. However, with that scale comes the challenge of

managing and securing the large number of devices.

Specialized Systems

As the name indicates, specialized systems are systems designed for special

purposes. Four primary types of specialized systems targeted by CompTIA are

the systems in medical devices, vehicles, aircraft, and smart meters. Each of

these categories has significant computer systems providing much of the

functionality control for the device, and each of these systems has its own

security issues.

Medical Systems

Medical systems is a very diverse group—from small implantable devices, such

as pacemakers, to multi-ton MRI machines. In between is a wide range of

devices, from those that measure vital signs to those that actually control vital

functions. Each of these has several interesting characteristics, and they all have

an interesting caveat—they can have a direct effect on a human’s life. This

makes security of these devices also a safety function.

Medical devices such as lab equipment and infusion pumps have been

running on computer controls for years. The standard of choice has been an

embedded Linux kernel that has been stripped of excess functionality and

pressed into service in the embedded device. One of the problems with this

approach is how to patch this kernel when vulnerabilities are found. Another,

related problem is that as the base system gets updated to a newer version, the

embedded system stays trapped on the old version. This requires regression

testing for problems, and most manufacturers will not undertake such laborintensive chores.

Medical devices are manufactured under strict regulatory guidelines that are

designed for static systems that do not need patching, updating, or changes. Any

change would force a requalification—a lengthy, time-consuming, and expensive

process. As such, these devices tend to never be patched. With the advent of

several high-profile vulnerabilities, including Heartbleed and Bash shell attacks,

most manufacturers simply recommended that the devices be isolated and never

connected to an outside network. In concept, this is fine, but in reality, this can

never happen, as all the networks in a hospital or medical center are connected.

A recent recall of nearly a half million pacemakers in 2017 for a software

vulnerability that would allow a hacker to access and change the performance

characteristics of the device is proof of the problem. The good news is that the

devices can be updated without removing them, but it will take a doctor’s visit to

have the new firmware installed.

Vehicle Systems

A modern vehicle has not a single computer in it, but actually hundreds of them,

all interconnected on a bus. The controller area network (CAN) bus is designed

to allow multiple microcontrollers to communicate with each other without a

central host computer. Before the CAN bus was invented, individual

microcontrollers were used to control the engine, emissions, transmission,

braking, heating, electrical, and other systems, and the wiring harnesses used to

interconnect everything became unwieldy. Robert Bosch developed the CAN bus

for cars, specifically to address the wiring harness issue, and when first deployed

in 1986 at BMW, the weight reduction was over 100 pounds.

Since 2008, all new U.S. and European cars must use a CAN bus, per SAE

regulations—a mandate engineers have willingly embraced as they continue to

add more and more subsystems. The CAN bus has a reference protocol

specification, but recent auto hacking discoveries have shown several interesting

things. First, in defending allegations that some of its vehicles could suddenly

accelerate without driver action, Toyota claimed that the only way to make a

vehicle accelerate quickly is to step on the gas pedal—that software alone won’t

do it. However, this was proven to be false. Hackers have demonstrated almost

complete control over all functions of the Toyota Prius using computers and

CAN bus commands. Second, every automobile manufacturer has

interpreted/ignored the reference protocol specification to varying degrees.

Finally, as demonstrated by hackers at DEF CON, it is possible to disable cars in

motion, over the Internet, as well as fool around with the entertainment console

settings and other systems.

The bottom line is that, to function properly, newer vehicles rely on multiple

computer systems, all operating semi-autonomously and with very little security.

The U.S. Department of Transportation is pushing for vehicle-to-vehicle

communication technology, so that vehicles can tell each other when traffic is

changing ahead of them. Couple that with the advances in self-driving

technology, and the importance of stronger security in the industry is clear. There

is evidence that this is beginning, that security is improving, but the pace of

improvement is slow when compared to typical computer innovation speeds.

Aircraft Systems

Aircraft also have a significant computer footprint inside, as most modern jets

have what is called an “all-glass cockpit,” meaning the old individual gauges and

switches have been replaced with a computer display that includes a

touchscreen. This enables greater functionality and is more reliable than the

older systems. But as with vehicles, the connecting of all of this equipment onto

busses that are then eventually connected to outside networks has led to a lot of

security questions for the aviation industry. And, as is true of medical devices,

patching the OS for aircraft systems is a difficult process because the industry is

heavily regulated, with strict testing requirements. This makes for systems that,

over time, will become vulnerable as the base OS has been thoroughly explored

and every vulnerability mapped and exploited in non-aviation systems, and these

use cases can port easily to aircraft.

Recent revelations have shown that the in-flight entertainment systems, on

standard Linux distros, are separated from flight controls not by separate

networks, but by a firewall. This has led hackers to sound the alarm over

aviation computing safety.

Smart Meters

Smart meters is the common name for the advanced metering infrastructure, a

program initiated by the Department of Energy to bring the functionality of

remote automation to meters in utilities. Real-time two-way communications,

computing infrastructure to analyze the data, and a whole host of new policies

and procedures to take advantage of the automation have provided a revolution

in utility operations. For electricity, this means real-time (with a granularity

measured in minutes, not a month like previous manual reads) usage data that

enables matching of supply and demand with greater efficiency. For all utilities,

the ability to read meters, change service, disconnect, reconnect, and detect and

manage outages provides cost savings and levels of service never possible with

the old manually managed meters. Managing the large-scale deployment of

infrastructure in a secure fashion requires an extensive cryptographic setup, with

some meters having multiple passwords for different levels of operation.

Multiply this by millions of meters, and this is not a trivial task to manage.

However, there are software packages designed to automate these elements as

well.

EXAM TIP Specialized systems are custom built to serve a purpose, and the

required level of security goes along with the purpose. If the data needs

protecting, then the same problems and solutions used to fix them apply. In most

specialized systems, the risks are significant and cryptographic solutions are

designed into the system to limit access to authorized users.

Voice over IP (VoIP)

Voice over IP—the transmission of voice communications over IP networks—is

now a commonplace method of providing telephone services. VoIP makes

telephone management as easy as an app in the enterprise, but it also brings

security risks and vulnerabilities. VoIP systems require protections from standard

traffic-based attacks such as denial of service, but also need protections from

spoofing. Suppose you get an internal phone call from Ms. Jones, the company’s

CFO, and your screen says “Ms. Jones,” but how do you know who is on the

line? If you have never heard Ms. Jones speak before, do you trust the voice, the

screen, or what? Authentication and the protection of the communication

channels have been the province of the telephone company, but in VoIP there is

no overarching phone company to manage these risks.

Additional risks include outsiders using your VoIP to connect to international

telephony services and offering free phone calls or using your phone service to

robocall people. Just as we have to secure systems like e-mail from outside,

unauthorized users, we need to do the same with VoIP services.

Heating, Ventilation, Air Conditioning

(HVAC)

Building-automation systems, climate-control systems, and HVAC (heating,

ventilation, and air conditioning) systems are all examples of systems that are

managed by embedded systems. Although these systems used to be independent

and stand-alone systems, the rise of hyper-connectivity has shown value in

integrating them. Having a “smart building” that reduces the use of building

resources in accordance with the number and distribution of people inside

increases efficiency and reduces costs. Interconnecting these systems and adding

in Internet-based central control mechanisms does increase the risk profile from

outside attacks. These outside attacks could result in HVAC malfunction or

failure, rendering a major office building uninhabitable due to heat and safety.

Although not specific to the HVAC system in one sense, Target corporation’s

2014 hack was begun when an HVAC vendor was compromised, leading to a

compromise of the Target network and access to its point-of-sale network. The

story of the hack made the news and cost Target hundreds of millions of dollars

and resulted in a significant number of executive changes. Cloud security service

provider Qualys said that its researchers have discovered that most of about

55,000 HVAC systems connected to the Internet over the past two years have

flaws that can be easily exploited by hackers.

Drones

Drones, or unmanned aerial vehicles (UAVs), represent the next frontier of

flight. These machines range from the small drones that hobbyists can play with

for under $300 to full-size aircraft that can fly across oceans. What makes these

systems different from regular aircraft is that the pilot is on the ground, flying

the device via remote control. UAVs have cameras, sensors, and processors to

manage the information, and even the simple hobbyist versions have

sophisticated autopilot functions. Because of the remote connection, UAVs are

networked and operated either under direct radio control (rare) or via a

networked system (much more common).

Multifunction Printers (MFPs)

Multifunction printers (MFPs), which combine a printer, scanner, and fax, have

embedded compute power to act as a print server, manage the actual printing or

scanning process, and allow complete network connectivity. These devices

communicate in a bidirectional fashion, accepting print jobs and sending back

job status, printer status, and other information to the computer. This has

decoupled printing from the computer, making the printer a stand-alone entity.

The system that runs all these functions was designed to provide maximum

functionality for the device, and security is more of an afterthought than a design

element. As such, these devices have been shown to be hackable and capable of

passing malware from the printer to the computer. These attacks still exist

primarily as a proof of concept as opposed to a real-world threat, which is

fortunate, because the current generation of security software does not monitor

printer activity to and from the computer very well.

Real-time Operating Systems (RTOSs)

Real-time operating systems (RTOSs) are designed for devices where the

processing must occur in real time and data cannot be queued or buffered for any

significant length of time. RTOSs are not general-purpose machines but are

programmed for a specific purpose. They still have to deal with contention, and

they have scheduling algorithms to deal with timing collisions, but in general an

RTOS processes each input as it is received, or within a specific time slice

defined as the response time. Examples of RTOSs range from something as

common as an anti-lock braking computer system in a car to something as

complex as a robotic system used on an assembly line.

Most general-purpose computer operating systems are capable of

multitasking by design. This includes Windows and Linux. Multitasking systems

make poor real-time processors, primarily because of the overhead associated

with separating tasks and processes. Windows and Linux may have interrupts,

but these are the exception, not the rule, for the processor. RTOS-based software

is written in a completely different fashion, designed to emphasize the thread in

processing rather than handling multiple threads.

The security implications surrounding real-time operating systems lie in their

timing. Should an event do something that interferes with the system’s ability to

respond within its time allotment, then the system itself can fail in its task. Realtime operating systems also tend to be specific to the degree that updates and

patches tend not to be common, as the manufacturer of the system does not

provide that level of support. As items such as cars become more networked,

these weaknesses are becoming apparent, and one can expect this situation to

change over time.

Surveillance Systems

Digital surveillance systems have entered the computing world through a couple

of different portals. First, there is the world of high-end digital cameras that have

networking stacks, image processors, and even 4K video feeds. These are used

in enterprises such as news organizations, which rely on getting the data live

without extra processing delays. What is important to note is that most of these

devices, although they are networked into other networks, have built-in virtual

private networks (VPNs) that are always on, because the content is considered

valuable enough to protect as a feature.

The next set of cameras reverses the quantity and quality characteristics.

Where the high-end devices are fairly small in number, there is a growing

segment of video surveillance cameras, including cameras for household

surveillance, baby monitoring, and the like. Hundreds of millions of these

devices are sold, and they all have a sensor, a processor, a network stack, and so

forth. These are part of the Internet of Things revolution, where millions of

devices connect together either on purpose or by happenstance. It was a network

of these devices, along with a default username and password, that led to the

Mirai botnet that actually broke the Internet for a while in the fall of 2016. The

true root cause was a failure to follow a networking RFC concerning source

addressing, coupled with the default username and password and remote

configuration that enabled the devices to be taken over. Two sets of fails,

working together, created weeks’ worth of problems.

EXAM TIP VoIP, HVAC, drones/UAVs, MFPs, and surveillance systems have

one weakness in common: access via the Internet. The same vector can be used

against any connected system, and without defenses, these are typically very

insecure systems. They have to be connected for functionality, and hence they

need basic protections like passwords.

System on a Chip (SoC)

System on a chip (SoC) refers to a complete computer system miniaturized on a

single integrated circuit, designed to provide the full functionality of a

computing platform on a single chip. This includes networking and graphics

display. Some SoC solutions come with memory, while others have the memory

separate. SoCs are very common in the mobile computing market (both phones

and tablets) because of their low power consumption and efficient design. Some

SoC brands have become household names because mobile phone companies

have advertised their inclusion in a system, such as the Snapdragon processor in

Android devices. Quad-core and eight-core SoC systems are already in place,

and they even have advanced designs such as quad plus one, where the fifth

processor is slower and designed for simple processes and uses extremely small

amounts of power. So when the quad cores are not needed, there is not

significant energy usage.

The programming of SoC systems can occur at several different levels.

Dedicated OSs and applications can be written for them, such as the Android

fork of Linux, which is specific to the mobile device marketplace. Because these

devices represent computing platforms with billions of devices worldwide, they

have become a significant force in the marketplace. The security implications of

SoC-based systems is associated not with the specifics of SoC, but in the fact

that they are ubiquitous in our technology-driven lives. Security issues are

handled by the device, not the specific SoC aspect itself.

Communication Considerations

Embedded and specialized systems are useful for a purpose, and many times

those purposes require communications across a network for other resources.

The communication considerations for embedded and specialized systems are

dependent on the service, what task it is doing, and the resources needed. The

methods of communication are wide and varied, and the choice is usually

dependent on the range needed and with whom the communication is needed.

For short-distance, local communications, certain technologies can excel. For

worldwide communications, others would work better. Adopting technology

already employed by users has advantages as well; for instance, why use a

specialty radio circuit in an environment that already has Wi-Fi?

5G

5G is the latest generation mobile radio-based network. It is designed to connect

virtually everyone and everything together, including machines, objects, and

devices, with a focus on higher data speeds and bandwidth. 5G networks are

more than just bigger pipes; the standard has many functional elements to

improve both performance and efficiencies. 5G is in the process of being rolled

out worldwide, and connectivity is via a cellular circuit—either a phone, modem,

or a chipset designed in a product.

Just as having a full-blown server is overkill for a simple sensor, 5G may be

overkill for many communication needs. If worldwide range, large bandwidth,

and low latency are important, then 5G may be warranted, but if not, there are

lower-cost alternatives.

Narrow-Band Radio

Narrow-band radio communications use narrow bands of frequencies for lowdata-rate communications. While a low data rate may seem to be a big problem,

not all systems have high-data-rate needs, and narrow-band radio offers

advantages in range and power utilization. Lower-power transmitters are the

norm, as are significantly longer ranges. So, if a company has a bunch of drilling

rigs over a large geographic area and needs to move relative small quantities of

data between them, then narrow-band radio can be the ideal solution.

Baseband Radio

Baseband refers to the original bandwidth produced by a signal. For typical

audio signals, it is 20–20,000 Hz. For a signal to be transmitted over a radio

circuit, it is usually encoded or modulated in a manner that can then be blended

with the radio wave, carrying the information in the changes on a radio wave.

Baseband radio refers to the signal that is being transmitted and represents a

single channel of communication. Broadband radio is when multiple signals are

bundled together for transmission, and equipment is typically required to

separate out the individual communications. Baseband radio, by design, is very

simple, as it only carries a single channel to manage communications across.

Subscriber Identity Module (SIM) Cards

A subscriber identity module (SIM) card is a device used to hold key information

needed to conduct communications across telecommunication networks. A SIM

card provides a means of identifying users and other key items of information

when using telecommunication networks. When accessing a telecommunication

network, one has to identify themselves for billing purposes. The SIM card

provides the information needed by the network to attribute the call. Elements

such as provider, serial numbers, and keys are stored on a universal integrated

circuit card that acts as a standard for storing and managing this information on

devices. SIM cards are important because they can contain user data and

authentication information as well as provide identity services. When one moves

a SIM card from one phone to another, the new hardware acts like the old

hardware with respect to connectivity and, to a degree, stored data.

Zigbee

Zigbee is a low-power mesh radio service used to connect sensors and basic

devices.

EXAM TIP Communication needs are common among a lot of devices, but the

methods vary. Understanding the limitations of the different methods and the

security options is important.

Constraints

Specialized and embedded systems have a different set of constraints that they

are designed to operate under. Typical constraints for these devices include

limitations on power, compute capacity, network throughput and bandwidth,

cryptography, and cost. Additional issues in items like authentication and trust

can also be driving factors. As these devices are built for a specific purpose,

these limitations are actually design elements and are part of the ability of the

system to perform its task in the expected environment.

Power

Electronic circuits take power to operate, and it comes from one of several

sources: a power supply connected to the grid, a battery, solar, or another type of

device. Power is a key driver in many embedded and specialized systems

because it is a true limiter. When the power supply is interrupted and no backup

power supply exists, the device stops functioning. Rechargeable lithium-ion

batteries have come a long way in the past few years, and for mobile devices

they form the primary supply. Power drives many design elements because extra

functionality that is not needed, including speed, only uses power and does not

add to the functionality of the unit.

Compute

The compute capability of embedded and specialized systems is another key

component that is matched to the task the device is designed to accomplish.

Compute performance is one of the major elements in the power equation, and

excess compute capacity results in more power drain and less useful life on a

battery charge. Microcontrollers, field programmable gate arrays (FPGAs), and

application-specific integrated circuits (ASICs), all discussed earlier in the

chapter, are valid options for the compute segment of a design, and each of these

comes with a wide range of capabilities. From tiny microcontrollers the size of a

grain of rice with very limited capabilities, to the ASICs designed for visual/lidar

processing in modern self-driving cars, the range of capabilities is wide. The key

point to remember is that compute power, power capacity, and useful lifetime

without external power are all locked in a battle, each taking from the other two.

Network

Network limitations are due to constraints from power and connectivity. Without

direct connectivity, networking requires a radio transceiver, and this increases

power demands. Therefore, where networking is needed, it comes at a cost.

There are a variety of methods used for networking, and the one chosen will be

the cheapest and best solution for the networking needs available at the time of

system design.

Leaving single-unit considerations aside, networking is the key value

component behind the Internet of Things revolution. The utility of networking

power is related to an exponential function associated with the number of nodes.

Hence, the greater the number of nodes, the larger the utility, and this growth is

exponential in nature. Larger deployments (think smart meters in a major

metropolitan area) deliver tremendous quantities of data, via a network, to a data

center on a regular and timely basis. Managing large data flows places a burden

on the central site, which if not properly planned for and executed becomes a

constraint on the overall system operation.

Cryptographic Functions

Cryptographic functions can be essential to secure data during transmission, but

this functionality comes at a price. The level of computational resources for

crypto functions can be substantial, thus becoming a constraint on the overall

system. Lightweight cryptographic algorithms are being developed to

specifically address these challenges, and these are covered in Chapter 16,

“Cryptographic Concepts.”

Inability to Patch

The inability to patch an item represents a security risk and a constraint. This is

typically caused by a series of design decisions predicated on producing items

that are not computers but rather single-purpose devices. While Raspberry Pi’s

and Arduinos may get patches from their developers, the embedded controller in

a surveillance camera is another story altogether. Simply put, the ecosystem for

most embedded devices is missing the means, the culture, and in many cases the

sheer ability to manage the patch process.

Authentication

Authentication is an important predicate to security functionality. The definitions

of confidentiality, integrity, and many other security attributes have the term

authenticated user in them. This make authentication an important property, but

given the non-computer ecosystem in which most specialized and embedded

devices function, there is a problem with directly adopting the concept of

authentication. This is not as significant a limitation as it may seem, however,

because unlike computers, which perform a multitude of different functions for

different users, specialized and embedded systems tend to perform a singular

function with an undefined user by design. There may be a need for an

administrative interface for some functions, but enabling this with a simple PIN

is not problematic, especially if defaults were taken into account during the

design and deployment. For more information on this topic, review “Weak

Defaults,” earlier in the chapter.

Range

In most cases, range is a function of power—one of the true limitations of many

specialized and embedded systems. One of the challenges of IoT deployments is

getting them to the Internet, because there range is unlimited. However, this

comes at the cost of security/risk.

Cost

The whole purpose behind developing specialized/embedded systems is that the

value is there. The functionality return for the cost of the unit justifies the design

and deployment, so cost is to a degree baked in. So, rather than viewing cost as a

constraint, it is the factor that drives the creation of these solutions. However,

cost is also an economic issue because extra functionality leads to extra cost, and

if this functionality isn’t needed in the final solution, money is wasted.

Implied Trust

Implied trust, by definition, is trust that has not been specifically set up but yet

exists. This is almost a given in many specialized systems because they are not

intended or designed to be general-purpose compute devices; therefore, the

thought processes associated with regular trust vis-à-vis computers and the

Internet do not exist. This makes for easier connectivity, but also opens doors for

an attacker.

EXAM TIP When questioned about constraints and specialized systems (not

general-purpose computers), remember the ecosystem that the device is intended

to run in and consider that when formulating the answer. Many times it is

different for specialized/embedded systems than for a general-purpose computer.

Chapter Review

In this chapter, you became acquainted with the security implications of

embedded systems, which have become ubiquitous in our everyday lives. The

chapter opened with a discussion of embedded systems in the form of the

Raspberry Pi, field programmable gate array (FPGA), and the Arduino

microcontroller platform. The chapter then presented the SCADA/ICS space and

how operational technology is its own world and one of significant size.

Examinations of these systems in facilities as well as industrial, manufacturing,

energy, and logistics settings were covered. The chapter then moved to the world

of smart devices and the Internet of Things, including sensors, smart devices,

wearable technology, facility automation, and weak defaults.

The chapter then covered specialized systems, including medical systems,

vehicles, aircraft, and smart meters. A discussion of Voice over IP, HVAC

systems, drones, and multifunction printers followed. Next, we looked at realtime operating systems, surveillance systems, and system on a chip, followed by

communication considerations for embedded and specialty systems such as 5G,

narrow band, baseband, SIM cards, and Zigbee. The chapter closed with an

examination of system constraints in these systems, including power, compute

performance, network functions, cryptographic functions, the inability to patch,

authentication, range, cost, and implied trust.

Questions

To help you prepare further for the exam, and to test your level of preparedness,

answer the following questions and then check your answers against the correct

answers at the end of the chapter.

1. Which of the following statements is not true?

A. Embedded systems are designed with a single control purpose in mind

and typically have no additional functionality.

B. Embedded systems are free of risk and security concerns.

C. Embedded is the name given to a computer that is included as an

integral part of a larger system.

D. Embedded systems can be as complex as the dozens of interconnected

embedded systems in a modern automobile.

2. Which of the following statements is true regarding the risk of nextgeneration vehicles?

A. There are minimal risks when next-generation automobiles share

information.

B. Passing traffic and other information between vehicles does not

increase security risks.

C. The sharing of navigation and other inputs between vehicles presents a

potential security issue.

D. Time-to-market and cost minimization have minimal impact on

potential risks being exploited.

3. Which of the following properly defines supervisory control and data

acquisition (SCADA)?

A. A scaled-down version of Linux designed for use in an embedded

system

B. The standard used for communicating between intelligent car systems

C. The risk created by connecting control systems in buildings

D. A system designed to control automated systems in cyber-physical

environments

4. Which of the following statements is true about smart devices and the

Internet of Things (IoT)?

A. The use of a Linux-type kernel as the core engine makes programming

more complex.

B. Mass production introduces significant security risks.

C. The scaling of the software development over large numbers of units

makes costs scalable, and functionality is paramount.

D. Security or anything that might impact new expanded functionality is

considered early and gets the focus and resources necessary.

5. Which of the following statements is true about HVAC and building

automation systems?

A. They have not been exploited to any significant degree yet.

B. Interconnecting these systems and using Internet-based central control

mechanisms increases the risk profile from outside attacks.

C. Having a “smart building” that reduces the use of building resources in

accordance with the number and distribution of people inside has not

increased efficiency or reduced costs.

D. The rise of hyper-connectivity has introduced no additional security

concerns.

6. Which of the following statements is not true about system on a chip?

A. It provides the full functionality of a computing platform on a single

chip.

B. It typically has low power consumption and efficient design.

C. Programming of SoC systems can occur at several different levels, and

thus potential risks are easily mitigated.

D. Because SoC represents computing platforms with billions of devices

worldwide, it has become a significant force in the marketplace.

7. What distinguishes real-time operating systems (RTOSs) from generalpurpose operating systems?

A. Unlike RTOSs, most general-purpose operating systems handle

interrupts within defined time constraints.

B. Unlike general-purpose OSs, most RTOSs are capable of multitasking

by design.

C. Unlike RTOSs, most general-purpose operating systems are

multitasking by design.

D. Unlike general-purpose OSs, RTOSs are designed to handle multiple

threads.

8. Which of the following statements is true about printers and multifunction

devices?

A. They rely on the computer to manage the printing and scanning

processes.

B. Because of their long history and widespread use, security is designed

into these products.

C. These devices communicate in a bidirectional fashion, accepting print

jobs and sending back job status, printer status, and so forth.

D. So far, they have not been shown to be hackable or capable of passing

malware to the computer.

9. Which of the following is a very important aspect to always remember

when dealing with security of medical devices?

A. They are still relatively new in their usage.

B. They can directly affect human life.

C. Security is not related to safety.

D. They are almost exclusively stand-alone devices, without Internet

connectivity.

10. Which of the following poses a significant potential risk of unmanned aerial

vehicles?

A. They have sophisticated autopilot functions.

B. They have cameras, sensors, and payloads.

C. Some models have a low price.

D. Because they are pilotless, their remote-control systems may be

networked and therefore vulnerable to potential risks.

Answers

1. B. Embedded systems are not free of risk or security concerns, as hackers

have demonstrated.

2. C. The sharing of navigation and other inputs presents a potential security

issue for next-generation vehicles. False information, when shared, can

cause problems.

3. D. SCADA is a system designed to control automated systems in cyberphysical environments.

4. C. The scaling of the software development over large numbers of units

makes costs scalable, and functionality is paramount in smart devices and

IoT.

5. B. Interconnecting HVAC and building automation systems and using

Internet-based central control mechanisms to manage them increases the

risk profile from outside attacks.

6. C. Programming of SoC systems can occur at several different levels, and

thus potential risks are difficult to mitigate.

7. C. One thing that distinguishes real-time operating systems (RTOSs) from

general-purpose operating systems is that most general-purpose operating

systems are designed for multitasking.

8. C. Printers and multifunction devices communicate in a bidirectional

fashion, accepting print jobs and sending back job status, printer status, and

so forth.

9. B. A very important aspect to always remember when dealing with security

of medical devices is that they can directly affect human life.

10. D. A significant potential risk of unmanned aerial vehicles is that, because

they are pilotless, their remote-control systems may be networked and

therefore vulnerable to potential risks.

CHAPTER 15

Physical Security Controls

In this chapter, you will

• Explore the importance of physical security controls

• Learn about important environment controls

Physical security is an important topic for businesses dealing with the security of

networks and information systems. Businesses are responsible for managing

their risk exposure, which requires securing a combination of assets: employees,

product inventory, trade secrets, and strategy information. These and other

important assets affect the profitability of a company and its future survival.

Companies therefore perform many activities to attempt to provide physical

security—locking doors, installing alarm systems, using safes, posting security

guards, setting access controls, and more.

Environmental controls play an important role in the protection of the systems

used to process information. Most companies today have invested a large

amount of time, money, and effort in both network security and information

systems security. In this chapter, you will learn about how the strategies for

securing the network and for securing information systems are linked, and you’ll

learn several methods by which companies can minimize their exposure to

physical security events that can diminish their network security.

Certification Objective This chapter covers CompTIA Security+ exam

objective 2.7: Explain the importance of physical security controls.

Bollards/Barricades

The primary defense against a majority of physical attacks is the barricades

between the assets and a potential attacker—walls, fences, gates, and doors.

Barricades provide the foundation upon which all other security initiatives are

based, but the security must be designed carefully, as an attacker has to find only

a single gap to gain access. Barricades can also be used to control vehicular

access to and near a building or structure. The simple post-type barricade that

prevents a vehicle from passing but allows people to walk past is called a

bollard.

EXAM TIP Bollards are sturdy posts often made of concrete or galvanized or

stainless steel. They are used to protect entry ways and prevent unauthorized

entry or vehicle ramming attacks.

Walls may have been one of the first inventions of humans. Once we learned

to use natural obstacles such as mountains to separate us from our enemies, we

next learned to build our own mountain for the same purpose. Hadrian’s Wall in

England, the Great Wall of China, and the Berlin Wall are all famous examples

of such basic physical defenses. The walls of any building serve the same

purpose, but on a smaller scale: they provide barriers to physical access to

company assets. In the case of information assets, as a general rule, the most

valuable assets are contained on company servers. To protect the physical

servers, you must look in all directions. Doors and windows should be

safeguarded, and a minimum number of each should be used in a server room

when they are all that separate the servers from the personnel allowed to access

them. It is very important that any transparent windows or doors do not allow

shoulder surfing from outside the server room. It is good to see people in the

room, just not what they type on their screens. Less obvious entry points should

also be considered: Is a drop ceiling used in the server room? Do the interior

walls extend to the actual roof, raised floors, or crawlspaces? Access to the

server room should be limited to the people who need access, not to all

employees of the organization. If you are going to use a wall to protect an asset,

make sure no obvious holes appear in that wall.

NOTE Windows or no windows? Windows provide visibility, allowing people

to observe activities in the server room. This can provide security if those doing

the observing have authority to see the activity in the server room. If those

outside do not have this authority, then windows should be avoided.

Another method of preventing surreptitious access is through the use of

windows. Many high-security areas have a significant number of windows so

that people’s activities within the area can’t be hidden. A closed server room

with no windows makes for a quiet place for someone to achieve physical access

to a device without worry of being seen. Windows remove this privacy element

that many criminals depend on to achieve their entry and illicit activities.

EXAM TIP All entry points to server rooms and wiring closets should be

closely controlled, and, if possible, access should be logged through an access

control system.

Access Control Vestibules

The implementation of a access control vestibule, also called a mantrap, is one

way to combat tailgating. An access control vestibule is composed of two

closely spaced doors that require the user to card through one and then the other

sequentially. Mantraps make it nearly impossible to trail through a doorway

undetected—if an intruder happens to catch the first door before it closes, he will

be trapped in by the second door, as the second door remains locked until the

first one closes and locks.

EXAM TIP An access control vestibule door arrangement can prevent

unauthorized people from following authorized users through an accesscontrolled door, which is also known as tailgating.

Badges

As organizations grow in size, it is not possible for everyone to know everyone

else by sight. Hence, some form of physical identification is needed to recognize

employees. A badge with a picture on it can enable others to quickly determine if

you are an employee or not. Visitors are given their own badge that identifies

them as a visitor. Radio-frequency identification (RFID) uses electromagnetic

fields to automatically identify and record information. RFID tags are widely

used in identification badges, replacing earlier magnetic stripe cards and making

them useable with just a swipe near a reader.

Alarms

Alarms serve to alert operators to abnormal conditions. Physical security can

involve numerous sensors, intrusion alarms, motion detectors, switches that alert

to doors being opened, video and audio surveillance, and more. Each of these

systems can gather useful information, but it is only truly useful if it is acted

upon. When one of these systems has information that can be of use to

operational personnel, an alarm is the easiest method of alerting personnel to the

condition. Alarms are not simple; if a company has too many alarm conditions,

especially false alarms, then the operators will not react to the conditions as

desired. Tuning alarms so that they provide useful, accurate, and actionable

information is important if you want them to be effective.

EXAM TIP Lighting, signs, fencing, and alarms are all items readily

associated with physical security. The proper answer to an exam question will be

based on the specific details of the question—watch for the clues and pick the

best answer based on the context of the question.

Signage

Signs act as informational devices and can be used in a variety of ways to assist

in physical security. Signage can provide information as to areas that are

restricted, or it can indicate where specific precautions, such as keeping doors

locked, are required. A common use of signs in high-security facilities is to

delineate where visitors are allowed versus secured areas where escorts are

required. Visual security clues can assist in alerting users to the need for specific

security precautions. Visual clues as to the types of protection required can take

the form of different-color name badges that signify the level of access, visual

lanyards that indicate visitors, colored folders, and so forth.

Cameras

Cameras are an important tool for security. The old adage “a picture is worth a

thousand words” consistently rings true, and this is especially true in security.

From recording evidence for later use, like taking pictures of equipment, serial

number panels, and so on, to collecting evidence at crime scenes, cameras

enable the re-creation of scenes at a later date. Cameras have been around for

over 100 years, but with the invention of digital photography, followed by the

addition of cameras to cell phones, today there are literally billions of cameras

worldwide taking tens of billions of photos. One of the interesting uses of this

technology is the ability to quickly share photos with others, allowing someone

to “see” far beyond normal range of eyesight. In 2020, when riots broke out

across the U.S., many protesters documented the police response using cell

phone cameras. While they were using their cameras to document the police,

they also captured images that law enforcement would later use to catch those

responsible for crimes.

Video cameras offer an even greater range of surveillance capability, and

closed-circuit TV cameras are covered in a later section.

Motion Recognition

Motion recognition is an important technology to limit the search time and

recording space associated with video images. Infrared (IR) radiation is not

visible to the human eye, but it can be used just like a light source to detect a

range of things. Motion from living creatures can be seen because of the heat

signatures of their bodies. Infrared detection is a technical means of looking for

things that otherwise may not be noticed. At night, when it is dark, someone can

hide in the shadows, but infrared light can point them out to IR-sensing cameras.

Infrared detectors can sense differences in temperature, which can be from a

person entering a room, even if that person is not visible due to darkness. IR

alarms are used extensively to monitor movement of people in areas where there

should be none.

Object Detection

Modern surveillance video systems come with some impressive software. Even

cameras sold to homeowners can scan video for movement and detect people,

cars, and other designated objects such as packages left on a porch. The use of

video software for object detection does not replace the human eye, but it

significantly enhances a guard’s ability to effectively use large banks of cameras

to cover a facility. The citywide video surveillance system in London was the

primary source of evidence that identified the terrorists who set off a series of

bombs across the city in 2005.

Closed-Circuit Television (CCTV)

Video surveillance is typically done through closed-circuit television (CCTV).

The use of CCTV cameras for surveillance purposes dates back to at least 1961,

when cameras were installed in the London Transport train station. The

development of smaller camera components and lower costs has caused a boon

in the CCTV industry since then.

CCTV cameras are used to monitor a workplace for security purposes. These

systems are commonplace in banks and jewelry stores—places with high-value

merchandise that is attractive to thieves. As the expense of these systems

dropped, they became practical for many more industry segments. Traditional

cameras are analog based and require a video multiplexer to combine all the

signals and make multiple views appear on a monitor. Digital, IP-based cameras

have changed that, as most of them are stand-alone units that are viewable

through a web browser, such as the camera shown in Figure 15-1.

Figure 15-1 IP-based cameras leverage existing IP networks instead of needing

a proprietary CCTV cable.

These IP-based systems add useful functionality, such as the ability to check

on the building from the Internet. This network functionality, however, makes

the cameras subject to normal IP-based network attacks. A DoS attack launched

at the CCTV system just as a break-in is occurring is the last thing that anyone

would want (other than the criminals). For this reason, IP-based CCTV cameras

should be placed on their own separate network that can be accessed only by

security personnel. The same physical separation applies to any IP-based camera

infrastructure. Older time-lapse tape recorders are slowly being replaced with

digital video recorders. While the advance in technology is significant, be

careful if and when these devices become IP-enabled, since they will become a

security issue, just like everything else that touches the network.

If you depend on a CCTV system to protect your organization’s assets,

carefully consider camera placement and the type of cameras used. Different iris

types, focal lengths, and color or infrared capabilities are all options that make

one camera superior to another in a specific location.

Industrial Camouflage

Camouflage is the specific act of rendering an item not readily observable.

Considered by many to be a military thing, camouflage began in nature, where

insects and animals have patterns making them seem to be different than they

really are. This same principle is used all the time to make things hide in plain

sight. Cell phone towers built to look like trees make them less conspicuous—

and generally improve the visual surroundings. In response to physical acts

against electrical substations, many utilities have put walls around the

substations, making the internal equipment no longer visible and less of a target.

NOTE If you want to see some industrial camouflage in action, use Street

View in Google Maps and look at these locations:

• 58 Joralemon Street, New York City, is a ventilation shaft and emergency

access to the New York subway.

• 640 Millwood Road, Toronto, Canada, is an electrical substation—one of

250 in the city.

• 51 W. Ontario Street, Chicago, Illinois, is another substation—this one by

Commonwealth Edison. The doors are fake and don’t open, and the

windows are actually vents.

Personnel

Physical security should be a part of a firm’s overall security program. Physical

security measures are those taken to ensure the separation of items to be

protected from all forms of physical risk. Personnel are an important part of this

equation—from guards to lobby workers who act as gatekeepers for visitors and

packages, people are part of the physical security system.

Guards

Security guards provide an excellent security measure, because guards are a

visible presence with direct responsibility for security. Other employees expect

security guards to behave a certain way with regard to securing the facility.

Guards typically monitor entrances and exits and can maintain access logs of

who has entered and departed the building. In many organizations, everyone who

passes through security as a visitor must sign a log, which can be useful in

tracing who was at what location and why.

Security personnel are helpful in physically securing the machines on which

information assets reside, but to get the most benefit from their presence, they

must be trained to take a holistic approach to security. The value of data typically

can be many times that of the machines on which the data is stored. Security

guards typically are not computer security experts, so they need to be educated

about the value of the data and be trained in network security as well as physical

security involving users. They are the company’s eyes and ears for suspicious

activity, so the network security department needs to train them to notice

suspicious network activity as well. Multiple extensions ringing in sequence

during the night, computers rebooting all at once, and strangers parked in the

parking lot with laptop computers or other mobile computing devices are all

indicators of a network attack that might be missed without proper training.

Robot Sentries

Guard duty is by and large boring work, and although guards aren’t highly paid

over time, having a number of guards can be expensive. Robot technology has

progressed to the point where robots can now perform many simple tasks, and

guard duty can be one of these tasks. Robot sentries can patrol empty buildings

and using sensors can detect the presence of unauthorized people. Robot sentries

can then report the issue to a manned station that can alert the proper authorities

for a response.

Reception

Reception areas are used as buffer zones between different areas of a facility,

segmenting the building into separate regions. Having a visitor check-in desk

allows control over visitors as well as enables functions like logging visitors,

managing deliveries, and providing escorts for visitors. In lower security

environments, this reception area may simply be someone at a desk, with no

physical barrier. In more secure facilities, the receptionist is not only responsible

for keeping logs, issuing access badges, and notifying escorts, but also controls

the door everyone must go through. In very highly controlled environments, the

actual door control is done remotely from the other side of the door so that

people can’t force their way past the receptionist.

Two-Person Integrity/Control

When tasks are critical, or failures could involve significant risk, the

organizational principle of separation of duties applies. This topic is fully

covered in Chapter 33, “Organizational Policies.” When there are physical tasks,

such as opening the door mentioned in the previous section, having two people

required to perform the task provides a means of checks and balances. Twoperson integrity/control is this principle in action: it is when two different people

have to perform respective tasks that are both necessary for the action to occur.

Person 1 can initiate a process, check IDs, enter data in a log, and issue a visitor

badge, while person 2 can control the door access. This way, a failure by either

person does not expose the process.

EXAM TIP Be able to explain important physical security controls, including

guards, robot sentries, reception, and two-person integrity/control.

Locks

Locks are a common security measure that are used with near ubiquity.

Everyone is familiar with using a lock to secure something. Many different lock

types are used in and around the computer security arena. There are types for

laptops and other mobile devices, for desktops, and even servers. Just as locks

can keep your car or bike from being stolen, they can secure computers as well.

Laptops are popular targets for thieves and should be locked inside a desk when

not in use, or secured with special computer lockdown cables. Laptop thefts

from cars can occur in seconds, and thieves have been caught taking laptops

from security screening areas at airports while the owners are distracted with the

screening process. If an organization uses desktop towers, it should use computer

desks that provide a space in which to lock the computer. In some cases,

valuable media is stored in a safe designed for that purpose. All of these

measures can improve the physical security of the computers themselves, but

most of them can be defeated by attackers if users are not knowledgeable about

the security program and do not follow it.

Biometrics

Biometrics is the measurement of biological attributes or processes with the goal

of identification of a party possessing those measurements. The most wellknown biometric factor is the fingerprint. Fingerprint readers have been

available for several years in laptops and other mobile devices, as shown in

Figure 15-2, and as stand-alone USB devices.

Figure 15-2 Newer laptop computers often include a fingerprint reader.

Other biometric measurements that can be used for physical security purposes

include the retina or iris of the eye, the geometry of the hand, and the geometry

of the face. When any of these are used for authentication, there is a two-part

process: enrollment and then authentication. During enrollment, a computer

takes the image of the biological factor and reduces it to a numeric value. When

the user attempts to authenticate, his or her feature is scanned by the reader, and

the computer compares the numeric value being read to the one stored in the

database. If they match, access is allowed. Since these physical factors are

unique, theoretically only the actual authorized person would be allowed access.

Biometrics are frequently used in physical security and are becoming nearly

ubiquitous for controlling access to mobile devices, such as phones and tablets.

For many physical security situations, the true question for access is, are you the

correct person who should have access? Using biometrics to confirm the identity

of the person being presented for access as the same person who went through

the identification phase at enrollment is a good way to answer this question. You

can’t loan your fingerprints, iris, or retina for a scan, or your hand for its

geometry. Biometrics bind the identification token to the person.

Biometrics are not foolproof. Some biometric measures can be duplicated to

fool a sensor, and in many cases, the actual biometric is converted to a number

that can also be intercepted and used in a software attack. Safeguards exist for

most biometric-bypass mechanisms, making them a usable security technology.

Electronic

Electronic locks are devices that impede a specific function unless a code is

entered. This code is compared to a stored secret, and if the correct code is

entered, the lock engages the mechanical stop and allows the mechanism to

open. Electronic locks have an advantage in that they are not as susceptible to

mechanical manipulation and bypass, yet they are still susceptible, in many cases

via the mechanism that updates the secret “combination.”

Physical

Physical locks have been used for hundreds of years; their design has not

changed much: a metal “token” is used to align pins in a mechanical device.

Physical locks have survived for years because they are low cost. Because all

mechanical devices have tolerances, it is possible to sneak through these

tolerances by “picking” the lock. Most locks can be easily picked with simple

tools, some of which are shown in Figure 15-3.

Figure 15-3 Lock-picking tools

Humans are always trying to build a better mousetrap, and that applies to

locks as well. High-security locks, such as the one shown in Figure 15-4, have

been designed to defeat attacks; these locks are more sophisticated than a

standard home deadbolt system. Typically found in commercial applications that

require high security, these locks are made to resist picking and drilling, as well

as other common attacks such as simply pounding the lock through the door.

Another common feature of high-security locks is key control, which refers to

the restrictions placed on making a copy of the key. For most residential locks, a

trip to the hardware store will allow you to make a copy of the key. Key control

locks use patented keyways that can only be copied by a locksmith, who will

keep records on authorized users of a particular key.

Figure 15-4 A high-security lock and its key

High-end lock security is more important now that attacks such as “bump

keys” are well known and widely available. A bump key is a key cut with all

notches to the maximum depth, also known as “all nines.” This key uses a

technique that has be around a long time but has recently gained a lot of

popularity. The key is inserted into the lock and then sharply struck, bouncing

the lock pins up above the shear line and allowing the lock to open. Highsecurity locks attempt to prevent this type of attack through various mechanical

means such as nontraditional pin layout, sidebars, and even magnetic keys.

Combination locks, which work via a rotating dial, are common on high-end

safes and can raise the security level substantially. In many cases, the only way

to bypass one of these locks is to physically bypass the lock itself through

drilling or other methods. Additional levels of safeguard exist, such as shatter

plates, which when broken engage pins that prevent the door from opening.

Cable Locks

Portable equipment has a principal feature of being portable. This can also be a

problem, as portable equipment—laptops, projectors, and the like—can be easily

removed or stolen. Cable locks provide a simple means of securing portable

equipment to furniture or another fixture in the room where the equipment

resides. Cable locks can be used by road warriors to secure laptops from casual

theft. They also can be used in open areas such as conference centers or rooms

where portable equipment is exposed to a wide range of visitors.

USB Data Blocker

USB connectors on computers offer a pathway for data to enter into the system.

Anyone who has physical access to a machine can plug in a USB device and

execute code from the device. There are a variety of ways to block USB ports or

render them inoperable, but in some cases, the USB port serves a secondary

function as a power source for external devices. The USB connection has four

conductors: two for power and two for data. If you block the data conductors,

you can still charge your device from a USB source without giving that device

any access to the data. When charging your phone in locations such as airports,

or other unknown power sources, the use of a USB data blocker protects the

phone but allows it to charge.

EXAM TIP A USB data blocker prevents attackers from infecting a device

with malware or stealing data. Also remember that turning off the AutoPlay

setting in the operating system will prevent malicious code from automatically

running when you plug in a USB or other external media device.

Lighting

Proper lighting is essential for physical security. Unlit or dimly lit areas allow

intruders to lurk and conduct unauthorized activities without a significant risk of

observation by guards or other personnel. External building lighting is important

to ensure that unauthorized activities cannot occur without being observed and

responded to. Internal lighting is equally important because it enables more

people to observe activities and see conditions that are not correct. As described

earlier in the “Bollards/Barricades” section, windows can play an important role

in assisting the observation of the premises. Having sensitive areas well lit and

open to observation through windows prevents activities that would otherwise

take place in secret. Unauthorized parties in server rooms are more likely to be

detected if the servers are centrally located, surrounded in windows, and well lit.

Fencing

Fencing serves as a physical barrier around property. It can serve to keep people

out or in, preventing the free movement across unauthorized areas. Fencing can

be an important part of a physical security plan. Properly employed, it can help

secure areas from unauthorized visitors. Outside of the building’s walls, many

organizations prefer to have a perimeter fence as a physical first layer of defense.

Chain-link-type fencing is most commonly used, and it can be enhanced with

barbed wire along the top. Anti-scale fencing, which looks like very tall vertical

poles placed close together to form a fence, is used for high-security

implementations that require additional scale and tamper resistance.

Inside a building, fencing can be used to provide a means of restricting entry

into areas where separate physical security policies apply. Material storage,

servers, networking gear, and other sensitive items can be separated from

unauthorized access with simple chain link fences. These areas are typically

called a cage, and entry/exit to the caged areas is via a gate. The gate allows

controlled access and makes it easier to monitor who and what enters and leaves

the controlled area. Gates are used for external fencing as well. Gates offer a

monitoring point for ingress and egress from a controlled area.

Fire Suppression

According to the Fire Suppression Systems Association (www.fssa.net), 43

percent of businesses that close as a result of a significant fire never reopen. An

additional 29 percent fail within three years of the event. The ability to respond

to a fire quickly and effectively is thus critical to the long-term success of any

organization. Addressing potential fire hazards and vulnerabilities has long been

a concern of organizations in their risk analysis process. The goal obviously

should be never to have a fire, but if one does occur, it is important to have

mechanisms in place to limit the damage the fire can cause. Fire suppression

systems are designed to provide protection against the damage from a fire that

spreads in a facility. Because they are suppression systems, they don’t prevent

the fire from occurring per se, but they do stop it once it begins.

Water-Based Fire Suppression Systems

Water-based fire suppression systems have long been, and still are today, the

primary tool to address and control structural fires. Considering the amount of

electrical equipment found in today’s office environment and the fact that, for

obvious reasons, this equipment does not react well to large applications of

water, it is important to know what to do with equipment if it does become

subjected to a water-based sprinkler system. The 2017 NFPA 75: Standard for

the Protection of Information Technology Equipment outlines measures that can

be taken to minimize the damage to electronic equipment exposed to water.

Clean-Agent Fire Suppression Systems

Carbon dioxide (CO2

) has been used as a fire suppression agent for a long time.

The Bell Telephone Company used portable CO2 extinguishers in the early part

of the 20th century. Carbon dioxide extinguishers attack all three necessary

elements for a fire to occur. CO2 displaces oxygen so that the amount of oxygen

remaining is insufficient to sustain the fire. It also provides some cooling in the

fire zone and reduces the concentration of “gasified” fuel.

Argon extinguishes fire by lowering the oxygen concentration below the 15

percent level required for combustible items to burn. Argon systems are

designed to reduce the oxygen content to about 12.5 percent, which is below the

15 percent needed for the fire but is still above the 10 percent required by the

EPA for human safety.

Inergen, a product of Ansul Corporation, is composed of three gases: 52

percent nitrogen, 40 percent argon, and 8 percent carbon dioxide. In a manner

similar to pure argon systems, Inergen systems reduce the level of oxygen to

about 12.5 percent, which is sufficient for human safety but not sufficient to

sustain a fire.

Handheld Fire Extinguishers

Although computer security professionals typically do not have much influence

over the type of fire suppression system that their office includes, they do need

to be aware of what type has been installed, what they should do in case of an

emergency, and what they need to do to recover after the release of the system.

One area that they can influence, however, is the type of handheld fire

extinguisher that is located in their area (see Table 15-1).

Table 15-1 Classes of Fires and Types of Suppression Methods

Automatic fire suppression systems designed to discharge when a fire is

detected are not the only systems you should be aware of. If a fire can be caught

and contained before the automatic systems discharge, it can mean significant

savings to the organization in terms of both time and equipment costs (including

the recharging of the automatic system). Handheld extinguishers are common in

offices, but the correct use of them must be understood; otherwise, disaster can

occur.

Fire Detection Devices

An essential complement to fire suppression systems and devices are fire

detection devices (fire detectors). Detectors may be able to detect a fire in its

very early stages, before a fire suppression system is activated, and sound a

warning that potentially enables employees to address the fire before it becomes

serious enough for the fire suppression equipment to kick in.

There are several different types of fire detectors. One type, of which there

are two varieties, is activated by smoke. The two varieties of smoke detector are

ionization and photoelectric. A photoelectric detector is good for potentially

providing advance warning of a smoldering fire. This type of device monitors an

internal beam of light. If something degrades the light (for example, by

obstructing it), the detector assumes it is something like smoke and the alarm

sounds. An ionization style of detector uses an ionization chamber and a small

radioactive source to detect fast-burning fires. Shown in Figure 15-5, the

chamber consists of two plates: one with a positive charge and one with a

negative charge. Oxygen and nitrogen particles in the air become “ionized” (an

ion is freed from the molecule). The freed ion, which has a negative charge, is

attracted to the positive plate, and the remaining part of the molecule, now with a

positive charge, is attracted to the negative plate. This movement of particles

creates a very small electric current that the device measures. Smoke inhibits this

process, and the detector will detect the resulting drop in current and sound an

alarm.

Figure 15-5 An ionization chamber for an ionization type of smoke detector

Both of these devices are often referred to generically as smoke detectors, and

combinations of both varieties are possible. For more information on smoke

detectors, see http://home.howstuffworks.com/home-improvement/householdsafety/fire/smoke2.htm. As both of these devices are triggered by the

interruption of a signal, without regard to why, they can give false alarms. They

are unable to distinguish the difference between the smoke from a kitchen fire

and burned toast.

Another type of fire detector is activated by heat. These devices also come in

two varieties. Fixed-temperature or fixed-point devices activate if the

temperature in the area ever exceeds some predefined level. Rate-of-rise or rateof-increase temperature devices activate when there is a sudden increase in local

temperature that may indicate the beginning stages of a fire. Rate-of-rise sensors

can provide an earlier warning but are also responsible for more false warnings.

A third type of detector is flame activated. This type of device relies on the

flames from the fire to provide a change in the infrared energy that can be

detected. Flame-activated devices are generally more expensive than the other

two types but can frequently detect a fire sooner.

Sensors

One of the first items in the security equation is detection. Detection of a specific

signal can then be compared to a reference as to if it is allowed or not. The

sensor element provides the detection aspect to the security system, enabling

decisions and resultant processes. For instance, a motion detector that is trained

to detect oncoming traffic can sense someone going the wrong way in a tunnel or

controlled exit space.

Motion Detection

When monitoring an area for unauthorized activity, one potentially useful tool is

a motion detector. In areas where there is little or no expected traffic, a motion

detector can alert an operator to activity in an area. Motion detectors come in a

variety of types, but most are based on infrared (heat) radiation and can detect

the changes of a warm body moving. They can be tuned for size, ignoring

smaller movement such as small animals in outdoor settings. Although not

useful in busy office buildings during normal daily use, motion detectors can be

useful during off-hours, when traffic is minimal. Motion detectors can be used to

trigger video systems, so they do not record large amounts of “empty” activity.

Video monitoring of the loading dock area in the back of the building can be

triggered in this fashion, using the motion detector to turn on cameras whenever

activity is occurring.

Noise Detection

Noise detection is a sensor method that listens for specific sounds. Ordinary

things can produce different sounds, and each of these can have a specific

spectral signature that can be used to hear some items while ignoring others.

Glass breakage has a specific sound, and sensors can be tuned to “hear” glass

breakage and provide an alert when it occurs. The use of sensors that target

events such as this and provide the information to a central alarm panel can

greatly increase the effectiveness of security personnel in monitoring a larger

facility.

Proximity Reader

Proximity readers are sensors that provide a signal at a specified distance. The

most common application of these are card readers connected to doors: you

“swipe” your card by placing it near the reader, and if the codes are correct, you

are granted access. However, these devices have much greater utility. A series of

proximity readers scattered throughout a facility can act as a reporting sensor,

monitoring guards as they traverse their appointed rounds. Guards can check in

to each point by interacting with the proximity reader, typically by swiping a

card near the device, and the device records their presence at that spot at that

time. With near field communication (NFC) and advanced Bluetooth via

smartphones, the uses of proximity readers beyond just paying for things is

growing exponentially. For example, proximity devices in bus stops can allow

your smartphone to get an updated bus schedule. The ability to sense and

communicate over short distances has almost endless possibilities.

Moisture Detection

Moisture, or water, can have significant detrimental effects on certain items.

Moisture detection sensors provide a remote means of monitoring everything

from water leaks to humidity problems. Water can cause damage to electronics,

artwork, and many other items. Being able to monitor moisture levels provides

the security team a means of detecting the potential for damage from items such

as leaking sprinklers or water leaks. As in all sensors, the objective is to provide

better “eyes and ears” for the security personnel, allowing 24/7 coverage of

issues, many times in remote areas, for conditions that may require attention.

Cards

Controlling physical access to a small facility can be achieved through door

locks and physical keys, but that solution is unwieldy for larger facilities with

numerous people coming and going. Many organizations rely on a badging

system using either tokens or cards that can be tied to automated ID checks and

logging of entry/exit. This can provide much greater detail in tracking who is in

a facility and when they have come and gone. Tokens and cards can provide a

serialized ID for each user, enabling user-specific logging. Originally designed

to augment payroll timecards, these electronic IDs have improved security

through the logging of employees’ in and out times. Tokens and cards offer the

same function as keys, but the system can be remotely updated to manage access

in real time, and users can have their privileges revoked without a company or

admin having to recover the token or card.

Temperature

Temperature sensors do exactly what you’d think: they sense temperatures. Part

of the physical security equation is preventing damage to the infrastructure in an

organization, and servers can be an important part of that infrastructure. Server

rooms are highly temperature-controlled areas, with hot and cold sides, as

servers tend to generate heat, and that heat needs to be removed. Hot and cold

aisles are covered in more detail in a later section in this chapter. Monitoring the

current temperature in server rooms requires temperature sensors, properly

placed to measure the actual temperature experienced by the servers. An

analytical monitoring solution can then alert the appropriate personnel when

certain temperature ranges are exceeded. In small facilities, one sensor for the

entire room might be sufficient; in larger server farms, there may be a sensor per

rack. In any case, the idea is the same: measure the temperature and report on

exceptions.

EXAM TIP The use of sensors as part of an overall physical security solution

is important. There are many things that need to be monitored, and using sensors

with automation to assist the security team in seeing out-of-range conditions is

important. The objective is to understand the importance of physical security,

and that includes the specific details measured by sensors.

Drones

The use of drones has risen sharply in the past couple of years. From

home/hobbyist models that can carry a small camera, to larger industrial rigs that

can carry larger cameras for longer periods, these devices have revolutionized

remote viewing of items. Drones are used by railroads to inspect tracks and used

by electric companies to inspect power lines. Their ability to go almost

anywhere and visually inspect things is a great resource. These offer interesting

use cases for both offense and defense in cybersecurity, because they can be used

to surveil physical facilities remotely, providing eyes on demand in a variety of

places you might not want a person to go to and in a timeframe that can’t be met

any other way.

Visitor Logs

Physical security visitor logs provide the same utility as computer logs for a

security investigation. They act as a record of what was observed at specific

points in time. Having roving guards check in at various places across a shift via

a log entry provides a record of the actual surveillance. Logs of visitors arriving

and departing, equipment received and shipped out, and so forth serve as a

record of the physical happenings in a facility.

Remote sensing of badges and equipment utilizing RFID tags can create an

automatic log of equipment movement, including information about when,

where, what, and who. Advanced capabilities such as these make inventory of

movable equipment easier, as its location is tracked and it can be scanned

remotely.

EXAM TIP ameras, IR detection, motion detection, and logs are all methods

associated with detection—and frequently after-the-fact detection at that. These

devices and methods provide valuable attribution fact patterns, even after the

actual event.

Faraday Cages

Electromagnetic interference (EMI) is an electrical disturbance that affects an

electrical circuit. EMI is due to either electromagnetic induction or radiation

emitted from an external source, either of which can induce currents into the

small circuits that make up computer systems and cause logic upsets. EMI can

plague any type of electronics, but the density of circuitry in the typical data

center can make it a haven for EMI. The amount of sensitivity to an EMI field

depends on a number of factors, including the length of the circuit, which can act

like an antenna. EMI is grouped into two general types: narrowband and

broadband. Narrowband is, by its nature, electromagnetic energy with a small

frequency band and, therefore, typically sourced from a device that is

purposefully transmitting in the specified band, such as a phone. Broadband

covers a wider array of frequencies and is typically caused by some type of

general electrical power use such as power lines or electric motors.

An example of shielding that can be employed is a Faraday cage or Faraday

shield, which is an enclosure of conductive material that is grounded. This can

be room-sized or built into a building’s construction; the critical element is that

there is no significant gap in the enclosure material. These measures can help

shield EMI, especially in high- radio-frequency environments. Faraday cages

can be item specific in size, so smaller systems that can encase just a single

smartphone are available.

EXAM TIP When it comes to shielding, understand the difference between a

Faraday cage (as a large open space) and EMI shielding on cables (very specific

shielding) and which is appropriate based on what is being protected from EMI.

Air Gap

Air gap is a term used to describe the physical and logical separation of a

network from all other networks. This separation is designed to prevent

unauthorized data transfers to and from the network. The flaw in this logic is that

users will move data by other means, such as a USB drive, to get their work

done. Frequently called “sneaker net,” this unauthorized bypassing of the air

gap, although ostensibly for the purpose of mission accomplishment, increases

system risk because it also bypasses checks, logging, and other processes

important in development and deployment.

Screened Subnet

The concept of a screened subnet (previously known as a demilitarized zone

[DMZ]) comes from military parlance where it represents an area that is not

“owned” by either side. This concept is used in networking to indicate a zone

where access controls are not as strict as the inside, or as open as the outside; it’s

a place of joint cooperation and controlled risk. This same concept works in

physical structures, where the lobby is like the outside world and anyone can

enter, then there are common hallways where employees mingle, and finally

there are special offices and server rooms where access is tightly controlled. The

common work areas are akin to the DMZ—an area of controlled risk.

Protected Cable Distribution

Cable runs between systems need to be protected from physical damage to the

cables and subsequent communication failures. This is accomplished by

protected distribution/protected cabling during the cable installation. This may

be something as simple as metal tubes or as complex a concrete pipes to run

buried cables. The objective is to prevent any physical damage to the physical

layer portion of the system. Protected distribution/protected cabling provides

physical safeguards to the cabling between systems, from all physical hazards,

including interception and tapping. Shielding cables, such as shielded twisted

pair cables, are designed to prevent electromagnetic interference from affecting

the signals on the wires in the cable. The protection of entire systems is covered

in the earlier section “Faraday Cages.”

Secure Areas

Secure areas are those areas where specific preventative measures are taken to

control access both to and from. Like many other physical security constructs,

there is a wide range of levels for secure areas. From those created by a simple

locked door, to those with special procedures and guards, secure areas can be

tailored to the security needs of an enterprise. The overall idea behind a secure

area is to limit information and people flow in and out of the area, and when it is

permitted it is under the proper level of control. Transport Security

Administration (TSA) creates a secure area when you go to an airport by

allowing only certain materials and people to pass the checkpoint.

Air Gap

As previously mentioned, the term air gap is used to refer to a system where

there is no direct connection to outside systems. An air-gapped network does not

have a connection to outside networks. An air gap for a network extends to all

physical connections, wired and wireless, and exists to protect a computer or

network from outside influences or to keep data from leaving the system.

Seemingly simple in principle, it is much harder in practice. If a system is air

gapped, how does data get in? What do you do with the results? How do you

maintain the system, provide updates, and so on? In practice, when air gaps are

used, they have to be monitored for connections that occur around them,

ensuring the system remains isolated.

EXAM TIP CompTIA lists air gap twice in Objective 2.7, so consider yourself

warned. An air gap is a security measure implemented to ensure that systems

within a secure network are totally isolated (not connected) from an unsecure

network such as the Internet.

Vault

A vault is a secured area that is designed to provide a specific level of security

for what is stored inside. This can be a physical space, with specific safeguards

such as walls that cannot be penetrated and doors that can be secured. A vault is

a larger item than most safes, typically room sized. For example, a bank vault is

used to store large sums of money and other valuables.

Safe

Safes are physical storage devices that are intended to impede unauthorized

access to their protected contents. Safes come in a wide variety of shapes, sizes,

and costs. The higher the level of protection from the physical environment, the

better the level of protection against unauthorized access. Safes are not perfect;

in fact, they are rated in terms of how long they can be expected to protect the

contents from theft or fire. The better the rating, the more expensive the safe.

There are times when a safe is overkill, providing a higher level of security

than is really needed. A simpler solution is secure cabinets and enclosures.

Secure cabinets and enclosures provide system owners a place to park an asset

until its use. Most secure cabinets/enclosures do not offer all of the levels of

protection that one gets with a safe, but they can be useful, especially when the

volume of secure storage is large.

Secure enclosures can provide security against some forms of physical access,

as in users, yet still provide the proper environmental controls and setting

necessary for operation. Safes cannot typically provide these levels of controls.

Hot and Cold Aisles

The trend toward smaller, denser servers means more servers and devices per

rack, putting a greater load on the cooling systems. This encourages the use of a

hot aisle/cold aisle layout. A data center that is arranged into hot and cold aisles

dictates that all the intake fans on all equipment face the cold aisle and that the

exhaust fans all face the opposite aisle. The HVAC system is then designed to

push cool air underneath the raised floor and up through perforated tiles on the

cold aisle. Hot air from the hot aisle is captured by return air ducts for the HVAC

system. The use of this layout is designed to control airflow, with the purpose

being never to mix the hot and cold air. This requires the use of blocking plates

and side plates to close open rack slots. The benefits of this arrangement are that

cooling is more efficient and can handle higher density.

NOTE Understanding airflow allows you to understand hot and cold aisles.

Cold air is produced by the HVAC equipment, and this cold air is sent to servers.

The servers shed their heat, making air hot, which is removed. The aisles keep

the hot air from mixing with the cold air, making the cooling efficient. You

wouldn’t leave a door open in summer with the air conditioning on, would you?

EXAM TIP Understand and be able to explain the importance of secure areas

such as air gap, vault, safe, hot aisle, and cold aisle.

Secure Data Destruction

When data is no longer being used, whether it be on old printouts, old systems

being discarded, or broken equipment, it is important to destroy the data before

losing physical control over the media it is on. Many criminals have learned the

value of dumpster diving to discover information that can be used in identity

theft, social engineering, and other malicious activities. An organization must

concern itself not only with paper trash, but also the information stored on

discarded objects such as computers. Several government organizations have

been embarrassed when old computers sold to salvagers proved to contain

sensitive documents on their hard drives. It is critical for every organization to

have a strong disposal and destruction policy and related procedures. This

section covers data destruction and media sanitization methods.

Burning

Burning is considered one of the gold-standard methods of data destruction.

Once the storage media is rendered into a form that can be destroyed by fire, the

chemical processes of fire are irreversible and render the data lost forever. The

typical method is to shred the material, even plastic discs and hard drives

(including SSDs), and then put the shred in an incinerator and oxidize the

material back to base chemical forms. When the material is completely

combusted, the information that was on it is gone.

Shredding

Shredding is the physical destruction by tearing an item into many small pieces,

which can then be mixed, making reassembly difficult if not impossible.

Important papers should be shredded, and important in this case means anything

that might be useful to a potential intruder or dumpster diver. It is amazing what

intruders can do with what appears to be innocent pieces of information.

Shredders come in all sizes, from little desktop models that can handle a few

pages at a time, or a single CD/DVD, to industrial versions that can handle even

phone books and multiple discs at the same time. The ultimate in industrial

shredders can even shred hard disk drives, metal case and all. Many document

destruction companies have larger shredders on trucks that they bring to their

client’s location and do on-site shredding on a regular schedule.

Pulping

Pulping is a process by which paper fibers are suspended in a liquid and

recombined into new paper. If you have data records on paper, and you shred the

paper, the pulping process removes the ink by bleaching, and recombines all the

shred into new paper, completely destroying the physical layout of the old paper.

Pulverizing

Pulverizing is a physical process of destruction using excessive physical force to

break an item into unusable pieces. Pulverizers are used on items like hard disk

drives, destroying the platters in a manner that they cannot be reconstructed. A

more modern method of pulverizing the data itself is the use of encryption. The

data on the drive is encrypted and the key itself is destroyed. This renders the

data nonrecoverable based on the encryption strength. This method has unique

advantages of scale; a small business can pulverize its own data, whereas it

would either need expensive equipment or a third party to pulverize the few

disks it needs to destroy each year.

Degaussing

A safer method for destroying files on magnetic storage devices (that is,

magnetic tape and hard drives) is to destroy the data magnetically, using a strong

magnetic field to degauss the media. Degaussing realigns the magnetic particles,

removing the organized structure that represented the data. This effectively

destroys all data on the media. Several commercial degaussers are available for

this purpose.

Purging

Data purging is a term that is commonly used to describe methods that

permanently erase and remove data from a storage space. The key phrase is

“remove data,” for unlike deletion, which just destroys the data, purging is

designed to open up the storage space for reuse. A circular buffer is a great

example of an automatic purge mechanism. It stores a given number of data

elements and then the space is reused. If a circular buffer holds 64 MB, for

example, once it is full, it overwrites the oldest material as new material is added

to the buffer.

Third-Party Solutions

Like many other elements of a security program, there are contractors that sell

data destruction as a service. These vendors can take advantage of scale,

increasing the capability while sharing the cost of equipment. However, this also

introduces a new form of data loss, through the use of the third party that has

access to the data before destruction. And, as with all third-party relationships,

what counts is what is in the contract. Therefore, a good security review of the

particulars in the contract is warranted, not just for legal issues but also technical

ones.

EXAM TIP This section covers several methods of data/media destruction, a

couple of which are used together. Learn the details of each method and look for

nonsense answer choices to narrow down the possible correct answers, such as

options that refer to pulping non-paper items or degaussing nonmagnetic media.

Chapter Review

In this chapter, you became acquainted with the principles of physical security

controls, including environmental controls. The chapter began by discussing

bollards/barricades, signage, cameras, CCTV, and industrial camouflage—all

items designed to restrict, guide, or monitor physical movement. From there the

chapter moved into security guards, robot sentries, locks, lighting, sensors,

drones/UAVs, and protected distribution for cables. These elements further

refine restrictions on movement and the ability to access system components.

The chapter then examined Faraday cages, air gap, and DMZ. Secure areas

including vaults, safes, hot aisles, and cold aisles were also covered.

The chapter closed with an examination of secure data destruction methods,

including burning, shredding, degaussing, and third-party solutions.

Questions

To help you prepare further for the CompTIA Security+ exam, and to test your

level of preparedness, answer the following questions and then check your

answers against the correct answers at the end of the chapter.

1. Why is physical security important to protecting data?

A. Physical access to data will negate the security advantages of the

cloud.

B. Information resides on physical assets, linking physical and

information security.

C. Social engineering can negate any information security controls.

D. None of the above.

2. Why is proper interior and exterior lighting important?

A. It can detect people who are where they don’t belong.

B. It shows who is in a restricted space.

C. It allows more people and activities to be observed.

D. It is needed for the use of closed-circuit television cameras.

3. Your organization has experienced multiple incidents of graffiti tagging and

people loitering in the parking lot despite the chain-link fence surrounding

it. What is the best solution to the issue?

A. “No Trespassing” signage

B. More guard stations

C. Additional external lighting

D. Changing the chain-link fencing to anti-scale fencing

4. After a physical security incident, what critical data can security guards

commonly provide?

A. Employee ID information

B. Access logs of who has entered and exited the building

C. Alarm codes

D. Blueprints showing unmonitored areas of the building

5. Alarms are effective only if which of the following is true?

A. They alert on abnormal conditions.

B. Every entrance is monitored with a sensor.

C. They are not tied to the information systems.

D. They are tuned to provide accurate and useful alerts.

6. You are implementing a test lab at your organization for early alpha

software development. To prevent any of the development code from

inadvertently getting put on production computers, what should you

implement?

A. Air gap

B. Strict firewalls

C. Protected distribution

D. Patch management

7. What is the security benefit of a Faraday cage?

A. Prevents attack by EMP

B. Prevents accessing a device using a wireless network or cell connection

C. Works better than anti-scale fencing

D. Prevents stack overflows by EMI

8. What is an example of a human-based screened subnet (DMZ)?

A. A visitor’s lobby that is separated from a company office by a

receptionist

B. Hallways between the company lobby and offices

C. A server room with a locked door

D. The networking cabinets in the facility

9. What is a primary problem with biometrics?

A. Technically, biometrics are difficult to implement.

B. The human body changes over time.

C. Biometrics are easily faked.

D. Biometrics can’t be loaned or delegated.

10. What should you do to protect your IP-based CCTV system from a DDoS

attack?

A. Reconfigure your firewalls.

B. Connect it to an intrusion detection system.

C. Require multifactor authentication to access the CCTV system.

D. Place all CCTV components on a separate network.

Answers

1. B. Information resides on physical assets, linking physical security with the

security of information.

2. C. Proper lighting allows more people and activities to be observed.

3. D. A change from chain-link fencing to anti-scale fencing to prevent

intruders from climbing the fence is the best solution.

4. B. Guards commonly have logs of who has entered and exited a building.

5. D. Alarms are effective only if they are tuned to provide accurate and useful

alerting information.

6. A. A lab environment can be air gapped from the rest of the network to

prevent software from being accidentally copied to production machines.

7. B. A Faraday cage can prevent accessing a device via radio frequency

waves, either from a wireless network or cell radio.

8. B. The lobby is part of the outside environment, so the hallways are the

better choice. Server rooms and networking rooms are the more secured

spaces.

9. B. Some biometric features can change over time, or medical conditions

can make them less reliable, thus forcing a re-identification phase to resync

a user and their biometric.

10. D. The CCTV system should be on a completely separate network, air

gapped if possible, with only security personnel having access