**EXAM TIP** Most implementations of Ethernet over UTP have a maximum cable length of 100 meters, partly to deal with attenuation.

**Crosstalk** *Crosstalk* is a phenomenon that occurs when electrical signals of one wire spill over to the signals of another wire. When electricity flows through a wire, it generates a magnetic field around it. If another wire is close enough, the second wire acts as an antenna that turns this magnetic field into an electric current. When the different electrical signals mix, their integrity degrades and data corruption can occur. UTP mitigates crosstalk by twisting the wires around each other. Because crosstalk is greatest wherever wires are parallel to each other, this twisting makes it harder for this condition to exist. Still, UTP is much more vulnerable to crosstalk than STP or coaxial because it does not have extra layers of shielding to help protect against it.

**Fire Rating of Cables** Just as buildings must meet certain fire codes, so must wiring schemes. A lot of organizations string their network wires in drop ceilings—the space between the ceiling and the next floor—or under raised floors. This hides the cables and prevents people from tripping over them. However, when wires are strung in places like this, they are more likely to catch on fire without anyone knowing about it. Some cables produce hazardous gases when on fire that would spread throughout the building quickly. Network cabling that is placed in these types of areas, called *plenum space*, must meet a specific fire rating to ensure the cable will not produce and release harmful chemicals in case of a fire. A ventilation system's components are usually located in this plenum space, so if toxic chemicals were to get into that area, they could easily spread throughout the building in minutes.

*Nonplenum* cables usually have a polyvinyl chloride (PVC) jacket covering, whereas plenum-rated cables have jacket covers made of fluoropolymers. When setting up a network or extending an existing network, it is important that you know which wire types are required in which situation.

Cables should be installed in unexposed areas so they are not easily tripped over, damaged, or eavesdropped upon. The cables should be strung behind walls and in the protected spaces, such as in dropped ceilings. In environments that require extensive security, wires can be encapsulated within *pressurized conduits* so if someone attempts to access a wire, the pressure of the conduit changes, causing an alarm to sound and a message to be sent to the security staff. A better approach to high-security requirements is probably to use fiber-optic cable, which is much more difficult to covertly tap.

**NOTE** While a lot of the world's infrastructure is wired and thus uses one of these types of cables, remember that a growing percentage of our infrastructure is not wired, but rather uses some form of wireless technology (Bluetooth, Wi-Fi, satellite, etc.), particularly to reach end devices.

## Bandwidth and Throughput

Whatever type of transmission you use over any given cable, there is a limit to how much information you can encode within it. In computer networks, we use two different but related terms to measure this limit. *Bandwidth* is the amount of information that theoretically can be transmitted over a link within a second. In a perfect world, this is the data transfer capability of a connection and is commonly associated with the number of available frequencies and speed of a link. Data *throughput* is the actual amount of data that can be carried over a real link. Throughput is always less than or equal to a link's bandwidth. In fact, it is most often the case that throughput is notably less than bandwidth. Why?

As mentioned, bandwidth is a theoretical limit determined by analyzing a medium (e.g., category 5 UTP cable) and a physical layer protocol (e.g., 100BaseT Ethernet) and then doing the math to calculate the maximum possible amount of data we could push through it. Now, of course, when you put that medium and protocol into a real environment, a multitude of issues come into play and make it hard to achieve that optimal data rate.

The throughput of our networks is affected by many factors. There could be EMI (or line noise) in the medium, as previously discussed. However, in a well-engineered facility and network, this should not be a big problem. Typically, you'll be more concerned about packet delays and losses. *Latency* is the amount of time it takes a packet to get from its source to its destination. This could be measured as either time to first byte (TTFB) or round-trip time (RTT). Latency can be caused by multiple factors, including

- **Transmission medium** Even though electricity and light move at the speed of light, it still takes time to get from one place to another. If your links are very long, or if the cables have too many imperfections, the medium itself will cause latency.
- **Network devices** Routers and firewalls take some time to examine packets, even if they're just deciding which outbound interface to use. If you have too many rules in your routing or security devices, this is invariably going to introduce delays.

To reduce latency, you should keep your physical links as short as possible. You should also look at how many hops your packets must take to get to their destinations. Virtual LANs (VLANs) can help keep devices that communicate frequently "closer" to each other. For international organizations, using a content distribution network (CDN), which we address later in this chapter, keeps most data close to where it is needed. Finally, the use of proxies can reduce latency by bringing frequently requested data closer to your users.

Another issue that negatively impacts your data throughputs (compared to a link's rated bandwidth) is congestion. Since some links in your network are shared, if you have too many packets moving around, it will inevitably bog things down. You may have a 1-GBps (bandwidth) connection to your home, but if every house in your neighborhood has one too and you all share a 1-GBps link from the local switch to the first router, your throughput will be way lower than advertised unless you log on when everyone else is sleeping. The best way to prevent congestion is through careful

design and implementation of your network. Keep your broadcast domains as small as possible, ensure that your shared links are able to support peak traffic rates, and consider prioritizing certain types of traffic so that if your staff decides to livestream news, that doesn't slow down your ability to get real work done.

# Network Devices

Several types of devices are used in LANs, MANs, and WANs to provide intercommunication among computers and networks. We need to have physical devices throughout the network to actually use all the protocols and services we have covered up to this point. The different network devices vary according to their functionality, capabilities, intelligence, and network placement. We will look at the following devices:

- Repeaters
- Bridges
- Switches
- Routers
- Gateways
- Proxy servers
- PBXs
- Network access control devices

The typical network has a bunch of these devices, and their purposes and operation can get confusing really quickly. Therefore, we will also look at network diagram techniques that can help us create different (simpler) views into complex environments. We'll also consider operational issues like power requirements, warranties, and support agreements.

## Repeaters

A *repeater* provides the simplest type of connectivity because it only repeats electrical signals between cable segments, which enables it to extend a network. Repeaters work at the physical layer and are add-on devices for extending a network connection over a greater distance. The device amplifies signals because signals attenuate the farther they have to travel.

Repeaters can also work as line conditioners by actually cleaning up the signals. This works much better when amplifying digital signals than when amplifying analog signals because digital signals are discrete units, which makes extraction of background noise from them much easier for the amplifier. If the device is amplifying analog signals, any accompanying noise often is amplified as well, which may further distort the signal.

A *hub* is a multiport repeater. A hub is often referred to as a *concentrator* because it is the physical communication device that allows several computers and devices to communicate with each other. A hub does not understand or work with IP or MAC addresses. When one system sends a signal to go to another system connected to it, the signal is broadcast to all the ports, and thus to all the systems connected to the concentrator.

**NOTE** Hubs are exceptionally rare nowadays but you may still come across them.

## Bridges

A *bridge* is a LAN device used to connect LAN segments (or VLAN segments) and thus extends the range of a LAN. It works at the data link layer and therefore works with MAC addresses. A repeater does not work with addresses; it just forwards all signals it receives. When a frame arrives at a bridge, the bridge determines whether or not the MAC address is on the local network segment. If it is not, the bridge forwards the frame to the necessary network segment. A bridge amplifies the electrical signal, as does a repeater, but it has more intelligence than a repeater and is used to extend a LAN and enable the administrator to filter frames to control which frames go where.

When using bridges, you have to watch carefully for *broadcast storms*. While bridges break up a collision domain by port (i.e., computers on the same bridge port are in the same collision domain), all ports are on the same broadcast domain. Because bridges can forward all traffic, they forward all broadcast packets as well. This can overwhelm the network and result in a broadcast storm, which degrades the network bandwidth and performance.

The international standard for bridges on Ethernet networks is IEEE 802.1Q. It describes the principal elements of bridge operation as follows:

- Relaying and filtering frames (based on MAC addresses and port numbers)
- Maintenance of the information required to make frame filtering and relaying decisions (i.e., the forwarding tables)
- Management of the elements listed (e.g., aging off forwarding table entries)


**EXAM TIP** Do not confuse routers with bridges. Routers work at the network layer and filter packets based on IP addresses, whereas bridges work at the data link layer and filter frames based on MAC addresses. Routers usually do not pass broadcast information, but bridges do pass broadcast information.

### Forwarding Tables

A bridge must know how to get a frame to its destination—that is, it must know to which port the frame must be sent and where the destination host is located. Years ago, network administrators had to type route paths into bridges so the bridges had static paths indicating where to pass frames that were headed for different destinations. This was a tedious task and prone to errors. Today, most bridges use *transparent bridging*.

In transparent bridging, a bridge starts to learn about the network's environment as soon as it is powered on and continues to learn as the network changes. It does this by examining frames and making entries in its forwarding tables. When a bridge receives a frame from a new source computer, the bridge associates this new source address and the

### Connecting Two LANS: Bridge vs. Router

What is the difference between two LANs connected via a bridge versus two LANs connected via a router? If two LANs are connected with a bridge, the LANs have been extended because they are both in the same broadcast domain. A router separates broadcast domains, so if two LANs are connected with a router, an internetwork results. An *internetwork* is a group of networks connected in a way that enables any node on any network to communicate with any other node. The Internet is an example of an internetwork.

port on which it arrived. It does this for all computers that send frames on the network. Eventually, the bridge knows the address of each computer on the various network segments and to which port each is connected. If the bridge receives a request to send a frame to a destination that is not in its forwarding table, it sends out a query frame on each network segment except for the source segment. The destination host is the only one that replies to this query. The bridge updates its table with this computer address and the port to which it is connected and forwards the frame.

Many bridges use the *Spanning Tree Protocol (STP)*, which adds more intelligence to the bridges. STP ensures that frames do not circle networks forever, provides redundant paths in case a bridge goes down, assigns unique identifiers to each bridge, assigns priority values to these bridges, and calculates path costs. This creates much more efficient frame-forwarding processes by each bridge. STP also enables an administrator to indicate whether he wants traffic to travel certain paths instead of others. Newer bridges implement the Shortest Path Bridging (SPB) protocol, which is defined in IEEE 802.1aq and is more efficient and scalable than STP.



## Switches

*Switches* are, essentially, multiport bridges that typically have additional management features. Because bridges are intended to connect and extend LANs (and not necessarily individual hosts), they tend to have few ports. However, if you take the exact same functionality and add a bunch of ports to it, you could use the ports to connect to each individual host or to other switches. Figure 14-5 illustrates a typical, hierarchical network configuration in which computers are directly connected to access switches within close proximity (100 m or less). Access switches are, in turn, connected to distribution switches, which usually connect different departments or floors in a building. This distribution layer is a great place to implement access control lists (ACLs) and filtering to provide security. Finally, the upper tier of core switches provides a high-speed switching and routing backbone for the organization and is designed to pass network traffic as fast as possible. In this layer, only switches are connected to each other (i.e., there are no computers directly connected to them).

On Ethernet networks, computers have to compete for the same shared network medium. Each computer must listen for activity on the network and transmit its data when it thinks the coast is clear. This contention and the resulting collisions cause

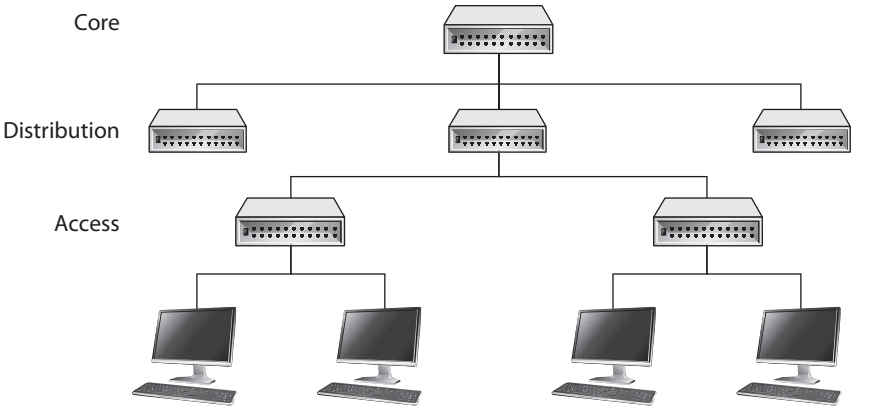



**Figure 14-5** Hierarchical model of a switched network

traffic delays and use up precious bandwidth. When switches are used, contention and collisions are not issues, which results in more efficient use of the network's bandwidth and decreased latency. Switches reduce or remove the sharing of the network medium and the problems that come with it.

Since a switch is a multiport bridging device where each port is connected to exactly one other device, each port provides dedicated bandwidth to the device attached to it. A port is bridged to another port so the two devices have an end-to-end private link. The switch employs full-duplex communication, so one wire pair is used for sending and another pair is used for receiving. This ensures the two connected devices do not compete for the same bandwidth.

Basic switches work at the data link layer and forward traffic based on MAC addresses. However, today's layer 3, layer 4, and other layer switches have more enhanced functionality than layer 2 switches. These higher-level switches offer routing functionality, packet inspection, traffic prioritization, and QoS functionality. These switches are referred to as *multilayered switches* because they combine data link layer, network layer, and other layer functionalities.

Multilayered switches use hardware-based processing power, which enables them to look deeper within the frame, to make more decisions based on the information encapsulated within the frame, and then to provide forwarding and traffic management tasks. Usually this amount of work creates a lot of overhead and traffic delay, but multilayered switches perform these activities within an application-specific integrated circuit (ASIC). This means that most of the functions of the switch are performed at the hardware and chip level rather than at the software level, making it much faster than routers.

**CAUTION** While it is harder for attackers to sniff traffic on switched networks, they should not be considered safe just because switches are involved. Attackers commonly poison cache memory used on switches to divert traffic to their desired location.

## Layer 3 and 4 Switches

Layer 2 switches only have the intelligence to forward a frame based on its MAC address and do not have a higher understanding of the network as a whole. A layer 3 switch has the intelligence of a router. It not only can route packets based on their IP addresses but also can choose routes based on availability and performance. A layer 3 switch is basically a router on steroids, because it moves the route lookup functionality to the more efficient switching hardware level.

The basic distinction between layer 2, 3, and 4 switches is the header information the device looks at to make forwarding or routing decisions (data link, network, or transport OSI layers). But layer 3 and 4 switches can use tags, which are assigned to each destination network or subnet. When a packet reaches the switch, the switch compares the destination address with its tag information base, which is a list of all the subnets and their corresponding tag numbers. The switch appends the tag to the packet and sends it to the next switch. All the switches in between this first switch and the destination host just review this tag information to determine which route it needs to take, instead of analyzing the full header. Once the packet reaches the last switch, this tag is removed and the packet is sent to the destination. This process increases the speed of routing of packets from one location to another.

The use of these types of tags, referred to as *Multiprotocol Label Switching (MPLS)*, not only allows for faster routing but also addresses service requirements for the different packet types. Some time-sensitive traffic (such as video conferencing) requires a certain level of service (QoS) that guarantees a minimum rate of data delivery to meet the requirements of a user or application. When MPLS is used, different priority information is placed into the tags to help ensure that time-sensitive traffic has a higher priority than less sensitive traffic, as shown in Figure 14-6.



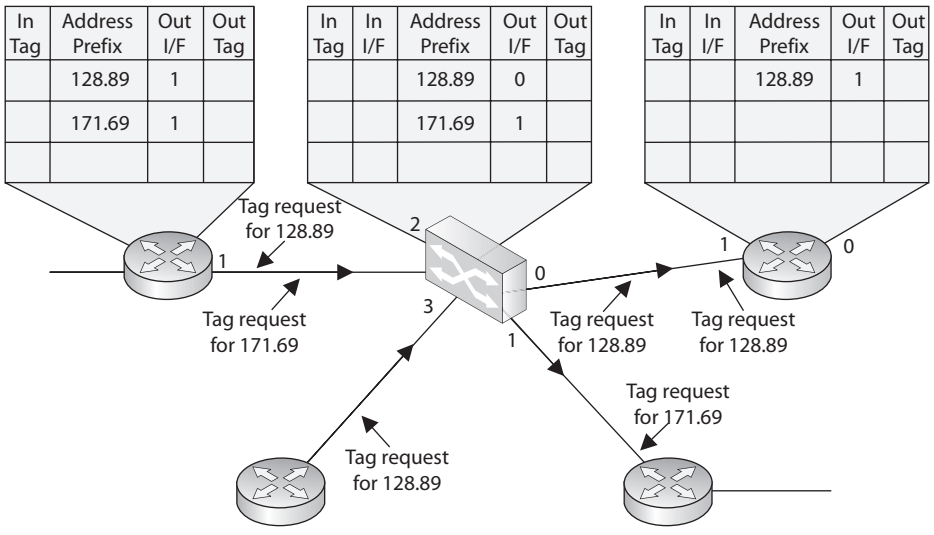



**Figure 14-6** MPLS uses tags and tables for routing functions.

Because security requires control over who can access specific resources, more intelligent devices can provide a higher level of protection because they can make more detail-oriented decisions regarding who can access resources. When devices can look deeper into the packets, they have access to more information to make access decisions, which provides more granular access control.

As previously stated, switching makes it more difficult for intruders to sniff and monitor network traffic because no broadcast and collision information is continually traveling throughout the network. Switches provide a security service that other devices cannot provide. VLANs (described in depth in Chapter 13) are an important part of switching networks, because they enable administrators to have more control over their environment and they can isolate users and groups into logical and manageable entities.

## Routers

We are going up the chain of the OSI layers while discussing various network devices. Repeaters work at the physical layer, bridges and switches work at the data link layer, and routers work at the network layer. As we go up each layer, each corresponding device has more intelligence and functionality because it can look deeper into the frame. A repeater looks at the electrical signal. The switch can look at the MAC address within the header. The router can peel back the first header information and look farther into the frame and find out the IP address and other routing information. The farther a device can look into a frame, the more decisions it can make based on the information within the frame.

*Routers* are layer 3, or network layer, devices that are used to connect similar or different networks. (For example, they can connect two Ethernet LANs or an Ethernet LAN to a Frame Relay link.) A router is a device that has two or more interfaces and a routing table, so it knows how to get packets to their destinations. It can filter traffic based on an access control list (ACL), and it fragments packets when necessary. Because routers have more network-level knowledge, they can perform higher-level functions, such as calculating the shortest and most economical path between the sending and receiving hosts.

A router discovers information about routes and changes that take place in a network through its routing protocols (RIP, BGP, OSPF, and others, as discussed in Chapter 11). These protocols tell routers if a link has gone down, if a route is congested, and if another route is more economical. They also update routing tables and indicate if a router is having problems or has gone down.

The router may be a dedicated appliance or a computer running a networking operating system that is dual-homed. When packets arrive at one of the interfaces, the router compares those packets to its ACL. This list indicates what packets are allowed in and what packets are denied. Access decisions are based on source and destination IP addresses, protocol type, and source and destination ports. An administrator may block all packets coming from the 10.10.12.0 network, any FTP requests, or any packets headed toward a specific port on a specific host, for example. This type of control is provided by the ACL, which the administrator must program and update as necessary.

What actually happens inside the router when it receives a packet? Let's follow the steps:

1. A packet is received on one of the interfaces of a router. The router views the routing data.
2. The router retrieves the destination IP network address from the packet.
3. The router looks at its routing table to see which port matches the requested destination IP network address.
4. If the router does not have information in its table about the destination address, it sends out an ICMP error message to the sending computer indicating that the message could not reach its destination.
5. If the router does have a route in its routing table for this destination, it decrements the TTL value and sees whether the maximum transmission unit (MTU) is different for the destination network. If the destination network requires a smaller MTU, the router fragments the packet.
6. The router changes header information in the packet so that the packet can go to the next correct router, or if the destination computer is on a connecting network, the changes made enable the packet to go directly to the destination computer.
7. The router sends the packet to its output queue for the necessary interface.



Table 14-3 provides a quick review of how routers differ from bridges and switches.

When is it best to use a repeater, bridge, or router? A repeater is used if an administrator needs to expand a network and amplify signals so they do not weaken on longer cables. However, a repeater also extends collision and broadcast domains.

Bridges and switches work at the data link layer and have a bit more intelligence than a repeater. Bridges can do simple filtering and separate collision domains, but not broadcast domains. A switch should be used when an administrator wants to connect multiple computers in a way that reduces traffic congestion and excessive collisions.

A router splits up a network into collision domains and broadcast domains. A router gives more of a clear-cut division between network segments than repeaters or bridges.

| Bridge/Switch | Router |
|---|---|
| Reads header information but does not alter it | Creates a new header for each packet |
| Builds forwarding tables based on MAC addresses | Builds routing tables based on IP addresses |
| Has no concept of network addresses | Assigns a different network address per port |
| Filters traffic based on MAC addresses | Filters traffic based on IP addresses |
| Forwards broadcast traffic | Does not forward broadcast traffic |
| Forwards traffic if a destination address is unknown to the bridge | Does not forward traffic that contains a destination address unknown to the router |

**Table 14-3** Main Differences Between Bridges/Switches and Routers

A router should be used if an administrator wants to have more defined control of where the traffic goes, because more sophisticated filtering is available with routers, and when a router is used to segment a network, the result is more controllable sections.

## Gateways

*Gateway* is a general term for software running on a device that connects two different environments and that many times acts as a translator for them or somehow restricts their interactions. Usually a gateway is needed when one environment speaks a different language, meaning it uses a certain protocol that the other environment does not understand. The gateway can translate mail from one type of mail server and format it so that another type of mail server can accept and understand it, or it can connect and translate different data link technologies such as Fiber Distributed Data Interface (FDDI) to Ethernet (both of which are discussed in Chapter 11).

Gateways perform much more complex tasks than connection devices such as routers and bridges. However, some people refer to routers as gateways when they connect two unlike networks (Token Ring and Ethernet) because the router has to translate between the data link technologies. Figure 14-7 shows how a network access server (NAS) functions as a gateway between telecommunications and network connections.

When networks connect to a backbone, a gateway can translate the different technologies and frame formats used on the backbone network versus the connecting LAN protocol frame formats. If a bridge were set up between an FDDI backbone and an Ethernet LAN, the computers on the LAN would not understand the FDDI protocols and frame formats. In this case, a LAN gateway would be needed to translate the protocols used between the different networks.

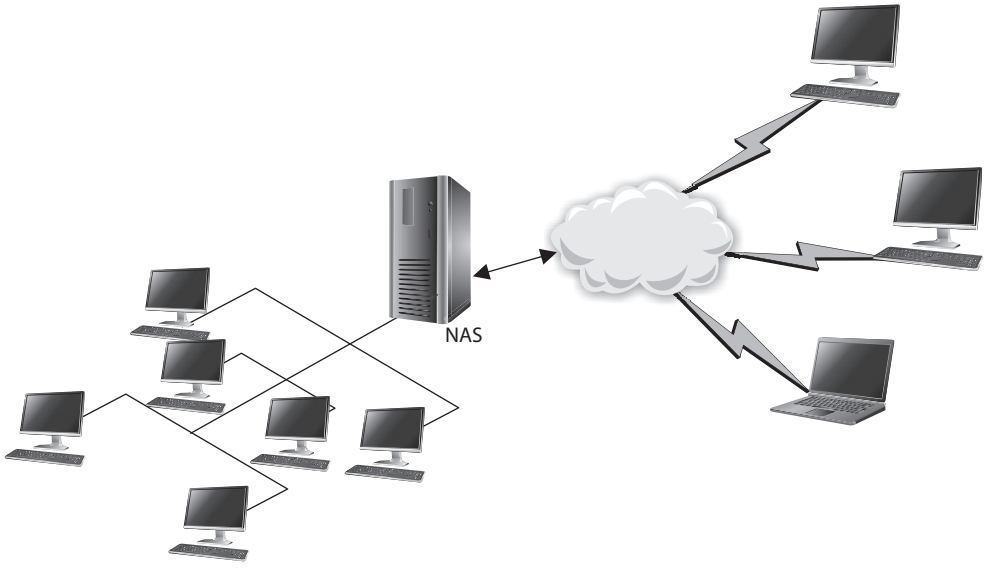



**Figure 14-7** Several types of gateways can be used in a network. A NAS is one example.

A popular type of gateway is an *e-mail* gateway. Because several e-mail vendors have their own syntax, message format, and way of dealing with message transmission, e-mail gateways are needed to convert messages between e-mail server software. For example, suppose that David, whose corporate network uses Sendmail, writes an e-mail message to Dan, whose corporate network uses Microsoft Exchange. The e-mail gateway converts the message into a standard that all mail servers understand—usually X.400—and passes it on to Dan's mail server.

## Proxy Servers

*Proxy servers* act as an intermediary between the clients that want access to certain services and the servers that provide those services. As a security professional, you do not want internal systems to directly connect to external servers without some type of control taking place. For example, if users on your network could connect directly to websites without some type of filtering and rules in place, the users could allow malicious traffic into the network or could surf websites your organization deems inappropriate. To prevent this situation, all internal web browsers should be configured to send their web requests to a web proxy server. The proxy server validates that the request is safe and then sends an independent request to the website on behalf of the user. A very basic proxy server architecture is shown in Figure 14-8.

The proxy server may cache the response it receives from the server so that when other clients make the same request, the proxy server doesn't have to make a connection out to the actual web server again but rather can serve up the necessary data directly. This drastically reduces latency and allows the clients to get the data they need much more quickly.

There are different types of proxies that provide specific services. A *forwarding proxy* is one that allows the client to specify the server it wants to communicate with, as in our scenario earlier. An *open proxy* is a forwarding proxy that is open for anyone to use. An anonymous open proxy allows users to conceal their IP address while browsing websites



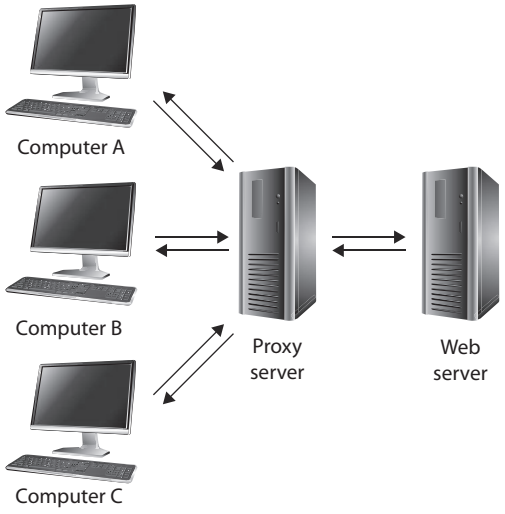



**Figure 14-8** Proxy servers control traffic between clients and servers.

or using other Internet services. A *reverse proxy* appears to the clients as the original server. The client sends a request to what it thinks is the original server, but in reality this reverse proxy makes a request to the actual server and provides the client with the response. The forwarding and reverse proxy functionality seems similar, but as Figure 14-9 illustrates, a forwarding proxy server is commonly on an internal network controlling traffic that is exiting the network. A reverse proxy server is commonly on the network that fulfills clients' requests; thus, it is handling traffic that is entering its network. The reverse proxy can carry out load balancing, encryption acceleration, security, and caching.

Web proxy servers are commonly used to carry out content filtering to ensure that Internet use conforms to the organization's acceptable use policy (AUP). These types of proxies can block unacceptable web traffic, provide logs with detailed information pertaining to the websites specific users visited, monitor bandwidth usage statistics, block restricted website usage, and screen traffic for specific keywords (e.g., porn, confidential, Social Security numbers). The proxy servers can be configured to act mainly as caching servers, which keep local copies of frequently requested resources, allowing organizations to significantly reduce their upstream bandwidth usage and costs while significantly increasing performance.

While the most common use of proxy servers is for web-based traffic, they can be used for other network functionality and capabilities, as in DNS proxy servers. Proxy servers are a critical component of almost every network today. They need to be properly placed, configured, and monitored.

**NOTE** The use of proxy servers to allow for online anonymity has increased over the years. Some people use a proxy server to protect their browsing behaviors from others, with the goal of providing personal freedom and privacy. Attackers use the same functionality to help ensure their activities cannot be tracked back to their local systems.

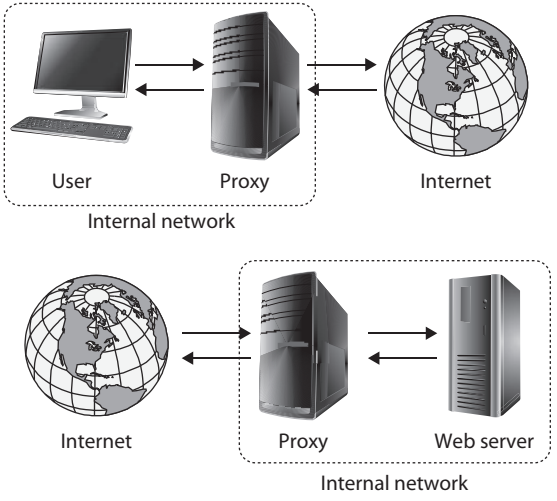



**Figure 14-9** Forward vs. reverse proxy services

### The Tor Network

Tor (originally known as The Onion Router) is a volunteer-operated network of computers around the world that work together to route encrypted web traffic. The goal of Tor is to keep your identity private online, or at least as close to private as is possible. (Misconfigurations or exploitable software on your local machine can still reveal your identity.) Every computer (or node) in Tor receives data from another node and passes it on to the next. Each node only knows where the encrypted data came from and where it's going next. After several hops, someone at the destination has no way of knowing who initiated the connection when you pop back up in the open Internet.

Tor can also provide access to so-called "hidden services" in the deep web that run only inside Tor. The infamous drug marketplace The Silk Road was an example of this. Tor is very popular among privacy advocates and people who live in countries that have strong censorship laws. However, Tor also is commonly used by criminal and even nation-state actors who want to protect their source location. Therefore, you should be extremely suspicious if you see Tor traffic in any enterprise network.



## PBXs

Telephone companies use switching technologies to transmit phone calls to their destinations. A telephone company's central office houses the switches that connect towns, cities, and metropolitan areas through the use of optical fiber rings. So, for example, when

### Putting It All Together: Network Devices

The network devices we've covered so far are the building blocks of almost any network architecture. Table 14-4 lists them and points out their important characteristics.

| Device | OSI Layer | Functionality |
|---|---|---|
| Repeater | Physical | Amplifies the signal and extends networks |
| Bridge | Data link | Forwards packets and filters based on MAC addresses; forwards broadcast traffic, but not collision traffic |
| Switch | Data link | Provides a private virtual link between communicating devices; allows for VLANs; reduces collisions; impedes network sniffing |
| Router | Network | Separates and connects LANs creating internetworks; filters based on IP addresses |
| Gateway | Application | Connects different types of networks; performs protocol and format translations |
| Web proxy | Application | Acts as an intermediary between clients and servers, typically to improve security and/or performance |

**Table 14-4** Main Differences Between Network Devices

Dusty makes a landline phone call from his house, the call first hits the local central office of the telephone company that provides service to Dusty, and then the switch within that office decides whether it is a local or long-distance call and where it needs to go from there. A *Private Branch Exchange (PBX)* is a private telephone switch that is located on an organization's property. This switch performs some of the same switching tasks that take place at the telephone company's central office. The PBX has a dedicated connection to its local telephone company's central office, where more intelligent switching takes place.

A PBX can interface with several types of devices and provides a number of telephone services. The voice data is multiplexed onto a dedicated line connected to the telephone company's central office. Figure 14-10 shows how data from different data sources can be placed on one line at the PBX and sent to the telephone company's switching facility.

PBXs use digital switching devices that can control analog and digital signals. While these modern exchanges are more secure than their analog predecessors, that in no way means PBX systems are free from vulnerabilities. Many PBX systems have system administrator passwords that are hardly ever changed. These passwords are set by default; therefore, if 100 companies purchase and implement 100 PBX systems from the PBX vendor ABC and they do not reset the password, a *phreaker* (a phone hacker) who knows this default password now has access to 100 PBX systems. Once a phreaker breaks into a PBX system, she can cause mayhem by rerouting calls, reconfiguring switches, or configuring the system to provide her and her friends with free long-distance calls. This type of fraud happens more often than most organizations realize because many of them do not closely audit their phone bills. Though the term is not used as much nowadays, phreakers are very much an issue to our telecommunications systems. Toll fraud (as most of their activities are called) associated with PBX systems are estimated to cost over $3 billion in annual losses worldwide, according to the Communications Fraud Control Association's (CFCA) 2019 Fraud Loss Survey.

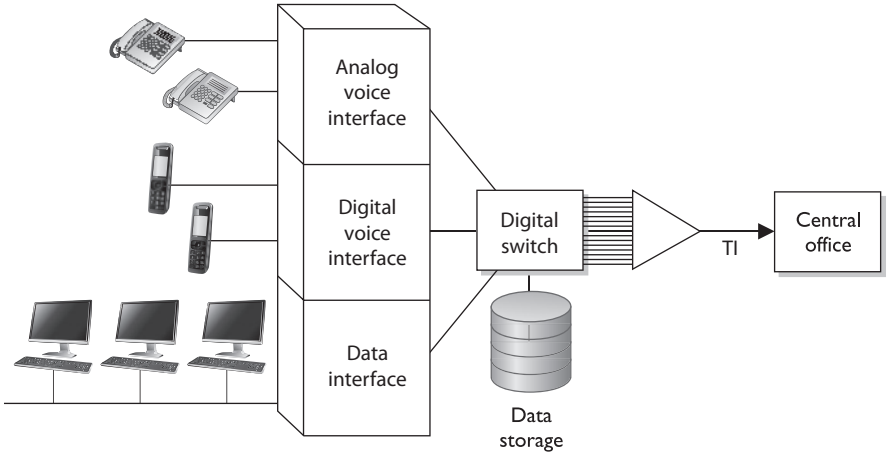



**Figure 14-10** A PBX combines different types of data on the same lines.

PBX systems are also vulnerable to brute force and other types of attacks, in which phreakers use scripts and dictionaries to guess the necessary credentials to gain access to the system. In some cases, phreakers have listened to and changed people's voice messages. So, for example, when people call Bob and reach his voicemail, they might hear not his usual boring message but a new message that is screaming obscenities and insults.

Unfortunately, many security people do not even think about a PBX when they are assessing a network's vulnerabilities and security level. This is because telecommunication devices have historically been managed by service providers and/or by someone on the staff who understands telephony. The network administrator is usually not the person who manages the PBX, so the PBX system commonly does not even get assessed. The PBX is just a type of switch and it is directly connected to the organization's infrastructure; thus, it is a doorway for the bad guys to exploit and enter. These systems need to be assessed and monitored just like any other network device.

So, what should we do to secure PBX systems? Since many of these systems nowadays ride on IP networks, some of the basic security measures will sound familiar. Start by ensuring you know all accounts on the system and that their passwords are strong. Then, ensure that your PBX is updated regularly and that it sits behind your firewall with the appropriate ACLs in place. Other security measures are more specific to a PBX. For example, consider separating your voice and data traffic through these systems by placing them on different VLANs. If one of the VLANs is penetrated, the other could remain secure. Also, limiting the rate of traffic to IP telephony VLANs can slow down an outside attack.



## Network Access Control Devices

*Network access control (NAC)* is any set of policies and controls that we use to, well, control access to our networks. The term implies that we will verify that a device satisfies certain requirements before we let it in. At its simplest level, this could just be user authentication, which was the theme of our discussion of the IEEE 802.1X standard when we were covering wireless network security in Chapter 12. The 802.1X protocol allows devices to connect in a very limited manner (i.e., only to the network authenticator) until we can verify the user credentials it presents.

To fully leverage the power of NAC, however, we should do much more. For starters, we can (and should) authenticate a device. Endpoint/device authentication should be familiar to you because you already use it whenever you establish an HTTPS connection to a web server. When a client requests a secure connection, the server responds with its certificate, which contains its public key issued by a trusted certificate authority (CA). The client then encrypts a secret session key using the server's public key, so only the server can decrypt it and then establish a symmetrically encrypted secure link. It is possible to configure a NAC device to authenticate itself in a similar manner, but also require the client device to do the same. Obviously, we'd need a certificate (and matching private key) installed on the client device for this to work. An alternative approach to using certificates is to use a hardware Trusted Platform Module (TPM) if the endpoint has one. We discussed TPMs in Chapter 9.

A common use of NAC is to ensure the endpoint is properly configured prior to it being allowed to connect to the network. For example, it is pretty common to check the version of the OS as well as the signatures for the antimalware software. If either of these is not current, the device may be placed in an untrusted LAN segment from which it can download and install the required updates. Once the device meets the access policy requirements, it is allowed to connect to the protected network.

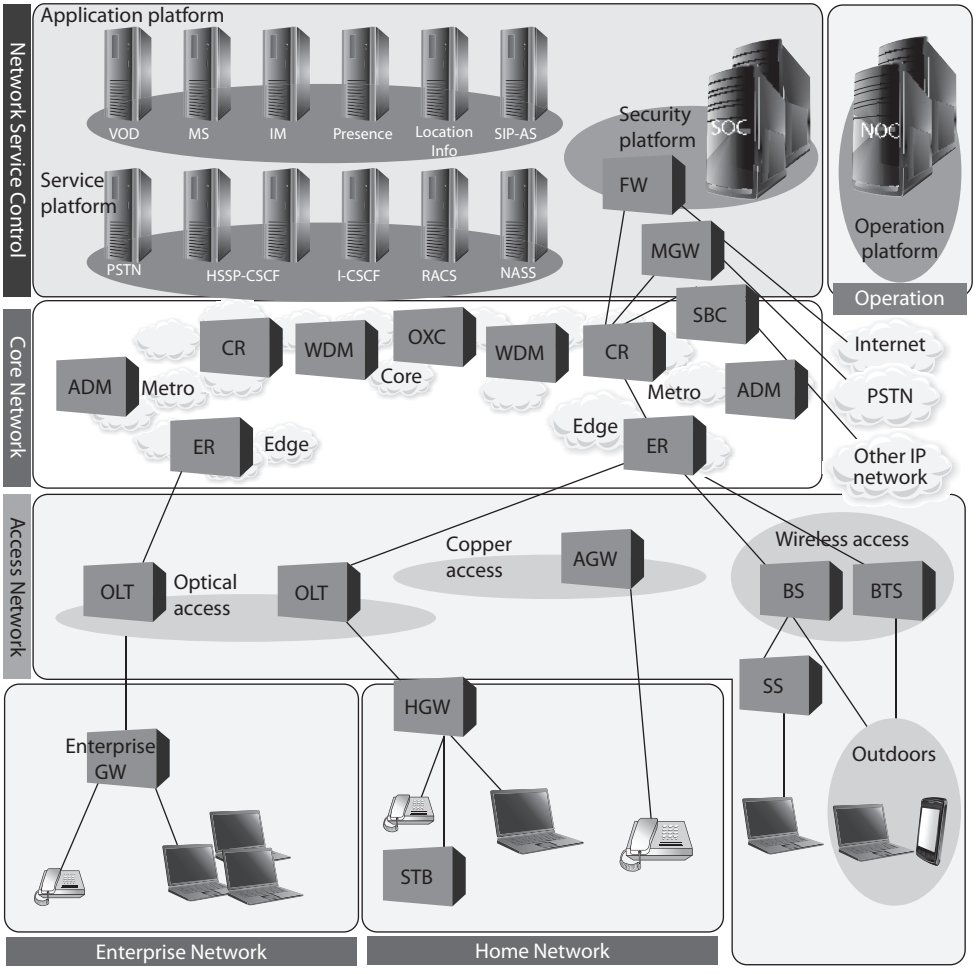
## Network Diagramming

In many cases, you cannot capture a full network in a diagram because of the complexity of most organizations' networks. Sometimes we have a false sense of security when we have a pretty network diagram that we can all look at and be proud of, but let's dig deeper into why this can be deceiving. From what perspective should you look at a network? Many possibilities exist:

- A cabling diagram that shows how everything is physically connected (coaxial, UTP, fiber) and a wireless portion that describes the WLAN structure
- A network diagram that illustrates the network in infrastructure layers of access, aggregation, edge, and core
- A diagram that illustrates how the various networking routing takes place (VLANs, MPLS connections, OSPF, IGRP, and BGP links)
- A diagram that shows how different data flows take place (FTP, IPSec, HTTP, TLS, L2TP, PPP, Ethernet, FDDI, ATM, etc.)
- A diagram that separates workstations and the core server types that almost every network uses (DNS, DHCP, web farm, storage, print, SQL, PKI, mail, domain controllers, RADIUS, etc.)
- A view of a network based upon trust zones, which are enforced by filtering routers, firewalls, and DMZ structures
- A view of a network based upon its IP subnet structure

But what if you look at a network diagram from a Microsoft perspective, which illustrates many of these things but in forest, tree, domain, and OU containers? Then you need to show remote access connections, VPN concentrators, extranets, and the various MAN and WAN connections. How do we illustrate our IP telephony structure? How do we integrate our mobile device administration servers into the diagram? How do we document our new cloud computing infrastructure? How do we show the layers of virtualization within our database? How are redundant lines and fault-tolerance solutions marked? How does this network correlate and interact with our offsite location that carries out parallel processing? And we have not even gotten to our security components (firewalls, IDS, IPS, DLP, antimalware, content filters, etc.). And in the real world,

whatever network diagrams an organization does have are usually out of date because they take a lot of effort to create and maintain.




The point is that a network is a complex beast that cannot really be captured on one piece of paper. Compare it to a human body. When you go into the doctor's office, you see posters on the wall. One poster shows the circulatory system, one shows the muscles, one shows bones, another shows organs, and another shows tendons and ligaments; a dentist's office has a bunch of posters on teeth; if you are at an acupuncture clinic, there will be a poster on acupuncture and reflexology points. And then there is a ton of stuff no one makes posters for—hair follicles, skin, toenails, eyebrows—but these are all part of one system.

So what does this mean to the security professional? You have to understand a network from many different aspects if you are actually going to secure it. You start by learning all this network stuff in a modular fashion, but you need to quickly understand how it all works together under the covers. You can be a complete genius on how everything works within your current environment but not fully understand that when an employee connects her iPhone to her company laptop that is connected to the corporate network and uses it as a modem, this is an unmonitored WAN connection that can be used as a doorway by an attacker. Security is complex and demanding, so do not ever get too cocky, and always remember that a diagram is just showing a perspective of a network, not the whole network.

## Operation of Hardware

Once you have your network designed and implemented, you need to ensure it remains operational. Keep in mind that one of the aspects of security is availability, which can be compromised not only by adversaries but also by power outages, equipment defects, and human error. Remember that all risks, not just the ones that come from human actors, should be addressed by your risk management program. This ensures that you can select cost-effective controls to mitigate those risks. In the sections that follow, we discuss three specific types of controls that protect the availability of your network components. These control types are redundant electrical power, equipment warranties, and support agreements on the operation of our network components.

### Electrical Power

Electrical power is essential to operating IT hardware, which, in turn, runs the software that provides IT services to our organizations. We already discussed this topic generally in Chapter 10, but we now return to it in terms of ensuring our critical systems have redundant power. To understand these power requirements, we need to first become familiar with three key terms that describe electricity:

- **Voltage** Measured in volts, this tells us what the *potential* electric force between two points in a circuit could be. You can think of volts as the water pressure inside a pipe.
- **Current** Measured in amps, this is the *actual* electric flow through the circuit. If you think of volts as the pressure inside a water pipe, you can think of current as the diameter of a valve attached to it; the bigger the valve, the faster the water can come out.
- **Power** There are two ways to measure power. We measure electrical power in watts, which we calculate by multiplying voltage by amperage. In other words, if your server rack is running on 240 volts and drawing 9 amps of current, it is consuming 2,160 watts or 2.16 kilowatts (kW). Another related term is kilowatt-hours (kWh), which is simply the amount of power consumed during a 1-hour period. So, that same server rack would draw 2.16 kWh in one hour, or 51.84 kWh in a day (assuming the current draw is constant).

What we actually care about is whether or not we have enough electric power to run our equipment. There are two ways to measure power: apparent and real. You can think of *apparent power* as the maximum amount of electricity that could get through a circuit in a perfect case. This value is simply the product of the voltage and current of a system, and is measured in volt-amps (VA). So, if you have a 120-volt computer that can draw up to 3 amps, its apparent power would be 360 VA.

Typically, however, the real power drawn by a system is less than its apparent power. This is because of certain complexities of alternating current (AC) circuits that we won't dive into. Suffice it to say that AC, which is the type of current produced from virtually every power outlet, is constantly changing. This variance means that the *real power* drawn by a server will be some value, measured in watts, equal to or (much more frequently) lower than the apparent power. Thankfully, we don't have to calculate this value; most computing equipment is labeled with the real power value in watts (or kilowatts).

Why should you care? Because real power (watts) determines the actual power you purchase from the utility company, the size of any backup generators you might need, and the heat generated by the equipment. Apparent power (VA) is used for sizing wiring and circuit breakers, so the former don't melt (or worse, catch fire) and the latter don't trip. The ratio of real power to apparent power is called the *work factor*, which can never be greater than one (since the denominator is the ideal apparent power).

With all this discussion under our belts, we can now (finally) talk about redundant power, which typically comes in the two forms presented in Chapter 10: uninterruptable power supplies (UPSs) and backup power sources. Suppose one of your organization's facilities has (what will eventually turn out to be) an extended power outage lasting multiple days. Your business continuity plan (BCP; covered in Chapter 2) should identify your mission-critical systems and determine how long they can remain unavailable before your organizational losses are intolerable. You would have addressed this in your facility planning (Chapter 10) by implementing a backup power source. Typically, there is a period between the start of a power outage and when the backup power source comes online and is usable. This is the amount of time during which your UPS systems will have to keep your critical assets running.

To determine how much power you need from your backup power source, you simply add up the power consumption of your critical assets (in kW), keeping in mind the need for cooling and any other supporting systems. Let's say this comes out to be 6 kW and your backup source is a generator. Since generators run optimally at 75 percent to 80 percent of their rated loads, you'd need an 8-kW generator or greater. You also want to factor in room for growth, which should be no less than 25 percent, so you end up getting a 10-kW generator. Now, suppose you also get an automatic transfer switch that will start the generator and transfer the load from critical circuits 60 seconds after the outage is detected. How much UPS capacity do you need?

Whereas the real power consumption that you used to estimate your generator needs probably came from actual readings of how many kilowatts your critical servers drew, your apparent power needs are probably higher because they capture peaks in consumption that are averaged out by real power readings. Remember that apparent power is at least as much as (and usually higher than) your real power. If you look at your equipment's

technical descriptions (or labels) you may see a value measured in volt-ampere (VA or kVA), and all you have to do is add up these values and get a UPS that is rated for that value. Alternatively, a good rule of thumb is to multiply your real power by 1.4 kWA (kilowatt-ampere) per kVA. The resulting number of kVAs should give you sufficient UPS capacity until the generator kicks in.

### Equipment Warranty

Of course, many other things can go wrong with our assets with or without power outages. Equipment failures due to manufacturing defects are, unfortunately, unavoidable in the long run. The good news is that most original equipment manufacturers (OEMs) provide a three-year warranty against such defects. However, you have to read the fine print and may want to upgrade the protections. Suppose that you have a critical server fail and you can only afford to have it down for 24 hours. The standard warranty includes next-day replacement delivery, so you're covered, right? Well, not if you factor in the time it'll take you to reconfigure the server, load up all the data it needs, and put it back into production. Since it is difficult and expensive to get better than next-day support, you may want to build in the cost of having a spare server (or two) in addition to the warranty to ensure you meet your maximum tolerable downtime (MTD).

Most OEMs also offer extended warranties at an additional cost. Depending on your hardware refresh cycle (i.e., how long you will operate equipment before replacing it with new systems), you may want to add one, two, or three more years to the base three-year warranty. This is usually cheaper to purchase when you buy the hardware, as opposed to purchasing it a year or two later. Seven to eight years after the initial purchase, however, warranty offers tend to expire, as the hardware will be too old for the OEM to continue supporting it.

### Support Agreements

Even if your hardware doesn't fail, it could become unavailable (or insufficiently available) with regard to supporting your organizational processes. For example, suppose that a server slows down to the point where your users sit around for several seconds (or even minutes) waiting for a response. This would not only be frustrating but also lead to a loss of productivity that could add up to significant financial losses. If you have a large and well-staffed organization, you probably have a resident expert who can troubleshoot the server and get it back to peak performance. If you don't have such an expert, what do you do?

Many organizations use support agreements with third parties to deal with issues that are outside the expertise of their IT or security staff. Sometimes this support can be provided by the OEM as part of the purchase of a system. Other times, organizations hire a managed services provider (MSP), who not only responds when things go badly but continuously monitors the systems' performance to detect and fix problems as early as possible. Most MSPs charge flat monthly fees per device and include 24/7 remote monitoring, maintenance, and, when needed, onsite support. Think of this as an insurance policy against loss of availability.

# Endpoint Security

An *endpoint* is any computing device that communicates through a network and whose principal function is not to mediate communications for other devices on that network. In other words, if a device is connected to a network but is not part of the routing, relaying, or managing of traffic on that network, then it is an endpoint. That definition leaves out all of the network devices we've discussed in the preceding sections. Endpoints include devices that you would expect, such as desktops, laptops, servers, smartphones, and tablets. However, they also include other devices that many of us don't normally think of, such as point of sale (POS) terminals at retail stores, building automation devices like smart thermostats and other Internet of Things (IoT) devices, and sensors and actuators in industrial control systems (ICS).

One of the greatest challenges in dealing with (and securing) endpoints is knowing they are present in the first place. While it would be extremely unusual (not to say frightening) for your routers and switches to unexpectedly drop in and out of the network, this is what mobile devices do by their very nature. The intermittent connectivity of mobile devices is also a problem when it comes to ensuring that they are properly configured and running the correct firmware, OS, and software versions. An approach to dealing with some of these issues is to use network access control (NAC), as discussed earlier in this chapter.

But mobile devices are not the only problem. Our increasing reliance on embedded systems like IoT and ICS devices poses additional challenges. For starters, embedded devices normally have lesser computing capabilities than other endpoints. You usually can't install security software on them, which means that many organizations simply



### Securing Endpoints

Endpoint security really boils down to a handful of best practices. Sure, you should thoroughly analyze risks to your endpoints and implement cost-effective controls as part of a broader risk management program, but if you don't take care of the basic "tackling and blocking," then whatever else you do won't really make much of a difference. Here's a short list to get you started:

- Know what every single endpoint is, where it is, who uses it, and what it should (and should not) be doing.
- Strictly enforce least privilege (i.e., no regular users with local admin rights).
- Keep everything updated (ideally, do this automatically).
- Use endpoint protection and response (EDR) solutions.
- Back up everything (ideally in a way that is difficult for an attacker to compromise).
- Export endpoint logs to a security information and event management (SIEM) solution.

create security perimeters or bubbles around them and hope for the best. Just to make things even more interesting, IoT and ICS devices oftentimes control physical processes like heating, ventilation, and air conditioning (HVAC) that can have effects on the health and safety of the people in our organizations.

## Content Distribution Networks

So far, our discussion of networking has sort of implied that there is *a* (singular) web server, a (singular) database server, and so on. While this simplifies our discussion of network foundations, protocols, and services, we all know that this is a very rare scenario in all but the smallest networks. Instead, we tend to implement multiples of each service, whether to segment systems, provide redundancy, or both. We may have a couple of web servers connected by a load balancer and interfacing with multiple backend database servers. This sort of redundant deployment can improve performance, but all clients still have to reach the same physical location regardless of where in the world they may be. Wouldn't it be nice if users in Europe did not have to ride transatlantic cables or satellite links to reach a server in the United States and instead could use one closer to them?

A *content distribution network (CDN)* consists of multiple servers distributed across a large region, each of which provides content that is optimized for users closest to it. This optimization can come in many flavors. For example, if you were a large streaming video distribution entity like Netflix, you would want to keep your movie files from having to traverse multiple links between routers, since each hop would incur a delay and potential loss of packets (which could cause jitter in the video). Reducing the number of network hops for your video packets would also usually mean having a server geographically closer to the other node, offering you the opportunity to tailor the content for users in that part of the world. Building on our video example, you could keep movies dubbed in Chinese on servers that are in or closer to Asia and those dubbed in French closer to Europe. So when we talk about optimizing content, we can mean many things.

Another benefit of using CDNs is that they make your Internet presence more resistant to distributed denial-of-service (DDoS) attacks. These attacks rely on having a large number of computers flood a server until it becomes unresponsive to legitimate requests. If an attacker can muster a DDoS attack that can send a million packets per second (admittedly fairly small by today's standards) and aim it at a single server, then it could very well be effective. However, if the attacker tries that against a server that is part of a CDN, the clients will simply start sending their requests to other servers in the network. If the attacker then directs a portion of his attack stream to each server on the CDN in hopes of bringing the whole thing down, the attack will obviously be diffused and would likely require many times more packets. Unsurprisingly, using CDNs is how many organizations protect themselves against DDoS attacks.

## Chapter Review

The physical components that make up our networks are foundational to our information systems. Without these cables and switches and routers, nothing else would work. This may seem obvious, but when was the last time you inspected any of them to ensure

that they are secure, in good condition, properly configured, and well supported by appropriate third parties? The two classes of threat actors with which we should concern ourselves in this context are attackers and nature. We take care of the first by applying the principles of secure design we've discussed throughout the book and, particularly, by physically securing these cables and devices as discussed in Chapter 10. As far as natural threats, we need to be on the lookout for the wear and tear that is natural over time and that can exacerbate small product defects that may not have been apparent during our initial inspections of new products. This boils down to having qualified staff that is augmented, as necessary, by third parties that provide warranty and support services.

## Quick Review

- Analog signals represent data as continuously changing wave values, while digital signals encode data in discrete voltage values.
- Digital signals are more reliable than analog signals over a long distance and provide a clear-cut and efficient signaling method because the voltage is either on (1) or not on (0), compared to interpreting the waves of an analog signal.
- Synchronous communications require a timing component but ensure reliability and higher speeds; asynchronous communications require no timing component and are simpler to implement.
- A baseband technology uses the entire communication channel for its transmission, whereas a broadband technology divides the communication channel into individual and independent subchannels so that different types of data can be transmitted simultaneously.
- Coaxial cable has a copper core that is surrounded by a shielding layer and grounding wire, which makes it more resistant to electromagnetic interference (EMI), provides a higher bandwidth, and supports the use of longer cable lengths.
- With twisted-pair cable, the twisting of the wires, the type of insulation used, the quality of the conductive material, and the shielding of the wire determine the rate at which data can be transmitted.
- Fiber-optic cabling carries data as light waves, is expensive, can transmit data at high speeds, is difficult to tap into, and is resistant to EMI and RFI. If security is extremely important, fiber-optic cabling should be used.
- Because it uses glass, fiber-optic cabling has higher transmission speeds that allow signals to travel over longer distances.
- Depending on the material used, network cables may be susceptible to noise, attenuation, and crosstalk.
- Line noise refers to random fluctuations in electrical-magnetic impulses that are carried along a physical medium.
- Attenuation is the loss of signal strength as it travels.
- Crosstalk is a phenomenon that occurs when electrical signals of one wire spill over to the signals of another wire.

- Bandwidth is the amount of information that can theoretically be transmitted over a link within a second.
- Data throughput is the actual amount of data that can actually be carried over a real link.
- A repeater provides the simplest type of connectivity because it only repeats electrical signals between cable segments, which enables it to extend a network.
- A bridge is a LAN device used to connect LAN segments (or VLAN segments) and thus extends the range of a LAN.
- A transparent bridge starts to learn about the network's environment as soon as it is powered on and continues to learn as the network changes by examining frames and making entries in its forwarding tables.
- Spanning Tree Protocol (STP) ensures that forwarded frames do not circle networks forever, provides redundant paths in case a bridge goes down, assigns unique identifiers to each bridge, assigns priority values to these bridges, and calculates path costs.
- The Shortest Path Bridging (SPB) protocol is defined in IEEE 802.1aq and is more efficient and scalable than STP; it is used in newer bridges.
- Switches are multiport bridges that typically have additional management features.
- Routers are layer 3, or network layer, devices that are used to connect similar or different networks.
- Routers link two or more network segments, where each segment can function as an independent network. A router works at the network layer, works with IP addresses, and has more network knowledge than bridges, switches, or repeaters.
- Gateway is a general term for software running on a device that connects two different environments and that many times acts as a translator for them or somehow restricts their interactions.
- A Private Branch Exchange (PBX) is a private telephone switch that is located on an organization's property and performs some of the same switching tasks that take place at the telephone company's central office.
- Proxy servers act as an intermediary between the clients that want access to certain services and the servers that provide those services.
- Network access control (NAC) is any set of policies and controls that restrict access to our networks.
- An endpoint is any computing device that communicates through a network and whose principal function is not to mediate communications for other devices on that network.
- A content distribution network (CDN) consists of multiple servers distributed across a large region, each of which provides content that is optimized for users closest to it.

## Questions

Please remember that these questions are formatted and asked in a certain way for a reason. Keep in mind that the CISSP exam is asking questions at a conceptual level. Questions may not always have the perfect answer, and the candidate is advised against always looking for the perfect answer. Instead, the candidate should look for the best answer in the list.

1. Which of the following is true of asynchronous transmission signals?
   - **A.** Used for high-speed, high-volume transmissions
   - **B.** Robust error checking
   - **C.** Used for irregular transmission patterns
   - **D.** More complex, costly implementation
2. Which of the following technologies divides a communication channel into individual and independent subchannels?
   - **A.** Baseband
   - **B.** Broadband
   - **C.** Circuit-switched
   - **D.** Crosstalk
3. What type of cabling would you use if you needed inexpensive networking in an environment prone to electromagnetic interference?
   - **A.** Fiber-optic
   - **B.** Unshielded twisted pair (UTP)
   - **C.** Plenum
   - **D.** Coaxial
4. Which of the following issues would be likeliest to cause problems in a cable tray where large numbers of cables run in parallel and close proximity?
   - **A.** Thermal noise
   - **B.** Line noise
   - **C.** Crosstalk
   - **D.** Attenuation
5. What problem is inevitable as the length of a cable run increases?
   - **A.** Thermal noise
   - **B.** Line noise
   - **C.** Crosstalk
   - **D.** Attenuation

6. What is the term for the maximum amount of data that actually traverses a given network link?
   - **A.** Latency
   - **B.** Bandwidth
   - **C.** Throughput
   - **D.** Maximum transmission unit (MTU)
7. Which protocol ensures that frames being forwarded by switches do not circle networks forever?
   - **A.** Open Shortest Path First (OSPF)
   - **B.** Border Gateway Protocol (BGP)
   - **C.** Intermediate System-to-Intermediate System (IS-IS)
   - **D.** Spanning Tree Protocol (STP)
8. Which standard specifically addresses issues in network access control?
   - **A.** IEEE 802.1Q
   - **B.** IEEE 802.1aq
   - **C.** IEEE 802.AE
   - **D.** IEEE 802.1X
9. Which of the following would not be considered an endpoint?
   - **A.** Point of sale (POS) terminal
   - **B.** Industrial control system (ICS)
   - **C.** Internet of Things (IoT) device
   - **D.** Multiprotocol Label Switching (MPLS) system
10. All of the following are good reasons to implement a content distribution network except for which one?
    - **A.** Reduced latency
    - **B.** Reduced total cost of ownership (TCO)
    - **C.** Protection against distributed denial-of-service (DDoS) attacks
    - **D.** Tailoring content to users around the world

## Answers

1. **C.** Asynchronous communications are typically used when data transfers happen at lower volumes and with unpredictable intervals. All other answers describe synchronous signaling, which is best suited for regular, high-volume traffic.

2. **B.** A broadband technology divides the communication channel into individual and independent subchannels so that different types of data can be transmitted simultaneously. A baseband technology, on the other hand, uses the entire communication channel for its transmission.
3. **D.** Coaxial cable has a copper core that is surrounded by a shielding layer and grounding wire, which makes it more resistant to electromagnetic interference (EMI). It is significantly cheaper than fiber-optic cable, which is the other EMI-resistant answer listed, while still allowing higher bandwidths.
4. **C.** Crosstalk is a phenomenon that occurs when electrical signals of one wire spill over to the signals of another wire. The more cables you have in close proximity, the worse this issue can be unless you use shielded cables.
5. **D.** Attenuation is the loss of signal strength as it travels. Regardless of which type of cabling is used, attenuation is inevitable given a long enough distance, which is why repeaters were invented.
6. **C.** Data throughput is the actual amount of data that can be carried over a real link. Bandwidth, on the other hand, is the amount of information that can theoretically be transmitted over a link within a second.
7. **D.** Spanning Tree Protocol (STP) ensures that forwarded frames do not circle networks forever, provides redundant paths in case a bridge goes down, assigns unique identifiers to each bridge, assigns priority values to these bridges, and calculates path costs. The other answers are all routing (layer 3) protocols.
8. **D.** The 802.1X protocol allows devices to connect in a very limited manner (i.e., only to the network authenticator) until the device and/or user can be authenticated. The other standards listed all pertain to layer 2 bridging and security.
9. **D.** An endpoint is any computing device that communicates through a network and whose principal function is not to mediate communications for other devices on that network. MPLS functionality is built into networking devices to help them move packets between endpoints more efficiently.
10. **B.** A content distribution network (CDN) consists of multiple servers distributed across a large region, each of which provides content that is optimized for users closest to it. This improves latency and localization. The very distributed nature of the CDN also provides DDoS protections. It all comes at significant costs and increases the complexity of deploying systems and content, which may require additional organizational resources apart from the service itself.

*This page intentionally left blank*

CHAPTER 15

# Secure Communications Channels

This chapter presents the following:

- Voice communications
- Multimedia collaboration
- Remote access
- Data communications
- Virtualized networks
- Third-party connectivity

*Mr. Watson—come here—I want to see you.*

—Alexander Graham Bell

Up to this point, we've treated all the data as if it were equal. While it is true that a packet is a packet regardless of its contents, there are a number of common cases in which the purpose of a communication matters a lot. If we're downloading a file from a server, we normally don't care (or even know about) the variation in delay times between consecutive packets. This variation, known as *packet jitter*, could mean that some packets follow each other closely (no variance) while others take a lot longer (or shorter) time to arrive. While packet jitter is largely inconsequential to our file download, it could be very problematic for voice, video, or interactive collaboration communications channels.

Implementing secure communications channels has always been important to most organizations. However, the sudden shift to remote working brought on by COVID-19 has made the security of these channels critical due to the convergence of increased demand by legitimate users and increased targeting by threat actors. In this chapter, we look at some of the most prevalent communications channels that ride on our networks. These include voice, multimedia collaboration, remote access, and third-party channels. Let's start with the one we're most accustomed to: voice communications.

# Voice Communications

Voice communications have come a long way since Alexander Graham Bell made that first call in 1876. It is estimated that 95 percent of the global population has access to telephone service, with most of those being cellular systems. What ties global voice networks together is a collection of technologies, some of which we've discussed before (e.g., ATM in Chapter 11 and LTE in Chapter 12), and some to which we now turn our attention.

## Public Switched Telephone Network

The traditional telephone system is based on a circuit-switched, voice-centric network called the *public switched telephone network (PSTN)*. The PSTN uses circuit switching instead of packet switching. When a phone call is made, the call is placed at the PSTN interface, which is the user's telephone. This telephone is connected to the telephone company's local loop via electric wires, optical fibers, or a radio channel. Once the signals for this phone call reach the telephone company's central office (the end of the local loop), they are part of the telephone company's circuit-switching world. A connection is made between the source and the destination, and as long as the call is in session, the data flows through the same switches.

When a phone call is made, the phone numbers have to be translated, the connection has to be set up, signaling has to be controlled, and the session has to be torn down. This takes place through the Signaling System 7 (SS7) protocol. Figure 15-1 illustrates how calls are made in the PSTN using SS7. Suppose Meeta calls Carlos. Meeta's phone is directly connected to a signal switching point (SSP) belonging to the telephone company (telco) that provides her service. Her telco's SSP finds the SSP of the telco providing Carlos's phone service and they negotiate the call setup. The call itself is routed over

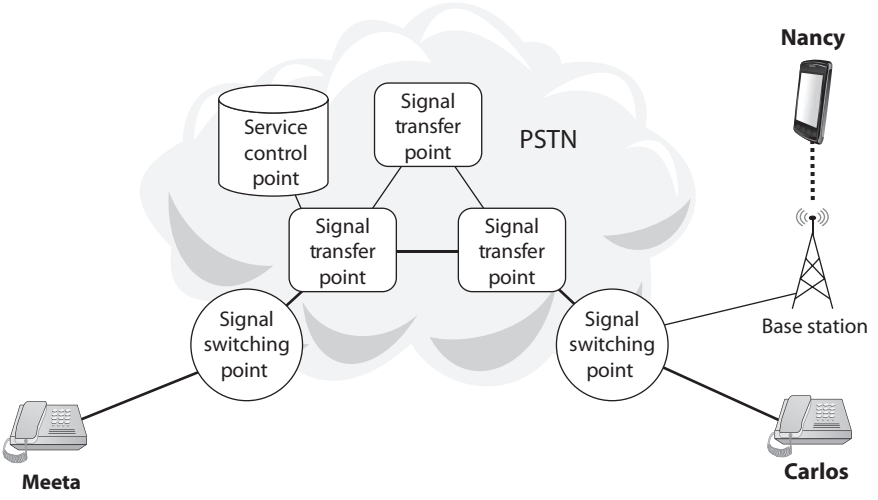



**Figure 15-1** Major components of a public switched telephone network

the two signal transfer points (STPs) that interconnect the two SSPs. STPs perform a similar function in a circuit-switched network as routers do in an IP network. If Meeta wanted to call (or conference in) Nancy on her mobile phone, her SSP could query a service control point (SCP), which controls advanced features such as finding mobile subscribers' SSPs and enabling conference calls involving multiple networks.

**NOTE** PSTNs are being replaced with IP telephony. In the UK, for example, the service provider BT announced that it will switch off its PSTN in 2025.

## DSL

It turns out that PSTN local loops (i.e., the telephone wires that go into our homes and offices) are able to support much more bandwidth than the small amount required for voice communications. In the 1980s, telcos figured out that they could transmit digital data at frequencies above those used for voice calls without interference. This was the birth of *digital subscriber line (DSL)*, which is a high-speed communications technology that simultaneously transmits analog voice and digital data between a home or business and the service provider's central office.

Figure 15-2 shows a typical DSL network. In the subscriber's home, a DSL modem creates a LAN to which computers and wireless access points can be connected. This modem, in turn, is connected to a DSL splitter if the home also has analog phone service. A bunch of DSL subscribers in the same neighborhood are then connected to a DSL access multiplexer (DSLAM) in the central office, where analog signals are sent to a voice switch (and on to the PSTN) and digital signals are routed out to the Internet. The tricky part is that the maximum distance between the DSLAM and the DSL splitter

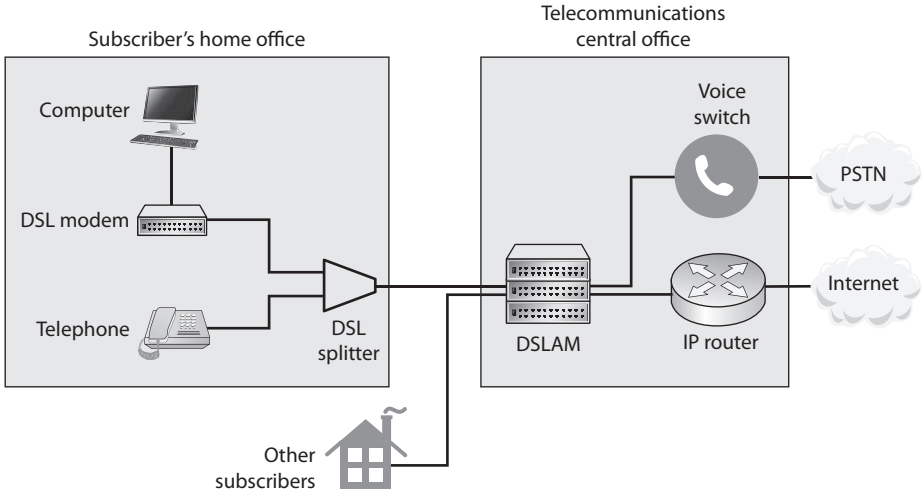



**Figure 15-2** DSL network

in the subscriber's home cannot be greater than about 2.5 miles unless you put extenders in place to boost the signal strength.

DSL offers two broad types of services. With *symmetric services*, traffic flows at the same speed upstream and downstream (to and from the Internet or destination). With *asymmetric services*, the downstream speed is much higher than the upstream speed. The vast majority of DSL lines in use today are asymmetric, because most users usually download much more data from the Internet than they upload. The following are some of the most common types of DSL service:

- **Asymmetric DSL (ADSL)** These lines allocate more bandwidth for downstream data than for upstream. The technology has gone through multiple upgrades, with ADSL2+ (ITU standard G.992.5) being the latest and fastest. It has data rates of up to 24 Mbps downstream and 1.4 Mbps upstream, but can only support distances of about a mile from the central office. ADSL is generally used by residential users.
- **Very high-data-rate DSL (VDSL)** VDSL is basically ADSL at much higher data rates (up to 300 Mbps downstream and 100 Mbps upstream). It is capable of supporting high-bandwidth applications such as HDTV, telephone services (Voice over IP), and general Internet access over a single connection.
- **G.fast** Since the biggest challenge with DSL is the length of the subscriber loop, why not run fiber-optic cable from the central office to a distribution point near the home and then finish the last few hundred feet using the copper wires that are already in place? This is what G.fast (ITU standards G.9700 and G.9701) does. It can deliver data rates of up to 1 Gbps.

### Dial-up Connections

Dial-up modems using PSTN were the dominant form of remote access in the early days of the Internet. Antiquated as they may seem, some organizations still have modems enabled, sometimes without the network staff being aware of them. For example, we once discovered that the facilities manager at a large school district installed a dial-up modem so he could control the HVAC systems remotely during inclement weather. Therefore, it is important to search for these systems and ensure no unauthorized modems are attached and operational.

If you find yourself using modems, some of the security measures that you should put in place for dial-up connections include

- Disable and remove nonessential modems.
- Configure the remote access server to call back the initiating phone number to ensure it is valid and authorized.
- Consolidate all modems into one location and manage them centrally, if possible.
- Whenever possible, implement use of two-factor authentication, VPNs, and NAC for remote access connections.

**NOTE** Despite being in wide use, DSL is an obsolescent technology. Major telecommunications companies around the world have announced plans to phase out DSL by 2025.

## ISDN

*Integrated Services Digital Network (ISDN)* is another technology that leverages legacy telephone lines to enable data, voice, and signaling traffic to travel over a medium in a digital manner previously used only for analog voice transmission. ISDN uses the same wires and transmission medium used by analog dial-up technologies, but it works in a digital fashion. If a computer uses a modem to communicate with an ISP, the modem converts the data from digital to analog to be transmitted over the phone line. If that same computer was configured to use ISDN and had the necessary equipment, it would not need to convert the data from digital to analog, but would keep it in a digital form. This, of course, means the receiving end would also require the necessary equipment to receive and interpret this type of communication properly. Communicating in a purely digital form provides higher bit rates that can be sent more economically.

ISDN is a set of telecommunications services that can be used over public and private telecommunications networks. It provides a digital, point-to-point, circuit-switched medium and establishes a circuit between the two communicating devices. An ISDN connection can be used for anything a modem can be used for, but it provides more functionality and higher bandwidth. This digital service can provide bandwidth on an



### ISDN Examined

ISDN breaks the telephone line into different channels and transmits data in a digital form rather than the old analog form. Three ISDN implementations are in use:

- **Basic Rate Interface (BRI) ISDN** This implementation operates over existing copper lines at the local loop and provides digital voice and data channels. It uses two B channels (at 64 Kbps each) to support user data or voice and one D channel (at 16 Kbps) for signaling, with a combined bandwidth of 144 Kbps. BRI ISDN is generally used for home and small office subscribers.
- **Primary Rate Interface (PRI) ISDN** This implementation has up to 23 B channels and 1 D channel, at 64 Kbps per channel. The total bandwidth is equivalent to a T1, which is 1.544 Mbps. This would be more suitable for an organization that requires a higher amount of bandwidth compared to BRI ISDN.
- **Broadband ISDN (BISDN)** This implementation can handle many different types of services simultaneously and is mainly used within telecommunications carrier backbones. When BISDN is used within a backbone, ATM is commonly employed to encapsulate data at the data link layer into cells, which travel over a SONET network.

as-needed basis and can be used for LAN-to-LAN on-demand connectivity, instead of using an expensive dedicated link.

Analog telecommunication signals use a full channel for communication, but ISDN can break up this channel into multiple channels to move various types of data and provide full-duplex communication and a higher level of control and error handling. ISDN provides two basic services: *Basic Rate Interface (BRI)* and *Primary Rate Interface (PRI)*.

BRI has two B channels that enable data to be transferred and one D channel that provides for call setup, connection management, error control, caller ID, and more. The bandwidth available with BRI is 144 Kbps, and BRI service is aimed at the small office and home office (SOHO) market. The D channel provides for a quicker call setup and process in making a connection compared to dial-up connections. An ISDN connection may require a setup connection time of only 2 to 5 seconds, whereas a modem may require a timeframe of 45 to 90 seconds. This D channel is an out-of-band communication link between the local loop equipment and the user's system. It is considered "out-of-band" because the control data is not mixed in with the user communication data. This makes it more difficult for a would-be defrauder to send bogus instructions back to the service provider's equipment in hopes of causing a denial of service (DoS), obtaining services not paid for, or conducting some other type of destructive behavior.

PRI has 23 B channels and one D channel, and is more commonly used in corporations. The total bandwidth is equivalent to a T1, which is 1.544 Mbps.

ISDN is not usually the primary telecommunications connection for organizations, but it can be used as a backup in case the primary connection goes down. An organization can also choose to implement *dial-on-demand routing (DDR)*, which can work over ISDN. DDR allows an organization to send WAN data over its existing telephone lines and use the PSTN as a temporary type of WAN link. It is usually implemented by organizations that send out only a small amount of WAN traffic and is a much cheaper solution than a real WAN implementation. The connection activates when it is needed and then idles out.

**NOTE** ISDN has lost popularity over the years and is now a legacy technology that is seldom used. Some organizations still rely on it as a backup for communications.

## Cable Modems

The cable television companies have been delivering television services to homes for years, and then they started delivering data transmission services for users who have cable modems and want to connect to the Internet at high speeds. *Cable modems* provide high-speed access to the Internet through existing cable coaxial and fiber lines. The cable modem provides upstream and downstream conversions.

Coaxial and fiber cables are used to deliver hundreds of television stations to users, and one or more of the channels on these lines are dedicated to carrying data. The bandwidth is shared between users in a local area; therefore, it will not always stay at a

static rate. So, for example, if Mike attempts to download a program from the Internet at 5:30 P.M., he most likely will have a much slower connection than if he had attempted it at 10:00 A.M., because many people come home from work and hit the Internet at the same time. As more people access the Internet within his local area, Mike's Internet access performance drops.

Most cable providers comply with *Data-Over-Cable Service Interface Specifications (DOCSIS)*, which is an international telecommunications standard that allows for the addition of high-speed data transfer to an existing cable TV (CATV) system. DOCSIS includes MAC layer security services in its Baseline Privacy Interface/Security (BPI/SEC) specifications. This protects individual user traffic by encrypting the data as it travels over the provider's infrastructure.

## IP Telephony

*Internet Protocol (IP) telephony* is an umbrella term that describes carrying telephone traffic over IP networks. So, if we have all these high-speed digital telecommunications services and the ability to transmit Voice over IP (VoIP) networks, do we even need analog telephones anymore? The answer is a resounding no. PSTN is being replaced by data-centric, packet-oriented networks that can support voice, data, and video. The new IP telephony networks use more efficient and secure switches, protocols, and communication links compared to PSTN but must still coexist (for now) with this older network. This means that VoIP is still going through a tricky transition stage that enables the old systems and infrastructures to communicate with the new systems until the old systems are dead and gone.

This technology gets around some of the barriers present in the PSTN today. The PSTN interface devices (telephones) have limited embedded functions and logic, and the PSTN environment as a whole is inflexible in that new services cannot be easily added. In VoIP, the interface to the network can be a computer, server, PBX, or anything else that runs a telephone application. This provides more flexibility when it comes to adding new services and provides a lot more control and intelligence to the interfacing devices. The traditional PSTN has basically dumb interfaces (telephones without much functionality), and the telecommunication infrastructure has to provide all the functionality. In VoIP, the interfaces are the "smart ones" and the network just moves data from one point to the next.

Because VoIP is a packet-oriented switching technology, the arrival times of different packets may not be regular. You may get a bunch of packets close to each other and then have random delays until the next ones arrive. This irregularity in arrival rates is referred to as *jitter*, which can cause loss of synchronicity in the conversation. It typically means the packets holding the other person's voice message got queued somewhere within the network or took a different route. VoIP includes protocols to help smooth out these issues and provide a more continuous telephone call experience.


**EXAM TIP** Applications that are time sensitive, such as voice and video signals, need to work over an isochronous network. An isochronous network contains the necessary protocols and devices that guarantee regular packet interarrival times.

Four main components are normally used for VoIP: an IP telephony device, a call-processing manager, a voicemail system, and a voice gateway. The *IP telephony device* is just a phone that has the necessary software that allows it to work as a network device. Traditional phone systems require a "smart network" and a "dumb phone." In VoIP, the phone must be "smart" by having the necessary software to take analog signals, digitize them, break them into packets, and create the necessary headers and trailers for the packets to find their destination. The *voicemail system* is a storage place for messages and provides user directory lookups and call-forwarding functionality. A *voice gateway* carries out packet routing and provides access to legacy voice systems and backup calling processes.

When a user makes a call, his VoIP phone sends a message to the *call-processing manager* to indicate a call needs to be set up. When the person at the call destination takes her phone off the hook, this notifies the call-processing manager that the call has been accepted. The call-processing manager notifies both the sending and receiving phones that the channel is active, and voice data is sent back and forth over a traditional data network line.

Moving voice data through packets is more involved than moving regular data through packets. This is because voice (and video) data must be sent as a steady stream, whereas other types of traffic are more tolerant to burstiness and jitter. A delay in data transmission is not noticed as much as is a delay in voice transmission. VoIP systems have advanced features to provide voice data transmission with increased bandwidth, while reducing variability in delay, round-trip delay, and packet loss issues. These features are covered by two relevant standards: H.323 and the Session Initiation Protocol (SIP).

**NOTE** A media gateway is the translation unit between disparate telecommunications networks. VoIP media gateways perform the conversion between TDM voice and VoIP, for example.

### VoIP vs. IP Telephony

The terms "IP telephony" and "Voice over IP" are used interchangeably, but there is a distinction:

- The term "VoIP" is widely used to refer to the actual services offered: caller ID, QoS, voicemail, and so on.
- IP telephony is an umbrella term for all real-time applications over IP, including voice over instant messaging (IM) and video conferencing.

So, "IP telephony" means that telephone and telecommunications activities are taking place over an IP network instead of the traditional PSTN. "Voice over IP" means voice data is being moved over an IP network instead of the traditional PSTN. They are basically the same thing, but VoIP focuses more on the telephone call services.

## H.323

The ITU-T *H.323* recommendation is a standard that deals with audio and video calls over packet-based networks. H.323 defines four types of components: terminals, gateways, multipoint control units, and gatekeepers. The *terminals* can be dedicated VoIP telephone sets, videoconferencing appliances, or software systems running on a traditional computer. *Gateways* interface between H.323 and non-H.323 networks, providing any necessary protocol translation. These gateways are needed, for instance, when using the PSTN to connect H.323 systems. *Multipoint control units (MCUs)* allow three or more terminals to be conferenced together and are sometimes referred to as *conference call bridges*. Finally, the H.323 *gatekeeper* is the central component of the system in that it provides call control services for all registered terminals.

## Session Initiation Protocol

An alternative standard for voice and video calls is the *Session Initiation Protocol (SIP)*, which can be used to set up and break down the call sessions, just as SS7 does for PSTN calls. SIP is an application layer protocol that can work over TCP or UDP. It provides the foundation to allow the phone-line features that SS7 provides, such as causing a phone to ring, dialing a phone number, generating busy signals, and so on. SIP is used in applications such as video conferencing, multimedia, instant messaging, and online gaming.

SIP consists of two major components: the *User Agent Client (UAC)* and *User Agent Server (UAS)*. The UAC is the application that creates the SIP requests for initiating a communication session. UACs are generally messaging tools and soft-phone applications that are used to place VoIP calls. The UAS is the SIP server, which is responsible for handling all routing and signaling involved in VoIP calls.

SIP relies on a three-way-handshake process to initiate a session. To illustrate how a SIP-based call kicks off, let's look at an example of two people, Bill and John, trying to communicate using their VoIP phones. Bill's system starts by sending an INVITE message to John's system. Since Bill's system is unaware of John's location, the INVITE message is sent to the SIP server, which looks up John's address in the SIP *registrar* server. Once the location of John's system has been determined, the INVITE message is forwarded to his system. During this entire process, the server keeps the caller (Bill) updated by sending his system a Trying response, indicating the process is underway. Once the INVITE message reaches John's system, it starts ringing. While John's system rings and waits for John to respond, it sends a Ringing response to Bill's system, notifying Bill that the INVITE has been received and John's system is waiting for John to accept the call. As soon as John answers the call, an OK packet is sent to Bill's system (through the server). Bill's system now issues an ACK packet to begin call setup. It is important to note here that SIP itself is not used to stream the conversation because it's just a signaling protocol. The actual voice stream is carried on media protocols such as the *Real-time Transport Protocol (RTP)*. RTP provides a standardized packet format for delivering audio and video over IP networks. Once Bill and John are done communicating, a BYE message is sent from the system terminating the call. The other system responds with an OK, acknowledging the session has ended. This handshake is illustrated in Figure 15-3.

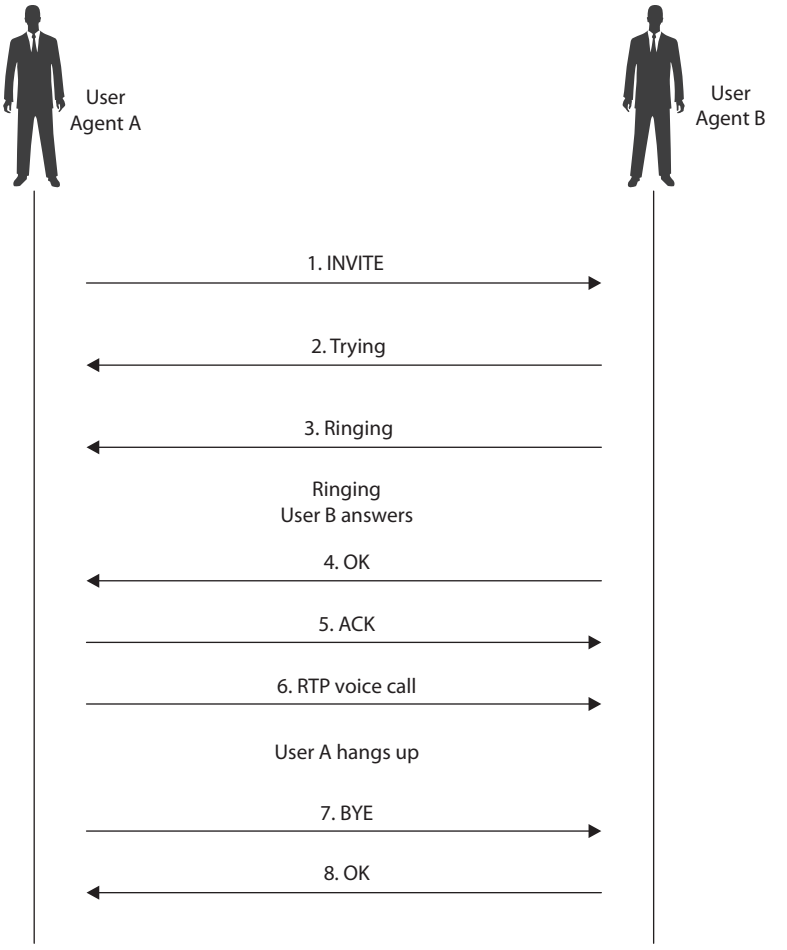
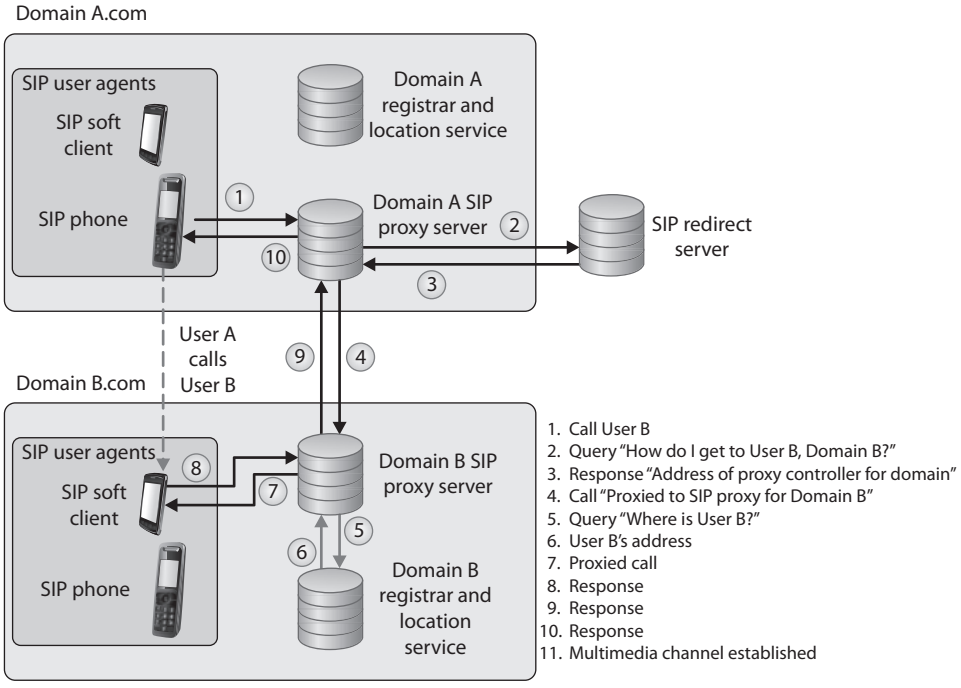



**Figure 15-3**
SIP handshake

The SIP architecture consists of three different types of servers, which play an integral role in the entire communication process of the VoIP system:

- **Proxy server** Is used to relay packets within a network between the UACs and the UAS. It also forwards requests generated by callers to their respective recipients. Proxy servers are also generally used for name mapping, which allows the proxy server to interlink an external SIP system to an internal SIP client.
- **Registrar server** Keeps a centralized record of the updated locations of all the users on the network. These addresses are stored on a location server.

- **Redirect server** Allows SIP devices to retain their SIP identities despite changes in their geographic location. This allows a device to remain accessible when its location is physically changed and hence while it moves through different networks. The use of redirect servers allows clients to remain within reach while they move through numerous network coverage zones. This configuration is generally known as an *intraorganizational* configuration. Intraorganizational routing enables SIP traffic to be routed within a VoIP network without being transmitted over the PSTN or external network.




## Streaming Protocols

The Real Time Protocol (RTP) is a session layer protocol that carries data in media stream format, as in audio and video, and is used extensively in VoIP, telephony, video conferencing, and other multimedia streaming technologies. It provides end-to-end delivery services and is commonly run over the transport layer protocol UDP. *RTP Control Protocol (RTCP)* is used in conjunction with RTP and is also considered a session layer protocol. It provides out-of-band statistics and control information to provide feedback on QoS levels of individual streaming multimedia sessions.

## IP Telephony Issues

VoIP's integration with the TCP/IP protocol has brought about some security challenges because it allows threat actors to leverage their TCP/IP experience to probe for flaws in both the architecture and the implementation of VoIP systems. Also involved are the traditional security issues associated with networks, such as unauthorized access, exploitation of communication protocols, and the spreading of malware. The promise of financial benefit derived from stolen call time is a strong incentive for most attackers. In short, the VoIP telephony network faces all the flaws that traditional computer networks have faced, plus the ones from legacy telephone systems too.

SIP-based signaling suffers from the lack of encrypted call channels and authentication of control signals. Attackers can tap into the SIP server and client communication to sniff out login IDs, passwords/PINs, and phone numbers. Once an attacker gets a hold of such information, she can use it to place unauthorized calls on the network. Toll fraud is considered to be the most significant threat that VoIP networks face, but illicit surveillance is also a threat for some organizations. If attackers are able to intercept voice packets, they may eavesdrop on ongoing conversations.

Attackers can also masquerade identities by redirecting SIP control packets from a caller to a forged destination to mislead the caller into communicating with an unintended end system. Like in any networked system, VoIP devices are also vulnerable to DoS attacks. Just as attackers would flood TCP servers with SYN packets on an IP network to exhaust a device's resources, attackers can flood RTP servers with call requests in order to overwhelm its processing capabilities. Attackers have also been known to connect laptops simulating IP phones to the Ethernet interfaces that IP phones use. These systems can then be used to carry out intrusions and DoS attacks. Attackers can also intercept RTP packets containing the media stream of a communication session to inject arbitrary audio/video data that may be a cause of annoyance to the actual participants.

Attackers can also impersonate a server and issue commands such as BYE, CHECKSYNC, and RESET to VoIP clients. The BYE command causes VoIP devices to close down while in a conversation, the CHECKSYNC command can be used to reboot VoIP terminals, and the RESET command causes the server to reset and reestablish the connection, which takes considerable time.

Combating VoIP security threats requires a well-thought-out infrastructure implementation plan. With the convergence of traditional and VoIP networks, balancing security while maintaining unconstrained traffic flow is crucial. VoIP calls can (and probably should) be encrypted over TLS. The use of authorization on the network is also an important step in limiting the possibilities of rogue and unauthorized entities on the network. Authorization of individual IP terminals ensures that only prelisted devices are allowed to access the network. Although not absolutely foolproof, this method can prevent rogue devices from connecting and flooding the network with illicit packets.

The use of secure cryptographic protocols such as TLS ensures that all SIP packets are conveyed within an encrypted and secure tunnel. The use of TLS can provide a secure channel for VoIP client/server communication and prevents the possibility of eavesdropping and packet manipulation.

### VoIP Security Measures Broken Down

Hackers can intercept incoming and outgoing calls, carry out DoS attacks, spoof phone calls, and eavesdrop on sensitive conversations. Many of the countermeasures to these types of attacks are the same ones used with traditional data-oriented networks:

- Keep patches updated on each network device involved with VoIP transmissions:
  - The call-processing manager server
  - The voicemail server
  - The gateway server
- Encrypt VoIP traffic whenever possible.
- Identify unidentified or rogue telephony devices:
  - Implement authentication so only authorized telephony devices are working on the network.
- Install and maintain
  - Stateful firewalls
  - VPN for sensitive voice data
  - Intrusion detection
- Disable unnecessary ports and services on routers, switches, PCs, and IP telephones.
- Employ real-time monitoring that looks for attacks, tunneling, and abusive call patterns through IDS/IPS:
  - Employ content monitoring.
  - Use encryption when data (voice, fax, video) crosses an untrusted network.
  - Use a two-factor authentication technology.
  - Limit the number of calls via media gateways.
  - Close the media sessions after completion.



# Multimedia Collaboration

The term *multimedia collaboration* is very broad and includes remotely sharing any combination of voice, video, messages, telemetry, and files during an interactive session. The term encompasses conferencing applications like Zoom, WebEx, and Google Meetings but also many other applications in disciplines such as project management, e-learning, science, telemedicine, and military. What distinguishes multimedia collaboration applications

is their need to simultaneously share a variety of data formats, each of which has different loss, latency, jitter, and bandwidth requirements. Of course, as we work to meet these performance requirements and allow maximum participation from authorized users (potentially around the world), we also have to ensure the security of this communication channel.

## Meeting Applications

Imagine this scenario: You are hosting an online leadership meeting with your international partners to discuss the year ahead. Suddenly, a participant with a name you don't recognize starts sharing pornographic images and hate speech for all to see. You've just been "Zoom-bombed." (A term that doesn't necessarily mean you were using that particular platform.) This is what happens when access controls to your online meeting are inadequate. Many naïve users of meeting applications simply share a link with their guests, usually via e-mail or some other messaging application. Anyone with that link could then join the call if other precautions aren't taken.

The rise in popularity of meeting applications and their increased importance to the business of our organizations have put them in the crosshairs of a wide range of attackers beyond the Zoom-bombing troll we described. To prevent these attacks, consider the following best practices for securing online meeting applications:

- *Don't use consumer-grade products.* There is much wisdom in the old adage "you get what you pay for." Consumer-grade products are much cheaper than enterprise-grade ones (or even free), but they lack most security controls that we need to secure our organizational meetings.
- *Use AES 256-bit encryption.* It is rare to be able to support true end-to-end encryption for online meetings because most service providers need access to the traffic for things like recording, closed captioning, and echo cancelation. Still, you should ensure all call traffic is encrypted between each participant and the service provider.
- *Control access to every meeting.* Enterprise-grade conferencing services can integrate with your identity and access management service to ensure strong authentication. Failing that, ensure that, at a minimum, each meeting is password-protected.
- *Enable the waiting room feature, particularly for external participants.* Many services place participants in a virtual waiting room when they sign in to the meeting until the host lets them in. This gives you an opportunity to screen each participant prior to allowing them to join. At a minimum, ensure participants cannot connect to the call before the host does.
- *Restrict participants' sharing of their screens or cameras as appropriate.* This is particularly important when the meeting involves external parties such as partners or clients. While cameras may be desirable for a variety of reasons, it is rare for all participants to need unfettered screen sharing. Either way, ensure this is a deliberate decision by the host or organizer and enforceable by the platform.

### Telepresence

Sometimes, you and other meeting participants need to do more than just see and hear each other and share slides remotely. *Telepresence* is the application of various technologies to allow people to be virtually present somewhere other than where they physically are. Consider a bomb disposal specialist trying to disarm an explosive device remotely using a robot, or a surgeon performing a delicate operation on a patient who would otherwise be inaccessible. The possibilities are endless and include the far more mundane applications that most of our organizations would consider, such as trade shows, pipeline inspections, and virtual reality (VR) training.

Because telepresence systems are not yet prevalent, there is no consensus yet on how to best secure them as a whole. Still, the secure design principles we've covered in this book (to which we'll return later in this chapter) apply to these systems.

- *Keep your software updated.* Online meeting software is no different than any other in the need for patch and update management. Even if you don't use dedicated clients and use web browsers to connect, you should ensure whatever you use is up to date.
- *Don't record meetings unless necessary.* It is helpful to record meetings, particularly when some participants cannot join in real time and must watch it later. However, the recordings can contain sensitive data that could be stolen or lead to other types of liability. If you do record the meeting, ensure it is for good reasons and that the recorded data is encrypted.
- *Know how to eject unwanted participants.* If you do get Zoom-bombed, that is not the time to figure out how to eject (and lock out) an offending participant. Ensure all hosts know how to do this beforehand and, while they're at it, learn also how to mute their microphones (and cameras) if needed.



## Unified Communications

While meeting applications like videoconferencing systems have received a lot of attention recently, there is a broader application of multimedia collaboration services known as *unified communications (UC)*. UC is the integration of real-time and non-real-time communications technologies in one platform. Real-time communications are those that are instantaneous and interactive, such as telephone and video conferencing. Non-real-time communications, on the other hand, don't require our immediate attention and are exemplified by technologies such as e-mail and text messaging. The whole point of UC is that it integrates multiple modes of communication, as shown in Figure 15-4.

One of the key features of UC is the concept of *presence information*, which is an indicator of a subject's availability and willingness to communicate. If you have ever used a platform like Slack or Microsoft Teams, you will have noticed the presence icon next to your teammates. It may show that they are available, sleeping, on a call, or on a meeting. Presence information allows you to choose how to interact with your colleagues. If you


**Figure 15-4** Unified communications components

need to get a message to Mohammed, who happens to be in a meeting, you can send him a text message. If, on the other hand, you see that Carmen is available, you may want to reach out to her on a voice or video call. Presence information can also show where in the world your colleagues are. For example, if you want to meet Bob and notice that he happens to be in the same city as you are, you may opt for a face-to-face meeting request.

Securing UC involves similar security controls that we would apply to any other communications platform, but with a couple of important caveats. For starters, UC relies on centralized data and access controls. This means that, whether your organization hosts its services on premises or in the cloud, there is a hub that supports and enables them. You want to ensure that this hub is adequately protected against physical and logical threats. Obviously, you want to protect your data, whether at rest or in motion, with strong encryption, but this will only get you so far if you allow anyone to access it. Consequently, you want to apply strict access controls that still allow the business processes to run efficiently. Finally, you want to ensure that demand spikes don't cause self-inflicted denial-of-service conditions. Instead, ensure that you have enough spare capacity to handle these inevitable (if rare) spikes.

## Remote Access

Remote access covers several technologies that enable remote and home users to connect to resources that they need to perform their tasks. Most of the time, these users must first gain access to the Internet through an ISP, which sets up a connection to the destination network. For many organizations, remote access is a necessity because it enables users to access centralized network resources; it reduces networking costs by using the Internet as the access medium instead of expensive dedicated lines; and it extends the workplace for employees to their home computers, laptops, and mobile devices. Remote access can streamline access to resources and information through Internet connections and provides a competitive advantage by letting partners, suppliers, and customers have closely controlled links.

## VPN

We discussed VPNs in Chapter 13 as a general concept, but let's circle back and see how to best employ them to provide secure remote connectivity for our staff members. VPNs are typically implemented using a client application that connects to a VPN server (commonly called a concentrator) in our organization. In a perfect world, you would have enough bandwidth and concentrator capacity to ensure all your remote staff members can simultaneously connect over the VPN. Then, you could enforce *always-on VPN*, which is a system configuration that automatically connects the device to the VPN with no user interaction. Obviously, this would only be possible with devices owned by the organization, but it can provide strong access controls if properly implemented. For even better results, you can implement a *VPN kill switch*, which automatically cuts off Internet access unless a VPN session is established.

Alas, things are usually a bit more complicated. Perhaps you don't have enough VPN capacity for your entire workforce, or you allow use of personal devices. If you cannot implement always-on VPN, the next best thing is to ensure you use multifactor authentication (MFA) and network access control (NAC). NAC is particularly important because you want to be able to check that the user device is safe before allowing it to access your corporate network. Since not everyone will be connecting to the VPN, you want to ensure that remote users have access to the resources they need and no others, possibly by putting them on the right VLANs and ensuring you have the right access control lists (ACLs) in your internal routers.

Regardless, you want to ensure your VPN systems (clients and concentrators) are updated and properly configured. Many clients allow you to select the cryptosystem to use, in which case you want to select the strongest option you can. Finally, carefully consider whether you will allow split tunnels.

A *VPN split tunnel* is a configuration that routes certain traffic (e.g., to the corporate data center) through the VPN while allowing other traffic (such as web searches) to access the Internet directly (without going through the VPN tunnel). The advantage of this approach is that users will be less likely to experience latency induced by an overworked concentrator. It also allows them to print to their local printer at home while on VPN. The disadvantage is that, should they pick up malware or otherwise become compromised on the Internet, the adversary will automatically get a free ride into your corporate network through the VPN. To prevent this from happening, you can enforce a *VPN full tunnel*, which routes all traffic through the concentrators.

### VPN Authentication Protocols

While we're talking about VPN configuration, let's go over some of the authentication protocols you may come across, so you know what each brings to the table.

**PAP** The *Password Authentication Protocol (PAP)* is used by remote users to authenticate over Point-to-Point Protocol (PPP) connections such as those used in some VPNs. PAP requires a user to enter a password before being authenticated. The password and the username credentials are sent over the network to the authentication server after a connection has been established via PPP. The authentication server has a database of user credentials that are compared to the supplied credentials to authenticate users. PAP is one

of the least secure authentication methods because the credentials are sent in cleartext, which renders them easy to capture by network sniffers. PAP is also vulnerable to man-in-the-middle attacks. Although this protocol is not recommended for use anywhere, some (improperly configured) systems can revert to PAP if they cannot agree on any other authentication protocol.

**EXAM TIP** PAP has been considered insecure for decades. If you see it on the exam, consider it a bad choice.

**CHAP** The *Challenge Handshake Authentication Protocol (CHAP)* addresses some of the vulnerabilities found in PAP. It uses a challenge/response mechanism to authenticate the user instead of having the user send a password over the wire. When a user wants to establish a PPP connection and both ends have agreed that CHAP will be used for authentication purposes, the user's computer sends the authentication server a logon request. The server sends the user a challenge (called a nonce), which is a random value. This challenge is encrypted with the use of a predefined password as an encryption key, and the encrypted challenge value is returned to the server. The authentication server also uses the predefined password as an encryption key and decrypts the challenge value, comparing it to the original value sent. If the two results are the same, the authentication server deduces that the user must have entered the correct password and grants authentication. The steps that take place in CHAP are depicted in Figure 15-5. Unlike PAP, CHAP is not vulnerable to man-in-the-middle attacks because it continues this challenge/response activity throughout the connection to ensure the authentication server is still communicating with a user who holds the necessary credentials.

**EXAM TIP** MS-CHAP is Microsoft's version of CHAP and provides mutual authentication functionality. It has two versions, which are incompatible with each other.

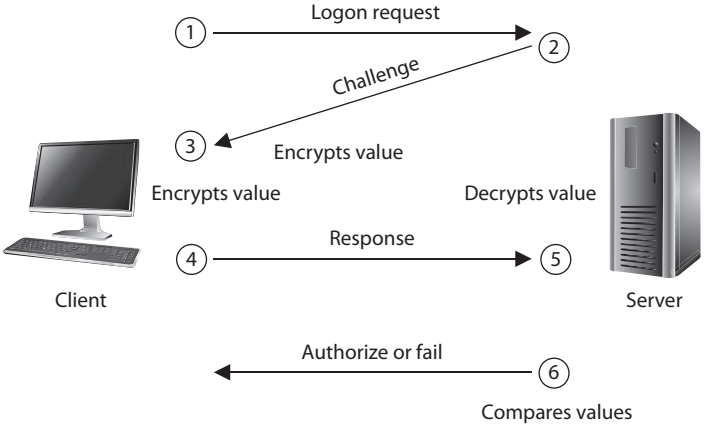



**Figure 15-5** CHAP uses a challenge/response mechanism instead of having the user send the password over the wire.

**EAP** The *Extensible Authentication Protocol (EAP)* is also supported by PPP. Actually, EAP is not a specific authentication protocol as are PAP and CHAP. Instead, it provides a framework to enable many types of authentication techniques to be used when establishing network connections. As the name states, it *extends* the authentication possibilities from the norm (PAP and CHAP) to other methods, such as one-time passwords, token cards, biometrics, Kerberos, digital certificates, and future mechanisms. So when a user connects to an authentication server and both have EAP capabilities, they can negotiate between a longer list of possible authentication methods.

**NOTE** EAP has been defined for use with a variety of technologies and protocols, including PPP, Point-to-Point Tunneling Protocol (PPTP), Layer 2 Tunneling Protocol (L2TP), IEEE 802 wired networks, and wireless technologies such as 802.11 and 802.16.

There are many different variants of EAP, as shown in Table 15-1, because EAP is an extensible framework that can be morphed for different environments and needs.

## Desktop Virtualization

*Desktop virtualization* technologies allow users to remotely interact with computers as if they were physically using them. In essence, these technologies present a virtual copy of a desktop that is running on some computer (physical or virtual) somewhere else



| Protocol | Description |
|---|---|
| EAP-TLS | Digital certificate–based authentication, considered one of the most secure EAP standards |
| EAP-PSK | Provides mutual authentication and session key derivation using a preshared key |
| EAP-TTLS | Tunneled TLS, which requires the server to have a CA-issued certificate, but makes this optional for the client |
| EAP-IKE2 | Internet Key Exchange version 2 (IKE2), which provides mutual authentication and session key establishment using asymmetric or symmetric keys or passwords |
| PEAPv0/EAP-MSCHAPv2 | Similar in design to EAP-TTLS but only requires a server-side digital certificate |
| PEAPv1/EAP-GTC | Cisco variant based on Generic Token Card (GTC) authentication |
| EAP-FAST | Cisco-proprietary replacement for Lightweight EAP (LEAP) based on Flexible Authentication via Secure Tunneling (FAST) |
| EAP-SIM | For Global System for Mobile Communications (GSM), based on Subscriber Identity Module (SIM), a variant of PEAP for GSM |
| EAP-AKA | For Universal Mobile Telecommunication System (UMTS) Subscriber Identity Module (USIM) and provides Authentication and Key Agreement (AKA) |
| EAP-GSS | Based on Generic Security Service (GSS), uses Kerberos |

**Table 15-1** EAP Variants

in the network. IT staff frequently use desktop virtualization to manage rack-mounted servers (without having to attach a monitor, keyboard, and mouse to each), to log into jump boxes, and to manage and troubleshoot user workstations. In some organizations, remote desktop solutions allow staff to work from home and, through their personal devices, securely use an organizational computer. The upside of desktop virtualization is that the asset is protected by the organization's security architecture but still is accessible from almost anywhere. There are two main approaches to desktop virtualization: remote desktops and virtual desktop infrastructure.

**NOTE** A *jump box* (also called a *jump host* or *jump server*) is a hardened host that acts as a secure entry point or gateway into a sensitive part of a network.

## Remote Desktops

Two of the most common approaches to providing remote desktops are Microsoft's *Remote Desktop Protocol (RDP)* and the open-source *Virtual Network Computing (VNC)* system. At a high level, both are very similar. They both require that a special server is running on the computer that will be controlled remotely and that the remote device has a software client installed and connected to the server, by default over port 3389 for RDP and 5900 for VNC. Although there are clients and servers for every major operating system, RDP is more common in Windows environments and VNC is more common in Linux environments.

The most important security consideration when deploying either RDP or VNC is to ensure that the connections are encrypted. Neither of these systems has robust security controls, so you have to tunnel them over a secure channel. If you are providing this service to remote users outside your organizational network, then you should ensure they are connected to the VPN. Having external RDP or VNC servers is a recipe for a security disaster, so their corresponding ports should be blocked at your firewall.

One of the advantages or disadvantages (depending on how you look at it) of RDP and VNC is that they allow a client to remotely control a specific computer. That computer must be provisioned somewhere on the network, specifically configured to allow remote access, and then must remain available. If it is powered off or is otherwise unavailable, there is nothing to remotely control.

## Virtual Desktop Infrastructure

By combining virtualization and remote desktop technologies, we can create an environment in which users access the desktops of virtual machines (VMs) that look and behave exactly as the users have configured them, but that can be spun up or down, migrated, wiped, and re-created centrally as needed. *Virtual desktop infrastructure (VDI)* is a technology that hosts multiple virtual desktops in a centralized manner and makes them available to authorized users. Each virtual desktop can be directly tied to a VM (very similarly to the remote desktops described in the previous section) or can be a composite of multiple virtual components, such as a desktop template combined with virtual

applications running on multiple different VMs. This flexibility allows organizations to tailor desktops to specific departments, roles, or even individuals in a scalable and resource-effective manner.

VDI deployments can be either persistent or nonpersistent. In a *persistent VDI*, a given user connects to the same virtual desktop every time and is able to customize it as allowed by whatever organizational policies are in place. In a persistent model, users' desktops look the same at the beginning of one session as they did at the end of the last one, creating continuity that is helpful for long-term use and for complex workflows. By contrast, users of a *nonpersistent VDI* are presented with a standard desktop that is wiped at the end of each session. Nonpersistent infrastructures are useful when providing occasional access for very specific purposes or in extremely secure environments.

VDI is particularly helpful in regulated environments because of the ease with which it supports data retention, configuration management, and incident response. If a user's system is compromised, it can quickly be isolated for remediation or investigation, while a clean desktop is almost instantly spawned and presented to the user, reducing the downtime to seconds. VDI is also attractive when the workforce is highly mobile and may log in from a multitude of physical devices in different locations. Obviously, this approach is highly dependent on network connectivity. For this reason, organizations need to consider carefully their own network speed and latency when deciding how (or whether) to implement it.



## Secure Shell

We don't always need a graphical user interface (GUI) to interact with our devices. In fact, there are many advanced use cases in which users, especially experienced and administrative ones, are more productive using a command-line interface (CLI). The tool of choice in many of these cases (particularly in Linux environments) is *Secure Shell (SSH)*, which functions as a type of tunneling mechanism that provides terminal-like access to remote computers. SSH is the equivalent of remote desktops but without the GUI. For example, the program can let Paul, who is on computer A, access computer B's files, run applications on computer B, and retrieve files from computer B without ever physically touching that computer. SSH provides authentication and secure transmission over vulnerable channels like the Internet.

**NOTE** SSH can also be used for secure channels for file transfer and port redirection.

SSH should be used instead of Telnet, FTP, rlogin, rexec, or rsh, which provide the same type of functionality SSH offers but in a much less secure manner. SSH is a program and a set of protocols that work together to provide a secure tunnel between two computers. The two computers go through a handshaking process and exchange (via Diffie-Hellman) a session key that will be used during the session to encrypt and protect the data sent. The steps of an SSH connection are outlined in Figure 15-6.

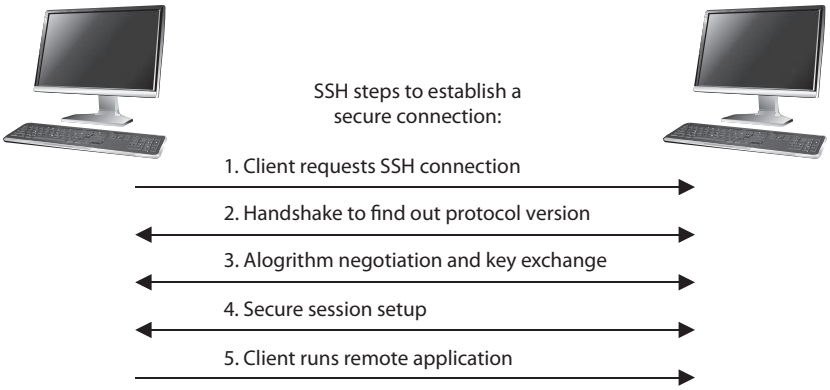
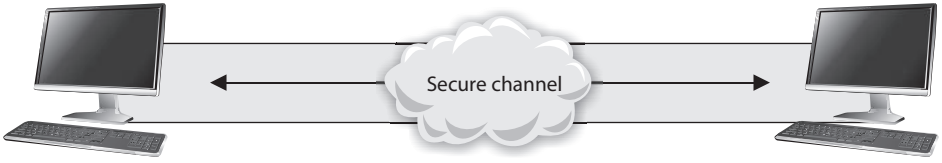



**Figure 15-6** SSH is used for remote terminal-like functionality.

**EXAM TIP** Telnet is similar in overall purpose to SSH but provides none of the latter's security features. It is insecure and probably not the right answer to any question.

Once the handshake takes place and a secure channel is established, the two computers have a pathway to exchange data with the assurance that the information will be encrypted and its integrity will be protected.




# Data Communications

Up to this point in this chapter, we've been focused on communications channels used by users. It is probably a good idea to also consider machine to machine data communications. Recall from Chapter 7 that there are multiple system architectures that require quite a bit of backend chatter between system components. For example, in an n-tier architecture, you may have an application server communicating quite regularly with a database. We must also map out and secure all these not-so-obvious data communications channels.

## Network Sockets

A *network socket* is an endpoint for a data communications channel. A socket is a layer 4 (transport) construct that is defined by five parameters: source address, source port, destination address, destination port, and protocol (TCP or UDP). At any given time, a typical workstation has dozens of open sockets, each representing an existing data communications channel. (Servers can have thousands or even tens of thousands of them.) Each of these channels represents an opportunity for an attacker to compromise our systems. Do you know what all your data channels are?

This is one of the reasons why understanding our systems architectures is so critical. Many systems use default installation configurations that are inherently insecure. In addition to the proverbial (weak) default password, a brand-new server probably includes a number of services that are not needed and could provide an open door to attackers. Here are some best practices for securing sockets-based communications channels:

- Map out every authorized data communications channel to and from each server.
- Apply ACLs to block every connection except authorized ones.
- Use segmentation to ensure servers that communicate with each other regularly are in the same network segment.
- Whenever possible, encrypt all data communications channels.
- Authenticate all connection requests.

One of the challenges of securing data communications channels is that they rely on service accounts that usually run with elevated privileges. Oftentimes, these service accounts are excluded from the password policies that are enforced for user accounts. As a result, service account passwords are seldom changed and sometimes are documented in an unsecure manner. For example, we know of organizations that keep a list of their service accounts and passwords on a SharePoint or Confluence page for their IT team. These passwords should be protected just like any other privileged account and securely stored in a password vault.

## Remote Procedure Calls

Moving up one level to the session layer (layer 5), a *remote procedure call (RPC)* allows a program somewhere in your network to execute a function or procedure on some other host. RPC is commonly used in distributed systems because it allows systems to divide larger tasks into subtasks and then hand those subtasks to other systems. Although the IETF defined an RPC protocol for Open Network Computing (ONC), the RPC concept can take many different forms in practice. In most networks (especially Windows ones), RPC services listen on TCP port 135. RPC use is ubiquitous in many enterprise environments because it is so powerful. However, by default, it doesn't provide any security beyond basic authentication.

If your organization uses RPC, then you should really consider upgrading its security. Secure RPC (S-RPC) provides authentication of both users and hosts as well as traffic encryption. As of February 9, 2021, Windows Active Directory (AD) systems require S-RPC. The IETF also released a standard for RPC security (RPCSEC) years ago, but because it is difficult to implement, it was never widely adopted. Instead, many organizations require TLS for authenticating hosts and encrypting RPC traffic. Other, vendor-specific implementations of RPC security exist, so you should research whatever versions are being used in your environment and ensure they are secure.

# Virtualized Networks

A lot of the network functionality we have covered in this chapter can take place in virtual environments. You should remember from our coverage of virtual machines (VMs) in Chapter 7 that a host system can have virtual guest systems running on it, enabling multiple operating systems to run on the same hardware platform simultaneously. But the industry has advanced much further than this when it comes to virtualized technology. Routers and switches can be virtualized, which means you do not actually purchase a piece of hardware and plug it into your network, but instead you deploy software products that carry out routing and switching functionality. Obviously, you still need a robust hardware infrastructure on which to run the VMs, but virtualization can save you a lot of money, power, heat, and physical space.

These VMs, whether they implement endpoints or networking equipment, communicate with each other over virtual networks that behave much like their real counterparts, with a few exceptions. In order to understand some of these, let us first consider the simple virtual infrastructure shown in Figure 15-7. Let's suppose that VM-1 is an endpoint (perhaps a server), VM-2 is a firewall, and VM-3 is an IDS on the external side of the firewall. Two of these devices (VM-1 and VM-3) have a single virtual NIC (vNIC), while the other one (VM-2) has two vNICs. Every vNIC is connected to a virtual port on a virtual switch. Unlike the real world, any data that flows from one vNIC

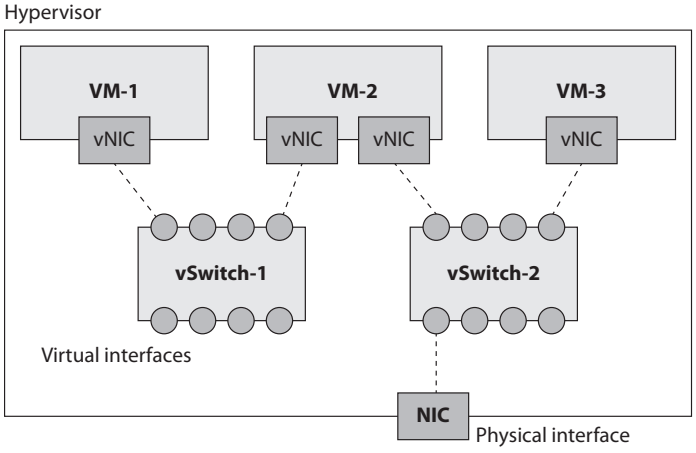



**Figure 15-7** Virtualized networks

to another vNIC is usually just copied from one memory location (on the physical host) to another; it only pretends to travel the virtual network.

The single physical NIC in our example is connected to vSwitch-2, but it could just as easily have been directly connected to a vNIC on a VM. In this virtual network, VM-2 and VM-3 have connectivity to the physical network but VM-1 does not. The hypervisor stores in memory any data arriving at the physical NIC, asks the virtual switch where to send it, and then copies it into the memory location for the intended vNIC. This means that the hypervisor has complete visibility over all the data traversing its virtualized networks, whether or not it touches the physical NIC.

It should come as no surprise that one of the greatest strengths of virtualization, the hypervisor, is potentially also its greatest weakness. Any attacker who compromises the hypervisor could gain access to all virtualized devices and networks within it. So, both the good and the bad guys are intensely focused on finding any vulnerabilities in these environments. What should you do to ensure the security of your virtualized networks and devices? First, just as you should do for any other software, ensure you stay on top of any security patches that come out. Second, beware of third-party add-ons that extend the functionality of your hypervisor or virtual infrastructure. Ensure these are well tested and acquired from reputable vendors. Last, ensure that whoever provisions and maintains your virtualized infrastructure is competent and diligent, but also check their work. Many vulnerabilities are the result of misconfigured systems, and hypervisors are no different.



## Third-Party Connectivity

We can't wrap up our discussion of securing the multitude of communications channels in our systems without talking about third parties. In Chapter 2, we covered the risks that third parties bring to our organizations and how to mitigate them. These third parties cover a broad spectrum that includes suppliers, service providers, and partners. Each of them may have legitimate needs to communicate digitally with our organizations, potentially in an automated manner. How can we provide this required connectivity to third parties without sacrificing our security? The answer can be found by applying the secure design principles we've been revisiting throughout the book:

- **Threat modeling** Always start by identifying the threats. What might malicious (or just careless) third parties be able to do with the communications channels we provide that would cause us harm? What are their likeliest and most dangerous actions? This deliberate exercise in understanding the threats is foundational.
- **Least privilege** Third parties will have legitimate connectivity requirements that we should minimally provide. If a contractor needs to monitor and control our HVAC systems remotely, we should segment those systems on the same VLAN and ensure that only specific calls from specific hosts to specific devices are allowed, and nothing more.

- **Defense in depth** Based on the threat model, we put in place controls to mitigate risks. But what happens if the first layer of controls fails to contain the threat? If that HVAC contractor is compromised in an island-hopping attack and the adversary is able to escape the VLAN, how do we detect the breach and then contain the attack?
- **Secure defaults** While ensuring that default configurations are secure is generally a best practice, it is particularly important on systems that will be used by third parties. One of the keys here is to enforce strict configuration management. For any system that will be accessible by a third party, we must ensure that all defaults are secure by testing them.
- **Fail securely** Speaking of testing, we should test the system under a range of conditions to see what happens when it breaks. For example, stress testing (under heavy usage loads), fuzzing, and power and network failure testing can show us what happens when a system fails. This is not specific to third-party systems, by the way.
- **Separation of duties** Giving third parties the least privileges needed actually makes separating duties easier. For example, it may be that the HVAC contractor does not normally start or stop the furnace, but this may be occasionally required. Because this can have an impact on our facility, the action must be approved by our site manager.
- **Keep it simple** This principle is centered on the statement of work (SoW) that describes the agreement with the third party and in the processes we build to support that work. A policy of "deny by default, allow by exception" can keep things simple, supports the least-privilege principle, and should be paired with a simple process for handling exceptions.
- **Zero trust** It goes without saying that we should not trust third parties when it comes to access to our systems. For every interaction of third parties with our systems, we must ensure that authentication, nonrepudiation, and audit controls are sufficient to detect and mitigate any threat (deliberate or otherwise) that they introduce into our environments.
- **Privacy by design** If we use this principle to guide the development of our entire security architecture (and we really ought to), then we really shouldn't have to do anything else to account for third parties using our systems, particularly if we couple privacy with least privilege in the first place.
- **Trust but verify** We already talked about auditability in the context of zero trust, but there is a difference between logging activities and analyzing those logs periodically (or even continually). What is the process by which our security staff verifies that the actions of third parties are appropriate? How are suspicious or malicious activities handled?
- **Shared responsibility** Finally, who is contractually responsible for what? As the saying goes, "good fences make good neighbors." It is important to define responsibilities in the service or partnership agreement so that there are no misunderstandings and, should someone fail, we can take financial or legal actions to recover our losses.

# Chapter Review

With this chapter, we have finished our coverage of the fourth domain of the CISSP Common Body of Knowledge, Communication and Network Security, by discussing the myriad of technologies that allow us to create secure communications channels in our organizations. Though most people (particularly in the technology fields) would not consider voice to be their primary means of communication, it remains important for many reasons, not the least of which is the fact that traditional voice channels are more commonly used nowadays for digital data traffic. It is important to understand how these technologies blend in different ways so that we can better secure them.

The COVID-19 pandemic forced most organizations around the world to quickly move toward (or improve their ability at) supporting a remote workforce largely based in home offices. While the news media regularly featured stories on the vulnerabilities and attacks on our multimedia collaboration and remote access systems, it is remarkable how well these held up to the sudden increase in use (and attacks). We hope that this chapter has given you a better understanding of how security professionals can continue to improve the security of these systems while supporting a remote workforce and third-party connectivity.



## Quick Review

- The public switched telephone network (PSTN) uses circuit switching instead of packet routing to connect calls.
- The Signaling System 7 (SS7) protocol is used for establishing and terminating calls in the PSTN.
- The main components of a PSTN network are signal switching points (SSPs) that terminate subscriber loops, signal transfer points (STPs) that interconnect SSPs and other STPs to route calls through the network, and service control points (SCPs) that control advanced features.
- A digital subscriber line (DSL) is a high-speed communications technology that simultaneously transmits analog voice and digital data between a home or business and a PSTN service provider's central office.
- Asymmetric DSL (ADSL) has data rates of up to 24 Mbps downstream and 1.4 Mbps upstream but can only support distances of about a mile from the central office without signal boosters.
- Very high-data-rate DSL (VDSL) is a higher-speed version of ADSL (up to 300 Mbps downstream and 100 Mbps upstream).
- G.fast is DSL that runs over fiber-optic cable from the central office to a distribution point near the home and then uses legacy copper wires for the last few hundred feet to the home or office. It can deliver data rates of up to 1 Gbps.
- Integrated Services Digital Network (ISDN) is an obsolescent pure digital technology that uses legacy phone lines for both voice and data.

- Basic Rate Interface (BRI) ISDN is intended to support a single user with two channels each with data throughput of 64 Kbps.
- Primary Rate Interface (PRI) ISDN has up to 23 usable channels, at 64 Kbps each, which is equivalent to a T1 leased line.
- Cable modems provide high-speed access to the Internet through existing cable coaxial and fiber lines, but the shared nature of these media result in inconsistent throughputs.
- Internet Protocol (IP) telephony is an umbrella term that describes carrying telephone traffic over IP networks.
- The terms "IP telephony" and "Voice over IP" are used interchangeably.
- Jitter is the irregularity in the arrival times of consecutive packets, which is problematic for interactive voice and video communications.
- The H.323 recommendation is a standard that deals with audio and video calls over packet-based networks.
- The Session Initiation Protocol (SIP) is an application layer protocol used for call setup and teardown in IP telephony, video and multimedia conferencing, instant messaging, and online gaming.
- The Real-time Transport Protocol (RTP) is a session layer protocol that carries data in media stream format, as in audio and video, and is used extensively in VoIP, telephony, video conferencing, and other multimedia streaming technologies.
- RTP Control Protocol (RTCP) is used in conjunction with RTP and is also considered a session layer protocol. It provides out-of-band statistics and control information to provide feedback on QoS levels of individual streaming multimedia sessions.
- Multimedia collaboration is a broad term that includes remotely and simultaneously sharing any combination of voice, video, messages, telemetry, and files in an interactive session.
- Telepresence is the application of various technologies to allow people to be virtually present somewhere other than where they physically are.
- Unified communications (UC) is the integration of real-time and non-real-time communications technologies in one platform.
- An always-on VPN is a system configuration that automatically connects the device to the VPN with no user interaction.
- A VPN kill switch is a system configuration that automatically cuts off Internet access unless a VPN session is established.
- A VPN split tunnel is a configuration that routes certain traffic through the VPN while allowing other traffic to access the Internet directly.

- The Password Authentication Protocol (PAP) is an obsolete and insecure authentication protocol that sends user credentials in plaintext and should not be allowed.
- The Challenge Handshake Authentication Protocol (CHAP) uses a challenge/response mechanism using the password as an encryption key to authenticate the user instead of having the user send a password over the wire.
- The Extensible Authentication Protocol (EAP) is a framework that enables many types of authentication techniques to be used when establishing network connections.
- Desktop virtualization technologies, such as remote desktops and virtual desktops, allow users to remotely interact with computers as if they were physically using them.
- Two of the most common approaches to providing remote desktops are Microsoft's Remote Desktop Protocol (RDP) and the open-source Virtual Network Computing (VNC) system.
- Virtual desktop infrastructure (VDI) is a technology that hosts multiple virtual desktops in a centralized manner and makes them available to authorized users.
- Secure Shell (SSH) is a secure tunneling mechanism that provides terminal-like access to remote computers.
- A network socket is an endpoint for a data communications channel, defined by five parameters: source address, source port, destination address, destination port, and protocol (TCP or UDP).
- Remote procedure calls allow a program somewhere in your network to execute a function or procedure on some other host.



## Questions

Please remember that these questions are formatted and asked in a certain way for a reason. Keep in mind that the CISSP exam is asking questions at a conceptual level. Questions may not always have the perfect answer, and the candidate is advised against always looking for the perfect answer. Instead, the candidate should look for the best answer in the list.

**1.** In which type of networks is the Signaling System 7 (SS7) protocol used?

**A.** Integrated Services Digital Network (ISDN)

**B.** IP telephony network

**C.** Real-time Transport Protocol (RTP) network

**D.** Public switched telephone network (PSTN)

2. Which of the following is true about the Session Initiation Protocol (SIP)?
   - **A.** Used to establish virtual private network (VPN) sessions
   - **B.** Framework for authenticating network connections
   - **C.** Session layer protocol for out-of-band statistics
   - **D.** Application layer protocol used in online gaming communications
3. Which of the following is not considered a best practice for securing multimedia collaboration platforms?
   - **A.** Don't record meetings unless necessary
   - **B.** Use consumer-grade products
   - **C.** Use AES 256-bit encryption
   - **D.** Restrict participants' sharing of their screens or cameras as appropriate
4. How could you best protect a unified communications (UC) platform?
   - **A.** Protect it as you would any other systems
   - **B.** Enable Password Authentication Protocol (PAP)
   - **C.** Use the Session Initiation Protocol (SIP) for every new session
   - **D.** Ensure the hub is protected against physical and logical threats

*Use the following scenario to answer Questions 5–7.* You are the CISO of a research and development company that is transitioning to a 100 percent remote workforce, so your entire staff will be working from home. You don't have enough laptops for all your staff, so those without one will be using their personal computers and printers for work. Your VPN concentrators are sufficient to support the entire workforce, and you will be requiring all staff members to connect to the VPN.

5. Which authentication protocol would be best for your VPN connections?
   - **A.** Password Authentication Protocol (PAP)
   - **B.** Challenge Handshake Authentication Protocol (CHAP)
   - **C.** Extensible Authentication Protocol (EAP)
   - **D.** Session Initiation Protocol (SIP)
6. Which of the following additional VPN configurations should you also enable?
   - **A.** Split tunneling
   - **B.** Full tunneling
   - **C.** VPN kill switch
   - **D.** Hybrid tunneling

7. Which of the following will best protect the confidentiality of your sensitive research data?
   - **A.** Secure Shell (SSH)
   - **B.** Virtualized networks
   - **C.** Virtual desktop infrastructure (VDI)
   - **D.** Remote Procedure Calls (RPC)
8. During a recent review of your enterprise architecture, you realize that many of your mission-critical systems rely on Remote Procedure Call (RPC). What measures should you take to ensure remote procedure calls are secured?
   - **A.** Implement ITU standard H.323
   - **B.** Tunnel RPC through Transport Layer Security (TLS)
   - **C.** Use the Password Authentication Protocol (PAP) for authentication
   - **D.** Enforce client-side authentication
9. Which of the following is not an advantage of virtual desktops?
   - **A.** Reduced user downtime during incident response
   - **B.** Support for both persistent and nonpersistent sessions
   - **C.** Support for both physical and remote logins
   - **D.** Better implementation of data retention standards



## Answers

1. **D.** The SS7 protocol is used in a PSTN to set up, control, and disconnect calls.
2. **D.** SIP is an application layer protocol used for call setup and teardown in IP telephony, video and multimedia conferencing, instant messaging, and online gaming.
3. **B.** Consumer-grade products almost always lack the security controls and management features that we need to properly secure multimedia collaboration platforms.
4. **D.** Securing UC involves similar security controls that we would apply to any other communications platform, but with a couple of important caveats. Unified communications rely on a central hub that integrates, coordinates, and synchronizes the various technologies. You want to ensure that this hub is adequately protected against physical and logical threats.
5. **C.** EAP is considered much more secure than both PAP (which is not secure at all) and CHAP. SIP does not provide authentication mechanisms at all.
6. **A.** Because your staff will be using printers on their home networks, you will have to enable split tunneling, which allows some traffic to be sent over the VPN and other traffic to go to the local network or to the Internet directly.

7. **C.** VDI allows your sensitive data to remain in your protected network even as users are able to work with it over a virtual desktop. Properly configured, this infrastructure prevents any sensitive research data from being stored on the remote user's computer.
8. **B.** Since many implementations of RPC lack security controls, many organizations require TLS for authenticating hosts and encrypting RPC traffic.
9. **C.** VDI is particularly helpful in regulated environments because of the ease with which it supports data retention, configuration management, and incident response through persistent and nonpersistent sessions. However, since VDI relies on VMs in a data center, there is not a computer at which a user could physically log in.

# PART V

# Identity and Access Management

*This page intentionally left blank*

CHAPTER 16

# Identity and Access Fundamentals

This chapter presents the following:

- Identification, authentication, authorization, and accountability
- Credential management
- Identity management
- Federated identity management with a third-party service

*The value of identity of course is that so often with it comes purpose.*

—Richard Grant

The concept of identity is foundational to controlling access to our assets because everyone (and everything) that touches them must have a legitimate purpose in doing so. What makes access control tricky is that most of us have multiple identities that depend on the context in which we find ourselves. A person could simultaneously be an asset owner, custodian, and processor (roles we discussed in Chapter 5), depending on which asset we consider and at what time. On top of the challenge of handling multiple identities, we also have to ensure that each identity belongs to the person claiming it.

In this chapter, we discuss the fundamentals of user identification, authentication, and authorization. We do this while considering a variety of real-world contexts, such as complex enterprise environments and the interaction with third parties. Of course, we must be able to verify that things are being done correctly, so we also talk about accountability in these efforts. This all sets the stage for the next chapter, in which we delve into how to actually manage identities and access.

## Identification, Authentication, Authorization, and Accountability

For users to be permitted to access any resource, they first must prove they are who they claim to be, have the necessary credentials, and have been given the necessary rights or privileges to perform the actions they are requesting. Once these steps are completed successfully, it is necessary to track users' activities and enforce accountability for their

actions. *Identification* describes a method by which a subject (user, program, or process) claims to have a specific identity (username, account number, or e-mail address). *Authentication* is the process by which a system verifies the identity of the subject, usually by requiring a piece of information that only the claimed identity should have. This piece could be a password, passphrase, cryptographic key, personal identification number (PIN), physiological characteristic, or token. Together, the identification and authentication information (for example, username and password) make up the subject's *credentials*. These credentials are compared to information that has been previously stored for this subject. If these credentials match the stored information, the subject is authenticated. But we are not done yet.

Once the subject provides its credentials and is properly authenticated, the system it is trying to access needs to determine if this subject has been given the necessary rights and privileges to carry out the requested actions. The system may look at an access control matrix or compare security labels to verify that this subject may indeed access the requested resource and perform the actions it is attempting. If the system determines that the subject may access the resource, it *authorizes* the subject.

Although identification, authentication, authorization, and accountability have close and complementary definitions, each has distinct functions that fulfill a specific requirement in the process of access control. A user may be properly identified and authenticated to the network, but may not have the authorization to access certain files on the file server. On the other hand, a user may be authorized to access the files on the file server, but until she is properly identified and authenticated, those resources are out of reach. Figure 16-1 illustrates the four steps that must happen for a subject to access an object.

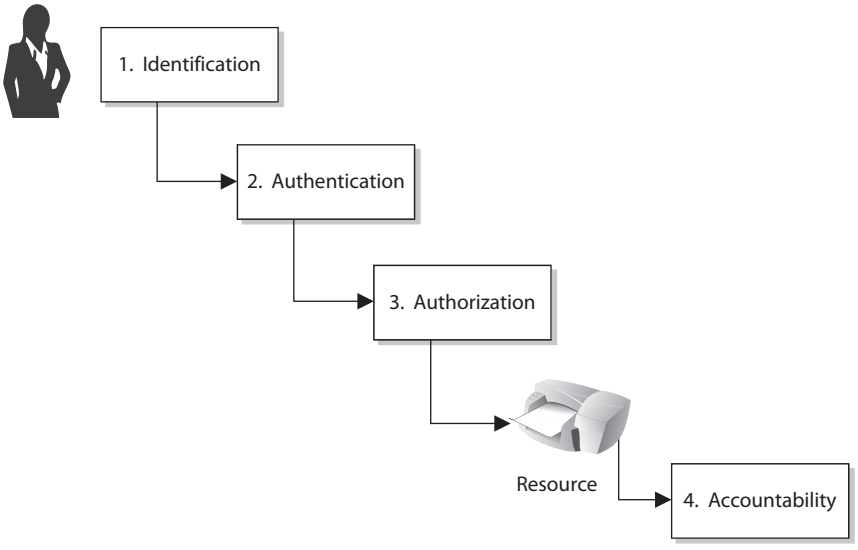



**Figure 16-1** Four steps must happen for a subject to access an object: identification, authentication, authorization, and accountability.

### Race Condition

A *race condition* occurs when processes carry out their tasks on a shared resource in an incorrect order. A race condition is possible when two or more processes use a shared resource, such as data within a variable. It is important that the processes carry out their functionality in the correct sequence. If process 2 carried out its task on the data before process 1, the result would be much different than if process 1 carried out its tasks on the data before process 2.

In software, when the authentication and authorization steps are split into two functions, there is a possibility an attacker could use a race condition to force the authorization step to be completed *before* the authentication step. This would be a flaw in the software that the attacker has figured out how to exploit. A race condition occurs when two or more processes use the same resource and the sequence of steps within the software can be carried out in an improper order, something that can drastically affect the output. So, an attacker can force the authorization step to take place before the authentication step and gain unauthorized access to a resource.

The subject needs to be held accountable for the actions taken within a system or domain. The only way to ensure accountability is if the subject is uniquely identified and the subject's actions are recorded.

*Logical access controls* are technical tools used for identification, authentication, authorization, and accountability. They are software components that enforce access control measures for systems, programs, processes, and information. The logical access controls can be embedded within operating systems, applications, add-on security packages, or database and telecommunication management systems. It can be challenging to synchronize all access controls and ensure all vulnerabilities are covered without producing overlaps of functionality. However, if it were easy, security professionals would not be getting paid the big bucks!




**EXAM TIP** The words "logical" and "technical" can be used interchangeably in this context. It is conceivable that the CISSP exam would refer to logical and technical controls interchangeably.

An individual's identity must be verified during the authentication process. Authentication usually involves a two-step process: entering public information (a username, employee number, account number, or department ID), and then entering private information (a static password, smart token, cognitive password, one-time password, or PIN). Entering public information is the identification step, while entering private information is the authentication step of the two-step process. Each technique used for identification and authentication has its pros and cons. Each should be properly evaluated to determine the right mechanism for the correct environment.

## Identification and Authentication

Once a person has been identified through the user ID or a similar value, she must be authenticated, which means she must prove she is who she says she is. Three main types of factors can be used for authentication: *something a person knows*, *something a person has*, and *something a person is*. Sometimes, these factors are combined with two additional factors: *somewhere a person is* (logical or physical location) and *something a person does* (behavioral factor). These location and behavioral factors may not be all that strong by themselves, but when combined with other factors they can significantly improve the effectiveness of the authentication process.

Something a person knows (knowledge-based authentication [KBA]) can be, for example, a password, PIN, mother's maiden name, or the combination to a lock. Authenticating a person by something that she knows is usually the least expensive method to implement. The downside to this method is that another person may acquire this knowledge and gain unauthorized access to a resource.

Something a person has (ownership-based authentication) can be a key, swipe card, access card, or badge. This method is common for accessing facilities but could also be used to access sensitive areas or to authenticate systems. A downside to this method is that the item can be lost or stolen, which could result in unauthorized access.

Something specific to a person (biometric authentication) becomes a bit more interesting. This is not based on whether the person is an American, a geek, or an athlete—it is based on a physical attribute. Authenticating a person's identity based on a unique physical attribute is referred to as biometrics.

*Strong authentication* contains two or all of these three methods: something a person knows, has, or is. Using a biometric system by itself does not provide strong authentication because it provides only one out of the three methods. Biometrics supplies what a person is, not what a person knows or has. For a strong authentication process to be in place, a biometric system needs to be coupled with a mechanism that checks for one of the other two methods. For example, many times the person has to type a PIN into a keypad before the biometric scan is performed. This satisfies the "something the person knows" category. Conversely, the person could be required to swipe a magnetic card through a

### One-to-One and One-to-Many

*Verification 1:1* is the measurement of an identity against a single claimed identity. The conceptual question is, "Is this person who he claims to be?" So if Bob provides his identity and credential set, this information is compared to the data kept in an authentication database. If they match, we know that it is really Bob. If the identification is *1:N (many)*, the measurement of a single identity is compared against multiple identities. The conceptual question is, "Who is this person?" An example is if fingerprints were found at a crime scene, the cops would run them through their database to identify the suspect.

reader prior to the biometric scan. This would satisfy the "something the person has" category. Whatever identification system is used, for strong authentication to be in the process, it must include multiple factors.

**TIP** Strong authentication is also sometimes referred to as *multifactor authentication (MFA)*, which just means that more than one authentication method is used. While two-factor authentication (2FA) is common, *three-factor authentication* (for example, smart card, PIN, and retinal scan) is sometimes used.

Identity is a complicated concept with many varied nuances, ranging from the philosophical to the practical. A person may have multiple digital identities. For example, a user could be JPublic in a Windows domain environment, JohnP on a Unix server, JohnPublic on the mainframe, JJP in instant messaging, JohnCPublic in the certification authority, and JohnnyPub on Facebook. If the organization that employs that user wants to centralize all of its access control, these various identity names for the same person may cause the security administrator undue stress.

**NOTE** *Mutual authentication* is when the two communicating entities must authenticate to each other before passing data. For example, an authentication server may be required to authenticate to a user's system before allowing data to flow back and forth.

While most of this chapter deals with user authentication, it is important to realize system-based authentication is possible also. Computers and devices can be identified, authenticated, monitored, and controlled based upon their hardware addresses (media access control) and/or Internet Protocol (IP) addresses. Networks may have network access control (NAC) technology that authenticates systems before they are allowed access to the network. Every network device has a hardware address that is integrated into its network interface card (NIC) and a software-based address (IP) that either is assigned by a Dynamic Host Configuration Protocol (DHCP) server or locally configured.

### Identification Component Requirements

When issuing identification values to users, the following should be in place:

- Each identifier should be unique, for user accountability.
- A standard naming scheme should be followed.
- The value should be nondescriptive of the user's position or tasks.
- The value should not be shared between users.

## Knowledge-Based Authentication

We start off our discussion of authentication methods by looking at the most commonly used approach: using something that a person knows. This knowledge-based approach typically uses a password, passphrase, or cognitive password. Let's take a closer look at each.

### Passwords

User identification coupled with a reusable password is the most common form of system identification and authorization mechanisms. A *password* is a protected string of characters that is used to authenticate an individual. As stated previously, authentication factors are based on what a person knows, has, or is. A password is something the user knows, and in order to ensure its effectiveness for authentication, it must be kept secret.

**Password Policies** Although passwords are prevalent, they are also considered one of the weakest security mechanisms available. Why? Users usually choose passwords that are easily guessed (a spouse's name, a user's birth date, or a dog's name), or tell others their passwords, and many times write the passwords down on a sticky note and hide it under the keyboard. To most users, security is usually not the most important or interesting part of using their computers—except when someone hacks into their computer and steals confidential information, that is. Then security is all the rage.

This is where password policies step in. If passwords are properly generated, updated, and kept secret, they can provide effective security. Password generators can be used to create passwords for users. This ensures that a user will not be using "Bob" or "Spot" for a password, but if the generator spits out "kdjasijew284802h," the user will surely scribble it down on a piece of paper and stick it to the monitor, which defeats the whole purpose. If a password generator is going to be used, the tools should create uncomplicated, pronounceable, nondictionary words to help users remember them so they aren't tempted to write them down.

If users can choose their own passwords, the operating system should enforce certain password requirements. The operating system can require that a password contain a certain number of characters, unrelated to the user ID, and not be easily guessable. The operating system can keep track of the passwords a specific user generates so as to ensure no passwords are reused. In March of 2020 the National Institute of Standards and Technology (NIST) updated its guidelines concerning passwords in SP 800-63B. These include the following recommendations:

- **Increased password length** The longer the password, the harder it is to guess. The recommended minimum password length is 8 characters for user-selected ones and 6 characters for computer-generated passwords. The maximum recommended length is 64 characters.
- **Allow special characters** Users should be allowed to use any special character, and even emojis, in their passwords. Special characters, however, should not be required.
- **Disallow password hints** On the surface, password hints may seem to make sense because they allow users to remember complex passwords and reduce reliance on password resetting features. However, they mostly help attackers.

If an attacker is after a password, she can try a few different techniques:

- **Electronic monitoring** Listening to network traffic to capture information, especially when a user is sending her password to an authentication server. The password can be copied and reused by the attacker at another time, which is called a *replay attack*.
- **Access the password file** Usually done on the authentication server. The password file contains many users' passwords and, if compromised, can be the source of a lot of damage. This file should be protected with access control mechanisms and encryption.
- **Brute-force attacks** Performed with tools that cycle through many possible character, number, and symbol combinations to uncover a password.
- **Dictionary attacks** Comparing files of thousands of words to the user's password until a match is found.
- **Social engineering** Falsely convincing an individual that she has the necessary authorization to access specific resources.
- **Rainbow table** Using a table that contains all possible passwords already in a hash format.

Certain techniques can be implemented to provide another layer of security for passwords and their use. After each successful logon, a message can be presented to a user indicating the date and time of the last successful logon, the location of this logon, and whether there were any unsuccessful logon attempts. This alerts the user to any suspicious activity and whether anyone has attempted to log on using his credentials. An administrator can set system parameters that allow a certain number of failed logon attempts to be accepted before a user is locked out; this is a type of *clipping level*. The user can be locked out for five minutes or a full day, for example, after the threshold (or clipping level) has been exceeded. It depends on how the administrator configures this mechanism. An audit trail can also be used to track password usage and both successful and unsuccessful logon attempts. This audit information should include the date, time, user ID, and workstation the user logged in from.



**NOTE** *Clipping level* is an older term that just means threshold. If the number of acceptable failed login attempts is set to three, three is the threshold (clipping level) value.

Policies can also specify other conditions that make passwords more difficult to exploit. Many organizations maintain a password history so users cannot reuse passwords within a certain timeframe. A variation on this is having the system remember the last *n* (where *n* is some number greater than or equal to one) passwords to prevent their reuse. Policies can also specify maximum age (that is, expiration) and minimum age (so the password can't be changed immediately to bypass the other policies) requirements.

As with many things in life, education is the key. Password requirements, protection, and generation should be addressed in security awareness programs so users understand

what is expected of them, why they should protect their passwords, and how passwords can be stolen. Users should be an extension to a security team, not the opposition.

**NOTE** Rainbow tables contain passwords already in their hashed format. The attacker just compares a captured hashed password with one that is listed in the table to uncover the plaintext password. This takes much less time than carrying out a dictionary or brute-force attack.

**Password Checkers** Several organizations test user-chosen passwords using tools that perform dictionary and/or brute-force attacks to detect the weak passwords. This helps make the environment as a whole less susceptible to dictionary and exhaustive attacks used to discover users' passwords. Many times the same tools employed by an attacker to crack a password are used by a network administrator to make sure the password is strong enough. Most security tools have this dual nature. They are used by security professionals and IT staff to test for vulnerabilities within their environment in the hope of uncovering and fixing them before an attacker finds the vulnerabilities. An attacker uses the same tools to uncover vulnerabilities to exploit before the security professional can fix them. It is the never-ending cat-and-mouse game.

If a tool is called a *password checker*, it is used by a security professional to test the strength of a password. If a tool is called a *password cracker*, it is usually used by a hacker; however, most of the time, these tools are one and the same.

You need to obtain management's approval before attempting to test (break) employees' passwords with the intent of identifying weak passwords. Explaining you are trying to help the situation, not hurt it, *after* you have uncovered the CEO's password is not a good situation to be in.

**Password Hashing and Encryption** In most situations, if an attacker sniffs your password from the network wire, she still has some work to do before she actually knows your password value because most systems hash the password with a hashing algorithm, commonly Message Digest 5 (MD5) or Secure Hash Algorithm (SHA), to ensure passwords are not sent in cleartext.

Although some people think the world is run by Microsoft, other types of operating systems are out there, such as Unix and Linux. These systems do not use registries and SAM databases but contain their user passwords in a file cleverly called "shadow." This shadow file does not contain passwords in cleartext; instead, your password is run through a hashing algorithm, and the resulting value is stored in this file. Unix-type systems zest things up by using salts in this process. *Salts* are random values added to passwords prior to hashing to add more complexity and randomness. The more randomness entered into the hashing process, the harder it is for the bad guy to decrypt and uncover your password. The use of a salt means that the same password can be encrypted into several thousand different hashes. This makes it much more difficult for an adversary to attack the passwords in your system using approaches like rainbow tables.

**Limit Logon Attempts** A threshold can be set to allow only a certain number of unsuccessful logon attempts. After the threshold is met, the user's account can be locked for a period of time or indefinitely, which requires an administrator to manually unlock

the account. This protects against dictionary and other exhaustive attacks that continually submit credentials until the right combination of username and password is discovered.

### Passphrase

A *passphrase* is a sequence of characters that is longer than a password (thus a "phrase") and, in some cases, takes the place of a password during an authentication process. The user enters this phrase into an application, and the application transforms the value into a *virtual password*, making the passphrase the length and format that are required by the application. (For example, an application may require your virtual password to be 128 bits to be used as a key with the AES algorithm.) If a user wants to authenticate to an application, such as Pretty Good Privacy (PGP), he types in a passphrase, let's say StickWithMeKidAndYouWillWearDiamonds. The application converts this phrase into a virtual password that is used for the actual authentication. The user usually generates the passphrase in the same way a user creates a password the first time he logs on to a computer. A passphrase is more secure than a password because it is longer, and thus harder to obtain by an attacker. In many cases, the user is more likely to remember a passphrase than a password.

### Cognitive Password

*Cognitive passwords* are fact- or opinion-based information used to verify an individual's identity. A user is enrolled by answering several questions based on her life experiences. Passwords can be hard for people to remember, but that same person will not likely forget the first person they kissed, the name of their best friend in 8th grade, or their favorite cartoon character. After the enrollment process, the user can answer the questions asked of her to be authenticated instead of having to remember a password. This authentication process is best for a service the user does not use on a daily basis, because it takes longer than other authentication mechanisms. This can work well for help-desk services. The user can be authenticated via cognitive means. This way, the person at the help desk can be sure he is talking to the right person, and the user in need of help does not need to remember a password that may be used once every three months.


**EXAM TIP** Knowledge-based authentication means that a subject is authenticated based upon something she knows. This could be a PIN, password, passphrase, cognitive password, personal history information, or through the use of a CAPTCHA, which is the graphical representation of data. A CAPTCHA is a skewed representation of characteristics a person must enter to prove that the subject is a human and not an automated tool as in a software robot.

## Biometric Authentication

*Biometrics* verifies an individual's identity by analyzing a unique personal characteristic, which is one of the most effective and accurate methods of verifying identification. Biometrics is a very sophisticated technology; thus, it is much more expensive and complex than the other types of identity verification processes. Biometric systems typically

base authentication decisions on physical attributes (such as iris, retina, or fingerprint), which provides more accuracy because physical attributes typically don't change, absent some disfiguring injury, and are harder to impersonate.

Biometrics is typically broken up into two different categories:

- **Physiological** This category of biometrics uses physical attributes unique to a specific individual to verify that person's identity. Fingerprints are a common example of a physiological trait used in biometric systems. Physiological is "what you are."
- **Behavioral** This approach is based on something an individual does uniquely to confirm her identity. An example is signature dynamics. Behavioral is "what you do."

A biometric system scans a person's physiological attribute or behavioral trait and compares it to a record created in an earlier enrollment process. Because this system inspects the grooves of a person's fingerprint, the pattern of someone's retina, or the pitches of someone's voice, it must be extremely sensitive. The system must perform accurate and repeatable measurements of anatomical or behavioral characteristics. This type of sensitivity can easily cause false positives or false negatives. The system must be calibrated so these false positives and false negatives occur infrequently and the results are as accurate as possible.

When a biometric system rejects an authorized individual, it is called a *Type I error* (false rejection rate [FRR]). When the system accepts impostors who should be rejected, it is called a *Type II error* (false acceptance rate [FAR]). The goal is to obtain low numbers for each type of error, but Type II errors are the most dangerous and thus the most important to avoid.

When comparing different biometric systems, many different variables are used, but one of the most important metrics is the *crossover error rate (CER)*. This rating is stated as a percentage and represents the point at which the FRR equals the FAR. This rating is the most important measurement when determining the system's accuracy. A biometric system that delivers a CER of 3 will be more accurate than a system that delivers a CER of 4.

**NOTE** Crossover error rate (CER) is also called equal error rate (EER).

What is the purpose of this CER value anyway? Using the CER as an impartial judgment of a biometric system helps create standards by which products from different vendors can be fairly judged and evaluated. If you are going to buy a biometric system, you need a way to compare the accuracy between different systems. You can just go by the different vendors' marketing material (they all say they are the best), or you can compare the different CER values of the products to see which one really is more accurate than the others. It is also a way to keep the vendors honest. One vendor may tell you, "We have absolutely no Type II errors." This would mean that their product would not allow

any imposters to be improperly authenticated. But what if you asked the vendor how many Type I errors their product had and the rep sheepishly replied, "We average around 90 percent of Type I errors." That would mean that 90 percent of the authentication attempts would be rejected, which would negatively affect your employees' productivity. So you can ask a vendor about their product's CER value, which represents when the Type I and Type II errors are equal, to give you a better understanding of the product's overall accuracy.

Individual environments have specific security level requirements, which will dictate how many Type I and Type II errors are acceptable. For example, a military institution that is very concerned about confidentiality would be prepared to accept a certain rate of Type I errors, but would absolutely not accept any false accepts (Type II errors). Because all biometric systems can be calibrated, if you lower the Type II error rate by adjusting the system's sensitivity, it will typically result in an increase in Type I errors. The military institution would obviously calibrate the biometric system to lower the Type II errors to zero, but that would mean it would have to accept a higher rate of Type I errors.

Biometric authentication is the most expensive method of verifying a person's identity, and it faces other barriers to becoming widely accepted. These include user acceptance, enrollment timeframe, and throughput. Many people are reluctant to let a machine read the pattern of their retina or scan the geometry of their hand. The enrollment phase requires an action to be performed several times to capture a clear and distinctive reference record. People are not particularly fond of expending this time and energy when they are used to just picking a password and quickly typing it into their console. When a person attempts to be authenticated by a biometric system, sometimes the system will request an action to be completed several times. If the system is unable to get a clear reading of an iris scan or cannot capture a full voice verification print, the individual may have to repeat the action. This causes low throughput, stretches the individual's patience, and reduces acceptability.

During enrollment, the user provides the biometric data (e.g., fingerprint, voice print, or retina scan), and the biometric reader converts this data into binary values. Depending on the system, the reader may create a hash value of the biometric data, or it may encrypt the data, or do both. The biometric data then goes from the reader to a backend authentication database where the user's account has been created. When the user needs to later authenticate to a system, she provides the necessary biometric data, and the binary format of this information is compared to what is in the authentication database. If they match, then the user is authenticated.

In Figure 16-2, we see that biometric data can be stored on a smart card and used for authentication. Also, you might notice that the match is 95 percent instead of 100 percent. Obtaining a 100 percent match every time is very difficult because of the level of sensitivity of the biometric systems. A smudge on the reader, oil on the person's finger, and other small environmental issues can stand in the way of matching 100 percent. If your biometric system was calibrated so it required 100 percent matches, this would mean you would not allow any Type II errors and that users would commonly not be authenticated in a timely manner.

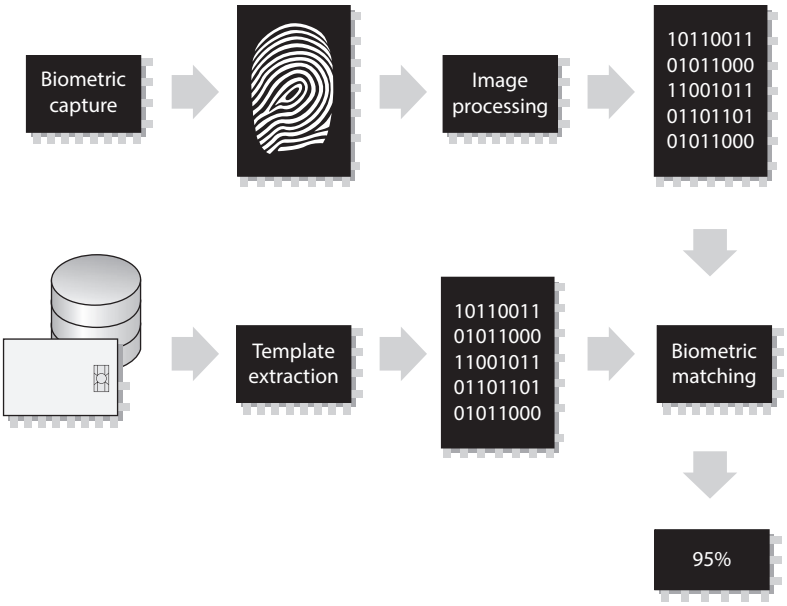



**Figure 16-2** Biometric data is turned into binary data and compared for identity validation.

### Processing Speed

When reviewing biometric devices for purchase, one component to take into consideration is the length of time it takes to actually authenticate users. From the time a user inserts data until she receives an accept or reject response should take five to ten seconds.

The following is an overview of the different types of biometric systems and the physiological or behavioral characteristics they examine.

## Fingerprint

Fingerprints are made up of ridge endings and bifurcations exhibited by friction ridges and other detailed characteristics called minutiae. It is the distinctiveness of these minutiae that gives each individual a unique fingerprint. An individual places his finger on a device that reads the details of the fingerprint and compares this to a reference file. If the two match, the individual's identity has been verified.

**NOTE** Fingerprint systems store the full fingerprint, which is actually a lot of information that takes up hard drive space and resources. The finger-scan technology extracts specific features from the fingerprint and stores just that information, which takes up less hard drive space and allows for quicker database lookups and comparisons.

## Palm Scan

The palm holds a wealth of information and has many aspects that are used to identify an individual. The palm has creases, ridges, and grooves throughout that are unique to a specific person. The palm scan also includes the fingerprints of each finger. An individual places his hand on the biometric device, which scans and captures this information. This information is compared to a reference file, and the identity is either verified or rejected.

## Hand Geometry

The shape of a person's hand (the shape, length, and width of the hand and fingers) defines hand geometry. This trait differs significantly between people and is used in some biometric systems to verify identity. A person places her hand on a device that has grooves for each finger. The system compares the geometry of each finger, and the hand as a whole, to the information in a reference file to verify that person's identity.

## Retina Scan

A system that reads a person's retina scans the blood-vessel pattern of the retina on the backside of the eyeball. This pattern is unique for each person. A camera is used to project a beam inside the eye and capture the pattern and compare it to a reference file recorded previously.

**NOTE** Retina scans are extremely invasive and involve a number of privacy issues. Since the information obtained through this scan can be used in the diagnosis of medical conditions, it could very well be considered protected health information (PHI) subject to healthcare information privacy regulations such as HIPAA.

## Iris Scan

The iris is the colored portion of the eye that surrounds the pupil. The iris has unique patterns, rifts, colors, rings, coronas, and furrows. The uniqueness of each of these characteristics within the iris is captured by a camera and compared with the information gathered during the enrollment phase. Of the biometric systems, iris scans are the most accurate. The iris remains constant through adulthood, which reduces the type of errors that can happen during the authentication process. Sampling the iris offers more reference coordinates than any other type of biometric. Mathematically, this means it has a higher accuracy potential than any other type of biometric.

**NOTE** When using an iris pattern biometric system, the optical unit must be positioned so the sun does not shine into the aperture; thus, when the system is implemented, it must be properly placed within the facility.

## Signature Dynamics

When a person writes a signature, usually they do so in the same manner and at the same speed each time. Writing a signature produces electrical signals that can be captured by

a biometric system. The physical motions performed when someone is signing a document create these electrical signals. The signals provide unique characteristics that can be used to distinguish one individual from another. Signature dynamics provides more information than a static signature, so there are more variables to verify when confirming an individual's identity and more assurance that this person is who he claims to be.

Signature dynamics is different from a digitized signature. A digitized signature is just an electronic copy of someone's signature and is not a biometric system that captures the speed of signing, the way the person holds the pen, and the pressure the signer exerts to generate the signature.

### Keystroke Dynamics

Whereas signature dynamics is a method that captures the electrical signals when a person signs a name, keystroke dynamics captures electrical signals when a person types a certain phrase. As a person types a specified phrase, the biometric system captures the speed and motions of this action. Each individual has a certain style and speed of typing, which translate into unique signals. This type of authentication is more effective than typing in a password, because a password is easily obtainable. It is much harder to repeat a person's typing style than it is to acquire a password.

### Voice Print

People's speech sounds and patterns have many subtle distinguishing differences. A biometric system that is programmed to capture a voice print and compare it to the information held in a reference file can differentiate one individual from another. During the enrollment process, an individual is asked to say several different words. Later, when this individual needs to be authenticated, the biometric system jumbles these words and presents them to the individual. The individual then repeats the sequence of words given. This technique is used so others cannot attempt to record the session and play it back in hopes of obtaining unauthorized access.

### Facial Scan

A system that scans a person's face takes many attributes and characteristics into account. People have different bone structures, nose ridges, eye widths, forehead sizes, and chin shapes. These are all captured during a facial scan and compared to an earlier captured scan held within a reference record. If the information is a match, the person is positively identified.

A naïve implementation of this technology could be fooled by a photograph of the legitimate user. To thwart this approach, the scanner can perform a three-dimensional measurement of the user's face by projecting thousands of infrared dots on it. This is how Apple's Face ID works.

### Hand Topography

Whereas hand geometry looks at the size and width of an individual's hand and fingers, hand topology looks at the different peaks and valleys of the hand, along with its overall shape and curvature. When an individual wants to be authenticated, she places her hand on the system. Off to one side of the system, a camera snaps a side-view picture of the

### Biometric Issues and Concerns

Biometric systems are not without their own sets of issues and concerns. Because they depend upon the specific and unique traits of living things, problems can arise. Living things are notorious for not remaining the same, which means they won't present static biometric information for every login attempt. Voice recognition can be hampered by a user with a cold. Retinas can detach. Someone could lose a finger. Or all three could happen. You just never know in this crazy world.

Some biometric systems actually check for the pulsation and/or heat of a body part to make sure it is alive. So if you are planning to cut someone's finger off or pluck out someone's eyeball so you can authenticate yourself as a legitimate user, it may not work. Although not specifically stated, this type of activity definitely falls outside the bounds of the CISSP ethics you will be responsible for upholding once you receive your certification.

hand from a different view and angle than that of systems that target hand geometry, and thus captures different data. This attribute is not unique enough to authenticate individuals by itself and is commonly used in conjunction with hand geometry.

## Ownership-Based Authentication

Authentication can also be based on something that the subject has. This is almost always some sort of physical or logical token. It can be a device such as a phone, identification card, or even an implanted device. It can also be a cryptographic key, such as a private key in public key infrastructure (PKI). Sometimes, access to the token is protected by some other authentication process, such as when you have to unlock your phone to get to a software-based token generator.

### One-Time Password

A *one-time password (OTP)*, also called a *dynamic password*, is used for authentication purposes and is valid only once. After the password is used, it is no longer valid; thus, it can't be reused if a hacker obtains it. The password is generated by a token device, which is something the person owns (or at least carries around). This device is the most common implementation mechanism for OTP and generates the one-time password for the user to submit to an authentication server. It is commonly implemented in three formats: as a dedicated physical device with a small screen that displays the OTP, as a smartphone application, and as a service that sends an SMS message to your phone. The following sections explain the concepts behind this technology.

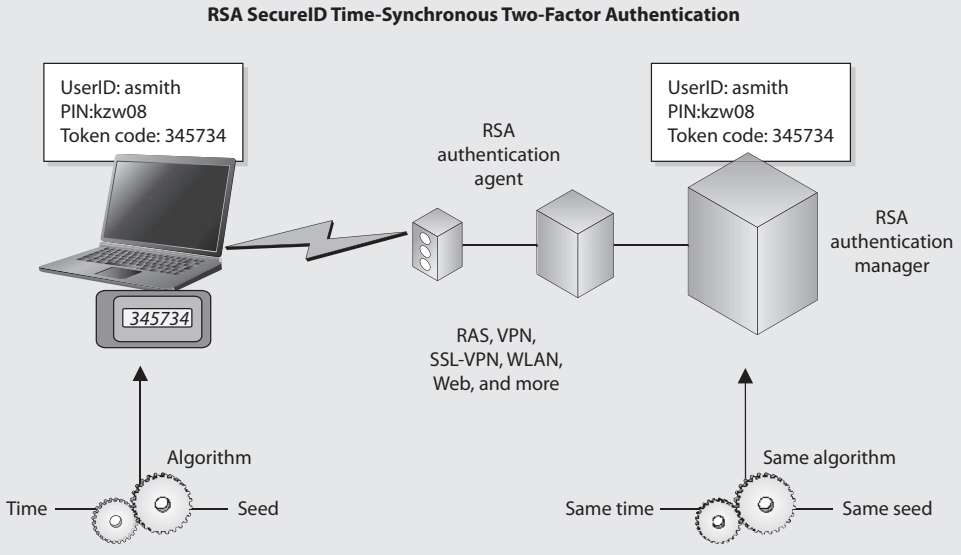
**NOTE** SMS was deprecated as a means of providing 2FA by the NIST in 2017. It is widely considered an insecure channel but is unfortunately still in common use.

**The Token Device** The token device, or password generator, is usually a handheld device that has a display and possibly a keypad. This hardware is separate from the computer the user is attempting to access. The token device and authentication service must be synchronized in some manner to be able to authenticate a user. The token device presents the user with a list of characters to be entered as a password when logging on to a computer. Only the token device and authentication service know the meaning of these characters. Because the two are synchronized, the token device presents the exact password the authentication service is expecting. This is a one-time password, also called a token, and is no longer valid after initial use.

**Synchronous** A *synchronous token device* requires the device and the authentication service to advance to the next OTP in sync with each other. This change can be triggered by time (e.g., every 30 seconds a new OTP is in play) or by simply going down a pre-agreed sequence of passwords, each of which is used only once before both the device and the server advance to the next one. The device displays the OTP to the user, who then enters this value and a user ID. The authentication service decrypts credentials and compares the OTP to the value it expects. If the two match, the user is authenticated and allowed to access the system.

## RSA SecurID

RSA SecurID, from RSA Security LLC, is a well-known time-based token. One version of the product generates the OTP by using a mathematical function on the time, date, and ID of the token card. Another version of the product requires a PIN to be entered into the token device.



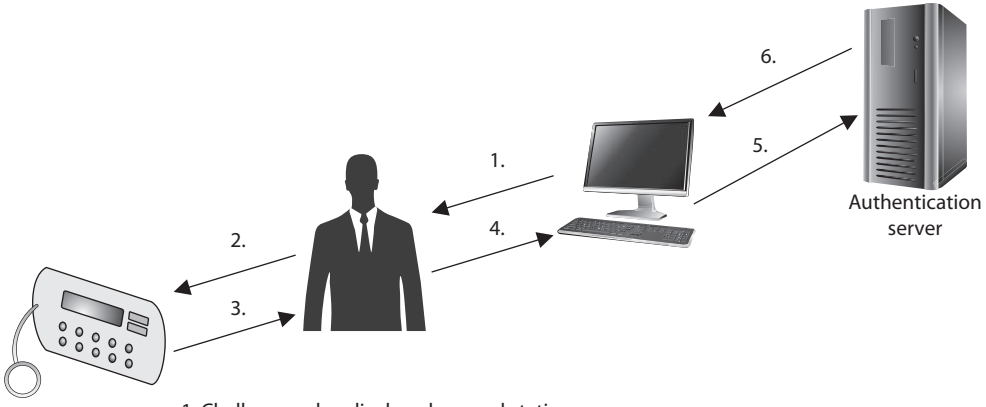
**EXAM TIP** Synchronous token-based OTP generation can be time-based or counter-based. Another term for counter-based is event-based. Counter-based and event-based are interchangeable terms, and you could see either or both on the CISSP exam.

**Asynchronous** A token device using an *asynchronous token*–generating method employs a challenge/response scheme to authenticate the user. In this situation, the authentication server sends the user a challenge, a random value, also called a *nonce*. The user enters this random value into the token device, which encrypts it and returns a value the user uses as an OTP. The user sends this value, along with a username, to the authentication server. If the authentication server can decrypt the value and it is the same challenge value sent earlier, the user is authenticated, as shown in Figure 16-3.

**EXAM TIP** The actual implementation and process that these devices follow can differ between different vendors. What is important to know is that asynchronous is based on challenge/response mechanisms, while synchronous is based on time- or counter-driven mechanisms.

Both token systems can fall prey to masquerading if a user shares his identification information (ID or username) and the token device is shared or stolen. The token device can also have battery failure or other malfunctions that would stand in the way of a successful authentication. However, this type of system is not vulnerable to electronic eavesdropping, sniffing, or password guessing.






1. Challenge value displayed on workstation.
2. User enters challenge value and PIN into token device.
3. Token device presents a different value to the user.
4. User enters new value into the workstation.
5. Value sent to authentication service on server.
6. AS sends an "allow access" response.

**Figure 16-3** Authentication using an asynchronous token device includes a workstation, token device, and authentication service.

If the user has to enter a password or PIN into the token device before it provides an OTP, then strong authentication is in effect because it is using two factors—something the user knows (PIN) and something the user has (the token device).

**NOTE** One-time passwords can also be generated in software, in which case a piece of hardware such as a token device is not required. These are referred to as *soft tokens* and require that the authentication service and application contain the same base secrets, which are used to generate the OTPs.

## Cryptographic Keys

Another way to prove one's identity is to use asymmetric cryptography and let the users' private keys show they are who they claim to be. Recall that the private key is kept secret by an individual and should never be shared. So, if the authentication server has (or gets a hold of) the user's public key, it can use that key to encrypt a challenge and send it to the user. Only the person owning the corresponding private key would be able to decrypt it and respond to it. Ideally, the user then encrypts the response using the server's public key to provide mutual authentication. This approach is commonly used in Secure Shell (SSH) instead of passwords, which are the weakest form of authentication and can be easily sniffed as they travel over a network.

## Memory Cards

The main difference between memory cards and smart cards is their capacity to process information. A *memory card* holds information but cannot process information. A *smart card* holds information and has the necessary hardware and software to actually process that information. A memory card can hold a user's authentication information so the user only needs to type in a user ID or PIN and present the memory card, and if the data that the user enters matches the data on the memory card, the user is successfully authenticated. If the user presents a PIN value, then this is an example of two-factor authentication—something the user knows and something the user has. A memory card can also hold identification data that is pulled from the memory card by a reader. It travels with the PIN to a backend authentication server.

An example of a memory card is a swipe card that must be used for an individual to be able to enter a building. The user enters a PIN and swipes the memory card through a card reader. If this is the correct combination, the reader flashes green and the individual can open the door and enter the building. Another example is an ATM card. If Buffy wants to withdraw $40 from her checking account, she needs to slide the ATM card (or memory card) through the reader and enter the correct PIN.

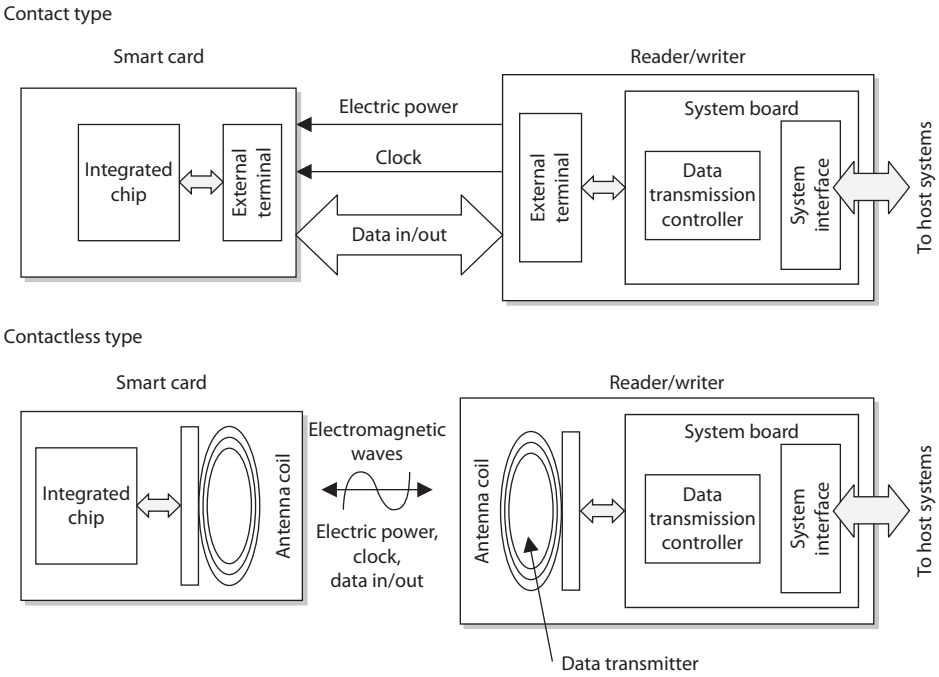
Memory cards can be used with computers, but they require a reader to process the information. The reader adds cost to the process, especially when one is needed per computer, and card generation adds cost and effort to the whole authentication process. Using a memory card provides a more secure authentication method than using a password because the attacker would need to obtain the card and know the correct PIN. Administrators and management must weigh the costs and benefits of a memory

token–based card implementation to determine if it is the right authentication mechanism for their environment.

## Smart Card

A *smart card* has the capability of processing information because it has a microprocessor and integrated circuits incorporated into the card itself. Memory cards do not have this type of hardware and lack this type of functionality. The only function they can perform is simple storage. A smart card, which adds the capability to process information stored on it, can also provide a two-factor authentication method because the user may have to enter a PIN to unlock the smart card. This means the user must provide something she knows (PIN) and something she has (smart card).

Two general categories of smart cards are the contact and the contactless types. The *contact* smart card has a gold seal on the face of the card. When this card is fully inserted into a card reader, electrical fingers wipe against the card in the exact position that the chip contacts are located. This supplies power and data I/O to the chip for authentication purposes. The *contactless* smart card has an antenna wire that surrounds the perimeter of the card. When this card comes within an electromagnetic field of the reader, the antenna within the card generates enough energy to power the internal chip. Now, the results of the smart card processing can be broadcast through the same antenna, and the conversation of authentication can take place. The authentication can be completed by using a one-time password, by employing a challenge/response value, or by providing the user's private key if it is used within a PKI environment.




PART V

**TIP** Two types of contactless smart cards are available: hybrid and combi. The hybrid card has two chips, with the capability of utilizing both the contact and contactless formats. A combi card has one microprocessor chip that can communicate to contact or contactless readers.

The information held within the memory of a smart card is not readable until the correct PIN is entered. This fact and the complexity of the smart token make these cards resistant to reverse-engineering and tampering methods. If George loses the smart card he uses to authenticate to the domain at work, the person who finds the card would need to know his PIN to do any real damage. The smart card can also be programmed to store information in an encrypted fashion, as well as detect any tampering with the card itself. In the event that tampering is detected, the information stored on the smart card can be automatically wiped.

The drawbacks to using a smart card are the extra cost of the readers and the overhead of card generation, as with memory cards, although this cost is decreasing. The smart cards themselves are more expensive than memory cards because of the extra integrated circuits and microprocessor. Essentially, a smart card is a kind of computer, and because of that it has many of the operational challenges and risks that can affect a computer.

Smart cards have several different capabilities, and as the technology develops and memory capacities increase for storage, they will gain even more. They can store personal information in a storage manner that is tamper resistant. This also gives them the capability to isolate security-critical computations within themselves. They can be used in encryption systems to store keys and have a high level of portability as well as security. The memory and integrated circuit also provide the capacity to use encryption algorithms on the actual card and use them for secure authorization that can be utilized throughout an entire organization.

**Smart Card Attacks** Smart cards are more tamperproof than memory cards, but where there is sensitive data, there are individuals who are motivated to circumvent any countermeasure the industry throws at them. Over the years, criminals have become very inventive in the development of various ways to attack smart cards. Smart card attacks tend to be special cases of the cryptanalysis techniques we discussed in Chapter 8. For example, attackers have introduced computational errors into smart cards with the goal of uncovering the encryption keys used and stored on the cards. These "errors" are introduced by manipulating some environmental component of the card (changing input voltage, clock rate, temperature fluctuations). The attacker reviews the result of an encryption function after introducing an error to the card, and also reviews the correct result, which the card performs when no errors are introduced. Analysis of these different results may allow an attacker to reverse-engineer the encryption process, with the hope of uncovering the encryption key. This type of attack is referred to as *fault generation*.

*Side-channel attacks* are nonintrusive and are used to uncover sensitive information about how a component works, without trying to compromise any type of flaw or weakness. So a noninvasive attack is one in which the attacker watches how something works and how it reacts in different situations instead of trying to "invade" it with more

### Interoperability

In the industry today, lack of interoperability is a big problem. An ISO/IEC standard, 14443, outlines the following items for smart card standardization:

- **ISO/IEC 14443-1** Physical characteristics
- **ISO/IEC 14443-2** Radio frequency power and signal interface
- **ISO/IEC 14443-3** Initialization and anticollision
- **ISO/IEC 14443-4** Transmission protocol

intrusive measures. Some examples of side-channel attacks that have been carried out on smart cards are *differential power analysis* (examining the power emissions released during processing), *electromagnetic analysis* (examining the frequencies emitted), and *timing* (how long a specific process takes to complete). These types of attacks are used to uncover sensitive information about how a component works without trying to compromise any type of flaw or weakness. They are commonly used for data collection. Attackers monitor and capture the analog characteristics of all supply and interface connections and any other electromagnetic radiation produced by the processor during normal operation. They can also collect the time it takes for the smart card to carry out its function. From the collected data, the attacker can deduce specific information she is after, which could be a private key, sensitive financial data, or an encryption key stored on the card.

*Software attacks* are also considered noninvasive attacks. A smart card has software just like any other device that does data processing, and anywhere there is software, there is the possibility of software flaws that can be exploited. The main goal of this type of attack is to input into the card instructions that will allow the attacker to extract account information, which he can use to make fraudulent purchases. Many of these types of attacks can be disguised by using equipment that looks just like the legitimate reader.

A more intrusive smart card attack is called *microprobing*, which uses needleless and ultrasonic vibration to remove the outer protective material on the card's circuits. Once this is completed, data can be accessed and manipulated by directly tapping into the card's ROM chips.

### Near Field Communications

*Near Field Communication (NFC)* is a short-range (i.e., a few centimeters) radio frequency (RF) communications technology that provides data communication on a base frequency of 13.56 MHz. Manufacturers of NFC devices abide by ISO/IEC 18092 for international interoperability. While this technology is perhaps best known for contactless payments using mobile phones, it is also used for contactless smart cards.

# Credential Management

Credential management deals with creating user accounts on all systems, assigning and modifying the account details and privileges when necessary, and decommissioning the accounts when they are no longer needed. In many environments, the IT department creates accounts manually on the different systems, users are given excessive rights and permissions, and when an employee leaves the organization, many or all of the accounts stay active. This typically occurs because a centralized credential management technology has not been put into place.

Credential management products attempt to attack these issues by allowing an administrator to manage user accounts across multiple systems. When there are multiple directories containing user profiles or access information, the account management software allows for replication between the directories to ensure each contains the same up-to-date information. This automated workflow capability not only reduces the potential errors that can take place in account management, it also logs and tracks each step (including account approval). This allows for accountability and provides documentation for use in backtracking if something goes wrong. Automated workflow also helps ensure that only the necessary amount of access is provided to the account and that there are no "orphaned" accounts still active when employees leave the organization. In addition, these types of processes are the kind your auditors will be looking for—and we always want to make the auditors happy!

**NOTE** These types of credential management products are commonly used to set up and maintain internal accounts. Web access control management is used mainly for external users.

Enterprise credential management products are usually expensive and can take time to properly roll out across the enterprise. Regulatory requirements, however, are making more and more organizations spend the money for these types of solutions—which the vendors love! In the following sections, we'll explore the many facets of a good credential management solution.

## Password Managers

Two of the best practices when it comes to password-based authentication are to use complex passwords/passphrases and to have a different one for each account; accomplishing both from memory is a tall order for most of us. A popular solution to address this challenge is to use software products that remember our credentials for us. These products, known as *password managers* or *password vaults*, come in two flavors: as a stand-alone application or as a feature within another application (such as a web browser). In either case, the application stores user identifiers and passwords in a password-encrypted data store. The user need only remember this master password and the application maintains all others. These products typically provide random password generation and allow the user to store other information such as URLs and notes. Most modern web browsers also provide features that remember the user identifiers and passwords for specific websites.

An obvious problem with using password vaults is that they provide one-stop-shopping for malicious actors. If they can exploit this application, they gain access to all of the user's credentials. Developers of these applications go to great lengths to ensure they are secure, but as we all know there is no such thing as a 100 percent secure system. In fact, there have been multiple documented vulnerabilities that allowed adversaries to steal these (supposedly secure) credentials.

## Password Synchronization

Another approach to credential management is to use password synchronization technologies that can allow a user to maintain just one password across multiple systems. The product synchronizes the password to other systems and applications, which happens transparently to the user. The goal is to require the user to memorize only one password, which enables the organization to enforce more robust and secure password requirements. If a user needs to remember only one password, he is more likely to not have a problem with longer, more complex strings of values. This reduces help-desk call volume and allows the administrator to keep her sanity for just a little bit longer.

One criticism of this approach is that since only one password is used to access different resources, the hacker only has to figure out one credential set to gain unauthorized access to all resources. But if the password requirements are more demanding (12 characters, no dictionary words, three symbols, upper- and lowercase letters, and so on) and the password is changed out regularly, the balance between security and usability can be acceptable.

## Self-Service Password Reset

Some products are implemented to allow users to reset their own passwords. This does not mean that the users have any type of privileged permissions on the systems to allow them to change their own credentials. Instead, during the registration of a user account, the user can be asked to provide several personal questions (first car, favorite teacher, favorite color, and so on) in a question-and-answer form. When the user forgets his password, he may be required to provide another authentication mechanism (smart card, token, etc.) and to answer these previously answered questions to prove his identity.

Products are available that allow users to change their passwords through other means. For example, if you forgot your password, you may be asked to answer some of the questions answered during the registration process of your account (i.e., a cognitive password). If you do this correctly, an e-mail is sent to you with a link you must click. The password management product has your identity tied to the answers you gave to the questions during your account registration process and to your e-mail address. If you do everything correctly, you are given a screen that allows you to reset your password.

**CAUTION** The product should not ask for information that is publicly available, as in your mother's maiden name, because anyone can find that out and attempt to identify himself as you.

## Assisted Password Reset

Some products are created for help-desk employees who need to work with individuals when they forget their password. The help-desk employee should not know or ask the individual for her password. This would be a security risk since only the owner of the password should know the value. The help-desk employee also should not just change a password for someone calling in without authenticating that person first. This can allow social engineering attacks where an attacker calls the help desk and indicates she is someone who she is not. If this were to take place, an attacker would have a valid employee password and could gain unauthorized access to the organization's jewels.

The products that provide assisted password reset functionality allow the help-desk individual to authenticate the caller before resetting the password. This authentication process is commonly performed through the use of cognitive passwords described in the previous section. The help-desk individual and the caller must be identified and authenticated through the password management tool before the password can be changed. Once the password is updated, the system that the user is authenticating to should require the user to change her password again. This would ensure that only she (and not she and the help-desk person) knows her password. The goal of an assisted password reset product is to reduce the cost of support calls and ensure all calls are processed in a uniform, consistent, and secure fashion.

## Just-in-Time Access

You probably don't want your general users having administrative privileges on their computers. However, if you apply the security principle of least privilege (described in Chapter 9), your users will probably lack the authorization to perform many functions that you would like them to be able to perform in certain circumstances. From having their laptops "forget" wireless networks to which they may have connected, to updating software, there are many scenarios in which a regular user may need administrative (or otherwise elevated) credentials. The traditional approach is to have the user put in a ticket and wait for an IT administrator to perform the action for the user. This is a costly way of doing business, particularly if you have a large organization.

*Just-in-time (JIT) access* is a provisioning methodology that elevates users to the necessary privileged access to perform a specific task. This is a way to allow users to take care of routine tasks that would otherwise require IT staff intervention (and possibly decrease user productivity). This approach mitigates the risk of privileged account abuse by reducing the time a threat actor has to gain access to a privileged account. JIT access is usually granted in a granular manner, so that it applies to a specific resource or action in a given timeframe. For example, if users need administrative rights to allow a conferencing application access to their desktop, they can be granted one-time access to change that particular setting in their systems and then it's gone.

## Registration and Proofing of Identity

Now let's think about how accounts are set up. In many environments, when a new user needs an account, a network administrator sets up the account(s) and provides some type

of privileges and permissions. But how would the network administrator know what resources this new user should have access to and what permissions should be assigned to the new account? In most situations, she doesn't—she just wings it. This is how users end up with too much access to too many resources. What should take place instead is implementation of a workflow process that allows for a request for a new user account. Since hardly anyone in the organization likely knows the new employee, we need someone to vouch for this person's identity. This process, sometimes called *proofing of identity*, is almost always carried out by human resources (HR) personnel who would've had to verify the new employee's identity for tax and benefit purposes. The new account request is then sent to the employee's manager, who verifies the permissions that this person needs, and a ticket is generated for the technical staff to set up the account(s).

If there is a request for a change to the permissions on the account or if an account needs to be decommissioned, it goes through the same process. The request goes to a manager (or whoever is delegated with this approval task), the manager approves it, and the changes to the various accounts take place.

Over time, this new user will commonly have different identity attributes, which will be used for authentication purposes, stored in different systems in the network. When a user requests access to a resource, all of his identity data has already been copied from other identity stores and the HR database and held in this centralized directory (sometimes called the *identity repository*). When this employee parts with the organization for any reason, this new information goes from the HR database to the directory. An e-mail is automatically generated and sent to the manager to allow this account to be decommissioned. Once this is approved, the account management software disables all of the accounts that had been set up for this user.

*User provisioning* refers to the creation, maintenance, and deactivation of user objects and attributes as they exist in one or more systems, directories, or applications, in response to business processes. User provisioning software may include one or more of the following components: change propagation, self-service workflow, consolidated user administration, delegated user administration, and federated change control.



### Authoritative System of Record

The authoritative source is the "system of record," or the location where identity information originates and is maintained. It should have the most up-to-date and reliable identity information. An *authoritative system of record (ASOR)* is a hierarchical tree-like structure system that tracks subjects and their authorization chains. Organizations need an automated and reliable way of detecting and managing unusual or suspicious changes to user accounts and a method of collecting this type of data through extensive auditing capabilities. The ASOR should contain the subject's name, associated accounts, authorization history per account, and provision details. This type of workflow and accounting is becoming more in demand for regulatory compliance because it allows auditors to understand how access is being centrally controlled within an environment.

User objects may represent employees, contractors, vendors, partners, customers, or other recipients of a service. Services may include e-mail, access to a database, access to a file server or database, and so on.

Great. So we create, maintain, and deactivate accounts as required based on business needs. What else does this mean? The creation of the account also is the creation of the access rights to organizational assets. It is through provisioning that users either are given access or have access taken away. Throughout the life cycle of a user identity, access rights, permissions, and privileges should change as needed in a clearly understood, automated, and audited process.

## Profile Update

Most companies do not just contain the information "Bob Smith" for a user and make all access decisions based on this data. There can be a plethora of information on a user that is captured (e-mail address, home address, phone number, and so on). When this collection of data is associated with the identity of a user, it is called a *profile*.

Profiles should be centrally located to enable administrators to efficiently create, edit, or delete these profiles in an automated fashion when necessary. Many user profiles contain nonsensitive data that users can update themselves (called *self-service*). So, if George moved to a new house, there should be a profile update tool that allows him to go into his profile and change his address information. Now, his profile may also contain sensitive data that should not be available to George—for example, his access rights to resources or information that he is going to be laid off on Friday.

You have interacted with a profile update technology if you have requested to update your personal information on any e-commerce website. These companies provide you with the capability to sign in and update the information they allow you to access. This could be your contact information, home address, purchasing preferences, or credit card data. They then use this information to update their customer relationship management (CRM) systems so they know where to send you their junk mail advertisements and spam messages!

## Session Management

A *session* is an agreement between two parties to communicate interactively. Think of it as a phone call: you dial your friend's number, she decides whether to answer, and if she does then you talk with each other until something happens to end the call. That "something" could be that you (or her) are out of time and have to go, or maybe one of you runs out of things to say and there's an awkward silence on the line, or maybe one of you starts acting weird and the other is bothered and hangs up. Technically, the call could go on forever, though in practice that doesn't happen.

Information systems use sessions all the time. When you show up for work and log onto your computer, you establish an authenticated session with the operating system that allows you to launch your e-mail client. When that application connects to the mail server, it establishes a different authenticated session (perhaps using the same credentials you used to log onto your computer). So, a session, in the context of information systems security, can exist between a user and an information system or between two

information systems (e.g., two running programs). If the session is an authenticated one, as in the previous two examples, then authentication happens at the beginning and then everything else is trusted until the session ends.

That trust is the reason we need to be very careful about how we deal with our sessions. Threat actors often try to inject themselves into an authenticated session and hijack it for their own purposes. Session management is the process of establishing, controlling, and terminating sessions, usually for security reasons. The session establishment usually entails authentication and authorization of one or both endpoints. Controlling the session can involve logging the start and end and anything in between. It could also keep track of time, activity, and even indicia of malicious activity. These are three of the most common triggers for session termination:

- **Timeout** When sessions are established, the endpoints typically agree on how long they will last. You should be careful to make this time window as short as possible without unduly impacting the organization. For example, a VPN concentrator could enforce sessions of no more than eight hours for your teleworkers.
- **Inactivity** Some sessions could go on for very long periods of time, provided that the user is active. Sessions that are terminated for inactivity tend to have a shorter window than those that are triggered only by total duration (i.e., timeout). For example, many workstations lock the screen if the user doesn't use the mouse or keyboard for 15 minutes.
- **Anomaly** Usually, anomaly detection is an additional control added to a session that is triggered by timeouts or inactivity (or both). This control looks for suspicious behaviors in the session, such as requests for data that are much larger than usual or communication with unusual or forbidden destinations. These can be indicators of session hijacking.



## Accountability

Auditing capabilities ensure users are accountable for their actions, verify that the security policies are enforced, and can be used as investigation tools. There are several reasons why network administrators and security professionals want to make sure accountability mechanisms are in place and configured properly: to deter wrongdoing, be able to track bad deeds back to individuals, detect intrusions, reconstruct events and system conditions, provide legal recourse material, and produce problem reports. Audit documentation and log files hold a mountain of information—the trick is usually deciphering it and presenting it in a useful and understandable format.

Accountability is enabled by recording user, system, and application activities. This recording is done through auditing functions and mechanisms within an operating system or application. Audit trails contain information about operating system activities, application events, and user actions. Audit trails can be used to verify the health of a system by checking performance information or certain types of errors and conditions. After a system crashes, a network administrator often will review audit logs to try and piece together the status of the system and attempt to understand what events could be attributed to the disruption.

Audit trails can also be used to provide alerts about any suspicious activities that can be investigated at a later time. In addition, they can be valuable in determining exactly how far an attack has gone and the extent of the damage that may have been caused. It is important to make sure a proper chain of custody is maintained to ensure any data collected can later be properly and accurately represented in case it needs to be used for later events such as criminal proceedings or investigations.

Keep the following in mind when dealing with auditing:

- Store the audits securely.
- Use audit tools that keep the size of the logs under control.
- Protect the logs from any unauthorized changes in order to safeguard data.
- Train staff to review the data in the right manner while protecting privacy.
- Make sure the ability to delete logs is only available to administrators.
- Configure logs to contain activities of all high-privileged accounts (root, administrator).

An administrator configures what actions and events are to be audited and logged. In a high-security environment, the administrator would configure more activities to be captured and set the threshold of those activities to be more sensitive. The events can be reviewed to identify where breaches of security occurred and if the security policy has been violated. If the environment does not require such levels of security, the events analyzed would be fewer, with less-demanding thresholds.

Without proper oversight, items and actions to be audited can become an endless list. A security professional should be able to assess an environment and its security goals, know what actions should be audited, and know what is to be done with that information after it is captured—without wasting too much disk space, CPU power, and staff time. The following is a broad overview of the items and actions that can be audited and logged.

**System-level events:**

- System performance
- Logon attempts (successful and unsuccessful)
- Logon ID
- Date and time of each logon attempt
- Lockouts of users and terminals
- Use of administration utilities
- Devices used
- Functions performed
- Requests to alter configuration files

**Application-level events:**

- Error messages
- Files opened and closed
- Modifications of files
- Security violations within applications

**User-level events:**

- Identification and authentication attempts
- Files, services, and resources used
- Commands initiated
- Security violations

The threshold (clipping level) and parameters for each of these items must be deliberately configured. For example, an administrator can audit each logon attempt or just each failed logon attempt. System performance can look at the amount of memory used within an eight-hour period or the memory, CPU, and hard drive space used within an hour.

Intrusion detection systems (IDSs) continually scan audit logs for suspicious activity. If an intrusion or harmful event takes place, audit logs are usually kept to be used later to prove guilt and prosecute if necessary. If severe security events take place, the IDS alerts the administrator or staff member so they can take proper actions to end the destructive activity. If a dangerous virus is identified, administrators may take the mail server offline. If an attacker is accessing confidential information within the database, this computer may be temporarily disconnected from the network or Internet. If an attack is in progress, the administrator may want to watch the actions taking place so she can track down the intruder. IDSs can watch for this type of activity during real time and/or scan audit logs and watch for specific patterns or behaviors.



## Review of Audit Information

Audit trails can be reviewed manually or through automated means—either way, they must be reviewed and interpreted. If an organization reviews audit trails manually, it needs to establish a system of how, when, and why they are viewed. Usually audit logs are very popular items right after a security breach, unexplained system action, or system disruption. An administrator or staff member rapidly tries to piece together the activities that led up to the event. This type of audit review is event-oriented. Audit trails can also be viewed periodically to watch for unusual behavior of users or systems and to help understand the baseline and health of a system. Then there is a real-time, or near real-time, audit analysis that can use an automated tool to review audit information as it is created. Administrators should have a scheduled task of reviewing audit data. The audit material usually needs to be parsed and saved to another location for a certain time period. This retention information should be stated in the organization's security policy and procedures.

Reviewing audit information manually can be overwhelming. Fortunately, there are applications and audit trail analysis tools that reduce the volume of audit logs to review and improve the efficiency of manual review procedures. A majority of the time, audit logs contain information that is unnecessary, so these tools parse out specific events and present them in a useful format.

An *audit-reduction tool* does just what its name suggests—reduces the amount of information within an audit log. This tool discards mundane task information and records system performance, security, and user functionality information that can be useful to a security professional or administrator.

Today, more organizations are implementing *security information and event management (SIEM)* systems. These products gather logs from various devices (servers, firewalls, routers, etc.) and attempt to correlate the log data and provide analysis capabilities. Reviewing logs manually looking for suspicious activity in a continuous manner is not only mind-numbing; it is close to impossible to be successful. So many packets and network communication data sets are passing along a network, humans cannot collect all the data in real or near real time, analyze it, identify current attacks, and react—it is just too overwhelming.

Organizations also have different *types* of systems on a network (routers, firewalls, IDS, IPS, servers, gateways, proxies) collecting logs in various proprietary formats, which requires centralization, standardization, and normalization. Log formats are different per product type and vendor. The format of logs created by Juniper network device systems is different from the format of logs created by Cisco systems, which in turn is different from the format created by Palo Alto and Barracuda firewalls. It is important to gather logs from various different systems within an environment so that some type of situational awareness can take place. Once the logs are gathered, intelligence routines need to be processed on them so that data mining can take place to identify patterns. The goal is to piece together seemingly unrelated event data so that the security team can fully understand what is taking place within the network and react properly.

**NOTE** Situational awareness means that you understand the current environment even though it is complex, dynamic, and made up of seemingly unrelated data points. You need to be able to understand each data point in its own context within the surrounding environment so that you can make the best possible decisions.

## Protecting Audit Data and Log Information

If an intruder breaks into your house, he will do his best to cover his tracks by not leaving fingerprints or any other clues that can be used to tie him to the criminal activity. The same is true in computer fraud and illegal activity. The intruder will work to cover his tracks. Attackers often delete audit logs that hold this incriminating information. (Deleting specific incriminating data within audit logs is called *scrubbing*.) Deleting this information can cause the administrator to not be alerted or aware of the security breach and can destroy valuable data. Therefore, audit logs should be protected by strict access control and stored on a remote host.

Only certain individuals (the administrator and security personnel) should be able to view, modify, and delete audit trail information. No other individuals should be able to view this data, much less modify or delete it. The integrity of the data can be ensured with the use of digital signatures, hashing tools, and strong access controls. Its confidentiality can be protected with encryption and access controls, if necessary, and it can be stored on *write-once media* (optical discs) to prevent loss or modification of the data. Unauthorized access attempts to audit logs should be captured and reported.

Audit logs may be used in a trial to prove an individual's guilt, demonstrate how an attack was carried out, or corroborate a story. The integrity and confidentiality of these logs will be under scrutiny. Proper steps need to be taken to ensure that the confidentiality and integrity of the audit information are not compromised in any way.

**NOTE** We cover investigative techniques and evidence handling in Chapter 22.

# Identity Management

*Identity management (IdM)* is a broad term that encompasses the use of different products to identify, authenticate, and authorize users through automated means. It usually includes user account management, access control, credential management, single sign-on (SSO) functionality, managing rights and permissions for user accounts, and auditing and monitoring all of these items. It is important for security professionals to understand all the technologies that make up a full enterprise IdM solution. IdM requires managing uniquely identified entities, their attributes, credentials, and entitlements. IdM allows organizations to create and manage digital identities' life cycles (create, maintain, terminate) in a timely and automated fashion. An enterprise IdM solution must meet business needs and scale from internally facing systems to externally facing systems. In this section, we cover many of these technologies and how they work together.

**NOTE** Identity and access management (IAM) is another term that is used interchangeably with IdM, though ISC² considers IdM to be a subset of IAM.

Selling identity management products is a flourishing market that focuses on reducing administrative costs, increasing security, meeting regulatory compliance, and improving upon service levels throughout enterprises. The continual increase in complexity and diversity of networked environments also increases the complexity of keeping track of who can access what and when. Organizations have different types of applications, network operating systems, databases, enterprise resource management (ERM) systems, customer relationship management (CRM) systems, directories, and mainframes—all used for different business purposes. Organizations also have partners, contractors, consultants, employees, and temporary employees. (Figure 16-4 provides a simplistic

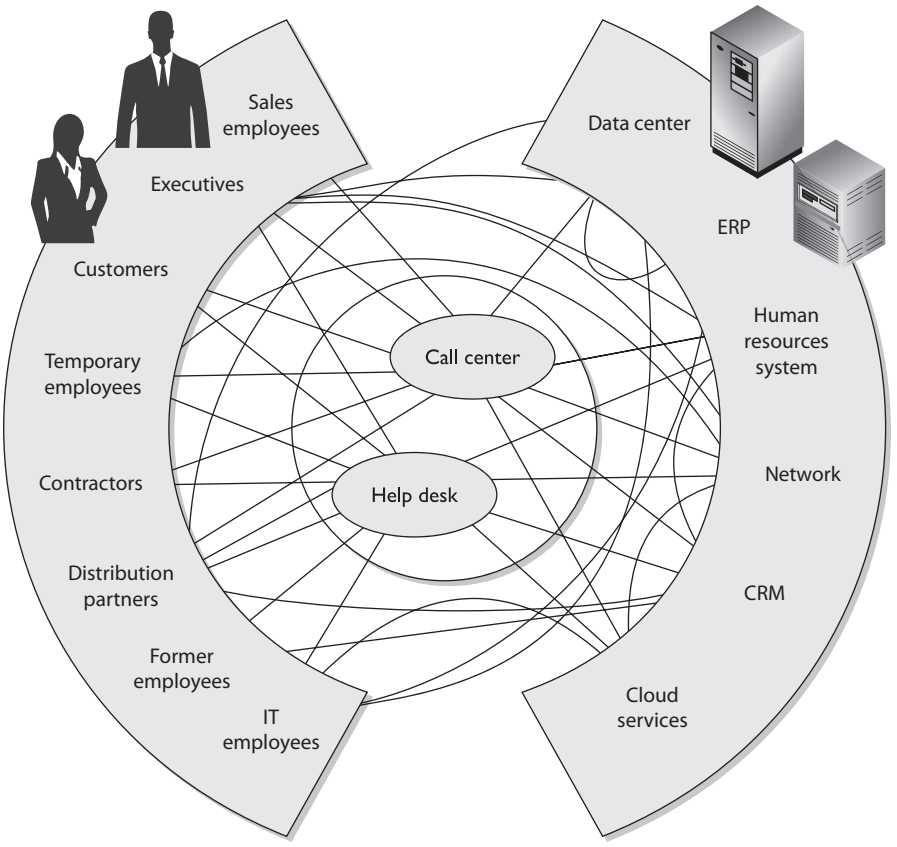



**Figure 16-4** Most environments are complex in terms of access.

view of most environments.) Users usually access several different types of systems throughout their daily tasks, which makes controlling access and providing the necessary level of protection on different data types difficult and full of obstacles. This complexity usually results in unforeseen and unidentified holes in asset protection, overlapping and contradictory controls, and policy and regulation noncompliance. It is the goal of IdM technologies to simplify the administration of these tasks and bring order to chaos.

The following are some of the common questions enterprises deal with regarding IdM implementation:

- What should each user have access to?
- Who approves and allows access?
- How do the access decisions map to policies?
- Do former employees still have access?
- How do we keep up with our dynamic and ever-changing environment?

- What is the process of revoking access?
- How is access controlled and monitored centrally?
- Why do employees have eight passwords to remember?
- We have five different operating platforms. How do we centralize access when each platform (and application) requires its own type of credential set?
- How do we control access for our employees, customers, and partners?
- How do we make sure we are compliant with the necessary regulations?

The traditional identity management process has been manual, using directory services with permissions, access control lists (ACLs), and profiles. This labor-intensive approach has proven incapable of keeping up with complex demands and thus has been replaced with automated applications rich in functionality that work together to create an IdM infrastructure. The main goal of IdM technologies is to streamline the management of identity, authentication, authorization, and auditing of subjects on multiple systems throughout the enterprise. The sheer diversity of a heterogeneous enterprise makes proper implementation of IdM a huge undertaking.

## Directory Services

*Directory services*, much like DNS, map resource names to their corresponding network addresses, allowing discovery of and communication with devices, files, users, or any other asset. Network directory services provide users access to network resources transparently, meaning that users don't need to know the exact location of the resources or the steps required to access them. The network directory services handle these issues for the user in the background.

Most organizations have some type of directory service that contains information pertaining to the organization's network resources and users. Most directories follow a hierarchical database format, originally established by the ITU X.500 standard but now most commonly implemented with the Lightweight Directory Access Protocol (LDAP), that allows subjects and applications to interact with the directory. Applications can request information about a particular user by making an LDAP request to the directory, and users can request information about a specific resource by using a similar request.

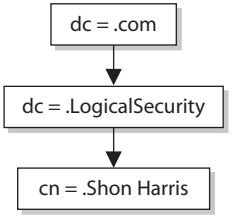
The objects within the directory are managed by a directory service. The directory service allows an administrator to configure and manage how identification, authentication, authorization, and access control take place within the network and on individual systems. The objects within the directory are labeled and identified with namespaces.

In a Windows Active Directory (AD) environment, when you log in, you are logging into a domain controller (DC), which has a hierarchical LDAP directory in its database. The database organizes the network resources and carries out user access control functionality. So once you successfully authenticate to the DC, certain network resources are available to you (print service, file server, e-mail server, and so on) as dictated by the configuration of AD.

How does the directory service keep all of these entities organized? By using *namespaces*. Each directory service has a way of identifying and naming the objects they manage. In LDAP, the directory service assigns distinguished names (DNs) to each object. Each DN

represents a collection of attributes about a specific object and is stored in the directory as an entry. In the following example, the DN is made up of a common name (cn) and domain components (dc). Since this is a hierarchical directory, .com is the top, LogicalSecurity is one step down from .com, and Shon is at the bottom.

```
dn: cn=Shon Harris,dc=LogicalSecurity,dc=com
cn: Shon Harris
```




This is a very simplistic example. Companies usually have large trees (directories) containing many levels and objects to represent different departments, roles, users, and resources.

A directory service manages the entries and data in the directory and also enforces the configured security policy by carrying out access control and identity management functions. For example, when you log into the DC, the directory service determines which resources you can and cannot access on the network.

## Directories' Role in Identity Management

A directory service is a general-purpose resource that can be used for IdM. When used in this manner it is optimized for reading and searching operations and becomes the central component of an IdM solution. This is because all resource information, users' attributes, authorization profiles, roles, access control policies, and more are stored in this one location. When other IdM features need to carry out their functions (authorization, access control, assigning permissions), they now have a centralized location for all of the information they need.

A lot of the information that is catalogued in an IdM directory is scattered throughout the enterprise. User attribute information (employee status, job description, department, and so on) is usually stored in the HR database, authentication information could be in a Kerberos server, role and group identification information might be in a SQL database, and resource-oriented authentication information may be stored in Active Directory on a domain controller. These are commonly referred to as *identity stores* and are located in different places on the network.

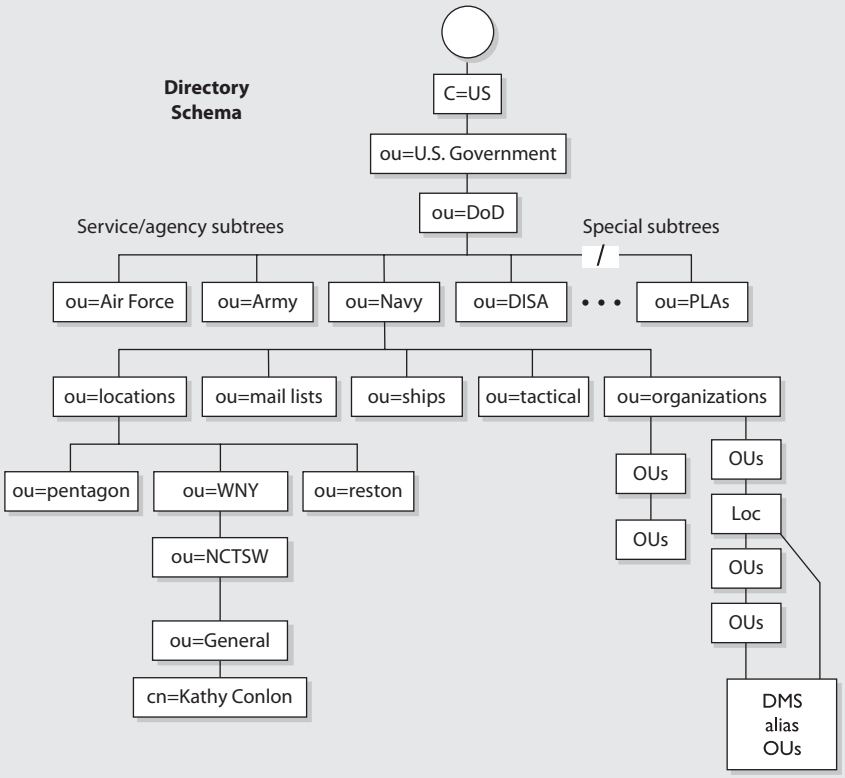
Something nifty that many IdM products do is create meta-directories or virtual directories. A *meta-directory* gathers the necessary information from multiple sources and stores it in one central directory. This provides a unified view of all users' digital identity information throughout the enterprise. The meta-directory synchronizes itself with all of the identity stores periodically to ensure the most up-to-date information is being used by all applications and IdM components within the enterprise.

## Organizing All of This Stuff

In an LDAP system, the following rules are used for object organization:

- The directory has a tree structure to organize the entries using a parent-child configuration.
- Each entry has a unique name made up of attributes of a specific object.
- The attributes used in the directory are dictated by the defined schema.
- The unique identifiers are called distinguished names.

The schema describes the directory structure and what names can be used within the directory, among other things. The following diagram shows how an object (Kathy Conlon) can have the attributes of ou=General, ou=NCTSW, ou=WNY, ou=locations, ou=Navy, ou=DoD, ou=U.S. Government, and C=US. Kathy's distinguished name is made up by listing all of the nodes starting at the root of the tree (C=US) all the way to her leaf node (cn=Kathy Conlon), separated by commas.




Note that OU stands for organizational unit. OUs are used as containers of other similar OUs, users, and resources. CN stands for common name.

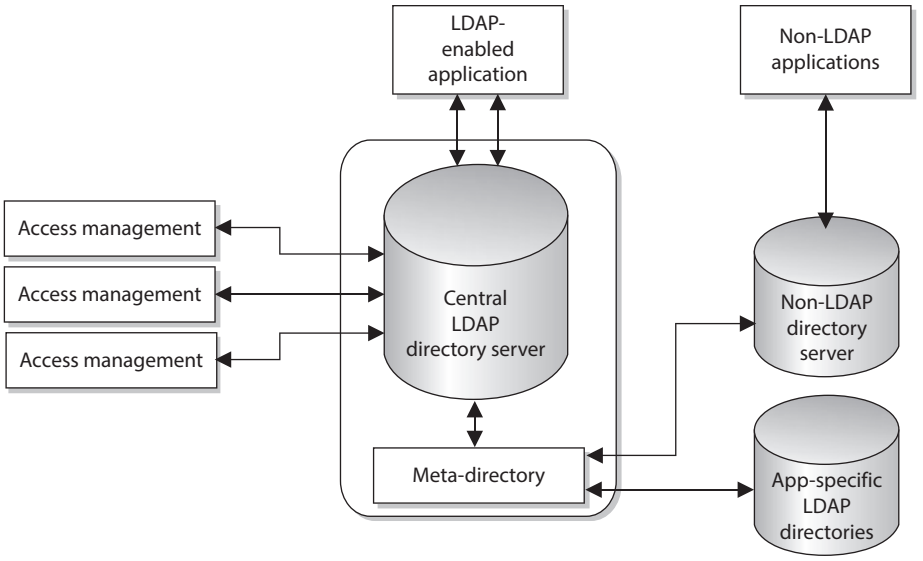



**Figure 16-5** Meta-directories pull data from other sources to populate the IdM directory.

A *virtual directory* plays the same role and can be used instead of a meta-directory. The difference between the two is that the meta-directory physically has the identity data in its directory, whereas a virtual directory does not and points to where the actual data resides. When an IdM component makes a call to a virtual directory to gather identity information on a user, the virtual directory points to where the information actually lives.

Figure 16-5 illustrates a central LDAP directory that is used by the IdM services: access management, provisioning, and identity management. When one of these services accepts a request from a user or application, it pulls the necessary data from the directory to be able to fulfill the request. Since the data needed to properly fulfill these requests is stored in different locations, the metadata directory pulls the data from these other sources and updates the LDAP directory.

## Single Sign-On

Employees typically need to access many different computers, servers, databases, and other resources in the course of a day to complete their tasks. This often requires the employees to remember multiple user IDs and passwords for these different computers. In a utopia, a user would need to enter only one user ID and one password to be able to access all resources in all the networks this user is working in. In the real world, this is hard to accomplish for all system types.

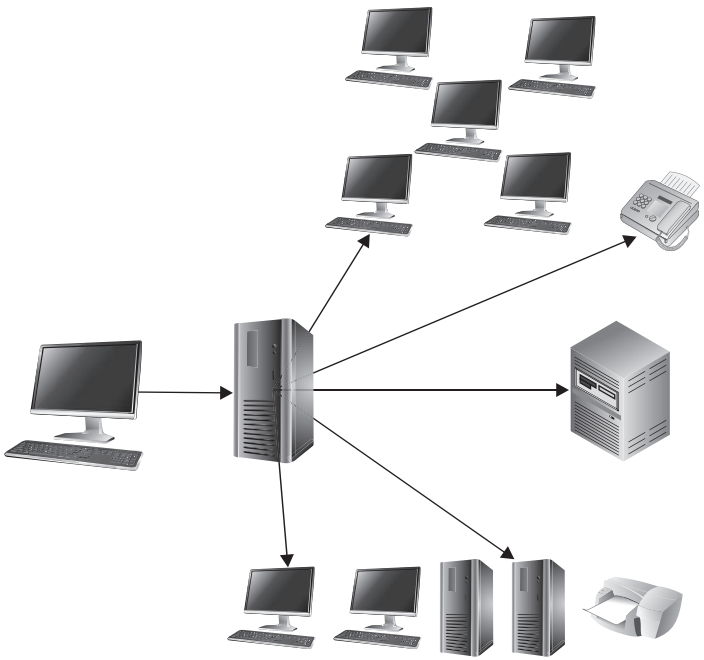
Because of the proliferation of client/server technologies, networks have migrated from centrally controlled networks to heterogeneous, distributed environments. The propagation of open systems and the increased diversity of applications, platforms, and operating systems have caused the end user to have to remember several user IDs and passwords just to be able to access and use the different resources within his own network. Although the different IDs and passwords are supposed to provide a greater level of

security, they often end up compromising security (because users write them down) and causing more effort and overhead for the staff that manages and maintains the network.

As any network staff member or administrator can attest to, too much time is devoted to resetting passwords for users who have forgotten them. More than one employee's productivity is affected when forgotten passwords have to be reassigned. The network staff member who has to reset the password could be working on other tasks, and the user who forgot the password cannot complete his task until the network staff member is finished resetting the password. Depending on the enterprise, between 20 percent and 50 percent of all IT help-desk calls are for password resets, according to the Gartner Group. Forrester Research estimates that each of these calls costs $70 in the United States. System administrators have to manage multiple user accounts on different platforms, which all need to be coordinated in a manner that maintains the integrity of the security policy. At times the complexity can be overwhelming, which results in poor access control management and the generation of many security vulnerabilities. A lot of time is spent on multiple passwords, and in the end they do not provide us with more security.

The increased cost of managing a diverse environment, security concerns, and user habits, coupled with the users' overwhelming desire to remember one set of credentials, has brought about the idea of *single sign-on (SSO)* capabilities. These capabilities would allow a user to enter credentials one time and be able to access all resources in primary and secondary network domains. This reduces the amount of time users spend authenticating to resources and enables the administrator to streamline user accounts and better control access rights. It improves security by reducing the probability that users will write down passwords and also reduces the administrator's time spent on adding and removing user accounts and modifying access permissions. If an administrator needs to disable or suspend a specific account, she can do it uniformly instead of having to alter configurations on each and every platform.





Single sign-on technology enables a user to enter credentials one time to be able to access all preauthorized resources within the domain.