Controlling the Crime Scene

Whether the crime scene is physical or digital, it is important to control who comes in contact with the evidence of the crime to ensure its integrity. The following are just some of the steps that should take place to protect the crime scene:

- Only allow authorized individuals access to the scene. These individuals should have knowledge of basic crime scene analysis.
- Document who is at the crime scene. In court, the integrity of the evidence may be in question if too many people were milling around the crime scene.
- Document who were the last individuals to interact with the systems.
- If the crime scene does become contaminated, document it. The contamination may not negate the derived evidence, but it will make investigating the crime more challenging.

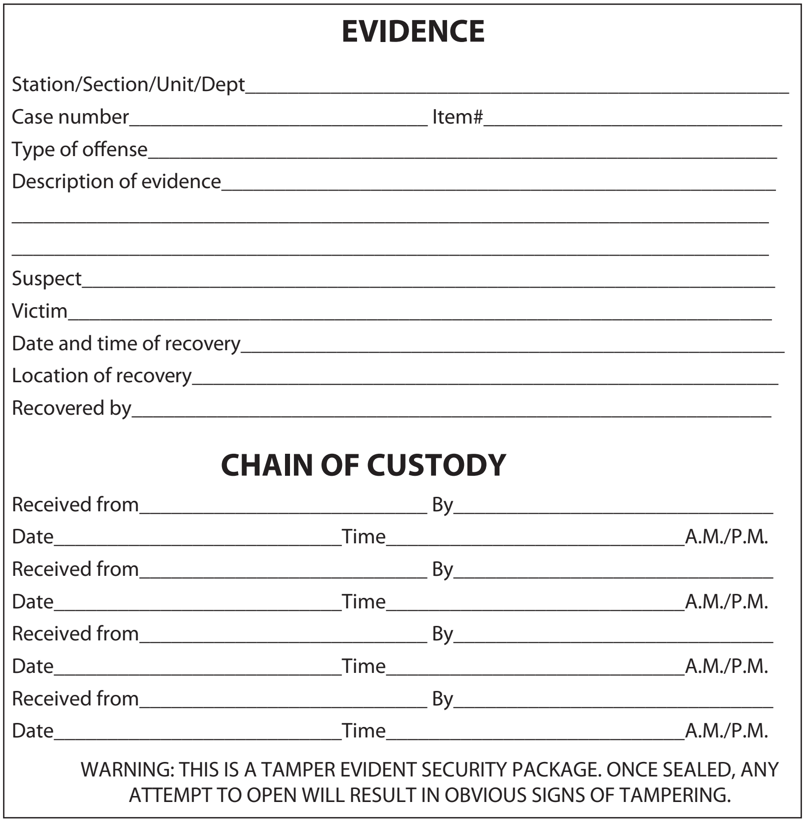
electronics. Thoroughness in identifying evidence is the most important consideration in this phase, and this may require you to think outside the box to ensure you don't miss or lose a critical evidentiary item.

Collection

Once you've identified the evidence you need, you can begin collecting it. Evidence collection is the process of gaining physical control over items that could potentially have evidentiary value. This is where you walk into someone's office and collect their computer, external hard drives, thumb drives, and so on. It is critical that you have the legal authority to do this and that you document what you take, where you take it from, and what its condition is at the time.

Each piece of evidence should be labeled in some way with the date, time, initials of the collector, and a case number if one has been assigned. The piece of evidence should then be placed in a container, which should be sealed (ideally with evidence tape) so that tampering can be detected. An example of the data that should be collected and displayed on each evidence container is shown in Figure 22-1.

After everything is properly labeled, a chain of custody log should be made for each container and an overall log should be made capturing all events. A *chain of custody* documents each person that has control of the evidence at every point in time. In large investigations, one person may collect evidence, another may transport it, and a third may store it. Keeping track of all these individuals' possession of the evidence is critical to proving in court that the evidence was not tampered with. It is not hard for a good defense attorney to get evidence dismissed from court because of improper handling. For this reason, the chain of custody should follow evidence through its entire life cycle, beginning with identification and ending with its destruction, permanent archiving, or return to owner.



EVIDENCE

Station/Section/Unit/Dept______________________

Case number______________ Item#______________

Type of offense______________________

Description of evidence______________________

____

____

Suspect______________________

Victim______________________

Date and time of recovery______________________

Location of recovery______________________

Recovered by______________________

CHAIN OF CUSTODY

Received from______________ By______________

Date______________Time______________A.M./P.M.

Received from______________ By______________

Date______________Time______________A.M./P.M.

Received from______________ By______________

Date______________Time______________A.M./P.M.

Received from______________ By______________

Date______________Time______________A.M./P.M.

WARNING: THIS IS A TAMPER EVIDENT SECURITY PACKAGE. ONCE SEALED, ANY ATTEMPT TO OPEN WILL RESULT IN OBVIOUS SIGNS OF TAMPERING.

Figure 22-1 Evidence container data

Evidence collection activities can get tricky depending on what is being searched for and where. For example, American citizens are protected by the Fourth Amendment against unlawful search and seizure, so law enforcement agencies must have probable cause and request a search warrant from a judge or court before conducting such a search. The actual search can take place only in the areas outlined by the warrant. The Fourth Amendment does not apply to actions by private citizens unless they are acting as police agents. So, for example, if Kristy's boss warned all employees that the management could remove files from their computers at any time, and her boss is not a police officer or acting as a police agent, she could not successfully claim that her Fourth Amendment rights were violated. Kristy's boss may have violated some specific privacy laws, but he did not violate Kristy's Fourth Amendment rights.

In some circumstances, a law enforcement agent is legally permitted to seize evidence that is not included in the search warrant, such as if the suspect tries to destroy the evidence. In other words, if there is an impending possibility that evidence might be destroyed, law enforcement may quickly seize the evidence to prevent its destruction.

This is referred to as *exigent circumstances*, and a judge will later decide whether the seizure was proper and legal before allowing the evidence to be admitted. For example, if a police officer had a search warrant that allowed him to search a suspect's living room but no other rooms and then he saw the suspect putting a removable drive in his pocket while standing in another room, the police officer could seize the drive even though it was outside the area covered under the search warrant.

EXAM TIP Always treat an investigation, regardless of type, as if it would ultimately end up in a courtroom.

Acquisition

In most corporate investigations involving digital evidence, the sort of Crime TV collection we just described will not take place unless law enforcement is involved. Instead, the IR team will probably be able to piece together a timeline of activities from various network resources and you may have to collect only a single laptop. In many cases you can probably acquire the evidence you need remotely without seizing any devices at all. Whatever the case, you ultimately need to get a hold of the data that will confirm or deny the claim that is being investigated, and you must do it in a forensically sound manner.

Acquisition means creating a forensic image of digital data for examination. Generally, speaking, there are two types of acquisition: physical and logical. In *digital acquisition*, the investigator makes a bit-for-bit copy of the contents of a physical storage device, bypassing the operating system. This includes all files, of course, but also free space and previously deleted data. In *logical acquisition*, on the other hand, the forensic image is of the files and folders in a file system, which means we rely on the operating system. This approach is sometimes necessary when dealing with evidence that exists in cloud services, where physical acquisition is normally not possible.

Before creating a forensic image, the investigator must have a medium onto which to copy the data, and ensure this medium has been properly purged, meaning it does not contain any preexisting data. (In some cases, hard drives that were thought to be new and right out of the box contained old data not purged by the vendor.) Two copies are normally created: a *primary image* (a control copy that is stored in a library) and a *working image* (used for analysis and evidence collection). To ensure that the original image is not modified, it is important to compute the cryptographic hashes (e.g., SHA-1) for files and directories before and after the analysis to prove the integrity of the original image.

The investigator works from the duplicate image because it preserves the original evidence, prevents inadvertent alteration of original evidence during examination, and allows re-creation of the duplicate image if necessary.

Acquiring evidence on live systems and those using network storage further complicates matters because you cannot turn off the system to make a copy of the hard drive. Imagine the reaction you'd receive if you were to tell an IT manager that you need to shut down a primary database or e-mail system. It wouldn't be favorable. So these systems and others, such as those using on-the-fly encryption, must be imaged while they are running.

In fact, some evidence is very volatile and can only be collected from a live system. Examples of volatile data that could have evidentiary value include

- Registers and cache
- Process tables and ARP cache
- System memory (RAM)
- Temporary file systems
- Special disk sectors

Preservation

To preserve evidence in a forensically sound manner, you must have established procedures based on legally accepted best practices, and your staff must follow those procedures to the letter. We've already covered two crucial steps in the chain of evidence and the use of hashes to verify that the evidence has not been altered. Another element of preserving digital evidence is ensuring that only a small group of qualified individuals have access to the evidence, and then only to perform specific functions. Again, this access needs to be part of your established procedures. In some cases, organizations implement two-person control of digital evidence to minimize the risk of tampering.

We introduced the topic of evidence storage in Chapter 10, but it bears pointing out that storage of media evidence should be dust-free and kept at room temperature without much humidity, and, of course, the media should not be stored close to any strong magnets or magnetic fields. Even if you don't have a dedicated evidence storage area, you should ensure that whatever space you commandeer is used strictly for this purpose, at least for the life of the investigation.

What Is Admissible in Court?

There are limits to what evidence can be introduced into a legal proceeding. Though the details will be different in each jurisdiction around the world, generally, digital evidence is admissible in court if it meets three criteria:

- **Relevance** Evidence must be relevant to the case, meaning it must help to prove facts being alleged. If a suspect is accused of murder, then a web search history for favorite vacationing spots is probably irrelevant. Judges typically rule on relevance of evidence.
- **Reliability** Evidence must be acquired using a sound forensic methodology that prevents alteration and ensures the evidence remains unaltered during the forensic examination. Multiple high-profile cases in recent years have had evidence rendered inadmissible because the chain of custody was broken.
- **Legality** The persons acquiring and presenting the evidence must have the legal authority to do so. If you have a court-issued search warrant, you must limit collection to whatever is spelled out in it. If you are conducting a workplace investigation, you must limit your collection to organization-owned assets, and only after legal counsel agrees.

The reliability of evidence is most often established by chains of custody and cryptographic hashing. But there is another element to reliability that excludes evidence deemed to be hearsay. *Hearsay evidence* is any statement made outside of the court proceeding that is offered into evidence to prove the truth of the matter asserted in the statement. Suppose that David is accused of fraud and Eliza tells Frank that David told her he was stealing from the company. Eliza's testimony in court would be admissible, but Frank normally wouldn't be allowed to testify about what Eliza claims to have heard because, coming from him, it would be considered hearsay.

Hearsay evidence can also include many computer-generated documents such as log files. In some countries, such as the United States, when computer logs are to be used as evidence in court, they must satisfy a legal exception to the hearsay rule of the Federal Rules of Evidence (FRE) called the business records exception rule or business entry rule. Under this rule, a party could admit any records of a business (1) that were made in the regular course of business; (2) that the business has a regular practice to make such records; (3) that were made at or near the time of the recorded event; and (4) that contain information transmitted by a person with knowledge of the information within the document.

It is important to show that the logs, and all evidence, have not been tampered with in any way, which is the reason for the chain of custody of evidence. Several tools are available that run checksums or hashing functions on the logs, which will allow the team to be alerted if something has been modified.

When evidence is being collected, one issue that can come up is the user's expectation of privacy. If an employee is suspected of, and charged with, a computer crime, he might claim that his files on the computer he uses are personal and not available to law enforcement and the courts. This is why it is important for organizations to conduct security awareness training, have employees sign documentation pertaining to the acceptable use of the organization's computers and equipment, and have legal banners pop up on every employee's computer when they log on. These are key elements in establishing that a user has no right to privacy when he is using organization equipment. The following banner is suggested by CERT Advisory:

> *This system is for the use of authorized users only. Individuals using this computer system without authority, or in excess of their authority, are subject to having all of their activities on this system monitored and recorded by system personnel.*
>
> *In the course of monitoring an individual improperly using this system, or in the course of system maintenance, the activities of authorized users may also be monitored.*
>
> *Anyone using this system expressly consents to such monitoring and is advised that if such monitoring reveals possible evidence of criminal activity, system personnel may provide the evidence of such monitoring to law enforcement officials.*

This explicit warning strengthens a legal case that can be brought against an employee or intruder, because the continued use of the system after viewing this type of warning implies that the person acknowledges the security policy and gives permission to be monitored.

NOTE Don't dismiss the possibility that as an information security professional you will be responsible for entering evidence into court. Most tribunals, commissions, and other quasi-legal proceedings have admissibility requirements. Because these requirements can change between jurisdictions, you should seek legal counsel to better understand the specific rules for your jurisdiction.

Digital Forensics Tools, Tactics, and Procedures

Digital forensics is a science and an art that requires specialized techniques for the recovery, authentication, and analysis of electronic data for the purposes of a digital criminal investigation. It is a fusion of computer science, IT, engineering, and law. When discussing computer forensics with others, you might hear the terms computer forensics, network forensics, electronic data discovery, cyberforensics, and forensic computing.

Forensics Field Kits

When a forensics team is deployed, the forensic investigators should be properly equipped with all the tools and supplies that they'll need to conduct the investigation. The following are some of the common items in forensics field kits:

- **Documentation tools** Tags, labels, forms, and written procedures
- **Disassembly and removal tools** Antistatic bands, pliers, tweezers, screwdrivers, wire cutters, and so on
- **Package and transport supplies** Antistatic bags, evidence bags and tape, cable ties, and others
- **Cables and adapters** Enough to connect to every physical interface you may come across

(ISC)^2 uses *digital forensics* as a synonym for all of these other terms, so that's what you'll see on the CISSP exam.

Anyone who conducts a forensic investigation must be properly skilled in this trade and know what to look for. If someone reboots the attacked system or inspects various files, this could corrupt viable evidence, change timestamps on key files, and erase footprints the criminal may have left. Most digital evidence has a short lifespan and must be collected quickly and in the *order of volatility*. In other words, the most volatile or fragile evidence should be collected first. In some situations, it is best to remove the system from the network, dump the contents of the memory, power down the system, and make a sound image of the attacked system and perform forensic analysis on this copy. Working on the copy instead of the original drive ensures that the evidence stays unharmed on the original system in case some steps in the investigation actually corrupt or destroy data. Dumping the memory contents to a file before doing any work on the system or powering it down is a crucial step because of the information that could be stored there. This is another method of capturing fragile information. However, this creates a sticky situation: capturing RAM or conducting live analysis can introduce changes to the crime scene because various state changes and operations take place. Whatever method the forensic investigator chooses to use to collect digital evidence, that method must be documented. This is the most important aspect of evidence handling.

Forensic Investigation Techniques

To ensure that forensic investigations are carried out in a standardized manner and the evidence collected is admissible, it is necessary for the investigative team to follow specific laid-out steps so that nothing is missed. Figure 22-2 illustrates the phases through a common investigation process and lists various techniques that fall under each phase. Each team or organization may come up with its own steps, but all should be essentially accomplishing the same things:

- Identification
- Preservation
- Collection
- Examination
- Analysis
- Presentation
- Decision

NOTE The principles of criminalistics are included in the forensic investigation process. They are identification of the crime scene, protection of the environment against contamination and loss of evidence, identification of evidence and potential sources of evidence, and the collection of evidence. In regard to minimizing the degree of contamination, it is important to understand that it is impossible not to change a crime scene—be it physical or digital. The key is to minimize changes and document what you did and why, and how the crime scene was affected.

Identification	Preservation	Collection	Examination	Analysis	Presentation
Event/crime detection	Case management	Preservation	Preservation	Preservation	Documentation
Resolve signature	Imaging technologies	Approved methods	Traceability	Traceability	Expert testimony
Profile detection	Chain of custody	Approved software	Validation techniques	Statistical	Clarification
Anomalous detection	Time synchronization	Approved hardware	Filtering techniques	Protocols	Mission impact statement
Complaints		Legal authority	Pattern matching	Data mining	Recommended countermeasure
System monitoring		Lossless compression	Hidden data discovery	Timeline	Statistical interpretation
Audit analysis		Sampling	Hidden data extraction	Link	
		Data reduction		Spatial	
		Recovery techniques			

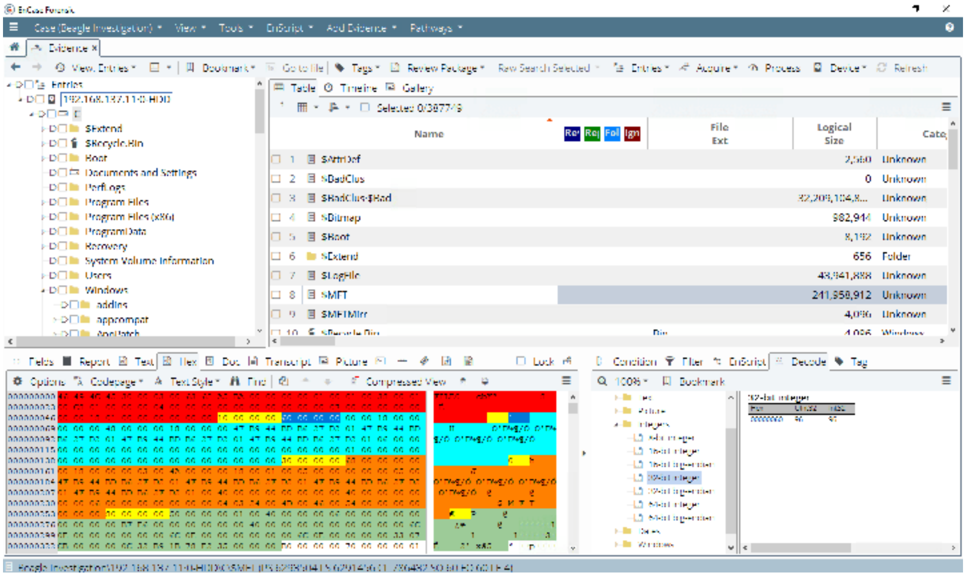
Figure 22-2 Characteristics of the different phases through an investigation process

During the examination and analysis process of a forensic investigation, it is critical that the investigator work from an image that contains *all* of the data from the original disk. It should be a bit-level copy, sector by sector, to capture deleted files, slack spaces, and unallocated clusters. These types of images can be created through the use of a specialized tool such as Forensic Toolkit (FTK), EnCase Forensic, or the dd Unix utility. A file copy tool does not recover all data areas of the device necessary for examination. Figure 22-3 illustrates a commonly used tool in the forensic world for evidence collection.

The next step is the analysis of the evidence. Forensic investigators use a scientific method that involves

- Determining the characteristics of the evidence, such as whether it's admissible as primary or secondary evidence, as well as its source, reliability, and permanence
- Comparing evidence from different sources to determine a chronology of events
- Event reconstruction, including the recovery of deleted files and other activity on the system

This can take place in a controlled lab environment or, thanks to hardware write-blockers and forensic software, in the field. When investigators analyze evidence in a lab, they are dealing with "dead forensics"; that is, they are working only with static data. Live forensics, which takes place in the field, includes volatile data. If evidence is lacking, then an experienced investigator should be called in to help complete the picture.



Entries
192.168.137.11-0 HDD
$Extend
$Recycle.Bin
Boot
Documents and Settings
PerfLogs
Program Files
Program Files (x86)
ProgramData
Recovery
System Volume Information
Users
Windows
addins
appcompat
AppPatch

Selected 0/387749

	Name	File Ext	Logical Size	Cate
1	$AttrDef		2,560	Unknown
2	$BadClus		0	Unknown
3	$BadClus·$Bad		32,209,104,8...	Unknown
4	$Bitmap		982,914	Unknown
5	$Boot		8,192	Unknown
6	$Extend		656	Folder
7	$LogFile		43,941,888	Unknown
8	$MFT		241,958,912	Unknown
9	$MFTMirr		4,096	Unknown
10	$Recycle.Bin	Bin	4,096	Windows

Figure 22-3 EnCase Forensic can be used to collect digital forensic data.

Finally, the interpretation of the analysis should be presented to the appropriate party. This could be a judge, lawyer, CEO, or board of directors. Therefore, it is important to present the findings in a format that will be understood by a nontechnical audience. As a CISSP, you should be able to explain these findings in layperson's terms using metaphors and analogies. Of course, the findings, which are top secret or company confidential, should be disclosed only to authorized parties. This may include the legal department or any outside counsel that assists with the investigation.

Other Investigative Techniques

Unless you work for a law enforcement agency, most of the investigations in which you will be involved are likely to focus on digital forensics investigative techniques. These techniques are applied when a device was compromised, or a malicious insider attempted to steal sensitive files, or something like that. All the evidence you need is probably in a device that you can get your hands on, so you can collect it, acquire it, analyze it, and get to the facts with just digital evidence. However, there may be other situations in which you'll need other types of evidence either in addition to or instead of 1's and 0's copied from some storage device. Interviews, surveillance, and undercover investigative techniques are some of the practices for acquiring evidence that you should be familiar with.

Interviews

Interviews can be effective for ascertaining facts when you have willing interviewees. Interviewing is both an art and a science, and the specific techniques you use will vary from case to case. Typically, interviews are conducted by a business unit manager with assistance from the human resources and legal departments. This doesn't, however, completely relieve you as an information security professional from responsibility during the interviewing process. You may be asked to provide input or observe an interview in order to clarify technical information that comes up in the course of questioning.

Whether you are conducting an interview or your technical assistance is needed for an interview, keep the following best practices in mind:

- *Have a plan.* Without a plan, the interview will be ineffective. Prepare an outline beforehand that focuses on getting the information you need from each interviewee. However, you should remain flexible and not read off a script.
- *Be fair and objective.* If you are conducting an interview, it is to get to the facts of an incident, not necessarily to reinforce whatever conclusions you may have already reached. Keep an open mind, focus on the facts, and try to avoid any biases.
- *Compartmentalize information.* Your interview plan should address what information you share with each interviewee, and what you don't share. You should not tell one interviewee what another said unless it's absolutely essential and legally permissible.
- *One interviewee at a time.* Interviewing multiple individuals together can introduce problematic group dynamics such as peer pressure. It can also lead interviewees to distort or suppress information.
- *Do not record the interview.* Recording devices can have a chilling effect on interviewees. Instead, have at least one notetaker in the room and, after the interview is complete, read back the notes to the interviewee to ensure their accuracy. If you must record the interview, ensure you comply with all applicable legal requirements (e.g., consent of all parties).
- *Keep it confidential.* Do your best to keep every aspect of the investigation under wraps. Even the fact that someone is being interviewed about an incident can have a damaging reputational effect for that person.

The employee interviewer should be in a position that is senior to the employee subject. A vice president is not going to be very intimidated or willing to spill his guts to the mailroom clerk. The interview should be held in a private place, in an environment conducive to making the subject relatively comfortable and at ease. If exhibits are going to be shown to the subject, they should be shown one at a time, and otherwise kept in a folder. It is not necessary to read a person their rights before the interview unless it is performed by law enforcement officers.

Surveillance

Two main types of surveillance are used when it comes to identifying computer crimes: physical surveillance and computer surveillance. *Physical surveillance* pertains to security cameras, security guards, and closed-circuit TV (CCTV), which may capture evidence.

Physical surveillance can also be used by an undercover agent to learn about the suspect's spending activities, family and friends, and personal habits in the hope of gathering more clues for the case.

Computer surveillance pertains to passively monitoring (auditing) events by using network sniffers, keyboard monitors, wiretaps, and line monitoring. In most jurisdictions, active monitoring may require a search warrant. In most workplace environments, to legally monitor an individual, the person must be warned ahead of time that her activities may be subject to this type of monitoring.

Undercover

Undercover investigative techniques are pretty rare in most corporate investigations, but can provide information and evidence that would be difficult to acquire otherwise. The goal of undercover work is to assume an identity that allows the investigator to blend into the suspect's environment to observe, and perhaps record, the suspect's actions.

A thin line exists between enticement and entrapment when it comes to capturing a suspect's actions. *Enticement* is legal and ethical, whereas *entrapment* is neither legal nor ethical. In the world of computer crimes, a honeypot is a good example to explain the difference between enticement and entrapment. Organizations put systems in their screened subnets that either emulate services that attackers usually like to take advantage of or actually have the services enabled. The hope is that if an attacker breaks into the organization's network, she will go right to the honeypot instead of the systems that are actual production machines. The attacker will be *enticed* to go to the honeypot system because it has many open ports and services running and exhibits vulnerabilities that the attacker would want to exploit. The organization can log the attacker's actions and later attempt to prosecute.

The action in the preceding example is legal unless the organization crosses the line to entrapment. For example, suppose a web page has a link that indicates that if an individual clicks it, she could then download thousands of MP3 files for free. However, when she clicks that link, she is taken to the honeypot system instead, and the organization records all of her actions and attempts to prosecute. Entrapment does not prove that the suspect had the intent to commit a crime; it only proves she was successfully tricked.

Forensic Artifacts

One of the grandfathers of forensic science, Dr. Edmond Locard, famously stated that "every contact leaves a trace." This principle, known as Locard's exchange principle, states that criminals always leave something behind at the crime scene. This fragmentary or trace evidence is a *forensic artifact*. A forensic artifact is anything that has evidentiary value. On a typical computer, the following are examples of forensic artifacts:

- Deleted items (in the recycle bin or trash)
- Web browser search history
- Web browser cache files
- E-mail attachments

- Skype history
- Windows event logs
- Prefetch files

Forensic artifacts can also be evidentiary items relating to network traffic. Network forensics is a subdiscipline that is focused on what happened on the network rather than on the endpoints. The tools used in network forensics are unique to that subdiscipline, and so are the artifacts for which the investigator looks. Tools used in network forensics include NDR solutions, SIEM systems, and the log files of any network device or server. They also include network sniffers that can capture full network frames. The following are some of the more useful network artifacts an investigator would be interested in:

- DNS log records
- Web proxy log records
- IDS/IPS alerts
- Packet capture (pcap) files

Finally, with the proliferation of mobile devices such as smartphones, tablets, and smartwatches, we must not overlook forensic artifacts stored on them. Unlike traditional computers, mobile devices are usually carried by their users around the clock. This means mobile devices tend to document multiple aspects of a person's life, some of which can serve as evidence of criminal activity.

Though mobile devices can be a treasure trove of information for the forensic investigator, they are not always easy to acquire and analyze. For starters, there are so many different models that no single tool can acquire all evidence from all devices. Staff expertise is similarly challenged by this diversity, because an investigator who is skilled at iPhone analysis may not be able to operate at the same level given an Android device. Just to make things more interesting, there is also the issue of encryption, which is prevalent in mobile devices these days.

Still, if forensic investigators can overcome these challenges, mobile devices are excellent sources of evidence for a variety of criminal activity. Among the most useful forensic artifacts found in them are

- Call logs
- SMS messages
- E-mail messages
- Web browser history

Reporting and Documenting

We already covered reporting in a fair amount of detail in Chapter 19. When it comes to investigations, however, there are some additional issues to consider. First and foremost, the need to document *everything* you do cannot be overstated. If you cannot account for or explain the why of any activities you undertook, it may render evidence inadmissible in court or even undermine the whole case. For this reason, many organizations assign investigators to work in teams of two, where one person documents while the other conducts the investigation. Most forensic analysis tools have a feature that automatically logs everything an investigator does with the tool.

Another issue that is particularly important in writing investigation reports is the need to remain completely logical and factual. Any conclusions you reach must follow logically from a sequence of facts that you spell out for the reader. For example, suppose that Carlos is one of your staff and is suspected of sending sensitive files to a competitor in hopes of landing a lucrative job with them. Even if you are sure he did it (after examining his computer), you should not just jump out and say so. Instead, you show how the forensic artifacts that you found, when arrayed on a timeline, substantiate the claim that Carlos sent sensitive files to a competitor. You'd start by establishing that he was logged into his computer, and then he logged into his personal e-mail account through a webmail interface, and then an e-mail was sent containing sensitive files x, y, and z, and then the e-mail was deleted from his sent items, and so on. It is ultimately up to the reader (presumably a senior manager or court official) to determine guilt or innocence. Your job is to establish the facts and determine whether or not they are consistent with the allegation.

Chapter Review

Incident management is a critical function for any organization. Odds are that if you are among the lucky few who haven't had a major incident yet, you will be faced with one in the near future. In fact, the IronNet 2021 Cybersecurity Impact Report found that 86 percent of respondents had a cybersecurity incident so severe in the previous year that it required a C-level or board meeting. Even if you've outsourced IR to a third-party service provider, you still need to have an incident management policy and an IR plan to guide the conduct of the entire organization before, during, and after an incident. The policy establishes authorities and responsibilities, while the plan specifies the procedures to be followed.

The other major topic we discussed in this chapter is investigations. Thankfully, the need to conduct investigations is fairly rare in most organizations. But therein lies the problem: if you hardly ever need to recall knowledge or practice skills, you are certain to lose them. This is why having detailed standard procedures for investigative work is absolutely essential. For example, evidence acquisition, as we saw, is a complex process that has very little room for errors, particularly if the evidence will end up in court (and we should always assume it will).

Quick Review

- A security event is any occurrence that can be observed, verified, and documented, whereas a security incident is one or more related events that negatively affect the organization and/or impact its security posture.

- A good incident response team should consist of representatives from various business units, such as the legal department, HR, executive management, the communications department, physical/corporate security, IS security, and information technology.
- Incident management encompasses seven phases according to the CISSP CBK: detection, response, mitigation, reporting, recovery, remediation, and lessons learned.
- The detection phase encompasses the search for indicators that an event has occurred and the formal declaration of the event.
- The response phase entails the initial actions undertaken to contain the damage caused by a security incident.
- The goal of the mitigation phase is to eradicate the threat actor from the affected systems.
- Incident reporting occurs at various phases of incident management.
- The aim of the recovery phase is to restore full, trustworthy functionality to the organization.
- In the remediation phase, the incident response team decides which security controls need to be deployed or changed to prevent the incident from recurring.
- The lessons learned phase is important to determine what needs to go into the incident response process and documentation, with the goal of continuous improvement.
- The incident management policy (IMP) establishes authorities and responsibilities across the entire organization, identifies the incident response (IR) lead for the organization, and describes what every staff member is required to do with regard to incidents.
- The incident response plan (IRP) gets into the details of what should be done when responding to suspected incidents, and includes roles and responsibilities, incident classification, notifications, and operational tasks.
- Incident classification criteria allow the organization to prioritize IR assets and usually consider the impact and type of the incident, and urgency with which the response must be started.
- A runbook is a collection of procedures that the IR team will follow for specific types of incidents.
- The four phases of evidence handling are identification, collection, acquisition, and preservation.
- Evidence collection is the process of gaining physical control over devices that could potentially have evidentiary value.
- A chain of custody documents each person that has control of the evidence at every point in time.
- Acquisition means creating a forensic image of digital data for examination.

- Evidence preservation requires maintaining a chain of custody and cryptographic hashes of all digital evidence, and also controlling access to the evidence.
- To be admissible in court, evidence must be relevant, reliable, and legally obtained.
- To be admissible in court, business records such as computer logs have to be made and collected in the normal course of business, not specially generated for a case in court. Business records can easily be deemed hearsay if there is no firsthand proof of their accuracy and reliability.
- Digital forensics is a science and an art that requires specialized techniques for the recovery, authentication, and analysis of electronic data for the purposes of a digital criminal investigation.
- In addition to forensic techniques, organizations sometimes use interviews, surveillance, and undercover investigation techniques.
- When looking for suspects, it is important to consider the motive, opportunity, and means (MOM).
- A forensic artifact is anything that has evidentiary value.

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. What are the phases of incident management?

A. Identification, collection, acquisition, and preservation

B. Detection, response, mitigation, reporting, recovery, remediation, and lessons learned

C. Protection, containment, response, remediation, and reporting

D. Analysis, classification, incident declaration, containment, eradication, and investigation

2. During which phase of incident management does the IR team contain the damage caused by a security incident?

A. Preservation

B. Response

C. Eradication

D. Remediation

3. During which phase of incident management are security controls deployed or changed to prevent the incident from recurring?

   A. Preservation

   B. Response

   C. Eradication

   D. Remediation

4. Which document establishes authorities and responsibilities with regard to incidents across the entire organization?

   A. Incident management policy

   B. Incident response plan

   C. Incident response runbook

   D. Incident classification criteria

5. After a computer forensic investigator seizes a computer during a crime investigation, what is the next step?

   A. Label and put it into a container, and then label the container

   B. Dust the evidence for fingerprints

   C. Make an image copy of the disks

   D. Lock the evidence in the safe

6. Which of the following is a necessary characteristic of evidence for it to be admissible?

   A. It must be real.

   B. It must be noteworthy.

   C. It must be reliable.

   D. It must be important.

7. Which of the following is *not* considered a best practice when interviewing willing witnesses?

   A. Compartmentalize information

   B. Interview one interviewee at a time

   C. Be fair and objective

   D. Record the interview

Use the following scenario to answer Questions 8–10. You recently improved your organization's security posture, which now includes a fully staffed security operations center (SOC), network detection and response (NDR) and endpoint detection and response (EDR) systems, centrally managed updates and data backups, and network segmentation using VLANs. It's the end of the workday and just as you are getting ready to go home your SOC detects a ransomware infection affecting at least two workstations in your marketing department. The SOC manager declares an incident and activates the IR team.

8. What should be your IR team's first action?

A. Determine the scope of the infection across the organization

B. Isolate the marketing VLAN from the rest of the network

C. Disconnect the infected computers from the network

D. Determine why the EDR system failed to protect the workstations

9. Using your NDR system, you determine the external hosts from which the malware was downloaded and with which the infected systems were communicating. As part of the remediation phase, which of the following is the next best action to take with this information?

A. Determine whether the external hosts you identified are related to the incident

B. Block traffic to/from the external hosts that you identified

C. Visit the remote hosts using a forensic workstation to acquire evidence

D. Share the address of the hosts with your partners as indicators of compromise (IOCs)

10. Luckily, this version of ransomware is buggy, and you find a security researcher's blog with detailed instructions for how to decrypt infected systems. Which of the following approaches will best mitigate the incident and make the affected systems operational again?

A. Follow the directions to decrypt the systems and remove the malware

B. Reinstall from a golden master and restore the data from backups

C. Reinstall from a golden master even though you have no backups

D. Restore the systems from the last known-good system backup

Answers

1. B. Incident management encompasses seven phases according to the CISSP CBK: detection, response, mitigation, reporting, recovery, remediation, and lessons learned.

2. B. The goal of containment during the response phase is to prevent or reduce any further damage from this incident so that you can begin to mitigate and recover. Done properly, this buys the IR team time for a proper investigation and determination of the incident's root cause.

3. D. In the remediation phase, you decide which control changes (e.g., firewall or IDS/IPS rules) are needed to preclude this incident from happening again. Another aspect of remediation is the identification of indicators of attack (IOAs) that can be used in the future to detect this attack in real time (i.e., as it is happening) as well as indicators of compromise (IOCs), which tell you when an attack has been successful and your security has been compromised.

4. **A.** The incident management policy (IMP) establishes authorities and responsibilities across the entire organization, identifies the incident response (IR) lead for the organization, and describes what every staff member is required to do with regard to incidents. The incident response plan (IRP) gets into the details of what should be done when responding to suspected incidents, and includes roles and responsibilities, incident classification, notifications, and operational tasks. A runbook is a collection of procedures that the IR team will follow for specific types of incidents.
5. **C.** Several steps need to be followed when gathering and extracting evidence from a scene. Once a computer has been confiscated, the first thing the computer forensics team should do is make an image of the hard drive. The team will work from this image instead of the original hard drive so that the original stays in a pristine state and the evidence on the drive is not accidentally corrupted or modified.
6. **C.** For evidence to be admissible, it must be relevant to the case, reliable, and legally obtained. For evidence to be reliable, it must be consistent with fact and must not be based on opinion or be circumstantial.
7. **D.** Recording devices can have a chilling effect on interviewees. Instead, have at least one notetaker in the room and, after the interview is complete, read back the notes to the interviewee to ensure their accuracy.
8. **B.** Having detected the incident, the next step is to respond by containing the damage that has been or is about to be done to your most critical assets. You could simply disconnect the infected systems from the network, but since there are multiple workstations and they are in the same department, it is probably better to isolate that entire VLAN until you can determine the true scope of the problem. Since this incident happened at the end of the workday, isolating the VLAN should have little or no impact on the marketing department.
9. **B.** In the remediation phase, you decide which security controls need to be put in place to prevent the attack from succeeding again. This includes controls that are hastily put into effect because you have high confidence that they will help in the short term. The situation in the question is a perfect example of when you bypass your change management process and quickly make changes to deal with the incident at hand. You probably want to share the IOCs with your partners (and perhaps your regional CERT), but that happens after you block the traffic.
10. **B.** You have a centralized backup system that was not affected, so you know you should have backups for all the workstations. The problem is that you may not know if any of the full-system backups also include the ransomware, so restoring systems from backups could bring you back to square one. It is best to reinstall the systems from golden masters and then restore only the data files. This process may take a bit longer, but it minimizes the risk of reinfection.

This page intentionally left blank

CHAPTER 23

Disasters

This chapter presents the following:

- Recovery strategies
- Disaster recovery processes
- Testing disaster recovery plans
- Business continuity

It wasn't raining when Noah built the ark.

—Howard Ruff

Disasters are just regular features in our collective lives. Odds are that, at some point, we will all have to deal with at least one disaster (if not more), whether it be in our personal world or professional world. And when that disaster hits, figuring out a way to deal with it in real time is probably not going to go all that well for the unprepared. This chapter is all about thinking of all the terrible things that might happen, and then ensuring we have strategies and plans to deal with them. This doesn't just mean recovering from the disaster, but also ensuring that the business continues to operate with as little disruption as possible.

As the old adage goes, no battle plan ever survived first contact with the enemy, which is the reason why we must test and exercise plans until our responses as individuals and organizations are so ingrained in our brains that we no longer need to think about them. As terrible and complex disasters unfold around us, we will do the right things reflexively. Does that sound a bit ambitious? Perhaps. Still, it is our duty as cybersecurity professionals to do what we can to get our organizations as close to that goal as realistically possible. Let's see how we go about doing this.

Recovery Strategies

In the previous chapters in this part of the book, we have discussed preventing and responding to security incidents, including various types of investigations, as part of standard security operations. These are things we do day in and day out. But what happens on those rare occasions when an incident has disastrous effects? That is the realm of disaster recovery and business continuity planning. *Disaster recovery (DR)* is the set of practices that enables an organization to minimize loss of, and restore, mission-critical technology infrastructure after a catastrophic incident. *Business continuity (BC)* is the set of practices that enables an organization to continue performing its critical functions through and after any disruptive event. As you can see, DR is mostly in the purview of safety and contingency operations, while BC is much broader than that. Accordingly, we'll focus on DR for most of this chapter but circle back to our roles in BC as cybersecurity leaders.

EXAM TIP As CISSPs, we are responsible for disaster recovery because it deals mostly with information technology and security. We provide inputs and support for business continuity planning but normally are not the lead for it.

Before we go much further, recall that we discussed the role of *maximum tolerable downtime (MTD)* values in Chapter 2. In reality, basic MTD values are a good start, but are not granular enough for an organization to figure out what it needs to put into place to be able to absorb the impact of a disaster. MTD values are usually "broad strokes" that do not provide the details needed to pinpoint the actual recovery solutions that need to be purchased and implemented. For example, if the business continuity planning (BCP) team determines that the MTD value for the customer service department is 48 hours, this is not enough information to fully understand what redundant solutions or backup technology should be put into place. MTD in this example does provide a basic deadline that means if customer service is not up within 48 hours, the company may not be able to recover and everyone should start looking for new jobs.

As shown in Figure 23-1, more than just MTD metrics are needed to get production back to normal operations after a disruptive event. We will walk through each of these metric types and see how they are best used together.

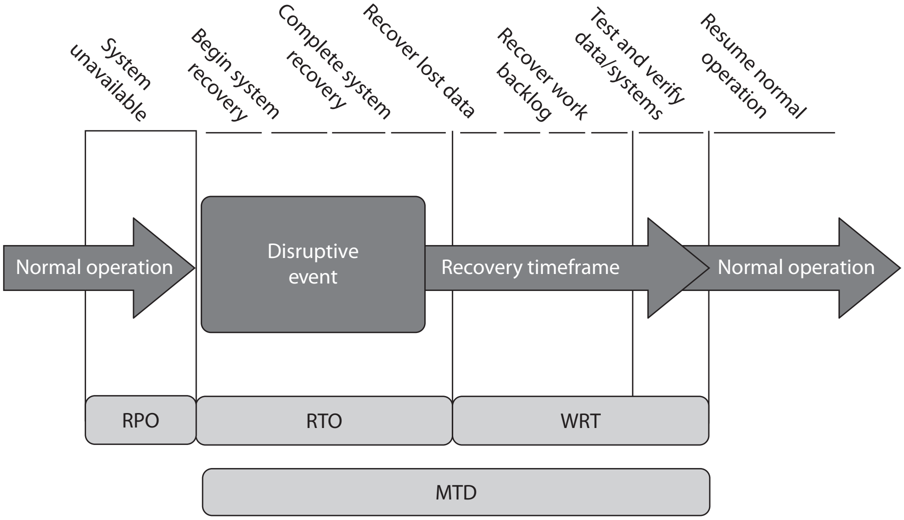


Figure 23-1 Metrics used for disaster recovery

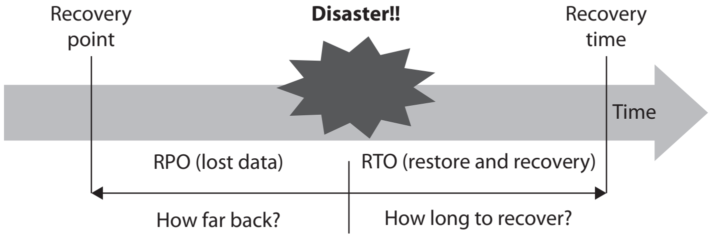
The *recovery time objective (RTO)* is the maximum time period within which a mission-critical system must be restored to a designated service level after a disruption to avoid unacceptable consequences associated with a break in business continuity. The RTO value is smaller than the MTD value, because the MTD value represents the time after which an inability to recover significant operations will mean severe and perhaps irreparable damage to the organization's reputation or bottom line. The RTO assumes that there is a period of acceptable downtime. This means that an organization can be out of production for a certain period of time and still get back on its feet. But if the organization cannot get production up and running within the MTD window, it may be sinking too fast to properly recover.

The *work recovery time (WRT)* is the maximum amount of time available for certifying the functionality and integrity of restored systems and data so they can be put back into production. RTO usually deals with getting the infrastructure and systems back up and running, and WRT deals with ensuring business users can get back to work using them. Another way to think of WRT is as the remainder of the overall MTD value after the RTO has passed.

The *recovery point objective (RPO)* is the acceptable amount of data loss measured in time. This value represents the earliest point in time at which data must be recovered. The higher the value of data, the more funds or other resources that can be put into place to ensure a smaller amount of data is lost in the event of a disaster. Figure 23-2 illustrates the relationship and differences between the use of RPO and RTO values.

The MTD, RTO, RPO, and WRT values are critical to understand because they will be the basic foundational measures used when determining the type of recovery solutions an organization must put into place, so let's dig a bit deeper into them. As an example of RTO, let's say a company has determined that if it is unable to process product order requests for 12 hours, the financial hit will be too large for it to survive. This means that the MTD for order processing is 12 hours. To keep things simple, let's say that RTO and WRT are 6 hours each. Now, suppose that orders are processed using on-premises servers, on a site with no backup power sources, and an ice storm causes a power outage that will take days to restore. Without a plan and supporting infrastructure already in place, it would be close to impossible to migrate the servers and data to a site with power within 6 hours. The RTO (that is, the maximum time to move the servers and data) would not be met (to say nothing of the WRT) and it would likely exceed the MTD, putting the company at serious risk of collapse.

Figure 23-2 RPO and RTO measures in use



Now let's say that the same company did have a recovery site on a different power grid, and it was able to restore the order-processing services within a couple of hours, so it met the RTO requirement. But just because the systems are back online, the company still might have a critical problem. The company has to restore the data it lost during the disaster. Restoring data that is a week old does the company no good. The employees need to have access to the data that was being processed right before the disaster hit. If the company can only restore data that is a week old, then all the orders that were in some stage of being fulfilled over the last seven days could be lost. If the company makes an average of $25,000 per day in orders and all the order data was lost for the last seven days, this can result in a loss of $175,000 and a lot of unhappy customers. So just getting things up and running (meeting the RTO) is just part of the picture. Getting the necessary data in place so that business processes are up to date and relevant (RPO) is just as critical.

To take things one step further, let's say the company stood up the systems at its recovery site in two hours. It also had real-time data backup systems in place, so all of the necessary up-to-date data is restored. But no one actually tested the processes to recover data from backups, everyone is confused, and orders still cannot be processed and revenue cannot be collected. This means the company met its RTO requirement and its RPO requirement, but failed its WRT requirement, and thus failed the MTD requirement. Proper business recovery means *all* of the individual things have to happen correctly for the overall goal to be successful.

EXAM TIP An RTO is the amount of time it takes to recover from a disaster, and an RPO is the acceptable amount of data, measured in time, that can be lost from that same event.

The actual MTD, RTO, and RPO values are derived during the *business impact analysis (BIA)*, the purpose of which is to be able to apply criticality values to specific business functions, resources, and data types. A simplistic example is shown in Table 23-1. The company must have data restoration capabilities in place to ensure that mission-critical data is never older than one minute. The company cannot rely on something as slow as backup tape restoration, but must have a high-availability data replication solution in place. The RTO value for mission-critical data processing is two minutes or less. This means that the technology that carries out the processing functionality for this type of data cannot be down for more than two minutes. The company probably needs failover technology in place that will shift the load once it notices that a server goes offline.

Data Type	RPO	RTO
Mission critical	Continuous to 1 minute	Instantaneous to 2 minutes
Business critical	5 minutes	10 minutes
Business	3 hours	8 hours

Table 23-1 RPO and RTO Value Relationships

What Is the Difference Between Preventive Measures and Recovery Strategies?

Preventive mechanisms are put into place not only to try to reduce the possibility that the organization will experience a disaster, but also, if a disaster does hit, to lessen the amount of damage that will take place. Although the organization cannot stop a tornado from coming, for example, it could choose to move its facility from Tornado Alley to an area less prone to these weather events. As another example, the organization cannot stop a car from plowing into and taking out a transformer that it relies on for power, but it can have a separate power feed from a different transformer in case this happens.

Recovery strategies are processes designed to rescue the company after a disaster takes place. These processes integrate mechanisms such as establishing alternate sites for facilities, implementing emergency response procedures, and possibly activating the preventive mechanisms that have already been implemented.

In this same scenario, data that is classified as "Business" can be up to three hours old when the production environment comes back online, so a less frequent data replication process is acceptable. Because the RTO for business data is eight hours, the company can choose to have hot-swappable hard drives available instead of having to pay for the more complicated and expensive failover technology.

The DR team has to figure out what the company needs to do to actually recover the processes and services it has identified as being so important to the organization overall. In its business continuity and recovery strategy, the team closely examines the critical, agreed-upon business functions, and then evaluates the numerous recovery and backup alternatives that might be used to recover critical business operations. It is important to choose the right tactics and technologies for the recovery of each critical business process and service in order to assure that the set MTD values are met.

So what does the DR team need to accomplish? The team needs to actually define the recovery processes, which are sets of predefined activities that will be implemented and carried out in response to a disaster. More importantly, these processes must be constantly reevaluated and updated as necessary to ensure that the organization meets or exceeds the MTDs. It all starts with understanding the business processes that would have to be recovered in the aftermath of a disaster. Armed with that knowledge, the DR team can make good decisions about data backup, recovery, and processing sites, as well as overall services availability, all of which we explore in the next sections.

Business Process Recovery

A *business process* is a set of interrelated steps linked through specific decision activities to accomplish a specific task. Business processes have starting and ending points and are repeatable. The processes should encapsulate the knowledge about services, resources, and operations provided by an organization. For example, when a customer requests to buy a book via a company's e-commerce site, the company's order fulfillment system must follow a business process such as this:

1. Validate that the book is available.
2. Validate where the book is located and how long it would take to ship it to the destination.
3. Provide the customer with the price and delivery date.
4. Verify the customer's credit card information.
5. Validate and process the credit card order.
6. Send the order to the book inventory location.
7. Send a receipt and tracking number to the customer.
8. Restock inventory.
9. Send the order to accounting.

The DR team needs to understand these different steps of the organization's most critical processes. The data is usually presented as a workflow document that contains the roles and resources needed for each process. The DR team must understand the following about critical business processes:

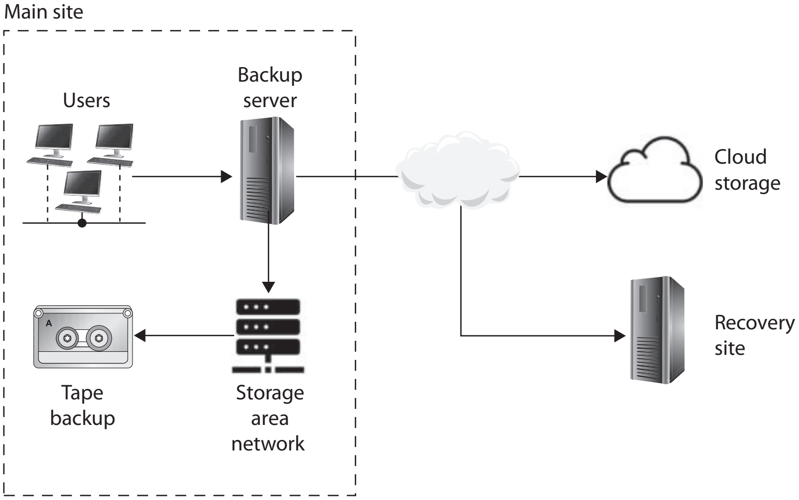
- Required roles
- Required resources
- Input and output mechanisms
- Workflow steps
- Required time for completion
- Interfaces with other processes

This will allow the team to identify threats and the controls to ensure the least amount of process interruption.

Data Backup

Data has become one of the most critical assets to nearly all organizations. It may include financial spreadsheets, blueprints on new products, customer information, product inventory, trade secrets, and more. In Chapter 2, we stepped through risk analysis procedures and, in Chapter 5, data classification. The DR team should not be responsible for setting up and maintaining the organization's data classification procedures, but the team should recognize that the organization is at risk if it does not have these procedures in place. This should be seen as a vulnerability that is reported to management. Management would need to establish another group of individuals who would identify the organization's data, define a loss criterion, and establish the classification structure and processes.

The DR team's responsibility is to provide solutions to protect this data and identify ways to restore it after a disaster. Data usually changes more often than hardware and software, so these backup or archival procedures must happen on a continual basis. The data backup process must make sense and be reasonable and effective. If data in the files changes several times a day, backup procedures should happen a few times a day or nightly to ensure all the changes are captured and kept. If data is changed once a month, backing up data every night is a waste of time and resources. Backing up a file and its corresponding changes is usually more desirable than having multiple copies of that one file. Online backup technologies usually record the changes to a file in a transaction log, which is separate from the original file.



The IT operations team should include a backup administrator, who is responsible for defining which data gets backed up and how often. These backups can be full, differential, or incremental, and are usually used in some type of combination with each other. Most files are not altered every day, so, to save time and resources, it is best to devise a backup plan that does not continually back up data that has not been modified. So, how do we know which data has changed and needs to be backed up without having to look at every file's modification date? This is accomplished by setting an *archive bit* to 1 if a file has been modified. The backup software reviews this bit when making its determination of whether the file gets backed up and, if so, clears the bit when it's done.

The first step is to do a *full backup*, which is just what it sounds like—all data is backed up and saved to some type of storage media. During a full backup, the archive bit is cleared, which means that it is set to 0. An organization can choose to do full backups only, in which case the restoration process is just one step, but the backup and restore processes could take a long time.

Most organizations choose to combine a full backup with a differential or incremental backup. A *differential process* backs up the files that have been modified since the *last full backup*. When the data needs to be restored, the full backup is laid down first, and then

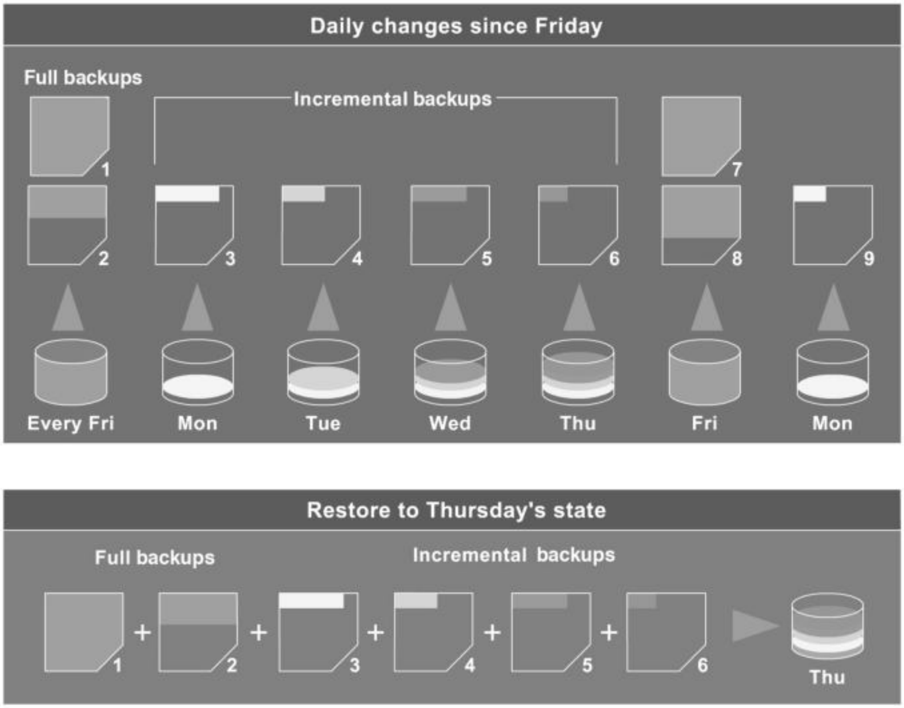


Figure 23-3 Backup software steps

the most recent differential backup is put down on top of it. The differential process does not change the archive bit value.

An *incremental process* backs up all the files that have changed since the *last full or incremental backup* and sets the archive bit to 0. When the data needs to be restored, the full backup data is laid down, and then each incremental backup is laid down on top of it in the proper order (see Figure 23-3). If an organization experienced a disaster and it used the incremental process, it would first need to restore the full backup on its hard drives and lay down every incremental backup that was carried out before the disaster took place (and after the last full backup). So, if the full backup was done six months ago and the operations department carried out an incremental backup each month, the backup administrator would restore the full backup and start with the older incremental backups taken since the full backup and restore each one of them until they were all restored.

Which backup process is best? If an organization wants the backup and restoration processes to be simple, it can carry out just full backups—but this may require a lot of hard drive space and time. Although using differential and incremental backup processes is more complex, it requires fewer resources and less time. A differential backup takes more time in the backing-up phase than an incremental backup, but it also takes less time to restore than an incremental backup because carrying out restoration of a differential backup happens in two steps, whereas in an incremental backup, every incremental backup must be restored in the correct sequence.

Whatever the organization chooses, it is important to not mix differential and incremental backups. This overlap could cause files to be missed, since the incremental backup changes the archive bit and the differential backup does not.

Critical data should be backed up and stored onsite *and* offsite. The onsite backups should be easily accessible for routine uses and should provide a quick restore process so operations can return to normal. However, onsite backups are not enough to provide real protection. The data should also be held in an offsite facility in case of disasters. One decision the CISO needs to make is where the offsite location should be in reference to the main facility. The closer the offsite backup storage site is, the easier it is to access, but this can put the backup copies in danger if a large-scale disaster manages to take out the organization's main facility and the backup facility. It may be wiser to choose a backup facility farther away, which makes accessibility harder but reduces the risk. Some organizations choose to have more than one backup facility: one that is close and one that is farther away.

Backup Storage Strategies

A backup strategy must take into account that failure can take place at any step of the process, so if there is a problem during the backup or restoration process that could corrupt the data, there should be a graceful way of backing out or reconstructing the data

Restoring Data from Backups: A Cautionary Tale

Can we actually restore data from backups? Backing up data is a wonderful thing in life, but making sure it can be properly restored is even better. Many organizations have developed a false sense of security based on the fact that they have a very organized and effective process of backing up their data. That sense of security can disappear in seconds when an organization realizes in a time of crisis that its restore processes do not work. For example, one company had paid an offsite backup facility to use a courier to collect its weekly backup tapes and transport them to the offsite facility for safekeeping. What the company did not realize was that this courier used the subway and many times set the tapes on the ground while waiting for the subway train. A subway has many large engines that create their own magnetic field. This can have the same effect on media as large magnets, meaning that the data can be erased or corrupted. The company never tested its restore processes and eventually experienced a disaster. Much to its surprise, it found out that three years of data were corrupted and unusable.

Many other stories and experiences like this are out there. Don't let your organization end up as an anecdote in someone else's book because it failed to verify that its backups could be restored.

from the beginning. The procedures for backing up and restoring data should be easily accessible and comprehensible even to operators or administrators who are not intimately familiar with a specific system. In an emergency situation, the same person who always does the backing up and restoring may not be around, or outsourced consultants may need to be temporarily hired to meet the restoration time constraints.

There are four commonly used backup strategies that you should be aware of:

- **Direct-attached storage** The backup storage is directly connected to the device being backed up, typically over a USB cable. This is better than nothing, but is not really well suited for centralized management. Worse yet, many ransomware attacks look for these attached storage devices and encrypt them too.
- **Network-attached storage (NAS)** The backup storage is connected to the device over the LAN and is usually a storage area network (SAN) managed by a backup server. This approach is usually centrally managed and allows IT administrators to enforce data backup policies. The main drawback is that, if a disaster takes out the site, the data may be lost or otherwise be rendered inaccessible.
- **Cloud storage** Many organizations use cloud storage as either the primary or secondary repository of backup data. If this is done on a virtual private cloud, it has the advantage of providing offsite storage so that, even if the organization's site is destroyed by a disaster, the data is available for recovery. Obviously, WAN connectivity must be reliable and fast enough to support this strategy if it is to be effective.
- **Offline media** As ransomware becomes more sophisticated, we are seeing more instances of attackers going after NAS and cloud storage. If your data is critical enough that you have to decrease the risk of it being lost as close to zero as you can, you may want to consider offline media such as tape backups, optical discs, or even external drives that are disconnected after each backup (and potentially removed offsite). This is the slowest and most expensive approach, but is also the most resistant to attacks.

Electronic vaulting and remote journaling are other solutions that organizations should be aware of. *Electronic vaulting* makes copies of files as they are modified and periodically transmits them to an offsite backup site. The transmission does not happen in real time, but is carried out in batches. So, an organization can choose to have all files that have been changed sent to the backup facility every hour, day, week, or month. The information can be stored in an offsite facility and retrieved from that facility in a short amount of time.

This form of backup takes place in many financial institutions, so when a bank teller accepts a deposit or withdrawal, the change to the customer's account is made locally to that branch's database and to the remote site that maintains the backup copies of all customer records.

Electronic vaulting is a method of transferring bulk information to offsite facilities for backup purposes. *Remote journaling* is another method of transmitting data offsite, but this usually only includes moving the journal or transaction logs to the offsite facility, not the actual files. These logs contain the deltas (changes) that have taken place to the individual files. Continuing with the bank example, if and when data is corrupted and needs to be restored, the bank can retrieve these logs, which are used to rebuild the lost data. Journaling is efficient for database recovery, where only the reapplication of a series of changes to individual records is required to resynchronize the database.

EXAM TIP Remote journaling takes place in real time and transmits only the file deltas. Electronic vaulting takes place in batches and moves the entire file that has been updated.

An organization may need to keep different versions of software and files, especially in a software development environment. The object and source code should be backed up along with libraries, patches, and fixes. The offsite facility should mirror the onsite facility, meaning it does not make sense to keep all of this data at the onsite facility and only the source code at the offsite facility. Each site should have a full set of the most current and updated information and files.

Another software backup technology is *tape vaulting*. Many organizations back up their data to tapes that are then manually transferred to an offsite facility by a courier or an employee. This manual process can be error-prone, so some organizations use *electronic tape vaulting*, in which the data is sent over a serial line to a backup tape system at the offsite facility. The company that maintains the offsite facility maintains the systems and changes out tapes when necessary. Data can be quickly backed up and retrieved when necessary. This technology improves recovery speed, reduces errors, and allows backups to be run more frequently.

Data repositories commonly have replication capabilities, so that when changes take place to one repository (i.e., database) they are replicated to all the other repositories within the organization. The replication can take place over telecommunication links, which allow offsite repositories to be continuously updated. If the primary repository goes down or is corrupted, the replication flow can be reversed, and the offsite repository updates and restores the primary repository. Replication can be asynchronous or synchronous. *Asynchronous replication* means the primary and secondary data volumes are out of sync. Synchronization may take place in seconds, hours, or days, depending upon the technology in place. With *synchronous replication*, the primary and secondary repositories are always in sync, which provides true real-time duplication. Figure 23-4 shows how offsite replication can take place.

The DR team must balance the cost to recover against the cost of the disruption. The balancing point becomes the recovery time objective. Figure 23-5 illustrates the relationship between the cost of various recovery technologies and the provided recovery times.

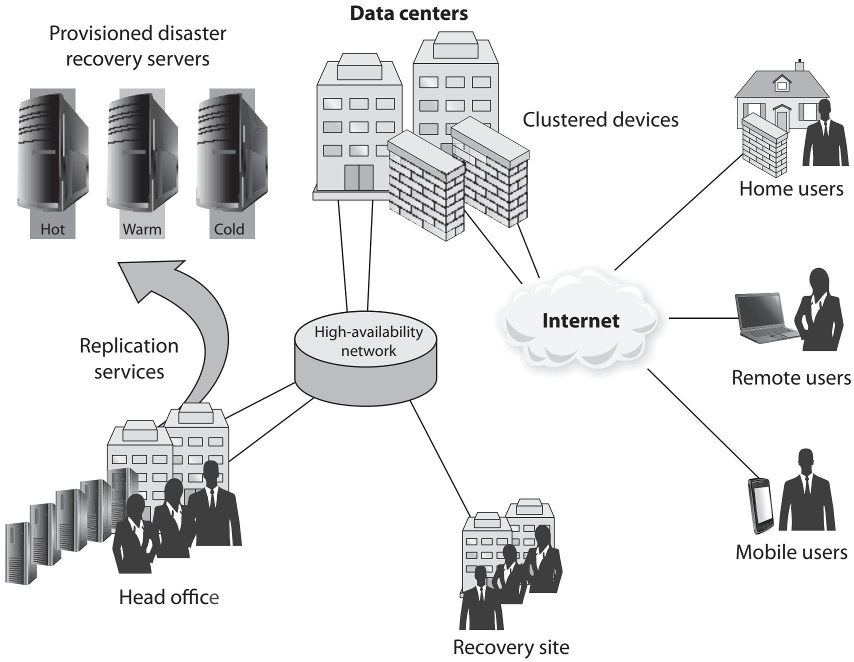


Figure 23-4 Offsite data replication for data recovery purposes

Choosing a Software Backup Facility

An organization needs to address several issues and ask specific questions when it is deciding upon a storage facility for its backup materials. The following list identifies just some of the issues that an organization needs to consider before committing to a specific vendor for this service:

- Can the media be accessed in the necessary timeframe?
- Is the facility closed on weekends and holidays, and does it only operate during specific hours of the day?
- Are the facility's access control mechanisms tied to an alarm and/or the police station?
- Does the facility have the capability to protect the media from a variety of threats?
- What is the availability of a bonded transport service?
- Are there any geographical environmental hazards such as floods, earthquakes, tornadoes, and so on that might affect the facility?
- Does the facility have a fire detection and suppression system?
- Does the facility provide temperature and humidity monitoring and control?
- What type of physical, administrative, and logical access controls are used?

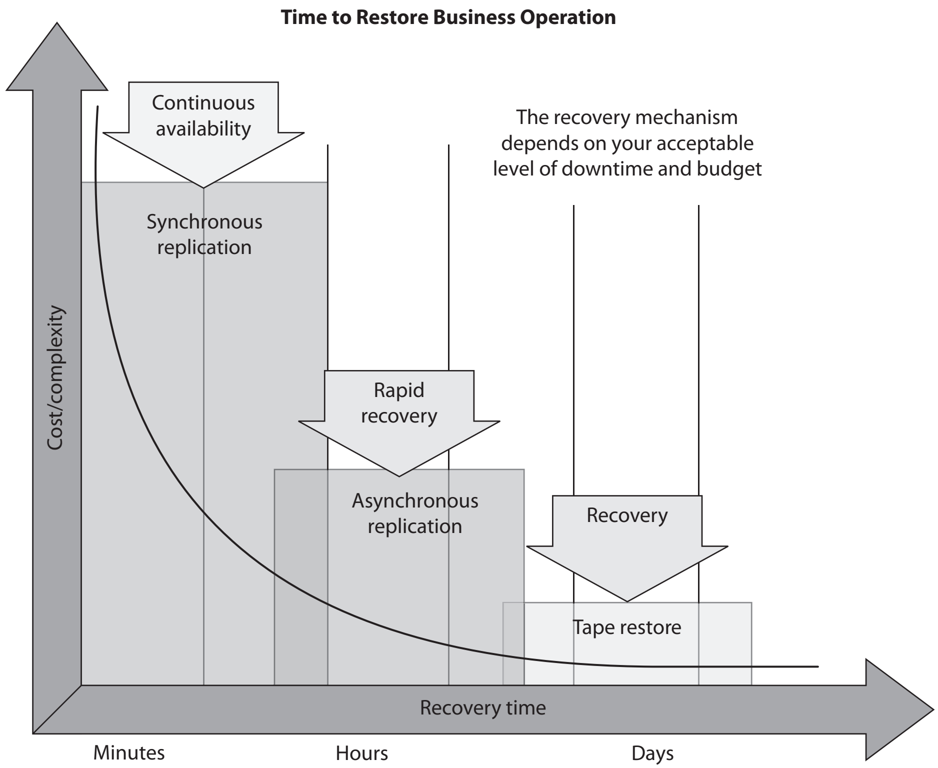


Figure 23-5 The criticality of data recovery will dictate the recovery solution.

The questions and issues that need to be addressed will vary depending on the type of organization, its needs, and the requirements of a backup facility.

Documentation

Documentation seems to be a dreaded task to most people, who will find many other tasks to take on to ensure they are not the ones stuck with documenting processes and procedures. However, without proper documentation, even an organization that does a terrific job of backing up data to an offsite facility will be scrambling to figure which backups it needs when a disaster hits.

Restoration of files can be challenging but restoring a whole environment that was swept away in a flood can be overwhelming, if not impossible. Procedures need to be documented because when they are actually needed, it will most likely be a chaotic and frantic atmosphere with a demanding time schedule. The documentation may need to include information on how to install images, configure operating systems and servers, and properly install utilities and proprietary software. Other documentation could include a calling tree, which outlines who should be contacted, in what order, and

Storing Business Continuity and Disaster Recovery Plans

Once the business continuity and disaster recovery plans are completed, where should they be stored? Should the organization have only one copy and keep it safely in a file cabinet next to Bob so that he feels safe? Nope. There should be two or three copies of these plans. One copy may be at the primary location, but the other copies should be at other locations in case the primary facility is destroyed. This reduces the risk of not having access to the plans when needed.

These plans should not be stored in a file cabinet, but rather in a fire-resistant safe. When they are stored offsite, they need to be stored in a way that provides just as much protection as the primary site would provide.

who is responsible for doing the calling. The documentation must also contain contact information for specific vendors, emergency agencies, offsite facilities, and any other entity that may need to be contacted in a time of need.

Most network environments evolve over time. Software is installed on top of other software, configurations are altered over the years to properly work in a unique environment, and service packs and patches are routinely installed to fix issues and update software. To expect one person or a group of people to go through all these steps during a crisis and end up with an environment that looks and behaves exactly like the original environment and in which all components work together seamlessly may be a lofty dream.

So, the dreaded task of documentation may be the saving grace one day. It is an essential piece of business, and therefore an essential piece in disaster recovery and business continuity. It is, therefore, important to make one or more roles responsible for proper documentation. As with all the items addressed in this chapter, simply saying "All documentation will be kept up to date and properly protected" is the easy part—saying and doing are two different things. Once the DR team identifies tasks that must be done, the tasks must be assigned to individuals, and those individuals have to be accountable. If these steps are not taken, the organization may have wasted a lot of time and resources defining these tasks, and still be in grave danger if a disaster occurs.

Human Resources

One of the resources commonly left out of the DR equation is people. An organization may restore its networks and critical systems and get business functions up and running, only to realize it doesn't know the answer to the question, "Who will take it from here?" The area of human resources is a critical component to any recovery and continuity process, and it needs to be fully thought out and integrated into the plan.

What happens if we have to move to an offsite facility that is 250 miles away? We cannot expect people to drive back and forth from home to work. Should we pay for temporary housing for the necessary employees? Do we have to pay their moving costs? Do we need to hire new employees in the area of the offsite facility? If so, what skill set do we need from them? These are all important questions for the organization's senior leaders to answer.

If a large disaster takes place that affects not only the organization's facility but also surrounding areas, including housing, employees will be more worried about their families than their organization. Some organizations assume that employees will be ready and available to help them get back into production, when in fact they may need to be at home because they have responsibilities to their families.

Regrettably, some employees may be killed or severely injured in the disaster, and the organization should have plans in place to replace employees quickly through a temporary employment agency or a job recruiter. This is an extremely unfortunate scenario to contemplate, but it is part of reality. The team that considers all threats and is responsible for identifying practical solutions needs to think through all of these issues.

Organizations should already have *executive succession planning* in place. This means that if someone in a senior executive position retires, leaves the organization, or is killed, the organization has predetermined steps to carry out to ensure a smooth transition to that executive's replacement. The loss of a senior executive could tear a hole in the organization's fabric, creating a leadership vacuum that must be filled quickly with the right individual. The line-of-succession plan defines who would step in and assume responsibility for this role. Many organizations have "deputy" roles. For example, an organization may have a deputy CIO, deputy CFO, and deputy CEO ready to take over the necessary tasks if the CIO, CFO, or CEO becomes unavailable.

Often, larger organizations also have a policy indicating that two or more of the senior staff cannot be exposed to a particular risk at the same time. For example, the CEO and president cannot travel on the same plane. If the plane were to crash and both individuals were killed, then the company could face a leadership crisis. This is why you don't see the president of the United States and the vice president together too often. It is not because they don't like each other and thus keep their distance from each other. It is because there is a policy indicating that to protect the United States, its top leaders cannot be under the same risk at the same time.

Recovery Site Strategies

Disruptions, in BCP terms, are of three main types: nondisasters, disasters, and catastrophes. A *nondisaster* is a disruption in service that has significant but limited impact on the conduct of business processes at a facility. The solution could include hardware, software, or file restoration. A *disaster* is an event that causes the entire facility to be unusable for a day or longer. This usually requires the use of an alternate processing facility and restoration of software and data from offsite copies. The alternate site must be available to the organization until its main facility is repaired and usable. A *catastrophe* is a major disruption that destroys the facility altogether. This requires both a short-term solution, which would be an offsite facility, and a long-term solution, which may require rebuilding the original facility. Disasters and catastrophes are rare compared to nondisasters, thank goodness.

When dealing with disasters and catastrophes, an organization has three basic options: select a dedicated site that the organization owns and operates itself; lease a commercial facility, such as a "hot site" that contains all the equipment and data needed to quickly restore operations; or enter into a formal agreement with another facility, such as a service bureau, to restore its operations. When choosing the right solution for its needs, the organization evaluates each alternative's ability to support its operations, to do it within an acceptable timeframe, and to have a reasonable cost.

An important consideration with third parties is their reliability, both in normal times and during an emergency. Their reliability can depend on considerations such as their track record, the extent and location of their supply inventory, and their access to supply and communication channels. Organizations should closely query the management of the alternative facility about such things as the following:

- How long will it take to recover from a certain type of incident to a certain level of operations?
- Will it give priority to restoring the operations of one organization over another after a disaster?
- What are its costs for performing various functions?
- What are its specifications for IT and security functions? Is the workspace big enough for the required number of employees?

To recover from a disaster that prevents or degrades use of the primary site temporarily or permanently, an organization must have an offsite backup facility available. Generally, an organization establishes contracts with third-party vendors to provide such services. The client pays a monthly fee to retain the right to use the facility in a time of need, and then incurs an activation fee when the facility actually has to be used. In addition, a daily or hourly fee is imposed for the duration of the stay. This is why service agreements for backup facilities should be considered a short-term solution, not a long-term solution.

It is important to note that most recovery site contracts do not promise to house the organization in need at a specific location, but rather promise to provide what has been contracted for somewhere within the organization's locale. On, and subsequent to, September 11, 2001, many organizations with Manhattan offices were surprised when they were redirected by their backup site vendor not to sites located in New Jersey (which were already full), but rather to sites located in Boston, Chicago, or Atlanta. This adds yet another level of complexity to the recovery process, specifically the logistics of transporting people and equipment to unplanned locations.

An organization can choose from three main types of leased or rented offsite recovery facilities:

- **Hot site** A facility that is fully configured and ready to operate within a few hours. All the necessary equipment is already installed and configured. In many cases, the remote data backup services are included, so the RPO can be down to an hour or even less. These sites are a good choice for an organization with a very small MTD. Of course, the organization should conduct regular tests (annually, at least) to ensure the site is functioning in the necessary state of readiness.

  The hot site is, by far, the most expensive of the three types of offsite facilities. The organization has to pay for redundant hardware and software, in addition to the expenses of the site itself. Organizations that use hot sites as part of their recovery strategy tend to limit them to mission-critical systems only.

- **Warm site** A facility that is usually partially configured with some equipment, such as HVAC, and foundational infrastructure components, but does not include all the hardware needed to restore mission-critical business functions. Staging a facility with duplicate hardware and computers configured for immediate operation is extremely expensive, so a warm site provides a less expensive alternate. These sites typically do not have data replicated to them, so backups would have to be delivered and restored onto the warm site systems after a disaster.

  The warm site is the most widely used model. It is less expensive than a hot site, and can be up and running within a reasonably acceptable time period. It may be a better choice for organizations that depend on proprietary and unusual hardware and software, because they will bring their own hardware and software with them to the site after the disaster hits. Drawbacks, however, are that much of the equipment has to be procured, delivered to, and configured at the warm site after the fact, and testing will be more difficult. Thus, an organization may not be certain that it will in fact be able to return to an operating state within its RTO.

- **Cold site** A facility that supplies the basic environment, electrical wiring, HVAC, plumbing, and flooring but none of the equipment or additional services. A cold site is essentially an empty data center. It may take weeks to get the site activated and ready for work. The cold site could have equipment racks and dark fiber (fiber that does not have the circuit engaged) and maybe even desks. However, it would require the receipt of equipment from the client, since it does not provide any.

  The cold site is the least expensive option, but takes the most time and effort to actually get up and functioning right after a disaster, as the systems and software must be delivered, set up, and configured. Cold sites are often used as backups for call centers, manufacturing plants, and other services that can be moved lock, stock, and barrel in one shot.

After a catastrophic loss of the primary facility, some organizations will start their recovery in a hot or warm site, and transfer some operations over to a cold site after the latter has had time to set up.

It is important to understand that the different site types listed here are provided by service bureaus. A *service bureau* is a company that has additional space and capacity to provide applications and services such as call centers. An organization pays a monthly subscription fee to a service bureau for this space and service. The fee can be paid for contingencies such as disasters and emergencies. You should evaluate the ability of a service bureau to provide services just as you would evaluate divisions within your own organization, particularly on matters such as its ability to alter or scale its software and hardware configurations or to expand its operations to meet the needs of a contingency.

NOTE Related to a service bureau is a *contingency supplier*; its purpose is to supply services and materials temporarily to an organization that is experiencing an emergency. For example, a contingency supplier might provide raw materials such as heating fuel or backup telecommunication services. In considering contingency suppliers, the BCP team should think through considerations such as the level of services and materials a supplier can provide, how quickly a supplier can ramp up to supply them, and whether the supplier shares similar communication paths and supply chains as the affected organization.

Most organizations use warm sites, which have some devices such as networking equipment, some computers and data storage, but very little else. These organizations usually cannot afford a hot site, and the extra downtime would not be considered detrimental. A warm site can provide a longer-term solution than a hot site. Organizations that decide to go with a cold site must be able to be out of operation for a week or two. The cold site usually includes power, raised flooring, climate control, and wiring.

The following provides a quick overview of the differences between offsite facilities.

Hot site advantages:

- Ready within hours or even minutes for operation
- Highly available
- Usually used for short-term solutions, but available for longer stays
- Recovery testing is easy

Hot site disadvantages:

- Very expensive
- Limited systems

Tertiary Sites

An organization may recognize the danger of the primary recovery site not being available when needed. This could be the case if the service provider assumes that not every customer will attempt to occupy the site at the same time, and then a major regional disaster affects more organizations than anticipated. It could also happen if a disaster affects the recovery site itself (e.g., fire, flood). Mitigating this risk could require a *tertiary site*, a backup recovery site just in case the primary is unavailable. The tertiary site is sometimes referred to as a "backup to the backup." This is basically plan B if plan A does not work out. Obviously, this is a very expensive proposition, so its costs should be balanced with the risks it is intended to mitigate.

Warm and cold site advantages:

- Less expensive
- Available for longer timeframes because of the reduced costs
- Practical for proprietary hardware or software use

Warm and cold site disadvantages:

- Limited ability to perform recovery testing
- Resources for operations not immediately available

Reciprocal Agreements

Another approach to alternate offsite facilities is to establish a *reciprocal agreement* with another organization, usually one in a similar field or that has similar technological infrastructure. This means that organization A agrees to allow organization B to use its facilities if organization B is hit by a disaster, and vice versa. This is a cheaper way to go than the other offsite choices, but it is not always the best choice. Most environments are maxed out pertaining to the use of facility space, resources, and computing capability. To allow another organization to come in and work out of the same shop could prove to be detrimental to both organizations. Whether it can assist the other organization while tending effectively to its own business is an open question. The stress of two organizations working in the same environment could cause tremendous levels of tension. If it did work out, it would only provide a short-term solution. Configuration management could be a nightmare. Does the other organization upgrade to new technology and retire old systems and software? If not, one organization's systems may become incompatible with those of the other.

If your organization allows another organization to move into its facility and work from there, you may have a solid feeling about your friend, the CEO, but what about all of her employees, whom you do not know? The mixing of operations could introduce many security issues. Now you have a new subset of people who may need to have privileged and direct access to your resources in the shared environment. Close attention needs to be paid when assigning these other people access rights and permissions to your critical assets and resources, if they need access at all. Careful testing is recommended to see if one organization or the other can handle the extra loads.

Offsite Location

When choosing a backup facility, it should be far enough away from the original site so that one disaster does not take out both locations. In other words, it is not logical to have the backup site only a few miles away if the organization is concerned about tornado damage, because the backup site could also be affected or destroyed. There is a rule of thumb that suggests that alternate facilities should be, at a bare minimum, at least 5 miles away from the primary site, while 15 miles is recommended for most low-to-medium critical environments, and 50 to 200 miles is recommended for critical operations, to give maximum protection in cases of regional disasters.

Reciprocal agreements have been known to work well in specific businesses, such as newspaper printing. These businesses require very specific technology and equipment that is not available through any subscription service. These agreements follow a "you scratch my back and I'll scratch yours" mentality. For most other organizations, reciprocal agreements are generally, at best, a secondary option for disaster protection. The other issue to consider is that these agreements are usually not enforceable because they're not written in legally binding terms. This means that although organization A said organization B could use its facility when needed, when the need arises, organization A may not have a legal obligation to fulfill this promise. However, there are still many organizations who do opt for this solution either because of the appeal of low cost or, as noted earlier, because it may be the only viable solution in some cases.

Organizations that have a reciprocal agreement need to address the following important issues before a disaster hits:

- How long will the facility be available to the organization in need?
- How much assistance will the staff supply in integrating the two environments and ongoing support?
- How quickly can the organization in need move into the facility?
- What are the issues pertaining to interoperability?
- How many of the resources will be available to the organization in need?
- How will differences and conflicts be addressed?
- How does change control and configuration management take place?
- How often can exercising and testing take place?
- How can critical assets of both organizations be properly protected?

A variation on a reciprocal agreement is a consortium, or *mutual aid agreement*. In this case, more than two organizations agree to help one another in case of an emergency. Adding multiple organizations to the mix, as you might imagine, can make things even more complicated. The same concerns that apply with reciprocal agreements apply here, but even more so. Organizations entering into such agreements need to formally and legally document their mutual responsibilities in advance. Interested parties, including the legal and IT departments, should carefully scrutinize such accords before the organization signs onto them.

Redundant Sites

Some organizations choose to have a *redundant site*, or mirrored site, meaning one site is equipped and configured exactly like the primary site, which serves as a redundant environment. The business-processing capabilities between the two sites can be completely synchronized. A redundant site is owned by the organization and mirrors the original production environment. A redundant site has clear advantages: it has full availability, is ready to go at a moment's notice, and is under the organization's complete control. This is, however, one of the most expensive backup facility options, because a full environment must be maintained even though it usually is not used for regular production activities until after a disaster takes place that triggers the relocation of services to the redundant site. But "expensive" is relative here. If a company would lose a million dollars if it were out of business for just a few hours, the loss potential would override the cost of this option. Many organizations are subjected to regulations that dictate they must have redundant sites in place, so expense is not a matter of choice in these situations.

EXAM TIP A *hot* site is a subscription service. A *redundant* site, in contrast, is a site owned and maintained by the organization, meaning the organization does not pay anyone else for the site. A redundant site might be "hot" in nature, meaning it is ready for production quickly. However, the CISSP exam differentiates between a hot site (a subscription service) and a redundant site (owned by the organization).

Another type of facility-backup option is a *rolling hot site*, or mobile hot site, where the back of a large truck or a trailer is turned into a data processing or working area. This is a portable, self-contained data facility. The trailer has the necessary power, telecommunications, and systems to do some or all of the processing right away. The trailer can be brought to the organization's parking lot or another location. Obviously, the trailer has to be driven over to the new site, the data has to be retrieved, and the necessary personnel have to be put into place.

Another, similar solution is a prefabricated building that can be easily and quickly put together. Military organizations and large insurance companies typically have rolling hot sites or trucks preloaded with equipment because they often need the flexibility to quickly relocate some or all of their processing facilities to different locations around the world depending on where the need arises.

It is best if an organization is aware of all available options for hardware and facility backups to ensure it makes the best decision for its specific business and critical needs.

Multiple Processing Sites

Another option for organizations is to have *multiple processing sites*. An organization may have ten different facilities throughout the world, which are connected with specific technologies that could move all data processing from one facility to another in a matter of seconds when an interruption is detected. This technology can be implemented within the organization or from one facility to a third-party facility. Certain service providers provide this type of functionality to their customers. So if an organization's data processing is interrupted, all or some of the processing can be moved to the service provider's servers.

Availability

We close this section on recovery strategies by considering the nondisasters to which we referred earlier. These are the incidents that may not require evacuation of personnel or facility repairs but that can still have a significant detrimental effect on the ability of the organization to execute its mission. We want our systems and services to be available all the time, no matter what. However, we all realize this is just not possible. *Availability* can be defined as the portion of the time that a system is operational and able to fulfill its intended purpose. But how can we ensure the availability of the systems and services on which our organizations depend?

High Availability

High availability (HA) is a combination of technologies and processes that work together to ensure that some specific thing is up and running most of the time. The specific thing can be a database, a network, an application, a power supply, and so on. Service providers have *service level agreements (SLAs)* with their customers that outline the amount of uptime the service providers promise to provide. For example, a hosting company can promise to provide 99 percent uptime for Internet connectivity. This means the company is guaranteeing that at least 99 percent of the time, the Internet connection you purchase from it will be up and running. It also means that you can experience up to 3.65 days a year (or 7.2 hours per month) of downtime and it won't be a violation of the SLA. Increase that to 99.999 percent (referred to as "five nines") uptime and the allowable downtime drops to 5.26 seconds per year, but the price you pay for service goes through the roof.

NOTE HA is in the eye of the beholder. For some organizations or systems, an SLA of 90 percent ("one nine") uptime and its corresponding potential 36+ days of downtime a year is perfectly fine, particularly for organizations that are running on a tight budget. Other organizations require "nine nines" or 99.9999999 percent availability for mission-critical systems. You have to balance the cost of HA with the loss you're trying to mitigate.

Just because a service is available doesn't necessarily mean that it is operating acceptably. Suppose your company's high-speed e-commerce server gets infected with a bitcoin miner that drives CPU utilization close to 100 percent. Technically, the server is available and will probably be able to respond to customer requests. However, response times will likely be so lengthy that many of your customers will simply give up and go shop somewhere else. The service is available, but its quality is unacceptable.

Quality of Service

Quality of service (QoS) defines minimum acceptable performance characteristics of a particular service. For example, for the e-commerce server example, we could define parameters like response time, CPU utilization, or network bandwidth utilization, depending on how the service is being provided. SLAs may include one or more specifications for QoS, which allows service providers to differentiate classes of service that are prioritized for different clients. During a disaster, the available bandwidth on external links may be limited, so the affected organization could specify different QoS for its externally facing systems. For example, the e-commerce company in our example could determine the minimum data rate to keep its web presence available to customers and specify that as the minimum QoS rate at the expense of, say, its e-mail or Voice over Internet Protocol (VoIP) traffic.

To provide HA and meet stringent QoS requirements, the hosting company has to have a long list of technologies and processes that provide redundancy, fault tolerance, and failover capabilities. *Redundancy* is commonly built into the network at a routing protocol level. The routing protocols are configured such that if one link goes down or gets congested, traffic is automatically routed over a different network link. An organization can also ensure that it has redundant hardware available so that if a primary device goes down, the backup component can be swapped out and activated.

If a technology has a *failover* capability, this means that if there is a failure that cannot be handled through normal means, then processing is “switched over” to a working system. For example, two servers can be configured to send each other “heartbeat” signals every 30 seconds. If server A does not receive a heartbeat signal from server B after 40 seconds, then all processes are moved to server A so that there is no lag in operations. Also, when servers are *clustered*, an overarching piece of software monitors each server and carries out load balancing. If one server within the cluster goes down, the clustering software stops sending it data to process so that there are no delays in processing activities.

Fault Tolerance and System Resilience

Fault tolerance is the capability of a technology to continue to operate as expected even if something unexpected takes place (a fault). If a database experiences an unexpected glitch, it can roll back to a known-good state and continue functioning as though nothing bad happened. If a packet gets lost or corrupted during a TCP session, the TCP protocol will resend the packet so that system-to-system communication is not affected. If a disk within a RAID system gets corrupted, the system uses its parity data to rebuild the corrupted data so that operations are not affected.

Although the terms fault tolerance and resilience are often used synonymously, they mean subtly different things. Fault tolerance means that when a fault happens, there’s a system in place (a backup or redundant one) to ensure services remain uninterrupted. *System resilience* means that the system continues to function, albeit in a degraded fashion, when a fault is encountered. Think of it as the difference between having a spare tire for your car and having run-flat tires. The spare tire provides fault tolerance in that it enables you to recover (fairly) quickly from a flat tire and be on your way. Run-flat tires allow you to continue to drive your car (albeit slower) if you run over a nail on the road. A resilient system is fault tolerant, but a fault tolerant one may not be resilient.

High Availability in Disaster Recovery

Redundancy, fault tolerance, resilience, and failover capabilities increase the reliability of a system or network, where *reliability* is the probability that a system performs the necessary function for a specified period under defined conditions. High reliability allows for high availability, which is a measure of its readiness. If the probability of a system performing as expected under defined conditions is low, then the availability for this system cannot be high. For a system to have the characteristic of high availability,

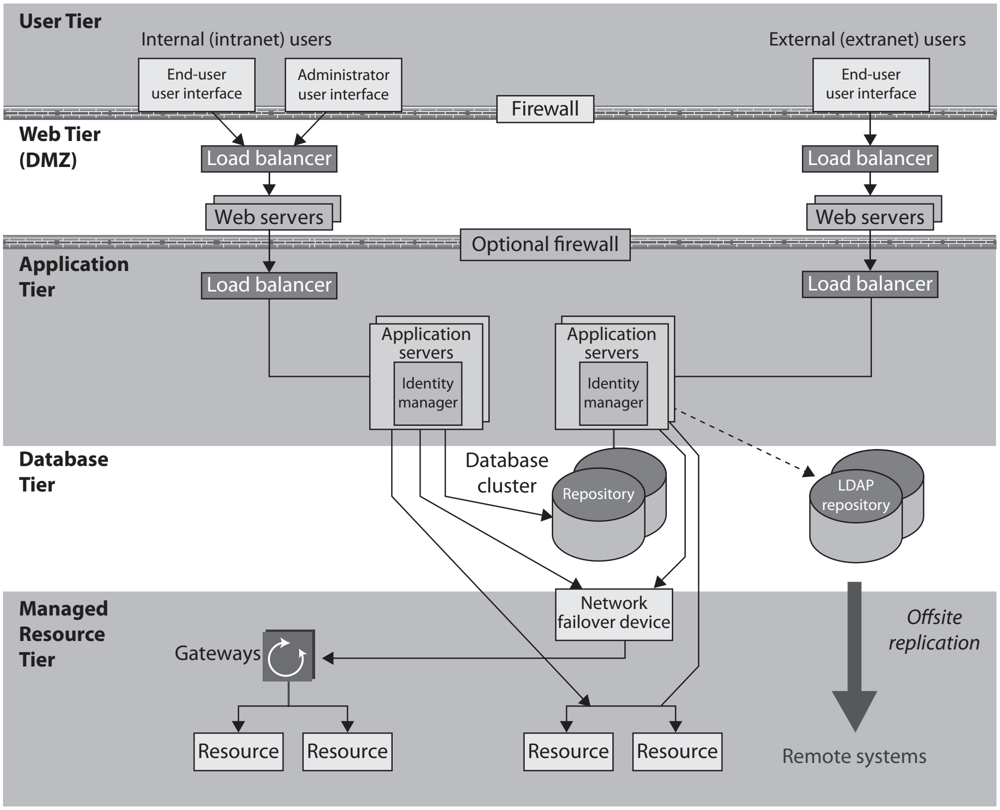


Figure 23-6 High-availability technologies

then high reliability must be in place. Figure 23-6 illustrates where load balancing, clustering, failover devices, and replication commonly take place in a network architecture.

Remember that data restoration (RPO) requirements can be different from processing restoration (RTO) requirements. Data can be restored through backup tapes, electronic vaulting, or synchronous or asynchronous replication. Processing capabilities can be restored through clustering, load balancing, redundancy, and failover technologies. If the results of the BCP team's BIA indicate that the RPO value is two days, then the organization can use tape backups. If the RPO value is one minute, then synchronous replication needs to be in place. If the BIA indicates that the RTO value is three days, then redundant hardware can be used. If the RTO value is one minute, then clustering and load balancing should be used.

HA and disaster recovery are related concepts. HA technologies and processes are commonly put into place so that if a disaster does take place, either the critical functions are likelier to remain available or the delay of getting them back online and running is low.

Many IT and security professionals usually think of HA only in technology terms, but remember that there are many things that an organization needs to have available to keep functioning. Availability of each of the following items must be thought through and planned:

- Facility (cold, warm, hot, redundant, rolling, reciprocal sites)
- Infrastructure (redundancy, fault tolerance)
- Storage (SAN, cloud)
- Server (clustering, load balancing)
- Data (backups, online replication)
- Business processes
- People

NOTE Virtualization and cloud computing are covered in Chapter 7. We will not go over those technologies again in this chapter, but know that the use of these technologies has drastically increased in the realm of business continuity and disaster recovery planning solutions.

Disaster Recovery Processes

Recovering from a disaster begins way before the event occurs. It starts by anticipating threats and developing goals that support the organization's continuity of operations. If you do not have established goals, how do you know when you are done and whether your efforts were actually successful? Goals are established so everyone knows the ultimate objectives. Establishing goals is important for any task, but especially for business continuity and disaster recovery plans. The definition of the goals helps direct the proper allocation of resources and tasks, supports the development of necessary strategies, and assists in financial justification of the plans and program overall. Once the goals are set, they provide a guide to the development of the actual plans themselves. Anyone who has been involved in large projects that entail many small, complex details knows that at times it is easy to get off track and not actually accomplish the major goals of the project. Goals are established to keep everyone on track and to ensure that the efforts pay off in the end.

Great—we have established that goals are important. But the goal could be, "Keep the company in business if an earthquake hits." That's a good goal, but it is not overly useful without more clarity and direction. To be useful, a goal must contain certain key information, such as the following:

- **Responsibility** Each individual involved with recovery and continuity should have their responsibilities spelled out in writing to ensure a clear understanding in a chaotic situation. Each task should be assigned to the individual most logically situated to handle it. These individuals must know what is expected of them, which is done through training, exercises, communication, and documentation. So, for example, instead of just running out of the building screaming, an individual must know that he is responsible for shutting down the servers before he can run out of the building screaming.

- **Authority** In times of crisis, it is important to know who is in charge. Teamwork is important in these situations, and almost every team does much better with an established and trusted leader. Such leaders must know that they are expected to step up to the plate in a time of crisis and understand what type of direction they should provide to the rest of the employees. Everyone else must recognize the authority of these leaders and respond accordingly. Clear-cut authority will aid in reducing confusion and increasing cooperation.
- **Priorities** It is extremely important to know what is critical versus what is merely nice to have. Different departments provide different functionality for an organization. The critical departments must be singled out from the departments that provide functionality that the organization can live without for a week or two. It is necessary to know which department must come online first, which second, and so on. That way, the efforts are made in the most useful, effective, and focused manner. Along with the priorities of departments, the priorities of systems, information, and programs must be established. It may be necessary to ensure that the database is up and running before working to bring the web servers online. The general priorities must be set by management with the help of the different departments and IT staff.
- **Implementation and testing** It is great to write down very profound ideas and develop plans, but unless they are actually carried out and tested, they may not add up to a hill of beans. Once a disaster recovery plan is developed, it actually has to be put into action. It needs to be documented and stored in places that are easily accessible in times of crisis. The people who are assigned specific tasks need to be taught and informed how to fulfill those tasks, and dry runs must be done to walk people through different situations. The exercises should take place at least once a year, and the entire program should be continually updated and improved.

NOTE We address various types of tests, such as walkthrough, tabletop, simulation, parallel, and full interruption, later in this chapter.

According to the U.S. Federal Emergency Management Agency (FEMA), 90 percent of small businesses that experience a disaster and are unable to restore operations within five days will fail within the following year. Not being able to bounce back quickly or effectively by setting up shop somewhere else can make a company lose business and, more importantly, its reputation. In such a competitive world, customers have a lot of options. If one company is not prepared to bounce back after a disruption or disaster, customers may go to another vendor and stay there.

The biggest effect of an incident, especially one that is poorly managed or that was preventable, is on an organization's reputation or brand. This can result in a considerable and even irreparable loss of trust by customers and clients. On the other hand, handling an incident well, or preventing great damage through smart, preemptive measures, can enhance the reputation of, or trust in, an organization.

The *disaster recovery plan (DRP)* should address in detail all of the topics we have covered so far. The actual format of the DRP will depend on the environment, the goals of the plan, priorities, and identified threats. After each of those items is examined and documented, the topics of the plan can be divided into the necessary categories.

Response

The first question the DRP should answer is, "What constitutes a disaster that would trigger this plan?" Every leader within an organization (and, ideally, everyone else too) should know the answer. Otherwise, precious time is lost notifying people who should've self-activated as soon as the incident occurred, a delay that could cost lives or assets. Examples of clear-cut disasters that would trigger a response are loss of power exceeding ten minutes, flooding in the facility, or terrorist attack against or near the site.

Every DRP is different, but most follow a familiar sequence of events:

1. Declaration of disaster
2. Activation of the DR team
3. Internal communications (ongoing from here on out)
4. Protection of human safety (e.g., evacuation)
5. Damage assessment
6. Execution of appropriate system-specific DRPs (each system and network should have its own DRP)
7. Recovery of mission-critical business processes/functions
8. Recovery of all other business processes/functions

Personnel

The DRP needs to define several different teams that should be properly trained and available if a disaster hits. Which types of teams an organization needs depends upon the organization. The following are some examples of teams that an organization may need to construct:

- Damage assessment team
- Recovery team
- Relocation team
- Restoration team
- Salvage team
- Security team

The DR coordinator should have an understanding of the needs of the organization and the types of teams that need to be developed and trained. Employees should be assigned to the specific teams based on their knowledge and skill set. Each team needs to have a designated leader, who will direct the members and their activities. These team leaders will be responsible not only for ensuring that their team's objectives are met but also for communicating with each other to make sure each team is working in parallel phases.

The purpose of the *recovery team* should be to get whatever systems are still operable back up and running as quickly as possible to reduce business disruptions. Think of them as the medics whose job is to stabilize casualties until they can be transported to the hospital. In this case, of course, there is no hospital for information systems, but there may be a recovery site. Getting equipment and people there in an orderly fashion should be the job of the *relocation team*. The *restoration team* should be responsible for getting the alternate site into a working and functioning environment, and the *salvage team* should be responsible for starting the recovery of the original site. Both teams must know how to do many tasks, such as install operating systems, configure workstations and servers, string wire and cabling, set up the network and configure networking services, and install equipment and applications. Both teams must also know how to restore data from backup facilities and how to do so in a secure manner, one that ensures the availability, integrity, and confidentiality of the system and data.

The DRP must outline the specific teams, their responsibilities, and notification procedures. The plan must indicate the methods that should be used to contact team leaders during business hours and after business hours.

Communications

The purpose of the emergency communications plan that is part of the overall DRP is to ensure that everyone knows what to do at all times and that the DR team remains synchronized and coordinated. This all starts with the DR plan itself. As stated previously, copies of the DRP need to be kept in one or more locations other than the primary site, so that if the primary site is destroyed or negatively affected, the plan is still available to the teams. It is also critical that different formats of the plan be available to the teams, including both electronic and paper versions. An electronic version of the plan is not very useful if you don't have any electricity to run a computer.

In addition to having copies of the recovery documents located at their offices and homes, key individuals should have easily accessible versions of critical procedures and call tree information. One simple way to accomplish the latter is to publish a call tree on cards that can be affixed to personnel badges or kept in a wallet. In an emergency situation, valuable minutes are better spent responding to an incident than looking for a document or having to wait for a laptop to power up. Of course, the call tree is only as effective as it is accurate and up to date, so verifying it periodically is imperative.

One limitation of call trees is that they are point to point, which means they're typically good for getting the word out, but not so much for coordinating activities. Group text messages work better, but only in the context of fairly small and static groups. Many organizations have group chat solutions, but if those rely on the organization's servers, they may be unavailable during a disaster. It is a good idea, then, to establish a communications platform that is completely independent of the organizational infrastructure. Solutions like Slack and Mattermost offer a free service that is typically sufficient to keep most organizations connected in emergencies. The catch, of course, is that everyone needs to have the appropriate client installed on their personal devices and know when and how to connect. Training and exercises are the keys to successful execution of any plan, and the communications plan is no exception.

NOTE An organization may need to solidify communications channels and relationships with government officials and emergency response groups. The goal of this activity is to solidify proper protocol in case of a city- or region-wide disaster. During the BIA phase, the DR team should contact local authorities to elicit information about the risks of its geographical location and how to access emergency zones. If the organization has to perform DR, it may need to contact many of these emergency response groups.

PACE Communications Plans

The U.S. armed forces routinely develop Primary, Alternate, Contingency, and Emergency (PACE) communications plans. The PACE plan outlines the different capabilities that exist and aligns them into these four categories based on their ability to meet defined information exchange requirements. Each category is defined here:

- **Primary** The normal or expected capability that is used to achieve the objective.
- **Alternate** A fully satisfactory capability that can be used to achieve the objective with minimal impact to the operation or exercise. This capability is used when the Primary capability is unavailable.
- **Contingency** A workable capability that can be used to achieve the objective. This capability may not be as fast or easy as the Primary or Alternate but is capable of achieving the objective with an acceptable amount of time and effort. This capability is used when the Primary and the Alternate capabilities are unavailable.
- **Emergency** This is the last-resort capability and typically may involve significantly more time and effort than any of the other capabilities. This capability should be used only when the Primary, Alternate, and Contingency capabilities are unavailable.

The PACE plan includes redundant communications capabilities and specifies the order in which the organization will employ the capabilities when communication outages occur.

Assessment

A role, or a team, needs to be created to carry out a *damage assessment* once a disaster has taken place. The assessment procedures should be properly documented in the DRP and include the following steps:

- Determine the cause of the disaster.
- Determine the potential for further damage.
- Identify the affected business functions and areas.
- Identify the level of functionality for the critical resources.
- Identify the resources that must be replaced immediately.
- Estimate how long it will take to bring critical functions back online.

After the damage assessment team collects and assesses this information, the DR coordinator identifies which teams need to be called to action and which system-specific DRPs need to be executed (and in what order). The DRP should specify activation criteria for the different teams and system-specific DRPs. After the damage assessment, if one or more of the situations outlined in the criteria have taken place, then the DR team is moved into restoration mode.

Different organizations have different activation criteria because business drivers and critical functions vary from organization to organization. The criteria may comprise some or all of the following elements:

- Danger to human life
- Danger to state or national security
- Damage to facility
- Damage to critical systems
- Estimated value of downtime that will be experienced

Restoration

Once the damage assessment is completed, various teams are activated, which signals the organization's entry into the *restoration phase*. Each team has its own tasks—for example, the facilities team prepares the offsite facility (if needed), the network team rebuilds the network and systems, and the relocation team starts organizing the staff to move into a new facility.

The restoration process needs to be well organized to get the organization up and running as soon as possible. This is much easier to state in a book than to carry out in reality, which is why written procedures are critical. The critical functions and their resources would already have been identified during the BIA, as discussed earlier in this chapter (with a simplistic example provided in Table 23-1). These are the functions that the teams need to work together on restoring first.

Many organizations create templates during the DR plan development stage. These templates are used by the different teams to step them through the necessary phases and to document their findings. For example, if one step could not be completed until new systems were purchased, this should be indicated on the template. If a step is partially completed, this should be documented so the team does not forget to go back and finish that step when the necessary part arrives. These templates keep the teams on task and also quickly tell the team leaders about the progress, obstacles, and potential recovery time.

NOTE Examples of possible templates can be found in NIST Special Publication 800-34, Revision 1, *Contingency Planning Guide for Federal Information Systems*, which is available online at https://csrc.nist.gov/publications/detail/sp/800-34/rev-1/final.

An organization is not out of an emergency state until it is back in operation at the original primary site or at a new site that was constructed to replace the primary original one, because the organization is always vulnerable while operating in a backup facility. Many logistical issues need to be considered as to when an organization should return from the alternate site to the primary one. The following lists a few of these issues:

- Ensuring the safety of employees
- Ensuring an adequate environment is provided (power, facility infrastructure, water, HVAC)
- Ensuring that the necessary equipment and supplies are present and in working order
- Ensuring proper communications and connectivity methods are working
- Properly testing the new environment

Once the coordinator, management, and salvage team sign off on the readiness of the primary site, the salvage team should carry out the following steps:

- Back up data from the alternate site and restore it within the primary site.
- Carefully terminate contingency operations.
- Securely transport equipment and personnel to the primary site.

The least critical functions should be moved back first, so if there are issues in network configurations or connectivity, or important steps were not carried out, the critical operations of the organization are not negatively affected. Why go through the trouble of moving the most critical systems and operations to a safe and stable alternate site, only to return them to a main site that is untested? Let the less critical departments act as the