Ransomware has become one of the most significant global cybersecurity threats, holding individuals, businesses, and even governments hostage by encrypting critical data and systems, demanding payment for decryption keys, and further extorting their victims by publicly disclosing their data. Imagine waking up to find all your files locked and a clock counting down until you pay up - or lose everything. This chapter begins the discussion of the evolution of ransomware, tracing its roots from early crypto-viral extortion schemes to its status as a multi-billion-dollar criminal enterprise. We will blend real-world cases and academic insights and explore how ransomware operates, why it’s so effective, and what can be done to protect against it. Whether you’re a tech novice or a seasoned professional, understanding ransomware is critical in today’s interconnected world.

This chapter will take you on a journey through the fundamentals of ransomware, starting with a clear definition of what it is - and just as crucially, what it isn’t. We’ll explore its history, from the first crude attempts at holding data hostage as early as 1989 to the complex, multi-layered attacks we see today. We’ll delve into traditional encryption ransomware, the growing trend of data exfiltration, and the increasingly common tactic of double extortion. Understanding who the attackers are is just as important; we’ll break down the roles of Initial Access Brokers (IABs) and ransomware groups to shed light on the criminal ecosystem. Finally, we’ll guide you through the typical sequence of a ransomware attack, arming you with the knowledge to better understand and combat this pervasive threat.

What Ransomware Is

Ransomware is malicious software (malware) designed to block access to a computer system or encrypt its data until a ransom is paid. At its core, ransomware infiltrates a computer system—often through deceptive means like phishing emails or malicious downloads - and then encrypts the victim’s files, making them inaccessible. In more advanced attack scenarios, threat actors will use a type of ransomware capable of performing surveillance within victim systems and networks before strategically encrypting devices in a coordinated, larger scale attack that not only encrypts user data files but also impacts the organization’s ability to conduct business by encrypting critical systems – including those that support email, phones, internet, and business applications (e.g. invoicing, warehouse, payroll, etc.). The attackers, who are usually cybercriminals seeking financial gain, then demand a ransom payment, typically in the form cryptocurrency like bitcoin, in exchange for a decryption key that can unlock the files. It’s a bit like a digital kidnapping, where it’s your files or systems that are held hostage. The appeal of ransomware to cybercriminals lies in its simplicity and effectiveness; with minimal effort and cost, they can disrupt entire organizations, from small businesses to large corporations, and even critical infrastructure like hospitals, water treatment plants, and government agencies.

What makes ransomware particularly frightening is the pressures it exerts on victims, including primarily the psychological and financial stresses for individuals, and financial, operational, reputational, and regulatory for organizations. The demand for ransom is often accompanied by a countdown clock, adding a sense of urgency and fear to the situation. Victims are told that if they do not pay within a specified time frame, the ransom could increase, their data will be permanently lost, or worse it will be published online for all to see. This threat of data loss or exposure can lead to panic, pushing victims to pay the ransom even when there is no guarantee that the attackers will actually provide the decryption key, that the files will be recoverable, and that they’ll honor their word in not publishing your data.

The evolution of ransomware over the years has also made it more dangerous. Early versions were relatively straightforward, simply locking users out of their systems until payment was made. However, modern ransomware has become more sophisticated, employing advanced encryption methods that are nearly impossible to break without the decryption key. Additionally, attackers have developed new tactics, such as exfiltrating sensitive data before encrypting it and then threatening to release it publicly if the ransom is not paid - a tactic known as double extortion. This evolution has turned ransomware from a mere nuisance into a weapon of choice for cybercriminals worldwide, leading to billions of dollars in losses annually (Conti, et. al., 2018 - https://doi.org/10.1016/j.cose.2018.01.001).

True accounting for the impact and losses associated with ransomware has become more complicated over the years, primarily due to the evolution of the reconnaissance and extortion tactics we see today. For example, the Federal Bureau of Investigation’s (FBI) annual internet crime complain center (IC3) report for 2023 reported almost $60 million in losses but noted that their data only captured the reported incidents. For those that did report, the losses did not include lost business revenue or the wages, expense, and third party remediation services related to recovering from the incident (IC3.gov, 2023. Federal Bureau of Investigation Internet Crime Report 2023. https://www.ic3.gov/Media/PDF/AnnualReport/2023\_IC3Report.pdf?trk=public\_post\_comment-text). However, if you expand your research to include other perspectives on ransomware, like tracking crypto payments, you will begin to see the magnitude of the problem. The crypto-tracing firm Chainalysis published data that show over $1.1 billion dollars in ransomware payments in 2023, a huge increase from the IC3 report and almost double the payments made in 2022 (https://www.chainalysis.com/blog/ransomware-2024/).

A graph of blue rectangular bars with numbers

AI-generated content may be incorrect.

Figure 1-1. Ransomware crypto payments between 2019 and 2023 as reported by Chainalysis (Source: https://www.chainalysis.com/blog/ransomware-2024)

In short, ransomware is a form of cyber extortion that preys on the fear of losing access to critical data. Its effectiveness and ever-evolving tactics make it one of the most challenging cyber threats to combat today. Understanding how ransomware works, and the tactics used by cybercriminals is crucial for anyone looking to protect themselves or their organization from this growing threat.

**How Does Ransomware Work?**

The short answer for how ransomware works is that it infiltrates one or more systems, identifies files of interest (typically documents, spreadsheets, photos, and email), encrypts the target files, and places a ransom note on the user’s desktop and in every file folder. The longer description, however, is much more technical and includes several alternatives depending on the variant of ransomware and the threat actor that is controlling it.

Ransomware is typically developed with high-level programming languages, like C, C++, JavaScript, and Python due to their flexibility and compatibility across various computer operating systems. Some have also been found to be written in more specialized programming languages like Go and Rust, which provide greater speed, efficiency, and additional compatibility across target systems. In any case, the ransomware developer begins with the fundamental objectives of getting the malware installed on victim computers without being detected, conducting an inventory, calling home to negotiate the encryption keys, encrypting the target files, and leaving a ransom note.

Avoiding detection is an important skill for any malware developer. For ransomware, the malicious code is commonly hidden within a seemingly legitimate file or application – like a Microsoft Word file named Invoice.doc or Resume.doc. This approach is described as a trojan horse, similar to the ancient Greek history of Troy, where something appears to be one thing – tricking victims into accepting it - and resulting in something malicious. In addition to hiding the ransomware within another file, the malware is also encrypted to further disguise it from anti-virus scanners. Well, those are simple examples of avoiding detection anyway.

The more sophisticated techniques that ransomware uses to avoid detection by security systems such as antivirus software, intrusion detection systems (IDS), and endpoint protection platforms a described in the following section. These techniques have evolved significantly over time to maintain their infection success, making modern ransomware increasingly stealthy and difficult to detect. This section provides an overview of the more complex and sophisticated methods ransomware uses to avoid detection:

*Encryption and Obfuscation*

Code Obfuscation: Ransomware often obfuscates or disguises its code to make it difficult for static analysis tools to understand. Techniques like encrypting strings, renaming functions to non-descriptive names, and using complex or redundant code structures to prevent easy identification of malicious code by anti-virus programs.

Packing or Encryption of Payloads: Attackers use packer and crypter programs to compress and encrypt the ransomware payload to prevent anti-virus solutions from analyzing the code until it is unpacked in memory on the target device, which can bypass signature-based detection systems. Traditional anti-virus programs don’t typically have AI or sandboxing capabilities (i.e., dynamic analysis) to identify the malware until its activities match an existing signature.

*Polymorphism*

Polymorphic Code: This technique involves the ransomware slightly changing its code each time it infects a new system or replicates itself. These minor changes are often automated and make it difficult for the signature-based detection methods of many antivirus programs to identify the malware, as each version of the ransomware looks different on the surface, even though it has the same core functionality.

Metamorphic Code: A more advanced form of polymorphism where the entire code structure changes without altering the malware’s functionality, further complicating detection by traditional security tools.

*Fileless Malware Techniques*

In-memory Execution: Ransomware increasingly operates filelessly, meaning it runs directly in the system’s memory without writing files to disk. By avoiding the filesystem, the ransomware can bypass traditional file-based detection methods that scan for known malicious file signatures.

Script-based Attacks: Ransomware may use PowerShell, Windows Management Instrumentation (WMI), or JavaScript to execute commands directly from memory, further reducing the likelihood of detection by antivirus systems focused on scanning files.

*Exploiting Legitimate Tools*

Living off the Land (LotL): Ransomware often exploits legitimate system tools and administrative utilities like PowerShell, PsExec, or Windows Script Host (WSH) to perform malicious actions. Because these tools are commonly used in enterprise environments for legitimate purposes, it is difficult for security software to block or flag them without risking false positives.

Dual-Use Software: Attackers may also abuse widely used software (e.g., Microsoft Office macros) to deliver ransomware payloads. Since these are legitimate applications, security systems may not flag their usage unless specific anomalies are detected.

*Stealthy Command and Control (C2) Channels*

Use of Trusted Platforms: Ransomware can communicate with its command-and-control servers via trusted platforms such as Google Drive, Dropbox, Google Sheets, or Slack. Since these platforms are widely used in enterprise environments, blocking them outright could disrupt business operations, making the ransomware’s communications harder to detect.

Use of the Tor Network: Many ransomware variants use the Tor network for encrypted and anonymous communication with C2 servers. This prevents tracking and monitoring by network security tools since the traffic is encrypted and obfuscated.

*Delaying Execution*

Delayed or Timed Execution: Some ransomware delays its execution until after it has infiltrated the system to avoid immediate detection by behavioral analysis tools. This can involve waiting for a specific trigger (like a system reboot) or implementing time delays before launching malicious actions.

Dormant Phases: Some ransomware lies dormant for a period after infection to evade sandbox environments and analysis systems that monitor malware for a short duration after execution. Once the ransomware detects it is no longer in a controlled environment, it activates its malicious functions.

*Behavioral Evasion*

Environment Detection: Ransomware often includes techniques to detect if it is running in a virtual machine (VM), sandbox, or other security research environments. If it detects such an environment, it may alter its behavior or terminate its execution to avoid analysis. Common methods include checking for the presence of debugging tools, unusual memory configurations, or indicators of VM use.

User-Agent Detection: Some ransomware checks for the type of system or user configuration before executing its payload. For example, it might verify whether the target machine is a specific operating system or belongs to a specific user profile, executing only when certain conditions are met.

*Bypassing Endpoint Security*

Disabling Security Software: Ransomware may attempt to disable security features such as antivirus, firewalls, and intrusion detection systems before launching its malicious payload. This can be achieved by modifying system settings or terminating security processes using administrative privileges gained during the infection process.

Privilege Escalation: Once ransomware gains access to a system, it may attempt to escalate its privileges to the level of a system administrator, allowing it to bypass user-level security measures and execute critical system commands without detection.

*Anti-Analysis Techniques*

Anti-Debugging: Ransomware can employ anti-debugging techniques that prevent analysts from stepping through the malware’s code using debugging tools. These techniques may include inserting code that detects the presence of a debugger and terminates or alters the malware’s behavior.

Control Flow Obfuscation: The malware’s control flow can be intentionally altered to make it difficult for security analysts or automated tools to follow the logical sequence of the ransomware’s actions. This includes adding meaningless loops, jumps, and conditional statements.

*Self-Destruction and Anti-Forensic Techniques*

Log Clearing: Some ransomware variants delete logs, system event records, and other traces of their activities, making it difficult for forensic investigators to track how the infection occurred.

File Deletion or Encryption of Logs: After completing its malicious actions, ransomware may delete itself or encrypt logs, making post-infection analysis difficult and reducing the ability of incident response teams to identify the attack’s origin or spread.

By using a combination of these techniques, ransomware is able to evade many detection mechanisms, successfully carry out its malicious activities, and remain hidden from traditional security tools.

Like most malware, however, common ransomware requires interaction. I say “common” because some ransomware attacks are triggered by the threat actor, not the victim; we’ll get to those later. Getting the victim to interact with something in order to open the door for the ransomware malware to download to the victim’s device is often accomplished through well-crafted phishing emails containing the malware or a link to a website containing the malware. These phishing emails appear to be legitimate messages related to the victim’s business, family, or topic of interest, or perhaps a free offer for something cool. Simply opening the email isn’t enough to trigger the malware; it requires opening an attachment or clicking a link. If it’s an attachment, then when the victim opens the file, the computer searches for the right application for viewing the attached file type (e.g., pdf, doc, xls, txt, zip, rar, etc.), and that operation often allows the ransomware to execute. If it’s a website link, clicking it will take the victim to a malicious web address where malware could be automatically downloaded and installed in milliseconds.

Once inside your computer, the malware conducts reconnaissance, inventorying your files and determining what privileges you have on your computer, the network, and possibly cloud environments like Microsoft or Google. The malware then begins encrypting files, making them inaccessible without a decryption key, which the attacker promises to provide in exchange for payment—often in hard-to-trace cryptocurrencies like Bitcoin (Bajpai, 2017- https://doi.org/10.1007/s11416-017-0304-4). The affected user is typically presented with a ransom note explaining the situation and providing instructions on making the payment.

One of the most notorious ransomware attacks in recent history was the WannaCry attack in 2017, which spread rapidly across the globe and infected more than 230,000 computers in over 150 countries (Mohurle & Patil, 2017 - https://doi.org/10.26483/ijarcs.v8i5.4022). This attack exploited a vulnerability in Microsoft Windows, and the attackers demanded $300 in Bitcoin for each infected machine. The impact was so severe that it caused disruptions in hospitals, businesses, and government agencies.

**Case Study – Polymorphic Ransomware**

Virlocker ransomware (a.k.a. VirLock or VirRansom, and Operation Global III)

VirLock was reportedly the first self-reproducing polymorphic ransomware, identified by Bleeping Computer in 2014 (Cimpanu, C., 2017). BleepingComputer.com. VirLocker Ransomware Returns Just as Virulent as Ever. https://www.bleepingcomputer.com/news/security/virlocker-ransomware-returns-just-as-virulent-as-ever/). This ransomware targets cloud or collaboration sites where its malware can spread through file sharing. The malware works by encrypting and repackaging victim files as an executable (e.g., your Word document .doc is now .doc.exe), but also infects each file with its ransomware. On the surface, each file looks normal and could be spread to coworkers or added to backups without knowing they are infected.

When a file is infected with VirLock, it is repacked into a Win32 PE file, and the .exe extension is added to its file name - unless it is already an executable file. When a victim opens an infected file, the ransomware decrypts the original file within its body, writes it to the current directory, and opens it. This behavior is one of the distinguishing characteristics of VirLock as compared to typical file-encrypting malware.

VirLock then installs itself by dropping two non-identical polymorphic instances into the %userprofile% and %allusersprofile% directories. It also adds entries to the Windows Run registry keys to ensure these instances are executed with Windows startup. These instances contain only the virus code, with no host file to decrypt, and are launched immediately. Recent variants of VirLock also drop a third instance that is registered as a Windows system service. This mechanism provides self-preservation for the malware, automatically restoring terminated processes and deleted files.

The dropped instances are responsible for executing the malicious payloads. One thread focuses on file infection, with Win32/VirLock searching for host files across local and removable drives, network shares, and accessible cloud folders to enhance its ability to spread. The file extensions targeted for infection vary between VirLock versions, with recent samples focusing on .exe, .doc, .xls, .zip, .pdf, .jpg, and many others.

Specific to cloud-based collaboration targets, VirLock was especially successful at infecting and spreading across international boundaries. In the example of just two users (A & B) collaborating through a cloud platform with some form of file synchronization enabled. If user A gets infected with VirLock and all of their local files are infected, any file synced with the cloud folder then also becomes infected and subsequently sync down to user B’s computer. When user B opens any of the sync’d infected files, their local computer and all its files are then infected. Each infected file becomes an infector. The following [Figure 1-2](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch01.html#randsomwareer_ch1_figure_2_1739263357420858) depicts a typical corporate example where users (black dots) interact with shared files (blue dots) – all of the users within the red boundary could become infected very quickly.

A diagram of a network

Description automatically generated

Figure 1-2. Example of VirLock spreading through typical Corporate file sharing collaboration platforms (Source: https://www.knowbe4.com/virlock-ransomware)

VirLock’s ransom note varies by version and is either a straightforward threat actor demand for payment ([Figure 1-3](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch01.html#randsomwareer_ch1_figure_3_1739263357420885) - VirLock Ransom Note), or the alternative Operation Global III note that pretends to be an official notice from law enforcement related to pirated software detection and the requirement to pay a fine – in bitcoin ([Figure 1-4](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch01.html#randsomwareer_ch1_figure_4_1739263357420914) - Operation Global III Notice of Software Piracy Fine).

A screenshot of a computer error

Description automatically generated

Figure 1-3. - VirLock Ransom Note from ESET Security, 12/22/2014 (https://www.welivesecurity.com/2014/12/22/win32virlock-first-self-reproducing-ransomware-also-shape-shifter/)

A screenshot of a computer

Description automatically generated

Figure 1-4. - Operation Global III Notice of Software Piracy Fine from ESET Security (https://www.welivesecurity.com/2014/12/22/win32virlock-first-self-reproducing-ransomware-also-shape-shifter/)

What Ransomware Isn’t

It’s essential to distinguish ransomware from other forms of cyber threats like viruses, worms, or spyware. While these types of malware can cause significant damage, they don’t typically involve holding data or systems hostage for financial gain. For instance, a virus might delete files, and spyware might steal data without the user’s knowledge, but ransomware uniquely combines the aspects of encryption and extortion (Scaife, et. al., 2016). This combination makes it particularly dangerous and lucrative for cybercriminals. Here’s what ransomware is not:

*Ransomware is Not Traditional Malware*

Traditional malware, such as viruses, worms, or trojans, are typically designed to corrupt files, steal data, or disrupt systems. Ransomware, on the other hand, is specifically designed to encrypt files or lock access to systems until a ransom is paid. The key difference lies in ransomware’s primary objective: financial extortion rather than pure data destruction or theft.

*Ransomware is Not Spyware*

Spyware is designed to secretly monitor user activity and collect sensitive information, such as passwords or financial details, often without causing immediate harm to the system. Ransomware, however, takes a more direct approach by openly restricting access to a user’s data or system and demanding payment for its restoration. While ransomware may sometimes steal data (especially in double extortion attacks), its primary goal is to force victims into paying a ransom, not covert surveillance.

*Ransomware is Not Adware*

Adware is a type of software that displays unwanted advertisements to users, typically for monetary gain through ad impressions or clicks. It is often more of a nuisance than a serious security threat. Ransomware, by contrast, is a critical and immediate security threat, causing operational disruption and financial loss by blocking access to essential files and systems.

*Ransomware is Not Phishing*

Phishing is a social engineering attack that aims to trick users into providing personal information or login credentials, often through fake emails or websites. While phishing may serve as a delivery mechanism for ransomware (e.g., by tricking users into downloading a ransomware payload), the two are distinct in their purpose and operation. Phishing is focused on data theft, whereas ransomware seeks financial extortion through encryption or system lockdown.

*Ransomware is Not a Data Breach*

A data breach occurs when unauthorized individuals gain access to private or confidential information. While some ransomware attacks may result in data being exfiltrated as part of a double extortion scheme (where attackers threaten to release stolen data), the primary goal of ransomware is to encrypt data and demand a ransom. In contrast, a data breach’s goal is typically to obtain and misuse data, often without locking or damaging the victim’s systems.

*Ransomware is Not a Denial-of-Service (DoS) Attack*

A Denial-of-Service (DoS) attack aims to make a service, network, or website unavailable to users by overwhelming it with traffic or exploiting vulnerabilities. Ransomware does not aim to flood or disrupt networks through traffic but instead targets individual systems or files, rendering them unusable through encryption or locking mechanisms. While both DoS and ransomware can cause operational disruption, the intent and methods are entirely different.

*Ransomware is Not a Network Takeover Attack*

Traditional network takeover attacks focus on gaining and maintaining unauthorized control over a network through stolen credentials, exploiting vulnerabilities, or using brute force to bypass security measures. The attacker then moves laterally across the network, escalating privileges, establishing backdoors, and maintaining persistence for as long as possible. Takeovers are often stealthy and designed to avoid detection, allowing the attacker to spy on the network or steal data over time, often for espionage, long-term data theft, or to disrupt operations. The attackers might steal intellectual property and sensitive data or use the network to launch further attacks without the victim’s knowledge.

By recognizing the distinct characteristics of ransomware compared to other malware, organizations, and individuals can create more targeted security strategies, improve detection and response protocols, and prepare more effectively for ransomware-specific threats. This understanding also allows for better training, investment in the right security tools, and the creation of response plans that minimize financial loss, reputational damage, and operational downtime in the event of an attack.

A Brief History of Ransomware

Over the years, ransomware has evolved significantly. Early forms of ransomware, like the AIDS Trojan, used weak encryption methods that could often be reversed without paying the ransom. However, modern ransomware uses strong encryption algorithms, making it nearly impossible to decrypt the data without the attacker’s key (Kharraz et al., 2015). Additionally, the rise of cryptocurrencies has made it easier for attackers to demand and receive payments without the transaction being traced.

The following [Figure 1-5](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch01.html#randsomwareer_ch1_figure_5_1739263357420935) and timeline outline ransomware’s history from its humble beginnings to its current state as a global cybersecurity threat.

A timeline of ransomware

AI-generated content may be incorrect.

Figure 1-5. Ransomware timeline from 1989 to 2022 (Esparza, 2022)

*The AIDS Trojan (1989): The Birth of Ransomware*

The first known ransomware attack, the AIDS Trojan (also known as the PC Cyborg Trojan), was created by evolutionary biologist Dr. Joseph Popp. This ransomware was distributed on over 20,000 infected floppy disks to attendees of the 1989 World Health Organization’s AIDS conference. When an attendee inserted the floppy disk into their computer, it became infected as the PC read and indexed the files it contained. Upon infection, the malware encrypted file names on the user’s system and displayed a ransom note ([Figure 1-6](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch01.html#randsomwareer_ch1_figure_6_1739263357420956)) demanding $189 to be sent to a P.O. Box in Panama to unlock the system (Esparza, 2022). Although easily defeated due to weak encryption methods, the AIDS Trojan marked the beginning of ransomware by introducing the core concept of data being held hostage for payment and laying the groundwork for the sophisticated ransomware campaigns we see today.

A black and white sign with text

AI-generated content may be incorrect.

Figure 1-6. AIDS Trojan ransom note (Esparza, 2022)

*Early 2000s: Evolution of Encryption Techniques*

Ransomware evolved slowly throughout the 1990s, but by the early 2000s, attackers began using stronger encryption techniques for holding victim data hostage. These early variants still relied on weak symmetric encryption algorithms, allowing some victims to recover data without paying the ransom. However, there was soon a shift to the use of asymmetric encryption (i.e., RSA), where the encryption and decryption keys are different, making data recovery without paying the ransom much more difficult.

One notable variant from this period is Archievus, a ransomware from 2006 that used RSA encryption. Victims were required to purchase a password from the attacker’s website to unlock their files. While Archievus lacked the widespread impact seen in later ransomware campaigns, it signified the beginning of more sophisticated encryption practices that would shape future attacks.

*The Rise of CryptoLocker (2013): Ransomware Goes Mainstream*

The CryptoLocker ransomware, which surfaced in 2013, marked a turning point in ransomware history. Distributed through malicious email attachments and botnets, CryptoLocker used RSA-2048 encryption, making file decryption virtually impossible without paying the ransom. CryptoLocker demanded ransoms in Bitcoin, allowing attackers to remain anonymous and offering payment through prepaid cards. Victims who refused to pay faced the permanent loss of their encrypted data ([Figure 1-7](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch01.html#randsomwareer_ch1_figure_7_1739263357420986)).

During this time, ransom demands for payment within a window of response (e.g., 72 hours) became not only prevalent but also necessary. The urgency for paying the ransom was intended to create stress for the victim. However, it was also typically aligned with the period of time for which the attackers rented the botnet servers used in the attack. The encryption key was permanently lost once the lease on the botnet server expired.

CryptoLocker infected over 250,000 systems during its run and collected an estimated $3 million in ransom payments before being disrupted by law enforcement. The sophistication and success of CryptoLocker set the stage for future attacks, with attackers increasingly favoring ransomware due to its profitability.

A screenshot of a computer

Description automatically generated

Figure 1-7. - CryptoLocker ransom note (Esparza, 2022)

*The Advent of Ransomware-as-a-Service (RaaS)*

By the mid-2010s, ransomware had become more accessible to a broader range of cybercriminals through Ransomware-as-a-Service (RaaS) providers on dark net forums. This model allowed attackers to purchase or lease ready-made ransomware kits on the dark web, enabling less technically skilled individuals to launch their own attacks in exchange for a share of the ransom. RaaS democratized ransomware, contributing to its explosion in volume and complexity. Notable RaaS variants include Cerber and Sodinokibi / Revil, which we will cover a little further down.

*Petya, WannaCry and NotPetya (2016 - 2017): Global Outbreaks*

The Petya ransomware campaign first emerged in 2016 and stood out for its unique method of attack. Unlike traditional ransomware, which typically encrypts individual files, Petya encrypted the Master Boot Record (MBR), rendering the entire system inoperable. Victims would see a fake “CHKDSK” screen before the ransomware demanded a payment of 0.9 Bitcoin (about $400 at the time) to regain access. Petya spread through infected email attachments and initially targeted organizations across the globe. Its early versions had limited success but were a precursor to later, more destructive variants.

In 2017, ransomware reached unprecedented levels of disruption with the WannaCry ransomware attack. WannaCry exploited a vulnerability in Microsoft’s SMB protocol (used for sharing files, printers, etc.), using an exploit known as EternalBlue, which was a spy tool developed by the U.S. National Security Agency (NSA) and leaked by the Shadow Brokers hacking group. WannaCry spread rapidly, infecting over 200,000 systems in 150 countries, including major organizations like the UK’s National Health Service (NHS). The attack demanded payment in Bitcoin but was poorly executed, with many victims unable to recover their data even after paying.

Shortly after WannaCry, a highly destructive variant of the Petya ransomware named NotPetya appeared. Though initially resembling ransomware, NotPetya was actually a wiper (malware designed to destroy data permanently), disguised as ransomware. It spread using methods similar to WannaCry but aimed to destroy data rather than hold it for ransom. Maersk, the global shipping giant, suffered $300 million in losses from the NotPetya attack, showcasing the devastating financial impacts of ransomware. Overall, estimated global damages for the NotPetya campaign were over $10 billion, with many victims unable to recover their systems even after attempting to pay the ransom.

*The Era of Double Extortion (2019–Present)*

Ransomware attacks became even more aggressive in 2019 with the rise of double extortion tactics, which we will cover in more detail later. Instead of just encrypting data, attackers began exfiltrating sensitive information before encrypting the victim’s data. If victims refused to pay the ransom, the attackers then threatened to leak the stolen data publicly. The Maze ransomware group was one of the pioneers of this approach, and other ransomware operators quickly followed suit.

By 2020, Revil, Ryuk, and other ransomware gangs were regularly making headlines for targeting large organizations with double extorsion ransomware campaigns. Total global ransom payments exceeded $692 million in 2020 (Chainanalysis, 2022), a 311% increase from the previous year, driven by high-profile attacks on healthcare, education, and government sectors.

*Colonial Pipeline and JBS Attacks (2021)*

In 2021, ransomware attacks on critical infrastructure demonstrated the potential for widespread societal disruption. The Colonial Pipeline attack, which was attributed to the DarkSide ransomware group, led to a major fuel shortage in the United States (Osborne, 2021). Colonial Pipeline paid a ransom of $4.4 million, although a significant portion was later recovered by law enforcement.

Shortly after, the world’s largest meat supplier, JBS, fell victim to a ransomware attack, causing temporary shutdowns of production plants in the U.S., Australia, and Canada. JBS paid the attackers $11 million in ransom, further highlighting the growing scale of ransom demands (Morrison, 2021).

Block-chain analysis was performed by the data analytics group Chainalysis, which identified the ransomware related cryptocurrency payments made between 2016 and 2021 ([Figure 1-8](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch01.html#randsomwareer_ch1_figure_8_1739263357421008)), as well as the top 10 ransomware strains by revenue for the year 2021 ([Figure 1-9](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch01.html#randsomwareer_ch1_figure_9_1739263357421032)).

A graph of a bitcoin value

Description automatically generated

Figure 1-8. - Graph of total global cryptocurrency payments made for ransomware demands (Chainalysis.com)

A graph of a number of blue bars

Description automatically generated

Figure 1-9. - Graph of the top 10 ransomware strains by revenue for the year 2021 (Chainanlysis.com)

*Current Trends: Supply Chain and Critical Infrastructure Attacks*

Today, ransomware has evolved into a highly sophisticated threat, with attacks increasingly targeting supply chains and critical infrastructure. Supply chain attacks where attackers infiltrate a trusted supplier to compromise multiple downstream organizations became more common in 2021 and 2022, as seen in the Kaseya ransomware attack, which affected hundreds of businesses globally (Hill, 2021).

The ransomware economy has also continued to grow into a multi-billion-dollar industry, with ransom demands regularly reaching millions of dollars. According to a 2021 Sophos Threat Report, the average ransom paid by mid-sized organizations in 2020 was $170,404, with recovery costs for affected organizations averaging $1.85 million. A Palo 2022 Alto Threat report with statistics for the same year found that the average global ransomware payments across all size organizations was $312,000 (Palo Alto, 2022). In 2021, there was a 144% increase in ransom demands (and an 85% increase in the number of victims), with payments reaching $570,000 (Palo Alto, 2022). The drastic increase over 2020 was attributed to the organization and sophistication of ransomware gangs and their use of double extortion tactics. There was a 71% increase in the average ransom paid in 2022, according to Palo Alto Networks’ report (Olson, 2022), reaching almost $1 million. By 2023, the continued increase in ransomware attacks impacted approximately 5,070 victims (a 55% increase over 2022) and an average ransom payment exceeding $1.7 million. This growth was attributed to new ransomware groups, like Cl0p and ALPHV, and new vulnerabilities, including the MOVEit zero-day flaw that impacted almost 300 organizations. The LockBit gang, however, was the most impactful ransomware group in 2023, with over 1,000 victims and an average negotiated payment of $282,000 (Help Net Security, 2024). The following [Figure](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch01.html#randsomwareer_ch1_figure_10_1739263357421049) shows the top 50 ransomware strains by median payment size (Chainalysis, 2024).

A close-up of several different colored squares

Description automatically generated

Ransomware attacks in recent years have also begun to target specific industries. For example, between the years 2020 and 2023, there was a drastic increase and marked focus on the Healthcare sector ([Figure 1-10](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch01.html#randsomwareer_ch1_figure_11_1739263357421069)). Analysis of this behavior attributed the attacks to a better understanding and strategy that threat actors developed to target victims that “must pay” to maintain operations, like hospitals, and the value of extorted data if they did not pay (FBI, 2023).

A graph of a number of workers

Description automatically generated with medium confidence

Figure 1-10. Infrastructure sectors affected by ransomware (FBI, 2023)

Ransomware has grown from a niche cybercrime in the late 1980s into a global menace with increasingly destructive capabilities. Its evolution, driven by advancements in encryption, the rise of RaaS, and the adoption of double extortion techniques, has made it one of the most significant cyber threats we face today. As ransomware continues to evolve, organizations must remain vigilant, continuously train and update their user base, invest in good cyber hygiene, the ability to detect anomalous activities with their systems 24 hours a day, and most importantly – implement an effective response and recovery program to mitigate the risks posed by this ever-growing threat.

Traditional Encryption Ransomware

Traditional encryption ransomware, or crypto-ransomware, operates by using encryption algorithms to lock victims out of their data, rendering files unusable unless a decryption key is purchased. The types of files targeted by ransomware are typically those essential business operations, such as documents, media files, and databases, or personal files like tax information or family photos. Attackers try to cause maximum stress and disruption, creating a strong incentive for victims to pay the ransom. The encryption process is usually done using advanced encryption algorithms like RSA and AES, making it nearly impossible for victims to break the encryption without the attacker’s decryption key or decryptor software​.

Once the encryption is complete, the ransomware displays a ransom note, often in a pop-up or a text file on the desktop. This note details the ransom demand, typically asking for payment in cryptocurrency such as Bitcoin to preserve the attacker’s anonymity. Victims are instructed to purchase the cryptocurrency and transfer the funds to the attacker. Some ransom notes include a countdown timer, adding psychological pressure to pay quickly. If the deadline isn’t met, the decryption key is often no longer available, or the ransom amount doubles, increasing the sense of urgency.

Despite the seemingly straightforward exchange and payment, there is no guarantee that the attacker will provide the decryption key once the ransom is paid. Studies show that even when victims comply with the demands, about 8% of those who pay never receive the decryption key (FBI, 2023). Additionally, even if the key is provided, the data recovery process may be incomplete, and some files could be permanently damaged. Security experts strongly advise against paying the ransom, as doing so also encourages attackers to continue their operations, potentially funding future attacks. Supporting this advice is another statistic that over 80% of victims who pay the ransom will be extorted again within 6 months (FBI, 2023).

The best defense against ransomware is a solid backup and recovery strategy, coupled with the ability to identify and respond to malware as quickly as possible 24 hours a day (this will be a recurring statement throughout the book). In recent years, many companies have implemented ransomware-specific recovery plans, for example, combining data restoration strategies with network segmentation and regular patch management to minimize the risk of infection. Additional methods and technologies will be discussed in more detail in subsequent chapters.

Exfiltration and Double Extortion

Ransomware attacks have evolved significantly over the years, with one of the most significant trends being the incorporation of data exfiltration for use as leverage in extorting the victims to pay the ransom. In these modern attacks, cybercriminals encrypt the victim’s files and steal sensitive information before demanding a ransom for both the decryption key and the promise not to leak the stolen data. This new approach, called double extortion, heightens the stakes for victims, particularly organizations that manage confidential customer, client, or patient data. The threat of public exposure or sale of this data on the dark web adds additional pressure, often compelling victims to consider paying the ransom.

The financial and regulatory or compliance implications of double extortion can be significant. While victims may believe they can recover their encrypted data from backups, the potential damage from a data leak can be far more damaging and longer lasting. Organizations are increasingly aware that the fallout from a data breach can lead to regulatory fines, loss of customer trust, and irreparable harm to their reputation. High-profile incidents involving healthcare providers and government agencies have underscored this risk, as leaked sensitive information can affect individuals’ lives and jeopardize national security. As a result, many organizations feel they have no choice but to comply with attackers’ demands.

Despite the push for improved security measures, ransomware payments’ legal and ethical implications remain complicated. Some governments have proposed legislation discouraging or prohibiting ransom payments, arguing that compliance only fuels the ransomware economy. This has led to a contentious debate about the moral implications of paying ransoms. Organizations must weigh the immediate need for data recovery against potential legal repercussions and the ethical responsibility not to support criminal enterprises.

Collaboration among various stakeholders is essential to combat ransomware attacks employing double extortion effectively. Information-sharing initiatives between businesses, cybersecurity firms, and law enforcement agencies can enhance collective defenses against these threats. By sharing intelligence on emerging threats and known vulnerabilities, organizations can better prepare themselves to defend against potential attacks. Moreover, developing a coordinated response strategy can help minimize the impact of a successful ransomware incident, allowing affected entities to recover more quickly and effectively.

**Case Study – Engineering Company Compromised and Infected with Ransomware**

It was a typical hot October Thursday morning in Texas; already 90 degrees by 10 am. My cell phone rings: it’s the cybersecurity manager for one of the largest engineering firms in the U.S. “Morning, this is Mike”. “Hey Mike, this is “Bob” (obviously changing Carl’s name to protect his identity). We just got this with ransomware, and we need help; about 80% of all our systems appear to be infected (approx. 2,300 computers and servers), and we aren’t sure what to do about our engineering repository, but we think it’s safe. Can you coordinate a call with our leadership team to help us respond to this?”

A few quick notes on this organization before we dig in. This billion-dollar company has only 2 dedicated cybersecurity employees (a manager and an engineer) who report to the Chief Information Officer (CIO). In late August of this same year, about 5 weeks prior to this incident, I moderated a ransomware tabletop exercise for this organization. The results from the tabletop identified several areas for improvement, including enhanced ability to detect and respond to malware, ensuring their critical systems are backed up offline, and testing their ability to recover from these backups – among other observations.

My firm coordinated the Incident Response (IR) call through our Microsoft Teams because the ransomware took down their network and phone system. Client executives and the IT team were all eagerly waiting to hear us explain to them what just happened, what to do next, how to respond, and lastly, that there was hope for getting through this. We could feel despair and anxiety coming through the phone, as they had no Incident Response Plan (except for “Call Mike”), and they took no action on the observations and recommendations from the recently completed ransomware tabletop exercise. In fact, it was later determined that the threat actors had compromised this organization’s network several months prior to this incident, and likely knew about the tabletop and the recommendations for improving the environment.

On the call, we asked the typical briefing questions, like what happened, who has been involved, and what you have done so far, and you educated us on all the business and operationally critical things. During this part of the conversation, it was determined through analyzing the ransom note and attributes of the encrypted files and the .sz40 file extension that the threat actor was likely the Lorenz Group. Also, during this initial call, we suggested involving federal law enforcement, like the FBI or Secret Service (collectively the Feds), due to their experience investigating related crimes, specifically the Lorenz gang. The Feds may already have a decryptor, which could provide intel for these threat actors and aid in forensic analysis. Involving them is a good demonstration of the company’s due diligence in responding to the incident. We received approval to contact the FBI, and the incident was transferred to a field office specializing in the Lorenz gang.

The Lorenz group is a double extortion threat actor. In this case, Lorenz compromised the victim’s network three months prior to the incident, conducted reconnaissance, compromised user accounts, and became knowledgeable about their business, revenues, insurance coverage, and vacation schedules. They strategically planned the infection for this particular Thursday because they knew the backup administrator was going on vacation the night before (Wednesday) and that Friday was a company holiday with limited coverage. In addition, their persistent access up to this point also allowed them to use the victim’s own network architecture to schedule the deployment and installation of the ransomware to all computers on the network at the time. At 8am Thursday morning, the automated delivery of the ransomware began, and within 2 hours the damage had been done. Some computers were off at the time, some employees noticed something weird and turned the computer off, and the remaining devices were completely infected. [Figure 1-11](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch01.html#randsomwareer_ch1_figure_12_1739263357421088) is an example of a Lorenz ransom note that is placed on the computer’s desktop and in every folder where files were encrypted.

A black marker on a red and white paper

Description automatically generated

Figure 1-11. Lorenz ransom note

Our next suggestion was to involve their cyber insurance provider and brief them on the situation. The Insurance company quickly connected us with their Ransomware Negotiator, who coincidentally works a lot with the Lorenz group. The Negotiator informed us that this was already the 4**th** Lorenz call he’s had today (it’s now 11 am). The group’s typical behavior is to compromise, learn and assess, exfiltrate your data, encrypt, and extort. He added that their extortion tactics involved a wall of shame website publishing their victims for the world to see, publishing victim data, selling victim data, and selling the access they gained in the initial compromise. The negotiations began, and the initial volley was on Friday afternoon to ensure the victim’s response was within 2 days of infection and prevent the initial posting on the wall of shame. The starting ransom demand was $8 million, and if they didn’t receive a response by Monday, it would double to $16 million. Sunday afternoon, we responded with $575,000. I will note here that between Thursday morning and Sunday afternoon, the victim’s critical engineering systems became infected, and there were no backups for this data – one of the observations from the tabletop (i.e., listen to the experts when they suggest improvements based on experience). I believe the attackers already knew the victim’s insurance coverage and countered with $4.5 million, which was allegedly paid. Payment was agreed to primarily because the Lorenz group offers not only the decryptor (unique to each infection), but also technical support to help you use it, tweak it, and test it. The money was paid, most of the victim’s data was recovered… except for the engineering data because it was a proprietary system that the decryptor would not work on. The company finally recovered approximately 7 weeks later at a total estimated cost of $7 million, including ransom payment, expenses, new systems, and lost revenue.

Understanding the Attackers: Initial Access Brokers (IABs) and Ransomware Groups

The ransomware ecosystem has also evolved, with specialized roles emerging within cybercriminal organizations. One such role is that of an Initial Access Broker (IAB). IABs play a crucial role in the ransomware ecosystem, serving as intermediaries between the attackers and the victims. These brokers specialize in gaining initial access to vulnerable systems and then selling that access to ransomware groups. IABs often exploit weaknesses in security protocols, using techniques such as phishing, social engineering, or exploiting unpatched software vulnerabilities to infiltrate networks. By focusing on the entry point, IABs allow ransomware groups to bypass the often more challenging task of breaching well-defended networks, thereby streamlining the attack process for cyber criminals.

IABs are a very organized and profitable part of the cybercrime landscape. They typically operate on a commission basis, earning a cut of the ransom collected from victims. This creates a symbiotic relationship where IABs provide the necessary access while ransomware groups can focus on what they do best: deploying the ransomware and negotiating the ransom. The profitability of this arrangement incentivizes IABs to continuously refine their techniques and tools, which can lead to more sophisticated and targeted attacks. However, some IABs may also directly affiliate with specific ransomware groups, leading to more collaborative attacks. This relationship can enhance the attack’s efficiency, as IABs may have insider knowledge about the vulnerabilities that are most likely to lead to successful breaches.

Alternatively, IABs also operate independently, obtaining access to victim systems and determining the value of their compromise, then posting the access for sale on dark net forums. In 2024, large organizations with revenues over $1 billion were the most targeted, making up 27% of all initial access listings for sale (Bleih, 2024). Most IAB posts on the dark net are between $500 and $2,000 for typical corporate access and upwards of $10,000 for high-value targets (Bleih, 2024). The three primary types of access being sold by IABs are remote desktop protocol (RD) at over 70%, VPN access credentials at 16%, and web shell access at 13%, and the credentials for U.S. organizations far outpace those of other countries as the most targeted (Bleih, 2024).

A graph showing the most targeted countries/regions

Description automatically generated with medium confidence

Figure 1-12. Initial Access Broker postings for the sale of compromised access by country (Bleih, 2024)

As the landscape of cybercrime continues to evolve, the role of IABs will likely grow more significant. Their ability to quickly adapt and exploit new vulnerabilities makes them formidable adversaries. Organizations must remain vigilant and agile, continuously updating their security measures to counter the changing tactics of IABs and ransomware groups. By understanding the dynamics of this relationship and implementing robust defenses, organizations can reduce their risk of falling victim to ransomware attacks and protect their sensitive data from exploitation.

Understanding the Typical Attack Sequence

Ransomware attacks have become a persistent threat in today’s digital landscape, targeting individuals and organizations alike with increasingly sophisticated methods. Understanding the typical attack sequence can provide valuable insights into how these attacks unfold and the steps that can be taken to mitigate their impact. From the initial reconnaissance phase, where attackers gather intelligence on their targets, to the final negotiation and recovery stages, each phase of a ransomware attack is carefully planned to maximize the threat actors’ chances of success. By breaking down this sequence, organizations can better prepare themselves to defend against these potential threats and respond effectively in the event of an attack.

A typical ransomware attack follows a structured sequence of steps, each crucial for the attacker’s success. Here’s a breakdown of the common phases.

**Reconnaissance**

Attackers begin by gathering information about their target. This may involve blind internet queries using, scanning the target’s network, identifying vulnerabilities, and understanding the organizational structure. They might use tools to probe for open ports, outdated software, or weak security practices that could be exploited. This phase can be broken down into several key activities:

* **Open-Source Intelligence (OSINT) Gathering**: Attackers often begin by collecting publicly available information from a variety of sources, such as social media profiles, company websites, and public databases. This information can reveal details about the organization’s infrastructure, employees, and technologies used, helping attackers to identify potential weak points.
* **Network Scanning**: Using tools like Nmap, attackers scan the target’s network to discover live hosts, open ports, and services running on those ports. This helps them understand the network’s layout and pinpoint systems with unpatched vulnerabilities or misconfigurations.
* **Identifying Software Vulnerabilities**: Attackers search for known vulnerabilities in the software and hardware used by the target. They may refer to databases like the National Vulnerability Database (NVD) or exploit kits that catalog these weaknesses. Understanding which systems are outdated or poorly secured allows attackers to tailor their approach for maximum effectiveness.
* **Employee Targeting**: Social engineering plays a significant role in reconnaissance. Attackers may identify key employees, especially those in IT or executive positions, and gather information that could facilitate phishing attacks. This could include analyzing email formats and internal communication styles or even using LinkedIn to identify roles and connections that may lead to more effective deception.
* **Mapping Out Network Architecture**: If attackers can gain initial access to a lower-level system, they may use tools to further probe the network for additional information, such as Active Directory structures or domain trusts. This internal mapping helps them plan for lateral movement once they establish a foothold.
* **Third-Party Risks**: Attackers may also assess third-party vendors or partners interacting with the target organization. Supply chain vulnerabilities can be a significant entry point, and attackers often look for ways to exploit less secure vendors to gain access to their primary target.

By conducting thorough reconnaissance, attackers can effectively customize their strategies, increasing their chances of success in the later stages of the attack. Organizations that understand this phase can take proactive measures, such as tightening access controls, enhancing employee training on social engineering, and regularly auditing their security posture to mitigate the risk of falling victim to ransomware.

**Initial Access**

Once reconnaissance is complete, attackers seek to gain entry into the target’s systems. This can be achieved through various methods, such as phishing emails containing malicious attachments or links, exploiting unpatched software vulnerabilities, or leveraging compromised credentials from previous data breaches. Initial access brokers may also facilitate this step by selling access to vulnerable networks. Here are some common methods used in this phase:

* **Phishing Attacks:** One of the most prevalent methods for gaining initial access is through phishing emails. Attackers craft what appear to be legitimate messages that trick recipients into clicking on malicious links or downloading infected attachments. These emails may impersonate trusted entities, such as colleagues, service providers, or reputable companies, increasing the likelihood of a successful breach.
* **Malware Delivery:** In some cases, attackers may use malware delivery techniques such as drive-by downloads, where malicious code is executed on a user’s system without their consent when they visit a compromised website. This method can infect systems quickly and stealthily, often without the target’s knowledge.
* **Exploiting Vulnerabilities:** Attackers often exploit unpatched software vulnerabilities or misconfigurations in applications and operating systems. They may use exploit kits or custom scripts to target known vulnerabilities, gaining unauthorized access to systems. Common targets include outdated web servers, content management systems, and network devices.
* **Credential Stuffing and Brute Force Attacks:** With access to stolen credentials from previous breaches or databases, attackers may employ credential stuffing techniques to log into accounts across various platforms. Alternatively, they might use brute force attacks to guess weak passwords and gain access to critical systems, particularly those with inadequate security measures.
* **Remote Desktop Protocol (RDP) Exploitation:** Many organizations use RDP for remote access, but weak configurations and poor password practices can leave these systems vulnerable. Attackers may scan for exposed RDP ports and attempt to log in using compromised credentials or exploit vulnerabilities in the RDP protocol itself.
* **Initial Access Brokers (IABs):** IABs specialize in breaching networks and selling access to ransomware groups. This relationship allows ransomware actors to focus on deploying their malware while relying on IABs to provide them with access to the target’s systems.
* **Supply Chain Attacks:** Attackers may infiltrate a target through its third-party vendors, leveraging weaker security protocols or outdated systems to gain entry. This method can be particularly effective, as organizations often overlook the security practices of their suppliers and partners.

Employing these techniques can give attackers a foothold within the target’s network, allowing them to execute the next stages of the ransomware attack.

**Case Study: The employee’s personal account was compromised, resulting in a business compromise.**

A client for that we provide 24/7 cybersecurity monitoring for called us in September of 2024 about a data breach that involved their sensitive data being posted for sale on a breach forum. Two things were determined fairly quickly; the source of the data was a system that our 24x7 monitoring team did not have visibility into, and secondly, we identified the user whose account was compromised. Speaking to the employee, they couldn’t recall work-related incidents, like phishing emails or clicking anything out of the ordinary, they commented that their personal email was having a lot of issues. We searched their personal email address in the recent data breach repositories and determined that their personal email was compromised in April of 2024. Talking with the employee again, they indicated that their personal email issues began in July 2024, which correlated to the circumstances surrounding the company’s data breach. We asked if they used the same password between home and work, to which they responded, “No, but I store many of my other passwords in my personal email account – including my work credentials.”

**Execution and Installation**

Attackers deploy the ransomware payload after gaining access. This “access” can be literally the attackers logging in with compromised credentials or an automated script that relies on the victim’s interaction with an email attachment or a link they clicked. Then the malware dropper file executes and begins its own reconnaissance to determine what it has access to, its privileges, and the type of host it is installed on (e.g., workstation, Point of Sale, Centrifuge) in preparation for encrypting files on the infected system. Attackers may also establish a foothold by installing additional malware, such as keyloggers or remote access tools, to maintain control over the system and facilitate further actions. Here’s a detailed breakdown of the key steps in this stage:

* **Payload Execution**: Attackers execute the ransomware payload on the compromised machine after obtaining access. This can occur in several ways depending on the method of entry. If access was gained through phishing, the payload might be disguised as a legitimate file (e.g., a document or a PDF) that triggers the ransomware upon opening. If the breach occurred through vulnerability exploitation or remote desktop access, attackers may manually run the payload once inside the system. Attackers often use scripting languages like PowerShell or command-line interfaces to automate this execution, bypassing traditional antivirus software.
* **Persistence Mechanisms**: Once the ransomware is executed, attackers aim to establish persistence in the system, ensuring that they maintain control even if the system is rebooted or if the initial payload is detected. Common persistence techniques include creating scheduled tasks, modifying registry keys, or embedding the malware in system startup processes. This persistence ensures that even if part of the ransomware is removed, it can continue to spread or remain dormant until reactivated.
* **Privilege Escalation**: After initial execution, attackers often attempt to escalate their privileges to gain deeper access to the network. Privilege escalation allows attackers to move from a user-level account to an administrative or system-level account, giving them full control over the infected machine or even the entire network. Attackers may exploit vulnerabilities in operating systems, use credential-stealing tools (like Mimikatz), or crack administrative passwords to gain higher privileges.
* **Installation of Additional Tools**: Attackers often install other malicious tools alongside the ransomware to enhance their control over the network and facilitate further attacks. Commonly installed tools include:
* **Credential Dumpers**: Tools like Mimikatz allow attackers to extract usernames and passwords from the system’s memory or cached files. This is essential for lateral movement across the network and gaining access to high-value systems.
* **Remote Access Trojans (RATs):** RATs provide attackers with ongoing remote access to infected systems, allowing them to execute commands, upload or download files, and monitor user activity in real-time.
* **Cobalt Strike:** A legitimate penetration testing tool often repurposed by attackers, Cobalt Strike helps in lateral movement, reconnaissance, and payload delivery. It also supports beaconing, enabling attackers to maintain communication with infected systems.
* **Command and Control (C2) Communication**: During this phase, the ransomware may communicate with a command-and-control server (C2) to receive instructions, download additional malicious files, or upload exfiltrated data. Attackers often use encrypted communications (such as HTTPS or TOR) to prevent detection by firewalls and intrusion detection systems. These communications enable the attackers to adjust their tactics in real time based on the network environment or to receive confirmation that their ransomware is ready for deployment.
* **Obfuscation and Evasion**: Attackers often employ obfuscation techniques to avoid detection by antivirus software and other security measures during the execution phase. They may encrypt the ransomware payload or use polymorphic code, which changes its appearance with each infection. Other evasion methods include sandbox detection (where the ransomware can detect if it’s running in a virtual environment and halt execution) and process hollowing, where the malware injects itself into legitimate processes to hide its activity.
* **Lateral Movement**: Once the ransomware is installed and attackers have administrative control, they begin moving laterally through the network to infect as many systems as possible. This will be discussed further in the next section.

**Lateral Movement and Data Exfiltration**

Once the ransomware is in place, attackers often move laterally within the network to infect additional systems and servers. They may also exfiltrate sensitive data before encryption, increasing their leverage during ransom negotiations. This stage may involve gathering information on critical files, databases, and other valuable data that could be used to pressure the victim.

**Lateral Movement**

Lateral movement involves attackers expanding their control beyond the initially compromised machine by navigating through the network to access other critical systems and data. This allows them to infect more devices, escalate privileges, and locate high-value data that could be encrypted or stolen. Key techniques used in lateral movement include:

* **Credential Dumping:** Attackers often use tools like Mimikatz or LSASS process injection to extract credentials from memory, allowing them to impersonate users or escalate their privileges. With administrative credentials in hand, they can access other machines, network resources, or even domain controllers.
* **Pass-the-Hash and Pass-the-Ticket Attacks:** Using stolen credentials, attackers may carry out pass-the-hash or pass-the-ticket attacks, which allow them to authenticate on other systems without needing to know the user’s plaintext password. These attacks exploit weaknesses in authentication mechanisms within the network, making lateral movement swift and less detectable.
* **Remote Execution:** Attackers often use legitimate tools and protocols, such as PowerShell, Windows Management Instrumentation (WMI), or PsExec, to move laterally. These tools allow them to run commands or scripts on remote machines, making it easier to distribute the ransomware across multiple systems. By using legitimate tools, attackers can evade detection, as their activities may appear normal to security teams and antivirus solutions.
* **Exploiting Trust Relationships:** Attackers target systems with trust relationships, such as domain controllers, file servers, or Active Directory (AD). Compromising the domain controller gives attackers complete control over the network, allowing them to distribute ransomware more widely and target sensitive resources, such as backups, HR files, or finance systems.
* **Pivoting and Network Mapping:**During lateral movement, attackers may perform internal reconnaissance to map the network and identify additional high-value systems, such as databases or servers. They may also set up “pivot points” or use tunneling to route malicious traffic through compromised systems, which helps avoid detection by security teams monitoring network traffic.
* **Network Propagation Techniques:** Attackers frequently use automated techniques, such as worms or self-spreading ransomware, to propagate across networks. These techniques exploit shared network drives, mapped drives, or known vulnerabilities (such as EternalBlue) to rapidly infect large portions of the network. Self-propagating ransomware like WannaCry and NotPetya used this method with devastating effects, spreading quickly within minutes.

**Data Exfiltration**

As ransomware groups increasingly adopt double extortion tactics, data exfiltration has become essential to their strategy. By stealing sensitive data before encryption, attackers gain additional leverage over victims, threatening to release or sell the stolen information if the ransom is unpaid. Key techniques for data exfiltration include:

* **Identifying and Targeting Sensitive Data**: Before exfiltration, attackers typically search for and identify critical data. This may include financial records, intellectual property, personally identifiable information (PII), or health records, depending on the organization. Attackers may also steal email archives, internal communications, or legal documents, knowing these could damage the organization’s reputation if leaked.
* **Encryption and Compression:** To avoid detection during the data exfiltration process, attackers often compress and encrypt the stolen data before sending it out of the network. Tools like 7-Zip or custom scripts are used to package large volumes of data into smaller, encrypted archives. This not only obscures the contents but also reduces the size of the data, making it easier to exfiltrate without triggering bandwidth-based alarms.
* **Stealthy Exfiltration Channels:** Attackers use various methods to transfer data out of the network stealthily:
* **FTP or SFTP:** Standard file transfer protocols are often used to send stolen data to attacker-controlled servers.
* **Cloud Storage Abuse:** Attackers may use popular cloud storage services like Google Drive, Dropbox, or AWS S3 buckets to upload stolen data, knowing that these services are commonly allowed by organizations’ firewall rules.
* **DNS Tunneling:** By embedding stolen data within seemingly normal DNS queries, attackers can bypass detection by security systems. DNS tunneling allows data to be exfiltrated in small chunks over time without raising suspicion.
* **HTTP/HTTPS Traffic:** Many attackers leverage HTTPS or custom command-and-control (C2) channels that use encryption to hide exfiltrated data within regular web traffic. This can make it difficult for security solutions to detect abnormal activity, especially when encrypted traffic is not inspected closely.
* **Disabling Security Mechanisms:** To further ensure successful data exfiltration, attackers may disable or evade security measures like data loss prevention (DLP) systems, intrusion detection systems (IDS), or firewalls. This could involve using administrative privileges to alter logging configurations, silencing alerts, or exploiting blind spots in network monitoring tools.
* **Timing and Throttling:** Attackers often conduct exfiltration in small increments over an extended period, known as data throttling, to avoid raising suspicion. By sending data in smaller chunks, they evade network monitoring tools that may flag large or unusual data transfers. Attackers may also execute exfiltration during off-hours, such as late at night or during holidays, when security teams are less likely to notice the abnormal activity.

**Combining Exfiltration with Encryption (Double Extortion)**

After data exfiltration is complete, attackers typically encrypt the victim’s files. This combination of stolen data and encrypted systems gives the attackers immense leverage in double extortion attacks. They can demand two types of ransoms:

* Payment for decryption to restore access to the victim’s systems and files.
* Payment to prevent data leaks, where they threaten to release or sell the stolen data if the victim refuses to pay.

This strategy forces organizations to face both operational disruption and reputational damage, as leaked data could lead to legal consequences, regulatory penalties, or loss of customer trust.

**Ransom Demand**

After the encryption process is complete and any sensitive data has been exfiltrated, attackers present their ransom demand. Victims receive a message detailing the ransom amount, typically in cryptocurrency, along with instructions on how to pay. The threat of data leaks or public exposure is often used to compel the victim to comply. If the ransom is not paid within a specified timeframe, attackers may threaten to release or sell the stolen data. A breakdown of how attackers execute this phase includes:

**Ransom Note Delivery**

Once the encryption process is complete, victims receive a ransom note or message that outlines the attackers’ demands. This message is typically delivered in one of the following ways:

* **Pop-up Messages:** The ransomware displays a prominent message on the victim’s screen, blocking access to the system and making it clear that files have been encrypted.
* **Text Files or HTML Files**: The ransomware may leave ransom notes as text or HTML files scattered across affected folders. These files provide instructions on contacting the attackers and paying the ransom.
* **Desktop Backgrounds**: Some ransomware variants go further by changing the victim’s desktop background to the ransom message, ensuring visibility when the user attempts to log in.

**Ransom Note Contents**

* **A Description of the Attack**: Attackers briefly explain what has happened, stating that the victim’s files have been encrypted and/or sensitive data has been stolen.
* **Payment Instructions**: Attackers provide specific instructions on making the payment, typically in cryptocurrency (e.g., Bitcoin or Monero), which offers anonymity for the attackers. They may also direct the victim to dark web portals where further communication can occur.
* **A Deadline**: To increase urgency, attackers often set a strict deadline for the ransom payment. The ransom note may include a countdown, and failure to pay within the specified time can lead to:
  + The ransom price doubles.
  + The decryption key is destroyed, resulting in permanent data loss.
  + The stolen data being publicly released or sold on the dark web (in double extortion attacks).
* **Proof of Decryption**: Attackers sometimes offer to decrypt a few files for free as proof that they can restore the victim’s data upon payment. This is intended to convince the victim that paying the ransom is a viable solution.

**Ransom Amounts**

The ransom demand is often carefully calculated based on several factors:

* **Size of the Organization**: Larger organizations are typically targeted with higher ransom amounts, often in the millions of dollars, as attackers know that the financial loss from downtime and data breaches can far exceed the ransom cost.
* **Industry:** Certain industries, such as healthcare, finance, and government, are more likely to be targeted with higher ransom demands due to the critical nature of their data and operations. For example, healthcare organizations are highly vulnerable because encrypted patient records can disrupt medical services, which increases the pressure to pay quickly.
* **Impact of Data Loss:** In double extortion attacks, if attackers have exfiltrated sensitive data, they may adjust their demands based on the potential damage caused by a data leak. The more valuable or sensitive the stolen information (e.g., financial data, trade secrets, customer information), the higher the ransom amount.
* **Precedents:** Attackers may base their demands on successful payments made by similar organizations in previous attacks, which are sometimes publicized in news reports or hacker forums.

**Cryptocurrency Payments**

Ransom payments are almost always requested in cryptocurrency, as they offer attackers greater anonymity and are difficult for law enforcement to trace. The process of paying the ransom usually involves:

* **Setting Up a Cryptocurrency Wallet**: Victims are instructed to create a cryptocurrency wallet and obtain the required amount of cryptocurrency, often Bitcoin or Monero. Attackers often provide detailed guides on how to do this, knowing that many victims may be unfamiliar with cryptocurrencies.
* **Payment through a Dark Web Portal**: Attackers typically use a secure dark web portal to manage the transaction and further communication. These portals may be accessed via the Tor network, ensuring encryption and anonymity. The portal will provide a unique payment address and track the payment progress.
* **Escrow or Payment Verification**: Once the ransom is paid, attackers may either release the decryption key immediately or use an escrow-like system where the payment is verified before the key is delivered.

**Double Extortion Threats**

In modern ransomware attacks, especially those employing double extortion, the ransom demand often comes with a secondary threat: releasing exfiltrated data. Attackers will threaten to publish or sell stolen sensitive data if the ransom is unpaid. These threats increase the pressure on victims to comply because the consequences extend beyond encrypted files:

* **Reputational Damage:** A public data leak can severely harm an organization’s reputation, especially concerning sensitive customer information or intellectual property. This can lead to long-term financial losses, regulatory penalties, and lawsuits.
* **Regulatory Fines:** For organizations that handle regulated data (e.g., healthcare records or financial information), a data breach can result in substantial fines under laws such as GDPR or HIPAA. The threat of data exposure in ransomware attacks forces organizations to weigh the ransom cost against potential legal repercussions.
* **Customer Trust:** Leaked customer data can damage trust in the organization, leading to loss of business and erosion of brand loyalty.
* **Ransom Discounts**: In some cases, attackers may offer “discounts” if the victim pays within a certain timeframe, further increasing the urgency and pressuring the organization to make a quick decision.

**Negotiation and Payment**

After a ransom demand has been issued, victims of a ransomware attack must decide whether to negotiate, pay the ransom, or attempt recovery through other means. The negotiation and payment phase can be complex and risky, as it involves communication with cybercriminals whose primary language is usually not English, potential financial losses, legal concerns, and operational consequences. This phase often marks a turning point in the victim’s response strategy, and the approach taken can have long-term effects on the organization’s security posture and reputation.

**Decryption and Recovery**

If the ransom is paid, the victim receives a decryption key, though there is no guarantee that it will work or that all data will be restored. Organizations may also face the arduous process of restoring systems from backups, if available, and strengthening their security to prevent future attacks. Post-incident, organizations often conduct thorough investigations to understand the attack vector and improve their defenses.

**Restoring from Backups**

If the ransom is not paid, the victim organization must rely on data restoring from backups in order to return systems to a fully operational state. In this scenario, every infected system must be rebuilt to ensure no traces of the ransomware remain. Key considerations during this step include:

* **Backup Integrity**: Before restoring from backups, organizations must ensure that they are clean and have not been infected by ransomware. To avoid contamination during the attack, it’s essential to have offsite, offline, or segregated backups.
  + **Testing Backup Recovery**: Regular testing of backup recovery processes can help avoid unforeseen issues during restoration, ensuring that backups are reliable and can be restored quickly.
* **Digital Forensics**:In some cases, digital forensics can be used to extract files from infected systems. If the infected system was powered off before the ransomware completed its infection, there may still be uninfected and unencrypted files that can be safely extracted. These files should still be scanned with anti-malware tools to ensure their integrity is intact.

Conclusion

Ransomware is a sophisticated and evolving threat that has caused significant damage across the globe. What began as simple encryption malware designed to lock victims out of their data has transformed into a highly sophisticated and multifaceted attack method. Modern ransomware attacks are no longer just about encrypting data; they often involve double extortion schemes, where attackers exfiltrate sensitive data and use the threat of public exposure as additional leverage to demand payment. Understanding the complexities of ransomware, as well as what ransomware is not (e.g., spyware or adware), provides the foundation for comprehending this growing menace.

The evolution of ransomware from simple encryption tactics into more dangerous double extortion attacks marks a significant shift in the cybercrime landscape. Threat actors continued to mature and evolve their tactics to maintain their success and monetary objectives. They no longer simply encrypt files; attackers now steal data, threatening its release if the ransom isn’t paid. This dual threat significantly raises the stakes for victims, especially organizations that handle sensitive data like personal information, health or financial records, or proprietary corporate information. Exfiltration is a key component of this extortion strategy, and understanding how data is extracted from compromised systems is essential for developing effective defenses.

A key player in the ransomware ecosystem is the Initial Access Broker (IAB), a threat actor who gains access to vulnerable systems and then sells that access to ransomware groups. By leveraging the skills of IABs, ransomware groups can execute attacks quickly and efficiently, making the ransomware-as-a-service (RaaS) model more accessible and profitable. Understanding these actors, their motivations, and how they collaborate offers crucial insight into the broader ransomware landscape.

The typical attack sequence for a ransomware incident follows a series of strategic steps that begin with reconnaissance, gaining initial access, followed by the execution of the ransomware payload, lateral movement within the network, data exfiltration, and finally, the ransom demand. Each phase of the attack is carefully orchestrated to maximize the potential payout for attackers and the disruption for victims. By understanding this sequence, organizations can better prepare their defenses, detect attacks earlier, and minimize damage.

Ransomware is a dynamic and evolving threat requiring constant vigilance and a comprehensive understanding of its technical components and the broader criminal networks driving these attacks. As ransomware continues to evolve, organizations must invest in advanced security technologies and stay informed about the tactics and motivations of attackers. By doing so, they can build resilient systems that are capable of withstanding the challenges posed by this ever-growing cyber threat.

Your Backup System Is Under Attack

If you learn nothing else while reading this book, please learn this: your backup system is in grave danger. Not only is it likely that it will be unable to perform its most vital function when you are attacked by ransomware. Not only is it also possible that it may be the reason that your organization ultimately ends up paying the ransom (because data was exfiltrated). You need to understand it’s quite possible that the backup system may be the attack vector through which the ransomware attack happens in the first place!

Did I get your attention? I hope I did, because the only thing worse than knowing that you’re not able to help during any events that would disable your organization, is knowing that your backup system was the ultimate cause. You are patient zero. You are the one that let the bad guys in the back door.

Why is the backup system under attack?

The real question is why wasn’t the backup system always the attack target? Backup systems have been a vulnerable part of the data center for a long time, so it’s surprising that it took threat actors this long to finally pay some attention to them. Once a threat actor understands just how vulnerable the typical backup system is, it may become the most interesting target in the environment! Yes, just like all security reporting, me acknowledging this in print will make it more true. Bad guys read O’Reilly books, too, right?

The full answer to the question of why backup systems are under attack is quite simple. The initial reason that threat actors began targeting the backup system was that a good backup and disaster recovery system was the number one reason their victims did not pay a ransom. If you have a good backup and DR system and know that you can easily recover your entire environment within the time period your organization has agreed to, why would you even consider paying a ransom?

**Let’s talk about RTO and RPO**

I need to make sure that we are all on the same page with regard to recovery time objective (RTO) and recovery point objective (RPO), and their related friends recovery time actual (RTA) and recovery point actual (RPA).

When designing a backup or disaster recovery system, you must first decide what requirements it is going to meet, and the most critical requirements are going to be RTO and RPO. The *recovery time objective*, or RTO, is the amount of time that we all agree that it should take to restore a system to working order. This isn’t just restoring the data, or even getting a database fully up and running. This is looking at an entire application and saying that the entire application should be able to accomplish its function again within a particular amount of time, such as two hours, 12 hours, or even two days. What that value is in your environment is entirely up to your organization and how much it will be negatively impacted by a particular application being down for a particular period of time.

Let’s consider two ends of a spectrum to help illustrate this point. I once worked with a paper mill and we were trying to determine what their RTO was. They said that their computer systems were not critical to the core function of the business, which was to produce paper. They felt that they could be down for weeks without any significant impact to the business, so their RTO was two weeks.

On the opposite end of the spectrum is a financial trading firm I worked with. They felt that they were losing millions of dollars every hour that they were down, so for them setting an RTO of one hour seemed totally reasonable. Obviously, since this place was essentially printing money, they had the money to pay for a system that could easily recover their environment within an hour.

Related closely to RTO is *recovery point objective*, or RPO. This is the amount of data everyone agrees is acceptable to lose in some sort of outage – as measured by time. In other words, we’re agreeing that we can lose 24 hours’ worth of data, or one hour’s worth of data. Again, thinking about the two businesses I diskussed in the previous paragraphs, the paper mill had a much looser RPO than the financial trading firm. The financial firm actually wanted to set their RPO at zero, meaning that even in the worst of outages, they should not lose any data. This is possible with certain technologies like synchronous replication, but it is obviously expensive. They didn’t seem to mind.

Closely related to RTO and RPO are RTA and RPA. RTO and RPO are the *objectives*, or requirements that we are going to try to meet in an actual outage. RTA and RPA are what the system is currently actually *capable* of doing. This allows conversations that sound something like this: our RTO is four hours, but we are nowhere near that. Our RTA is more like 24 hours. We will need to significantly rearchitect our system to have our RTA be anywhere near our RTO of four hours.

To return to the topic of ransomware, let’s think about how a typical ransom demand is given. You are told to pay $10 million within 72 hours, and if you don’t, the $10 million becomes $20 million. This is the threat actor wanting you to think that you really need to pay the ransom now, rather than attempting a recovery that would likely take longer than 72 hours.

But if your RTO and RTA are under 24 hours, why would you pay the ransom? This might even be true even if your RTA is nowhere near your RPO, if we’re talking about a ransomware attack. For example, let’s say you are able to easily restore your systems well within your RTO. Even if that’s the case, it is unlikely that you will be able to successfully remove the ransomware from your environment *and* restore your systems within that amount of time. (There will obviously be more about this during the section on what you actually have to do when responding to a ransomware attack.)

Even if you’re unable to meet your RTO (when you include the amount of time it takes to clean the environment of malware) you still might decide not to pay the ransom for many reasons, such as the ones we will diskuss in chapter x. Obviously, if you’re going to do this, you must be absolutely sure that you can successfully restore your entire environment. Because if you tell the threat actors to go pound sand, it’s going to be very difficult to go back to them two or three weeks later after you have cleaned the environment of malware and then realized that you’re unable to recover.

To once again get back to the question at hand, the mere fact that your backup and disaster recovery systems offer you the option of not paying the ransom is the number one reason threat actors began targeting these systems in the first place. Ransomware threat actors want to disable your backup and disaster recovery system so that it cannot be used against them in an attack.

However, there is more to why they are very interested in your backup and DR system. They have also diskovered the backup system is a very valuable resource in conducting a *double extortion attack*. As diskussed in [Chapter 1](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch01.html#randsomwareer_ch1_what_is_ransomware_1739263357444487), a double extortion attack is one where they have both encrypted and exfiltrated your data. You must pay the ransomware or you will never get your data back, and your sensitive data will be exposed to the world. These two “features” of the backup system have moved from being just another computer system in your computing environment to the most popular person at the party.

The 2024 Crowdstrike threat report mentioned that threat actors have adapted the methods that they use to gain initial access due to the existence of endpoint detection and response (EDR) sensors in many places. They are now beginning to target edge devices that they believe are either unmanaged or possibly on monitored. Is it any surprise that backup storage is on the short list of types of devices that they’re trying to attack? They believe – justifiably so – that the likelihood of it being monitored is much lower than other systems.

Let me say it again: your backup system is in grave danger. It has become a prime target for threat actors because they see it as both public enemy number one and simultaneously their greatest ally trying to successfully extort money from you. Let this serve as a wake-up call to you and the cybersecurity team.

**They want to disable the backup/DR system**

As I previously mentioned, if you are able to successfully restore your environment from backup, you’re much less likely to pay the ransom; therefore, the first goal of the threat actor is disabling your backup system. They want to injure your star quarterback so he cannot throw the winning touchdown. They want to “sweep the leg” so Johnny is unable to win the karate tournament. If they can disable your backup system, it will be useless in responding to their ransomware attack.

Yes, they understand that the moment they disable the backup system, they are showing their hand. At this point, they have moved from the reconnaissance phase to the attack phase. However, before they disable it, they will definitely perform reconnaissance on your backup system to see if and how they will be able to disable it when the time comes.

They will determine what type of backup software you are running, which will allow them to then immediately tailor their attack for that backup system. They know the vulnerabilities and the strengths of different backup packages, as well as the vulnerabilities of different versions of those packages. So once they determine you’re using a particular backup system, they’re going to see if they can determine what version of it you’re running. They’re also going to see whether or not you have configured it in a secure way or a non-secure way.

Their attempts to disable your backup system take a variety of forms. The easiest way to take your backups out of the picture is simply delete them. If they can directly access and delete backup files on disk, they are home free. These people are smart and will be able to figure out if you have multiple copies of your backups, and they will attempt to delete all copies.

However, since this is the ransomware world, what they’re more likely to do is to encrypt your backups just like they do other servers. If they can gain sufficient operating system-level privileges to do so, they will probably encrypt the backup files (if they can access them), or they might choose to encrypt critical files on the operating system, which would cause the backup system to simply become inoperable. This creates a catch 22 situation where you cannot restore your environment without your backup server but you cannot restore your backup server without the decryption key.

They might also attempt to gain administrative control of the backup software. (It’s quite normal to have an authentication system for the backup system that is separate from the backup server.) If they can do that, their day just got better. If they have administrative control over the backup software, they can easily delete all your disk backups by simply setting the expiration date of the backups to today’s date. Expiring backups deletes them out of the backup catalog, and if the backups happen to reside on disk, it will delete them from the disk.

If your backups are on tape, deleting them this way would not actually delete the backups, since they still reside on tape. However, if you have administrative control, you also have the ability to electronically relabel the tapes. This doesn’t mean physically changing the paper label on the front of the tape; it means putting a new digital label on the front of the actual tape itself. Backups always write to tape in a sequential manner, and the last file on the tape always places an end-of-data marker at the end of the last file. By putting a new digital label on the front of the tape, you place an end-of-data marker after that label, which renders the rest of the data on the tape unreadable.

Although it is entirely possible to easily relabel all of the tapes in a tape library, you can take some solace in the thought that this is probably unlikely to happen for two reasons. The first reason is that this is a ransomware attack and they want you to pay the ransom, and rendering all of your backup tapes completely useless isn’t going to help their cause. (Unlike encryption, there is no turning back from relabeled tapes.) The second reason that they are unlikely to actually erase all your tapes this way is that this requires even more specialized knowledge that they probably do not have. But it is important to understand that this is a danger. As we will diskuss in later chapters, this is why we are starting to talk more and more about immutable storage options for backups.

**The backup system as an attack surface**

In addition to the backup system being a target because of its ability to restore your data, it’s also important to understand that the backup system itself is also an additional attack surface. This means that the backup system is another area of vulnerability itself, meaning that it can also assist the attackers in furthering their goals.

**Exfiltration**

As we diskussed in [Chapter 1](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch01.html#randsomwareer_ch1_what_is_ransomware_1739263357444487), double extortion attacks that include both encryption and exfiltration of your data are becoming increasingly common. Outside of the backup system, exfiltration is easy to detect, and we give some advice on how to do that in Chapter X. But if you have administrative control over the backup system – or at the very least can impersonate the owner of a particular data set – you can easily transfer any amount of data that you’d like to without setting off any alarm bells. Think about it; *the backup system’s sole purpose is to transfer large amounts of data to other systems.*

If a threat actor would like to steal a customer database, they can simply tell the backup system to restore the database to a system that they control. This could be a VM in the cloud, or simply a file system that they have access to anywhere outside the victims firewall. Just like most environments do not monitor outgoing traffic, almost no one monitors the backup system for restores! An attacker that has the appropriate access to the backup system would be able to kick off a gigantic restore and no one would notice.

Think about this. Not only is the backup system this wonderful tool that you can use for exfiltration of data without anyone noticing, it is also the central repository for all valuable data. If you want to exfiltrate data, why hack dozens of systems when all you need to do is hack one and use that system to transfer any other system’s data that you would like to wherever you want it to go?

**Reconnaissance**

Suppose a threat actor would like to run a number of attempts to take control of the system in your environment. Perhaps they would like to run brute force password attempts without being noticed. Perhaps they would like to attempt to exploit possible vulnerabilities that a given system might have. What if they were able to boot your system to alternate media to give them read access to everything on most filesystems that you would find on such a system?

Of course they wouldn’t want to try any of this on a live system inside your environment, right? They would set off all sorts of alarms and be immediately detected. However, if they’ve got control of your backup system, they can easily restore the VM of whatever server they would like to attack to a network outside your firewall, boot that VM, and attack it all they want without setting off any alarms. They could even keep a copy of it to be used to restore again if any of their attempts to hack it lock up the VM. The ability to restore any server to outside the environment is an incredibly powerful one that is absolutely a tool in the threat actor’s arsenal.

**Privilege escalation**

If a threat actor has administrative access over the backup software, they can also use it to enable privilege escalation attacks. A privilege escalation attack is necessary when a threat actor has access to a system, typically via stolen credentials, but the account that they have access to does not have enough privileges to allow them to do what they want. The threat actor would want to *escalate* the privileges that they have, which would give them the power to do what they actually want to do. Let’s say you have the ability to log into a basic, non-administrator account, and what you need is an administrator account. To get from one to the other you need a privilege escalation attack.

You might be wondering how a backup system could be used to do this. It’s actually quite simple; remember that a backup system can read and write any file on any system in your environment that has been configured for backup. Perhaps you might want to backup the security account manager (SAM) file in Windows or the password file in Linux, edit it and change the password to NULL, back it up again, then restore it in place.

But with most backup systems, it’s actually easier than that. Create a script that can be run on the command line that gives you the privileges you want, back that script up, restore it to the target system, and then tell the backup software product that it is a pre-backup script that must be run before each backup. When you run a backup, it will run that script as the backup service account, which has administrator access. The threat actor just gave themselves whatever access they wanted.

**File access**

Even if they don’t have administrative access to the backup system, perhaps they might be able to exploit the backup service account to access data they’re trying to steal. To understand what I’m talking about, let me first explain what a service account is and what’s so special about the backup service account.

A service account is an account that was created to allow another process to access either data or another process on the computer upon which it is created. For example, backup software often communicates via the backup service account, meaning an account called *backup* that is used for this purpose. The backup service account is a rather unique one because it is specifically designed for a process that is going to access hundreds of thousands of files every night. Therefore, the backup service account allows you to access any file on the system without logging said access. If the service account that you use for backup is not configured in a secure manner, it might allow a threat actor to actually use that service account to access files as if they were the backup system – and no one would know.

**Listen to the hacker**

Duanne Laflotte is an offensive cybersecurity expert, or what we used to call a white-hat hacker, or an ethical hacker. He is hired by organizations to test their security. I recently had the opportunity to talk to him about this issue on my podcast, The Backup Wrap-up. This is a slightly edited transcript of that conversation.

*W. Curtis Preston What do you know about backup systems as an attack surface?*

Duane Laflotte: We were doing a penetration test two weeks ago in an organization where we breached it over the backup system. So they were all virtualized, of course. They were backing up all of their VMs and we got access to the backup manager because the password for the backup manager was weak. It was actually default passwords. because people think to themselves, it’s a backup manager, what do I care? What are they gonna do, restore it? And that’s what we did. We actually took the backup of the domain controller and pulled it over the internet to us and restored it in my own lab. And then were able to tear it apart, pull every single username and password.

So I would be careful. That repository is just as sensitive as your primary network. It’s not only your path to recovering from disaster, but from an attacker. I’m always looking for backup systems, and what I can pull out of those systems, right?

Backup accounts should have strong passwords and should be audited. Backup systems should be audited for who’s trying to log in, et cetera. Backup service accounts that are running on boxes, we’ve seen far too often, just have weak passwords. And it’s super easy for us to then compromise.

Backup is awesome, actually. The backup service on Windows gives you the ability to read any file without being audited. So you have all these auditing tools looking for users like reading files and opening secure files. But if you can request the service account *backup*? You can touch anything and nobody ever sees it. So from a surface of attack standpoint, backups are like a win button for us. We’re always looking for, Hey, do they have a backup system? Is there an account we can compromise that has backup rights? Because if so, you know – money – we can go open any file we want and nobody will know we were there. So yeah, I would absolutely say the surface of attack is large there.

So many cool things you could do. Privilege escalation from ransomware can be done through backups. I mean, there’s so many cool things!

How could this happen?

If you’ve been paying attention the last few paragraphs, you’re probably a little freaked out right about now. You might be asking the question how is it possible that such a powerful system is not protected with an immense amount of security and monitored for the incredible vulnerability that it is? I feel your pain. In fact, I have felt your pain for the last 30 years. The answer to this question is a combination of human nature and people simply not thinking deviously enough.

Backup, recovery, and disaster recovery have always been extremely difficult, thankless jobs. No one remembers the millions of backups you got right – they only remember the one restore you got wrong. What this translates into is that nobody wants the job.

This means that very few people are going to be thinking about backups when you are in important meetings. No one is going to raise their hand and ask important questions about whether or not a new system is being backed up, or whether or not the backup system is using modern cybersecurity methods. You know why? Because they know the answer is likely that no one has any idea, so why don’t you look into that? Next thing you know you are in charge of a giant backup modernization product – the last thing you want.

This also means that the backup job is often given to the junior person. Just after the conversation recorded in the transcript above, Duane and I both said that backup was the first job that we got. Backup is often the first job that many people get in IT, because it’s the hot potato that no one wants to hold, but the new person will take the job because they want a job. And then as soon as they get any type of seniority and you hire a new person, they will pass the job to them.

I got my first job in backup in 1993 because Ron Rodriguez wanted to become a regular system administrator, so that meant giving the job to me. I, of course, am incredibly thankful for that opportunity, especially since it launched a great career. But the reality remains – I was given an incredibly powerful position at a $35 billion bank with only a few months of UNIX experience.

The junior person running the backup system is rarely trained in cybersecurity. They probably get the same cybersecurity training that all new employees get, but they do not get the specialized type of training that they should be getting as the administrator of one of the most powerful systems in your computing environment from a cybersecurity perspective.

Since no one wants to volunteer themselves to look into the backup system, and the person administering the backup system is usually a junior person, the result is a system that is often overlooked by the typical cybersecurity processes found in your organization. Hopefully this book will help to change that. Hopefully a cybersecurity professional is reading this and realizing that someone better take a look at the backup system and the person running it. Perhaps the backup administrator is reading it and realizing that they’re in way over their head from a cybersecurity perspective and can gain a lot of organizational brownie points by bringing this to the attention of the cybersecurity team.

**Cybersecurity wasn’t part of the plan**

People that have grown up in today’s cybersecurity world may look upon this whole scenario and look down upon those who wrote backup software and those of us who have used it. It’s important to understand that no one was really thinking about cybersecurity like we do today.

Backup software was way less secure than it is today, and required you to login as root or Administrator to use it. Before commercial backup software, the only way we could get centralized backup done was to allow running commands as root on other servers without a password! We used a command called RSH, and we configured it in such a way that the backup server root account could run whatever command it wanted on each of the systems that it was to backup. It was literally the only way we can get the job done from a central place.

I remember configuring backups in the late 90s at a major company that is now a household name. They had the ability to RSH and RCP (remote shell and remote copy) from any server to any other server in the world! Just think about that. Cybersecurity simply wasn’t part of the plan.

Even commercial backup software had giant security holes in it that would shock modern day users. The best example I can think of is a major commercial backup software package that had a command that would allow you to use the protocols built into the tool to easily copy any file from anywhere to anywhere. As mentioned previously in this chapter, you could do the same thing by backing up a file, editing it, backing it up again, and then restoring it to another server. This command did all of that in one step! That command lived on well into the 21st century before they finally took it out.

When I think about this, I reminisce about security on planes and how we handled things like hijackers back in the day. Hijackers would be allowed to do just about anything they wanted to, including landing planes and taking off, as long as they held hostages at gunpoint. Then 9/11 happened. For the first time the plane itself was used as a weapon. The hijackers were more than willing to sacrifice their lives, along with everyone else on board, in order to accomplish their purpose. Once that happened, we began to treat hijackers completely differently.

Backup systems have now had their 9/11 moment. You can easily see that they are a huge vulnerability and should be given the cybersecurity attention they deserve. Don’t let this valuable tool become a weapon that is used against you.

**Disk backups made it worse**

Over the last 20 or so years, the entire computing industry has moved from tape to disk as the main target for backups. I was and am a huge proponent of this change; however, it’s also important to acknowledge that moving from tape to disk for backups made the cybersecurity problem worse.

Backups stored on tape are simply harder to get to from a cybersecurity perspective. Manipulating a tape library, and reading and writing from a tape, is specialized knowledge that the typical system administrator or hacker might not have. Being a sequential medium also means that it takes a long time to interact with thousands of backups on hundreds of tapes. Each tape is going to take a few minutes to load before you can do anything with it, such as erasing it by writing a new label. If anyone was near the tape library, they would hear all of the tape movement and begin to wonder what was going on. Perhaps this would allow them to stop a cyber attack that was going after the backup tapes.

Contrast this to backups stored on a file system. They are simply files just like any other files, and they can easily be accessed, deleted, or encrypted by anyone with the appropriate credentials. You could easily impact thousands of backups in a matter of seconds or minutes without anyone having any idea what you’re up to – especially considering that no one is probably going to even see what you’re doing. You can inflict a lot of damage in a very short period of time and all you need is a privileged account on the backup server, or the storage system being used as a backup target.

**Today’s vendors often exaggerate**

Some backup vendors are exacerbating the problem by exaggerating their cybersecurity capabilities. As you will see in later chapters of this book, we will be recommending the use of truly immutable backup storage. We want at least one copy of your backup to be stored on a device that even you cannot override – even if you are a superuser and have a really good reason.

Having your backups on truly immutable storage ensures that you will have at least one copy available to you even after the worst cyber attack. This is why many people are starting to ask their backup vendors whether or not they have immutable backups. Virtually all of them are now saying yes, but what they are calling immutable is often a variety of things. All of the following types of backup storage are being called immutable. Remember that outside of computing circles, the word immutable simply means cannot be changed.

*Restored/Retrieved file is same as it was restored*

This is what I would call “old school immutability.” Someone familiar with a particular backup system could testify in court that the backup system is immutable, and that you can prove with a chain of custody and various logs that the file that was restored is exactly the same as the file that was backup. It is not possible to change it without it being noticed. However, that doesn’t mean that someone can’t modify the backed up item: it just means that if they modify a backup you will know.

*Backups cannot be directly modified (unless you have root)*

There are products that support the use of the immutable flag in the Linux file system, which means that when they write backups to this file system, the software tells Linux that the file is immutable and that it cannot be deleted until the assigned expiration date, which the system would specify a time of creation. Again, this sounds great. And again, it *is* great, but it has limitations. The immutable flag in Linux is changeable *if you have root*. If the threat actor wants to delete backups on this system, and they have root, they can simply unset the immutable flag and then delete or encrypt all the files they want.

*Backups cannot be directly modified/encrypted/deleted*

Some backup software products make the claim that once they have written backups to storage, they cannot be directly modified or encrypted via that storage. This means that even if you login with a privileged account on the backup server, you would not be able to directly access the files and encrypt or delete them. This sounds great, right? It is good, but it’s also important to understand whether or not the backup system itself can modify those files. If you can log into the backup administrator account and expire those backups (as mentioned previously in this chapter), then so could a threat actor if they gain control of the backup system.

*Backups cannot be modified even by privileged account*

The only functionality that I would consider truly immutable is one where backups cannot be deleted by anyone at any time until they have expired. That includes privileged accounts like root or Administrator. That includes people with administrator accounts within the backup software itself. This includes the customer calling in to the backup software company and saying that they really need those backups deleted. If no one can delete or prematurely expire any backups for any reason, no matter how privileged this person might be, then I would say that such a system is truly immutable. Unfortunately, all of the above are referred to as immutable.

Sounds pretty bleak

Hopefully I’ve painted a very scary picture for you of just how poorly your backups are probably secured – or not. Hopefully you are now very scared about your backup system being used as an attack service, or you’re worried that all of your backups are a few commands away from being deleted. If any of that is true, then I’ve accomplished my goal for this chapter.

The good news is that there are responses to each of these vulnerabilities, and all you have to do is follow our suggestions. I wish you the best of luck and I mean that with zero sarcasm.

Backup and Recovery Basics

It occurred to me as I was writing this book that its target audience may not be familiar with many basic backup and disaster recovery (DR) concepts. That’s the purpose of this chapter. Feel free to skip it if you have the basics down, or read it just to make sure you have a solid foundation upon which to build a solid DR system that can recover from a ransomware attack.

Defining Backup System Requirements

First let’s focus on the business and political aspects of building or enhancing your data protection system. While technical solutions are important, getting buy-in from stakeholders who’ll benefit from (or fund) the system is equally crucial.

Data protection isn’t the exciting part of IT. It reminds organizations of their vulnerabilities and requires resources that don’t directly contribute to the product you’re selling. You’re essentially selling an insurance policy that no one wants to buy. Despite this challenge, your data protection plan will be one of your organization’s most critical investments.

Before allocating a significant amount of money to data protection, ensure your plan properly addresses organizational needs. An insurance policy is worthless if it won’t cover your losses when disaster strikes. Let’s create a structured process and develop tools for building an effective plan.

**What Does Your Organization Do?**

Success requires more than just technical expertise in data protection systems. You need to understand your organization’s purpose and any external requirements from laws and regulations.

Consider:

* Are you a government organization? What IT-dependent services do you provide?
* Do you operate through an e-commerce model?
* Are you a commercial business with R&D and manufacturing cycles?
* Are there external requirements governing how you store certain data types?

Each aspect has different requirements, and as the data protection person, you’ll need to architect a solution that protects everything for everyone.

Start by understanding your organization and its services or products to grasp your data’s importance. Find someone (or multiple people) willing to explain what your organization does, its product portfolio, and how it serves its customers. This complete picture will guide your journey.

**Build Yourself a Framework**

Data protection affects all organizational aspects. You’ll need input and approval from various technical and non-technical groups. Prepare to recruit teams for review boards that will provide requirements and feedback on service design and operation.

**Document Templates**

All documentation should follow a standard template including:

1. Purpose statement: Concise explanation in 1-2 paragraphs
2. Executive summary: Key information for approvers
3. Revision history: Track changes with dates, authors, and modification details
4. Sign-off page: Ensure accountability from approvers and SMEs
5. Policy/scope: Define specific subjects addressed
6. Glossary: Clarify terminology for non-SMEs
7. Appendices: Supporting information

**Review/Advisory Boards**

Create several review phases and follow them throughout your requirements gathering process. Here is an example of the various phases of review.

*Requirements Review*

Include department representatives and senior management sponsors (like the CIO) to ensure organizational approval.

*Design Review*

The design review board (DRB) should include technology-specific team members who understand organizational implementation. Include representatives from systems, database, storage, network engineering, and cyber-security. The process should include both preliminary design review (PDR) and production readiness review (PRR).

*Operations Review*

Involve operational teams who’ll run the production service to ensure they understand their responsibilities. This should result in a comprehensive runbook.

*Change Review*

The change advisory board (CAB) reviews all production changes to protect organizational integrity.

*Project Management*

Engage the project management office early to track work, coordinate resources, and ensure smooth delivery.

**Collecting Requirements**

Identifying key stakeholders and understanding their needs is crucial. Two critical metrics drive any data protection plan: recovery point objective (RPO) and recovery time objective (RTO). I’ll define these in more detail later in this chapter, but for now just understand it’s about how fast you can recover (RTO) and how much data you can lose (RPO).

*Find the Subject Matter Experts*

Create a list of internal customers by department, remembering that everyone is a data protection customer. Find subject matter experts (SMEs) who can describe each group’s activities, protection needs, and criticality. Key categories include:

*Data Creators*

Understand where data originates, whether from manufacturing systems, artistic teams, or customer-facing departments. This helps determine data recreation complexity and appropriate RPO levels. Remember multiple data entry points and varying change rates based on internal workflows. Work with data creators to understand RPO implications and get insights into transaction volumes and storage requirements.

*Executives*

Executive staff provide insight into organizational operating speed and timeline expectations. While they typically want everything protected constantly, they’re best at helping understand organizational flow and priorities. Be prepared for discussions about acceptable downtime and cost constraints to determine RTO.

*Compliance and Governance*

Ensure compliance with relevant laws and regulations, particularly regarding privacy (like GDPR and CCPA). Consult legal or governance teams to confirm your design meets requirements for backup and archive access.

**Review Requirements**

After collecting requirements, get everyone aligned by reviewing the requirements you believe you have gathered from them. Make sure you understand and can present your understanding of all of the following:

* Data location and volume
* Data generation and change rates
* Service/product creation timelines
* Organizational outage tolerance

Present these findings to key stakeholders, including management representatives and technology team members. Define the problem clearly and present each department’s requirements. Focus on verification rather than solution design. In other words, the only thing you should be worried about at this point is ensuring that you have accurately captured the requirements you will use to build your backup system.

**Service Level Agreements**

Establish service level agreements (SLAs) based on agreed-upon RPOs and RTOs. Consider physical constraints of network resources, storage devices, and potential tape usage. Remember data protection typically uses more network bandwidth than other services.

*Data Classification*

Consider classifying data based on importance levels. Not all data is equally critical, and some might be disposable without affecting operations. While many systems use a single “important” classification, taking time to determine what’s critical, important, nice-to-have, or expendable directly impacts RPO and RTO. There is no time this is more crucial to know than during a ransomware attack.

*Design and Build Your System*

After gathering requirements, move forward with system design. Create multiple approaches with different advantages and disadvantages.

*Draw up Multiple Designs*

Start with a “money is no object” solution achieving defined RPO and RTO. Then create “value-engineered” alternatives that meet objectives with caveats, trading upfront costs for potential backend expenses.

*Review the Designs*

Present your designs to the DRB, including technical teams and data creator SMEs. Cover data storage locations, encryption methods, bandwidth expectations, and operational requirements. Iterate based on feedback and conduct proof-of-concept testing if needed.

**Document and Implement the New System**

Once the system has been decided on, it’s time to implement it. As you’re implementing it, make sure to document it. Documentation completeness is crucial. It’s a good idea to start with a chart that defines responsibilities. A common format is RACI: responsible, accountable, collaborator, and informed.

* Responsible: Those doing the work
* Accountable: Those held accountable for completion
* Collaborator: Those who must be consulted
* Informed: Those kept updated on progress

**Operations Review / Documentation**

Create comprehensive operations documentation including:

* Requirements and design documents
* Operational procedures and runbooks
* Contact information for vendors and key personnel
* Incident tracking and resolution procedures

Consider maintaining printed copies for emergencies when electronic documentation might be inaccessible. But make sure to store them in a loose-leaf notebook to allow you to update them by updating a few pages.

**Implement the New System**

Present complete documentation to the CAB for final approval. Be patient with their process and address any concerns thoroughly. Once approved, implement the system and ensure all changes go through CAB review.

Backup & Recovery Basics

Now that I’ve covered how to gather requirements and prepare to design and implement a backup and DR system, we need to explain some basic backup & recovery concepts. I’ll start with discussing the all-important concept of recovery testing, followed by the concept of backup levels. We then look at many backup system metrics, especially the concepts of RTO and RPO and how they (more than anything else) determine backup design. I then talk about image-level vs file-level backups, and how the contents of backups are selected. The first, and possibly most important basic backup concept, however, is that all backups must be tested.

**Recovery Testing**

The only reason we back up things is to be able to restore them. And the only way you’re going to know if you can restore the things you’re protecting is to test your ability to do so. Regular recovery testing should be a fundamental part of your backup system.

Testing serves two crucial purposes: validating the backup system and its documentation while training your personnel. If the first time they’re executing a big restore is during a production crisis, such a restore will be much more stressful and error-prone. If they’ve performed such restores many times during testing, they should be able to follow their usual procedure confidently.

You should regularly test the recovery of anything and everything you’re responsible for, from small items to very large systems. The frequency of testing each item should relate to how often a restore of such a thing happens in production. While a few times a year might be appropriate for a big DR test, you should be restoring individual files and VMs at least once a week per person.

The cloud has made testing much easier by eliminating resource constraints. You don’t have to fight for resources to use for recovery; you just configure the appropriate resources in the cloud and then restore to those resources. This is especially true of large DR resources; it should be very easy to configure everything you need to do a full DR test in the cloud. Regular testing in this environment will make production recoveries much smoother. Tests should also include restoring common items in SaaS services, including users, folders, and individual files or emails.

**Note**

A backup isn’t a backup until it’s been tested!

Ben Patridge

**Backup Levels**

There are essentially two very broad categories of what the backup industry calls backup levels: backing up everything (full backup) or backing up only what has changed (incremental backup). Each of these broad types has variations that behave slightly differently. Most backup levels are throwbacks to the tape era, but understanding their definitions remains valuable.

*Traditional Full Backup*

A traditional full backup copies everything from the system being backed up (except anything specifically excluded) to the backup server. This includes all files in a filesystem (unstructured data) or all records in a database (structured data). It requires significant I/O, which can create a substantial performance impact on your application.

*Traditional Incremental Backup*

A traditional incremental backup captures all filesystem files or database records that have changed since a previous backup. Unless otherwise specified, incremental backups are full-file incremental backups, meaning they will backup an entire file or record if the modification time has changed or its archive bit has been set in Windows. Even if only one byte changed, the complete file gets backed up. Several other types of incremental backups exist:

*Typical incremental backup:*

Backs up all data changed since the previous backup, regardless of type. Whether the previous backup was full or incremental, the next incremental backup will only capture changes since then. This is the most common approach.

*Cumulative incremental backup:*

Backs up all data changed since the last full backup. While requiring more I/O and bandwidth than typical incrementals, it simplifies restores by needing only the full backup and latest cumulative incremental. This saves time compared to typical incrementals, which require restoring the full backup and each subsequent incremental.

*Incremental backup with levels:*

Uses numbered levels (0-9) where 0 represents a full backup, and 1-9 represent other incremental levels. An incremental backup of a certain number will backup everything changed since a previous backup one level down. This allows various combinations for different results, such as monthly level 0, weekly level 1, and daily increasing levels. These are now very uncommon, and are only here for completeness.

*Block-level incremental backup:*

Only backs up bytes or blocks that have changed since the last backup. Something must track which bytes or blocks have changed, typically through changed block tracking (CBT). This requires significantly less I/O and bandwidth than full-file approaches. It’s especially popular in hypervisor backups, where the hypervisor maintains a bitmap of changed blocks.

*Source-side deduplication:*

An extension of block-level incremental backup that applies additional processing to new or changed blocks before sending them to the backup server. It identifies if “new” blocks have been seen before anywhere in the backup system, saving even more time and bandwidth than standard block-level incremental backups.

*Synthetic full backups*

Created to solve the restore inefficiency of traditional incremental approaches while avoiding the load of regular full backups. Three main methods exist:

* Synthetic full via copying: Creates a full backup by copying data from available backups based on the backup system’s catalog of current versions. Can be run anytime without impacting backup clients but may create I/O load on backup storage.
* Virtual Synthetic full: Possible only with target deduplication systems, creating a full backup by pointing to existing blocks without data movement. Requires backup product integration but is very efficient.
* Forever incremental: Modern approach that never requires another full backup synthetic or otherwise. Requires disk as the primary target and stores each changed item as separate objects. Every backup behaves as a full backup for restores without creating one. This is now the most common type of incremental backup in modern systems.

**Metrics**

Several crucial metrics determine backup system design and maintenance. They govern everything from system design to performance monitoring and capacity planning.

**Recovery Metrics**

No metrics are more important than those involving recovery. Nobody cares how long backups take; they only care about restore speed. Two metrics determine if your backup system works: restore speed and data loss during recovery.

*Recovery Time Objective (RTO):*

The agreed-upon time allowed for a restore after an incident. RTO length typically depends on potential financial losses during downtime. Financial trading firms might seek near-zero RTO, while other organizations might accept weeks of downtime.

For government or non-profit organizations, RTO calculations might focus on overtime costs for catching up after outages rather than lost revenue. Multiple RTOs across an organization are normal, with tighter RTOs for critical applications.

Importantly, RTO starts when the incident occurs and ends when business returns to normal not just when data restoration completes. This includes hardware ordering, logistics, and post-restore activities.

*Recovery Point Objective (RPO):*

The acceptable data loss after an incident, measured in time. Like RTO, multiple RPOs throughout an organization are common depending on data criticality.

Most organizations settle on values higher than one hour, typically 24 hours or more, because tighter RPOs require more frequent backups. There’s little point in agreeing to a one-hour RPO while running daily backups you’ll achieve 24-hour RPO at best.

*Negotiating RTO and RPO:*

Most organizations initially want zero RTO and RPO no downtime or data loss. This isn’t technically possible even with the best systems and would be prohibitively expensive to approach.

Present cost estimates for requested RTOs and RPOs. Organizations justifying tight recovery objectives should provide potential outage cost analysis. Negotiate between technical feasibility, affordability, and current capabilities.

*Recovery Time Actual (RTA) and Recovery Point Actual (RPA):*

These metrics measure how well you meet RTO and RPO during actual recoveries or tests. Most organizations’ RTAs and RPAs fall short of objectives. Either adjust objectives or redesign the backup system there’s no point having tight objectives if actuals aren’t close.

Testing is essential because most organizations rarely perform production recoveries. Regular testing reveals:

* Backup system reliability
* Resource requirements for large recoveries
* Actual RTA and RPA values
* Recovery process familiarity

Having participated in large-scale recoveries without knowing system capabilities, I can confirm the first question is always “how long will this take?” Without regular testing, you can’t answer confidently.

**Capacity Metrics**

These metrics track the capacity of your system to back up and restore data, like how much it can store, and how fast it can back it up, how fast it can restore it, how long data is being kept, and others.

*License/workload usage:*

Track license utilization to anticipate needs. Monitor workload numbers even without license implications to understand backup system growth.

*Storage capacity and usage:*

Monitor available capacity and usage trends. Failing to do so can force emergency decisions violating organizational policies, like deleting older backups prematurely.

Cloud object storage simplifies this by automatically growing. However, cloud block storage still requires volume management and monitoring. Additionally, cloud block volumes charge for provisioned capacity versus object storage’s usage-based pricing, making it harder to manage and costing more than object storage.

*Throughput capacity and usage:*

Backup systems have finite daily backup volume capabilities, usually measured in MB/s or TB/hour. Monitor this to prevent backups extending into workday hours. If you’re using tape, tape throughput monitoring is particularly crucial; backup speed must match tape drive minimums to prevent device failure. Consult the manufacturer for details.

Cloud offers both advantages and challenges here. While cloud throughput is virtually unlimited with properly designed systems, many cloud-based products using traditional backup software in VMs face bandwidth limitations. You’ll need to upgrade VM types or add instances to increase throughput. Some systems can automatically scale bandwidth with needs.

The cloud’s main throughput limitation is your site’s upload bandwidth. Even with byte-level replication or source deduplication, you may exceed site capacity, requiring bandwidth upgrades or vendor changes.

*Compute capacity and usage:*

Whether in the datacenter or the cloud, backup system capability depends on underlying compute power. Insufficient processing capability in backup servers or databases can slow backups into workday hours. Monitor system performance to prevent this.

Cloud-native backup systems can automatically scale compute resources. Some even scale up and down throughout the day, reducing costs. However, traditional backup software running in cloud VMs lacks this capability, requiring manual compute additions and associated licensing.

*Backup Window*

Traditional backup systems significantly impact primary system performance during backup. Full backups are resource-intensive, and even incremental backups can be challenging if they’re full-file incrementals. You should agree in advance on allowable backup times.

Traditionally, a typical window might be 6 PM to 6 AM Monday-Thursday, and 6 PM Friday to 6 AM Monday for environments with minimal weekend activity. However, this is highly environmentally dependent. Monitor window utilization to see if you’re approaching window limits, then either reevaluate the window or redesign the system.

Let backup products handle scheduling within their window rather than over-engineering with external schedulers. The built-in scheduler typically manages resources more efficiently.

Organizations using block-level incremental-forever techniques often don’t need strict windows. These methods run briefly (minutes) and transfer small amounts (megabytes), with minimal performance impact. Such systems typically run throughout the day, as frequently as every five minutes, making the backup window point moot.

*Backup & Recovery Success and Failure*

Track backup and recovery quantities and success rates. While 100% success is ideal, it’s unlikely especially for backups. Monitor trends to identify improvements or degradation.

Address all failures. Successfully re-running failed backups or restores resolves immediate concerns, but track failures for trend analysis.

*Retention*

While not technically a metric, monitor retention policy adherence. IT shouldn’t determine retention periods; these should come from legal, organizational, and regulatory requirements.

Modern systems use multiple storage tiers. Retention should specify duration for each tier, not just overall retention. Review system settings against organizational policies periodically and adjust as needed.

**Using Metrics**

It’s not good to just have metrics; you need to use them to show your system is meeting those metrics. Publishing metrics builds backup system confidence. Be sure to share:

* Performance against design goals
* Backup and recovery success rates
* Capacity forecasts
* RTO/RPO achievement capability (i.e. RTA/RPA)

Being transparent about RTA and RPA limitations serves everyone better than hiding shortcomings.

**Item vs Image-level Backups**

There are two very different ways we tend to back up data. We can either back up items or images. An item-level backup backs up discrete items like files or objects, typically using in-system agents. Advantages include:

* Easy to understand
* Straightforward implementation
* Direct file selection
* Simple restore process

An image-level backup backs up entire drives or volumes at the block level. It’s very popular in virtualization environments. Benefits include:

* Faster backups and restores
* More efficient incremental backups
* Better disaster recovery capabilities
* Simplified bare-metal recovery

Modern systems often combine both approaches, performing image-level backups while maintaining file-level restore capabilities through mounting or indexing.

**Backup Selection Methods**

Understanding how systems are included in backups is crucial for ensuring complete protection. First, ensure servers and services are registered with your backup system. No backup system automatically detects new SaaS applications or physical servers, though some can detect new VMs in monitored environments. Two broad categories exist for backup inclusion.

*Selective inclusion*

Administrators specify which filesystems, databases, or objects to back up. Example: backing up only D:\ drive or specific databases.

*Selective exclusion (automatic inclusion)*

Backs up everything except specifically excluded items. Example: backing up all filesystems except /tmp or user media directories.

While selective inclusion might save storage, it’s riskier than selective exclusion. Configuration changes require backup reconfiguration, or new resources never get backed up.

Some people may feel they are saving space using selective inclusion and not including things like the OS drive. However, with deduplication, excluding operating systems saves less space than you might think. The OS is typically stored only once across all backups. Consider leaving default configurations and letting deduplication handle efficiency.

**Tag-based and Folder-based Inclusion**

Tag- and folder-based inclusion is another method, and is popular in virtualization and cloud environments, where new VMs or databases can receive tags or folder placements indicating their type. This can automatically apply appropriate backup policies.

Important: If you’re using tag-based inclusion, be sure to create a default backup policy for untagged resources. Monitor this policy for new systems that might need different backup approaches. Without a default policy, this functionality risks leaving systems unprotected.

**Backup & Archive Myths**

“You don’t need to backup RAID”

False. RAID only protects against physical device failure. It can’t help with file deletion, ransomware encryption, or database corruption. RAID protects the volume, not the filesystem above it.

“You don’t need to backup replicated data”

False. Replication copies everything including mistakes and corruption. If someone drops a database table accidentally, replication ensures that mistake reaches all copies.

“You don’t need to backup IaaS/PaaS”

False. Cloud vendors typically provide backup capabilities but don’t automatically back up your resources. While cloud resources offer high availability, this doesn’t protect against user mistakes or attacks.

“You don’t need to backup SaaS”

False. Major SaaS vendors rarely include comprehensive backup. Check your service contract backup, recovery, and restore terms typically aren’t included. Built-in features like versioning and recycle bins don’t replace proper backups.

“Backups should be stored for many years”

False. Traditional backup products aren’t designed for long-term retention. Without proper archive functionality, using backups for long-term storage creates expensive e-discovery challenges. Most restores come from recent backups; consider 18-month retention for regular backups and proper archive systems for longer retention.

“Tape is dead”

False. While I haven’t used tape for primary backups in years, tape sales continue growing. Tape excels at long-term archives, offering reliable, cost-effective storage with minimal power requirements. Cloud providers often use tape behind the scenes for cold storage.

If you’re designing or operating a backup system without agreed-upon SLAs for RTO and RPO, you’re asking for trouble. Take time now to document these requirements, remembering they should come from the organization, not IT. If you know your RTA and RPA don’t meet objectives, speak up.

Regarding myths, claims that you don’t need to back up certain technologies are almost never true. Get backup coverage in writing from any vendor claiming it’s included.

Most modern backup products use block-level incremental backups forever, making traditional backup levels less relevant. Default to selective exclusion for safety and efficiency. Focus on protection first, cost second no one ever got fired for backing up too much data, but plenty have been fired for backing up too little.

A commercial backup and recovery solution can take various forms, from traditional methods not fundamentally different than those used 30 years ago to techniques developed in recent years. This chapter provides a comprehensive view of available options and their pros and cons.

Backup Types

There are myriad ways you can meet the backup requirements you agreed to in your environment. This is my best attempt to summarize the pros and cons of the various methods.

**Is Everything Backup?**

Many people have a narrow view of backup, immediately thinking of tape drives and batch processes encapsulating files into formats like tar, dump, or commercial backup formats. While they might expand this to include disk targets, they often don’t consider replication, snapshots, or continuous data protection as backup methods. I respectfully disagree.

I define backup broadly as any copy of data stored separately from the original that can restore the original system if damaged. Not everything mentioned here qualifies as backup when used alone. Replication and snapshots, for instance, aren’t valid backup methods individually. However, combined they create near-continuous data protection (near-CDP), a robust solution.

**The 3-2-1 rule**

First, anything called backup in this chapter must conform to the 3-2-1 rule. What is that, you ask? The 3-2-1 rule is fundamental to backup design, like E=MC**2** is to physics. Use it as one of the ways to verify if your backup design is proper.

The rule states: have at least three versions of your data on two different media, one of which should be somewhere else. Let’s break this down:

*“Three versions of your data”*

Some include the original in those versions; others do not. Multiple versions protect against corrupted backups or unnoticed mistakes. Three is the minimum, not maximum. Modern systems often create many more versions - office apps create multiple daily versions, database logs create thousands of versions, and laptop backups might run every minute.

*“On two different media”*

Never store all backups on the same media as the original. For example, using Mac OS Time Machine to backup one partition to another partition on the same drive is risky - if the drive fails, you lose both original and backup. Backups should be on different drives, different computers, and preferably physically distant from the original system. The best thing to do is to at least two different types of media, like disk and cloud, tape and disk, etc.

**Making Sense of a Big Landscape**

We’ll categorize backup solutions based on how they restore data:

1. Traditional restore: Copies data from backup to the system after initiating restore
2. Instant recovery: Makes backup usable as primary without performing traditional restore

**Backup Methods Supporting a Traditional Restore**

*Traditional Full and Incremental Backups*

The most common approach starts with an initial full backup followed by incremental or cumulative incremental backups until the next full backup. Historically, organizations performed weekly full backups with daily incrementals. Many now use monthly fulls with weekly cumulative incrementals and daily incrementals.

Advantages:

* Simple to understand
* Proven method
* Widely supported

Disadvantages:

* Repeated full backups waste resources
* Restore process is inefficient
* Must restore full backup then each incremental sequentially

*File-level Incremental Forever*

This approach performs one initial full backup followed by perpetual incremental backups at the file level. Key characteristics:

* Never performs another full backup
* More efficient than traditional methods
* Incompatible with tape
* Faster restores by avoiding redundant data transfers
* Built from scratch for this purpose

Requirements for true incremental forever:

* Single initial full backup
* Only incremental backups afterward
* Never requires synthetic fulls
* Built specifically for this approach

*Multiplexing*

A tape-writing method addressing speed mismatch between tape drives and backup streams. Modern approaches use:

* Multiple backup streams interleaved together
* Larger chunk sizes for better restore performance
* Careful logistics and memory management
* Improved restore capabilities compared to traditional multiplexing

*Block-level Incremental Forever*

Operates at bit or block level rather than file level. Requires:

* Changed block tracking (CBT)
* Bitmap maintenance of changed data
* Integration with systems providing CBT
* Disk-based storage

Benefits:

* Reduced data transfer
* Efficient for remote backups
* Perfect for VM backups
* Lower network bandwidth requirements

*Source Deduplication*

Source deduplication performs deduplication at the backup source before data transfer. Key aspects:

* Operates as incremental forever by design
* Never backs up previously seen chunks
* More efficient than block-level incremental
* Ideal for remote backups and cloud targets

Example scenario: Consider 100 Windows VMs after patch day. While file-level incremental backs up all new files and block-level backs up changed blocks, source deduplication only backs up new blocks from the first VM, recognizing identical blocks across all 100 VMs.

Benefits:

* Most efficient network usage
* Excellent for remote systems
* Perfect for cloud backups
* Reduces storage requirements

Disadvantage:

* May require changing backup software

**Methods Supporting Instant Recovery**

These methods support recovery without traditional restore operations. The restored system becomes instantly available upon recovery, making these ideal for tight RTOs and RPOs.

*Replication*

This is a well-established concept using source and target volumes. Types of replication include:

Synchronous replication, which replicates before acknowledging writes

* Provides highest protection level
* Impacts performance
* Distance limitations affect feasibility
* Often impractical beyond 300 miles

Asynchronous Replication, which queues changes for replication, then replicates them as it has bandwidth

* Minimal performance impact
* Potential lag between copies
* Better for longer distances
* May support write coalescing when behind

Database Replication operates at transaction level in the database, replicating changes at the transaction level from one database to another.

* Usually asynchronous
* Different from block-level approaches
* Focused on database consistency

Limitations of replication by itself

* No “back button” for mistakes
* Replicates errors and corruption
* Doesn’t satisfy 3-2-1 rule alone
* Requires additional protection methods

*Continuous Data Protection (CDP)*

Combines asynchronous replication with change logging, essentially creating “replication with a back button.”

Benefits

* Instant recovery capability
* Point-in-time recovery options
* Operational and disaster recovery in one system
* Infinite recovery points within retention period

Challenges

* Resource intensive
* Expensive implementation
* Complex point-in-time selection
* Higher system requirements

*Snapshots*

Virtual snapshots rely on primary volumes for most data. Popularized by NAS vendors, they provide historical versions while using minimal space. Though “minimal” is relative, as they occupy expensive primary storage space.

Advantage

* Space-efficient versioning
* User-directed restores possible
* Excellent backup sources
* Perfect for pre-upgrade protection

Disadvantages

* Relies on source volume
* Doesn’t satisfy 3-2-1 rule alone
* Vulnerable to source corruption
* Primary storage cost impact

*Near-Continuous Data Protection (near-CDP)*

Combines snapshots and replication to create a complete backup solution. Common implementations:

* Snapshot primary volume and replicate
* Replicate to secondary and snapshot there
* Integrated with hypervisor systems
* Often included in HCI appliances

Benefits

* Instant recovery capability
* Multiple version support
* Off-site copies possible
* Complete backup solution
* Integration with virtualization

Challenges

* Potential vendor lock-in
* Single vendor risk
* May need additional protection layer
* Often requires scripting for application integration

**Deciding on a Backup Method**

No perfect backup and recovery method exists. Each approach has pros and cons. To narrow your choices, determine which advantages you can’t live without and which disadvantages you won’t accept.

**Do you need to change?**

The grass often looks greener with other backup solutions. Before switching, ask yourself some key questions:

* Does current system meet RTOs and RPOs?
* Are costs within budget?
* How are operational costs?
* How much troubleshooting time required?
* What’s the cost of switching, including training?

Consider consulting product experts before replacement. Most backup system problems stem from configuration (wetware) rather than software or hardware issues. The cost of expert consultation is usually less than system replacement.

**Advantages/Disadvantages of Different Approaches**

The following list is just a few of the things to consider when deciding on a backup system. Do not do this lightly or quickly.

* Recovery Time Considerations:
  + Traditional restore methods can’t match instant recovery RTOs
  + Verify zero RTO requirement before choosing instant recovery
  + An eight-hour RTA is acceptable if you have a 24-hour RTO
  + Don’t pay for speed you don’t need
* Recovery Point Considerations:
  + Traditional approaches limited to 24-hour RPO due to low backup frequency
  + Many methods achieve one-hour RPO
  + An RPA of zero is possible, but extremely expensive
  + Match method to actual requirements.
  + Consider cost vs. benefit
* Virtualization Support:
  + Most methods handle VMware and Hyper-V well
  + Specialized solutions needed for AHV or KVM
  + Consider hypervisor backup roadmap
  + Verify specific platform support. Do not take vendor’s word for it.

**Cloud Considerations**

If your environment is heavily into cloud computing (who isn’t these days), you have even more things to consider. The following, again, is a brief list of such things.

*Primary backup storage type (object vs. block)*

Object is much cheaper (10X less or more) than block, but you need to make sure your backup system can properly handle object storage. It’s a myth that it’s slower; however, some backup systems may require block storage.

*Cloud-native vs. traditional design*

A backup design that fully leverages how the cloud works will save you money in the long run. Simply lifting and shifting backup servers into the cloud costs much more.

*Resource scaling capabilities*

Find out whether or not the system you are considering has the ability to automatically scale up and down without you having to manually do it. This gives you performance when you need it, and saves you money when you don’t.

*Bandwidth requirements*

If you’re going to back up to the cloud, make sure you know how much bandwidth you need and how much you have.

Takeaways

This chapter provided a foundational understanding of backup and disaster recovery concepts, starting with the critical business aspects of building stakeholder buy-in and gathering requirements. We covered the framework needed to design and implement a backup system, including documentation templates, review boards, and the importance of finding subject matter experts.

We then explored core backup concepts, from basic backup levels to essential metrics like RTO and RPO. The chapter emphasized that recovery testing is fundamental - a backup isn’t truly a backup until it’s been tested. We examined different backup methods, from traditional full/incremental approaches to modern techniques like source deduplication and near-CDP.

The 3-2-1 rule emerged as a foundational principle: maintain at least three versions of your data on two different media types, with one copy kept off-site. This rule serves as a fundamental benchmark for evaluating any backup design.

Whether choosing traditional restore methods or instant recovery solutions, the key is matching your approach to organizational requirements while considering factors like virtualization support and cloud integration. Remember: there’s no perfect backup solution - focus on finding the best match for your specific needs while ensuring compliance with fundamental principles like the 3-2-1 rule.

Test Questions:

1. Your organization wants an RTO and RPO of zero. What should your response be? a) Immediately implement synchronous replication b) Present cost estimates and negotiate realistic objectives based on business impact c) Tell them it’s impossible and suggest alternatives d) Implement CDP
2. Which of these violates the 3-2-1 rule? a) Three backup copies on different storage arrays, one off-site b) Two backup copies on tape, one in the cloud c) Two backup copies on the same disk array as the original d) Five backup copies across three different media types
3. When implementing a new backup system, which review board should give final approval before going live? a) Design Review Board (DRB) b) Requirements Review Board c) Operations Review Board d) Change Advisory Board (CAB)
4. Your backup success rate is 98%, but you’ve never tested restores. What should you do first? a) Celebrate the high backup success rate b) Implement more backup jobs to reach 100% c) Begin regular recovery testing to establish RTA/RPA metrics d) Switch to a different backup solution
5. Which backup selection method is safer for ensuring complete protection? a) Selective inclusion b) Selective exclusion c) Tag-based inclusion without a default policy d) Manual selection

Answers:

1. b - Present costs and negotiate realistic objectives
2. c - Two copies on same array violates “two different media” requirement
3. d - CAB should give final approval for production changes
4. c - Recovery testing is essential for verifying system effectiveness
5. b - Selective exclusion ensures new resources aren’t misse

Regarding ransomware, the reality is that no defense is entirely foolproof. Even with the best preventative measures, attackers constantly evolve their tactics, and breaches can still occur. What truly determines whether an organization survives a ransomware attack isn’t just its ability to prevent it but its capacity to respond and recover effectively. Organizations that invest in robust incident response plans, secure backup strategies, and rapid recovery capabilities are far more effective at minimizing downtime, reducing financial losses, and restoring operations with minimal disruption. In an era where cybercriminals continually outpace security technologies, preparation for a swift and effective response is as critical as prevention.

Consider the story of a small accounting firm hit by ransomware during tax season. Despite having strong firewalls and anti-malware software, attackers exploited an unpatched vulnerability to gain access to the network, deploy ransomware, and encrypt all of their client files. However, because the firm invested in regular, offline backups and trained employees on response protocols, they restored their systems within days without paying the ransom. While the attack was disruptive, their recovery measures saved them from financial ruin and maintained client trust. This underscores a fundamental truth: prevention can reduce risk, but a comprehensive recovery plan is what ensures survival when the worst happens.

Ransomware is not just a technological issue; it’s an operational challenge that requires a balanced approach. While proactive defenses such as patch management, endpoint protection, and employee training are vital, equal attention must be given to preparing for the possibility of a breach. This chapter will explore how to fortify your organization’s response capabilities alongside preventative measures. By integrating robust backup systems, incident response planning, and network segmentation into your strategy, you can significantly reduce the impact of ransomware, ensuring a faster and more effective recovery. Let’s dive into how to build resilience in the face of this evolving cyber threat.

Why Preparation is the Best Defense

When it comes to ransomware, the best way to minimize its impact is not just through defense mechanisms but through preparation. In today’s world, where cyberattacks are on the rise, expecting ransomware to hit your organization is a more realistic approach than assuming it won’t happen. As ransomware attacks grow in frequency and sophistication, organizations need to think ahead, preparing for the possibility of an attack rather than waiting for one to happen. Prevention is important, but a solid response and recovery plan can make a difference when an attack occurs. As noted in Chapter 1, ransomware is a major threat, causing significant operational, financial, and reputational damage to organizations, especially when unprepared.

The evolution of ransomware attacks, including tactics like double extortion and data exfiltration, highlights the need for a comprehensive, multi-layered approach to cybersecurity. Today, attackers encrypt data and steal it, threatening to release it unless a ransom is paid. This evolving threat landscape means relying on a single security measure is no longer enough. Organizations must adopt a holistic cybersecurity strategy that blends prevention, detection, response, and recovery strategies to mitigate the risks effectively. According to the FBI’s 2023 report, double extortion attacks have significantly increased in recent years, illustrating just how much the nature of ransomware attacks has changed.

A great example of effective ransomware preparedness comes from a small to mid-sized company with a solid backup and incident response plan. When attacked, their quick response and well-tested recovery plan allowed them to restore critical systems from backups and avoid paying the ransom. They also utilized their cyber insurance to cover their incident response expenses, significantly reducing the financial burden. This scenario shows that being prepared doesn’t just minimize damage, it can also ensure an organization can recover more effectively and at a lower cost. Such preparedness isn’t just about reacting to a breach and planning for the unexpected.

While prevention measures like strong firewalls, endpoint protection, and employee training are important, they are just part of the solution. The reality is that ransomware attacks often get through even the most fortified defenses. That’s why response and recovery plans are critical. According to the 2022 study conducted by Palo Alto Networks’ Unit 42, the average ransom payment in 2022 was close to $1 million, demonstrating the high cost of being unprepared for such attacks. But what’s not often reported is the additional financial impact of a ransomware attack on the organization, such as overtime pay for incident responders, hiring consultants, legal fees, regulatory penalties, loss of productivity, loss of business, and replacement hardware.

The following sections will explore the essential components of ransomware preparedness and response. These include proactive cybersecurity measures, such as vulnerability management, regular software updates, strong password policies, multi-factor authentication, and the critical role of employee training. We’ll also delve into the importance of backup systems, incident response planning, network segmentation, and more. By understanding the importance of these steps, you can better prepare your organization to handle a ransomware attack when it does happen.

Proactive Cybersecurity Measures

**Cyber Hygiene Basics**

Implementing and maintaining good cyber hygiene is fundamentally critical in mitigating the success of a ransomware attack. It starts with the basics, such as having a good data backup strategy, regularly updating and patching software and systems, vulnerability management, system hardening, email filtering, multi-factor authentication, user training, enhanced anti-malware protection, and continuous monitoring. All of these require an organization to understand itself, the technologies it relies on, the location and value of its data, and its users. If implemented strategically, these basics can provide effective layers of protection against cyber threats like ransomware.

**System Hardening**

Ransomware attackers are like burglars, always looking for an easy way in. System hardening is the process of locking every door, securing every window, and putting alarms in place to make sure they can’t get in. It’s about reducing the attack surface, which simply means eliminating unnecessary vulnerabilities or opportunities in your systems, applications, and devices. The smaller your footprint and the more secure your environment, the harder it is for ransomware to find a way to strike.

The first step in system hardening is removing what you don’t need. Every application, service, or account running on your network is a potential entry point for attackers. Start by understanding the systems in your environment, the types of users, and the applications they (systems and users) need to do their job. Document this information in a Systems Hardening guide and use it to disable unnecessary services, uninstall unneeded software, and remove inactive user accounts, templates, demo files, etc. For instance, old software versions often contain unpatched vulnerabilities, new servers and workstations come with unnecessary software, and many devices have services running that don’t support the system’s purpose or role (e.g., RDP, SMB, PowerShell, Bluetooth, web server, Xbox Game Console, etc.). Keeping your systems lean not only improves performance but also reduces risks.

Next, define the acceptable configuration of each device and application. Default settings are rarely secure and are often widely known, making them an easy target. Take the time to configure your systems, devices, and software to meet industry security standards and their intended purpose. This includes enabling workstation firewalls, turning off unnecessary features, and hardening access permissions. This is especially important for any Internet-facing systems (i.e., no RDP on external systems and no Internet-facing admin services for external network devices). For example, Microsoft and other vendors provide baseline security configurations that can be tailored to your organization’s needs. Regular configuration reviews also help ensure no weak spots are left exposed or could develop over time.

Implement network and systems policies to define and enforce encryption, password, and authentication requirements. Encrypt endpoints wherever possible with BitLocker or similar tools. Require complex, unique passwords and enforce multi-factor authentication (MFA) for accessing your organization’s network and all business-critical applications. Once the systems and application hardening guidelines are finalized, create golden images of each type and secure them offline for use in new deployments and recovery efforts. These golden images provide the base restore points upon which data restore can be applied, which is critical when recovering from ransomware.

System hardening isn’t about making systems impossible to use, it’s about finding the balance between security and functionality. Removing unnecessary components, enforcing secure configurations, and strengthening user authentication make it significantly harder for ransomware to infect your systems.

The National Institute of Standards and Technology (NIST) has a National Checklist Program (NCP) for just about any technology. The NCP was established by the NIST Special Publication 800-70 and is the United States government repository of publicly available security checklists. Each checklist or benchmark provides detailed low-level guidance for setting security configurations of operating systems and applications. The NCP is free to use and can be found at https://ncp.nist.gov/repository.

**Data Backups**

An effective data backup strategy can be the silver bullet response to a ransomware attack, among many other types of incidents. The next section on the role of backups in ransomware defense, as well as several other chapters of this book, touches on the various data backup approaches, techniques, strategies, and recovery capabilities that can help save an organization infected with ransomware if the critical systems are identified, data are securely archived and capable of being restored in a timely manner.

**Patch Management**

Patch management controls, like enforcing automatic Windows updates and keeping track of all the third-party application updates (e.g., Java, Adobe, Microsoft Office), are key to mitigating a network intrusion-based ransomware attack. Attackers often exploit known network and application layer vulnerabilities in sophisticated ransomware attacks, so ensuring all software is patched and up to date can reduce the chances of these vulnerabilities being exploited. As noted in Chapter 1, vulnerabilities in outdated systems have been a consistent entry point for ransomware. They may provide an internal foothold from which attackers can coordinate a much larger scope infection.

I’ll add here that the effectiveness of a good patch management program requires a complete and updated inventory of all the systems, applications, and devices within your organization. If you don’t know they exist, you can’t ensure they’re patched, secured, backed up, etc. In addition, without an effective patch management program, you cannot ensure the integrity or security of the devices on your network, and you will continue to identify deficiencies through vulnerability scans.

**Vulnerability Management**

When it comes to ransomware, cybercriminals are always looking for the easiest way in. They thrive on vulnerabilities, those tiny cracks in your digital armor, like an outdated application, a forgotten software patch, or misconfigured systems. Vulnerability management is the practice of identifying, prioritizing, and fixing these weak spots before attackers can exploit them. Think of it as a regularly scheduled home inspection.

The foundation of vulnerability management starts with an effective patch management program, which is then assessed by conducting regular vulnerability assessments. These assessments help you identify missed patches and any new potential risks across your network, applications, and devices. For example, a vulnerability scanner can find outdated software, weak encryption protocols, unnecessary services running, or open ports that could provide attackers with a foothold. While the volume of potential vulnerabilities may seem overwhelming, prioritization is key. Not all risks are created equal, and you must focus on the vulnerabilities that pose the greatest threat to critical systems or are actively exploited in the wild.

Most vulnerability scanners, including the free Nessus scanner, provide internet reference links for additional vulnerability details and the Common Vulnerability Scoring System (CVSS) score for each. The CVSS score is a good initial indicator to help you prioritize your remediation efforts. This doesn’t mean lower-scoring vulnerabilities shouldn’t be reviewed or considered for remediation. So, it depends, and you may have to invest more time upfront on the initial scan results to truly understand the findings and become more familiar with your environment. This will lead to less effort and better prioritization in subsequent scan results.

* The first Nessus scan I ran on a production network resulted in over 15,000 pages of findings, which took our team of four about a month to review, validate, and prioritize.

Beyond patches, identifying system misconfigurations is another value of vulnerability management. Even the most secure systems can be exposed if they’re not set up correctly. Default passwords, user accounts with excessive permissions or improperly configured firewalls can all serve as entry points for ransomware. Regular configuration reviews and adherence to best practices like least privilege access can significantly reduce your attack surface.

Vulnerability management isn’t a one-time effort, it’s an ongoing, continuous process. Cyber threats are constantly evolving, and new vulnerabilities emerge almost daily. That’s why many organizations integrate vulnerability management into their broader cyber hygiene practices. You can stay one step ahead of attackers by combining automated tools, like vulnerability scanners, with human oversight and consistent updates. In the fight against ransomware, vulnerability management keeps us from becoming complacent while helping identify new and evolving vulnerabilities.

**Email Filtering**

Email systems have evolved over the years from on-premise systems that had to be purchased, maintained and patched, backed up, licensed, and managed. Today, most email systems are hosted by a third party, like Microsoft O365 or Gmail. The good news is that, depending on their configuration and use, these cloud-based email systems can help filter out spam email, suspicious attachments, and prohibited content.

Configuring the O365 “strict” email filtering setting, for example, will heavily scrutinize incoming emails against a set of spam and phishing criteria and put suspected messages in quarantine instead of the user’s inbox or junk folder. This is an important out-of-sight out-of-mind control that will keep curious users from opening these emails and clicking attachments or links even when they know not to. Though users are getting better at identifying bad or suspect emails, human nature (curiosity) still leads a larger percentage of these users to interact with the bad email anyway.

**Passwords and MFA**

Another critical component of cyber hygiene is enforcing strong password policies. Weak or reused passwords are prime targets for attackers looking to exploit access points. Multi-factor authentication (MFA) is one of the best defenses against unauthorized access. By requiring more than just a password, MFA adds an additional layer of protection that makes it significantly harder for attackers to gain access, even if they have compromised a password.

MFA, however, is just a tool and must be appropriately configured and used to be effective. The implementation of MFA must include technical controls to ensure that it is deployed and enforced, such as an active directory GPO and internet browser configurations that prevent saving the MFA token. In addition, users must be trained on the importance of MFA and prohibited from saving MFA tokens, selecting “trust this device”, or other means of circumventing the additional authentication step.

**Case Study – Network Compromise Leads to Ransomware infection due to stolen MFA Token**

A mid-sized software-as-a-service company became infected with ransomware, impacting all of its production systems and the majority of its employee workstations. The attackers demanded several million dollars in ransom, claiming that they had complete control of the Company’s network domain and access to its insurance and financial records.

How did this happen? The Company’s Chief Information Security Officer (CISO) was confident in their cybersecurity program, the layers of protection they’d implemented, and even the 24x7 monitoring they do around the clock… nothing tripped any alarms. The CISO reviewed all the firewall logs, anti-malware logs, and system configurations, looked for rogue devices, and tried to find patient zero. Nothing added up until an employee commented about something weird they experienced a few days prior.

User: I’m not sure if it’s related to what’s happening, but a few days ago, last Wednesday, I received an email from our Company health insurance provider. The email looked ok to me; it looked like other emails from them. In the email, there was a link to a “new employee benefits portal” and a message that if I signed up for their fitness program, I would receive an Apple Watch. I tried clicking the link from my phone, but I got an error message about using my computer to access the site. I then logged into webmail and clicked the link from my computer. The Insurance site looked familiar, and I registered for the program, but when I clicked “submit” nothing happened. So, it was the combination of the error message on my phone and the webpage that didn’t do anything that made me feel like it was a scam.

CISO: Thank you, I’ll check the email and webpage to see if it’s related. You mentioned logging into webmail but didn’t say if you had to use MFA.

User: It didn’t ask me for MFA. In the past, I clicked “Remember this PC” or “Trust this PC,” and I haven’t had to enter the MFA again.

A review of the email determined that it was a cleverly crafted phishing email that avoided being filtered out by the email system. The link in the email led to a website that ran code to determine whether a user was connecting via mobile device or computer. If the connection was a computer, then the website attempted to steal the user’s store (remembered) MFA token from the Internet browser. With the stolen MFA token, the attacker can log in to the company’s network as if they were a legitimate user and without triggering any alarms or alerts. The attackers then accessed the Company’s network and used this foothold to begin their attack.

**Employee Training**

While technology plays a significant role in defending against ransomware, human error remains one of the most common causes of successful attacks. Phishing emails, which trick users into clicking malicious links or downloading infected attachments, remain a primary method of delivering ransomware. Employee training is essential to help individuals recognize suspicious emails, links, or attachments. By creating a cybersecurity awareness culture, employees become the first line of defense against these threats.

The importance of continuous employee education cannot be overstated. Training should not be a one-time annual event but an ongoing process that reflects the evolving nature of ransomware attacks. As ransomware evolves, so must the training programs, ensuring that employees know the latest tactics and threats. This proactive approach can significantly reduce the likelihood of an attack being successful, as informed employees are less likely to fall victim to phishing scams and other social engineering techniques and are less likely to circumvent the cybersecurity controls we put in place to protect them.

**Advanced Endpoint Protection**

In addition to employee training, organizations should deploy advanced endpoint protection tools. Traditional antivirus software is no longer sufficient to protect against modern ransomware attacks, as these tools often fail to detect new or polymorphic strains of malware. Next-generation anti-malware solutions, which leverage artificial intelligence (AI) and machine learning (ML), are more capable of identifying and blocking ransomware before it can cause harm. These solutions continuously monitor systems for abnormal behavior, allowing them to detect ransomware payloads based on their activity rather than known signatures.

Firewalls and Intrusion Detection Systems (IDS) also play a critical role in protecting endpoints. A well-configured firewall can block malicious traffic before it reaches an organization’s network, while an IDS can identify unusual network activity and alert security teams to potential intrusions. By monitoring for signs of exploitation or suspicious behavior, these tools enable organizations to detect ransomware early in the attack lifecycle, allowing for a quicker response before the attack spreads. In addition, strict firewall and IDS configurations such as geo-IP blocking and blocking known bad IP addresses can possibly prevent a device infected with ransomware from being able to contact its master (a.k.a. the command-and-control systems) to negotiate the encryption keys.

**Continuous Monitoring**

Ransomware doesn’t just appear out of nowhere, and it usually leaves clues. Attackers often move silently through networks, probing for weak spots, testing credentials, or exfiltrating data before launching their final strike. This is where continuous monitoring comes in. Think of it as having a security guard who never sleeps, constantly watching for suspicious activity and sounding the alarm before things spiral out of control. By proactively monitoring your systems, networks, and endpoints, you can catch the warning signs of ransomware early and stop it in its tracks.

At its core, continuous monitoring involves using tools to collect and consolidate data from sources on the network. This data helps observe network activity, system performance, and user behavior 24/7. Solutions like Security Information and Event Management (SIEM) systems and Endpoint Detection and Response (EDR) platforms play a vital role here. They collect and analyze massive amounts of data in real-time, flagging unusual behavior such as unexpected file encryption, unauthorized access attempts, or large amounts of data being transferred outside your network. For example, if an employee’s account suddenly starts accessing critical systems, it shouldn’t; continuous monitoring tools will alert your security team before it becomes a full-scale attack.

Another key benefit of continuous monitoring is its ability to identify zero-day threats or vulnerabilities that haven’t been publicly disclosed or patched yet. Ransomware groups frequently exploit zero-days to bypass traditional defenses, making continuous monitoring a crucial line of defense. The ability to detect abnormal behavior in real-time can mean the difference between catching ransomware early and watching it wreak havoc across your systems. Monitoring tools can recognize suspicious activity by correlating data patterns and spotting anomalies even when a ransomware variant is brand new.

Continuous monitoring, however, still requires skilled people to configure the platforms, review the alerts, and escalate issues to be addressed. Automated systems are powerful, but they’re most effective when paired with trained security teams who can investigate alerts and act quickly. A well-monitored system combines real-time detection with rapid incident response. For example, if ransomware has been detected encrypting files, the response team can isolate the affected device, shut down network access, and prevent the infection from spreading. Though this chain of events could be automated through a Security Orchestration and Automated Response (SOAR) tool, the impact and scope of the response it typically more accurate and limited when done by skilled people.

Continuous monitoring allows you to see threats as they emerge and respond before they escalate. It’s not just about catching the attack but about building confidence that your systems are being watched and protected at all times. This is especially true when, in the absence of continuous monitoring, most company employees are sleeping, and threat actors are just clocking in. We talk more about these monitoring and detection tools in Chapter ##.

**Comprehensive Protection with Layered Security**

In today’s threat landscape, where ransomware is becoming increasingly sophisticated, relying on a single security measure is no longer effective, and it really hasn’t been for some time. A layered security approach, combining proactive measures such as employee training, advanced endpoint protection, and secure data backups, offers a more robust defense against ransomware attacks. This multi-faceted strategy ensures that even if one defense layer is bypassed, others will still provide protection.

By integrating these tools and practices, organizations can create a security infrastructure that minimizes vulnerabilities and maximizes their ability to prevent or quickly contain a ransomware attack. These layers aren’t just for prevention; they are also designed to give us (network defenders) the opportunity and time to identify the attack and slow down the attackers so that we can begin responding before they reach their target.

An analogy I often use is the visualization of a suburban house. There is usually a street, a curb, a sidewalk, a yard, maybe a fence, motion lights, the walls and doors of the house, locks on the windows and doors, maybe an alarm system or dog, and the people inside. These are all layers between a potential threat actor and the valuables inside the house. None of these layers can truly deter or stop a determined criminal from getting what they want. But they can help identify the attack, slow them down, and give the occupants time to respond.

The Role of Backups in Ransomware Defense

When it comes to ransomware defense, backups are your ultimate safety net. Imagine this: a ransomware attack encrypts all your critical business files overnight, halting operations and leaving you staring at a ransom demand. What do you do? You can restore your systems without paying a dime if you have complete and secure data backups. But if you don’t, or if your backups have been compromised, you’re at the mercy of the attacker. That’s why backups are a key component of any ransomware preparedness plan.

First and foremost, it’s critical that you have a well-defined backup strategy. There are several types of backups and each serves a specific purpose, described in [Chapter 3](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch03.html#chapter3_ransomeware_backup_and_recovery_basics_1746150934875416) – Backup and Recovery Basics. For example, full backups capture everything, and incremental or differential backups only record changes. Combining these methods is often the best approach for balancing efficiency and thoroughness. Additionally, secure storage is essential. Offline and immutable backups are critical to ransomware defense, as attackers often target connected backup systems during an attack.

However, having backups isn’t enough; they must be tested regularly. Imagine going to restore your data only to find out that the backup is incomplete or corrupted, or the method or restore isn’t working (e.g., tape drive, download, sync, etc.). Testing ensures you’re not just backing up your files but that you can actually recover them when it counts. Backup testing should be as routine as the backups themselves; without testing, backups are just a false sense of security.

Protecting your backups is also essential. Store them offline or in systems designed to be ransomware-resistant, such as immutable storage that attackers cannot alter or delete. Use encryption and access controls to safeguard them from unauthorized access. Alternatively, save critical data to USB or network-attached storage and keep it offline (unplugged) and in a fireproof safe between backups. By prioritizing backups, testing them regularly, and ensuring they’re protected, you’re building a critical line of defense against ransomware that can mean the difference between paying the ransom and a significantly less costly recovery.

Incident Response Planning

When the ransomware attack does happen, the clock starts ticking, and every moment counts. An effective incident response plan (IRP), and more specifically a ransomware playbook, can be calm within the storm and provide a coordinated recovery. An IRP ransomware playbook outlines what to do when the worst happens, including who to call, what steps to take, information about critical systems, and how to limit the damage. Without a clear plan, panic can set in, causing costly delays and wasted effort.

The first step in creating an IRP is assigning clear roles and responsibilities. Who will lead the response? Who are the stakeholders (i.e., management, system owners, etc.)? Who will communicate with stakeholders? Who will contact legal counsel, insurance, or law enforcement? These roles must be defined in advance so that when an attack occurs, everyone knows their job. Communication protocols are also essential. For instance, how will you notify employees, customers, and external partners, and what will you say? An IRP should include templates for internal memos, press releases, and ransom negotiation protocols to ensure consistent messaging.

Cyber insurance also plays an important role in modern ransomware response. Many organizations now rely on insurance policies to cover ransom payments, recovery costs, and legal fees. However, understanding your policy is critical. What’s covered, and what isn’t? Will the insurer handle negotiations with the attackers? Having this clarity before an attack is essential. Additionally, cyber insurers often require proactive measures like regular risk assessments and employee training to maintain coverage.

Law enforcement is also an important resource when dealing with a ransomware incident. Agencies like the FBI have extensive experience handling ransomware cases and may be able to assist in recovery. While reporting an incident can be daunting, it’s often a legal requirement and can provide critical support. As highlighted in Chapter 1, a multi-pronged approach combining preparation, insurance, and external collaboration can make all the difference in surviving a ransomware attack.

Network Segmentation and Access Control

Ransomware’s ability to spread across a network is what makes it so devastating. Once it gains access to a single system, it can quickly move laterally, encrypting files on shared drives, databases, and connected devices. Suppose the attack was a result of a network or application vulnerability that gave the attacker access to the network. In that case, the infection could be coordinated across all accessible network assets at the same time. This is where network segmentation and access control become crucial. Think of segmentation as building firewalls within your network that limit or contain the damage of the attack or infection to just that affected segment. It’s a proactive strategy that limits ransomware’s reach and protects your most critical systems.

Network segmentation involves dividing your network into smaller, isolated zones based on their function or sensitivity. For example, an organization might separate financial systems, customer databases, and employee workstations. If ransomware infects one zone, it cannot easily jump to others. Segregating high-value assets from less sensitive areas is one of the most effective ways to prevent malware from spreading during an attack.

Combining segmentation with access control further reduces ransomware’s ability to exploit user accounts. This includes using least-privilege access principles, where employees and systems are only granted the permissions they absolutely need to perform their tasks. If ransomware compromises a low-privilege account, its ability to spread is significantly limited. Multi-factor authentication (MFA) adds another layer of security by ensuring that even stolen credentials cannot be easily used.

Another approach to network defense is the adoption of a Zero Trust Architecture. This approach assumes that no user or device inside or outside the network should automatically be trusted. Instead, every access request is verified, monitored, and logged. As ransomware tactics grow more sophisticated, Zero Trust could be the right approach for some organizations. By combining network segmentation, strict access control, and continuous verification (Zero Trust), you can significantly reduce the impact of ransomware.

Post-Attack Recovery

When the dust settles after a ransomware attack, the real work begins. At this stage, every decision you make can impact how quickly and effectively you can get things back in working order. Whether you’re restoring systems from backups, decrypting files, or conducting an investigation, post-attack recovery is all about balancing urgency with caution. As noted earlier in this chapter, preparation plays a pivotal role, but even the best-laid plans require thoughtful execution after an attack.

The first priority is to restore your data and systems. If you’ve maintained secure, offline backups, this process can be straightforward. But it’s critical to approach recovery carefully. Before restoring any data, ensure that your systems are completely clean. You don’t want to reintroduce the ransomware into your network. This step requires coordination with IT teams or external forensic experts. As discussed in Section 3, testing your backup restores regularly ensures you’re not relying on faulty or incomplete backups when the time comes.

If your organization decides to pay the ransom, you’ll likely be working with a decryption key provided by the attackers. While this may unlock your files, there’s no guarantee that the process will work seamlessly or that your data won’t be corrupted. Moreover, paying the ransom doesn’t eliminate the possibility of attackers retaining or leaking your data, especially with the rise of double extortion tactics. As highlighted in Chapter 1, “paying a ransom is no guarantee of data recovery or security going forward.”

**Case Study – The Decryption Key Doesn’t Work**

A Fortune 500 organization suffered a ransomware attack that infected its core production systems. Management decided to pay the ransom primarily because the IT team indicated that the core system backups were corrupt and unusable for recovering the environment. Management coordinated with the company’s insurance provider to negotiate the ransom payment, and the threat actors provided the decryptor tool and basic technical support.

As the IT team and the threat actors worked to use the decryptor on the core production systems, they soon determined that the tool was not working, and the files remained unusable. This continued for two days until the threat actors suggested the company pay them an additional $100,000 for continued support, or they would disengage. The company decided not to pay the additional sum and continued to be offline for the next three weeks while the systems and data were rebuilt.

An after-action review identified the absence of decryptor testing (proof of life) prior to paying the ransom, which would normally have determined whether the tool was worth paying for.

Once your systems are operational, the next step is to conduct a forensic investigation to identify how the attack occurred. Was it a phishing email? An unpatched vulnerability? Understanding the root cause is essential to prevent future incidents. This process often reveals vulnerabilities that go unnoticed, offering valuable lessons for improving your defenses. For example, if ransomware spreads through a lack of network segmentation, this insight can guide critical updates to your infrastructure.

Recovery is not just about systems, it’s about learning and adapting. Document every step of the recovery process, from initial detection to resolution, and use it to refine your incident response plan and ransomware playbook. Share lessons learned across your organization and consider running post-attack simulations to test the effectiveness of your remediation efforts. Ransomware is a relentless and evolving threat, but every recovery strengthens your resilience for the future. If you’ve survived one, then you are more capable of surviving the next attack. Figure #.# is an example of a ransomware timeline for key events in responding to this type of attack. Not all of these activities apply to each organization or even different ransomware attacks but are intended to be a reference for your ransomware playbook, which we will talk more about in Chapter ##.

A diagram of a company

AI-generated content may be incorrect.

Figure 4-1. Ransomware Timeline of Key Events (Tuma, 2024)

Takeaways

Ransomware is an ever-present threat, but with the right preparation, response, and recovery strategies, you can significantly minimize its impact and even turn an attack into a learning opportunity. As we discussed throughout this chapter, effective ransomware defense requires a comprehensive approach that combines proactive measures like backups and network segmentation with the ability to respond swiftly when an attack happens.

Building resilience starts with being prepared. By implementing strong cyber hygiene practices and securing your backup systems, you’re laying the foundation for a robust defense. But preparation isn’t just about technology; it’s also about people. Training employees to recognize threats and act responsibly can close critical gaps in your defenses. At the same time, having a well-thought-out incident response plan and incorporating all the right resources ensures your organization is ready to respond effectively and recover quickly.

While no defense is perfect, the strategies outlined in this chapter provide a roadmap for reducing the risks and costs of ransomware attacks. Network segmentation and access controls can limit the impact and spread of a ransomware attack. Even if ransomware strikes, a solid post-attack recovery plan will allow you to bounce back stronger than before, learning from the experience and improving your defenses.

Ultimately, ransomware preparedness and response aren’t just about avoiding downtime or saving money. They’re about ensuring the long-term success and security of your organization. As the cyber threat landscape continues to evolve, staying vigilant and proactive will help you not only survive but thrive in the face of future challenges.

**References**

Tuma, S. (2024) The GC+CISO Connection. Self-published. ISBN 979-8-9913969-05 [*https://gccisoconnection.com/product/gccisoconnectionbook/*](https://gccisoconnection.com/product/gccisoconnectionbook/)

Minimize blast radius

Imagine you’re sitting at your desk, going about your day, when you begin to notice your desktop files changing and can’t access them. Then a ransom note pops up and demands payment in exchange for your data. It’s a nightmare scenario and one that’s becoming all too common. But ransomware doesn’t just appear out of thin air. It finds its way into environments where users don’t have proper awareness training, technical defenses are weak or outdated, security gaps have been overlooked, and attackers can exploit vulnerabilities without being detected.

That’s why effective training and technical measures are critical to reducing the impact of a ransomware attack, the blast radius. Think of them as ways to enhance your digital armor and increase your hit points by limiting the effectiveness of an attack. This chapter assumes compromise (i.e., the user clicked it, the vulnerable system was compromised, etc.) and focuses on implementing the controls that will diminish the impact of the attack while also slowing it down so we can react. Those include technical strategies, such as access controls, endpoint hardening, network segmentation, real-time monitoring, and threat detection. We’ll also discuss practical tips for isolating infections, mitigating lateral movement, and building long-term resilience against ransomware threats.

This isn’t about implementing flashy, overcomplicated solutions or spending endless amounts of money on the latest cybersecurity solutions. Instead, it’s about understanding your technology environment, layering proven defenses, understanding the risks, and making smart, strategic investments in your security posture. By the end of this chapter, you’ll have a clear roadmap for reducing the technical vulnerabilities ransomware attackers love to exploit and the confidence to defend your systems when it matters most.

Lastly, this chapter recommends some fairly technical solutions at a high level and is not intended to be a lengthy technical manual. Please use this information to search for and obtain the technical information, best practices, and procedures needed to implement the pieces that are applicable to your environment.

Incident Response Effectiveness

Regardless of which, if any, of the recommendations in this chapter you implement, please apply some forethought to the types of incidents that present the greatest impact to your organization (e.g., ransomware or data breach) and then determine what resources you currently have, the resources you need, and a plan to fill in the gaps for how to effectively respond to those incidents. Even the simplest thought for preparation is better than none and typically evolves into good discussions with stakeholders and broader realization of risk and areas for improvement.

The speed and effectiveness of your response determine whether the impact is minimal or catastrophic. A well-prepared incident response (IR) plan can help contain the ransomware, mitigate damage, and restore operations as quickly as possible.

This section outlines the essential steps of a ransomware-specific incident response strategy, focusing on containment, eradication, recovery, and future prevention.

**Know Thy Self**

It is near impossible to effectively defend an environment with which you are not familiar. It is imperative that you know what you are protecting, the ingress and egress points, the priority and dependencies of different assets, and the capabilities of your team and support partners.

Things to consider:

* Develop and maintain a detailed asset list that includes servers, workstations, network architecture, applications, cloud assets, and their respective stakeholders and criticality.
* These inventories can be incorporated into several of the other sections in this chapter, including EDR, vulnerability assessments, and network monitoring for example. The output from these activities will help maintain the accuracy of your inventories.
* Knowing your environment, your business, the data, systems, and applications required to support the company and the resources you have will allow you to develop a plan for incident response.
* Developing the incident response plan will also identify gaps in resources and capabilities that must be addressed “today”. These gaps are commonly skill and technologies, but may also include cyber insurance coverage, law enforcement contacts, and outside legal counsel.

**Preparing for a Ransomware Incident**

An effective response starts long before an attack happens. Organizations need a clear, well-documented incident response plan (IRP) that outlines roles, responsibilities, and response procedures.

Key preparation steps:

* Develop a Ransomware Playbook:
  + Define specific actions for IT, security, legal, and executive teams.
  + Include decision trees for scenarios such as partial vs. full network compromise.
* Assemble an Incident Response Team (IRT):
  + Assign roles to security analysts, system administrators, legal counsel, public relations, and senior management.
  + Conduct periodic training exercises and tabletop simulations.
* Establish Communication Protocols:
  + Define secure communication methods in case internal systems are compromised.
  + Consider using out-of-band communication channels like encrypted messaging apps (e.g., Signal) or external email services.
* Engage External Support:
  + Establish relationships with cybersecurity firms, legal advisors, and law enforcement before an attack occurs. Many of the professional services firms that support incident response will do a $0 retainer in order to get all the paperwork out of the way ahead of an incident. Note: doing the paperwork while the house is on fire is typically a 2 to 6 hour delay.

Incident response will only be as effective as the resources and knowledge you bring to the bear. Even if it’s just an outline in a notebook with phone numbers, IP addresses, and important steps it is better than trying to objectively and calmly think through those things during the stress an potential chaos of a ransomware event. Create or update your IRP today, including all of the applicable resource, contacts, and procedures.

Endpoint Security – The First Line of Defense

When it comes to ransomware attacks, endpoints are often the most vulnerable and likely entry points. Whether it’s a workstation, a smartphone, or an IoT device, attackers frequently target these systems because they’re the weakest link in most organizations’ defenses. The good news is that a well-configured endpoint security strategy can significantly reduce the risk of an infection and contain its spread if ransomware does find a way in.

In this section, we’ll explore practical, technical recommendations for securing endpoints and turning them from liabilities into fortified checkpoints.

**Deploying Endpoint Detection and Response (EDR) Tools**

Modern ransomware strains are designed to evade traditional antivirus solutions. That’s where Endpoint Detection and Response (EDR) comes into play. Unlike basic antivirus software, EDR tools monitor endpoint activity in real-time, looking for suspicious behavior such as unusual file encryption, changes to registry keys, or attempts to disable security tools. Newer EDR solutions also incorporate AI and sandboxing in their analysis capabilities, letting the files execute in a safe environment to determine how it behaves.

Key steps for implementing EDR:

* Before deploying an EDR solution, the organization must define policy, objectives, and how EDR will be utilized and enforced. An analogy for this would be ensuring that once you buy the car, you are familiar with the motor vehicle laws that govern its use. Applicable policies include Acceptable Use, Information Security, and Incident Response.
  + **Acceptable Use Policy**–Effective policy must include acceptable use to define how the organization’s devices can and cannot be used, and how to communicate issues, incidents, and observations to the appropriate personnel. The appropriate personnel to report incidents to are often the organization’s Information Security Team, which is governed by the Information Security Policy and Incident Response Plan.
  + **Information Security Policy** – This policy is typically comprehensive of all information security objectives and controls across the company, such as anti-malware, vulnerability management, and system hardening. These policy components must be defined and adhered to for an EDR solution be proactively effective and reactively timely. EDR-specific provisions typically define the solution’s configuration, technical policy for end-point restrictions (e.g., USB device connections, file downloads, and installations), management, monitoring, and response (e.g., quarantine, auto-deletion).
  + Management of the EDR solution is critical to ensure that it is maintained, updated, and monitored. Monitoring is important to identify evolving user behaviors, such as continuous policy violations or alerts, and timely responses to identified threats.
  + **Incident Response Plan** – The plan identifies the people, processes, and technologies necessary to respond to incidents. It includes contact information, roles, responsibilities, escalation paths, and expectations. The plan would ideally also include threat-specific playbooks for the most likely and impactful threats, like ransomware, and include detail for specific response requirements.
* Choose a solution that integrates with your organization’s broader security stack. As with any new technology, it must play well with your existing environment and users, and be manageable. Considerations for choosing the EDR should include an analysis of its capabilities (centralized management console, policies, AI, sandboxing, offline mode), the overhead it creates for the host, and integration with a SIEM.
* Configure the EDR to automatically isolate suspicious files, applications, and devices to prevent lateral movement. The effectiveness of an EDR solution is in its ability to timely identify and quarantine or prevent malicious activities. The EDR solution must be configured to be as aggressive as possible in quarantining files and devices. Implementing this approach will require coordinate and patience with users, ensuring that all trusted files and applications are whitelisted during the initial implementation and follow a formal process for reviewing future changes.
* Manage the EDR. Implementing the EDR is not a point-in-time activity and must be managed on a daily basis. Once the baseline policies are implemented, the EDR must be reviewed daily to ensure it is current, endpoints are updated and checking in, and alerts are timely reviewed and responded to.
* Use the EDR’s threat intelligence capabilities to keep up with the latest ransomware tactics. Well-established EDR solutions, those big names with millions of deployments around the world, collect tons of data and perform extensive analysis that helps make the EDR solution better and contributes to threat intelligence reports and updates. Managing the EDR solution also involves staying abreast of the current and evolving threats to endpoints.

**Using Next-Gen Anti-Malware Solutions**

Antivirus and anti-malware software are an important layer in a ransomware defense strategy. While these tools alone may not be sufficient to stop sophisticated ransomware campaigns, they are critical for detecting and blocking known threats, particularly in the early stages of an attack. When properly configured and maintained, they act as a strong safety net, reducing the risk of infection and giving your organization valuable time to respond.

**Role of Real-Time Scanning**

Real-time scanning is one of the most effective features of antivirus and anti-malware tools. It works by continuously monitoring your systems and files for any suspicious or malicious activity. When ransomware attempts to execute or encrypt files, real-time scanning can detect and block the malware before bad things happen. EDR solutions that incorporate sandboxing and AI have become very effective at real-time scanning protection.

Real-time scanning differs significantly from the traditional approach of scanning the entire device every time. Instead, the EDR does a full system scan on day one to establish a baseline and then only scans what changes or is interacted with. This approach makes better use of device resources and limits the impact on the user.

**Advanced Features**

Many anti-malware tools now include advanced capabilities specifically designed to combat ransomware. When selecting or upgrading your solutions, look for the following features:

* **Heuristic Analysis**:  
  Detect threats based on their behavior rather than relying solely on known signatures. This is critical for identifying new or unknown ransomware strains.
* **Ransomware Rollback**:  
  Some modern solutions provide a rollback feature that allows systems to revert to a pre-infection state. This is particularly useful for minimizing downtime and recovering encrypted files.
* **Machine Learning Integration**:  
  AI-powered threat detection analyzes patterns in malware behavior to detect and block zero-day ransomware attacks.
* **File Integrity Monitoring (FIM)**:  
  Monitor critical files and systems for unauthorized changes, such as encryption attempts or sudden file renaming.
* **Cloud-Based Threat Intelligence**:  
  Leverage real-time updates from global threat intelligence networks to identify and block the latest ransomware variants.

**Configuration Best Practices**

Even the best anti-malware solution can fail if not configured correctly. To maximize the effectiveness of an EDR, consider the following best practices:

* Enable Full Coverage:
  + Ensure real-time scanning is enabled on all endpoints, including desktops, laptops, mobile devices, and servers.
  + A specific note here to address the gap in coverage at most organizations is with mobile devices. We allow users to receive company email and files on their mobile devices, but we often fail to require equal anti-malware protection on these devices. Even if the devices are not owned by the company, anti-malware protection must be required by policy before the user can access company data from these devices.
  + Extend coverage to cloud-based storage solutions where ransomware can reside.
* **Scan Network Shares**:  
  Ransomware often targets shared drives. Enable scanning of network shares to identify and block threats attempting to propagate through shared folders.
* **Quarantine Suspicious Files**:  
  Configure the software to immediately isolate and quarantine suspicious files rather than attempting to delete them. Quarantine allows security teams to analyze the file before deciding on the next steps.
* **Schedule Regular Full-System Scans**:  
  Set up regular full-system scans during off-peak hours to catch dormant malware or threats that may have evaded real-time detection.
* **Whitelist Known Safe Applications**:  
  Reduce false positives by whitelisting trusted applications while maintaining strong monitoring for unknown or unauthorized software.

**Ensuring Anti-Malware Are Up-to-Date**

Ransomware evolves rapidly, and attackers are constantly finding new ways to bypass outdated software. Keeping your anti-malware tools updated is critical to ensuring they can recognize and respond to the latest threats.

* **Automatic Updates**:  
  Enable automatic updates for both the EDR software and its threat signature database across all endpoints.
* **Vendor Communication**:  
  Stay in contact with your software vendor to receive notifications about major updates or newly discovered vulnerabilities.
* **End-of-Life Management**:  
  Replace outdated software no longer supported by the vendor, as it will lack updates and pose a security risk.

**Integrating Antivirus with a Broader Security Ecosystem**

While antivirus and anti-malware solutions are essential, they work best when integrated with other security measures. For example:

* Feed the EDR threat data into a **Security Information and Event Management (SIEM)** system to analyze patterns and uncover broader attack campaigns.
* Use them alongside **firewall and network monitoring tools** to block malicious payloads before they reach endpoints.

**Common Mistakes to Avoid**

* **Relying Solely on Default Settings**:  
  Default configurations may not align with your specific environment or needs. Customize settings for your unique use cases and company workforce environment.
* **Delaying License and Maintenance Renewals**:  
  Expired EDR software is effectively useless. Ensure timely renewals to maintain uninterrupted protection.
* **Ignoring Alerts**:  
  Don’t treat EDR alerts as noise; each alert deserves investigation to ensure threats aren’t missed, and any accepted anomaly is documented.
* **Running Multiple Antivirus Programs**:  
  Installing multiple antivirus solutions on the same device can cause conflicts and reduce the overall effectiveness of both.

By effectively utilizing antivirus and anti-malware solutions with real-time scanning and the advanced features available today, organizations can significantly reduce their risk of ransomware infections. While these tools are not a standalone solution, they are a foundational part of a comprehensive ransomware defense strategy. Combined with other endpoint and network security measures, they can provide critical protection against even the most determined attackers.

**Importance of Endpoint Hardening**

Endpoints are only as strong as their configuration. These configurations must be intentional, documented, and periodically reviewed to ensure they remain relevant and maintained. Policies, such as an IT Asset Policy and/or the organization’s Information Security Policy, should also require documented configuration standards. By hardening these devices, you can reduce the likelihood of exploitation and increase the ability to identify deviations from the build standards.

Key hardening techniques:

* **Role-based configurations**: Each endpoint should be built and configured for its purpose, and these build standards should be documented, approved, and used for all new device builds.
* **Disable Unused Ports and Services**: Minimize the attack surface by disabling unnecessary features like Remote Desktop Protocol (RDP) and file sharing.
* **Enforce Application Whitelisting**: Allow only pre-approved applications to run, blocking unauthorized or malicious software.
* **Control USB Device Access**: Use tools to restrict unauthorized USB storage devices, which are common vectors for ransomware delivery.

**Role-based Configurations**

Endpoints, such as servers, workstations, laptops, smartphones, and IoT devices, are common entry points for ransomware attacks. By hardening endpoints, you can eliminate weaknesses that attackers exploit to deliver ransomware, execute malicious code, or propagate an infection across a network. Endpoint hardening involves reducing the attack surface, securing configurations, and limiting opportunities for attackers to exploit vulnerabilities.

For each endpoint or endpoint type (e.g., accounting PC or Webserver), the organization should define the appropriate hardware configuration, BIOS settings, operating system (O/S), O/S configuration, authorizations, third-party applications, backup and recovery strategy, and network connectivity and placement. Ideally, the secure build guideline would also require a build checklist to be completed and archived for each device.

**Disabling Unused Ports and Services**

Many endpoints come with unnecessary ports, services, software, and features enabled by default, creating potential entry points for attackers. Removing or disabling these reduces the number of pathways ransomware can use to infiltrate a device.

Using the documented secure build guideline for each device, begin disabling or removing everything that is unnecessary or doesn’t support the device’s purpose.

Key steps:

* **Turn Off Unneeded Services**:  
  Services like remote desktop, file sharing, and print sharing are often not required by all endpoints. The National Institute of Standards and Technology (NIST) offers best practice guides for various devices. In January 2025, the US Cybersecurity and Infrastructure Security Agency (CISA) published guidance for IoT devices.
* Services running on a device consume resources like memory and processing but are also potential targets for system exploitation. FTP and Telnet services, for example, are configured by default to always listen for an incoming connection; if an authentication configuration file wasn’t created, null access may be allowed. Disabling unnecessary services, like FTP and Telnet, will often disable unnecessary communication ports, like ports 21 and 23.
* **Disable Unnecessary Ports**:  
  Communication ports are used by specific services and applications to establish a connection between devices on a network. These ports can also be used by malware to send and receive data during an attack. Use a network scanner, like Wireshark, to identify open ports and either disable the associated service or applicable or close the ports that aren’t in use, such as FTP (port 21), Telnet (port 23), or unused RDP (port 3389).
* **Host Firewall Configuration**:  
  Configure endpoint firewalls, such as the Windows firewall, to block inbound and outbound traffic on unused ports and restrict traffic to trusted IPs. This can be managed as part of the secure build guideline and the deployment of active directory group policy (GPO).

**Enforcing Application Whitelisting**

Ransomware often relies on executing unauthorized or malicious programs on endpoints. Application whitelisting is one of the most effective ways to prevent this, as it only allows pre-approved programs to run on a device – as you no doubt defined in your role-based secure build guideline.

Steps for implementing application whitelisting:

* Only allow software to be obtained from legitimate sources, ideally centralized with the IT or technical department for approval, procurement, and deployment. This workflow will also ensure an accurate inventory of software, licensing, uses, and stakeholders.
* Use tools like Microsoft AppLocker or third-party solutions to create and manage whitelists. An EDR solution can also support whitelisting applications through policy.
* Block executable files from running in directories commonly used by ransomware, such as temporary folders (%AppData%, %Temp%).

**Controlling USB Device Access**

USB devices are a common vector for ransomware delivery. Attackers can use USB drives loaded with malicious payloads to infect endpoints. Infected mobile devices that connect to PCs through a USB port may cause an infection. Limiting USB device access minimizes the risk of malware infection and spread.

Best practices:

* **Restrict USB Access**:  
  Use endpoint management tools, like Microsoft InTune, or operating system policy to disable USB ports unless explicitly required. Alternatively, the PCs could be configured to allow USB power charging but not to allow for transferring data. Some EDR solutions can also enforce policy that restricts the use of USB devices.
* **Enable Read-Only Mode**:  
  Configure USB ports to allow only read access where necessary, preventing malicious files from being written to the endpoint. There aren’t very many use cases for a read-only USB port, and this approach really only prevents the spread of malware from the PC to the USB, not necessarily the other way around.
* **Monitor USB Usage**:  
  Use data loss prevention (DLP) solutions to track and control USB device activity. Some EDR and SIEM solutions can alter on USB usage if the event log data is available.

**Hardening Operating System and Software Configurations**

Default settings in operating systems and applications often prioritize usability over security. Hardening these configurations ensures your endpoints are less vulnerable to ransomware attacks.

Key steps:

1. Secure Operating Systems:
   1. Disable unnecessary features like Windows PowerShell, which is often exploited by ransomware.
   2. Enable security features such as BitLocker encryption (for Windows) or FileVault (for macOS).
   3. Turn on secure boot to ensure only trusted software loads during startup.
2. Limit Macro Execution:
   1. Ransomware is commonly delivered through malicious macros in Office files. Disable macros unless required, and enforce policies to only allow macros signed by trusted developers.
3. Remove Bloatware:
   1. Pre-installed applications that aren’t needed increase the attack surface. Uninstall or disable unnecessary software.

**Enforce Strong User Permissions**

Improperly configured user, system, and service accounts often provide attackers with an easy way to escalate privileges or access sensitive data.

Steps to enforce strong permissions:

1. Principle of Least Privilege:
   1. Ensure that accounts only have the access and permissions they need for their roles. Don’t let a vendor persuade you to provide an account, like a service account, with administrative privileges during deployment or troubleshooting – these are often targets of threat actors and malware.
   2. Regularly audit and revoke unused permissions.
2. Limit Administrative Access:
   1. Reduce the number of users with administrator privileges to an absolute minimum.
   2. These administrator accounts must also be named, like MikeAdm or SuperMike instead of using the built-in admin accounts. This creates accountability and traceability for admin activities, also known as nonrepudiation.
   3. Use separate accounts for administrative tasks and everyday use.
   4. Use separate credentials between the production environment and any test or development environments. Also, Test/Dev credentials should not have access to any production environments.
3. Implement Role-Based Access Control (RBAC):
   1. Define access levels based on job roles, ensuring that employees have access only to what’s necessary for their responsibilities. This can be done on a case-by-case basis in smaller organizations. In larger organizations these roles should be built into templates with input from stakeholders and approved by management. In either scenario, be consistent with your approach to limiting access and you will then only need to manage the approved exceptions.

**Regular Security Patching**

Outdated software is one of the most significant risk factors for ransomware infections. Unpatched endpoints leave vulnerabilities open for exploitation.

Effective patch management:

* Define a formal Patch Management Policy and Procedure. This establishes approval from stakeholders for the how and when patches are applied, and accountability for getting it done.
* Create and maintain an inventory of all the software in the organization, including firmware, operating systems, databases, and apps.
* Ideally you want to automate patch deployment across all endpoints using endpoint management tools. These can be expensive so alternatives are upgrade your remote management tools like Kaseya or Manage Engine to add the patch management module.
* Another good practice is to incorporate patch management into your vulnerability management process. After every vulnerability scan (OpenVAS or Nessus are free scanners), use the results related to missing patches to help prioritize your efforts.
* Prioritize patches for critical vulnerabilities, especially those with active exploits in the wild. However, don’t discount non-critical findings because even the mediums and lows can add up to potential entry points.

**Endpoint Isolation**

If an endpoint is compromised, quick isolation can prevent ransomware from spreading to other systems.

Key isolation strategies:

* Use endpoint detection and response (EDR) tools to automatically isolate infected devices.
* Configure VLANs or network segmentation to limit communication between endpoints.
* Develop incident response playbooks with clear procedures for isolating infected devices.
* Implement a SIEM or OpenXDR solution, like Stellar Cyber, that has security orchestration and automated response (SOAR) capabilities. Then configure the SOAR playbook to automatically address compromised user accounts and endpoints. SOAR playbooks could also modify firewall rules and cloud infrastructure configurations where applicable.

**Monitoring Endpoint Activity**

Continuous monitoring helps identify suspicious activity early, reducing the chance of a full-scale ransomware attack.

Best practices:

* Deploy endpoint monitoring tools to log file changes, application behavior, and network activity. Ideally these endpoint tools, like an EDR, would send their logs to a SIEM to be correlated along with events and logs from other data sources.
* Enable alerts for common ransomware behaviors, such as mass file encryption, registry modifications, or attempts to disable security tools.
* Regularly review logs and reports for indicators of compromise (IOCs).
* The 24x7 monitoring is often a missing component in most organizations, especially small to medium sized businesses. Threat actors know this and will wait until after hours or weekends to make their move. One of the important key to surviving an incident is timely identification that leads to timely and appropriate response.

Implementing these endpoint-hardening techniques, organizations can significantly reduce the attack surface available to ransomware. Hardened endpoints make it more difficult for attackers to compromise systems and limit the potential damage if an attack does occur. In the next sections, we’ll explore how network security and other technical measures can further improve your ransomware defense strategy.

Network Security – Limiting Lateral Movement

While endpoints are often the initial targets of ransomware attacks, a compromised network can amplify the impact, enabling ransomware to spread laterally and disrupt operations on a much larger scale. Strengthening network security is a critical part of reducing the impact of ransomware, as it provides the digital equivalent of locks, speed bumps, and alarms to keep attackers out or slow them down and contain them if they get in.

In this section, we’ll cover essential network security measures to reduce the risk and impact of ransomware, prevent lateral movement, and maintain control over your IT environment.

**Network Segmentation: Containing the Blast Radius**

One of the most effective ways to limit the spread of ransomware is through network segmentation. By dividing your network into smaller, restricted segments, you can reduce the ability of ransomware to move laterally and infect other systems.

Key steps to implement segmentation:

* **Create Zones Based on Sensitivity**:  
  Separate critical systems (e.g., financial databases, production servers) from general-purpose systems like user workstations or guest Wi-Fi.
* **Use VLANs**:  
  Implement Virtual Local Area Networks (VLANs) to logically separate traffic within your network. Common examples include putting the VoIP phone traffic on its own network, different floors of an office building, and the data backup network.
* **Restrict Communication**:  
  Limit communication between segments to only what is absolutely necessary. For example, end-user devices should not directly communicate with servers unless explicitly required, and the production network should not directly communicate with the test or development environments.
* **Micro-segmentation**:  
  For high-security environments, use micro-segmentation to enforce communication policies at the application level.

Most of these are easier to implement in virtual or cloud environments that were built with these in mind. However, most network devices are capable of creating VLANs and their interfaces are fairly intuitive.

**Firewalls and Traffic Control**

Firewalls act as the first line of defense in monitoring and filtering network traffic to prevent unauthorized access. A properly configured firewall can block ransomware-related activity, such as downloading payloads or communicating with command-and-control (C2) servers.

Best practices for firewalls:

* **Deploy Next-Generation Firewalls (NGFWs)**:  
  NGFWs provide advanced features like deep packet inspection, intrusion prevention, and application awareness to identify and block malicious activity. The effectiveness of these capabilities requires good maintenance and management to ensure bad-IP lists are current, rules are updated and appropriate, and any changes follow a formal review and approval process.
* **Restrict Outbound Traffic**:  
  Block outbound connections to unapproved or suspicious IP addresses and domains to prevent ransomware from communicating with C2 servers. Some firewall vendors provide updates to know bad-IP and C2 server IPs, which can change several times a day. There are also subscription services that ensure you have the most accurate list of IPs to block.
* **Use Geo-Blocking**:  
  If your organization doesn’t interact with certain regions, block traffic to and from those areas to reduce exposure to international ransomware threats. Though this highly effective at stopping the automated attacks, if bad guys really wanted to attack you then they will find a compromised host within your geo location and use that as an attack point.
* **Log and Monitor Activity**:  
  Regularly review firewall logs for unusual traffic patterns or blocked attempts that may indicate reconnaissance or attempted ransomware delivery. Most firewalls have a good dashboard, but the view is focused on network traffic without context like user behavior, EDR alerts, etc. Ingesting your firewall logs into a SIEM or OpenXDR platform is the most effective way of monitoring your firewall while also correlating those events with the cybersecurity events from other data sources (yes, I will continue to harp on getting a SIEM-type solution throughout the book).

**Network Monitoring and Behavioral Analytics**

Continuous monitoring of your network is essential for detecting ransomware-related activity early. Threat hunting adds a proactive layer by identifying potential threats before they escalate. The behavioral analytics capability is a newer and evolving approach to detecting the precursors to a ransomware infection before it happens.

Steps for effective monitoring and threat hunting:

1. **Deploy SIEM Tools**:  
   Use Security Information and Event Management (SIEM) solutions to aggregate and analyze logs from across your network. The more data sources the better, and should include at a minimum, the firewall, Active Directory or other authentication source, endpoint data (EDR), and server logs.
2. **User and Event Behavior Analytics (UEBA):**Some SIEM or OpenXDR platforms are UEBA capable and establish baselines for each user, device, and network segment. Even if a behavior isn’t yet a security concern, if it deviates from the baseline, it will alert you. Examples include robotic beaconing, which may hit an IP address 500 times a second, which is not humanly possible and likely abnormal for Bob’s laptop. Another example might be Sally downloading large files or files with extensions that she has not done before.
3. **Monitor for Common Indicators of Compromise (IOCs)**:  
   Examples include large numbers of encrypted files, unusual lateral movement, or unauthorized privilege escalation.
4. **Monitor Common Attack Vectors**:  
   Remote access (RDP, VPN, RMM), email, all internet-facing systems, and user downloads.
5. **Automate Alerts and Response**:  
   Set up automated alerts for activities associated with ransomware, such as mass file renaming or spikes in outbound traffic. Ideally, someone would be available 24x7 to receive these alerts and respond accordingly. Alternatively, the SIEM could be configured with a SOAR playbook to automatically respond.

**DNS Filtering**

Ransomware is often delivered through malicious websites or phishing emails with links to compromised domains. DNS filtering and web security tools can block these threats before they reach users.

Best practices for DNS filtering:

* **Block Known Malicious Domains**:  
  Use threat intelligence feeds, like DarkTrace or OpenDNS, to block access to domains associated with ransomware campaigns.
* **Enforce Safe Browsing Policies**:  
  Implement an Acceptable Use Policy to define what is acceptable internet use and what is expressly prohibited. Restrict access to prohibited, high-risk categories, such as file-sharing, adult content, or newly registered domains.
* **Monitor DNS Queries**:  
  Look for suspicious DNS queries, such as frequent lookups for random domain names, which may indicate ransomware beaconing to C2 servers.

**Network Vulnerability Assessments**

Routine assessments help identify weaknesses that ransomware attackers could exploit. This is important because ransomware is not only delivered by email, and the most significant ransomware attacks typically start with a network intrusion that allows attackers time to perform reconnaissance and staging for their coordinated ransomware deployment.

Best practices for vulnerability assessments:

*Conduct a Comprehensive Internal and External Network Vulnerability Assessment*

All internal network assets and all internet-facing network assets must be scanned in order to fully assess the network. You wouldn’t hire a home inspector and ask that they only review the front of the house, you want to know where all the potential issues are before you decide to live there.  
  
Do these assessments as often as possible as technology changes every day, attackers find new ways of exploiting our systems every minute, and our users present a constant risk to the network. The appropriate frequency should be unique to your environment and the availability of resources (time, skill, budget, etc.), but I would recommend at least quarterly plus anytime the environment changes (e.g., new switch, firewall, internet facing application, etc.).

Effective vulnerability assessments will find open ports, running services, software versions and missing patches, and configuration issues among other things. Of course, conducting the assessment is only part of the process… now verify, prioritize, and fix these issues.

*Network Penetration Testing:*

At least annually, simulate an attack on your network. An effective penetration test will be conducted by a third-party familiar with threat actor tactics, techniques, and protocols (TTPs). The approach will include conducting reconnaissance for find data from the internet about your company, the networks, DNS leaks, registrations, social media, etc,.and use this information to design the simulated attack. During the attack, vulnerabilities are exploited to gain access and demonstrate the extent and impact of unauthorized access. The exercise also assesses your ability to detect the attack and the effectiveness of your response to stopping it.

By implementing these network security measures, you can significantly reduce the risk and impact of a ransomware attack. A well-segmented, monitored, and actively defended network are a substantial part of an effective ransomware defense strategy. Next, we’ll explore how access control and identity management play a important role in ransomware prevention and containment.

Access Control & Privilege Management

Ransomware attacks often exploit weak or misconfigured access controls to gain entry and escalate privileges within an organization. Attackers frequently use stolen credentials, brute-force techniques, or privilege escalation to move laterally and maximize damage. Implementing strong access control and identity management practices is critical to minimizing ransomware risks and limiting its spread.

In this section, we’ll cover essential strategies to secure user accounts, enforce strict authentication, and prevent ransomware operators from exploiting identity-based weaknesses.

**Implementing the Principle of Least Privilege**

One of the most effective ways to limit ransomware impact is by restricting user permissions to the minimum level required for their job functions. The more privileges an account has, the more damage an attacker can do if they compromise it.

Best practices for enforcing Least Privilege:

* Restrict Administrative Privileges:
  + Minimize the number of users with administrative rights.
  + Use separate admin accounts instead of giving standard user accounts admin privileges.
  + Don’t allow users to be local administrators.
* Role-Based Access Control (RBAC):
  + Assign permissions based on job roles rather than individuals.
  + Regularly review and update role assignments to ensure users only have necessary access.
* Monitor and Audit Privileged Accounts:
  + Use privilege management tools to track administrative access and detect anomalies. SIEM platforms can alert on privileged access policies, especially one that has UEBA capabilities.
  + Implement just-in-time (JIT) access, granting admin privileges only when needed and revoking them automatically after use.

**Enforcing Multi-Factor Authentication (MFA)**

MFA is one of the better defenses against credential-based attacks. Even if an attacker steals a password, they cannot access the account without an additional authentication factor. Of course, the implementation of MFA and related polices must also be effective.

Key MFA recommendations:

* Require MFA for All Access Points:
  + Enforce MFA on everything possible, including domain authentication, remote access (VPN, remote desktop, etc.), cloud services (O365, social media, backups, etc.), and applications.  
      
    MFA must be accompanied by well-designed technical and administrative policies to be effective. The administrative policy must define how MFA will be used, like all logins will require MFA every time. This will drive the technical policy that prohibits a device or user from being “trusted” or “remembered” on subsequent logins that wouldn’t then require MFA.
* Use Phishing-Resistant MFA Methods:
  + Hardware security keys (e.g., YubiKey), MFA or Authenticator applications, or biometrics are more secure than SMS-based authentication, which can be intercepted.
* Monitor Failed MFA Attempts and Credential Theft:
  + Unusual patterns of failed MFA logins may indicate an attacker attempting to bypass security. SIEM platforms can alert on MFA failures and abnormalities (UEBA) as part of their authentication policies.
  + Subscribe to dark web monitoring services or solutions (e.g., RecordedFuture, Searchlight) for alerts related to possible credential theft of employees, administrators and executives. Something to consider here and to incorporate in company password policy is to require that company passwords be especially unique and not coincidental with any other account. The policy could also extend to vendors, especially managed service providers who may use the same credentials across multiple clients.

Enforcing strict access control policies and strengthening identity management, organizations can significantly reduce the risk of ransomware infiltrating their network through compromised credentials. In the next section, we’ll explore incident response strategies to contain ransomware and mitigate its damage when an attack occurs.

Data Protection Strategies

Ransomware primarily focuses on two tactics, encrypting your data so that it can’t be used, and or stealing your data and using it for leverage. In both scenarios the focus is your data. So, what can be done to limit the impact of ransomware on your data?

**Data Encryption**

Encrypting your data is a common recommendation for protecting it and making it useless to attackers. However, it is important that you understand how and when encryption protects your data, especially in a ransomware attack.

Data encryption on end points, like Microsoft BitLocker for workstations and severs, is only effective when the storage drive is not mounted and the user is not logged in. This is important to note when a user logs into a workstations and becomes infected with ransomware that also steals the data from the device. Even if BitLocker is used to encrypt the workstation storage drive, the data for that user is unencrypted while the user is logged in and would also be available to the ransomware in this unencrypted state. Tying this back to ensuring users’ access levels are appropriate, if they are logged in as the local administrator then the ransomware would have access to all the data on the device.

**Virtualization**

Virtualization can be a powerful tool for protecting your data through frequent snap-shot backups, but also by using the virtual host environment to isolate infected systems.

1. **Isolation and Containment**  
   Virtual machines (VMs) can be isolated from one another and the host system. If a VM becomes infected with ransomware, it can be isolated, preventing it from spreading to other VMs or other parts of the network.
2. **Snapshots and Rollbacks**  
   Virtualization platforms often support snapshots, which are point-in-time copies of a VM’s state. A VM that becomes infected can be rolled back to a previous snapshot very quickly. This minimizes downtime and data loss and is a very effective response if you are confident in the integrity of the snapshot being used (i.e., it’s not also infected).
3. **Segmentation**: By segmenting critical applications and data into separate VMs, you can limit the impact of a ransomware attack.

**Data Restore Testing**

Ensure that whatever your data backup strategy is (VM snapshots, cloud sync, tape, etc.), that you frequently test your ability to recover your data. Many times we become complacent in the “backup completed successfully” message without ever testing our ability to recovery or realizing what may be involved with that recovery effort, including time, additional storage space or VMs, network bandwidth, or even additional software licenses in some cases. Again, these are all things that should be well understood and tested before an incident occurs.

Incident Containment and Isolation Techniques

When a ransomware incident is detected, timely and effective containment is essential to minimize damage and prevent the attack from spreading further. This section outlines practical techniques for isolating compromised systems, disabling vulnerable network functions, and automating the containment process to help security teams respond quickly and efficiently.

**Rapid Isolation of Infected Devices**

The first and most critical step in incident response is to immediately isolate any device that shows signs of ransomware infection. Quick isolation prevents the malware from propagating across the network and encrypting additional files.

* Target Identification
  + Determine which devices are infected, their purpose, the data they contain, the type of user(s), system owner(s), and what the device(s) have access to.
  + Knowing these attributes will aid in making quick decisions about related users, devices, and data.
* Disable the Infected User’s Account
  + Consider disabling the user(s) associated with infected systems and forcing a revocation of all of their authenticated devices and MFA tokens. This could aid in limiting the ransomware’s access to network resources as well as any unauthorized external connections.
* Power Off
  + If possible, the most effective action to take on a system infected with ransomware is to completely power it off, unplug it, take out the battery, etc. Ransomware encryption does not work without power and the remainder of the data and any evidence (e.g., dropper file or log data) may be salvageable.
  + Some potential investigative evidence may be lost, such as volatile memory, but the loss of critical data may outweigh this.
* Disconnection and Quarantine
  + If the system can’t be powered off then disconnect the affected device from all network connections (wired, wireless, VPN) as soon as an infection is detected.
  + If feasible, physically unplug network cables or disable Wi-Fi adapters.
  + Move the infected device to a dedicated quarantine VLAN or a separate network segment designed for compromised systems.
  + Ensure that the quarantine environment restricts access to sensitive resources and limits lateral movement.

**Disabling Network Shares and Stopping Compromised Processes**

Ransomware often exploits shared network resources to move laterally and encrypt files across multiple systems. Disabling these shared services and stopping malicious processes can significantly reduce the potential spread of the infection.

* Disable Network Shares
  + Immediately unmount or disable network drives and shared folders that the infected device can access. Knowing what the user of the infected device has access to may be helpful.
  + Temporarily restrict file sharing protocols (e.g., SMB) on both the affected device and the network to block unauthorized file access.
* Terminate Malicious Processes
  + Use Endpoint Detection and Response (EDR) tools or manual process management (e.g., Task Manager on Windows, command-line utilities on Linux) to identify and kill any processes associated with ransomware.
  + Monitor for recurring malicious processes and enforce policies that prevent unauthorized application execution.

**Blocking External Communication to Command-and-Control Servers**

Ransomware frequently relies on external command-and-control (C2) servers to receive instructions or exfiltrate data. Blocking these communications is vital to stopping the attack’s progression and hindering the attackers’ ability to control the malware. Remember, though the device is infected it may not have access to the C2 servers needed to negotiate the encryption keys needed to begin encrypting the victim’s data.

* Firewall and Proxy Configurations
  + Update firewall rules to block outbound traffic from infected devices to known malicious IP addresses and domains associated with ransomware C2 servers.
  + Utilize web proxies and DNS filtering to intercept and block requests to suspicious or untrusted domains.
* Network Monitoring:
  + Deploy intrusion detection systems (IDS) or a SIEM with IDS and UEBA capabilities to monitor unusual outbound traffic patterns.
  + Immediately block any traffic that deviates from established norms, especially if it appears to be attempting to connect to external C2 infrastructure.

**Automating Containment with SOAR Tools**

Security Orchestration, Automation, and Response (SOAR) platforms can significantly enhance incident response by automating repetitive tasks, standardizing containment procedures, and reducing reaction times during a ransomware attack. There are several standalone SOAR solutions as well as several SIEM platforms that incorporate a SOAR capability.

* Automated Isolation
  + Integrate SOAR tools with network and endpoint security systems to automatically isolate devices as soon as ransomware indicators are detected.
  + Use automated scripts to disconnect infected systems from the network and apply quarantine measures without delay.
* Process Automation
  + Automate the disabling of network shares, termination of malicious processes, and update of firewall rules to block external communications—all triggered by predefined indicators of compromise (IOCs).
  + Configure SOAR playbooks that guide the response team through a consistent containment workflow, ensuring that no critical step is missed.
* Real-Time Alerts and Reporting
  + Leverage SOAR platforms to provide real-time alerts to incident responders and generate detailed reports on containment actions.
  + Use these insights to continuously refine automated playbooks and improve overall response effectiveness.

Effective incident containment and isolation are crucial to minimizing the damage caused by a ransomware attack. By rapidly isolating infected devices, disabling network shares and halting malicious processes, blocking external communications to C2 servers, and leveraging automation through SOAR tools, organizations can significantly reduce the impact of ransomware. These techniques not only help prevent further spread but also ensure that incident response efforts are efficient and well-coordinated, paving the way for a successful eradication and recovery process.

Malware Identification

Understanding the specific strain of ransomware affecting a system is crucial for an effective response and recovery. Different ransomware variants exhibit unique behaviors, encryption methods, and attack vectors. Identifying the strain enables security teams to determine the best course of action, whether it’s finding a known decryption tool, understanding the attack’s origin, or preventing further damage.

**Identifying the Ransomware Strain**

**Check the Ransom Note**

Most ransomware leaves a ransom note in each folder where data was encrypted. The ransom note typically includes a introduction and instructions for payment. The wording, structure, and contact details often indicate the strain. A few examples include:

* Ryuk: Uses plain-text ransom notes with an email contact.
* REvil: Uses a structured ransom note with a Tor link for negotiations.
* DarkSide: Provides a unique victim ID and an extortion message.

**Analyze Encrypted File Extensions**

Many ransomware families append unique extensions to encrypted files (e.g., your ContactList.xls is not named ContactList.xls.locky and the icon changed). Checking these extensions can help identify the variant. File extension examples include:

* **.locky** extension is created by Locky ransomware
* **.zepto** belongs to the Zepto ransomware (variant of Locky)
* **.encrypted** is a Generic extension, but often linked to specific strains like Crypt0L0cker

**Use Ransomware Identification Tools**

Several online tools allow users to upload ransom notes or encrypted files to determine the ransomware strain. Popular ransomware identification tools include:

* No More Ransom ID Ransomware (nomoreransom.org)
* MalwareHunterTeam’s ID Ransomware  
  (id-ransomware.malwarehunterteam.com)
* VirusTotal (virustotal.com for analyzing ransomware executables)

**Check System Logs and Network Traffic**

Investigating logs can reveal clues about how the ransomware executed, including the processes involved and network connections made. Key areas to investigate:

* Windows Event Logs (to detect suspicious processes)
* Firewall logs (to see C2 communication attempts)
* SIEM alerts (if endpoint detection had any alters or if it was disabled by the malware)

**Analyze the Ransomware Binary**

If possible, security teams can analyze the ransomware executable using sandbox environments or malware analysis tools. Tools for malware analysis include:

* Any.Run (interactive malware sandbox – any.run is the website)
* Hybrid Analysis (hybrid-analysis.com)
* Cuckoo Sandbox (cuckoo.cert.ee)

**Determining Recovery Options**

Some ransomware families have publicly available decryption tools, while others do not. Identifying the strain helps responders assess whether recovery without paying a ransom is possible. For example, ransomware like **STOP/Djvu** has partial decryption tools available, whereas **LockBit** does not.

* Search for a decryptor in online databases like No More Ransom Project (nomoreransom.org). If a decryptor exists, organizations can avoid ransom payments and expedite data recovery.
* Contact your insurance provider. Insurance companies often have ransomware negotiators that work with specific ransomware gangs and know the intimate details of their operations, tactics, and expectations.
* Contact law enforcement. The FBI and US Secret Service also have ransomware experts and investigators that have work numerous cases. Federal law enforcement can provide guidance, suggestions for additional resources, and in some cases the decryptor that another victim received that may work on your infected systems.

**Understanding the Attack Mechanism**

Different ransomware families use varied infection techniques, such as exploiting vulnerabilities, phishing emails, or Remote Desktop Protocol (RDP) attacks. Knowing the strain helps organizations focus on patching the critical security gaps to prevent reinfection.

* Conti and Ryuk ransomware gangs, for example, often use RDP for initial access and then use Emotet malware as a delivery mechanism for the ransomware payloads.
* In addition, ransomware gangs have their own tactics, techniques, and protocols (TTPs), and their particular ransomware’s behavioral signature or indicators of compromise (IOCs). Knowing these TTPs and IOCs will significantly help with the investigation, identification of infected systems, and the remediation of possible ingress points.
* Some ransomware strains attempt to move laterally across networks, while others focus only on local encryption. Identifying the type helps security teams contain the infection more effectively. WannaCry, for example. exploited the EternalBlue vulnerability to spread rapidly, while Maze ransomware exfiltrated data before encrypting files on the infected system.

Identifying the ransomware strain is a critical step in mitigating the impact of an attack. It provides insights into potential recovery methods, attack vectors, and ways to prevent further damage. By using ransom notes, file extensions, identification tools, and log analysis, organizations can gain a clearer picture of the attack and take informed actions toward remediation and future protection.

Fortifying the environment

We’ve already stated that this book starts from an “assume breach” position, and we do that for several reasons. First, there are many books that will attempt to explain how to prevent ransomware. So we didn’t feel another book on that topic would be helpful. In contrast, there didn’t appear to be another book like this one – a book specifically designed to help you respond to a ransomware attack.

Finally, and perhaps most importantly, this book starts from an assumed breach position because we believe the chances of you successfully defending yourself against every ransomware attack are virtually nil. Like terrorist attacks, you must be successful 100% of the time; the threat actors only have to be successful once. The nicest way to say this is that their chances of success are much greater than your chances of success. It reminds me of the poster from despair.com, reprinted with permission below. (I think the posters are hilarious and I highly recommend them to those with a twisted human like mine.)

A group of people running on a track

AI-generated content may be incorrect.

The chances of one threat successfully infecting your organization with ransomware are much greater than your organization successfully battling off every single attack 100% of the time. Since your odds of needing to respond to a ransomware attack are therefore very high, you should simply assume that you’re going to need to do so and prepare accordingly. In other words, this book begins from an assume breach position because you should begin from an assume breach position. It’s simply the most logical starting point when beginning a ransomware defense strategy.

It may not matter when you invent a time machine, but it certainly matters when you prepare your organization to respond to ransomware. I’m reminded of the old Dramamine commercials, which said “the time to take Dramamine is too late to take dramamine.” So before your organization is infected with ransomware, take the time to follow the suggestions of this chapter.

Engage with cyber professionals

One of the best things that you can do to prepare for a cyber attack is to engage with cyber professionals in advance of the attack. I’m reminded of stunt drivers saying, “I’m a professional; do not try this at home.”

**Find a blue team now**

When you are in the middle of an attack, there is no time to spare. This is not the time to start googling phrases like “cyber defense professionals,” or “ransomware blue team.” (In cyber defense, the defense team is called the *blue team*, and the *red team* is the one who you contract to attack you in order to test your defenses.) You should already have a relationship with someone who can help you defend against the attack. You should have a hotline that you can call to engage your team. Depending on the size of the company, you may even have mobile numbers to call.

Please understand that ransomware recovery is a complicated process fraught with peril. There are many steps in the process where you must decide which tack to take – and taking the wrong tack can place your organization on a path from which you cannot come back. For example, the use of certain tools and certain techniques might seem like a good idea to a novice, but they would be immediately detected by the threat actor. This might cause them to escalate the situation, making it worse and giving you less time to successfully defend against the attack.

It’s also easy in the midst of such an attack to take things personally and respond in a war of words with the threat actor. Perhaps you might want to curse them out in an email and suggest that their mother was a hamster and their father smells of elderberries. Perhaps you want to make emphatic statements like you’re never going to give them money, or something to that effect. It’s very easy and very human to become angry in the midst of a cyber attack and to want to say something that you wouldn’t be able to take back – and could drastically change the behavior of the threat actors.

A cyber professional does this all the time and is also not personally under attack. This allows them to be dispassionate in their responses and also gives them insight into how certain responses might be received from a threat actor. Understand that ransomware attacks are not usually done by individual threat actors; they are usually done by large ransomware organizations and there are only so many of them. So an experienced cyber professional will have likely dealt with most of them. This means that experienced professionals might know how specific ransomware groups might respond to different scenarios. This would include things like how a particular threat actor might respond to negotiating the ransom, or whether or not they’re likely to detect your forensic response.

They will know what you can say and what you cannot say, as well as what tools you can and cannot use to attempt to respond to the attack. Using cyber professionals gives you a much greater chance of successfully defending against the attack without having to pay the ransom.

**Get a red team, too**

Once you have a blue team that can help you defend against an actual cyber attack, it’s time to hire a red team to help reduce the chances of the threat actors. While it is our position that you cannot stop all attackers and attacks, hiring a red team is a great way to at least make it harder for the threat actors.

Red teams can conduct everything from basic automated port scans to physical penetration tests. They can start by making sure you’re not doing anything obvious to make yourself a target, and help you slowly harden your infrastructure. Make sure to have them specifically target the backup and DR infrastructure, because (as this chapter is going to say repeatedly) it is your last line of defense. It needs to be as solid as it can be.

**Find a cyber insurance carrier**

One of the other teams you’re going to want on your side is a cyber insurance provider. Perhaps it goes without saying, but insurance only works if you get it before an attack – sooner is better than later.

Consider the company Spectra, which (interestingly enough) is actually a robotic tape library provider in Colorado. A few years ago they suffered a ransomware attack and were quite pleased to realize that they had contracted a cyber insurance company only one month prior to the attack. The cyber insurance company honored the contract and was instrumental in their successful defense against the attack. Another interesting note about this story is that they, as we say, eat their own dogfood. Spectra is a tape company and its backups were on tape. They successfully defended themselves against a ransomware attack without paying the ransom – and they did it with tape.

Creating the relationship in advance gives you the advantage of time. The first thing that you have time to do properly is to search for and vet potential companies. Cyber insurance providers and cybersecurity professionals are not all created equal, and you need time to properly examine the differences between these different providers and select which one is most appropriate for your business.

You have time for a full request for proposal (RFP) process. Reach out to different vendors with your requirements and let them propose to you how they would help you. When developing your RFP, make sure to include stakeholders in your organization, such as the legal department and whatever resources you have in the cyber defense category.

If you are a reasonably-sized organization with a CISO, they will help you create the questions for your RFP. If you are a small organization without a CISO, you will need to be a little bit more creative. You might even have to hire a consultant to help you craft it. But again, doing this in advance gives you so much more time to do things properly. Waiting until the moment of disaster to start contacting cyber defense professionals just means you’re likely to get whoever has the best SEO – which may have little to do with whether or not they’re any good at cyber defense. And, of course, insurance companies will just laugh at you if you call them in the midst of an attack.

Once you have finished your RFP process and successfully vetted and selected both a cyber defense team (a.k.a. your blue team) and a cyber insurance provider, you will need to create contracts with them on what they will charge you for the different services that they will provide. This is another process that can take time. At a minimum it is probably going to take several days to review contracts and agree on pricing. That’s not a problem if you’re doing things in advance; it’s a disaster if you’re doing it in the middle of a cyber attack.

So if you do not already have a cyber insurance carrier and a blue team, it’s time to find them and hire them immediately. I would even suggest that you do this in two rounds. Assuming it doesn’t include a large upfront investment or a severely long-term contract, I’d do a very quick RFP for both types of companies and quickly hire the ones that feel right in the beginning. Then once you have them under contract, start a second RFP round to ensure you made the right decision. But at least you’ll have someone on your side if the worst happens while you’re trying to find the perfect team.

Identify forensic tools

Just like the cyber defense professionals mentioned in the previous section, you will want to find time to select these tools and become familiar with their operation. And just like with the cyber professionals, I wouldn’t take too long to pick your first tool. Pick one and make sure someone knows how it works before you spend weeks or months finding just the right tool.

**Forensic imaging tools**

Here’s a list of things you should consider when picking a forensic tool to use in the midst of a ransomware attack.

*Accuracy is number one*

If your imaging tool isn’t giving you a perfect bit-for-bit copy, you might as well not bother. Make sure it captures everything – deleted files, hidden data, all of it. When you’re defending against an attack (or presenting evidence of one), you need to be 100% certain your copy is identical to the original. No exceptions.

*Live capture*

It’s essential that your forensic imaging tool is capable of capturing an image of a live system. This allows you to capture an image of what’s in memory before you shut down an infected system. You can then examine what’s in that image passively without it being active on your network.

*Speed*

Your imaging tool needs to handle large drives and servers without breaking a sweat. I’ve seen cases crater because someone picked a tool that couldn’t handle the volume. Losing the fight with a threat actor because your tool is crawling is not a place you want to be.

*Rock-solid reliability*

Your imaging tool needs to work every single time, no matter what you throw at it. I don’t care if it’s an ancient Windows machine or the latest SSD. When you’re in the field, you can’t afford tools that throw tantrums over different operating systems or file formats. Reliability isn’t a feature it’s a requirement.

*Legal defensibility*

If you want to be able to assist law enforcement in prosecuting your threat actors, your imaging process needs to be bulletproof. That means proper audit trails, documented chain of custody, and adherence to standards like the Scientific Working Group on Digital Evidence (SWGDE). If you’re not worried about assisting law enforcement, this may be a lower priority for you.

**Log analysis tools**

When investigating a ransomware attack, the quality and capability of your log analysis platform can make the difference between understanding what happened and being left in the dark. The following is a set of criteria for evaluating these platforms as well.

*Data Ingestion and Normalization*

The foundation of any effective log analysis platform is its ability to ingest and normalize data from diverse sources. Your environment likely generates logs from numerous systems everything from network devices and firewalls to server logs, security tools, authentication systems, and cloud services. The platform must be able to ingest these varying formats and normalize them into a consistent, searchable format. It also needs to maintain the original log data while creating these normalized versions because sometimes normalized data obscures crucial details present in the original logs, leading investigators down incorrect paths. Your platform should give you the best of both worlds normalized data for efficient searching and the original logs for verification.

*Search Capabilities and Performance*

When investigating an incident, time is critical, and your analysis platform must provide robust search capabilities that operate at scale. Beyond basic full-text search, you need support for regular expressions, field-based filtering, and complex Boolean logic. The platform should allow you to save searches and create templates for common investigation patterns. Also be sure to test search performance at scale. You don’t want a large investigation to cripple your search.

*Timeline Analysis and Correlation*

Understanding the sequence of events during an attack is perhaps the most crucial aspect of incident investigation. Your platform should provide clear visual timeline representations while correlating events across multiple sources. Don’t forget time zone normalization, which becomes critical when dealing with geographically distributed systems or cloud services. The system should also help identify gaps in logging data, as these could indicate either system issues or deliberate log deletion by attackers.

*Event Classification and Automated Analysis*

While human analysis remains crucial, modern platforms should provide both manual and automated analysis capabilities. The ability to classify and tag events during analysis helps track investigation progress and facilitates collaboration among team members. But remember that automated analysis features like anomaly detection and pattern recognition should augment human analysis not replace it. Relying entirely on automated analysis can cause you to miss crucial evidence.

*Integration and Scalability*

Your log analysis platform shouldn’t exist in isolation. It should integrate seamlessly with your other security tools, including threat intelligence platforms, SIEM systems, and case management systems. These integrations should be bidirectional, allowing both import and export of relevant data. Scalability becomes increasingly important as log volumes grow. The platform must handle increasing data volumes while maintaining performance under load. Support for multiple concurrent analysts is essential during large-scale investigations.

*Cost Considerations and Evaluation*

Remember to consider both initial and ongoing costs. This includes licensing, storage requirements, training needs, and support costs. But also understand that cost may sometimes take a back seat to functionality and performance. Don’t attempt to save money by using inadequate tools. Also be sure test everything thoroughly with your actual log volumes and verify performance with historical data – before signing that purchase order. Perhaps most importantly, verify support responsiveness, remembering that they will be at their best during testing. It only goes downhill from there.

Before you spend any money on forensic tools, be sure to check out the often free forensic tools available at these websites.

[*https://www.magnetforensics.com/free-tools/*](https://www.magnetforensics.com/free-tools/)

[*https://www.sans.org/tools/*](https://www.sans.org/tools/)

[*https://www.exterro.com/ftk-product-downloads*](https://www.exterro.com/ftk-product-downloads)

**Learn your tools**

Once you select your forensic tools, it’s time to try them out. You want to be very familiar with how a tool behaves well before you need it. Start with some non-production systems and see how long it takes to image them and see how easily you can analyze those images. Try to analyze a large set of logs to see what you can find out. This is another area where a red team can be useful. They can conduct a plan attack against your environment, which you should be able to detect using your forensic tools. The more you use these tools in simulated attacks, the better you will be at using them during an actual attack. Just like everything else in this chapter, waiting until you need a tool to figure out how to use the tool is definitely not the way to go. Having said that, I still maintain that the best thing to do is to develop a relationship with the blue team and bring them in when an actual attack occurs; however, there’s nothing wrong with you learning how to use the tools as well.

Prepare your backup system

Just as this book assumes that your environment will eventually be compromised by ransomware, you also have to assume that your backup system will be compromised as well. You need to assume that a threat actor will gain the username and password for the most powerful accounts in your backup system. Therefore, you need to do what you can to stop them from using that login information or at least minimize the damage that they would be able to do with that account. As you will read in the coming pages, the biggest thing you need to be able to prevent is the deletion of the backups that you will need to respond to the attack. The good news is that stopping that particular attack is actually relatively easy.

I like to tell people that when it comes to cyber security and ransomware attacks, you should consider your backup system as Helm’s Deep. For those unfamiliar with the amazing trilogy *The Lord of The Rings* by JRR Tolkien, Helm’s Deep is a fortress valley, with a huge wall blocking its entrance and a castle at one end. When the kingdom of Rohan comes under attack by the evil wizard Saruman’s massive army, King Théoden leads his people there for a desperate last stand.

The defenders, outnumbered five to one, hold off the enemy until Saruman’s forces breach the wall using explosives. Just when defeat seems certain at dawn, the wizard Gandalf arrives with reinforcements, and a mysterious forest of walking trees appears to trap and destroy the enemy army.

For those who have seen *The Fellowship of The Ring* movie, I want you to think of yourself as Gandalf standing in front of the Balrog saying “You shall not pass!”



Every element of the backup system should be designed with security in mind. The following are a list of non-negotiable security design elements for your backup system.

Some of the advice given here is very similar to the previous chapter, but the purpose is different, and the urgency is higher. Things that may be optional in other areas of your environment (e.g. MFA) are table stakes here in Helm’s Deep.

Remember that when we are talking about the backup system and its security, we are talking about the main backup server, any media servers, any disk arrays or purpose-built backup appliances (e.g. dedupe arrays), any cloud accounts where backups may be running (e.g. AWS EC2), any cloud storage systems where backups are stored (e.g. AWS S3) and any tape libraries. We’re also talking about the physical security where any hardware or media is found. If we’re talking about a SaaS application, the good news is that most of this is taken care of for you, and all you have to worry about is the security of the app itself; however, some apps allow you to use other cloud providers’ storage. If so, you must include that.

**Taking backups out of the equation**

The first thing a threat actor is going to want to do is delete or encrypt your backups, in order to remove them as an option for you. That forces you to rely on them as your only way of recovering your data, which means you’ll be paying the ransom. Since encrypting your backups can take some time and deletion can be done near instantaneously, what they are most likely to do is attempt to delete your backups.

There are two ways a threat actor can delete your backups. The first method would be to target the backup data itself by directly accessing and deleting files on disk, in the cloud, or on backup tapes. This means that you need to be sure to properly secure all of the ways that you can access backup data outside of the actual backup software or as secure as the backup system.

The second method a threat actor might use to delete your backups is to get your backup system to do it for them. If they can become a privileged user in the backup system, they can tell it the backups in question have expired. Most backup systems will then immediately delete any disk based copies of expired backups.

If you have backups on tape, it will be a bit harder for them to delete them, as backup data on tape must be manually erased before it will disappear. Even if they delete the tape out of the backup software database, the data on the tape remains and can be re-ingested in a crisis. Depending on how proficient they are with your particular backup application, they may try to actually delete the data on the tapes as well.

Depending on the backup software in question, deleting backups on tape can be a one or two-step process. Some backup software packages will allow them to simply tell the tape library to re-label the tapes it assumes that if you ask it to re-label it then you want the backups on it delete. That would be the one step process.

Other packages will not allow you to re-label the tape if it knows it has backup data on it. In that case, the first step would be the same as with disk – tell it the backups on a given tape are expired. Then you need to tell it to put a new electronic label on the front of the tape. If the tapes are in a tape library and accessible to the backup system, that will happen almost as quickly as deleting backups on disk. Once a tape has a new electronic label at the beginning of the tape, it places an end of data (EOD) mark directly after the label, and all of the data on the rest of the tape is completely inaccessible – effectively deleting all the data on that tape in a few seconds. This of course means that one rather effective way of preventing this is to simply not keep all of your tapes online and accessible via the tape library.

**Friendly Fire**

Many years ago, I saw firsthand how quickly all of the tapes in a tape library can be rendered useless. It was at a Legato NetWorker client many years ago. I was one of two specialists on site, and the other one was a Legato employee – and we were using a new version of NetWorker. Prior to this new version, relabeling a large number of tapes took a long time, since it only used one drive at a time. A large relabeling operation also had to be babysat, because you had to say yes to the “are you sure?” message for each tape.

But this new version of NetWorker had a new feature called “fast and silent” that allowed you to use all of the tape drives in the tape library to do a relabeling operation, and without being bugged by the “are you sure” notifications. It was indeed a very nice feature if you had to re-label a large number of tapes.

Unfortunately, however, this feature came out at a time where there also was a bug where if you double-clicked on one tape in the library, it would actually select all of the tapes in the library.

This consultant needed to relabel one tape – so he double-clicked it. Because of this bug, he actually selected all tapes in the tape library. He also clicked the fast and silent option because he didn’t want to deal with the relabeling notification for that tape. Yes, in hindsight, it seems silly to click that button for just one tape, but that’s what he did. Clearly he was distracted, and he did not realize what he was doing at the moment he did it. But what he did was – in just a few clicks – relabel (i.e. erase) every tape in the tape library of a production customer.

A few minutes later, he realized what he had done, but the damage had already been done. To save face for Legato, they blamed me – since I was a contractor and not a Legato employee. (I was also nowhere near the keyboard, so there’s that. I fell on my proverbial sword and quietly exited the client.)

The point of this story is not about how I was “done dirty” by the other consultant. It’s to see just how easy it is to delete all backups in a tape library. It’s even easier if the backups are on disk.

All of these attempts to delete your backup data require privileged access. You need root or Administrator in order to delete backup files on disk – or at least I would hope so. You also need root or Administrator to do the kinds of commands that would re-label tapes outside of the backup system. And of course you’re going to need privileged access to the backup system in order to convince it that the backups should be expired early and deleted, or that backup tapes should be relabeled.

This risk is why we must do our best to prevent unauthorized privileged access to the backup system and its software, and to limit the damage that a particular user can do if their account is breached.

**Role-based administration**

Before we start talking about locking down various privileged accounts, it’s important to discuss the concept of role-based administration. It is a natural outcome of implementing the concept of least privilege discussed in chapter 4. The idea is to create roles that have various levels of power within the backup system and to assign those roles to different people that perform various functions.

It’s hard to fathom at this point, but at one point in the backup world we didn’t care at all about cyber security. Anyone who needed to do anything in the backup system was essentially all-powerful. Backups needed to run as root or administrator, so the backup software had to run as root or administrator. If a separate user ID was required to operate the backup system, that user id was also all-powerful. It could create backup configurations, schedule backups, perform restores, expire backups and relabel tapes – virtually anything was possible if you had administrator privilege to the backup system. I used to joke about how you should be friends with your back up administrator because they had root everywhere not only did they have the power to delete the primary copies of all of your data, they also had the power to wipe out the backup and recovery system. (It reminds me a little bit of how when I went in the military I remember my dad telling me that I should be nice to dispursint because they signed my paychecks.)

We have learned over time this is a bad idea. To help you understand why it’s such a bad idea, here’s a quick list of some of the operations a typical backup person might perform, and how they could be used by a threat actor for nefarious purposes.

*Create backup configurations*

A threat actor could create a custom backup configuration that backs up all of the sensitive data they are looking for, such as password files, important spreadsheets with configuration information. No one would look twice at new backup configurations. Also, if they have the ability to create backup configurations, that also means that they can change existing backup configurations, such as reducing the retention period of certain backups, which could have the effect of deleting existing backups made with that policy.

*Setting retention policies*

It’s not quite a Backup system thing, but it is related. Think about the data protection capabilities in.SaaS products like Microsoft 365. I complain a lot about how you need to back up such things, but many people rely on the retention capabilities of these products as a method of backup. It’s not back up as far as I’m concerned, but stick with me. One of the things a threat actor might do is reduce your number of versions down to one, which again has the net effect of deleting all previous versions of files. Once again, this is probably not something that anyone is watching for either.

*Manually expiring backups*

This is easily the most dangerous operation. If a threat actor can perform this function, they can use it to expire all backups in your environment in a matter of seconds. And if those backups are on disk, they will be deleted as well.

*Perform restores*

This task could be used to exfiltrate data. Even if they were unable to create custom backup configurations, they can easily restore data to a location that they could control – even a location outside of your firewall. This is because most modern backup software has the ability to restore any file to anywhere. Since most environments don’t monitor restores, a threat actor could probably do this without anyone noticing.

*Expire backups and relabel tapes*

This was already mentioned in the previous section. This could be used to delete every backup you have in a matter of minutes. Even if you replicated those backups to another location, the backup software will delete those backups as well.

Scheduling backups that have already been configured and monitoring those backups are relatively harmless functions. Even manually rerunning already-configured backups is relatively harmless. This means that a backup operator could be allowed to do their job without needing access to the more dangerous functions of creating new backups, performing restores, and expiring backups prematurely.

This means you could create a backup operator role that only has access to the safer operations, and a backup administrator role with access to the riskier operations. The backup administrator role could be locked up behind tighter security and receive much tighter scrutiny when it is being used.

**Secure your logins**

The first and foremost method of stopping a threat actor is login security. If they cannot login to the backup system or its systems, they will be unable to take it out of the ransomware equation. Therefore, if your backup server, backup storage, or backup software allows you to login to a privileged account simply by entering a username and password … that needs to be addressed immediately. This means that all a threat actor needs to compromise your back up environment is to steal one set of credentials. Sadly, there are innumerable ways that can be done; therefore, we must discuss ways to make that login more secure.

**Passwords and password managers**

The authors of this book believe strongly in the value of password managers. A password manager lets you generate unique, random, complex passwords for every site you use, without having to remember them all. It can also warn you about reused passwords, weak passwords, and compromised accounts. They are not perfect, but the benefits of password managers far outweigh the concerns.

I can’t tell you how many times someone has told me they don’t need a password manager because they have a “system.” Maybe they take their dog’s name and add some numbers, or they use the same base password and just change one character for each site. I hate to break it to you, but that’s not a system – that’s a disaster waiting to happen.

I get all the objections. “I don’t trust putting all my passwords in one place.” “What if the password manager gets hacked?” “I don’t want to rely on a company that might go out of business.” These are all valid concerns, but here’s the reality: using a password manager is still far, far safer than not using one. Let me explain why.

First, let’s address the elephant in the room. You’re already storing all your passwords in one place – your brain. And let’s be honest, your brain is terrible at generating and remembering truly random, unique passwords for every site you use. That’s why you’re probably reusing passwords, maybe with slight variations. The problem is that when one site gets breached (and they do, all the time), attackers immediately try those same passwords on other sites. I’ve seen entire organizations compromised because someone reused their LinkedIn password for their corporate account.

“But what if I forget my master password?” Well, what if you forget all your other passwords? At least with a password manager, you only have to remember one really good password instead of dozens of mediocre ones. And yes, you should store your master password somewhere secure – like in a physical safe or with other important documents. Just don’t stick it on a Post-it note on your monitor.

Password managers let you do two things really well: create a unique password for each login, and create really long hard-to-guess passwords. You should obviously be doing that first part, but it’s quite hard to do without a password manager. But I want to talk a little bit about the second thing: creating really long passwords. Did you know that a 13-word password can be guessed in months by a good password cracker? Check out the chart in [Figure 6-1](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch06.html#learningransomware_ch06_figure_3_1752802563950010), included here with permission from bitwarden.

A screen shot of a test chart

AI-generated content may be incorrect.

Figure 6-1. Password strength test chart. Courtsey of bitwarden

At this point, you should be creating passwords at least 16 characters in length, using a series of random characters, numbers, and mixed-cased letters. This is extremely difficult to do without a password manager, but with one it’s a breeze. [Figure 6-2](https://learning.oreilly.com/library/view/learning-ransomware-response/9781098169572/ch06.html#learningransomware_ch06_figure_4_1752802563950029) shows a screen capture from my password manager, Dashlane, that shows me creating a 40-character password with all kinds of complexity, including “similar looking characters,” which makes it near impossible to read and copy. Use your password manager to create the longest passwords your system will accept, and you’ll never have to worry about some quantum computing system guessing your password.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 6-2. Screen capture from password manager Dashlane

All password managers are not created equal. Let me break down your options.

*Fully Online Password Managers*

These are services like LastPass, 1Password, Dashlane, or Bitwarden’s cloud service. They store your encrypted password vault in the cloud, making it accessible from anywhere. The key thing to understand is that these services never see your actual passwords – everything is encrypted on your device before it gets sent to their servers. Even if they get hacked (which has happened), the attackers just get encrypted gibberish without your master password. LastPass had a breach in 2022, and you know what? If their users had strong master passwords, their vaults were still safe. That’s because the encryption happens on your device, not their servers. It’s like having a locked safe – even if someone steals the safe, they still can’t get in without the combination.

*Downloadable Software Password Managers*

These are applications like KeePass that store your password vault locally on your computer. The advantage here is that you’re not relying on a cloud service – you control where your vault lives. The downside is that you’re responsible for backing it up and syncing it between devices (although there are ways to do this using your own cloud storage). Think of it like keeping your money in a safe at home instead of at a bank. You have complete control, but you’d better have a good backup plan if your house burns down.

*Open Source Password Managers*

Then there’s the “trust but verify” crowd who want to see exactly how their password manager works. Enter open source options like KeePassXC (a KeePass variant) or Bitwarden’s open source version. These let you examine the code yourself or rely on the community to vet it for security issues. I particularly like Bitwarden because it gives you options. You can use their cloud service for convenience, or you can self-host the entire system if you want complete control. It’s like having the recipe for Coca-Cola and being able to make it yourself instead of relying on the company.

And if you’re using MFA (which you absolutely should be – we had that talk already, remember?), many password managers can store your MFA tokens too. Just make sure you keep separate backups of those MFA seeds – you don’t want to lose access to everything if something goes wrong.

I’m not going to tell you which password manager to use. That’s a personal choice based on your specific needs and threat model. But I will tell you this: not using a password manager in 2025 is like not using a seat belt because you’re worried about getting trapped in your car. Could it happen? Sure. But you’re a lot more likely to get hurt by not using one.

Still not convinced? Let me put it this way: I’ve been in IT security for decades, and I’ve never seen anyone get hacked because they used a password manager. But I’ve seen plenty of people get hacked because they didn’t use one.

**Multi factor authentication**

Let’s start with what most of you are probably already familiar with: multi-factor authentication, or MFA. Some folks call it 2FA (two-factor authentication), but it’s the same idea. The core concept here is that you’re using more than one way to prove you are who you say you are.

Think about it this way. When you go to your bank‘s website, and try to get a new debit card, they don’t just ask for your password, right? They want to see your driver’s license, and they might even ask you questions about your account history. That means they’re using multiple *factors* to verify your identity. MFA works the same way in the digital world.

With MFA, you start with the traditional username and password combo (something you know), but then you need something else. Usually, the other factor is either something you have (like your phone) or something you are (like your fingerprint). A common setup these days is getting a code sent to your phone via text message or using an authenticator app that generates *one time passwords*, which is why they are called *OTP* apps. Two popular free OTP apps are Google Authenticator and Authy.

There are also physical devices that can display one time passwords as well, but the proliferation of smartphones, and the fact that most of us have it with us at all times is why app-based MFA has become so popular. It reminds me of the question people ask photographers, “What is the best camera?” Their answer is always, “the one you have with you.” We want to make authentication hard for the threat actor, not for us. While physical MFA devices may be more secure, we don’t always have them with us.

Also common is MFA built into a complete IAM (identity and access management) system like Okta. The server side runs in the cloud, and it is paired with its own app on a smartphone. Once you login, the app notifies you that someone is trying to access your login, and you simply have to approve that request.

The biggest challenge with MFA is an *MFA fatigue* attack, where a threat actor bombards the user with so many MFA requests that they eventually approve one of them. There are several well-known attacks that were believed to have been accomplished via this method. The good news is that their are no known exploits for the next topic: passkeys.

**Passkeys and FIDO**

Now, let’s talk about *FIDO* (Fast Identity Online) and passkeys. FIDO is the overall framework, and passkeys is simply the most common FIDO-compliant system for IAM. Passkeys accomplish the same thing as passwords and MFA – proving you are who you say you are using multiple factors – but they do it all in one smooth step instead of the back-and-forth dance we do with MFA.

When you use passkeys, you’re combining something you have (the security key) with something you are (e.g. fingerprint or face scan) or something you know (e.g. PIN). No codes to type in, no waiting for text messages, no opening another app. You simply authenticate with the passkey system and it logs you into the application.

The really clever bit about FIDO is that it’s not just more convenient – it’s actually more secure than traditional MFA. That’s because FIDO uses some pretty sophisticated cryptography to make sure that even if someone somehow intercepts your login attempt, they can’t use that information to pretend to be you somewhere else. It’s like having a key that changes shape every time you use it, and only works with the specific lock you’re trying to open.

FIDO is the real deal. It’s backed by all the major tech companies (Google, Microsoft, Apple, etc.), and it’s becoming more common every day. If your organization isn’t using it yet, they probably will be soon.

**Downsides to FIDO and Passkeys**

As of this writing, there are no known exploits of FIDO-compliant systems. Zero. Nada. Zilch. While traditional username/password systems get compromised daily, and even MFA systems occasionally get phished or hijacked through various social engineering tricks, FIDO has remained solid. This isn’t just luck – it’s by design. The whole system was built from the ground up to prevent the types of attacks that plague other authentication methods. When people say something is “unhackable,” I usually roll my eyes, but FIDO is about as close as we’ve gotten so far.

One valid concern is that if you lose access to your passkey device – you’re locked out! In other words, some see it as *too* secure! Be sure to take this into account when picking a passkey system and designing your implementation.

One show-stopper is that you can’t just decide to use FIDO and make it happen by sheer force of will. The system you’re logging into has to support it. If you’re trying to log into a legacy application that only supports username and password, that’s what you’re stuck with. Same goes for MFA – if the system only supports basic password authentication, that’s all you get. So while FIDO might be the gold standard, you should use the most secure method that your system supports. A password plus MFA is way better than just a password, even if it’s not as slick as FIDO. Remember, security isn’t about being perfect – it’s about being as secure as you reasonably can be with the tools you have available.

**Login security summary**

Securing access to your backup infrastructure requires a comprehensive approach across your entire backup ecosystem - the applications, servers, and storage systems. Your organization’s entire recovery capability hinges on these accounts remaining secure, so you must treat these credentials with the same gravity as nuclear launch codes - and I’m not exaggerating here.

Each administrator must have their own unique account for proper accountability and auditing. I don’t care how much easier it might seem to share accounts among team members; I’ve seen that convenience turn into catastrophe more times than I can count. The moment you create a shared account, you’ve lost the ability to track who did what in your backup system, and that’s a risk you simply cannot afford to take.

Implement the strongest authentication methods available for each component of your backup system. Where passkeys are supported, use them. Where they aren’t, ensure you’re using properly managed complex passwords with robust MFA. The key is consistency - apply these security measures across every element of your backup infrastructure, from the management console to the storage systems themselves.

I would look into Okta and its competitors. Okta is a platform that uses the latest in security technology to allow you to login to multiple things via a single sign-on to their platform. They support (and require) MFA, and also support passkeys. This would allow you to take the security of your backup system to the next level, even if parts of your system don’t support passkeys yet – Okta will manage it. I’m mainly discussing Okta here due to its popularity, and the fact that most cloud-based backup products support using it to login to their systems. Okta is likely to support securely logging into just about anything you would need as part of your backup infrastructure. The biggest challenge I have with Okta, however, is that a standalone implementation of it just for your backup system might be very costly. If you’re using it elsewhere in your environment, perhaps you can negotiate something with Okta to allow you to have one billing account, but two separate cloud administration accounts. As you’ll see later in this chapter, you must segregate your backup system from the rest of your security infrastructure.

Another product I would look into is Yubikey, which are physical hardware keys that support MFA and FIDO-compliant passkeys using biometrics. The thing I like best about Yubikey is the converse of my concern with Okta; Yubikey scales down very well, as you can have a complete system for around $100 of hardware. I can’t go into details here, because their choices are very broad, and it highly depends on your infrastructure.

If any component of your backup system - whether application, server, or storage - doesn’t support MFA and/or passkeys, seriously consider replacing it. The integrity of your backup system is your last line of defense. When everything else fails, these security measures could be all that stands between you and complete data loss.

**Update your software**

This is mainly true only if you are running your own backup system; if you’re using a SaaS vendor they will handle server upgrades. This should be obvious but often isn’t: keep your backup infrastructure updated. I can’t tell you how many times I’ve walked into a data center and found backup servers running operating systems and application versions that went end-of-life years ago. Or perhaps they are running current versions, but their patches are not up to date. The excuse is always the same: “We don’t want to break anything that’s working.”

Your backup server is one of the most privileged systems in your environment. It needs access to pretty much everything to do its job. Running it on an out-of-date system is like giving the keys to your house to someone you know is a burglar. Sure, they might not rob you today, but it’s probably not going to end well.

What you need to do:

* Keep your backup server’s operating system patched and current
* Update your backup software regularly (even if the current version “works fine”)
* Update any agents or components running on protected systems (this applies even if you’re using a SaaS-based product)
* Keep your backup target systems (like deduplication appliances) current
* Monitor security advisories (e.g. [*https://cve.mitre.org/*](https://cve.mitre.org/)) for your backup infrastructure

“But Curtis,” you say, “what if an update breaks something?” That’s why you test updates in a lab environment first. Yes, it’s more work. Yes, it takes more time. But it’s a lot less work than trying to explain to your CEO why all the backups are gone because you got hit with a vulnerability that was patched six months ago.

**Segregate all backup infrastructure**

Your backup system should share nothing in common with your production system, except (perhaps) the network. Your backup system should share no security infrastructure in common with your production systems. It should be on separate servers, inside a separate virtual datacenter (if using virtualization), on separate storage, etc.

*Most importantly*, it must not share your corporate network’s IAM system. It doesn’t matter if it’s OKTA, Entra ID, Active Directory, or LDAP. If that system is compromised, it might also compromise your backup system. Use a completely separate IAM system. I’d prefer it to be a different technology altogether, such as using Okta when the rest of your company uses Entra ID, or vice versa. But I understand why you might want to use the same technology as the rest of the company; just make sure it is a completely separate administrative account with no crossover to the production side. There should be no “super user” account that controls both environments.

**Shut off RDP**

The Remote Desktop Protocol (RDP) is one of the most common attack vectors used by ransomware groups. Ransomware attacks via RDP are so common that some of us say RDP stands for *Ransomware Distribution Protocol*. If you absolutely must have RDP enabled on your backup servers (and you probably don’t), consider securing it using these methods.

*Restrict it to specific IP ranges*

First, restrict access to specific IP ranges inside your LAN. (Obviously, RDP must never be allowed externally.) I mean completely locked down - no connections allowed unless they’re coming from known, trusted addresses.

*Require VPN access first*

Even inside your LAN, consider a separate firewall that requires VPN access before allowing any RDP connection. This means potential attackers can’t even see your RDP port unless they’ve already compromised that VPN.

*Enable Network Level Authentication*

Network Level Authentication forces users to authenticate before establishing an RDP session, not after. This might sound like a small distinction, but it prevents attackers from exploiting vulnerabilities in the RDP protocol itself.

*Use non-standard ports*

You should also consider moving RDP to a non-standard port. Yes, this is security through obscurity, and no, it won’t stop a determined attacker. But it will keep your servers from showing up in automated port scans, which is how many attacks begin. Think of it as locking your back door even though someone could still break a window - it’s not perfect, but it’s better than leaving the door wide open.

*Disable RDP entirely*

The best solution is to simply disable RDP entirely and use your backup software’s built-in console access. Modern backup solutions have spent considerable time and effort building secure remote access capabilities. Use them. They’re almost always more secure than RDP, and they’re certainly more appropriate for backup system management.

*Lock down SMB*

Server Message Block (SMB) security is another critical area we need to discuss, particularly when it comes to backup servers. I’ve seen too many organizations treating their backup servers like any other file server, and that’s a recipe for disaster.Just like RDP, SMB is also used to deploy ransomware and to interrogate your network for vulnerable information. Lock down SMB access to only what’s absolutely necessary.

*Shut down SMBv1*

First and foremost, if you’re still running SMBv1, stop reading right now and go disable it. I mean it. This version of the protocol has been obsolete for years and was the primary attack vector for WannaCry and several other devastating ransomware variants. There is absolutely no excuse for running SMBv1 in a modern environment. If you have legacy applications that require it, those applications need to be upgraded or replaced - *immediately*.

*Restrict SMB to certain IPs*

Just like with RDP, you need to restrict SMB access to specific IP ranges. Your backup server has no business accepting SMB connections from arbitrary network segments. Lock it down to only the specific systems and networks that absolutely require this access. There are many documented incidents where unrestricted SMB access gave attackers free rein to explore - and eventually encrypt - entire backup repositories.

*Use modern authentication methods*

Modern authentication methods are crucial here. This means using at least SMBv3 with encryption enabled, and implementing strong authentication protocols. The days of NTLM and basic password authentication should be long behind us. If your backup software supports it, consider using its native protocols instead of SMB.

*Implement monitoring*

Finally, you must implement robust monitoring for unusual SMB access patterns. This means watching for things like excessive file access, unusual access times, or connections from unexpected sources. Modern ransomware often behaves differently than legitimate backup processes - it tends to access files much faster and in different patterns than normal backup operations. Having proper monitoring in place can alert you to an attack in progress before it’s too late.

**Watch Everything: Monitoring Your Backup Environment**

You know what’s worse than getting hit with ransomware? Finding out you got hit three months ago and nobody noticed. This is why monitoring your backup environment is crucial.

What you should be watching for:

*Unusual patterns in data transfer (especially outbound)*

Your backup system should have a fairly predictable pattern of data movement. You know roughly how much data you back up each day and where it’s supposed to go. When you see unexpected outbound data transfers, especially large ones, alarm bells should start ringing. Attackers can use backup systems as exfiltration points precisely because organizations often don’t monitor these patterns carefully enough.

*Excessive restore operations*

Sudden spikes in restore operations are a red flag. Your restore patterns should be predictable - either scheduled tests or legitimate restores that follow your documented processes. If you’re seeing numerous concurrent restores or restores of unusual scope, that could indicate someone trying to extract data through your backup system.

*Mass deletion attempts*

Any attempt to delete multiple backups simultaneously should trigger immediate alerts. Normal backup deletion follows retention policies and happens in a predictable pattern, and it’s usually not done manually. There are many documented incidents where attackers tried to delete entire backup sets before launching their main attack. Your monitoring system needs to catch these attempts before they succeed.

*Failed login attempts*

Failed authentication attempts often precede successful breaches. Pay special attention to failed logins happening outside normal maintenance windows or from unusual source addresses. I’ve seen attackers probe backup systems for weeks looking for vulnerabilities in authentication. Every failed login attempt tells a story - make sure you’re reading those stories.

*Configuration changes*

Your backup system’s configuration should be relatively stable. Any changes to backup policies, retention settings, or security configurations should be documented and approved. Unexpected configuration changes, especially those reducing retention periods or disabling security features, often precede attacks. Monitor these changes religiously and verify every single one.

*Access from unusual locations or times*

Backup systems typically operate on well-defined schedules. When you see access attempts outside normal maintenance windows or from unexpected locations, you should get concerned. This is especially true for administrative access - your backup administrators should have predictable work patterns. Any deviation from these patterns deserves investigation. Access during odd hours is often the first sign of compromise.

Set up alerts for anything that looks suspicious. Yes, you’ll probably get some false positives. But I’d rather investigate ten false alarms than miss one real attack.

**Encryption**

Let’s talk about encryption. If you’re not encrypting your backups, stop reading right now and go fix that. I’ll wait. Are you back? Good. Here’s what you need to know about backup encryption.

**At-Rest Encryption**

Your backup data needs to be encrypted anywhere it sits - whether that’s on disk, tape, or in the cloud. This is your first and most critical line of defense against unauthorized access. Start with AES-256 encryption at a minimum; anything less is simply inadequate for modern security requirements. Skimping on encryption strength to save processing overhead is a false economy if I’ve ever seen one.

Every single copy of your backups must be encrypted - no exceptions. This includes primary copies, replicas, and especially anything stored offsite. And here’s the crucial part: you need to regularly test recovery operations with your encrypted backups. I’ve seen too many organizations discover during a crisis that their encryption strategy had holes in it.

**In-Flight Encryption**

Data in motion is just as vulnerable as data at rest - maybe more so. Every piece of backup traffic moving across your network needs encryption. Implement TLS 1.2 or higher for all network communications - version 1.1 and below have known vulnerabilities that make them unsuitable for backup traffic. This applies to everything: backup agent communications, replication traffic, and especially any data moving between sites.

Ensure your backup server-to-agent communications are encrypted with modern protocols. I’ve investigated incidents where organizations encrypted their backup data but left the control channel vulnerable - that’s like putting a steel door on a cardboard frame.

**Key Management**

This is where I see organizations stumble most often. You can have the strongest encryption in the world, but it’s worthless if you manage your keys poorly. First rule: never, ever store encryption keys alongside the data they protect. I once worked with a company that stored their encryption keys in a text file on their backup server - don’t be that company.

Implement a secure process for key rotation. This means having clear procedures for generating new keys, securely distributing them, and maintaining access to older backups encrypted with previous keys. Keep secure backups of your encryption keys, stored separately from your backup environment. Document every aspect of your key management procedures thoroughly, and store that documentation securely - but not with either the keys or the backups.

Remember: encryption isn’t just about making your data unreadable to unauthorized users. It’s about ensuring your entire backup and recovery process remains secure and reliable. Test everything regularly, document meticulously, and never assume your encryption is working without verifying it.

**Stay out of user space**

This recommendation will come as a complete shock to many. Keep your disk-based backups out of user space. This means I should not be able to login to your backup server and see your backups, especially not in D:\backups! Keeping them where you can see them (in user space) is a bad idea for multiple reasons.

First, it makes your backups vulnerable to the same attacks that might hit your user systems. Second, it often means your backups are accessible to regular users, which violates the principle of least privilege. And third, it makes it really easy for ransomware to find and encrypt your backups.

If no one has ever told you this before, you might find yourself wondering how you’re supposed to do that. It’s usually not that hard.

*Use direct storage connections (like Veeam’s direct SAN access)*

One of the most secure ways to store your backups is to bypass the file system entirely. Modern backup solutions like Veeam offer direct storage access capabilities that connect straight to your storage infrastructure. This isn’t just about performance - it’s about security. When your backups never touch the file system, they’re essentially invisible to file-system based attacks. I’ve seen organizations survive ransomware attacks specifically because their backups were stored this way, completely out of reach of the malware that encrypted their production systems.

*Store backups in dedicated backup appliances*

Purpose-built backup appliances exist for a reason. These systems are designed from the ground up to securely store and manage backup data. They typically include their own storage, their own operating system, and their own security controls. These appliances often include features like built-in encryption, deduplication, and immutability - all crucial for secure backup storage. I’ve worked with organizations that tried to save money by using general-purpose storage for their backups, only to spend far more recovering from a security incident that a dedicated appliance would have prevented.

*Use object storage instead of file shares*

Object storage represents a fundamental shift in how we store backup data. Unlike traditional file systems, object storage systems provide inherent security features that make them ideal for backup storage. They’re not mountable as drives, they’re not browsable like file shares, and many offer built-in immutability features (more on immutability later). I’ve seen organizations completely transform their backup security postures simply by moving from traditional file shares to object storage. Plus, most modern object storage systems also scale better than enterprise file storage.

*Keep backups completely separate from user-accessible areas*

This is perhaps the most crucial principle of all. Your backups should exist in their own isolated environment, completely separate from your production systems and user access. I’ve investigated too many incidents where backups were compromised simply because they were stored in areas that users could access. Think about it - would you store your insurance policies in the same safe as the valuables they’re insuring? Of course not. Your backups deserve the same level of separation and protection.

Here’s what proper backup storage looks like in practice: separate networks, separate authentication, separate storage systems, and separate access controls. Yes, this makes things slightly more complicated to manage. Yes, it might cost a bit more. But I can tell you from experience that the cost of properly segregating your backups is nothing compared to the cost of losing them in an attack.

Remember, your backup storage isn’t just another place to put data - it’s Helm’s Deep. It’s your last line of defense against data loss. Every decision about where and how to store your backups should start with security considerations. If you’re still storing backups in file system locations or on general network shares, you need to rethink your approach. The threat landscape is too sophisticated, and the stakes are too high, to continue with these outdated practices.

**Immutable storage**

Complete immutability is the best way to ensure that your backups cannot get encrypted or deleted in a ransomware attack. *Immutable* simply means *cannot be changed*. In backup parlance, it means that once written, backups cannot be changed, encrypted or deleted by anyone – including the administrator. It can, of course, expire based on its retention period, after which it will automatically get deleted. But it will not and cannot be expired or deleted before that time.

If even you, the backup administrator, cannot manually expire and delete the backups, then neither can a threat actor. If there is a back door that allows you to delete backups in certain special circumstances, then that’s not truly immutable. Therefore, avoid the temptation to enable such features.

*Good: file systems that support immutability*

One way to implement immutable storage is to use an on-premises file system that has built-in immutability features at the file system level, such as Linux. This means the immutability is enforced by the file system itself, not just the application or storage layer. While this approach can be effective, remember that anyone with root access to the file system can typically disable these controls. That’s why you need additional security measures around these systems.

*Better: Purpose-built appliances with immutability features*

These systems, such as those provided by Oootbi, are designed specifically for backup storage and include built-in immutability capabilities. When you write data to these appliances, you can specify a retention period during which the data cannot be modified or deleted - by anyone. And I mean anyone - not even a system administrator with root access can alter this data until the retention period expires.

*Best: Immutable storage in the cloud*

On-premises immutable storage is a solid start - but I don’t think it’s enough. There’s still that risk of physical access, although I will admit that is a very remote possibility in a typical ransomware attack. Cloud immutable storage adds another crucial layer of protection, and it does so in a way that’s both elegant and cost-effective. Think of it as your insurance policy’s insurance policy.

Let’s talk about the major players. AWS led the charge with S3 Object Lock, and it’s still one of the most robust implementations available. Once you enable Object Lock on a bucket and set a retention period, that data is untouchable - even if someone compromises your AWS root credentials. Azure followed suit with their Blob Immutable Storage, offering similar capabilities but with their own twist on the implementation. Google Cloud’s Object Lock rounds out the big three, providing comparable features for those invested in the Google ecosystem.

What makes cloud immutable storage particularly compelling is its inherent separation from your primary site. I’ve investigated incidents where organizations lost both their primary and local backup storage because both were affected by the same disaster - whether that was a flood, fire, or particularly aggressive ransomware attack. Cloud immutable storage would have saved them.

Here’s another reality check: these cloud providers have security teams that are larger than most IT departments. They have physical security that would make Fort Knox envious. When you store your immutable backups in their facilities, you’re essentially hiring their security team to protect your data.

The scalability aspect is crucial too. I’ve watched organizations struggle to predict and provision enough local immutable storage, often either over-provisioning (and wasting money) or under-provisioning (and compromising their backup strategy). With cloud storage, you pay for what you use and can scale instantly when needed.

And let’s talk about costs. Yes, cloud storage has ongoing operational costs that can add up. But when you factor in the total cost of ownership for local immutable storage - hardware, maintenance, power, cooling, floor space, and the expertise to manage it all - cloud storage can come out ahead. I’ve helped organizations run these numbers, and they’re frequently surprised by how cost-effective immutable storage in the cloud can be.

Just remember: while immutable storage in the cloud is powerful, it’s not magic. You still need to properly configure it, monitor it, and most importantly, test your ability to recover from it regularly. I’ve seen too many organizations assume their cloud backups were working correctly, only to discover problems during a crisis.

**Creating a Disaster Recovery Plan**

Having secure backups is great, but you also need a plan for how to use them when everything goes sideways. This means having multiple DR options and knowing how to orchestrate them.

**Full Hot Site**

Let’s talk about the Cadillac of disaster recovery solutions - a *hot site*. This is exactly what it sounds like - a complete duplicate of your production environment, sitting there ready to take over at a moment’s notice. This isn’t just backup hardware gathering dust; it’s a fully operational duplicate of your environment that’s kept in sync with production continuously. Yes, it’s expensive. Yes, it’s complex to maintain. But for organizations that simply cannot tolerate downtime - think financial trading systems or critical healthcare applications - it’s the only real option. When you need to be back online in minutes instead of hours or days, this is how you do it.

*Duplicate hardware/infrastructure*

A hot site requires a complete mirror of your production environment. Every server, every switch, every storage array needs to be replicated. I’ve seen organizations try to cut corners here, only to discover during a disaster that some crucial piece of infrastructure wasn’t duplicated. You’re either fully redundant or you’re not - there’s no middle ground with a hot site.

*Regular replication*

Your hot site needs to stay synchronized with production, usually through continuous replication. This means substantial bandwidth requirements and sophisticated replication technologies. You’re typically looking at synchronous or near-synchronous replication with recovery point objectives measured in seconds or minutes.

*Automated failover capabilities*

Manual failover procedures are too slow and error-prone for a proper hot site. You need automated systems that can detect failures and initiate failover without human intervention. I’ve seen too many organizations with beautiful hot sites that still required hours of manual work to activate. That’s not a hot site - that’s a very expensive warm site.

*Regular testing procedures*

If you’re investing in a hot site, you’d better be testing it regularly - and I mean really testing it, not just checking that the lights are on. I recommend quarterly full failover tests at minimum. Yes, it’s disruptive. Yes, it’s expensive. But it’s nowhere near as expensive as discovering your hot site doesn’t work during an actual disaster.

*Clear activation criteria*

You need crystal clear, documented criteria for when to activate your hot site. I’ve investigated incidents where organizations delayed failing over to their hot site because no one was sure who could make the call. Every minute of indecision costs you money and potentially data.

**Cold Site Recovery**

A cold site is fundamentally different from a hot site, and it’s the more common approach. Instead of maintaining a duplicate environment, you’re essentially starting with an empty room and a set of backup copies. When disaster strikes, you’ll be building your environment from scratch using these backups. It’s much slower than a hot site, but it’s also *far* less expensive to maintain. This approach makes sense for organizations that can tolerate hours or days of downtime in exchange for lower ongoing costs.

*Where will you get hardware?*

This is the first question you need to answer with a cold site strategy. You need guaranteed access to hardware that matches your specifications, usually through vendor agreements. I’ve seen organizations assume they could just order hardware when needed - then discover during a disaster that their required equipment had a 12-week lead time.

*How will you handle networking?*

Your network infrastructure needs to be just as recoverable as your servers and data. This means documented IP schemes, routing configurations, and firewall rules.

*What’s your process for declaring a disaster?*

You need clear procedures for who can declare a disaster and under what circumstances. This needs to be documented and practiced.

*How will you handle DNS and other external services?*

External services like DNS need to be part of your recovery plan. I’ve seen perfectly executed server recoveries fail because no one remembered to update DNS records. Document every external dependency and how to handle it during recovery.

**Cloud Recovery**

Cloud recovery has revolutionized disaster recovery over the past decade, offering a middle ground between hot and cold sites. Instead of maintaining your own recovery facility, you’re leveraging the virtually unlimited resources of a cloud provider. The beauty here is flexibility - you can choose exactly how “hot” or “cold” you want your recovery to be, and you can adjust it on the fly based on your needs and budget. You’re essentially trading capital expenses for operational expenses, and gaining a level of scalability that’s simply impossible with traditional DR approaches.

*Understand cloud costs*

Cloud recovery can be cost-effective, but only if you understand the pricing model. Egress charges can be especially shocking if you’re not prepared. I’ve seen organizations get hit with massive bills during recovery testing because they didn’t understand the cost implications of moving large amounts of data out of the cloud.

*Test your recovery procedures regularly*

Cloud recovery procedures need regular testing just like any other DR strategy. The cloud adds its own complexity - API changes, service updates, security requirements. I recommend monthly testing of critical system recovery procedures at minimum.

*Document network requirements*

Cloud networking is different from on-premises networking. You need to document exactly how your recovered environment will connect to users and other systems. I’ve seen organizations successfully recover their servers to the cloud, only to realize they had no plan for how users would access them.

*Have clear procedures for accessing cloud resources*

Cloud access during a disaster needs to be carefully controlled. This means having documented procedures for accessing cloud consoles, APIs, and recovered systems. I’ve investigated incidents where organizations compromised security during recovery because they hadn’t planned their access procedures properly.

*Maintain security during recovery*

Security often takes a back seat during recovery operations - and that’s a massive mistake. Your recovered environment needs to be just as secure as your production environment. I’ve seen organizations recover from one disaster only to create another by standing up insecure systems in their rush to restore operations.

**Failback**

In DR planning, everyone focuses on failing over, but failing back is just as critical. You need a clear plan for how to return to normal operations once the disaster is over. I’ve worked with organizations that executed their failover perfectly but had no plan for getting back to normal operations.

Final Thoughts

Securing your backups isn’t a one-time thing; it’s an ongoing process. The threat landscape is constantly changing, and your security measures need to evolve with it. Don’t fall into the trap of thinking “it won’t happen to me” or “we’re too small to be a target.” I’ve seen organizations of all sizes get hit, and the ones that survive are the ones that took security seriously before they needed it.

I’ll say it again: Your backups are Helm’s Deep –  your last line of defense. If they’re not secure, nothing else matters. Treat them accordingly.

And if anyone tells you that securing backups is too much work or too expensive, ask them how expensive it would be to lose everything. I’ve never met anyone who regretted having too much security, but I’ve met plenty who regretted not having enough.

Stay safe out there, and keep those backups secure. Your future self will thank you.

Make plan preparing inevitable

Ransomware has come a long way, and not in a good way. What started as an annoyance targeting individual users became a full-blown criminal operation aimed at businesses, governments, and even critical infrastructure. In recent years, we’ve seen major ransomware attacks bring down supply chains, cripple healthcare systems, and cost organizations billions in ransom payments, recovery costs, and lost productivity. The growing sophistication and frequency of these attacks make one thing clear: you need a solid incident response plan (IRP) in place before your number comes up in the ransomware game.

A good ransomware response plan is like a playbook for disaster. It lays out exactly how your team will detect, contain, remove, and recover from an attack. Without one, you’re looking at longer downtime, bigger financial losses, and a greater chance of losing sensitive data, or having it leaked to the public. A well-executed plan can mean the difference between getting back on track quickly or being stuck in damage-control mode for weeks (or longer).

But you also can’t stop with just having a plan on paper. The real test is whether your team can execute it under pressure. That’s where tabletop exercises and war games come in. Running a simulation of a ransomware attack in a controlled environment helps you see how prepared your team really is. It’s the best way to spot gaps in your plan, sharpen decision-making, and make sure everyone knows their role when things get messy.

In this chapter, we’ll break down how to build a ransomware response plan that works, from setting clear goals and defining roles to developing containment and recovery strategies. We’ll also talk about how to design and run effective tabletop exercises and war games so you can test and improve your plan before a real attack happens. By the end, you’ll have a clear roadmap for handling ransomware with confidence.

Setting Objectives and Scope

When you’re putting together a ransomware incident response plan (IRP), getting your objectives and scope nailed down is the first priority. Without them, things can get confusing and your response might drag on, costs could pile up, and your reputation could take a hit. Let’s walk through how to set clear goals and boundaries so your team can handle an attack smoothly and successfully.

**Defining the Goals of the Incident Response Plan**

So, what’s the point of this plan? It’s all about giving your team a clear target to aim for when ransomware strikes. Without that focus, everyone’s just guessing, and that’s a recipe for trouble. Here are a few targets you might want to focus on:

* *Minimize Downtime*—Get your business back online quickly. No one wants operations grinding to a halt for longer than necessary.
* *Protect Key Data*—Keep sensitive information safe from compromise or theft. That’s a top priority.
* *Limit the Damage*—Reduce the financial and reputational fallout, including ransoms, legal headaches, or bad media coverage.
* *Stay Compliant*—Make sure you are meeting any applicable regulatory requirements, like SOX, PCI, GDPR, HIPAA, or CCPA to name a few. If you have compliance requirements, then having an IRP isn’t optional.
* *Preserve Evidence*—Hold onto logs or files that could help with investigations later, whether it’s for insurance, legal, law enforcement or internal reviews.
* *Get Stronger*—Learn from the experience to improve your processes, skills, and defenses for next time.

Something to keep in mind while you build out your IRP goals is to use SMART. SMART stands for Specific, Measurable, Achievable, Relevant, and Time-bound. It’s a straightforward way to keep things practical. For example:

* Restore critical systems within four hours of detecting the ransomware.
* Contain the spread within 30 minutes of the first alert.
* Notify affected stakeholders, like customers or regulators, within an hour of confirming the incident.

These activities set expectations and can be measured. Did you hit that four-hour mark? Were we able to stop it in 30 minutes? It keeps your team on track and gives you a way to gauge how you’re doing. Clear goals, expectations, and metrics will help everyone keep a pulse on the progress of the response.

**Figuring Out What the Incident Response Plan Covers**

Alright, so what is your IRP actually going to address? The scope is like setting the boundaries, including what we’re protecting as well as what’s off the table or out of scope. It must fit how your organization runs, the tech you’re using, and where the trouble spots might be. Some things to consider:

* What Types of Ransomware Are We Dealing With?
  + Ransomware is not just one thing; it has different flavors and attack types, and your plan needs to be ready for all of them. There is the old-school encryption kind that locks up your files and demands a payout to in order to unlock ‘them. Then there is Locker ransomware that just kicks you out, no fancy encryption required. Double extortion ransomware is yet another, and it encrypts your stuff while the attackers siphon your data out to the dark net *and* threaten to post it publicly if you don’t pay. Ransomware-as-a-Service (RaaS) is a franchise anyone can buy into, where bad guys rent out their tricks to other bad guys or wannabe cybercriminals. Lastly, we have wiper ransomware that will “burn it all down” if you don’t pay, forensically wiping everything with no way back. Knowing this helps you spot it, stop it, and clean up the mess.
* Which Systems Are We Covering?
  + Time to make a list of what’s in scope, like endpoints, servers, network gear, OT, and backups. Endpoints include laptops, mobile devices, and that dusty desktop in the break room. Servers that are needed to run your critical processes or house important data, whether they are in your closet or up in the cloud. Network stuff includes firewalls, devices that support remote access (VPN), and routers. Operational tech (OT) includes factory machines or IoT devices. And backups of your data, whether they are in the cloud, offline, or at a managed service provider. This rough framework should help you to start thinking about how things are connected, where they are, and most importantly, how and where they are backed up.
* How’s This Going to Affect the Business?
  + Now we need to consider the impact. What are the critical things we need to keep critical processes running, like orders or payroll? What data is required to support these processes, such as customer details, financial records, and HR data? Is there any legal stuff to worry about, like GDPR or HIPAA rules? And the financial impact, including ransom costs, downtime, lost deals, 3**rd** party labor, and overtime? Determine where the biggest impacts are, and you’ll have a head start on prioritizing your IRP focus.
* Who’s In and Who’s Out?
  + Is it the whole company, every office worldwide, or just one location? What about subsidiaries or third-party partners, are they part of this? And vendors or supply chain vendors, what’s their role if things go sideways? Lay it out, create a flowchart, diagram, what ever helps clarify things so nobody’s standing around like, “Uh, am I supposed to help or what?”

**Metrics of Effectiveness**

How do you tell if this plan’s a winner or a dud? You need some solid checkpoints and success criteria to see how you’re holding up. Here’s a few to consider:

* Mean Time to Detect (MTTD)—How fast can you catch the problem?
* Mean Time to Respond (MTTR)—How quick can respond and begin dealing with the situation?
* Mean Time to Restore (MTTR)—How long until things are back to normal?
* Containment Rate—How many of these attacks did you stop?
* Data Recovery Rate—How much of your encrypted data did you get back?

Keep an eye on these over time, and you’ll spot if you’re getting better or if you need to tweak something, like more practice or better gear.

**Matching the Plan to What the Business Actually Cares About**

This incident response plan (IRP) can’t be all about what IT thinks should be in-scope, it has to include the whole company’s priorities. Tie the IRP into a company-wide (enterprise) strategy and risk approach, and it’ll feel like everyone’s included. Here’s how to make it work:

* **Get the Bosses Involved**—Bring in the C-suite so they’re rooting for it (and tossing some budget your way).
* **Align with Disaster Recovery Plans**—Make sure your IRP aligns with your broader “everything’s gone wrong” Disaster Recovery Plan (DRP) or Business Continuity Plan (BCP).
* **Pull in the other Teams**—Legal, compliance, HR, communications and get their input so we are considering and including all possible perspectives.
* **Find the Balance**—Keep the business rolling while locking things down; security can’t just steamroll operations, even during an incident.

**Key Takeaways from this Section**

* Clear goals keep your response sharp and focused.
* Scope nails down what systems and stuff you’re protecting.
* Success metrics tell you if your IRP is effective or where to improve.
* Linking your IRP to business priorities gets everyone on board with the IR Team.

Assembling Your Incident Response Team (IRT)

Your ransomware incident response plan (IRP) is only as good as the people on your Incident Response Team (IRT). You need a tight and cohesive Incident Response Team that can hit the ground running and deal with a ransomware mess like it’s no big deal. And it’s not just the IT crew, you must include others from across the business to handle everything from stopping the attack to keeping the lawyers and PR people from freaking out. So, how do you set up the roles, keep everyone ready around the clock, and make sure nobody’s stepping on toes when it all goes down? The following sections will help us figure it out.

**Sorting Out Who Does What**

First things first, everyone needs to know their job. You can’t afford to be scrambling to figure out who’s responsible for what when ransomware is encrypting your data and locking you out of critical systems. Define these roles now, so when it’s game time, your team’s not just standing there like, “Wait, is that on me?” Here are the roles you need to assign:

* Incident Response Leader
  + This is your quarterback, the one calling the plays, steering the ship, etc. They are the go-to person, keeping tabs on everything from spotting the attack to getting things back on track. They are talking to the response team, management, maybe even law enforcement or outside experts. Pick someone who doesn’t crack under pressure, communicates well, and can be effective in both technical and business conversations.
* Technical Response Team
  + These are your technical subject matter experts. You need:
  + *Cybersecurity Analysts*—the ones playing detective, figuring out where it started and how bad the situation is.
  + *Network Engineers*—the lockdown crew, shutting off systems, collecting data from the firewall, and stopping it from spreading.
  + *System Admins*—the repair crew, killing processes, pulling backups, and getting stuff running again.
  + *Malware Analysts*—the people familiar with ransomware, determining what variant you’re working with, and possible decryption options (these are usually part of the law enforcement, insurance, or expert incident response teams).
  + *Threat Intel Analysts*—the ones doing internet and dark net searches to see who’s behind it and what they might do next; understanding their tactics, techniques, and procedures (TTPs).
* Legal and Compliance Team
  + You need the legal team in on this because ransomware can be a messy tangle of laws. They’re sorting out if paying’s even an option, handling those “we got hacked” notices (like SOX, GDPR, or HIPAA stuff), and talking with law enforcement if it escalates. They’re your shield from bigger trouble, so don’t hesitate to involve them as soon as you know there’s trouble. It would be great if you worked with your legal contact to develop and obtain preapproval for the communication templates (email, press release, internal communications, etc.) that the communications and public relations team will use.
* Communications and Public Relations
  + When everyone’s eyes are on you, these are your smooth talkers. They’re keeping the team calm with clear updates, prepping what to say to the press, and reassuring customers it’s under control. They’ll even tackle the social media chaos.
* Executive Leadership
  + Big decisions need big players, like whether to pay the ransom or shut everything down. Their perspective can weigh security against keeping the business alive, brief the board, approve requests for additional help, and involve insurance. Identify the primary contact from the Leadership Team, with at least one backup contact.
* Human Resources (HR)
  + HR is your people department. They’re helping to address staff questions and payroll snags, communicating new updates, and setting up training later so this doesn’t happen again. They’ve got the human side covered. Make sure you get to know your primary HR contact and at least one alternative contact.
* Third-Party Vendors and Partners
  + Sometimes you need extra resources, specific skills, or just the A-Team. These may include Managed Security Service Providers (MSSPs), professional incident response firms, and or insurance and ransom negotiators. They bring experience and perspectives gained from other incidents, extra know-how, and extra hands. Identify these resources today, and get the paperwork out of the way so they’re ready when you need them.

**Who’s Calling the Shots**

You need to be clear about “who’s in charge?”. There is no time for arguing when the clock’s ticking. Here’s how it should break down:

* **Primary Incident Lead**—The main shot-caller, making moves and signing off on the big stuff.
* **Backup Incident Lead**—Jumps in if the lead is MIA, even if they’re on a beach somewhere.
* **Departmental Leads & Stakeholders**—Each group (IT, legal, comms) has a go-to person keeping their team looped in and on-task.
* **Executive Sponsor**—The primary executive contact that monitors the response activities, updates the leadership team and board, and provides approvals as needed.

**Making Sure You’re Covered 24/7**

Ransomware doesn’t care about your sleep schedule; though most coordinated attacks happen on Thursday or Friday morning, the response efforts will no doubt carry into the early morning hours. Your team has to be ready anytime. Here’s how:

* **Shift Rotation**—Set up a schedule so a contact from each of the IT, legal, and comms teams is always capable of responding. Rotate them so nobody is “always on”.
* **Contact List**—Keep a list of everyone’s primary and backup phone numbers, and alternate contact methods like Signal or the Microsoft Teams app on their smartphone.
* **Alert System**—Get a loud “RANSOMWARE!” alarm that goes off when it hits. This could the subject line of the email, an SMS text message, phone call, etc.. Train everyone to jump when they hear it. No holes, no slack—24/7 is the deal for the IRT members.

**Cross-Training and Alternate Plans**

These attacks can drag on for days and sometimes weeks. Your response can’t be hindered because someone is on vacation, sick, or unresponsive. Cross-training among IRT members can be critical for an effective response where one or more members aren’t in the game:

* **Cross-train**—Share basic skills and responsibilities across teams to establish familiarity with activities that may need to be reassigned if the primary resource isn’t available.
* **Write It Down**—Keep playbooks and quick reference guides handy so that any IRT member can access them and follow basic procedures.
* **Practice Swaps**—Run drills where alternate resources from other teams take over. Can they handle it? This exercise can build your IRT’s confidence.

**Teaming Up with Outside Help**

You don’t have to fight this alone; build your ecosystem of support to call upon when things get bad.

* **Law Enforcement**—Get friendly with the FBI, USSS, CISA, or your local police department. They’ve got skills, resources, and experience.
* **Insurance Company**—Know your policy and understand what it covers. Have a claims plan and a contact ready. Insurance typically wants to be notified as soon as you’ve determined there is a problem, even if it turns out to be a false positive.
* **Response Firms**—Connect with a professional incident response firm to back you up. Your insurance company may already have pre-approved firms to use, so reach out and get connected. Even if you don’t need their whole IRT, it is often very helpful to have an extra set of eyes and ears from someone that’s been in the trenches and can be your IR Coach or Advisor.
* **Recovery Pros**—These data experts may be necessary to help your IRT with decryption, data restoration, or rebuilding. The key here is to be certain that they understand your environment, any data dependencies, and especially your recovery priorities.

Build these connections now, reach out, meet for coffee, get direct phone numbers, and get the paperwork out of the way as soon as possible. Support vendors do not typically do work without a signed agreement, and getting those reviewed and approved can take hours if not days.

**Key Takeaways**

* Your IRT is a full squad, including IT, legal, communications/HR, executive leadership, and outside help when needed (i.e., special teams).
* Establish a clear chain of command to address any confusion.
* 24/7 coverage and backups mean you’re never caught off guard.
* Outside help brings the heavy hitters when you need them.
* Cross-training keeps your team flexible and ready.

Developing Response Procedures

When ransomware hits, you need a plan that’s easy to follow because the timeliness and effectiveness of your IRT reaction can mean the difference between a bump in the road and a total train wreck. We need less chaos and confusion for your business, data, and reputation. With some clear, no-nonsense steps in place, your IRT can handle this from catching it early to getting things back on track. We must have procedures to address are detection, containment, cleanup, recovery, and keeping everyone on the same page. Write this stuff down, test it, and make sure your team knows it like the back of their hand.

**Initial Detection and Assessment**

Detecting ransomware quickly is critical. The earlier you identify it, the more likely you are to contain it and keep things from really getting messy.

* Detection Tools
  + You need solid tech and some sharp instincts. Things like a good EDR, SIEM, IDS, and XDR systems. Endpoint Detection and Response (EDR), ideally with AI and sandboxing capabilities, can keep an eye on your user devices for anomalous behavior like odd combinations of system file calls and PowerShell use, or robotic beaconing. SIEM systems chew through all of your logs to spot patterns that need to be investigated. Intrusion Detection Systems (IDS) can automatically shun or block weird network traffic. Extended Detection and Response (XDR) tools can correlate all of these capabilities into one platform, some with the use of AI, machine learning, and User Behavior Analytics (UBA), to identify anomalous behavior of devices and users, and conduct automated response like quarantining a device, blocking traffic, or disabling user accounts. Threat Intel Feeds can also help get ahead of a ransomware attack and contribute to the investigation of a potential or current attack. One last note on detection tools: they must be configured and managed well, monitored 24x7, and have defined SLAs in order to be truly effective.
* Ransomware Clues to Watch For
  + What does ransomware look like? CISA has a stop ransomware website (CISA.gov/stopransomware) with tons of information and indicators for various ransomware. A few common symptoms to look out for include: the obvious file extension changes to .locked, .crypt, etc., system performance issues (likely due to the encryption process), spikes in outbound network traffic, files with double extensions, like .pdf.exe, unusual use of PowerShell, or anti-malware or EDR services disabled on an endpoint. If you see any of these, it’s time to act.
* Size It Up
  + Once you see it, figure out what you’re dealing with: Which systems are compromised? What kind of ransomware is it—encrypting, locking, or the double-extortion data-stealing kind? How fast is it moving? If you can figure these things out quickly then you’ll know how to better react.

**Isolation and Containment of Affected Systems**

Your focus at this point needs to be keeping it from getting worse. Lock down, or ideally shut down, the compromised systems, turn off remote access, make sure your backups are safe, and start blocking IP addresses and services at the firewall.

* First Moves
  + Pull those infected systems offline by yanking the plug or disassociating it from the network (revoke IP, certificate, etc.) asap. If you can’t determine how it’s communicating to the command and control (C2), then you have to consider killing the internet connection so it can’t get further instructions or payloads (this step should be preauthorized incase it’s needed in the middle of the night). Review all privileged account activities and consider resetting passwords, revoking MFA tokens, and all current sessions. And, if possible, shut down everything you can before they become infected. Remember, ransomware can’t infect offline systems, and encryption doesn’t work without power.
* Network Tricks
  + Set up zones ahead of time to keep the critical stuff separate from the chaos. Build some digital walls-VLANs. Network segmentation can save your bacon if it’s done right. Also, consider limiting the IP ranges that can even reach your firewall (e.g., if you don’t care about international IP traffic, then block it), subscribe to known-bad IP lists and update your firewall rules, and please do not allow administrative services to be exposed to the internet. Lastly, create a whitelist for all remote access to limit the attack surface even if credentials are stolen.

**Notification, Communication, and Escalation Protocols**

When ransomware attacks occur, you have to let the appropriate people know. Your IRT needs to loop in the right people, inside and out, without wasting time. Time is a recurring theme for effective response.

* Inside Scoop
  + Kicking off the IR plan—the Incident Response Leader declares the incident and gets it moving. Reach out to the IRT through approved methods, including alternate email (not your work email), text, phone calls, Teams (if it’s available), or encrypted apps like Signal. Give your executive contact the rundown, including what’s hit, how bad, what’s next. Provide a similar briefing to the IRT and get the response rolling.
* Outside Calls
  + Confirm with your executive contact about connecting with Insurance, outside legal counsel, and law enforcement. Someone other than the IRT Leader is typically responsible for reaching out to these contacts. Depending on your organization’s preference, contacting law enforcement is either pre-authorized for the IRT Leader or determined by your executive or legal counsel. In either case and as soon as it’s appropriate, reach out to your law enforcement contacts (USSS, FBI, or CISA) and get them looped in. Contacting regulators is also typically not the responsibility of the IRT, and usually lies with Legal or a Compliance department and should be tracked as an IR activity. You also typically have a period of time before these notifications are required (e.g., 48 to 72 hours after incident validation). Notifications to customers, partners, etc., are not typically part of the IRT Leader’s role, but providing the necessary information is essential, such as who was affected, what type of data and how many records, etc.
* Steer the Story
  + A communications specialist should take point on handling any outside communications, like social media posts, phone inquiries, news media, etc. Have some “we’re handling it” messaging ready and preapproved by legal for the press, but don’t overshare. Stick to facts—no wild guesses—to keep your reputation intact and avoid legal headaches later.

**Engaging External Response and Support**

Sometimes, your IRT can’t do it alone, and you need some pros to help with the heavy lifting.

*Response Firms*

Keep a professional IR team on speed dial. These guys eat ransomware for breakfast. Get to know them today, or more importantly, help them get to know your environment and what matters most. They can be a significant augmentation to your IRT, bringing resources for coaching, containment, cleanup, and forensic investigation and recovery.

*Law Enforcement*

Federal agencies often have a lot of good intel on ransomware actors and sometimes some good recovery advice, tricks, and possibly a decryptor from previous cases. If you do pay the ransom, coordination between you, your insurance, and these Cybercrime units can track the ransom payments or crypto trails.

*Insurance Carrier / Broker*

Your insurer has a list of approved go-to firms, including IRT, cybersecurity, digital forensics, technical remediation, and lawyers. In addition, if management decides to pay the ransom, it’s your insurer that will typically negotiate and make the payment. Coordinate as much as you can through your insurance to make sure the bills get covered in your claim. Don’t mess that up.

**Deciding on Ransom Payment**

Paying the ransom should be your last resort and the IRT needs to ensure that executives have a clear and current understanding of the likelihood and timeline for recovery so they can make the best-informed decision possible. The decision to pay a ransom should not be made by the IRT, but there are several scenarios where the IRT may heavily suggest the need for a decryptor in order to recover. Think it over hard, it is not a comfortable decision to make. Things to consider:

* Legal Angle—Can you even do it, or is it against some rule?
* Will It Work?—Any chance they’ll give you a legitimate decryptor, or just take your money?
* Double Dip Risk—Could they still leak your stuff after you pay? Or attack you again via some dormant back door, unpatched vulnerability, etc. A large percentage of those who pay the ransom will be attacked again within the same year.
* Backup Check—Can you dodge it and restore from the backups you have? Can we ensure that the restore doesn’t also include the ransomware?

**Key Takeaways**

* Catch it quick and figure it out—speed is your friend.
* Lock it down fast to stop the bleeding, spread, and encryption.
* Talk straight and swift—communicate effectively to update everyone that needs to know.
* Outside help is a lifeline when the IRT needs a boost.
* Paying is a roll of the dice—don’t rush the decision.

Data Recovery and Remediation

Alright, the ransomware’s out the door—contained, wiped out, done. Now comes the messy part: picking up the pieces. Recovery is no walk in the park. It’s about getting your business back on its feet by restoring your systems to a usable state and ensuring this doesn’t happen again. You’ve got to be smart about your recovery efforts, or you’re just begging for more downtime and stress. The upside? With a solid plan, you can bounce back quicker, keep your data safe, and be more resilient for the next incident. Here are a few thoughts:

**Assessing the Damage and Recovery Scope**

Before you start patching things up, you’ve got to know what was impacted or infected. It’s like checking the wreckage after a crash—your Incident Response Team (IRT) needs to dig in and see how bad the damage is.

* What Got Hit? Determine which systems and related data was infected. Then assess how the loss of these systems impacts the business—can you still take orders, make product, provide services, or pay bills? Now is also the time to start prioritizing your recovery efforts. Customer information and financials beat that random marketing folder most days (i.e., unless you are a marketing company).
* Is It Still Good? Make sure that what doesn’t appear to be infected is truly ok to keep. Whether it’s an unaffected system or the data you are restoring to a rebuild server, it is imperative to use one or more anti-malware tools that are different than the one you were using before the attack. Ransomware and the initial breach payloads love sneaking into backups so that you self-infect during remediation. These safety checks should be done in a disconnected / air-gapped environment to ensure that any system that continues to be or becomes re-infected doesn’t spread back onto the production network.
* Are We Ready? Double-check your backups, and please say they are not encrypted. Once your rebuild (stressing a strict rebuild policy and not a clean and restore approach), each system and data is restored to them, ensure they are thoroughly tested before introducing them back into your production environment. Take this opportunity to start fresh, rolling back all rogue software, permissions, services, and privileges and limiting the environment to only what the business and users need.

**Restoring Data from Backups**

If your backups are good, they are your golden ticket to getting back to normal. But don’t just assume they are clean and complete or that this is a simple restore job.

* Keep It Offline—Make sure your pot of gold is untouchable until you are certain that the ransomware threat has been addressed. Even then, the first phase of recovery should be to limit access to your backup for only those systems or recovery network segments you are working with – not the whole network.
* Pick Your Battles—Start with the essentials (prioritized from your earlier efforts), like Internet access, email, your ERP (financial) system, or manufacturing and shipping. Create a restore plan for what is expected to come back online and when, and communicate this to management to be disseminated at their discretion.
* Check Your Work—Make sure those backups are good, the restore is completed without errors or corruption, and the rebuilt systems are stable. Take notes on any issues and how they were addressed. Start to breathe easier, but don’t pat yourself on the back (or ask for a raise) until you see it all working like it was before “the big one”, which is how this will likely be referred to in the stories you tell your grandkids.

**Evaluating Decryption Options (and possibly life choices)**

Ok, so we’ve determined that the backups are hosed, encrypted, kaput, or heaven forbid you simply just didn’t have any. Now you need to start thinking about how to “decrypt” the mess this ransomware has made, initially focusing on how to accomplish this without paying the ransom.

* Free Tools First—Check the threat intel sites, like CISA or NoMoreRansom.org, and reach out to professionals or law enforcement for available decryption keys. Ransomware experts could have some hacks too and though they often charge a hefty fee, it’s often significantly less than the ransom demand.
* Talking to the Bad Guys—Do not contact the ransomware actors directly on your own. If you are truly stuck without a recovery option and thinking of paying the ransom, the ideal next step is to work with your insurance provider. If you don’t have insurance, then you should find an insurance carrier that will help you get through this. Insurance companies often have the resources and negotiators you need at this point. Alternatively and at a minimum, get your company’s legal or outside counsel and law enforcement involved. You do not want to shoulder the risks associated with doing this on your own.

Much like real-life ransoms, you must ask for proof of life before paying anything. This requires the attacker to accept encrypted files you send to decrypt and prove that your money will be well spent. Make sure you send them examples of critical files, not just your resume. But fair warning and in any event, paying the ransom is a crapshoot, and they might ditch you or bale when you find out the decryptor doesn’t work on everything.

**Managing Ransom Payment Considerations**

Paying the ransom should be your last resort (though I’ve seen several organizations able to recover and management still decided to pay the ransom – go figure). Get the executives involved, and make sure the legal team and insurance are all together for this event.

* Legal Considerations - Make sure that your ransom payment isn’t illegal, like sending money to a restricted group (OFAC rules). Your insurance provider is accustomed to determining this and filing an exception if needed. Also, work with legal to determine if the attack constitutes a data breach, and if any exfiltrated data should be considered compromised (e.g., if the exfiltrated data was encrypted by you before it was stolen then there is a chance that it isn’t compromised).
* Money vs. Mess – Provide management with the information they need to support the decision to recover on your own or pay the ransom, including data for assessing the difference between the ransom price and how long it will take the IRT to get things back online. Just remember that there is another factor that must remain front of mind – there’s no honor among thieves and they could and likely will just take the money.
* Double Whammy - If they also stole your unencrypted data in addition to the ransomware (double extorsion) most executives lean towards paying the ransom in an effort to keep things from getting worse (e.g., posting your data on the internet, listing your company on a wall of shame monitored by the media, or selling it to bidders on the dark net). There is absolutely no guarantee (zero or negative percent) that the attackers will destroy any data they took from you – how would you prove that? If the data they took from you was readable, you are just going to have to face the music and start assessing the impact this data breach is going to have – invest your money (hopefully insurance money) in that instead of paying the ransom.

**Rebuilding and Restoring Systems**

The data is flowing from your backups like a warm lazy river of comfort, so now is the time to tweak your system configurations, network restrictions, firewall rules, etc. Let’s lock things down.

* Start Fresh - Don’t trust anything ransomware touched. Wipe it and rebuild from scratch with clean installs, and patch every gap before putting it back on the production network.
* Tighten Things Up - Add MFA to everything you can and require MFA for every authentication (i.e., prohibit the browser from saving the token, trusting the device, etc.). Review user access profiles and limit who has access to what. Review all remote access and restrict it to the extent possible. Reset everyone’s passwords and enforce a strong policy for the new ones. There will be no more open doors.
* Ease Back In - Bring your fresh newly built, and better-secured systems back online and watch for any hiccups. Monitor your “new” EDR solution (since the last one didn’t help much) and run a vulnerability scanner to look for any remaining weak spots, and keep an eye on your network traffic for anything weird.

**Forensic Investigation and Root Cause Analysis**

We can’t forget to go back and figure out how this mess started. Digital forensic analysis and old-fashioned investigative curiosity can help look back at the facts, events, and data. A few things we should be looking at and the insurance company will be interested in, include:

* How did the ransomware get in? – Was it a Phishing email (that would be so great if it was really that simple)? Was it stolen credentials that then led to compromised remote access (often due to scrapped MFA tokens saved in a browser)? Or a network vulnerability we may have missed?
* Preserve the Evidence – During your IR, make sure you are grab all the logs, files, traffic and anything requested by insurance, or law enforcement. Make sure they are stored offline and secured from misplacement or damage.
* Plug the Holes – Review your public/internet facing systems for any signs of abnormal activity, vulnerabilities, or other malware.

**Key Takeaways**

* Offline backups are your lifeline – make sure you do them, and often (Stressing offline).
* Paying the ransom is a last resort that has no guarantees.
* Rebuild everything that was impacted from scratch, patch, harden, and scan.
* Figure out why it happened and how, then fix as soon as possible.

Post-Incident Review and Continuous Improvement

So, you survived a ransomware attack - nice work! Your T-shirt is in the mail! You locked it down, got your systems back, and hopefully did it all without paying the ransom. Solid moves, but don’t book your cruise just yet. The real win comes after the post-incident review. Think of it like a team huddle after a rough game. Get everyone together, review all details of what happened, compare notes, and figure out how to get better next time. Because, yeah, I hate to break it to you but there’s always a next time. Ransomware keeps evolving, and your IR plan has to keep up. Here are a few things to consider for a post-incident review:

**Conducting a Post-Incident Review**

First up: get everyone involved in the response effort together, including your IRT, IT, experts, executives, legal, HR, and any outside help. At this point, however, this is typically an internal review that often does not include insurance or outside counsel – we’ll provide them a brief once we have are ducks in a row. Remember, this is not about pointing fingers or placing blame, it’s about working together to put all the puzzle pieces together. Keep the review session casual and comfortable, order lunch, and maybe do it offsite. Here’s what we need to figure out:

* Timeline – Start compiling events and activities into a consolidated timeline. Try to determine when the attack started, how long it took until someone noticed, and everything else up until you stopped the attack.
* Wins – Capture all the strong points, like automated alerts, and quarantine. Solid procedures, like employee notifications, and following the IR Plan. And any impressive moves that someone did to improvise, innovate, or otherwise save the day.
* Setbacks – Were there any delays, gaps, or missteps? Issues with communications, contacts, configuration settings, or backups?
* Ah-Hah Moments – Were there any moments that came as a surprise (both good and bad) that need to be considered for improving the IR Plan, fill gaps in staffing, skills, or technical capabilities, etc.? Did anything blindside you?
* Damage Report - How long were you down? What did it cost? What was the fall out (e.g., delayed orders, lost customers, penalties, etc.)?
* People Power – Incident Response is nothing without people. From the first person to observe or report the problem, to every last contributor working to stop the attack. Everyone in the organization must be treated like part of the solution, and given access to the resources to play their part. This means awareness training, understanding of procedures to follow, roles and responsibilities, etc. Were there any gaps in your people power?

**Updating the Incident Response Plan**

Time to remediate and look for ways to improve the IR Plan. Your review is going to find gaps, outdated information, and opportunities to improve the processes. Examples often include escalation delays or confusion, containment issues, or communications issues that include email, Teams, or phone outages. Here’s how to shore it up:

* Tune the Steps – Review the critical activities, including identification, impact analysis, containment, and recovery. Now that you’ve been through the fire, look for ways to improve timeliness, effectiveness, communication, and confidence.
* Who’s on First - Review roles and responsibilities. Add new roles, responsibilities, and IR Team members. Modify current roles and responsibilities by expanding where necessary and, in some cases, reducing what’s expected from others. Create a matrix to capture IR team role names with their respective Responsible, Accountable, Consulted, and Informed responsibilities (a.k.a., the RACI Matrix).
* Improve Communications – Communications can always improve. Consider what you and the organization could do better with regard to communications among the IRT, from the IRT to management and third parties, and from management to employees, contractors, employees, and the media.
* Remediation Management – All concerns identified during the incident must be remediated as soon as possible, and someone needs to take point on managing these to completion. Bad guys already know you are vulnerable and will try again and or tell others about it. This must include technical concerns, procedural improvement, staffing and skill gaps, agreements and contracts for future support, etc.

**Conducting a Root Cause Analysis**

There are so many questions that we would the answers to, but do we have the time, resources, and management support to get there? A root cause analysis can help answer the questions about what was affected, why did things happen the way they did, what kind of ransom was it, and how were the attackers able to get in? Time to play detective.

* Ingress Point – How did the ransomware get it? Phishing? Weak RDP? Compromised Account? Exploited vulnerability in a perimeter system? A priority should be to find the ingress point, patient zero, and anything else that would help you figure out how this started (e.g., weak passwords, unpatched software).
* Blind Spots – How did the attack sneak past our detection and defenses? Gaps in capabilities? Missed alert? Once you determine how they got it, you should be able to figure out why you were alerted sooner (i.e., before the ransomware started doing its thing). Document the gaps or deficiencies and start looking at tools, tech, skills, and services that will address any blind spots.
* Attacker TTPs – What tools or tactics did they use? Any related IP addresses? Did you find the dropper file? These are a few pieces of the Tools, Tactics, and Procedures (TTPs) that could help identify the attacker and allow you to incorporate these into your alerting policies. Law enforcement would also benefit from this information in further pursuing these criminals.

**Training and Awareness Updates**

Alright, you’ve figured out how they snuck in (right?). Now it’s time to make sure it’s not a repeat performance and review how you train your users and keep them current on attack trends, attacks, and what to keep an eye out for. Time to get the crew up to speed.

* Phishing 101 – Statistically, most ransomware is delivered via phishing emails. If it was an email that gotcha, ramp up the training. There are several options here, ranging from custom exercises to automated awareness platforms that include videos, quizzes, and scheduled recurring phishing exercises. Continuous training and exposure to current events will help with the user-focused attack vector.
* Lock It Down – Implement and enforce MFA on everything, but at least Windows authentication and cloud environments, and remote access – every time and every login. Review all privileged access accounts, reduce as appropriate, and implement logging and alerting for these accounts.
* Speak Up – Make sure everyone is comfortable with the “See something, Say something” culture. Make sure everyone knows the expectations and the process for reporting suspicious events or incidents.
* Remote Access – Review your remote access methods, users, and configurations. Shore up the remote access methods and improve where you can. Reduce the number of remote access users to the extent possible and whitelist IP addresses if possible. If you provide remote access to vendors, consider turning this off until support is needed.

**Testing and Refining the Plan**

Updating the IR Plan should be done at least annually, and after any incident, but don’t forget to take it a step further and walk through it with your team. From an audit perspective we consider these two different activities: design and operation. Writing and updating the IR Plan is a design activity, which can be assessed as effective from a read through for completeness. Actually testing the IR Plan is an operational assessment to ensure that a) it was designed well; and b) it operates as intended. Operational testing is typically referred to as either a tabletop exercise (TTX) or a wargame.

* Tabletop Exercises - Run a moderated simulation of a ransomware attack to test your updated IR Plan. There are resources available online from CISA and other organizations to help you design and run these TTXs. There are also professional services firms that can custom build, moderate, and facilitate these for you. A quick internet search will get your started. The important outcome from the TTX is to determine if additional updates need to be made, or is the IR Plan good – for now?
* Backup Restore Testing – As often as possible, backups should be tested for integrity, restore metrics, and variations of file, volume, and whole system recovery. Document these activities so you get credit in an audit or post-incident review, and address any issues you run into. This will also help solidify your recovery time objectives (RTO) and give you the information you need to set expectations during an IR.
* Test Your Tools – Did you update your toolset, configurations, or capabilities after the breach? Make sure to test these as well, ensure they are operating effectively, identifying threats faster, and providing the information we need to better respond.
* Cross Train – Switch up the roles and test your alternates to make sure everyone if comfortable and capable in their backup responsibilities.
* Keep fine-tuning – Ransomware is always changing, and so are the TTPs of the attackers, so your IR Plan has to keep improving as well.

**Key Takeaways**

* A solid post-incident review will highlight what rocked and what tanked.
* Everyone’s input, notes, and perspectives will help catch the things you might overlook.
* Root cause analysis is important and can shed light on the “how and why”.
* Training keeps the team and the employees sharp, aware, and ready.
* Periodic testing proves you’re not just talking a big game, the IR Team can actually deliver.

Planning an Annual Ransomware Tabletop Exercise

Planning an annual ransomware tabletop exercise might sound like a daunting task, but it’s really about bringing people together to think through a messy, all-too-real scenario in a low-stress way. This section will walk you through the nuts and bolts of setting it up, figuring out who needs to be there, what goals you’re aiming for, and how to keep it practical and engaging. It’s less about fancy tech and more about getting everyone on the same page.

**Defining Exercise Goals and Objectives**

Okay, so you’re setting up a ransomware tabletop exercise. Step one is figuring out what your success criteria is. This isn’t just busywork; it’s your chance to see how the IR team and other stakeholders handle a ransomware scenario. You need some clear targets to keep this on point. Here are a few to focus on:

*Testing IRP Effectiveness*

First up: how solid is your Incident Response Plan (IRP) when it’s go-time? You want to know if your IR Team can grab the playbook, talk it through, and pull in the right resources and tools without fumbling. Think of it as a practice game; does your plan actually hold water, or is it just a fancy document created to check a box? Realize also that the TTX is often the first exposure most IR team members have to the Plan; it is absolutely ok if it isn’t perfect – that’s one of the primary objectives of the exercise.

*Spotting Holes in the Game Plan*

While you’re at it, take note of things that weren’t considered or documented in the IR Plan. Are there parts of the response that feel off or confusing? Maybe the team’s stumped on something; dig into those and talk through it now. Better to address the weaknesses now than when ransomware’s already tearing through your network.

*Getting Everyone in Sync*

Ransomware’s not just IT’s problem (I say this a lot). It will eventually involve management, internal legal, external legal, HR, PR, insurance and risk, compliance… pretty much everybody. This exercise is your shot to see if all the stakeholders can get through it. Is the information flowing, can decisions be made, or are people tripping over each other? If the teams aren’t on the same page, a real hit could turn into chaos. Pro-tip, if you expect there will be contention or political issues, then you should consider having a third party moderate and run the exercise.

**Selecting Participants**

Choosing the right participants for your ransomware tabletop exercise is important because you want to make sure all the key players are involved. This isn’t just about the IT team, it’s about bringing together all the different departments that will play a role during an actual ransomware attack.

*Including All Key Stakeholders*

First and foremost, you need to bring in all the essential teams, like IT, security, legal, PR, and executive leadership for example. Each of these groups will have a critical role during a real attack, so it’s important that they practice how they’ll work together during the exercise. Having them involved will help ensure that everyone is on the same page and knows their specific responsibilities when the time comes. A RACI diagram can help visualize this and be a good reference for the participants.

*Involving External Participants Where Appropriate*

Depending on your organization’s situation, it can also be valuable to include external participants. For example, if you work with Managed Security Service Providers (MSSPs) or cybersecurity experts, invite them to participate. You might also want to bring in the insurance company or regulators if applicable, especially if your industry has specific compliance requirements related to data breaches. These outside perspectives can provide valuable insights and make the exercise feel more realistic.

By carefully selecting the right mix of participants, you’ll ensure that the exercise covers all the bases and gives you a thorough test of how your response plan holds up under pressure. Some organizations have two exercises with the first being a scrimmage among the tighter team that builds IR Team confidence for the second exercise with the broader participation.

**Scenario Development**

Creating a solid, realistic ransomware attack scenario is the heart of your tabletop exercise. You want to simulate a situation that could actually happen. This should be something that feels real and challenging, so your team can respond effectively when the time comes. Here are a few objectives:

*Crafting Realistic Ransomware Attack Scenarios*

When designing your scenario, think about the different types of ransomware you might face. It could be encryption ransomware that locks up critical files, or maybe something more complex like double extortion, where attackers not only encrypt data but threaten to release it publicly unless they get paid. You want to mix things up and consider all possibilities. A precursor to defining the scenario may be an exercise in determining the most likely threat vector and impacts specific to your organization.

*Considering Different Attack Vectors*

Think about how the attack could enter your system. Maybe it’s through a phishing email that tricks an employee into downloading malicious software. Or perhaps it’s via an exposed RDP (Remote Desktop Protocol) connection that gives attackers a way into your network. The attack vector sets the stage for how the incident unfolds and helps your team prepare for multiple types of entry points and what defensive controls and capabilities that come into play.

*Simulating Business Impact*

It’s critical to think about the real-world consequences to your business. What happens if systems are down for hours or days? How does a system outage impact your daily operations? For instance, consider how a disruption might affect order processing, manufacturing, customer service, or financial reporting. You want the exercise to reflect the chaos and the pressure of trying to keep things running while managing the attack.

*Adding Complexity and Time Pressure*

A good scenario isn’t just a straightforward “here’s what happens, now respond” situation. You want to inject a bit of chaos by adding unexpected elements as the exercise goes on, like a second ransom demand, the absence of an IR Team member, additional systems being compromised, or even fake media inquiries. Putting time pressure on participants can also help simulate the stress of a real attack, forcing them to make quick decisions and respond effectively under pressure. A third-party moderate is good at throwing wrenches into the mix.

By creating a detailed and multifaceted scenario, you’ll ensure that your team gets a true-to-life experience. They’ll practice dealing with multiple layers of complexity and learn how to think on their feet.

**Facilitating the Exercise**

Now that your scenario is ready, it’s time to run the tabletop exercise. This is where your IR team puts the plan into action, and the goal is to create an environment where they can think critically and make decisions as if it’s an actual ransomware attack. Here are some tips for facilitating the exercise.

*Designate a Moderator/Facilitator*

You’ll need someone to lead the exercise and keep it on track; a moderator who understands ransomware, your ransomware response plan, and the scenario. Their job is to guide the discussion, ensure everyone stays involved, and introduce unexpected twists, when appropriate, to test how the team handles pressure. They’re essential for keeping the exercise focused and productive. You should also designate a scribe who will observe, take notes, capture ah-hah moments and action items, and summarize the exercise from an observer’s perspective.

*Encourage Active Participation*

The point here is to get everyone engaged; no one should be sitting quietly. From IT to executives to PR, every participant needs to take part in tackling the crisis. As the facilitator, prompt people to share their thoughts, offer solutions and work together. It’s about team building and making sure everyone’s voice is heard in shaping the response.

*Document Key Decisions and Bottlenecks*

As the exercise unfolds, keep track of what’s happening. If you have a scribe, they will capture the decisions made, the steps taken, and any issues or delays that pop up. This could be uncertainty about roles or slow communication between teams. Recording these details is critical for the debrief afterward, giving you clear insights into what’s working and what needs adjustment.

*Keep the Pressure On*

To make it realistic, add some urgency. Ransomware attacks often come with deadlines for payment, but there is also a timeline for recovery so introduce surprises, like a change ransom demand or a new system going down. Use these to push the team to act quickly. This helps represent the stress of a live incident, showing how well they adapt and stay focused when things get intense.

By facilitating the exercise with these steps, you’ll turn it into a valuable learning opportunity. It’s less about getting everything perfect and more about getting it out on the table an preparing your team to respond confidently when it counts.

Conducting a Ransomware War Game

Conducting a ransomware war game takes things up a notch from the planning you’ve already mapped out in your annual tabletop exercise. While the tabletop was all about talking through scenarios in a calm, discussion-based setting, like a strategy huddle over coffee or lunch, a war game throws you into the deep end with a live, hands-on simulation that feels closer to the real deal. In this section, we’ll walk through how to conduct a war game, including key players, the mock attack, and keeping it real without letting it get out of hand. It’s your chance to pressure-test your IR team’s reflexes.

**Difference Between a Tabletop and a War Game**

When you’re gearing up to handle a ransomware mess, you’ll hear about two big ways to practice: tabletop exercises and war games. Both are super useful, but they’re not the same thing, each has its own benefits and purpose. Let’s break it down.

*Tabletop Exercises (Strategic and Procedural Focus)*

A tabletop is like a group huddle with snacks. You grab your team, sit around, and talk through a ransomware attack step by step, following your incident response plan. It’s all about hashing out what you’d do, who’s handling what, and how you’d keep everyone in the loop. The point is to verify that the plan actually makes sense, content is accurate and current, and that the steps clear. It’s less about doing stuff and more about nailing the strategy, ensuring it’s comprehensive, and making sure you’re all on the same page.

*War Games (Tactical and Operational Focus)*

War games are next-level intensity. This is way more hands-on like you’re in the middle of an actual ransomware attack. Instead of just chatting, your team is at their keyboard and taking action, making decisions as the incident unfolds. It’s fast, it’s stressful, and it might even create some competition between the “red teams” (the pretend bad guys) and the “blue teams” (your defenders). This one’s about seeing how your IR team holds up when the fire starts and how quickly they can put it out.

**Setting Up the War Game Environment**

Setting up a ransomware war game that actually works is all about making it feel real, like, *really* real. So the more people you can get involved in the exercise, the better. A few tips include:

*Live Attack Simulation vs. Discussion-Based*

Tabletops are all about talking it out, strategy, decision tree conversation, etc. A war game is intended to crank things up by simulating a live attack. With a live simulation, you’re mimicking that gut punch of a real ransomware attack. Systems might actually go offline, calls are made to administrators to react, and the stopwatch starts tracking your recovery efforts. It’s about putting actual people in a pseudo-real situation and tracking the real response.

If that’s too much, you can still run it discussion-style by walking through each move step by step. Either way, the trick is to make it feel as close to the real deal as you can get.

*Real-World Consequences and Time-Based Decision-Making*

A solid war game makes the objectives and consequences feel real. You want scenarios where every decision is stressful and time is a constant influence, you’re short on options, and what you decide could impact the outcome. It’s about getting your team to think and move like they would when it’s all on the line. A critical system going down, for example, and time’s ticking - how fast can the team respond? Or a ransom demand pops up on another critical system - how does the team reach and handle things when the fire spreads?

*Involving Red Team/Blue Team Dynamics*

Creating Red and Blue teams can be fun and competitive. The red team plays the bad guys, looking for ways to deploy their ransomware, while the blue team is your defenders, monitoring your systems, investigating anomalies, and putting out fires. The red team is there to poke holes and push the blue team into tough spots, adding some real-time tension you can’t fake in a conversation-based exercise. It’s a great way to spot weaknesses in your defenses that could detrimental in a live attack.

*Playing in a Sandbox*

Building an environment for a ransomware war game is all about creating a sandbox that’s safe, realistic, and realistic enough to feel like the real thing. You want a setup where your team can stumble, experiment, and learn, so start by creating or carving out an isolated virtual environment. Think of it like a digital escape room: it’s got to feel authentic but stay contained. Here’s how to pull it together, with some tools and resources to consider:

* First, set up a virtual network using something like VMware or VirtualBox -both are good for quickly creating virtual machines (VMs). You’ll want a mix of systems to mirror your real-world setup, like a couple of Windows workstations, a server or two (maybe one running Active Directory), and even a Linux box if that’s in your environment. Keep it simple but representative. Don’t overcomplicate it with 50 machines if your company only uses five. Install basic software that your team knows, like Windows, Office apps, or whatever tools your business runs on, so it feels familiar when the “attack” happens.
* Next, you need the ransomware feel without the actual danger. Tools like the open-source “RansomSimulator” from KnowBe4 or custom PowerShell scripts can fake an encryption process by locking files or popping up ransom notes without damaging your VMs. If you have some coders on your team, they can customize scripts found on GitHub repositories like “Invoke-Malware” to simulate file encryption or network propagation. Just make sure everything related to this exercise is done in the sandbox, including testing these scripts. For the network-level stuff, use tools like Kali Linux on a separate VM to mimic an attacker conducting pings to flood or spoof traffic with something like hping3. It’s low-tech but effective for stress-testing your detection.
* Data’s your bait, so seed the environment with dummy files, like fake customer lists, financial spreadsheets, or even a mock HR database. Tools like Mockaroo can generate realistic-looking datasets quickly. Scatter these across your VMs, maybe on a shared drive, to give your team something tangible to “protect” or “lose.” If you’ve got logs from a real system, anonymize them with a script and import them into a lightweight SIEM (if you don’t have one) like Graylog or ELK Stack to let your defenders practice spotting the bad stuff.
* Resources-wise, look for freebies where you can. NIST’s Cybersecurity Framework has a ransomware appendix (Special Publication 800-53) with scenarios to use. CISA’s Ransomware Guide has attack patterns you can replicate, like phishing emails or credential theft. If you have some money to spend, consider a platform like AttackIQ or SafeBreach to automate some of your war game with preloaded ransomware behaviors.
* Keep it fenced off by using a dedicated VLAN or air-gap it if you’re paranoid. Test everything yourself first - launch the “attack,” watch it unfold, and make sure it’s tough but not impossible. Your intent isn’t to humiliate the team but to make them better. Once everything is ready, hand the keys to your red team and let the war game begin.

Building an environment that cranks up the heat will show you how your IR team holds up when it’s crunch time, how your tools and plans stack up, and where you might need to level up with more tech, training, or tweaks.

Setting up a war game is no joke, it’s intense and takes time to build - but it is hands-down one of the best ways to get your team ready for the real thing. With a live attack feel, SLAs, and that red/blue team push-and-pull, you’re giving everyone a workout that’s tough but also a total game-changer.

**Evaluating Team Performance**

The dust has settled from your ransomware war game, now it’s time to figure out how your team did. This part is important because it’s where you review what went well, what didn’t, and what needs some work. You’re not just checking how they tackled the attack but also how they held it together and made calls when the heat was on. Here’s how to break it down:

*Identifying Weaknesses in Detection and Response*

First up, look at how fast and sharp your team was at catching the ransomware and jumping on it. Did they find it early enough to keep it from spreading? Could they determine what kind of ransomware it was and take the right steps to shut it down?

Dig into where things got messy. Was the early detection phase a scramble? Did they miss some red flags? Maybe containment took too long, or they went down the wrong path. Those hiccups are very valuable to see and they show you exactly what to tighten things up for next time.

*Testing Communication and Decision-Making Under Pressure*

Ransomware hits hard and fast, so you’ve got to see how your team handled the stress. Did they communicate to the right people in a timely manner? Did they keep their cool and stick to the incident response plan, or was it chaos? How tight were the teams - IT, PR, legal, etc.?

In a war game, communications are everything. One slip or delay in passing the ball can significantly set you back. Keep an eye on the lines of communication, including email, phone, secure chat, whatever, and if it actually worked.

*Assessing the Readiness of Backup and Recovery Plans*

Your backup and recovery game plan is your lifeline to getting back online. During the war game, check how effective they were at pulling it off. Did they know right away where the backups were and how to get recovery started? Any holdups or headaches getting data back?

Look at how they stuck to the recovery steps in the IR Plan, like double-checking the backups weren’t trashed and making sure systems were good before flipping the switch. Were there any issues or gaps in these activities that could bite you in a real attack?

*Debriefing and Feedback*

Once you’ve sized it all up, sit down with the team for a debrief. This is your chance to hear from everyone. Talk about any curveballs or rough spots, and get their thoughts on what they would change next time.

That back-and-forth is what sharpens your IRP and levels up your game plan for the real deal.

Bottom line: sizing up your team’s war game performance is about looking for the weak spots, seeing how they perform under pressure, and making sure your recovery is on point. Dig into this stuff, and you’ll know your IR team is ready to roll when ransomware hits.

Post-Exercise Review and Improvement

Welcome to the “Post-Exercise Review and Improvement” phase, where the real payoff starts. This isn’t about patting yourselves on the back or pointing fingers; it’s about really assessing what worked, what flopped, and how to get better before it really happens. In this section, we’ll walk through debriefing your team to hear their feedback, update the IR plan to plug holes and boost training, and track how your IR team’s readiness stacks up over time.

**Conducting a Debrief**

The debrief is a big deal because it’s your chance to take a step back, review how things went, and highlight the good and the bad.

*Gathering Participant Feedback*

Get feedback from everyone involved. Every perspective counts. Start with some laid-back questions like, “What went well during the exercise?” or “Where’d things get crazy?” Hearing from the whole team paints the full picture. These debriefs can be done one-on-one or in a group setting.

*Analyzing Performance Gaps and Strengths*

Once you’ve got everyone’s feedback, you need to build the big picture. Were did the team stumble or couldn’t make a decision? Maybe some steps flawed or dragged on too long? Maybe there were moments where everything was awesome. This is when you spot the weaknesses and the wins. You need to celebrate them both in order to get better.

*Actionable Takeaways*

The whole point of this debrief is to walk away with stuff you can actually use. What needs a tweak, and how do you make it happen? Maybe it’s tightening up your incident response steps, improving communications, or getting the team some extra training. This is your shot to beef up the plan for next time.

**Updating the IRP**

After the debrief, it’s time to take all the insights you’ve gathered and update your Incident Response Plan (IRP). The goal here is to make sure that your plan reflects the lessons learned during the exercise and is ready to handle real-life situations even better.

*Adjusting Response Procedures Based on Findings*

Based on the feedback and what you observed during the exercise, look for areas where your response procedures might need tweaking. Maybe certain steps were too slow or not as clear as they should have been. Adjust those steps to make them smoother and more efficient. It’s all about making sure your team knows exactly what to do at each stage of a ransomware attack, without hesitation.

*Closing Security and Operational Gaps*

Did the exercise reveal any gaps in your security infrastructure? Perhaps there were vulnerabilities that went unnoticed or weren’t addressed quickly enough. Take a hard look at where your security systems fell short and make a plan to patch those gaps. Also, assess your operational workflows to ensure that your business continuity can keep running smoothly if the worst happens.

*Enhancing Training and Awareness Programs*

Sometimes, weaknesses are tied to knowledge gaps within the team. If you found areas where team members were unclear on their roles or the tools at their disposal, it’s a good idea to beef up your training. Ensure that everyone is up to speed on the latest procedures, tools, and technologies to respond to a ransomware attack. Regular training sessions can help everyone feel more confident when the real deal happens.

**Tracking Progress and Maturity**

After the debrief and the updates to your IR Plan, there is still work to do. Now it’s about keeping tabs on how things are shaping up over time – your IR maturity. This is how you know you’re actually getting better and improving your readiness with every exercise or real-deal incident.

*Setting Key Performance Indicators (KPIs)*

KPIs are your way of tracking if this plan’s really doing its job. Think of them as your scorecard. You could track things like how fast your team spots a ransomware attack, how quickly you lock down systems, or the effectiveness of communications. These numbers give you hard proof of what’s working and where you’re still improving.

*Monitoring Incident Response Readiness Over Time*

Wrapping up one exercise doesn’t mean you’re “all good”. You’ve got to keep an eye on your team’s readiness. Conduct regular check-ins on the status of remediated processes or capabilities and if they are working. If the same problems keep popping up, that’s your cue to dig deeper in the next round of training or practice.

*Looking for Trends and Patterns*

As you keep at it, watch for any patterns that stick out. Is your team performing well every time, or are there still some stumbles? Spotting those trends lets you zero in on the stuff that needs more attention.

Conclusion

When it comes to ransomware preparedness, the work never really stops. It’s about creating a mindset and a culture where everyone in your organization understands the importance of being ready for a cyberattack and always working to improve how you respond.

**Importance of Continuous Improvement and Adaptation**

The cyber threat landscape is constantly changing. Ransomware tactics evolve, and attackers find new ways to gain access to systems. That’s why your incident response plan can’t be static. You need to keep improving and adapting your processes based on new information, insights from exercises, and lessons learned from real incidents. This mindset of continuous improvement will ensure that when a ransomware attack hits, you’re not only prepared but also resilient enough to handle it.

**Encouraging an Incident Response Culture Across the Organization**

Building an incident response culture means making cybersecurity a priority at every level of your organization. It’s not just up to the IT team; everyone from leadership to your employees should understand their role in preventing and responding to ransomware. Encourage regular training, run exercises often, and make sure everyone is clear on their responsibilities in case of an attack. When an incident happens, everyone will know how to react swiftly and effectively, which can make all the difference in reducing the impact.

**Next Steps: Scheduling the Next Exercise and Reviewing IRP Updates**

Once your current plan is updated and your team has learned from past exercises, it’s time to start thinking ahead. Schedule the next tabletop or war game to keep things current. Review your IRP periodically to ensure it still reflects the current threat landscape and your organization’s evolving needs. The goal is to keep refining and strengthening your defenses so you’re always prepared for whatever comes next.