Practical Ideas for Ransomware Resilience

with VMware Cloud Infrastructure

vmware®

January 2023

Table of Contents

Introduction

Ransomware is a major threat, and many organizations are looking for ways to protect themselves. This discussion covers steps that virtual infrastructure administrators can take to make their environments more resilient against ransomware attacks. These ideas include both technical controls and people and process changes. Defending against ransomware requires a mix of approaches, including returning to information security basics, implementing new technology, and making organizational changes.

Some of these ideas have benefits beyond just ransomware defense, such as improving system resilience, encouraging automation, and fostering communication among staff. While this discussion is not comprehensive, it is intended to start changing attitudes and approaches to security based on tactics used by ransomware attackers and cybercriminals. Every environment is different, and for a full assessment, organizations should consider engaging a cybersecurity firm that specializes in penetration testing and incident response.

Disclaimer

This document is intended to provide general guidance for organizations that are considering VMware solutions. The information contained in this document is for educational and informational purposes only. This document is not intended to provide advice and is provided "AS IS." VMware makes no claims, promises, or guarantees about the accuracy, completeness, or adequacy of the information contained herein. Organizations should engage appropriate legal, business, technical, and audit expertise within their specific organization for review of requirements and effectiveness of implementations.

Business Continuity

Organizations that assume a breach will happen tend to be the most prepared if it does. This mindset shift is crucial for combating new attacks. It is not a judgement of an organization's security practices, but a statistical reality in the field of information security. While defenders must always be at 100%, attackers only need to get lucky once. Complex and constantly changing environments, involving people and trust, combined with sophisticated attackers, make it a matter of time before someone clicks on something they shouldn't, and their desktop becomes compromised.

Everything Will Be Down

Ensure that attacks are covered in your organization's disaster recovery & business continuity planning. Plan for an "everything down" scenario. Proactively engage a security consultancy that specializes in ransomware incident response and has a record of successful ransom negotiations with attackers.

Plan to restore services and functionality without paying the attackers. Paying the attackers does not guarantee anything:

- These are criminals and will act in their own best interest despite whatever they promised you.
- Organizations often find that the decryption program for their data does not work correctly or works too slowly to be helpful. The decryption program may also install additional malware that is different than the original attack, so that your organization can be attacked again. Do not run untrusted software.
- Studies have shown that organizations that pay get attacked again, soon after the first attack, by other cybercriminal gangs. The first group of attackers will share vulnerability information with their criminal partners and they act quickly before you have a chance to protect yourself.
- Paying criminals funds their criminal activities, further exacerbating the problem. Payments to criminals may be illegal in your country, state, or municipality, and are usually not covered by insurance.

As part of your organization's business continuity plan your systems should be classified according to their business criticality. This will help you assign correct security controls to systems, assess dependencies, and prioritize work during an incident.

Ensure that contact information and roles & responsibilities documents are stored in a place that will be accessible if IT systems are offline. Some organizations, with otherwise terrific business continuity plans, have found themselves hampered because their plans were stored on systems that were inaccessible because of the outage!

Practice

It is an unfortunate truth that people who are good at incident response simply have endured many incidents. While it is impractical in most organizations to intentionally cause a disaster just to practice recovery, there are other ways to simulate an incident. IT staff expected to participate in a real situation should already be familiar with recovery processes to speed recovery and reduce stress during an incident. Many organizations use tabletop exercises to familiarize staff with incident response situations as well. Lab and test environments can also be helpful for practicing recoverability.

Asset Inventory

In a situation where details matter, and a single device with a vulnerability can open the door to a full-blown attack, it is very important to know what is running in your environment.

Configuration Management Databases

Some organizations have Configuration Management Databases (CMDBs) that serve as asset inventories. If you don't, consider starting out with a simple spreadsheet. The Center for Internet Security (CIS) also has a sample asset management spreadsheet as part of their 18 core security controls.

Asset inventories should include operating system type and version, hardware type and model, and network addresses at a minimum. Take care when including information that the asset inventory or CMDB wouldn't be authoritative for. For example, the authoritative source for DNS information for an IP address is not a CMDB, it is the Domain Name Servers themselves. If DNS information is stored in the CMDB it is likely to not be synchronized with the authoritative source. Another example would be patch and update levels for operating systems, as those are details best queried by a vulnerability scanner or other tool.

Auditing Devices

Asset inventories should list all infrastructure, including devices that are not directly visible to workloads, like network and storage switches. Tools like ping and nmap are commonly used for auditing, but you can also use ARP tables on network switches and physical network switch port data as well. Consider that a rogue device on the network may not have an IP address, and many legitimate devices will not respond to ICMP echos (pings).

PowerCLI can easily generate lists of virtual machines and their properties, as well as ESXi hosts and their properties. Carbon Black Workload, which is a terrific tool for assessing vulnerabilities in virtualized workloads, can also be used to generate asset maps and reports.

Improve Patching

Patching is the only way to remove a vulnerability from an environment, but patching processes are fundamentally broken in many organizations. Long change control processes for approving security patches are an attacker's best friend. While a Change Advisory Board is taking 90 days to approve a change, the attacker is using the vulnerability to turn a small malware infection on a desktop into a full-blown attack on your organization's infrastructure.

Use an asset inventory and the prioritization from the business continuity planning process to determine the level of availability and security needed for systems. Business critical systems, such as identity management systems, should be designed so they can be updated in minutes from the release of a patch.

If a system cannot be patched during business hours it won't survive an actual outage, either. Patching can be positioned as a small, scheduled test of the resilience of individual services and systems.

Patching ESXi

VMware vSphere and VMware Cloud have many features in them to turn infrastructure patching into a low-effort, automated task that can be done during normal business hours. vMotion moves workloads between hosts in a way that most workloads never notice, and improvements in vSphere 7 & 8 ensure that the most sensitive workloads move gracefully. Distributed Resource Scheduling (DRS) works with vMotion to automatically coordinate workload relocation during patching. vSphere Lifecycle Manager manages and remediates hosts and host configurations from the firmware all the way up, making patching something you can start in the morning and walk away from. For an even easier approach, VMware Cloud on AWS turns those responsibilities over to VMware Site Reliability Engineers to handle, freeing your staff time to work on other issues.

Patching vCenter Server

Patching vCenter Server does not bring workloads down but does affect the management interface that administrators use to interact with the environment. Systems that connect to vSphere should be set to automatically reconnect and resume operations when vCenter Server becomes available again.

People & Process

Organizations that use ITIL methodologies to classify changes should consider making all patches either a Standard change or an Emergency change. Standard changes are well-documented, repeatable changes and can be applied at any time. This approach works well for most product updates lower than "Important" and "Critical." Indeed, patching should be a process that is repeatable, reliable, and can be done with minimal disruption through staged rollouts and good system design for availability (using vMotion and DRS).

Emergency changes should be used for security vulnerabilities considered "Important" or "Critical." Often the difference between those two levels is whether the attacker needs to be authenticated or not. In environments that use external authentication sources, such as Active Directory, an attacker may already be able to authenticate. This turns an Important vulnerability into a Critical one. Assuming an attacker has already breached the corporate perimeter is an approach many organizations have started using to improve their own security processes. Assuming the attacker is already close by, and can already authenticate, changes how organizations respond to new security threats and disclosures.

Improve and Separate Your Infrastructure Authentication

From an authentication & authorization perspective there are three big realities that work against organizations when building and operating infrastructure systems:

1. Phishing attacks and credential compromise are a primary method for attackers to gain access to an organization. Attackers gain control of a user account and then use that to conduct attacks from the inside. Virtualization Admins are not immune to this, and the compromise of an administrator's account can be devastating. The attacker can deactivate other security controls, exfiltrate data, and encrypt workloads at will.
2. Centralized identity providers like Microsoft Active Directory are a major target for attackers. This is not a criticism of Active Directory; this issue exists for any centralized identity provider. In Active Directory's case, it is a target precisely because it is such a successful way to manage enterprises. Attackers know that if they can break into Active Directory they can achieve their own nefarious goals much faster.
3. Infrastructure systems often host the identity providers that they rely on for authentication and authorization, and in the case of Active Directory, often DNS, DHCP, and Certificate Management, too. Controlling these dependencies is crucial for the "everything down" scenario that should be in an organization's business continuity plans.

Dedicated Administration Accounts

IT infrastructure should be managed with separate admin accounts that are not used for regular tasks. This way, if an attacker uses stolen credentials, it will show up as failed login attempts in logs. Monitor authentication systems for both failed and successful admin login attempts and audit them regularly to detect any successful attacks. After all, if an attacker has valid credentials, there will be no failed login events.

Authentication Isolation

Authentication for infrastructure systems and devices should be isolated from general purpose authentication sources used by desktops, so that a breach does not automatically mean a compromise of the infrastructure. This can be done in a variety of ways, from local authentication on discrete infrastructure devices to a separate, purpose-built infrastructure authentication system inside the secure management perimeter that centralizes infrastructure admin logins and offers an opportunity to introduce multifactor authentication.

Avoid Authorization Groups

Attackers who compromise an identity source can often add themselves to authorization groups, and simply log into systems they should not otherwise have access to. Additionally, reliance on central identity systems means that the administrators of those systems are potentially infrastructure administrators, too, as they can add themselves to infrastructure access groups at will. Some regulatory compliance efforts, such as for PCI DSS and NIST 800-171, flag those identity management admins as "in scope" for audits and compliance actions. Organizations that do not wish their domain admins – rogue or legitimate – to be storage, firewall, vSphere, or other admins should reconsider the use of domain groups for authorization.

Most infrastructure, including vSphere, allows authorization to be done on the systems themselves, such as through the use of SSO groups. This has the advantage of no dependencies on other systems but may be harder to manage. Techniques for automation of account management can be employed, though recent attacks that made headlines remind us to protect automation systems as well.

Protect Management Interfaces

Just as you would not let anyone in your organization wander freely through your data center, you should also protect your virtualized data center from casual or malicious access. Organizations that assume that an attacker is inside their corporate perimeter also add internal controls to ensure that only specific authorized devices can access the management interfaces and systems.

Network Isolation

VMware strongly recommends that virtual infrastructure management interfaces be on a dedicated VLAN or network segment, separate from workloads and other systems. This includes ESXi, vCenter Server, NSX Manager, SDDC Manager, and other components that should have a very small user population. Access to these network segments should be restricted to only staff with infrastructure administration duties. Similarly, the VMware Cloud Console, and vCenter Server deployed in a cloud SDDC, should be restricted to administrators.

Hardware management controllers should be isolated as well, with comprehensive access controls to the network they are attached to. Hardware management controllers should not present emulated USB NICs and other interconnections to ESXi that serve as back doors for attackers.

Manage vSphere via vCenter Server

vSphere management activities should be through vCenter Server where they are subject to the Role-Based Access Control (RBAC) permission model. ESXi should be configured according to best practices where SSH is disabled and normal Lockdown Mode is enabled. If SSH access is needed for support and troubleshooting it should be enabled for the troubleshooting operation and then returned to its stopped and off state.

Multifactor Authentication

The nature of ESXi limits the scope of authentication, but vCenter Server is more flexible. Multifactor authentication into vCenter Server can be introduced through either Smart Cards, LDAP, or vSphere Identity Federation. Third-party identity services have LDAP proxies that can be configured as identity sources for vCenter Server, prompting users when a login attempt is made. A more modern approach is available with vSphere Identity Federation, where direct connections to Microsoft Active Directory Federation Services (ADFS) in vSphere 7+, and Okta services in vSphere 8, can be made. Many authentication and identity services plug into ADFS.

Virtual Machine Consoles

The vSphere Client proxies virtual machine console connections through vCenter Server if others need access. That access can be controlled using the RBAC permission model in vCenter Server. There is no longer a need to open ESXi management interfaces to workload administrators if they do need direct console access.

Ecosystem Connectivity

Large-scale, newsworthy attacks via management systems remind us that any other system or tool that has rights to change the virtual environment must be considered an administration method and protected accordingly. For example, vRealize Operations Manager can have connectors configured to allow the automatic rightsizing of a virtual machine. This is a powerful and easy-to-use management feature, but access to it should be considered carefully. If your organization doesn't continuously right-size workloads it may be better to leave the connectors configured as read-only.

Mythical Single Pane of Glass

Plugins and integrations between infrastructure systems sell the dream of a single pane of glass to manage an infrastructure. On paper it seems like a great idea, but in practice these approaches add friction to patches and upgrades, force administrators to qualify many major infrastructure components together, and even offer attackers convenience in deleting backups and disabling security controls.

Just as it is recommended to place separate IT infrastructure components on their own network segments, manage each through their own interfaces.

Separate Workload Administration Access

Virtual infrastructure does not exist to serve itself; it is there to enable workloads. However, every person you authorize into vCenter Server & Cloud Console is a path to a potential compromise, either through phishing and compromised credentials or insider threats.

Use SSH and RDP Directly

Workload administrators should access virtual machines through the workload's own network interface, via SSH or RDP. This simplifies access control rules and auditing and makes workload management subject to NSX intrusion detection and other monitoring.

Workloads do not typically need to be managed from the console of the virtual machine. Administrators who need Microsoft Windows console access can connect with the "/console" switch for the Remote Desktop client.

Ensure Backup & Restore Functionality

Paying a ransom does not guarantee a good outcome. Decrypters do not work, work very slowly, or may contain additional backdoors and malware. A way to avoid paying a ransom while ensuring recoverability is to have a reliable and secure backup solution as a last line of defense. Backup vendors have best practices for implementing their own software, but we will discuss some ideas and considerations here in the context of ransomware.

System Volumes vs. Data Volumes

The time between a breach and a full-blown ransomware attack can be quite long. Studies have shown that the average time between a breach and containment of the breach can be 300 or more days. Attackers are patient and will work to ensure that you must pay the ransom. This means that deleting or corrupting your backups is a high priority. Over a period that long it is likely your backup systems will capture the results of the attack. Restoring a backup may also restore infected systems, and/or restore systems to a vulnerable state. Consider how you would recover systems in a questionable state, as well as how you would assess the reliability of such systems.

Consider changes to how workloads are configured. For instance, even if the operating system and application files are corrupted by malware it may be possible to remount or reattach the data to a fresh installation of an application. Database

servers are a good example of this, where the data files and the application are held separately. If separate virtual disks are used this process is easier.

Separation & Isolation

Keep your backup systems separate from corporate identity systems. Some of VMware's backup vendor partners strongly recommend never joining your backup systems to a central identity provider. Do not join your backup systems to a separate infrastructure authentication solution, either. If the infrastructure is breached you don't want attackers having access to the backups, losing your last line of defense and your chances of recovery.

Backup systems should be an island, extremely isolated, managed by few staff, monitored heavily, but also protected against outside influence from centralized monitoring and management systems. Yes, this flies in the face of most "one size fits all" corporate policy.

Air Gaps & Immutable Backups

Use immutable backups and/or backup "air gaps" where possible. Immutable backups are backups that cannot be altered. This has typically been implemented with specialized backup media like "WORM" – Write Once, Read Many – devices. Most current backup solutions also have software-based approaches, but care should be taken to ensure that they are truly immutable. Air gaps are where backups are intentionally taken offline, and often offsite as well, to protect against threats like fire, flood, burglary, and such. There are several strategies for this, each dependent on the software and the implementation.

Note that there may be legal implications to changes in backup strategies, especially considering GDPR and data retention issues. Many larger organizations have specific data retention policies, not just to ensure that data is retained for certain periods of time, but also that it is destroyed when that time is over.

More Information

VMware maintains a Ransomware Resource Center with links to other trusted sites, documents, and more. Visit it at:

https://core.vmware.com/ransomware

https://core.vmware.com/practical-ideas-ransomware-resilience

About the Author

Bob Plankers works in the Cloud Platforms group at VMware, focusing on all forms of security and compliance from VMware Cloud to on-premises vSphere. Prior to joining VMware, he spent more than two decades leading cross-organizational teams that designed, built, and operated reliable, secure, and compliance-oriented IT infrastructures worldwide, focusing not just on technological solutions, but also the people and process aspects, too.