WHITEPAPER

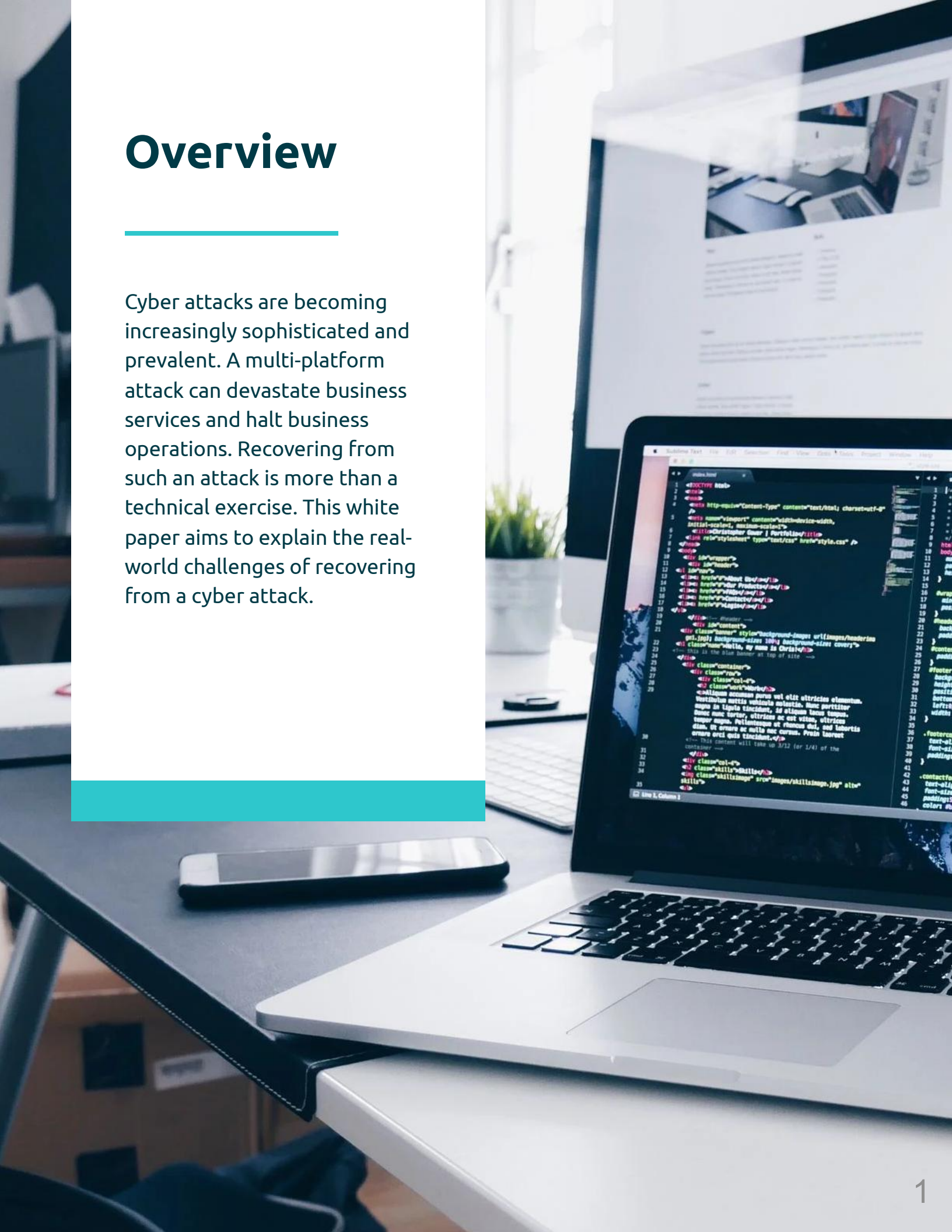
Cyber Recovery

A Complete Guide

VAULT SENTINEL™

Overview

Cyber attacks are becoming increasingly sophisticated and prevalent. A multi-platform attack can devastate business services and halt business operations. Recovering from such an attack is more than a technical exercise. This white paper aims to explain the real-world challenges of recovering from a cyber attack.



Cyber attacks

While ransomware dominates headlines, it is not the only mechanism used to attack an organization. Wiperware, for instance, can have a devastating effect by wiping everything from a device, sometimes rendering it physically unusable. Windows is the most common platform for ransomware attacks, but malicious actors are increasingly developing multi-platform ransomware that can compromise an organization's platform resiliency.

Real World Problem Example

June 2017: A.P. Moller-Maersk experienced a nation-state cyber attack.

- At the time, this type of attack was unheard of but has since become more common. The attack was based on wiperware.
- Maersk's cyber hygiene was above average.
- The attack was software supply chain-based and utilized three zero-day exploits.
- All Windows devices within the company were destroyed, approximately 50,000 end-user devices and 6,000 servers, including critical infrastructure such as DHCP, AD, firewalls, switches, and backup servers.
- All network-stored design or recovery documents were also destroyed.
- Maersk was not the intended target.

Adam Banks, Former Group CTIO A.P. Moller Maersk

Cyber attacks

Real World Solution Challenges

To effectively recover business services, the first step was to rebuild the network and technical services, such as Active Directory and backup servers.

- After building the network and foundational infrastructure, the next step was to attempt recovery from backups. Unfortunately, the new backup servers couldn't re-index the SANs, rendering the raw backups unusable.

- Consequently, each business service had to be manually rebuilt and reinstalled, followed by the reconstruction and population of necessary data from surviving mainframe data.

- Business services had to be restored on a server-by-server basis using knowledge and experimentation due to the destruction of all documentation.

This effort required 10,000 technical resources and took 10 days to get critical services live, 30 days to get all services live, and 45 days to restore full capacity. Supporting systems such as development and testing were not fully recovered for 100 days.

The incident highlighted a significant gap in the market: the need for a solution that is effectively 'offline' and capable of recovering services rather than servers in a vanilla environment, with no pre-configuration required

Adam Banks, Former Group CTIO A.P. Moller Maersk

Cyber attacks

An excerpt from Mimecast

Criminals are nothing but opportunists, and recent months have seen numerous gangs develop multi-platform malware. Early adopters include:

1. BlackCat malware, written in Rust, which emerged in late 2021. Microsoft notes that the RaaS group has mounted “successful attacks **against Windows and Linux devices** and VMWare instances''. Tools analysing Rust are not as sophisticated as those designed for C, making analysis harder.
2. Blackbasta, which may be linked to Conti or **Russian group FIN7**, is another RaaS group with multiple attacks under its belt; its malware has Windows and Linux versions.
3. Luna, **which operates on Windows, Linux and ESXi**, and is also written in Rust. Its operators claim only to work with Russian-speaking partners.
4. A new variant of **RansomExx ransomware**, RansomExx2, has been written in Rust, with a Linux version already operational and a Windows equivalent believed to be in development.
5. RedAlert, which emerged in **mid 2022**, strikes at both Windows and Linux VMWare ESXi servers.
6. Deadbolt **attacks network-attached storage devices**. It is written in Golang but uses an HTML ransom note and a Bash script for decryption.

What to protect?

The default position is to protect the Windows estate, as this is predominantly the major platform in an organization. However, malicious actors have been developing multi-platform attacks, with Linux ransomware first seen in 2019 and VMware ESXi attacks first seen in 2023.

National security agencies worldwide have observed attempted attacks on network and storage layers. The principle in designing a cyber recovery strategy must be that anything connected to a network is vulnerable, including:

- Telephony, network devices (switches, routers, files, load balancers)
- Foundational infrastructure services (DNS, DHCP)
- Security devices (privileged access applications and appliances)
- Databases, mid-range platforms, virtualization platforms (VMware, Hyper-V)
- Desktops

All of these have vulnerabilities that state-sponsored groups and criminal organizations seek to exploit. Documentation could be lost, and incident and crisis management applications may be unavailable, as well as communications.

This means that a recovery plan must assume that everything has been compromised and is unavailable.

We have disaster recovery so we're ok

Most organizations have robust disaster recovery plans and solutions. However, recovering from a cyber attack is a different scenario. Disaster recovery is for known events.

Organizations have multiple data centers to provide resilience from the loss of a physical data center. Each data center has multiple machine halls and technical resiliency such as clustering and load balancers. A cyber attack, however, will propagate as far as it can across the entire network, nullifying these resiliency capabilities.

This replication between resilient data centers only spreads the attack, compromising resiliency measures.

Cyber recovery, on the other hand, is a different scenario, attempting to plan for an unknown event. It is unknown what platform, location, technology, or devices will be compromised during an attack, and the degree of propagation is also unknown.

The cyber recovery plan must operate on the principle that after the attack, all the usual resiliency measures have been compromised, and everything is lost or unavailable.

What to recover?

Minimum Viable Business or Minimum Viabale Company

The entire production estate is not required immediately after an attack. The priority must be to recover a minimum service for customers.

An MVB (Minimum Viable Business) comprises the minimum components required to run a business service, typically 20% or so of an organization's production estate. The focus must be on core business services rather than restoring the entire production environment.

Restoring technology alone does not restore business services. The organization must clearly understand how each business service should be recovered.

This minimum viable business must include everything needed to allow the organization to be operational at a minimum level. Recovering technology in a data center or cloud is only part of the story. If call centers, head offices, regional offices, branches, manufacturing plants, and internet channels cannot function, then the business cannot operate.

All of these must be included in a recovery plan.

We've got backups, so we're ok

Backups

In the majority of cyber attacks, backups remain intact since all storage platforms have immutability functionality. However, access to these backups is often lost during an attack.

Utilizing backups requires a stack of foundational infrastructure, including the vendor's backup and restore infrastructure, backup library, and backup catalog, as well as foundational infrastructure services such as Active Directory, DNS, and DHCP. Devices are also needed to access all of this, most or all of which may have been compromised in the attack. This presents a key issue: how to access backups?

Backups generally cover server instances and databases but often do not include the network and storage layers, telephony, some underlying technologies, access control systems, manufacturing control systems, or industrial control systems.

Recovering from a cyber attack is essentially an enterprise bare metal recovery. Before backups can be used for restoration, each server must have its operating system installed, and virtualization platforms must have their underlying hypervisor installed and configured.

It's ok, we replicate everything

Replication is a common technique for resiliency, but replicating production to another site or device only replicates the compromise.

Applications can be installed to detect ransomware attacks by identifying when a device starts being encrypted. However, this means the ransomware has been on the device for a while, sometimes weeks before detonation.

Replicating production simply copies the ransomware, propagating it and undermining resiliency. Even with snapshots of the replicated target, multiple copies of each device are required to ensure recovery to a clean state.

So what to do?

Cyber attacks are designed to propagate widely and compromise as many devices as possible. On a typical organization's estate, most devices are vulnerable.

In the infamous NotPetya attack on Maersk in 2017, 55,000 devices were compromised in six minutes, affecting Active Directory, DNS, DHCP, email, mobile phones, build servers, backup servers, application servers, databases, documentation, incident, and crisis management systems.

If it's connected to production, it's vulnerable

To ensure a cyber recovery environment is secure, it must be completely offline, eliminating the threat of compromise. Many vendors use the term "Air Gap," but what does that really mean?

A true air gap is a physical separation between two devices. In legacy systems, this meant storage in a separate room or building. Modern systems often use the term despite having connections between devices, relying on disabling and enabling ports instead of maintaining a physical air gap.

Principles of a Cyber Vault

There are principles and design considerations for a cyber vault

1. The Vault must be isolated from production
2. The credentials for the Vault must not be shared outside of the Vault.
3. Ingress of an object to the Vault must be a pull from the Vault and Egress of an object from the Vault must be a push from the Vault.
4. Objects in the vault must be immutable
5. The Vault must have the ability to have multiple logical Vaults each with their own credentials
6. The Vault must have the ability to encrypt the objects in each logical Vault using different keys for each Vault
7. There must not be any TCP connection open to the Vault from other environments
8. There must be no persistent and accessible compute in the Vault
9. Objects in the Vault must be Self Inflating with the ability to recover without any intermediary applications.
10. The Vault must have the ability to run analytics against the objects to check integrity without starting the object.
11. The Vault must have a Management Application that tracks objects in each of the technical Vaults
12. The Management Application must be isolated from production and not have any application interface to any other applications and the only connectivity to the Management Application must be the used who have to use Multi Factor Authentication.
13. The Vault must have the capability to map objects to other objects and to services and services to processes
14. The Vault must have the ability to recover objects, services and processes
15. Any changes or recovery actions in the Management Application must have a two person authorization

Offline

Storing copies offline can be achieved in various ways, from building a physical environment in a data center to storing data in the cloud.

Securing another copy of devices and data in another part of a physical data center presents many recovery challenges. The physical infrastructure must be built at great expense and must be as secure, if not more secure, than production. Physical offline environments are expensive, difficult to manage in terms of capacity and performance, and costly to maintain. Upgrades to infrastructure and operating systems need to align with production.

Creating isolated environments in this offline setup proves very difficult.

Business services span not only physical data centers but also countries and continents. Recovering a major business service could require recovery in multiple global locations.

Building an environment in the cloud would be completely offline but would need to be separate from the organization's production environments in the cloud. A connection from production to the offline cloud environment is necessary, raising the issue of securing it against attacks from on-premise production.

Managing an offline repository poses challenges. Allowing access to an offline environment inherently introduces potential access points for bad actors.

Management

In the aftermath of a cyber attack, demonstrating control and management over the crisis is crucial.

There are two key parts to recovery: the technical recovery and the management of the crisis. The technical recovery involves taking copies of devices and databases in a format that can be used without foundational infrastructure and storing them securely.

Understanding what teams to mobilize, what checks to run on the production environment, identifying the attack and its entry point, plugging the compromise, and preparing a recovery environment are essential steps.

Then, identifying the devices, configurations, databases, and applications that make up the multiple business services that need to be recovered, and understanding how long recovery will take, are critical.

All this information must be immediately available to the crisis management team to recover business services in the shortest possible time while demonstrating effective management of the situation.

The situation an organization finds itself in during a cyber attack is unique. They must perform an enterprise bare metal recovery on the entire estate, recovering thousands of devices and bringing components back in the correct sequence to bring services online.

Tracking the recovery of these devices and services across multiple locations and countries is a complex and difficult task.

How do I know what I'm recovering is valid?

The bottom line is, you don't, but this risk can be minimized. Anti-ransomware, malware, and virus products should be run against copies of devices and data in the offline environment. However, since the attack could be a sophisticated zero-day vulnerability or a multi-platform attack, there is no guarantee of recovering a clean copy.

The objective in the offline environment is to mitigate as much risk as possible. Running these products against copies in the offline environment during ingress, and possibly during egress, although this would slow recovery, provides added reassurance that the copies being recovered are clean. It is key to run different products in the offline environment from those run in production.

The ideal situation involves running intelligent products against copies in the offline environment for servers, device configurations, and databases. Identifying an infected or compromised copy is not sufficient. The organization must know what business services the compromised device affects.

Many products should be run in production to identify and alert the start of the encryption process of ransomware. However, it is crucial to understand that this is not the point of infection, which could have occurred many weeks or months earlier. Restoring such copies without remediating the infection will only reinfect the entire recovered production environment.

Identifying an infected or compromised copy is not sufficient. The organsaition will need to know what a business services the compromised device affects.

What about the teams?

An organization's teams are dedicated and understand the environment. The challenge is when was the last time they recovered anything from bare metal? Recovering the minimum service and then the entire infrastructure will likely require hundreds or thousands of additional engineers.

So how will the team know what to do? Although these teams will be highly skilled, they will not have faced such a scenario before and will need as much assistance as possible. The additional engineers will not know the customer's environment and will need to be guided through the tasks.

They need to know what devices to recover, where to recover them from, how to recover them, the tasks needed for recovery, and the sequence of tasks. Restoring technology alone does not restore business services, so service and application specialists need to know in what sequence to bring up devices and services.

Recovery could span multiple countries, with language barriers and time zones to overcome. With a multi-platform attack, multiple technical disciplines are needed, with the recovery sequence essential to a successful recovery.

Where do I recover too?

Recovering back to the compromised production environment is most common, but there are options. Restoring to the compromised production environment can only begin after the compromise has been identified, the environment cleaned, and the organization is confident that restoration will not reinfect production.

In a complex multi-platform attack, this could take days. Other options include building a recovery environment on-premise, but this can be very expensive and restricts the organization to a single data center. Providing similar functionality in multiple data centers and countries could be prohibitively expensive.

Restoring to the cloud is an option, but only if this has been designed beforehand. Simply starting copies of servers in the cloud will not restore service. The security and access model will need to be redefined to operate in the cloud.

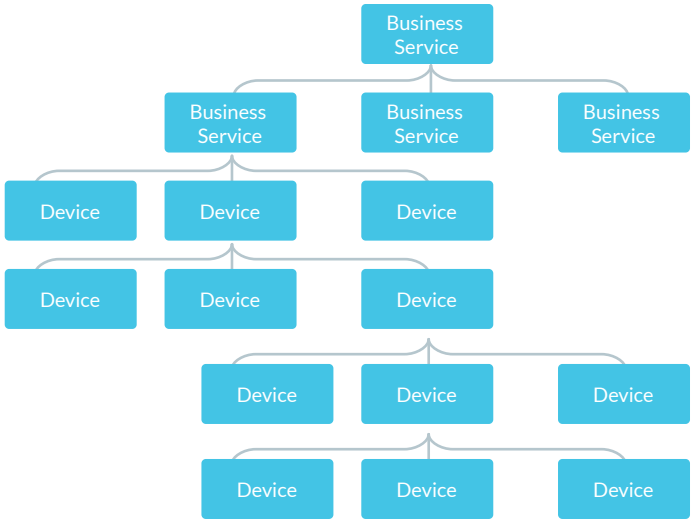
When losing the desktop estate, starting thousands of desktop instances in the cloud might seem attractive, but a device is needed to access these virtual machines, and those devices may have been compromised.

How do I know I can recover and how long will it take?

This is a key question that an organization's board, regulators, stakeholders, and shareholders will ask. Any recovery solution must be tested regularly.

The offline environment must be kept evergreen to keep the RPO and RTO to a minimum. Rehearsals are key to understanding if individual and multiple devices can be restored technically and to understand each step in the recovery of a device and each step in each device of the recovery of a business service.

These rehearsals should validate not only the technical aspects of the recovery but also how long it takes to recover business services. Understanding what to recover is extremely difficult during the chaos of a recovery. Spending the time to understand the intricacies of a service hierarchy is key.



Business services

Restoring technology alone does not restore business services. After a cyber attack, it is effectively an enterprise bare metal recovery, as none of the devices can be trusted. Once the technical recovery is complete, business services need to be brought online and started in sequence.

In addition to the many engineers working on the recovery, application and service specialists will need to know the sequence to bring the entire business service hierarchy online, testing each component as they go.

Detailed procedures are required to bring services back online in the correct sequence with the correct checkpoints.

The channels that serve the minimum viable services need to be brought back online with priority to provide customers with a minimum service.

Reconciliation

Enterprise business services usually have many data stores across different data platforms. These data stores need to be kept aligned to the transaction level for the correct functioning of services. Post-recovery, there is a good chance that these data stores might not be aligned. It is crucial that the crisis management team has access to information regarding the alignment to make business decisions on how to proceed. The decision could be between the time it would take to correct the alignment, which could cost significant money, or bringing services back online in a state that the organization knows and understands, and working on reconciliation afterward.

Regulatory requirements will need to be met with sufficient evidence.

It is key to have the functionality to provide a clear position on data reconciliation before bringing services back online.

Conclusion

Recovering from a sophisticated and catastrophic cyber attack is not a simple technical restore exercise. The complexity of national and global organizations makes the recovery process intricate, involving the recovery of individual devices and coordinating many engineers, application specialists, and service specialists.

As with everything, planning is key, with regular rehearsals and red team/blue team events necessary. Many vendors focus on the technical aspects, but this is only part of the challenge an organization will face after a complex, sophisticated multi-platform cyber attack. An organization should have as many tools and processes completely offline, with no connection to production, so they have everything needed to recover services.

Developing a robust and successful cyber recovery plan involves viewing the problem as if everything is lost and nothing is accessible. Only then can you be confident that you have a plan and solution that allows the organization to recover in a timely manner.

Authors

Andrew Kirkby

Andrew Kirkby is a cyber recovery specialist having led cyber recovery programs for a number of financial institutions.

Adam Banks

Adam is the former Group CTIO of A.P. Moller Maersk who was in the role during the infamous 2017 cyber attack.
Adam is seen as a expert in the cyber recovery sector and regularly presents at cyber recovery events.

VAULT SENTINEL