Technical white paper

Understanding real-time encryption detection with HPE Zerto Software

Contents

Introduction

Ransomware continues to wreak havoc on organizations of all sizes, often forcing a lose-lose decision between paying a ransom or losing untenable amounts of data. The ability to detect ransomware quickly and take action mid-attack remains a key part of any multi-faceted defense-in-depth security strategy.

Hewlett Packard Enterprise uses an innovative approach to ransomware detection based on a real-time encryption analyzer that can give one of the earliest warnings that the detonation phase of a ransomware attack has started. The encryption analyzer was co-developed with HPE in 2021, extensively tested and improved before becoming generally available in API-only mode in 2022 with HPE Zerto Software 9.7, and then added to the HPE Zerto GUI and fully launched in 2023 with HPE Zerto 10.0.

This white paper describes how the encryption analyzer works, how it compares to alternatives and where it outperforms them, and provides recommendations for using HPE Zerto's inline detection features in production environments.

Options for detecting ransomware

Ransomware detection has historically relied on three major approaches spanning both direct and indirect methods. The first approach is signature-based and is a direct method that uses malware definitions to scan for known variants. It has the advantage of generating relatively few false positives—since this approach is not based on inference, confidence levels can be high when a variant is detected. The downside is that a library of ransomware definitions must be created and kept up to date. As malware evolves, it can be challenging to keep up; by very definition, this approach is always playing catch-up to threat actors. This is especially true with "polymorphic ransomware," such as VirLock, that changes shape or how it presents itself—a moving target for more rigid detection mechanisms.

A second common approach, also categorized as direct, uses "honeypots" to lure attackers into unexpectedly revealing themselves. By setting traps throughout IT environments, ransomware can be detected when it's detected by those traps. Rather than infecting or encrypting a real server, the ransomware has instead tripped an alarm so that immediate action can be taken. Unfortunately, ransomware variants are not always so easily fooled and can sidestep the limited frame of focus inherent to honeypots. This approach also requires keeping up with attack strategies in a way similar to the first method. If the honeypots are not laid out adequately or extensively enough, they may not be enough to catch/trap sophisticated or widespread attacks.

The third approach, while indirect, is among the most common. This approach is sometimes termed "signatureless" and is anomaly- or pattern-based. It relies on any number of signals—often called indicators of compromise (IOCs)—that may mean malware is lurking or an attack is underway. The signals used can greatly vary depending on the vendor and solution. Examples include:

– Suspicious elevation of privileges on a user account
– Unexpected logins from an unusual geography
– Sudden surge in CPU usage
– Deviations in I/O patterns or network traffic
– Anomalous encryption events

This latter approach is the one HPE Zerto is using: assessing data patterns and analyzing entropy to detect unusual encryption that may indicate the locking phase of a ransomware attack has begun.

Comparison to other anomaly-based detection

The breadth and variety of IOCs makes anomaly detection a popular approach. Many of the vendors developing these solutions are purpose-built data security or cybersecurity companies who can detect ransomware quickly and with high degrees of confidence. These solutions are a key part of any well-rounded security stack.

Unfortunately, these strong solutions are decoupled from recovery, a key phase in mitigating and responding to ransomware. Restoring good, clean data is essential to getting back to business as usual.

The result has been a move from data protection vendors who specialize in recovery to focus on detection as well. The combination of a traditional security practice with a traditional protection practice enables enterprises to realize the best of both domains.

However, HPE Zerto's approach to real-time encryption detection (RED) is unique among data protection vendors. This is because the HPE Zerto Software offers encryption analyzer is built on a foundation of continuous data protection (CDP), a technology widely considered, even by our competitors, to be the best way to reduce data loss and downtime. HPE Zerto's proprietary CDP engine has been battle-tested and proven at scale, resulting in thousands of customers achieving RPOs of seconds whether protecting 70 or 700 VMs—or even 7,000 VMs simultaneously.

CDP is essential to recovery, but this foundation of block-based (not file-based), hypervisor-level replication unlocks five primary benefits for encryption detection as well, including:

1. **Real-time:** HPE Zerto's always-on, near-synchronous replication engine unlocks the ability to analyze data on a virtually real-time basis with high degrees of granularity. Users no longer need to wait until a backup is run—or worse, wait until the backup window ends—to start scanning. Unfortunately, a periodic approach also means the dataset being analyzed will always be dramatically larger than real-time block inspection—increasing the detection time even further. Instead, HPE Zerto enables IT and SecOps teams to take action mid-attack as malicious encryption is happening. The result is dramatically less data loss and faster time to begin recovery and remediation and get back to business as usual. Refer to the case study callout below to learn more about how real-time detection could play out in an average attack.
2. **Agnostic:** HPE Zerto is completely indifferent to the type of file being encrypted. Detectors that aren't block-based must often make assumptions about the type of data to look for or the type of encoding. Some alternative scanning solutions even require specific file formats, such as their proprietary backup format, to function. HPE Zerto's agnostic insights are an advantage here: the encryption analyzer is inspecting, assessing, and dynamically adjusting as the I/O comes in regardless of what the data is, how it's encoded, how large it is, or where it's coming from.
3. **Relative:** Another key piece to real-time encryption detection is that it's constantly adjusting to the ever-changing conditions of the environment. A continuous, moving training period helps HPE Zerto learn what normal write patterns are to home in on what may be anomalous and what may be expected encryption. The encryption analyzer is dynamically adaptive—it does not make assumptions about the digital estate and does not need manual updating based on new or different environmental variables. Furthermore, if a false positive does occur and IT has validated that all is well, the training period restarts to ensure it's as up to date as possible and doesn't rely on data models that don't reflect current realities.

"Zerto will play a key role in delivering effective data protection and disaster recovery in an environment of increased cyber threats. Its real-time ransomware detection puts us in a much stronger position to both identify and mitigate ransomware attacks. This gives us confidence that we can proactively meet the risks presented by ransomware and achieve the business goals we have in place."

– **Network admin**, manufacturing customer

4. **Agentless:** As with HPE Zerto's real-time replication, RED does not use any agents on protected VMs. This has two benefits during normal operations: first, it reduces management and admin overhead since agents do not have to be installed, configured, updated, or otherwise maintained; second, it improves performance by ensuring HPE Zerto consumes no overhead on each protected server. More critically, an agentless solution delivers a key advantage during an attack: without an agent, there is no HPE Zerto component/service on a VM that can be disabled or hijacked by threat actors. Ransomware, such as Conti and others, commonly scans infected machines to find security and backup agents and disable them.
5. **Lightweight:** The block-level, real-time nature of the encryption analyzer means there are very light infrastructure requirements. There are no additional components to deploy and configure—HPE Zerto uses the existing virtual replication appliances (VRAs) and virtual managers that are required for normal CDP. The burden of the inspection and analysis, as detailed out later in this paper, fall to the VRA, but the overhead here is negligible: less than 3 MB of memory needed per volume, with a total per-VRA maximum of 10% of memory and more typical usage in the 3-5% range. Encryption analyses also do not interfere with normal replication. Under high write rate or infrastructure bottleneck, if an I/O cannot be both analyzed and replicated, it will be dropped from the analysis pool to prioritize replication. One of the key results is that repeated performance tests have shown no statistically significant drop in the maximum I/O that a VRA can handle.
6. **API-first:** Lastly, although not specific to CDP-based detection, the encryption analyzer carries a final differentiator. The analyses and metrics that HPE Zerto uses are also exposed via our open REST, Swagger-based API to enable integration with an organization's larger security stack. Many alternatives to HPE Zerto lock their detection inside a black box so that all reporting and analysis is unavailable outside of the solution. Unfortunately, this makes IT teams wholly reliant on the vendor rather than enabling each business to customize the solution, integrate with others, or apply existing workflows (for example, in-house AI/ML tooling) to the detection dataset. More on HPE Zerto's detection API is covered later on this paper.

Case study: Why real-time detection matters

Real-time encryption detection can help minimize the scale or blast radius of ransomware's impact phase. Some businesses may misunderstand how much and how fast ransomware typically encrypts, but the numbers tell the story.

An internal analysis of 116 globally diverse ransomware attacks spanning 43 different ransomware variants uncovered that a median dataset of 183.5 GB was compromised. A separate study from Splunk, An Empirically Comparative Analysis of Ransomware Binaries, found the average ransomware can encrypt a gigabyte of data in 47.7 seconds. This means in a typical attack, the full encryption detonation would be estimated to take 2 hours and 26 minutes.

Unfortunately, waiting for a nightly backup to run and then scanning those copies means the average ransomware has already finished encrypting the entire dataset 12 or 24 hours beforehand—in a race against time, the attackers are miles down the road before there'd be any alerts that there's an issue.

HPE Zerto, on the other hand, can detect and alert within seconds. If detected within 15 seconds, for example, not only is the average ransomware not finished encrypting, it would've only managed to encrypt about 300 MB out of the 183.5 GB—about a 99.8% savings in amount of locked data.

The sooner you can detect, the sooner you can take action: that's why real-time encryption detection has real-world ramifications.

How it works

The encryption analyzer works in three main phases: collection, inspection, and reaction (CIR). Together the CIR process makes up the combined real-time encryption detection in HPE Zerto.

Collection

In the collection phase, virtual replication appliances copy every I/O to an in-memory buffer. When the buffer has enough data to perform meaningful analysis, then the second phase, inspection, begins. The collection phase includes a moving training period for each volume of each VM being protected with HPE Zerto—these training

periods are independent of each other so that if one is reset (see below), others are not affected. Training for a volume is paused if there is no replication and resumed when there is traffic again.

To avoid performance impact, data writes are collected prior to being compressed as well as prior to replication since it is the source VRA, not target VRA, performing the encryption detection using the CIR process. Lastly, as noted earlier, keep in mind that replication performance will never be impeded because of the collection phase. If there are not sufficient resources to both collect a sample for analysis and perform continuous replication to the target site, HPE Zerto will always prioritize replication and drop I/Os from the collection buffer as needed. HPE Zerto will also not assess any self-generated I/O and only collect replication I/Os from protected VMs.

"Our IT systems are at the heart of our services, so protecting our systems and the data of our individuals with Zerto is vital. We recently tested Zerto 10 and believe that real-time ransomware detection with help us deliver even greater protection from cyber threats."

– **IT director**, healthcare customer

Inspection

The inspection phase takes place on the virtual manager and uses two proprietary, patent-pending algorithms in concert together to analyze the sample buffer previously collected. The two algorithms can be nicknamed RED-C and RED-E for ease of reference. RED-C uses a cumulative sum (CuSum) test for randomness, while the RED-E algorithm assesses the entropy of the sample dataset. Both cycle through the collected data simultaneously in tranches of 21 MB.

By tackling encryption analysis from multiple perspectives, HPE Zerto can increase the odds of accurate detection and reduce false positives. RED-E, in particular, is unique in the data protection space because it measures relative entropy by continuously adapting to the incoming data patterns. This means RED-E uses dynamic thresholds that are independent of the type of data being written, whether text, images, binaries, and so on. This is critical since Base64 encoding is a popular method ransomware, such as the Big Head variant, can use to spoof low entropy. HPE Zerto's innovations with RED-E combat this by detecting regardless of whether a binary-to-text encoding scheme has been used by the malware. The inspection process is also effective regardless of the encryption technique, whether AES, ChaCha20, Salsa20, and so on.

From the two algorithms together, HPE Zerto assesses both the level of confidence that an encryption event is expected or anomalous, as well as the level of severity. To ensure production-readiness, these RED algorithms have also been successfully stress-tested against multiple ransomware variants, including ones like Thanos, Cerber, DarkSide, and more.

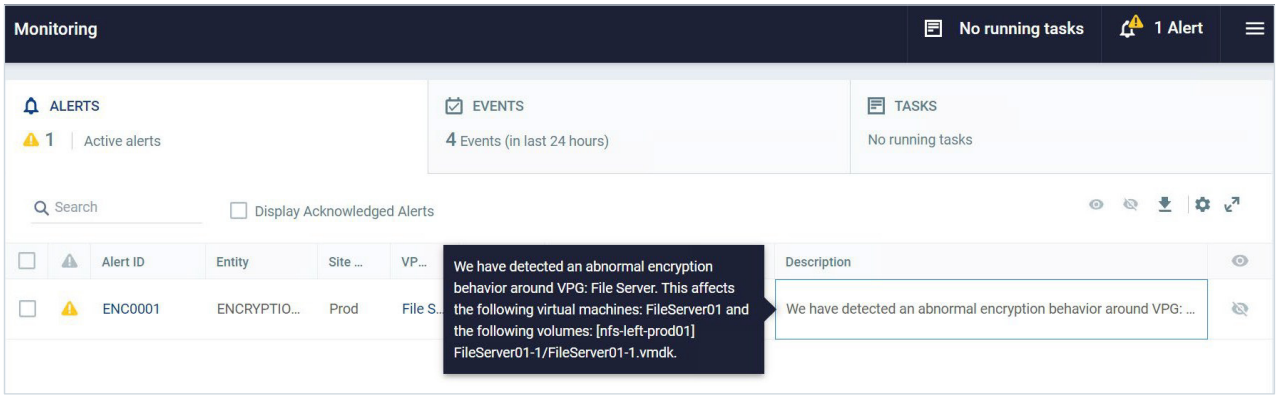
Reaction

The reaction phase is the culmination of the CIR process. It has three main parts: alerting, tagging, and user response.

False positives can become frustrating for IT admins and potentially lower the efficacy of the tool: the more false positives there are, the less trust the administrators have in it, and the more likely a real alert will go unnoticed or get dismissed as just noise. HPE Zerto understands that a ransomware attack is likely going to impact multiple volumes across multiple VMs, and potentially not simultaneously. The random patternation of a ransomware attack is what makes it so hard to detect. HPE Zerto scores its encryption detection based across a site score and not an individual VM or volume, which ensures false positives are lower and accuracy is higher. The site score collects and analyzes all the data across a whole site with regards to encryption detection and will only generate alerts when it feels there is a chance of ransomware attack across multiple vectors that could affect multiple disks, multiple VMs, or a combination of both.

This intelligence will lower alerts due to false positives and breed confidence into the solution allowing fast concise action to be taken when alerts are triggered as the trust level is likely to be much higher.

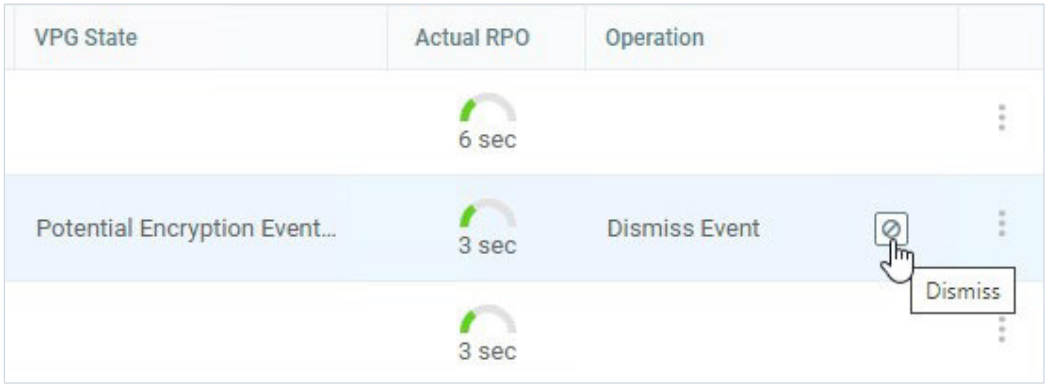
1. HPE Zerto generates alert ENC0001 if the overall site score reaches its threshold. This can be altered with the help of HPE Zerto support to be finely tuned to your environment's uniqueness. The alert is available via GUI and API and includes which volume, VM, and virtual protection groups (VPGs) are affected.



2. Concurrently with alerting, HPE Zerto sets all VPGs' impacted State to "Potential Encryption Event" and tags journal checkpoints in two places: both the time the encryption was detected and the clean checkpoint to allow for a more confident recovery

3. Lastly, user response is required to either validate or invalidate the detection event. By design, HPE Zerto does not automatically take action on a VPG in response to detection; for example, HPE Zerto will not automatically pause replication, change journal hard limit, disconnect NICs, and more. Given the chances of a false positive (even if low) and the complex, multi-faceted nature of a ransomware attack, automatic intervention would be too disruptive for HPE Zerto to execute without user input first. However, these containment and remediation steps can be scripted if so desired and thus be more tailored to each organization's own security processes rather than be vendor-imposed.

   Use the HPE Zerto API or the VPGs pane in the virtual manager GUI to clear the ENC0001 encryption alert; dismissal is done on a per-VPG basis, not for the overall site. Dismissing an alert removes the journal tags, resets the training period, and clears the I/O collection buffer. It is recommended to clear the alerts only after they have been correlated with other solutions in the security stack, such as tools for observability, network monitoring, and so on.



The last part of the reaction phase is always user-initiated: recovery and restoration of encrypted files, folders, and VMs. HPE Zerto has three key recovery options for these scenarios: file restore, VM restore, or full failover of one or more VMs or applications. Some details of each are below:

1. **File-level restore:** Give the user ability to browse individual files and folders, be selective about what data they want to restore and then recover that data directly back into the VMs where it is situated, all without any agents to manage or control.
2. **Instant VM restore:** Instantly recover individual VMs directly back onto production grade storage and compute—no staging areas or mounting onto appliances, instant VM restore from HPE Zerto gives the user confidence that they can bring their VMs back onto the same high performance infrastructure that their applications need without compromise or secondary actions such as vMotions or lengthy migrations.
3. **Full failover:** Unlock true mass data recovery, failover a subset of an applications, whole apps, multiple applications and even whole sites with just four clicks. Failovers are fully automated and orchestrated, ensuring all VMs are back online on the production-grade computer with the correct tier of storage, connected to the right networks and configured with a new IP address, if needed. All of this happens on completely disparate sites, including public cloud, with only minutes of downtime.

Using the detection API

HPE Zerto's API-first development approach applies to the real-time encryption detection features as well. By not locking away the detection analyses, organizations can enhance their defense-in-depth capabilities by integrating HPE Zerto with existing cybersecurity solutions. Since we're not using a closed black box where customers have no access to detection data, HPE Zerto's unique block-level detection can be combined with other solutions such as EDRs, SIEMs, SOARs, or AI/ML toolsets.

An example of what's possible with HPE Zerto's API is freely available on GitHub. Known as the HPE Zerto Resilience Observation Console (zROC), this open-source project uses Prometheus and Grafana to visualize a number of HPE Zerto metrics, including for encryption detection.

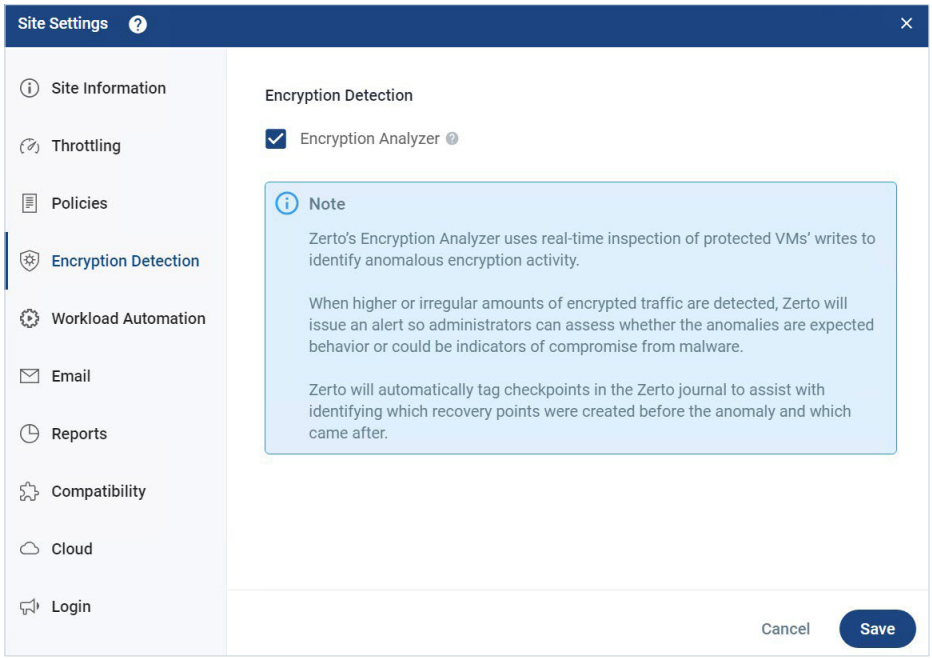
HPE Zerto provides seven API endpoints for the encryption analyzer:

Table 1. Seven API endpoints for the encryption analyzer

Type	Endpoint	Description
GET	/v1/encryptionDetection/suspected/volumes	Get a list of volumes that are suspected to have an encryption event
GET	/v1/encryptionDetection/suspected/vms	Get a list of VMs that are suspected to have an encryption event
GET	/v1/encryptionDetection/suspected/vpgs	Get a list of VPGs that are suspected to have an encryption event
GET	/v1/encryptionDetection/metrics/volumes	Get a list of volumes with encryption data
GET	/v1/encryptionDetection/metrics/vms	Get a list of VMs with encryption data
GET	/v1/encryptionDetection/metrics/vpgs	Get a list of VPGs with encryption data
POST	/v1/encryptionDetection/dismissEvent	Dismiss encryption event (resolve alert, clear tagged checkpoint)

Considerations and limitations

1. The encryption analyzer requires a virtual manager appliance running HPE Zerto 9.7 or higher; HPE Zerto 10 is strongly recommended.
2. A VM must be protected via a VPG for the encryption analyzer to inspect and assess its writes.
3. To avoid dropped I/Os during the collection phase, ensure VRAs are not sized lower than the minimum requirements outlined in HPE Zerto Technical Documentation.
4. Since VRAs transmit their RED metrics to the virtual manager, the training period will reset for all volumes if the virtual manager is rebooted, but not if a VRA is rebooted.
5. The encryption analyzer is currently only available with HPE Zerto for VMware vSphere®, including VMware® on public cloud, and is not yet available for HPE Zerto on Microsoft Azure, Hyper-V or Amazon Web Services. The vSphere version must be currently supported by HPE Zerto as outlined in the Interoperability Matrix.
6. The encryption analyzer is enabled by default, but this can be toggled on or off via the virtual manager GUI under Site Settings → Encryption Detection.



Learn more at

HPE.com/Zerto

Visit HPE.com

Chat now

a00134978ENW, Rev. 1

HEWLETT PACKARD ENTERPRISE

hpe.com