DELL Technologies

Whitepaper

Machine Learning Analytics to Detect Data Corruption Due to Ransomware

CyberSense® for Dell PowerProtect Cyber Recovery

Powered by Index Engines

Real-time cyber security solutions are designed to protect from an attack however, these solutions are not 100% effective and corporate data is still corrupted daily. CyberSense adds a layer of protection to these real time solutions and finds corruption that occurs when an attack has successfully penetrated the data center.

CyberSense helps you get back to business.

June 2022

Table of Contents

Why CyberSense? Why Now

66%

of Organizations were hit by ransomware in 2021[1]

Real-Time Security is Not Enough

Ransomware is circumventing existing security products and attacks are growing

20 Days

Average downtime after a ransomware attack in 2021, up from 15 in 2020[2]

Downtime is Increasing

Organizations are struggling to restore after an attack

***Sabbath Ransomware Gang* Targets Critical Infrastructure, Backups**[3]

Backups Are Not Reliable

Backups and backup data can easily be corrupted with new variants

Organizations have heavily invested in real-time anti-virus products and data center security applications. Over time, many add dozens of these in line applications to their data center, however, ransomware attacks and resulting downtime are on the rise.

How? Cyber criminals are circumventing existing security by upgrading attack variants to be more sophisticated. In the past they relied on very basic variants that encrypted files, appended extensions, and other obvious types of behavior that security tools were detecting. In 2021, the attacks involved more sophisticated corruption including partial encryption, hands-on keystroke changes and content changes that hidden inside files and databases and are harder to detect.

And they're attacking critical data assets, including:

- Core infrastructure including Active Directory, DNS, and LDAP
- Production databases including Oracle, DB2, SQL, IRIS and SharePoint
- User files including contracts, financial documents, and intellectual property
- Backups, encrypting and corrupting backup images.

Why? To make it as challenging as possible to recover so they can demand higher ransoms.

[1] Sophos "The State of Ransomware 2022": https://www.sophos.com/en-us/content/state-of-ransomware

[2] Statista "Average duration of downtime after a ransomware attack from 1st quarter 2020 to 4th quarter 2021: https://www.statista.com/statistics/1275029/length-of-downtime-after-ransomware-attack/

[3] https://securityintelligence.com/news/sabbath-ransomware-critical-infrastructure-backups/

CyberSense: The Last Line of Defense When Existing Security Solutions Fail

Most organizations have dozens of security tools, such as firewalls, deployed to help fend off cyberattacks and protect their data. One of the most common tools used is antivirus software. Antivirus software runs on the production environment which is the primary attack surface for ransomware. In real-time, it's designed to scan programs and files as they are created or changed and compares them to known virus signatures. Backup software that has integrated virus scan can be another layer of protection from cyberattacks. Scanning the backup content for malware in the production environment provides an additional defense, but it's still directly on the attack surface for cyberattacks.

Antivirus software is imperative, but it's not 100% effective. What if there is no malware to be detected? Many attacks are hands-on keystrokes with manual intervention. Therefore, the data isn't corrupted and antivirus software will be completely oblivious to the attack. Cyber criminals spend massive resources developing technology that circumvents these security tools tools as well as focusing on attacking the production backup environment. Including backup environments that have integrated virus scans. A last line of defense is needed to detect when corruption begins to occur, diagnose the activity, and facilitate the restoration process.

CyberSense is not a replacement for existing real-time security applications or virus scans. It compliments these applications and is fully integrated with Dell PowerProtect Cyber Recovery. Within the cyber recovery vault, CyberSense delivers petabyte-class scanning of backup images designed to check the integrity of data and detect suspicious behavior including encryption, mass deletions and data corruption. CyberSense continually checks your data for signs of ransomware corruption and will alert you when corruption begins. When an attack occurs, CyberSense provides post attack forensic reports to diagnose the damage and report on the last known good files to facilitate a rapid recovery.

At the core of CyberSense are analytics and machine learning. CyberSense generates thousands of analytics by looking at observations of your data over time and then feeds them to powerful machine-learning algorithms to make a deterministic decision on whether the data looks good or suspicious. When data is found to be corrupted it's tagged. A report of these corrupted files is provided along with the last good backup copy. It will know where the corrupt data is, where the last good version of the data is, and what backup sets the data was in to streamline the recovery process.

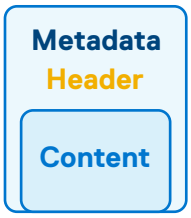
With CyberSense, organizations are not at the mercy of the cyber criminals. They can utilize existing disaster recovery resources to quickly return the business to a steady state. For confident detection and recovery, CyberSense is your only option.

Only CyberSense:

- Validates the integrity of data
- Provides confidence that data is good
- Alerts when signs of corruption are detected
- Reports on the last good backups for restoration

The CyberSense Advantage

CyberSense is the only solution of its type to deploy full-content analytics on all protected data. This is the only way you can be confident that your data has integrity and that cyber criminals are not circumventing your data security tools, hiding their tracks and covertly corrupting your data. In order to understand how bad actors corrupt data, you first need to understand the anatomy of a file.



A file contains three main components:

- **Metadata:** document properties including file size, extension and name.
- **Header:** The content header that includes true file type and document structure.
- **Content:** The content of the file or pages within a database.

Cyber criminals will initially deploy basic types of malware, if they can get away with it. Basic types of attacks are focused on corruption that impacts the file metadata. Changes include appending a file extension with .wcry or .locky or a large change in file size due to file corruption. These basic types of corruption are easy to detect. Vendors with metadata only analytics are just seeing obvious signs of corruption that impact the file properties. When these attacks are detected, cyber criminals will move on to more advanced types of attacks. Advanced attacks are focused on hiding corruption inside a file or page of a database. These attacks are difficult to detect. They include malware that corrupts a header or the content of a file or database. Other types of corruption change the file structure and encrypt or partially encrypt the internal contents of a file or page of a database to avoid detection.

Both types of advanced corruption are not obvious to detect if metadata alone is used for analysis. There are even ransomware, such as JigSaw, that will append a file name with a valid file extension such as .fun to obfuscate the file. A metadata scan would detect this as a valid extension and result in a false negative. CyberSense content analytics would read the header and detect that the file type is not accurate.

Vendors with only metadata-based analytics compensate for a high-level view of how files are changing by incorporating thresholds. Thresholds will alert when the number of file changes, in size for example, or deletions, goes above the norm for the day. Thresholds along with metadata properties make an educated guess that the behavior is suspicious and will send an alert. This will result in false positives.

CyberSense utilizes over 200 full-content-based analytics that are indicative of the types of corruption resulting from a ransomware attack. Other solutions are based on metadata, and not full content. As such, they will have only metadata-based evidence, which will max out at about 12 analytics properties, versus CyberSense with over 200 analytics. This comprehensive insight does not need to guess at suspicious behavior but is deterministic with the diagnosis with 99.5%[1] confidence in finding corruption due to an attack.

If Databases are critical to your business, then CyberSense is your only option

[1] Based on Dell analysis of publicly available data, June 2022. Actual results may vary.

Powerful Machine Learning

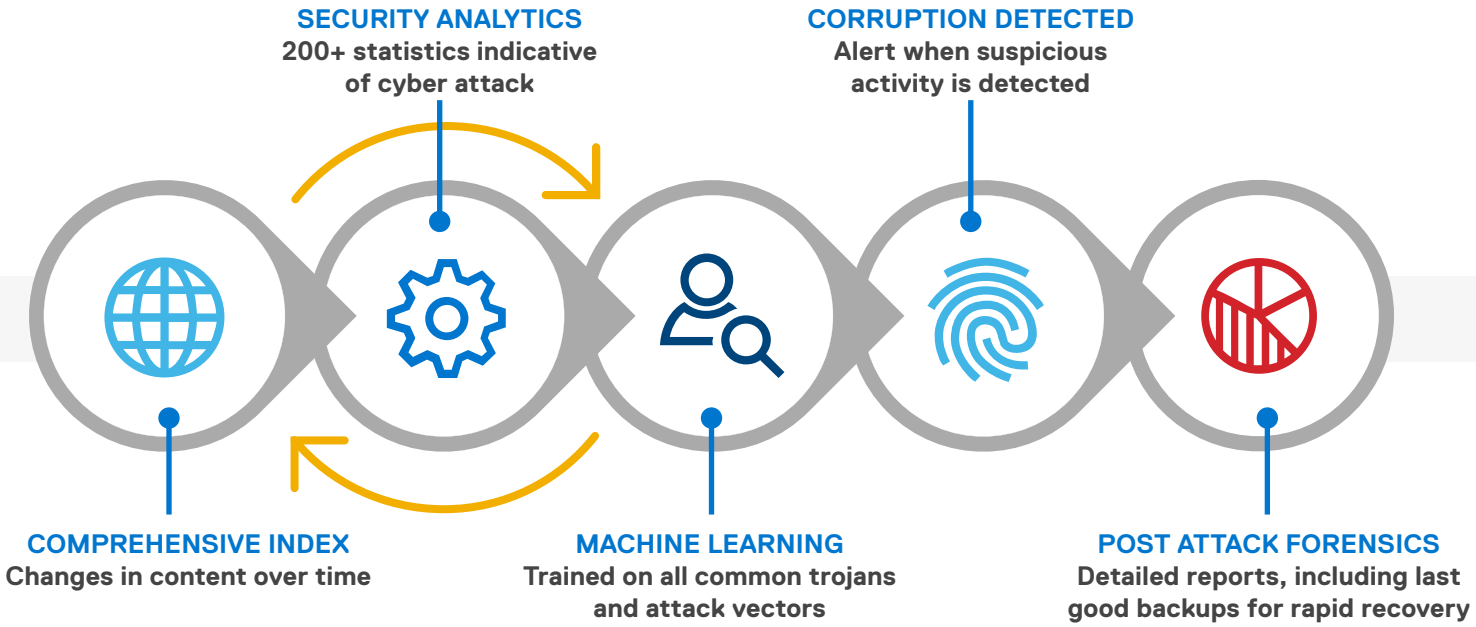
Over 200 analytics
Data observations over time
Trained on thousands of malwares

- **Detects unusual patterns**
- **Distinguishes user activity versus ransomware corruption**
- **Minimizes false positives and negatives**

Powerful and deterministic machine-learning is at the heart and brain of CyberSense. CyberSense combines over 200 analytics – over 20x as many as competitors - with data observations that get more intelligent over time with more observations. The machine-learning is trained on thousands of malware infections to find unusual patterns of behavior and distinguish user activity vs ransomware, while minimizing false positives and negatives.

Machine learning gets additional education from research in the CyberSense labs, including new attack variants. Additionally, machine learning is updated based on real-life data from existing CyberSense clients. This is achieved by exporting CyberSense analytics from the Cyber Recovery vault, via a secure data diode, and delivering them to the CyberSense development team. These analytics are used to further educate the machine learning model and the resulting updates are then available to all CyberSense clients.

Below is the orchestration of the CyberSense workflow once data is synced to the vault.



1. **Comprehensive Index:** Backup image is indexed at the content level
2. **Security Analytics:** Over 200 analytics indicatives of corruption collected for each image
3. **Machine Learn:** Utilizes the analytics, compares current backup with previous
4. **Repeat:** If no corruption, previous steps repeat with every new replicate backup image
5. **Corruption Detected:** If corruption is detected an alert will be sent to the dashboard
6. **Post Attack Forensics:** Reports that detail who, what, where and when to help with recovery

CyberSense Vs. the Others

Many other backup vendors have added some type of metadata analytics to their products. On paper it may sound good and on screen it may look easy, but they provide a false sense of confidence that will lead to false negatives and false positives. Users may start to ignore them and perceive them as 'normal'. These analytics will prove to be unreliable when faced with today's advanced attacks.

CyberSense: Full Content Analytics

Over 200 analytics that look for signs of corruption within files, databases and core infrastructure

Others: Metadata Analytics

Metadata analytics that cap at ~12 and are limited to files only. Other critical data assets such as databases not included

CyberSense: Advanced Machine Learning

Trained on thousands of malware and trojans to detect unusual patterns indicative of ransomware attacks

Others: (Un)Educated Guess

Use of basic analytics and thresholds to determine if unusual behavior has occurred

CyberSense: Comprehensive Insight

Deep analytics fed to trained machine learning models provide 99.5%[1] confidence in corruption detection

Others: Limited Insight

Metadata analytics plus thresholds provide low levels of insight that can easily be circumvented producing false negatives

[1] Based on Dell analysis of publicly available data, June 2022. Actual results may vary.

We see attacks today that others are missing. Some examples of variants that fall in the gray area where light weight analytics tools fail include:

- **Slow Corruption:** hides below the threshold radar
- **Appended File Extensions:** need to read file header to confirm
- **Partial Encryption:** subtly corrupts file content
- **Mass Deletions/Creations:** number of files do not change

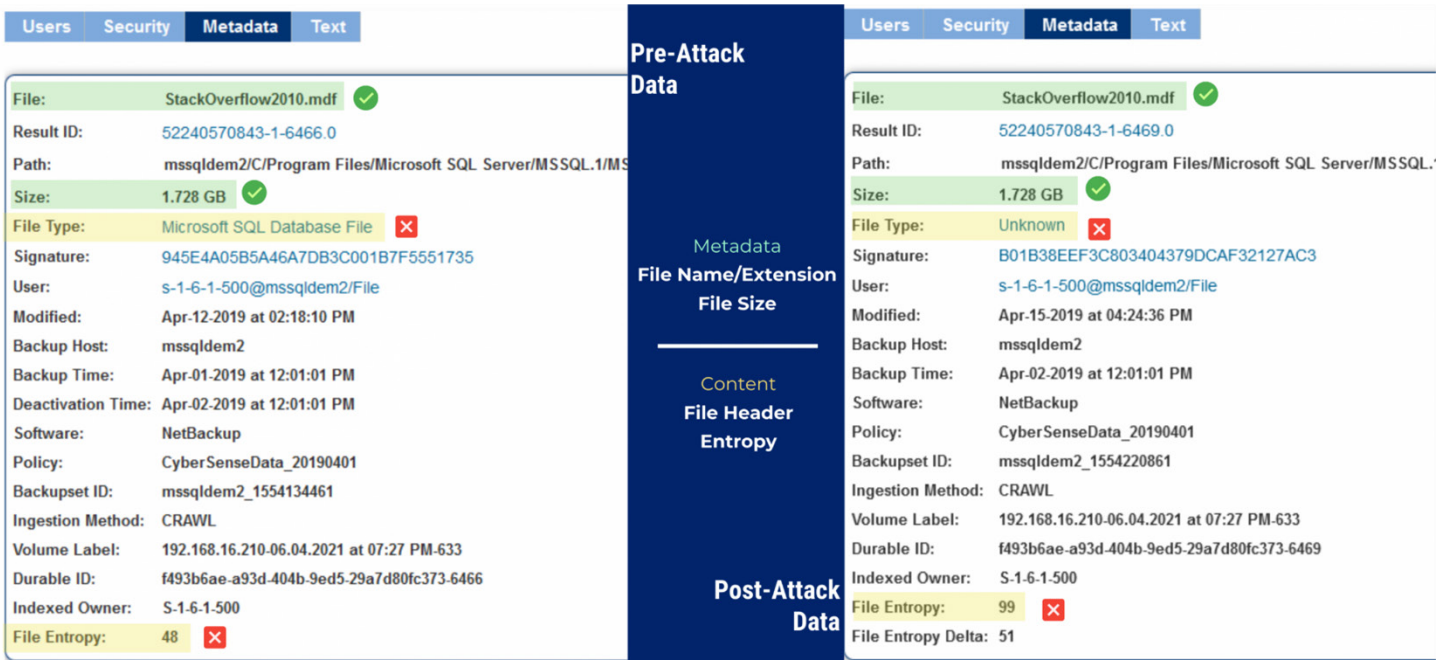
With CyberSense, you're analyzing backups and how data changes over time, every observation, every backup, is a view into data. The challenge is to determine what changes are normal user activity and what are indicative of ransomware corruption.

Files can be added, deleted or encrypted both by users in normal everyday activity, as well as by cyber criminals perpetrating malicious activity. Determining which is which is the challenge. Appending a file extension such as .encrypted is obviously cyber corruption. Creating a series of files is obviously normal user behavior. However, there is a wide gray area where cyber criminals dwell where they make changes that are hard to detect and determine if it is normal user behavior or malicious corruption. It is this gray area where newer ransomware variants focus because they are harder to detect. With CyberSense's full-content analytics, intelligence, and machine learning to check this data, it can detect corruption even in the gray area of file changes.

How do we know CyberSense delivers such a high accuracy in detecting corruption do to ransomware or cyberattacks? CyberSense is continually tested against the latest malware and trojans. Over 20 million backup sets are tested with CyberSense to find corruption. When a new variant is found in research, academia, and other sources such as VirusTotal, those variants are downloaded, executed to corrupt data, and tested with CyberSense. Beyond this testing, CyberSense has access to many analytics from end user installs to validate the confidence of the corruption alerts. All together this delivers 99.5%[1] level of accuracy in detecting corruption for existing and new ransomware variants. With other solutions, the obvious signs of corruption and user data may be detected. However, they will struggle and often fail to uncover corruption in the gray area where cyber criminals hide their tracks and make it a challenge to detect resulting in false positives and more troublesome, false negatives.

[1] Based on Dell analysis of publicly available data, June 2022. Actual results may vary.

CyberSense in Action



Post Attack Recovery

An example of advanced corruption is the AlphaLocker ransomware. This ransomware will encrypt the file contents while maintaining the original metadata. In the image you can see an example of a pre and post attack version of a file.

In checking the file information, you can see highlighted in green the file name and size have remained the same before and after the attack. A basic metadata analytics tool would see these properties and assume the file has integrity.

Looking deeper, using CyberSense's content-based analytics, you will see highlighted in yellow that the file header is now corrupted, and the entropy score has increased to 99 signifying encryption. The machine learning in CyberSense will recognize the file cannot be typed and the internal structure validated along with a high entropy score will then generate an alert.

There are many examples of ransomware on the market today that avoid detection by basic metadata only analytics. Some examples are JigSaw and CyrpMIC. Both types of ransomware will avoid metadata changes and focus on corrupting the internal content of a file or page of a database.

Beyond detecting corrupted data as a result of ransomware, ensuring that the malware is not restored during a recovery process is of utmost importance. When a recovery is underway, CyberSense can search the relevant backup sets to determine if a file or directory exists. Using a file signature, file name, or directory/path, a search can be executed on the backups before they are restored. If ransomware is detected, these files or directory can be deleted, and further corruption prevented.

Know

When it Happened.

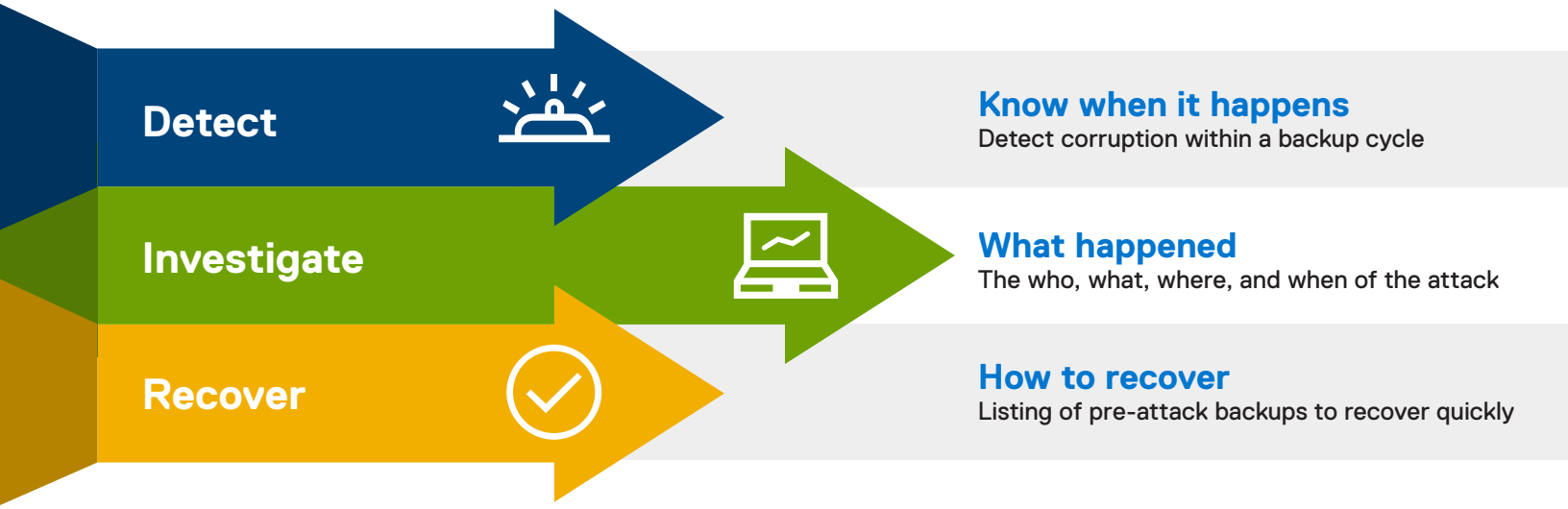
What Happened.

Who it Happened to.

How to Recover.

When an attack occurs, it is critical to diagnose and recover from the attack and get business operations back to a steady state as quickly as possible. When backup data is synced to the vault, it is indexed to determine if corruption is detected. If corruption is detected, an alert is sent out so your organization can begin the recovery process quickly and minimize business down time.

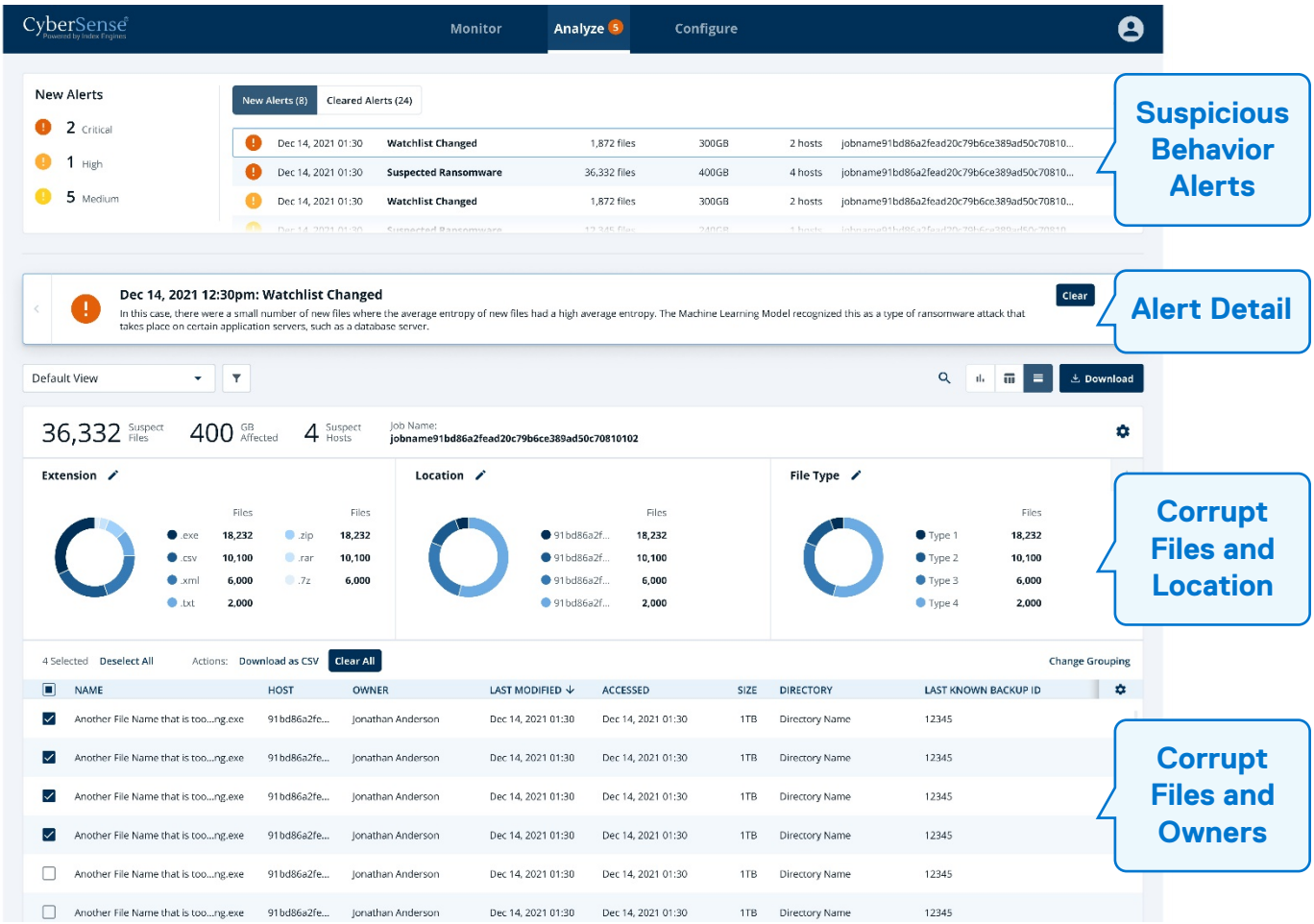
Minimize Business Downtime



CyberSense will provide the following insight:

- **Who** was impacted and how much damage was done on what servers.
- **What** was attacked including a listing of files by path, owner and department.
- **Where** is the source of the attack including the breached user account and ransomware.
- **When** was the last good backup of the corrupted files.

With comprehensive reporting from CyberSense the mystery of the attack can be exposed. All the details you need will be available to determine the most confident and reliable course of action. In the Analyze dashboard below, you will see suspicious behavior alerts, alert details, corrupt files owners and more. This dashboard will give you a complete view of your critical data's integrity in order to help you resume your business as quickly as possible.

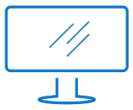


Conclusion

Multiple layers of protection are needed for organizations to be cyber resilient. Antivirus software is just one layer and very effective at detecting virus signatures. However, cyberattacks that aren't using virus signatures can be missed and cause drastic business disruption. CyberSense's full content analytics, machine learning and unusual patter recognition detection is needed in addition to antivirus applications to withstand cyberattacks. CyberSense for Dell PowerProtect Cyber Recovery provides the last line of defense that will streamline the recovery process from a cyberattack. **Detect, Diagnose and Recovery.**

No paying exorbitant ransoms!
No being at the mercy of cyber criminals!

Get back to business with CyberSense.



Learn more about CyberSense for Dell PowerProtect Cyber Recovery

Contact a Dell Technologies Expert

View more about Dell PowerProtect Cyber Recovery

Join the conversation with #PowerProtect