CyberSense by INDEXENGINES

Innovative AI Strategies Supporting Trusted Data Recovery

CyberSense AI White Paper

Impact of Ransomware

The impact of ransomware is devastating. It is clear why cyber resiliency has become a top priority for IT organizations worldwide. When bad actors, such as cyber-attackers, infiltrate the data center, they cause chaos and destruction, demanding significant ransoms by locking down and corrupting critical business data. This disruption can cripple organizations, halting daily operations.

For example, production databases like Oracle, SAP HANA, and Epic/IRIS, when unavailable, force companies to revert to manual record-keeping and worst-case bankruptcy. User files, including contracts and spreadsheets, when deleted or unavailable, result in the loss of intellectual property. Core infrastructure, such as Active Directory and networking applications, when corrupted, terminates normal business operations. Bad actors know what, when, and how to attack to create maximum impact, forcing organizations to pay exorbitant sums to unlock their data assets with the hope to resume normal operations as quickly as possible.

Evolution of Ransomware Variants

Ransomware variants, thousands of which exist in the black market, modify data to make it unusable. In the past, these attacks corrupt or encrypt targeted files using brute force methods, with data encryption being a popular approach. However, these methods are no longer effective. More sophisticated variants are now the norm. Bad actors have developed variants that make detection a challenge. They use AI-based tools to build ransomware variants that stealthily corrupt data to avoid detection. Variants include intermittent encryption, where small portions of files and databases are encrypted to avoid detection. Other variants use encryption algorithms such as XOR or encode encrypted data using base64, which can hide the corruption from detection using compression rates or increases in entropy. Cyber criminals know that many organizations use behavioral based analytics and supporting tools to detect changes in compression rates or file encryption, so they have developed advanced variants to circumvent these detection tools by changing their behaviors. They are smart, and with AI at their disposal, they are getting smarter.

The combination of advanced variants, which corrupt data, and polymorphic variants, which are nearly impossible to detect, requires new and innovative approaches to securing data and ensuring its integrity when recovery is needed.

Ransomware variants are constantly evolving. Building a resiliency model that looks for ransomware executables or indicators of compromise based on behavioral analysis coupled with encryption detection is often flawed and can be evaded. For example, new polymorphic ransomware changes its code or appearance to evade traditional security measures by frequently mutating variants, allowing them to live freely in the data center and execute when needed. The combination of advanced variants, which corrupt data, and polymorphic variants, which are nearly impossible to detect, requires new and innovative approaches to securing data and ensuring its integrity when recovery is needed.

Understanding the Role of AI with Data Integrity

Innovation today must leverage AI. It is not just a buzz word, but a critical tool for analyzing large volumes of data to ensure integrity. The advantages of AI are clear: efficiency of processing, speed of results, scalability to handle large volumes of data, pattern recognition to uncover hidden insights, automation of repetitive tasks, predictive analytics to analyze historical data and look forward, and confident decision-making based on comprehensive analysis. AI was built to solve complex challenges such as ransomware and support smarter recovery from an attack to minimize business impact and avoid the loss of critical data assets.

The most critical use of AI for cyber resiliency is its ability to monitor data in real-time and over periods of time to ensure data integrity. Many times, this is a critical last line of defense against ransomware but notably, the most impactful on the business. Without data integrity (or clean data), organizations will fail, and there have been cases where companies have gone bankrupt as a result.

Simply looking for indicators of compromise is insufficient; you need to look for patterns of behavior based on real-world activity.

Data Integrity is Different than Data Protection

Organizations protect data using standard data protection strategies; however, these strategies do not validate data integrity—they simply take a copy of what exists, whether it is clean from corruption or not. Many data protection and storage vendors have added "cyber" capabilities to their products that look for indicators of compromise or unusual data modifications. These approaches are loaded with false positives as they lack the insight over time needed to confidently determine if data modifications result from a bad actor or a normal user. When organizations face false alerts, often it must be manually investigated to determine if it is a true ransomware attack. As these alerts continue, they will eventually be ignored as the majority will be false. Remember, bad actors know what, when, and how to attack to create maximum impact – which manual intervention cannot detect.

As a result, leveraging AI must be integrated within the cyber resiliency solution, whether it includes their production data environment, the backup or data protection, or recovery strategies. AI-powered data integrity should be a standard offering for both primary and secondary storage, delivering ongoing confidence in the chosen cyber recovery strategy.

Post-attack, the analysis of these environments should include detailed forensics on what servers and data were impacted, analysis of the attacked data, and the last clean version of the data for recovery. This is the true value of data integrity, allowing for faster recovery and minimizing data loss. Without AI-powered cyber resilience, this is impossible.

CyberSense Approach

AI driven models require several components, including a comprehensive set of analytics collected at each observation for decision making, data for training and testing, proprietary algorithms and models, parameters, and variables to adjust the models, training, continual evaluation, and finally deployment and monitoring. Early on, CyberSense determined that the most confident approach to analyzing data integrity, with negligible false positives/negatives rates (0.001% false positives as of 2024), was to analyze actual ransomware variants. To understand what cyber criminals do, you must research and understand the tools they use as they evolve. Simply looking for indicators of compromise is insufficient; you need to look for patterns of behavior based on real-world activity. This is how CyberSense trains its AI driven models.

The CyberSense Research Lab supports the analysis of real-world ransomware variants using an automated approach. The lab has architected an automated process that includes:

Data: Accumulate millions of files, backups, and snapshots of thousands of types of data for testing. These data corpuses are representative of customer environments.

Proprietary algorithms and models: CyberSense utilizes several machine learning models that inspect how data changes over time.

Parameters and variables: As testing against actual ransomware variants proceeds, the machine learning model is updated and adjusted to maintain a high level of accuracy.

Deployment and monitoring: CyberSense does not release a new update to the software until it undergoes rigorous testing and maintains a 99.99% accuracy in detecting data corruption.

Training and continual evaluation: Hundreds of ransomware variants are accumulated daily, and an automated testing process detonates the variant behaviors and uses CyberSense to scan both clean and infected data sets to test and fine-tune the models. Additionally, anonymized customer analytics output is utilized for both training and evaluation.

Instead of detecting specific ransomware variants, CyberSense utilizes generalization to evaluate and then categorize thousands of variants into a few dozen classes based on patterns of behavior that bad actors use to corrupt files and databases. Patterns based on years of research. These patterns also predict future variant behavior, as when cyber criminals modify their approach, it will fit into one of the predefined classes. The generalized classes include patterns representing types of obfuscation and encryption deployed today. With new variants analyzed daily, models can easily be adjusted in the rare event of a new and unique pattern emerging.

AI-Based Machine Learning Models:

AI-based machine learning is a sophisticated approach to ensuring data integrity. A common example of how AI can be used may help clarify how AI can deliver a high level of accuracy based on patterns of analysis. Take the example of facial recognition models. You can train a facial recognition model using a few different pictures of a person's face. The common example is when you train your phone to unlock after scanning your face. The training process will take several pictures of your face. These are point-in-time snapshots. When you request your phone to unlock, it will take a current point in time snap of your face and determine if it matches the pattern. It has learned the pattern of your face, which can be different from the images it was trained on, but AI has determined with high accuracy that it is in fact you.

CyberSense provides reliable protection, much like how your phone recognizes you today, not just based on the photos it was originally trained on.

CyberSense operates on a similar model, identifying patterns learned through the regular detonation of real ransomware to analyze and use for training. This innovative approach ensures confidence in data integrity and recovery after an attack. By continuously understanding, maintaining, and predicting ransomware patterns, CyberSense provides reliable protection, much like how your phone recognizes you today, not just based on the photos it was originally trained on.

CyberSense uses a more holistic approach to analyzing how data evolves.

Automated Analysis Process

With its AI engine continually trained and tested for accuracy, CyberSense scans data by integrating with point-in-time copies through backups or snapshots. It establishes an initial view of files, databases, and core infrastructure on day one, then compares subsequent scans to detect changes with each new observation. This enables high-accuracy detection of patterns indicative of ransomware corruption, including slow attacks that result in slight changes over time.

Unlike solutions that simply compare snapshots or backups, CyberSense rebuilds server content from incremental backups or snapshots to analyze how servers evolve over time. Solutions that simply analyze data changes from one incremental backup or snapshot to the next result in incomplete view of data comparisons. CyberSense uses a more holistic approach to analyzing how data evolves. By comparing servers and not just incremental copies, CyberSense has advantages in detecting slow-moving attacks or data deletions that will not appear in backups.

CyberSense leverages hundreds of content-based analytics for each host analyzed to identify corruption within files and databases, such as altered file headers or corrupted database pages. This unique content-focused analysis is performed after every scan, ensuring precise detection of ransomware behavior. Combined with ongoing training on emerging variants, CyberSense achieves 99.99% confidence in identifying attacks, detecting corrupted data, and providing a forensic report to guide recovery efforts.

Real-World Scenarios Bypass Conventional Tools

One of the more sophisticated variants on the market today is Expiro/Xpiro and Virlock, which use byte substitution or XOR encryption with an 8-byte key to corrupt data. These variants deploy a pattern of corruption that CyberSense detects. They modify the file header, changing it to an executable, which requires content-based header analysis to detect. Additionally, they use byte substitution to surgically replace individual bytes in the file, corrupting the content and circumventing threshold and entropy analysis. CyberSense identifies these changes by analyzing the content and understanding the pattern of corruption.

Beyond byte substitution, variants that deploy intermittent encryption within files and databases pose an even greater challenge for detection by conventional tools. These variants have been responsible for some of the most impactful attacks in recent times, variants such as LockBit, BlackCat (ALPHV), and Black Basta. Tools that focus solely on metadata changes or rely on analytics without inspecting file content are likely to miss these attacks, leading to false negatives.

CyberSense is an innovation that is proven daily to customers, including identifying malicious activity from the most sophisticated variants. Exabytes of data are scanned daily across the globe, providing confidence that data has integrity. There are many approaches to detecting data corruption due to ransomware, but each has its flaws. Looking for specific variants that continually change or indicators of compromise that can be easily circumvented by advanced variants is not enough. Patterns are consistent, and with detailed training, CyberSense can confidently predict new patterns. This approach provides robust data resiliency.

125,000
Data Samples Collected for Testing

94, 100
Infected with Ransomware

94, 097
Successfully Deleted

99.99%
Verified Rate for Accurate Detection

Conclusion

In the face of increasingly sophisticated ransomware attacks, traditional data protection strategies are no longer sufficient. The evolution of ransomware, with its advanced and polymorphic variants, necessitates innovative approaches to ensure data integrity and maintain business operations. CyberSense leverages AI to provide a robust solution for detecting and mitigating the impact of ransomware.

By integrating AI-based machine learning models, CyberSense can analyze large volumes of data with 99.99% accuracy, identifying generalized patterns of corruption that traditional methods might miss. The CyberSense Research Lab's continuous analysis and training on real-world ransomware variants ensure that the system remains effective against emerging threats. This approach not only detects data corruption but also provides detailed forensics and recovery options, minimizing downtime and data loss.

CyberSense's unique ability to scan and analyze data at the binary level, combined with its content-based analytics, offers unparalleled confidence in data integrity. This innovation supports smarter recovery from ransomware attacks, providing organizations with the resilience needed to protect their critical data assets and maintain normal business operations.

In conclusion, CyberSense represents a significant advancement in the field of cyber resiliency. Its AI-driven approach to data integrity and recovery sets a new standard for protecting against ransomware, ensuring that organizations can confidently navigate the challenges of today's digital landscape.

CyberSense can analyze large volumes of data with 99.99% accuracy, identifying generalized patterns of corruption that traditional methods might miss.

Get an inside look at how CyberSense trains our AI for accurate ransomware detection

To add CyberSense AI to your cyber recovery strategy, contact: info@indexengines.com

INDEX ENGINES