Likewise®

Whitepaper

Monitoring Unstructured Data

Uniting Multi-Protocol Storage and Cross-Platform Access Control for File Activity Monitoring and Context-Aware Security

Table of Contents

By Steve Hoenisch, Likewise Software

Executive Summary

Unstructured data is rapidly piling up in file servers and NAS systems – and 40 percent of it tends to be sensitive information: intellectual property, confidential data, and company secrets.

Monitoring the data, however, is harder than storing it. Because the unstructured data of Unix users is typically stored in separate silos from the data of Windows users, it is problematic to track the data with a single, standardized monitoring system.

The problem of monitoring across silos is compounded by incompatible identity management systems for Unix and Windows users, making it hard for standardized monitoring to link users with their identities. Compliance regulations, disclosure laws, and risk management, though, require security that identifies threats based on user identities.

This white paper argues that a multi-protocol file server or NAS system with an integrated cross-platform access control system is a blueprint to efficiently and effectively monitor unstructured data.

First, it frees you from the silos of platform-specific storage, empowering you to monitor all the stored data without regard for storage protocol.

Second, it secures the unstructured data by applying a common security model to it, enabling the monitoring system to associate access with user identities and organizational roles.

Third, it establishes the foundation for a high-performance file activity monitoring system that can track unstructured content in a security-aware context of user

identities, patterns of access, and file change events.

The result is an identity-aware, cross-platform storage system that makes it easy to secure unstructured data from internal threats, monitor user access, track changes to sensitive files, and generate reports that demonstrate regulatory compliance with evidence.

Introduction

Unstructured data is growing faster than all other types of data, industry analysts say. It will increase by as much as 800 percent during the next five years – and more than 40 percent of it, a survey by the Aberdeen Group found, is sensitive. If the data is sensitive, it must be protected: Compliance regulations, disclosure laws, and risk mitigation mandate security.

Meanwhile, industry analysts report that IT security managers are evaluating database activity monitoring tools to comply with regulations and to manage risks associated with structured data stored in databases. Database activity monitoring, however, focuses on structured data without placing a corresponding emphasis on rapidly growing file repositories, exposing a large security hole. For comprehensive security, activity monitoring tools should thus be implemented to perform the same security functions -- only in relation to unstructured data. But the monitoring system, to be effective, must identify threats based on user identities across data silos.

Data silos proliferate because there are two main protocols – NFS and CIFS – for accessing file servers and NAS systems. Unix users access data on file servers by using NFS, while Windows users access data on servers by using CIFS. The inability to interoperate between the two protocols creates data silos, which are difficult to monitor with a common system. Ad hoc systems that add a monitoring layer for file events frequently lead to performance issues.

More importantly, though, most monitoring frameworks fail to tie file events to user identities on both Unix and Windows computers. Just as there are different access protocols, there are different, incompatible access control systems for Unix and Windows users that impede the association of identities with events. When monitoring is integrated with a cross-platform identity management system, it has the power to link events with user identities.

Furthermore, the integration of the identity management system with the activity monitoring system is a prerequisite for effective exception monitoring – analyzing suspect events at the nexus of user identity, access, and activity.

In the past, performance issues and poor resulting data have made implementing a useful identity-aware file activity monitoring system impractical. To ensure that events do not consume too much network traffic or bog down systems, the monitoring should ultimately take place as part of the file server. For fast write speeds and horizontal scalability, events should be pushed to a NoSQL database. In a high-volume enterprise, the result of tight integration coupled with a NoSQL database is a system that scales well to deliver high performance.

This white paper argues that a multi-protocol file server with an integrated cross-platform access control system establishes a powerful identity-aware framework to efficiently and effectively monitor unstructured data.

Why Monitor Unstructured Data?

The main business reasons to monitor unstructured data are as follows:

- **Protect stored secrets, confidential information, intellectual property, and private data. The goal is to control the secrets so they don't get into the wrong hands, which** could hurt your business prospects or undermine your competitive advantage. Private data and confidential information must be protected to meet compliance regulations and disclosure laws.
- **Demonstrate compliance with regulations, legal requirements, and internal security policies.** Some of the key regulations in the United States are the Health Insurance Portability and Accountability Act (HIPAA), the Payment Card Industry Data Security Standard (PCI DSS), Sarbanes Oxley (SOX), the International Traffic in Arms Regulations (ITAR), and the Federal Information Security Management Act of 2002 (FISMA).
- **Mitigate risk of security breeches, data loss, fraud, noncompliance, and legal problems.** Monitoring can detect potential sources of data loss, fraud, incorrect entitlements, inappropriate access attempts, and anomalies that are indicators of risk – especially when the monitoring system can associate data access with user identities.
- **Ensure the quality, integrity, and availability of important unstructured data.** Implementing strong controls over who can access, modify, or delete important unstructured data helps ensure its quality and integrity. Meantime, when important data is highly available, it can save time and spur innovation. In a HIPAA environment, highly available data with the right access controls can save lives.
- **Reduce costs associated with records management, storage, and**

security. Although some unstructured data might need to be stored securely for years for such reasons as complying with government regulations and obtaining patents, storing everything can increase storage costs. The real value of monitoring, however, lies in its potential to radically cut the costs associated with security problems.

Use Cases

The business justifications above manifest themselves in the following use cases, which demonstrate the need to monitor unstructured data.

- **Determine who can access sensitive data and link user identifiers to people and their organizational roles.** Here's an example use case: As a compliance manager, you want to generate a report that lists the identities and entitlements of those who can access sensitive data, and you want to be able to map their identities to their organizational roles so you can vet it for inappropriate access rights.
- **Show what directories and files were changed when and by whom, especially in the face of identities that are inconsistent across access control systems.** Here's an example: As an auditor, you cannot reconcile the identities of some Windows users with the identities of some Unix users, even though you believe they might be the same people accessing the same content from different machines.
- **Track changes, such as modifications of files or security descriptors, as well as attempts to delete content.** Example: As the security manager, you want to monitor sensitive engineering plans that are protected from access by those without permission to view them. In addition, you want to track denied attempts to access or change files so you can respond proactively.
- **Monitor controls that protect the integrity of critical unstructured data.** Example: To support HIPAA compliance at a hospital, you must monitor the controls that protect the accuracy and completeness of patient data. You need to monitor who can and does make changes to patient records and what changes they make.
- **Tag files with metadata to flag important content for exception-based management.** Metadata – including metadata for compliance, data governance, or records management – can mark data as sensitive or as falling under a compliance regulation, such as HIPAA. Example: As an IT security manager at a hospital, you want to filter documents covered by

HIPAA from those that are not, even though both get stored in a NAS system.

- **Generate a security alert when sensitive files or folders are accessed, modified, or deleted.** Example: As the records manager charged with storing and protecting sensitive files, you want an email alert when certain files are accessed or changed.
- **Inspect application data on file servers.** Example: As an IT auditor, you find out that your IT department is increasingly migrating application workloads to filers because it is easier to provision file-based storage than block-based storage. As a result, you have a heightened need for exception-based monitoring of application data.
- **Provide context-aware security where the context is the intersection of content, event, access, and identity.** Example: As an IT security officer in charge of compliance at a defense company, you seek to design a file server for sharing classified ITAR-controlled documents. You must be able to control who can access and view which content, see who changes what, monitor for exceptions, and produce reports to demonstrate compliance with ITAR. You must be able to prove that employees who are foreign nationals cannot view the documents.
- **Use an analytics engine to discover hidden patterns that could reveal internal threats or find new value in the data.** Example: As a researcher, you want to analyze all the data from the monitoring system to extrapolate patterns that can help predict future threats.

In addition, the following requirements ensure that the infrastructure has the flexibility to monitor events and generate reports in a way that fulfills a diverse, dynamic set of needs.

- Be easy to deploy and maintain without requiring extensive customization.
- Create custom dashboard displays for exception-based management. You should, for example, be able to formulate your own queries, including searches using Boolean operators to formulate logical statements of inclusion, exclusion, and so forth.
- Support near real-time collection and analysis of file events from multiple repositories, including for example NetApp storage systems.
- Have an architecture that is flexible enough to accommodate complementary technologies for data loss prevention, such as interoperable storage encryption and document archiving.

Avoiding Performance Issues

In the past, performance issues and poor resulting data have made implementing a useful identity-aware file activity monitoring system impractical. Such systems are frequently undermined by three main performance issues that must be avoided:

- Avoid network performance issues by an over-reliance on technologies such as sniffers, agents, and shims that lack close integration with the storage system. Example: To try to monitor the sensitive data spread out across your many file servers, you put in place network sniffers and shims that analyze files in motion for sensitive data. But users complain that access to the servers is sluggish and network administrators complain that it bogs down the network.
- Avoid file server performance issues traditionally associated with collecting file events. Example: On your Windows file servers, you tried to use the built-in Windows event-logging system to capture file events, but doing so degraded the server's performance and consumed too much memory.
- Avoid database performance and scalability issues associated with the write speeds and clustering requirements of SQL databases. Example: To store millions of access requests and file events, you bring in database administrators to add clusters of SQL databases, but find that the SQL databases are difficult and expensive to scale.

To ensure that events do not consume too much network traffic or bog down systems, the monitoring should ultimately take place close to the data, that is, as part of the file server. In a high-volume enterprise, the result of tight integration is a system that scales well to deliver high performance.

Fulfilling these requirements by positioning content in its rightful security context raises the following question: What kind of architecture for a file server would make identity-aware file activity monitoring a reality without degrading the performance of the network, the file server, or the database? The next section proposes an answer.

Architecture

The following components provide the architecture for a file server that supports a universal approach to monitoring unstructured data:

- A multi-protocol, cross-platform file server or NAS system that supports CIFS and NFS to accept connections from both Windows and Unix computers.
- An integrated authentication engine that can authorize users with Active Directory, NIS, or LDAP.
- An integrated application for marking and tracking sensitive folders and files.

- A secure event monitoring subsystem with collectors and forwarders that record, manage, and transmit file activity events.
- A NoSQL database for event processing and advanced analytics.
- A SQL data store for reports.
- An auditing and reporting console.
- An events dashboard.

Multi-Protocol File Server Accessible by Windows and Unix

At the foundation is a file server that is multi-protocol and cross-platform: It supports both the SMB/CIFS and the NFS protocols, making it usable simultaneously by Windows and Unix or Linux clients. A cross-platform, multi-protocol file server solves the interoperability problem that separates the data of Unix users from the data of Windows users, providing a consolidated approach to storage for users of all types of computers.

Cross-platform incompatibilities have also been a hindrance to applying a uniform set of security policies. In the past, just as there have been different, incompatible storage systems for Unix and Windows users, there have also been different, incompatible identity management systems for Unix and Windows users. Unix clients have tended to use NIS or LDAP, while the de facto standard for Windows clients is Microsoft Active Directory.

Secure Cross-Platform Access Control

In this architectural schematic, therefore, the file server includes an integrated identity management service to authenticate users with Active Directory, NIS, or LDAP – a component that, when combined with the multi-protocol file server, lays the architectural foundation for solving many of the problems in monitoring unstructured data.

The overall result is twofold. First, it frees your users from the bounds of platform-specific storage, empowering you to monitor all the stored data from a single system. Second, it secures the unstructured data by applying a common security model to it, enabling the monitoring system to associate data access with user identities and roles.

The integrated identity service also lets you control access to sensitive unstructured data and, as described below, monitor those controls for compliance.

Collect Access Data and File Evnts for Analytics and Reports

The event collectors and forwarders form the event monitoring subsystem. On

the file server, the event collectors record data about viewing, moving, copying, modifying, or deleting directories or files. The collectors also capture changes to security descriptors.

Over a secure connection, the event forwarders send the file events on to the NoSQL database, where they are stored in a highly flexible format. The NoSQL database allows the events to be manipulated for a variety of purposes, including big-data analytics, forecasting, and business intelligence.

In this way, the NoSQL database becomes the basis for a powerful, flexible analytics engine that can correlate content types, sensitivity levels, modification attempts, security descriptors, user entitlements, access patterns, and patterns in content. The analytics system can, for instance, use data about past access patterns and file events to hypothesize about future patterns. These inferences can identify files that might contain sensitive material and need to be flagged for inspection.

The NoSQL system, meanwhile, interfaces with a SQL Server database that segments frequently used data into columns and rows for reports and custom queries. The SQL Server also makes the data available to the dashboard for near real-time display of file events, especially exceptions, so that they can be acted on to deal with threats, breeches, and compliance violations. For extra security, the solution can easily be combined with interoperable storage encryption.

Performance

Millions of file events can easily overwhelm the network and the monitoring system. Because of the sheer number of events generated as a multitude of users access and modify files, performance is a requirement that must be considered up front – but all too frequently it is not, and it is only after implementation that performance issues emerge: networks slow down, databases overwhelm disk space, dashboards freeze.

In an enterprise environment with 50 million objects stored across a 25-node array, for example, more than 2 million objects can be modified a day, with the number of events for access events and file views being much higher.

The performance of the event monitoring system plays a key role in how efficiently end-user components that rely on events will function. To be expedient and relevant, exception monitoring and compliance reports depend on how fast events are collected and correlated.

The NoSQL database adds a unique high-performance layer: It digests events with write speeds faster than SQL databases and, more importantly, can easily scale

horizontally to handle more events.

To ensure that events do not consume too much network traffic or bog down systems, monitoring ultimately should take place as part of the file server. When the monitoring is handled by the file server and is built with performance in mind, it ensures that the system scales to deliver high performance in high-traffic environments.

Features

The architecture outlined above exposes the following features and methods to track and monitor unstructured data.

Classify and Track Sensitive Files Tied to Identities and Owners

The application lets you mark sensitive files, associate them with the identities of their owners, and track changes by user. Records managers who are charged with managing confidential information in unstructured files, for example, can use the identity service to limit access to specific users and map file changes to those users.

Email Alerts

When monitoring detects certain defined security events or exceptions, alerts can prompt administrators to take action, such as sending an automated email to inform a user about a potential violation of policy.

Reporting Console

Reporting can mitigate security threats, identify vulnerabilities, inspect access rights, show patterns of access and change, and double-check levels of protection – all of which can help comply with regulations such as PCI, SOX, and HIPAA.

Compliance often leads to the deployment of security information and event monitoring tools (SIEM). Yet few organizations have tied reports to SIEM tools. Even fewer have integrated their reporting tools with their identity management and access control systems. As a result, the tools cannot report on user access and activity and detect exceptions based on one of the most important IT security factors, an authenticated identity.

Linking the reporting system to SIEM tools as well as the identity management system lets you show who owned and modified sensitive files over time.

Dashboard

For internal and external threat monitoring, the dashboard displays, in near real-time, file events correlated with identities and permissions. The dashboard's

exception-based management proactively monitors access to servers and changes to tracked files. The result: real-time situational awareness of what's happening to your sensitive unstructured data.

Situational awareness of the changes being made to tracked files, such as an attempt to change a file's security descriptors, is in effect file activity monitoring, or FAM. It can help comply with the file integrity monitoring stipulated in PCI DSS requirement 11.5 – raising an alert for unauthorized changes to content files.

More importantly, however, file activity monitoring is, for unstructured data, the cornerstone of threat monitoring and risk management. The next sections explain why.

File Activity Monitoring

In the architecture outlined above, the event monitoring subsystem makes possible a file server with integrated high-performance file activity monitoring. Similar to database activity monitoring, FAM refers to tools that can identify and report on file access patterns that could be noncompliant, fraudulent, or illegal. More broadly, the tools can also be used for discovery and classification, vulnerability analysis, intrusion prevention, and risk management.

Increasingly, industry analysts report that security managers are looking at database activity monitoring tools to comply with regulations and to manage security risks associated with structured data stored in databases. Doing so, however, focuses on structured data without placing a corresponding emphasis on rapidly growing file repositories, exposing a security hole. FAM technologies should likewise be evaluated and implemented to perform the same functions – only in relation to unstructured data.

"One or more vendors may be adding capabilities for activity monitoring of unstructured data, to enable enterprises to understand what is happening with their Windows or Unix file shares, for example," Jeffrey Wheatman of Gartner says in The Future of Database Activity Monitoring. "Gartner believes this is an important potential development, and one that enterprises considering DAM solutions should follow closely, because a narrow focus on structured data is a long-standing weakness of DAM technology."

File activity monitoring is at its most powerful when it is tied to identity management. In fact, the integration of the identity management system with the activity monitoring system is a precondition for effective exception monitoring. It is effective because it records exceptions at the nexus of user identity, resource access, and file activity.

Architecturally, when the file activity monitoring system is part of the file server, as it is in the architecture outlined above, it can exploit its close ties to the server to track activity at both the level of user access and the level of the file, heightening visibility into changes to content in a security-aware context.

Content and Context

The importance of file activity monitoring highlights the shift in IT toward contextualized security – in this case, viewing content in the context of identity, entitlements, access patterns, sensitivity levels, file events, and other factors related to security.

A file server with an architecture that includes an identity service and file activity monitoring can collect supplemental information – data that can be combined in different ways in near-real time for situational awareness:

- **Identity:** Authentication transactions, business roles of users and groups, entitlements, permissions.
- **Access:** Whether access is granted or denied, type of access (read or write), time of access, IP address and type of client requesting access, etc.
- **Content and metadata:** Tracked directories or files, files marked sensitive, types of files such as spreadsheets or Word documents, directory name, file name, files marked for a compliance regulation like ITAR.
- **Event:** Actions such as read, write, modify, copy, move, or delete a file or directory; changes to security descriptors, permissions, etc.

When identity, access, content, and event are tracked at the file server, monitoring is enriched by contextualized security data – the correlations that take place at the intersection of users with known roles and entitlements accessing tracked content to perform logged events. The data lights up a dashboard with context-aware security events and exceptions that can be used for decision making, troubleshooting, forensics, and compliance auditing.

The result is that you can monitor the data in context and then use the information to dynamically adjust your security policies to improve compliance, mitigate risk, and cut costs. If, for instance, the file events show that someone from marketing is accessing sensitive financial information in an accounting folder but throwing an exception because of a role-content mismatch, you can get the security settings changed to disallow access.

These kinds of real-time, context-driven policy correctives stem from the protection and control architecture for information life-cycle management outlined above – an architecture that fuses user-based policies, role enforcement, access control, and context-awareness of when and how unstructured data is viewed and modified.

Using Big-Data Analytics to Mitigate Risks

In the events that are generated when you track content in the context of identity and access, there lies a huge amount of data that describes patterns of access, activity, and change – data that becomes an input to an important use that progressive auditors can exploit to mitigate risk in the future: Analytics.

An analytics system can use the data about past access patterns and file activities to hypothesize about future patterns. Such inferences can identify sensitive files that might need to be tracked. The data can also be correlated in unexpected ways to produce innovative results – you can find new value in your old data.

A solution that integrates big data analytics with file server technology heightens the strategic importance of IT: All of a sudden, IT is poised to provide services like legal discovery, classification, and data governance. Such services can increase competitive advantage and revenue while improving security.

Big Data and Business Intelligence

"Big data has quickly emerged as a significant challenge for IT leaders," Michael Cooney, citing Gartner research, writes in Network World. The architecture prescribed in this white paper can turn an IT challenge into an IT opportunity. In particular, the NoSQL server that's included in the storage architecture can be used to analyze unstructured data by using a distributed data processing technique Google invented, called MapReduce. The business value of big data is there to be exploited in myriad ways, such as exploiting the data to find new value that increases revenue or cuts costs.

Conclusion

A multi-protocol file server or NAS system that includes an integrated cross-platform identity management service to control access provides the architectural basis to effectively and efficiently monitor unstructured data.

First, it frees you from the silos of platform-specific storage, enabling you to monitor all the stored data without regard for storage protocol.

Second, it secures the unstructured data by applying a common security model to

it, enabling the monitoring system to associate data access with user identities and roles.

Third, it establishes the foundation for a powerful high-performance activity monitoring system to track unstructured data in a security-aware context of user identities, patterns of access, and file change events.

The result is an identity-aware storage system that makes it easy to secure unstructured data from threats, monitor user access, track changes to sensitive files with situational awareness, and generate reports that demonstrate regulatory compliance.

Ten Steps to Effective Monitoring

Mismanagement of unstructured data can put your reputation at risk, lead to privacy violations, and result in costly compliance incidents. In large organizations, administrators and users are frequently unaware of regulatory requirements for sensitive data. Unless automated systems are put in place to force adherence and to monitor for lapses, users will inadvertently subvert those requirements.

Here's a ten-step program to organize, protect, and monitor your unstructured data.

1. Identify your IT monitoring requirements in relation to compliance regulations, disclosure laws, industry standards, and internal security policies.
2. Find your secret, toxic, confidential, and otherwise sensitive unstructured data.
3. Consolidate your sensitive unstructured data to a cross-protocol file server or NAS system that can be accessed by Windows as well as Unix, Linux, and Mac clients.
4. Integrate the file server with an identity management system that can provide cross-platform access control for Windows, Mac, Linux, and Unix users.
5. Implement an identity-aware security incident and event monitoring system, or SIEM, to track access to sensitive data.
6. Tightly integrate the monitoring system with the file server and the identity management system to ensure scalability. It must perform well even in enterprises with heavy network traffic and a deluge of user activity.
7. Make sure the monitoring system includes a dashboard to display near real-time security events and exceptions.
8. Include a module to generate compliance reports. Make sure the reporting system can create custom reports as well as reports from predefined templates for regulations like Sarbanes-Oxley, PCI DSS, FISMA, and HIPAA.
9. Use an analytics engine to aggregate all the monitoring data so you can look for new patterns that reveal vulnerabilities, predict threats, detect aberrant patterns, and find innovative ways to mitigate risk.
10. Standardize monitoring across all your file servers and NAS systems. In its survey on Securing Unstructured Data, the Aberdeen Group says, "The standardization of audit, analysis, and reporting is an emerging capability with respect to unstructured data." The Aberdeen Group adds: "Standardization in this area improves visibility, provides a common point of reference, and reduces the ongoing cost of operations compared to non-standard, ad hoc methods."

Consolidating storage of sensitive unstructured data to file servers governed by a common access control system can identify who has access to what data, providing a framework to monitor the data to demonstrate compliance, mitigate risks, and cut costs.

Next Steps

For more information on Likewise, visit the web site at www.likewise.com. To contact the sales team, call (800) 378-1330 or email info@likewise.com.

About Likewise Storage Services

Likewise Storage Services provides a platform for cross-platform network access to files used in OEM storage products built on Linux- and Unix-based devices. Whether you're building a cloud storage offering, cloud gateway, a traditional NAS device, or another application where you need to provide secure access to files across a network, Likewise Storage Services can help. To learn more visit www.likewise.com/products/likewise_storage_services.

About Likewise

Likewise makes an integrated software platform, Likewise Storage Services, for identity, security and storage used by market-leading OEM storage vendors including Riverbed, EMC and HP. In addition, Likewise Data Analytics and Governance is an application which helps enterprise IT organizations mitigate risk and drive greater value from their unstructured data. Likewise Data Analytics and Governance ties identity and other contextual data to unstructured data as metadata for better analytics, governance and compliance, entitlement management, and performance management. Likewise enables organizations to provide both access to and control of their data across mixed network environments. More information is available at the company's website, www.likewise.com.

Likewise Software
15395 SE 30th Place, Suite #140
Bellevue, WA 98007
USA