WHITE PAPER (DIGITAL GUARDIAN)

How To Protect Unstructured Data

Identification, Classification, And Management

Complete Data Protection Against Insider and Outsider Threats with One Agent

Introduction

An organization's sensitive information includes intellectual property, trade secrets, consumer information, and financial data. While protecting data from misuse is difficult, unstructured data types are often the hardest to manage. Unlike structured data in databases and fixed formats, unstructured data resides in documents, spreadsheets, and email messages. It can also include image, audio, and video files not organized or segregated within a defined resource such as a file share or document management system. Fortra™'s Digital Guardian® has developed a data protection model to control unstructured data by identifying and applying policy-based controls over the data's lifecycle. Digital Guardian provides organizations with tools to classify unstructured information automatically, analyze its uses and risks, and enforce user-specific policies with precision and persistence.

Unstructured Data: The Hidden Risk

To control unstructured data, one must first identify where it resides and how it is used. A typical approach to information discovery by data loss prevention (DLP) tools is to search for files by content. This "snapshot" approach provides a moment in time inventory that is out of date almost immediately and cannot monitor data use or enforce usage policies. Managing unstructured data requires a methodology that incorporates an understanding of how data is used across the enterprise and applies intelligent controls to secure the data while maximizing its value, including:

- Knowing how sensitive unstructured data is defined and where it is located at all times
- Controlling who has access to sensitive data and how it can be used
- Retaining a complete audit trail of all activities with sensitive data
- Communicating and reinforcing usage policies to end users

Identifying Unstructured Data with Digital Guardian

A data protection solution for unstructured data must include a policy-driven program to classify data at its point of creation or discovery, and ensure its appropriate use at all times. Traditional solutions classify data at rest and rely heavily on network DLP or system "crawlers" to scan servers and workstations using content pattern matching. This often results in false positives (and false negatives), making it difficult to apply policy controls without affecting normal workflow. In addition, network scanning is impossible when users are offline or within a virtual environment.

Accurately and consistently identifying unstructured data, particularly those in unrecognizable forms (for example, pictures or formulae), requires a "reference monitor."[1] In this case, the reference monitor is an agent on the host machine that ensures a file's characteristics are understood in full operational context before classifying the file. Identifying unstructured data using a reference monitor ensures high accuracy whether the user is online, offline, or accessing information virtually.

Digital Guardian's reference monitor agents operate at the kernel of the operating system, allowing full contextual visibility of user, application, and data activities, including the creation of new unstructured data or actions upon existing data. This visibility allows Digital Guardian to classify unstructured data based on contextual parameters such as file path, application name, web application URL, and/ or file content. Since unstructured data is often derived from an application or database (e.g. through a database query), context-based rules are ideal for classifying newly created data. In addition, Digital Guardian supports use of context and content inspection rules simultaneously.

[1] The "reference monitor" security concept was first described in "The Trusted Computer System Evaluation Criteria (TCSEC)" (i.e. The Orange Book), a 1983 Defense Department study which states that policies are most accurately enforced when data is identified within its full operational context.

This is particularly useful when identifying data types that can produce high false positive rates, such as Personally Identifiable Information (PII). The combination of content inspection and contextual identification creates a more accurate method for identifying unstructured data than either method alone, and in some cases it can eliminate false positives altogether.When installed on a host machine, the Digital Guardian agent immediately identifies data in use or at rest. Using Digital Guardian to identify and inventory unstructured data locally provides a substantial performance and cost advantage over remote scanning. It can be configured to scan when the file is accessed, or as a background process during off-peak use. If a Digital Guardian agent is not installed on the host, unstructured data can be evaluated remotely using the DG eDiscovery Agent.

Adding Structure to Data

Classification effectively provides structure to unstructured sensitive data, using the content and context of its creation, access, revision, or transmission. It allows organizations to monitor the data lifecycle and apply appropriate policy enforcement in real time. Accurate, persistent classification supports policies and controls to maximize the value of information use without impeding authorized workflow.

Digital Guardian applies classification tags in real time when unstructured data is created. Once applied, tags propagate from a source file to any new file that is created using any portion of its parent. For example, if a JPEG classified as sensitive is copied to a Word document, the document will automatically inherit the classification of the JPEG. Once classified, Digital Guardian agents on any system will enforce usage policies for each user.

Classifying data with Digital Guardian is substantially more secure and scalable than network-based DLP, which must rescan content as it is used. Classification tags identify data permanently by policy or by a change to the data itself. Tags cannot be obfuscated and persist through encryption, embedding, or compressing data in other formats.

Managing Unstructured Data Usage

Digital Guardian's architecture makes it possible to maximize data sharing while minimizing risk. Its kernel level agents provide visibility and policy enforcement as data is used and where unstructured data is most at risk. Control polices detect risk and enforce controls based on the context of each data transaction. Policies are a collection of alert and control rules based on dozens of factors, including user identity, file source or destination, and network connection. For example, a user may have the ability to copy and save sensitive information in most cases, but have those privileges restricted if they are logged into their Facebook account. Enforcement controls include blocking actions, silent alerting, automatic file/email encryption, user warnings, user prompting, and data masking.

Prompts make users accountable for their own policy compliance. A prompt requires a user to acknowledge policies before allowing a risky action. For example, a user wishing to copy sensitive information to a USB drive will be served a dialog box reminding him or her of company policy and (optionally) requiring justification for the action. The goal is to ensure that users are aware of their role in protecting information and encourage self-compliance. This increases security through greater awareness while substantially reducing support and administrative costs.

Automatic Encryption for Secure Collaboration

One of Digital Guardian's most cost-effective control options in data sharing environments combines classification with policy-based file encryption. Digital Guardian provides integrated encryption for files, network shares, removable devices, and email. Encrypted data can be accessed transparently by authorized users or password-protected for use by authorized third parties on devices that lack a Digital Guardian agent.

Digital Guardian's patented key management system uses a file's classification and its transaction context to ensure access polices are persistently and accurately applied. For example, unstructured data classified as IP can be immediately encrypted when identified. Its decryption rights can be restricted solely to authorized users on Digital Guardian-protected systems. This policy awareness can be used to implement logical file segregation, even if the user is offline or working in a virtual environments, without requiring additional infrastructure or firewalls to support separation-of-duty requirements.

Decision Support & Information Assurance

Digital Guardian combines real-time alerting and forensic-quality reporting to provide enterprisescale visibility into unstructured data usage and risks. Business managers and security staff can be notified immediately if trending and outlier analysis shows changes in the use of classified unstructured data.

It provides evidence-based decision support when measuring policy compliance or verifying the success of new or modified policies. Digital Guardian allows new risks to be identified quickly and accelerates incident response times.

Conclusion

The need for organizations to manage all their highvalue information — particularly that in irregular and unstructured form — is critical to success in highly competitive and regulated markets. Beyond simply knowing where data resides at a moment in time, companies need to know how and under what conditions data is used at all times. This allows control policies that minimize risk while supporting business objectives.

Digital Guardian leverages the full transactional context in which data is used to identify and classify unstructured data when content alone is not reliable. From this classification process, organizations can analyze actual usage patterns to quantify the risk to unstructured data and design appropriate policies.

Digital Guardian's autonomous host agents are based on a "reference monitor" concept to monitor, audit, and enforce policy online, offline, or in a virtual workspace. Digital Guardian allows organizations to adopt a data protection program using a phased approach that respects business processes while discovering, classifying, monitoring, analyzing, and iteratively decreasing the risk to unstructured sensitive information.

About Digital Guardian

At Digital Guardian, we believe in data. We know that within your data are your company's most valuable assets. The sum total of innovations, plans and potential. We protect your company's sensitive information like it's our own so you can minimize risk without diminishing returns.

For over 10 years we've enabled data-rich organizations to prevent data loss at the endpoint. Our expert security team and proven Digital Guardian platform radically improve your defense against insider and outsider threats.

Hundreds of customers across a wide range of industries rely on Digital Guardian to protect their critical information at the point of risk. Seven of the top ten IP holders and five of the top ten auto companies trust us with the integrity of their most valuable and vulnerable data. We take pride in knowing that, at this very moment, Digital Guardian agents are securing the sensitive data of the world's most inventive, influential companies.

FORTRA™

Fortra.com

About Fortra

Fortra is a cybersecurity company like no other. We're creating a simpler, stronger future for our customers. Our trusted experts and portfolio of integrated, scalable solutions bring balance and control to organizations around the world. We're the positive changemakers and your relentless ally to provide peace of mind through every step of your cybersecurity journey. Learn more at fortra.com.