

WHITEPAPER · UNSTRUCTURED DATA SECURITY

Protection and Compliance for Unstructured Data in the Multi-Cloud Enterprise

The Next Frontier of Data Loss Protection

A Confidencial Point of View
CONFIDENCIAL, INC.

confidencial.io

INTRODUCTION

Rethinking Data Loss Protection

Data Loss Protection (DLP) efforts have historically been founded on a broad spectrum of tools and processes used to secure sensitive data from being exfiltrated, misused, or accessed by unauthorized users. DLP software helps enterprises to classify regulated, confidential, and business critical data. It also helps to identify violations of policies defined by the enterprise or within a predefined policy framework. In some cases, the policies are driven by regulatory compliance such as HIPAA, PCI-DSS, or GDPR. Often the policies are driven by trade secrets, sensitive intellectual property, and national security information classified, or are required to be protected by court orders or legal actions.

Once violations are identified, DLP enforces remediation with alerts, encryption, and other protective actions to prevent end users from accidentally or maliciously sharing data that could put the organization at risk. DLP solutions also assist with monitoring and controlling endpoint activities, filtering data streams on corporate networks, and monitoring data in the cloud to protect data at rest, in motion, and in use.

DLP also provides reporting to meet compliance and auditing requirements and identifies areas of weakness and anomalies for forensics and incident response.

INTRODUCTION (CONTINUED)

However, a gap in every DLP solution is the inability to assess and see what the human eye has seen, irrevocably losing visual capture data that would allow an administrator to determine the information that is at risk of being transported and disseminated by the person. There is no current way of determining what they read or saw, how much of it was inspected or understood, and aside from legal means as a deterrent, this data cannot be unseen or stopped from being shared.

Currently enterprises rely on the time-proven method of blacking out words, sentences, or paragraphs that need to be redacted. This "old school" DLP approach does work to some extent, but it comes with the significant drawback that information redacted using this method is redacted for *everyone*, including users that would otherwise be permitted to see it.

What if we could highlight words, sentences, or paragraphs within a document and protect them with **military-grade encryption algorithms? All done at the source copy of the document and with permissions set such that the same document is redacted by encryption differently for each user depending on the user's defined roles, permissions, and policies.**

Those solutions are here and every DLP strategy should embed this as part of their enterprise solution.

01 — LANDSCAPE

State of Data in the Modern Enterprise

80%

of enterprise data is unstructured — PDFs, Office docs, emails, and messages. In regulated industries, this poses a massive challenge for data protection teams.

Recent reports indicate¹ that up to 80% of data in the enterprise is in an unstructured form such as PDF, Microsoft Office documents, emails, and messages in enterprise messaging apps, etc. In regulated industries especially, unstructured data poses a significant challenge to those charged with preventing the leakage or loss of such data. Compounding the problem is the fact that the amount of unstructured data is doubling every two years².

Unstructured data is a concerning source of leaks and data loss, for example, *47% of financial services employees say they have downloaded, saved, or sent work-related documents to their personal accounts before leaving or after being dismissed from a job*³. These figures do not account for the countless number of DLP incidents where data is viewed inadvertently by unauthorized users and later shared.

We expect this problem to become more pressing, especially as enterprises transform to be more distributed in both infrastructure (90% of enterprises use multi-cloud⁴) and operation (with hybrid-office work becoming the norm after the pandemic). Regardless of technological advances, the premise never changes. It's about unauthorized viewing of sensitive data that can't be unseen.

¹ <https://www.fisglobal.com/en/fintech2030/connectivity/unstructured-data-banking>

² <https://www.nanalyze.com/2022/08/investing-in-unstructured-data/>

³ <https://www.tessian.com/blog/insider-threat-statistics/>

⁴ <https://www.forbes.com/sites/googlecloud/2022/03/04/90-of-companies-have-a-multicloud-destiny-can-conventional-analytics-keep-up/>

02 — THE PROBLEM

Challenges of Protecting Unstructured Data

Developing an effective approach to protecting unstructured data, and ensure compliance with security policies and standards, faces the following challenges:

Proactive Protection

The protection approach should be proactive, and not only applied when data loss is happening or detected. A proactive solution prevents data loss from happening in the first place by protecting documents at the time of origination.

High Coverage

To be secure during transmission and storage, encryption should be applied as much as possible, and decryption should only be applied temporarily, while in use by an end-user for processing.

Storage Agnostic

The protection approach, once applied, should work on any cloud storage or transmission platform or medium and protect the documents without an additional burden on admins or users.

Inside & Outside Enterprise

The protection should be seamlessly effective whether the content resides and is transmitted only inside the enterprise, or whether it is transmitted and shared outside the enterprise.

Visibility and Insight

The approach should have built-in features to enable the enterprise to know events that occurred on protected documents and glean insights into enterprise-wide patterns of data usage.

Anyone charged with implementing an enterprise DLP strategy and architecture will confirm that the above-listed criteria are difficult to implement.

03 — WHAT GOOD LOOKS LIKE

Requirements for Adoption

For rapid and effective adoption, the ideal protection solution must have the following characteristics:

Easy Deployment

Seamless integrations with existing enterprise infrastructure and business applications

Low Friction

Minimal user effort, transparent operation, no workflow disruption

Zero Trust

No single point of failure, enterprise-controlled keys, no third-party exposure

Standards Compliant

NIST & ISO compliant, post-quantum ready cryptography

Automation

Built-in automation for high data volumes and complex sharing patterns

Fair Pricing

Flexible, usage-based pricing that scales with enterprise size and needs

⁵ <https://csrc.nist.gov/projects/post-quantum-cryptography>

04 — RETURNS BY ROLE

Benefits to Security and IT, Legal, and Compliance

Proactively and effectively protecting unstructured data as it travels and lives inside and outside the enterprise provides the following benefits and returns to these roles and departments:

CISO, CIO, and Head of IT

C-level executives and directors, especially CISOs and CIOs, and even boards of directors are increasingly paying attention to cybersecurity and need to have periodic visibility (at different time frames) into the state of unstructured documents in their enterprise, including what is shared with whom, inside and outside the enterprise.

Compliance

Concrete demonstration of how sensitive data in unstructured documents is protected. Accurate logs provide situational awareness and help reduce dark data footprint (estimated at 55% of enterprise data in 2022)⁶.

Legal

Defensible proof (backed by cryptography) that certain content was only viewable by certain employees and business partners. Demonstrates authenticity and integrity of documents as they travel.

⁶ https://www.splunk.com/en_us/form/the-state-of-dark-data.html

⁷ <https://www.darkreading.com/risk/most-companies-pass-on-breach-costs-to-customers>

05 — ENTERPRISE IMPACT

Benefits to the Enterprise at Large

Proactively and effectively protecting unstructured data provides significant enterprise-wide returns:

Avoiding Lost Productivity

Hundreds of hours of valuable employee time are regularly lost due to data leaks. Platforms that prevent leaks obviate the need for extensive forensics and assessments.

Avoiding Fines

The average cost of a data breach⁷ reached \$4.4 million globally (13% increase since 2020) and \$9.4 million in the US. Fines continue to increase as more regulations are enacted.

Protecting Brands

Brand damage can cost an enterprise tens or hundreds of millions in lost revenues and spending to rebuild reputations. Prevention is significantly more cost-effective than remediation.

CONCLUSION

Getting It Done

Solution providers that can address the requirements for adoption and provide this much-needed benefit to the enterprise are beginning to emerge, although in a limited manner. The use of military-grade cryptography and easy integration with a broad array of other platforms is a must.

At Confidential.io, we offer the critical ability to enable safe and predictable sharing of information in which privacy is preserved in accordance with the data access, authorization, and sensitivity policies set for the documents and the roles of the viewers. Regardless of whether that information is shared internally across the organization, with third-party providers, or directly with customers, information such as sensitive internal documents, IP, PII, PCI, HIPAA, legal mandates, court orders, national security guidelines, or competitive advantage IP, the data must be protected.

By using policy-based encryption technology developed within DARPA, Confidential enables compliance-based, cryptographically-enforced access to different sections within a sensitive document. Routine information remains viewable, but sensitive information can only be viewed by the individuals, groups, or roles you specifically authorize, allowing the document to be shared widely, yet protected with exquisite encryption.

CONFIDENTIAL

The file is the perimeter.

Protect your crown jewels wherever they go — with policy-based, cryptographically-enforced access that travels with every document, inside or outside the enterprise.