Confidencial

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

The Next Frontier of Data Loss Protection

April 2023

Introduction

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.

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.

State of Data in the Modern Enterprise

Recent reports indicate[1] 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

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

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 years[2]

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-cloud[4]) 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.

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. Protecting the data at the source and when the document is created.
- **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, then encrypted again upon closing or before sharing.
- **Agnostic to Storage or Transmission Mechanism:** 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 and Outside the 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 also have built-in features to enable the enterprise to know events that occurred on protected documents (whether they occur inside or outside the enterprise) and be able to glean insights into enterprise-wide patterns of data usage; this will facilitate compliance and auditability of the enterprise.

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

Requirement for Adoption

[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/?sh=794ce10d5d89

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

- **Easily and Quickly Deployable:** Out of the box, seamless Integrations with existing enterprise infrastructure such as IDP and IAM tools (e.g., Active Directory, Azure AD, and Okta) and business applications and platforms (e.g., MS Office, PDF viewers and editors, email clients, messaging, and collaboration platforms) is essential to efficiently apply protections and demonstrate effectiveness to encourage adoption across different industries and jurisdictions.
- **Low Friction Usability:** For widespread, regular adoption, the solution should impose minimal additional effort from end-users (if they must be involved at all). This approach should be completely transparent and take place in the background. If an end-user must be involved, it should only require a few clicks and not disrupt their workflows or require them to learn new, complex tools or procedures. This is especially true for workflows that involve collaborators outside the enterprise – end-users must be able to share documents with external collaborators without a reduction in security or an increase in complexity.
- **Zero Trust:** No single entity (services, servers, users) should be a single point of failure, and enterprise content should never be stored or exposed to third parties, including the solution provider itself. Also, if any cryptography is used, keys should be under the control of the enterprise and never reside with the content (as is, unfortunately, often the case).
- **Compliance with Standards:** When cryptography is used, it should be used as a building block. It should be compliant with existing national and international standards from well-known organizations such as NIST and ISO, and easily upgradable to be post-quantum to ensure a seamless transition to newly developed standards[5].
- **Automation:** Various forms of automation must be built-in into the protection approach to keep up with the high rate of unstructured data generation in the enterprise, and the large number of entities such data could be shared with (inside and outside the enterprise). Automation is essential to be able to protect the Petabytes of historical and archived data large enterprises typically store.
- **Fair Pricing:** The pricing should be flexible and able to accommodate various sizes of enterprises, and be charged based on usage, required features, and integrations. None of us can overlook that price and affordability are consistent factors that must be considered.

Benefits to Enterprise 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:** Such departments will be able to concretely demonstrate how sensitive data in unstructured documents is protected inside the enterprise and even as it is shared outside it. Accurate logs will provide situational awareness of who had access to which

[5] https://csrc.nist.gov/projects/post-quantum-cryptography

data, and if automated destruction is applied, it can help lessen the footprint of dark data, which is an increasingly challenging issue enterprises are grappling with. (estimated to be 55% of enterprise data in 2022)[6].

- **Legal:** Legal departments will have defensible proof (backed by cryptography), ensuring that certain content (even at the object level inside a document) was only viewable by certain employees and business partners, and can also demonstrate the authenticity and integrity of documents as they travel inside and outside the enterprise.

Benefits to the Enterprise at Large

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

- **Avoiding Lost Productivity:** Hundreds of hours of valuable employee time are regularly lost due to data leaks, especially in regulated industries such as defense, banking, and biotech. Such time is spent on forensics and assessment, investigations, sanitizing existing devices and infrastructure, or setting up new ones. Platforms that can prevent such leaks will obviate the need for such forensics and assessments, or at least lessen the forensics effort required by providing visibility and monitoring capabilities.
- **Avoiding Fines:** The average cost of a data breach[7] continued to rise in 2022 and has reached an average of $4.4 million globally (13% increase since 2020) and is now $9.4 million in the United States. Fines are only going 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.

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 Confidencial Inc, 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, Confidencial 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.

---

[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