FORTRA

WHITE PAPER (Titus)

Improving DLP Accuracy with Data Classification

Rich, Persistent Metadata From Titus Enables DLP Solutions To Make Better Policy Decisions.

Data security is on every IT professional's mind these days, regardless of industry or geographical location. Given the increase in remote workers, the intricacy of most organizations' IT infrastructure, and the number of different types of devices connecting to network resources, information security threats are an ever-present reality. The fact that we are creating data at warp speed and that data is circulating everywhere inside and outside of organization walls only compounds the situation. Ensuring that sensitive personal information and critical business data doesn't fall into the wrong hands is a constant job.

Threats can stem from a few different sources:

- Sensitive attachments get distributed that should have been kept confidential.
- Documents get saved on shared folders where they can be easily viewed by anyone.
- Emails get sent to the wrong people.
- Ransomware and criminal hackers seek to steal data for illegal intent.
- Bad actors within an organization sometimes gain access to information they really shouldn't have access to and use it for malicious purposes.

In an effort to curtail some of these threats, new data protection and consumer privacy regulations are emerging across the globe, from the California Consumer Privacy Act (CCPA) in the U.S. to the General Data Protection Regulation (GDPR) in Europe, and the Lei Geral de Proteção de Dados (LGPD) in Brazil. All are aimed at keeping citizens' sensitive personal information secure as well as protecting proprietary business information from being hijacked for illegal use.

But keeping information protected and adhering to strict privacy regulations requires organizations to better understand the types of sensitive data they have and where it resides so they can then take steps to prevent data loss. These tasks can be daunting at first glance, especially if you think about trying to manually cull through terabytes of data to identify the most sensitive.

How DLP Technologies Can Help

Fortunately, there are technologies to help with the process of data identification and ways to automatically restrict data from unintentional exfiltration and from falling into the wrong hands.

Data loss prevention (DLP) technologies keep an eye out for sensitive data leaving an organization and alert users. These solutions can be configured to:

- Scan documents for sensitive data
- Halt exfiltration of specified information types
- Implement basic data handling policies

DLP solutions work by scanning data to look for designated word or number patterns that might indicate sensitive information.

Gaining momentum in the mid-2000s, DLP technologies provided organizations with a way to protect their intellectual property and other data at the endpoint, lessoning the burden on users to ensure sensitive data stays within the organization.

Over the years, as cybercrime increased and data became a hot commodity, DLP technologies evolved to allow organizations to apply labels to files and instate overarching data handling rules as protection against a breach.

DLP technologies can certainly implement your organization's data policies to a certain degree. For example, if an HR employee writes an email to a new hire and attaches a contract that contains the new hire's social security number, the organization's DLP software would quickly scan the email and its attachment for sensitive patterns. Upon discovery of the social security number, the DLP solution would alert the user before allowing the document to be sent.

DLP Does Have Limitations

In the scenario above, HR would likely not want to send the new hire's social security number out into cyberspace unprotected, but they probably might want that contract to get to the new hire. They would have to find a secure workaround, but they probably would want that contract to require additional technologies. All in all, the process could be frustrating both for the HR employee and the new hire waiting to review the contract.

While DLP solution providers have been improving their offerings over the past few years by incorporating cloud access security brokers (CASBs) and encryption capabilities, their data classification features are still very rudimentary. Several limitations persist:

- Time-consuming, cumbersome policy implementation
- Rigid labeling structure offering only two levels of detail
- Cannot understand data context
- Too many false positives on sensitive data

Implementing detailed organizational data policies within a DLP solution can be time-consuming, cumbersome, and often insufficient. Most bundled classification capabilities are more like simple labeling features with only one or two levels and a rigid structure. While DLP technologies can recognize unstructured data to some degree, they lack contextual understanding. Context accuracy is highest when data is being created — the creator usually understands the data better than anyone. DLP technologies, however, are typically used to scan data that has moved beyond the creation stage.

Most DLP technologies use regular expressions, or regex, as their method of search. Sort of a precursor to machine learning, regex are algorithms that look for a sequence of characters or common patterns in content, such as social security or passport numbers. Regex can identify specified words and phrases but can't understand grammar, syntax, or context. This method of search-and-replace will inspect your data character by character looking for a pattern that has been requested, as in a search-and-replace function.

Context is critical, however, to understanding how specific information should be used. Every organization creates data that is unique to its business, including intellectual property, R&D information, customer data, and more. DLP technologies cannot differentiate between types of information or understand how that information is used within an organization.

Without that unique contextual understanding, DLP technologies can falsely assume that certain data is confidential, also known as "false positive data identification." Users are then either alerted that they should take a look at the data and consider not sending it outside of the organization, or the DLP technology may restrict its movement entirely.

For example, say an organization has a mergers and acquisitions (M&A) project code named "Chicago," with documentation that includes stock prices, dates, dollar amounts, and names of senior executives. The IT team wouldn't do something so drastic as to configure their DLP solution to flag or block emails that contain the word "Chicago," but it would be a significant challenge to program the DLP to accurately distinguish sensitive M&A documents from, let's say, regular business emails and documents pertaining to executives flying to Chicago for a trade show.

Overly sensitive technologies are obviously disruptive to users' workflow and can be a cause of frustration. These technologies may also fail to recognize and protect some sensitive data, resulting in "false negatives."

Privacy regulations such as the GDPR, CCPA, LGPD, and others add another layer of complexity to the story, requiring companies to know granular details about the data they have stored in their systems. Unfortunately, DLP technologies on their own have a limited ability to track and understand unstructured data on such a detailed level. DLP solutions can look at content and analyze it, but the technologies stop at validation. They simply cannot offer the deeper contextual understanding around the data without access to rich, persistent metadata.

Because of these limitations, a significant number of people using DLP technologies have become frustrated enough to switch to Monitor Mode, where IT teams disable the enforcement of policies for end users and use simple Log Reporting to keep track of what's being shared.

> "Without that unique contextual understanding, DLP technologies can falsely assume that certain data is confidential, also known as "false positive data identification."

How Titus Complements and Improves DLP

Securing data cannot be achieved by the act of one technology. The growing volumes and diversity of data require a seemingly disparate set of purpose-built tools working together to make data security achievable. Working together requires industry collaboration and technological integration to create a more valuable security ecosystem.

> "Titus enables DLP technologies to piggyback onto Titus policies, which offer deep customization options in order to accommodate virtually any requirement."

Many DLP solution providers have now recognized that their enterprise-level customers need more flexibility and specificity when it comes to data classification. These DLP companies are now partnering with Titus to apply context to data and enable more accurate and effective data security decisions.

Titus complements DLP technologies in several ways:

- Understands context based on machine learning algorithms and user feedback
- Adds intelligent metadata that your other security solutions can access
- Applies metadata to define context of information
- Offers flexible schemas for unlimited, customized data classification and information handling policies

Titus adds much needed intelligence to DLP products by balancing machine learning–based automation and user feedback with expert systems and pattern-matching technologies to determine context and apply detailed metadata. The application of metadata at creation makes data identification much more accurate. Trying to look at a document down the road to determine its sensitivity level can be more challenging, especially for users other than the original document creator. Classifying data at creation empowers organizations to move from watch mode to true data protection mode where data privacy and security policies can be applied more strategically.

In essence, Titus enables DLP technologies to piggyback onto Titus policies, which offer deep customization options in order to accommodate virtually any requirement, even in the most modern enterprise or governmental environments. By providing context and more granular details about your data, Titus metadata enables DLP solutions to make better policy decisions.

As discussed earlier, false positive identification of sensitive data can disrupt business and overwhelm your users with unnecessary alerts and blocked emails that might lead to employees downgrading their security policies or trying less secure workarounds to get the job done. Titus adds intelligence and context through its metadata so that your other security investments can work more accurately and effectively.

Titus' virtually unlimited schema flexibility allows you to custom build your data security strategy around your existing processes and workflows. The Titus policy engine specializes in identifying sensitive data and can orchestrate security actions across a number of data conditions (content, author, etc.) and workflow events (send, save, open, etc.). If certain conditions are met within a file's metadata, Titus can also set policy to instruct a third-party application to encrypt the data as it moves outside of the organization.

A DLP policy engine on its own doesn't offer true data classification; rather, it offers rudimentary tagging capabilities without applying the information as persistent metadata. Titus can scan for, and embed, an unlimited number of data details. Setting actions around product release dates, customer information, and other parameters allows Titus policies to be exponentially more sophisticated in the way they handle information.

For instance, a DLP solution might be able to label a document as "Confidential," but the label itself doesn't contain enough detail or nuance when implementing security policies or taking an action. The document could contain sensitive HR information, proprietary R&D details, financial information, or a range of other types of confidential data. Titus metadata adds that layer of context, which instructs all of the players in the security ecosystem and enables them to take more granular actions in terms of handling instructions.

> "Metadata is essentially data about the data, and properly applied, it works to inform all of your other security solutions."

Titus offers a far more robust policy engine driven by machine learning that can provide automated context for a wide range of data types.

How Titus Metadata Informs DLP Technologies

Metadata is essentially data about the data, and properly applied, it works to inform all of your other security solutions. The ability to parse such granular details about data is what allows a wide range of security solutions to make use of it.

Metadata doesn't need to use the word "confidential" to indicate sensitivity level. Instead, IT admins and data stewards can write metadata using language that can't be easily discovered by unauthorized people or systems — sort of a secret code that's unique to the business. DLP technologies add a layer of protection by reading that metadata and stopping any exfiltration when certain secret code words are encountered.

You might think that unlimited levels of metadata could make a data protection solution overly complicated for users. The beauty is that users are typically only seeing a few select labels as they interact with data. Metadata is primarily a machine-readable format that works in the background to inform other systems to use the extra information to drive more granular policy actions and remediations.

If an email contains sensitive data or has an attachment that contains sensitive information, Titus will alert the user before allowing it to be sent. At other exfiltration points, however, such as uploading a doc to the cloud or downloading a file to a USB device, an organization's DLP technologies can read the Titus embedded metadata and either alert the user to the sensitive information or stop the exfiltration.

Titus' flexible schema design enables organizations to design a solution using detailed data policies to fit their specific needs. A law firm would need a different set of vcompany or a healthcare organization. In addition, that flexible schema gives users various levels of interaction. In some cases, users may not have to interact with the classification process at all, instead relying on Titus automation capabilities. But in organizations where information can fall into multiple categories with multiple levels of variation, or in organizations where fully automated systems are prohibited (e.g., military or government), users may require numerous drop-down menus to help them choose the appropriate labels and more granular sensitivity levels. Titus schema allows for a virtually unlimited number of menus and fields.

Schemas evolve over time, and organizations can refine their policies as business requirements evolve. In some cases, it makes sense to introduce a data classification initiative gradually with simple options so as not to overwhelm users. This approach allows you to educate users about which data types should be kept internal and what's okay to be sent externally. Over time, you can adjust the schema to parse more granular levels of differentiation among data sensitivity levels. As users work with the solution, selecting from the drop-down menus, they are also informing the policy engine and eventually it can do much of the work in the background.

Build Your Security Ecosystem

One of Titus' core business values has always been to create technologies that foster an open ecosystem. Titus understands that data identification and classification solutions are part of an overall data protection strategy.

The agnostic and open design of Titus' data protection solutions underpins their unparalleled interoperability. Titus acts as the policy broker at the heart of your security ecosystem, applying the rich, persistent metadata that enables all your data protection solutions to work better together — DLP technologies as well as CASBs, firewalls, encryption technologies, and others. This gives organizations the flexibility to use the technologies of their choice to guard against security threats, stop unintended exfiltration of sensitive data, and ensure compliance with data privacy regulations.

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.