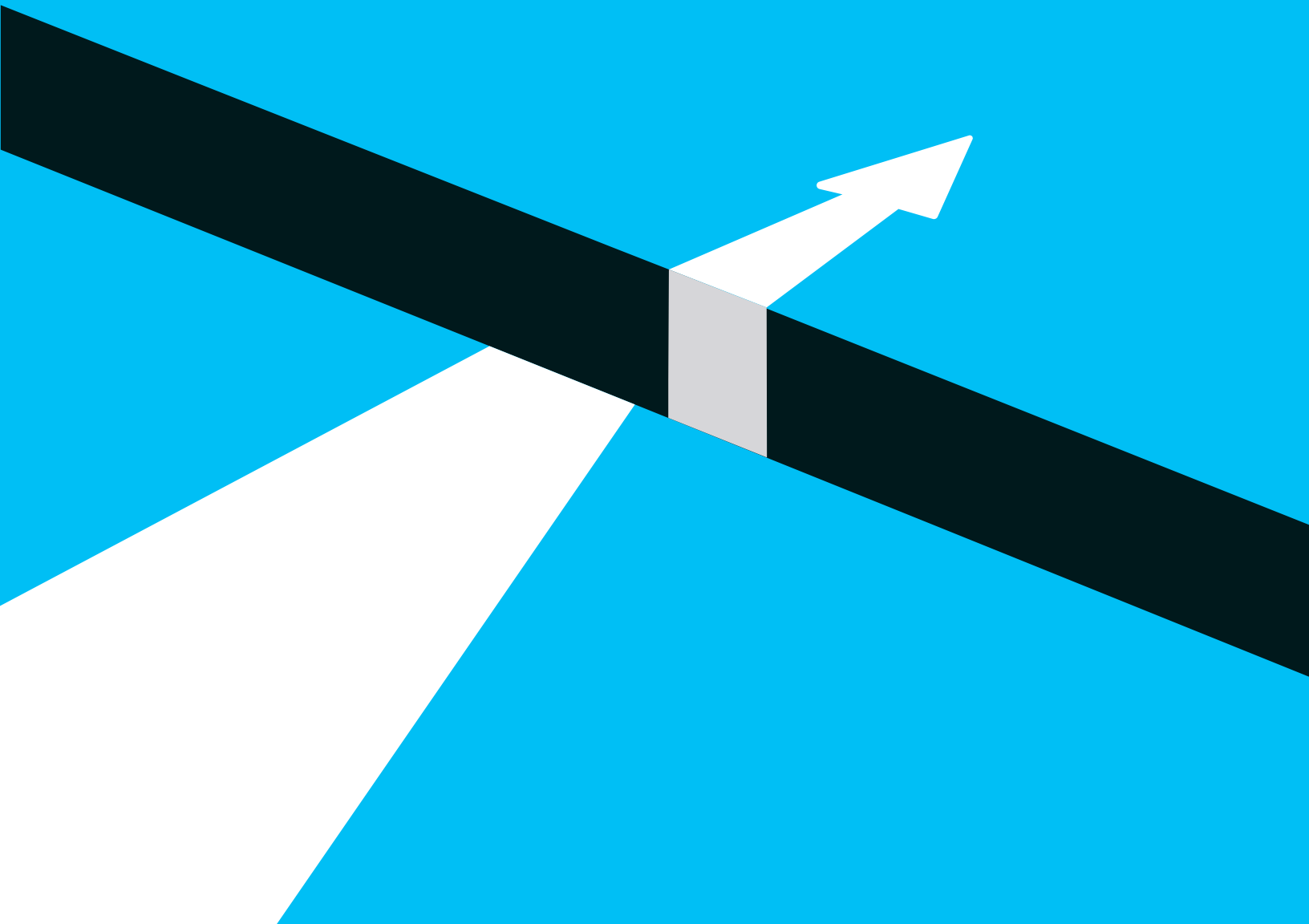


CYBERHAVEN

ADDRESSING THE TOP **5 GAPS** IN **DLP**



Data loss prevention (DLP) has a well-deserved reputation for being challenging to implement and largely ineffective for securing an organization’s sensitive data. Enterprises regard it as a checkbox tool for compliance, a necessary evil that has low accuracy, is difficult to configure, annoys users by interfering with their work, and can make systems slow and unstable.

While some enterprise security teams have grown accustomed to these limitations, the problems of DLP have only gotten worse in recent years. The shift to the cloud has transformed how enterprise data is created, shared, and ultimately stored. At the same time, Shadow IT, and collaborative and remote work has made data sprawl even more difficult to control, amplifying risks that DLP was never designed to address in the first place. Collectively, these issues have magnified the traditional problems of DLP tools, making them even more ineffective while their policies have grown ever more complex.

These are the top 5 challenges with implementing DLP, according to the Insider Threat Report from Cybersecurity Insiders:



DLP's reputation aside, the goal of preventing data leakage remains as critical as ever in our hyper competitive economy. What every company must achieve is abundantly clear—maintaining control over how sensitive data is shared and used with as little overhead and user friction as possible, and preventing breaches of both IP and PII data wherever it is located. Moreover, leaks of high-value data can cause substantial brand damage and/or degradation of competitive advantage, making companies increasingly interested in DLP for more than compliance.

Cyberhaven has pioneered a new approach known as Data Detection and Response that allows organizations to simply and easily protect any type of data or content, anywhere in the enterprise. Built on innovative data tracing technology, this approach enables security teams to address each of the five main gaps in the prior generation of DLP products.

Dynamic Data Tracing automatically records and reports on all data movement without data classification or tagging. Continuous, complete, and contextual visibility into the behavior and movement of all data, across on-premises and cloud environments, delivers a transformative level of data protection for any organization.

The TOP 5 DLP Gaps

-  Complex and incomplete policies
-  Constrained by data classification
-  Limited visibility
-  High false positives
-  Not user-friendly



GAP 1

COMPLEX AND INCOMPLETE POLICIES

Creating DLP policies requires you to predict the future—you need to know what the data that you are trying to protect looks like (i.e., what patterns it contains), which channels will be used for exfiltration, and under what conditions the data is allowed to be shared externally.

A typical DLP implementation can take 6 months or more of tuning before going live, and often the resulting policies are purposely relaxed to avoid triggering too many false positives. Unsurprisingly, the demand for managed services for investigating DLP incidents is growing steadily.

As an example, consider a seemingly simple scenario: preventing documents containing sensitive IP, such as designs for a new device or a new vaccine formula, from being shared externally. Step 1 is to identify the textual pattern—you might look for “confidential” markers or the IP’s codename in all document types supported by DLP (MS Office, PDF, text, etc.). Step 2 is to figure out which channels are allowed to send this info outside the organization, by which users, and under what circumstances. Step 3 is to determine how to deal with exceptions. There are always legitimate scenarios for sharing data externally, such as for an upcoming product launch or when collaborating with supply chain partners, and you do not want to trigger an incident alert each time there is a content match. Thus, one exception might be based on allowing a particular user group to share externally, and another on the volume of matched documents in a particular email or web form post. And with every change in business workflow (e.g., new project code names added to the same sensitive IP or a new cloud application for editing files), you will likely have to create a new exception. With increased collaboration, it’s harder than ever to know with certainty the content of the sensitive data to protect and how to tune DLP policies to establish a low rate of false positives.

More importantly, it is very hard to test DLP policies. The knobs provided are essentially just the content of the file and the exfiltration channel; the absence of additional context makes policies very complex, as they must overcompensate to differentiate between legitimate sharing and a data breach. Moreover, the security team does not have a bird’s-eye view of data flows or the ability to look at historical user and data activity in order to test their policies before deploying them.



GAP 2

CONSTRAINED BY DATA CLASSIFICATION

DLP uses content inspection to identify and classify sensitive data based on pattern matching or fingerprinting. When that does not work, data has to be manually tagged.

Content inspection is rarely useful for identifying intellectual property, though. One of the main challenges is the difficulty of keeping the patterns in sync with the underlying IP, which is constantly evolving and growing. Furthermore, DLP content inspection only works for a few file types (PDF, most common Office formats, images), while others are simply ignored - Apple Pages files for example. Even simple patterns like a physical mailing address often lead to false positives.

At least in principle, manual tagging is a good solution to the limitations of content inspection. However, it has two main problems. First, the tags are often lost or incorrectly propagated when users handle the data, for example uploading it to a cloud storage service or converting it to other formats. Second, tagging often ends up relying on content inspection anyway to prompt users to manually apply a certain data sensitivity tag. This process often results in users unsure of what classification tags to use, and the workflow is error-prone, inconsistent, and completely futile against malicious insiders.



GAP 3

LIMITED VISIBILITY


Existing DLP solutions offer very limited visibility into all of the data locations and movements within the modern enterprise. These legacy products merely watch select egress channels; they do not evaluate any pre-egress activities, and they cannot track IP in the cloud applications that hold an ever-growing volume of enterprise data.

DLP products inspect data when it traverses an egress point, such as email being sent to an external address. They cannot evaluate pre-egress risk factors such as high data sprawl, users hoarding sensitive data on their endpoints, and unauthorized internal access.



The number of data egress channels in a typical medium-sized enterprise is overwhelming – thousands of users interacting with thousands of endpoints and hundreds of cloud apps. DLP products attempt to widen their coverage by tapping into network, email, endpoint, and some SaaS traffic. The security team typically deploys several DLP products to cover these various channels, and unfortunately even after all of them are deployed, many gaps remain.

The various tools deployed to handle the many egress channels include email DLP, endpoint DLP and network DLP.

| PRODUCT | GOAL | ISSUES |
|---|--|--|
|  Email DLP | Prevent data exfiltration through corporate emails and attachments | No visibility into personal email and webmail |
|  Endpoint DLP | Prevent data exfiltration for any endpoint application | Performance and stability problems with user machines due to content inspection of each file |
|  Network DLP | Prevent data exfiltration of anything going across the email and webmail | Only works when endpoints are connected to corporate network; loss network that is not caught by email and endpoint DLP of data semantics and understanding when reverse-engineering a stream of packets |

Some DLP products go further in an attempt to acquire additional context by monitoring user activity in the associated apps (email, browser, instant messaging, file explorer) under the assumption that data leaks do not happen in discrete isolated events. They face two major technical issues, though:

VISIBILITY IS STILL LIMITED:

DLP products can't see what happens to data once it is accessed by an application. Can that application send the data outside the enterprise? Can it transform the data into a format that is difficult for the DLP product to parse?

USER FRICTION:

User endpoints become slow and unstable because these products hook into the operating system and end-user apps, frustrating users who just want to get their job done.



DLP was designed before the advent of the cloud, so Cloud Access Security Broker (CASB) products have assumed responsibility for governing data that goes to cloud applications. While CASB products can provide visibility into corporate cloud usage, they lack context about the data's origin and cannot track data across different SaaS environments. To fill this gap, CASB products are often integrated with traditional DLP for content inspection and thus inherit all of their limitations around accuracy.

Any of these DLP blind spots can be easily exploited by motivated malicious insiders. And unfortunately, careless employee acts can be equally costly to organizations when coupled with external threats like phishing. Tracking data automatically across all cloud services and endpoint apps could address the visibility gap that DLP is struggling with.



GAP 4

HIGH FALSE POSITIVES

High false positives rates have long been associated with DLP technology. While vendors do not publish benchmarks on their accuracy, the number of methods implemented over the years to improve accuracy is revealing - fuzzy matching, fingerprinting, exact partial matching, policy tuning, and exception management, just to name a few. Of these, the most promising is exact matching (for instance, matching exact partial records such as client IDs from a DB containing customer information), but this technique only works when you can precisely identify sensitive data and hardly works for CAD files, other non-text data, or any file format that DLP is unable to decipher.

A false positive rate of 10% is considered very good for a DLP policy, yet achieving that is often very complex or even impossible. To meet compliance requirements, low thresholds may be set to trigger alerts on everything that could possibly indicate a violation of policy or data egress. If an enterprise has 50 policies enabled and each policy fires 10 times per day (a conservative estimate), a 10% rate means 50 false positives per day, and each can take hours to investigate properly. As a result, many alerts are ignored because there are simply too many for the investigation team to process. The temptation to ignore alerts because so many of them are false positives is among the top reasons why current DLP technology, despite years of incremental improvement, does not effectively prevent most data leaks.



GAP 5

NOT USER-FRIENDLY

Organizations understandably resist rules and policies that employees perceive as inhibiting their work, and 23% of respondents in the Insider Threat Report acknowledged that DLP initiatives “impede employee productivity and collaboration.” DLP is synonymous with blocking in many scenarios, and when users are blocked from sending or sharing information, they find workarounds based on social media apps or the thousands of shadow IT cloud services that help them be more productive. This trend has only accelerated as the COVID crisis has forced employees to work from home.

Another point of user friction is the fact that widening DLP coverage makes it more intrusive and unreliable. For instance, DLP endpoint agents hook into the operating system kernel and into applications themselves, making endpoints brittle; the result is unresponsive applications, incompatibility with OS upgrades, and blue screens. Thus, such in-line hooking and blocking remains a feature that is demonstrated by vendors in POCs, works in a highly controlled environment, and quickly gets disabled in production after a flood of user complaints.

For an effective DLP program, these are important questions to ask:

Can you create policies quickly?

Do you have access to all the context you’d wish for when creating DLP policies?

Do DLP false positives get in the way of employee productivity?

Do your DLP policies have negligible false positives rates?

Does DLP cover all data in the cloud and on employee endpoints?

Do your DLP policies require months of tuning?

ADDRESSING DLP GAPS WITH DATA DETECTION AND RESPONSE

Cyberhaven protects high-value data through out the data lifecycle while gaining insight into how that data is used and where it resides, using a new approach called Data Detection and Response (DDR). Cyberhaven's DDR monitors all your data across on-premise and cloud environments, enabling you to automatically detect improper handling without the hassle of tagging or classification. Cyberhaven extracts and records metadata from each user interaction so that you can inspect any data flow in real time or retrospectively with just a few clicks. The result is effective data protection without impeding employee productivity and dramatically reduced incident investigation time and cost.

In addition to monitoring, Cyberhaven also implicitly classifies data based on what really matters when writing data leak policies or when doing data leak and insider threat investigations. Rather than relying solely on content inspection, Cyberhaven provides a rich, automatically inferred context around each data flow, including how the data was created and where it was stored throughout its lifetime. The result is transformative effectiveness in data protection for modern organizations.

Cyberhaven does not have any of the limitations of DLP discussed above:



NO COMPLEX AND INCOMPLETE POLICIES: with Cyberhaven you don't need to write policies ahead of time. And the policies you do need are easy to write and well-defined.



UNCONSTRAINED BY DATA CLASSIFICATION: Cyberhaven uses dynamic data tracing rather than scanning content.



UNLIMITED VISIBILITY: Cyberhaven monitors and controls the flow of sensitive data through SaaS apps, endpoints, and email, starting from creation through egress.




NO FALSE POSITIVES: Cyberhaven has zero false positives. By design. Period.



USER-FRIENDLY: Cyberhaven is designed to avoid crashing computers or applications, and because it does not have false positives, it does not get in the way of productive employee work.

CYBERHAVEN: DATA DETECTION AND RESPONSE VS DLP

| Cyberhaven: Data Detection and Response |  | DLP |
|---|---|---|
| Not required | Content inspection | Extensive and imprecise |
| Gradual and education-centric | Remediation | Rigid and disruptive |
| Unified (on-premises & SaaS) | Coverage | On-premises only / siloed / egress-only |
| Contextual (source, destination, people, action, content) | Policies | Content-based, egress only |
| High fidelity, 100% accurate | Alerts | Low quality, "alert fatigue" common |

KEY BENEFITS OF CYBERHAVEN

1

Delivers comprehensive visibility by tracing all data activity automatically, across all data silos (including endpoints, email, and an ever-growing number of SaaS applications).

2

Creates a complete and contextual journal of how each piece of sensitive data was accessed and exposed.

3

Educates users in real time with configurable and escalating enforcement actions.

4

Visually highlights unexpected destinations for your sensitive data, to accelerate investigations.

5

Monitors sensitive data movements within cloud-based applications and platforms such as Office 365, SharePoint, G Suite, and Salesforce.

6

Makes deployment of sensors easy and rapid and avoids data classification.