TERAMIND

More Than Just a Data Problem:

The Future of DLP is User-Centric

Fully securing organizational data in the enterprise remains an elusive challenge.

The process covers data classification, making architectural changes to separate sensitive data from the rest of your network in strategies like network segmentation, leveraging tools available within SaaS platforms such as data tagging, and more. Enterprise IT security teams require compliance with such processes as a rule. Yet, almost daily, another story emerges of how an enterprise has experienced significant data loss through an insider attack, leakage, or outright theft.

Data Classification Makes Traditional DLP Difficult

CISA instructional guides and Enterprise Information Management principles clearly establish the first step of organizing and protecting enterprise data as classifying it. That's absolutely correct. The next step is creating secure places in the network for sensitive datasets to live, and adding security controls to protect it.

Given the ever-changing nature of data, though, classifications and taxonomies begin to decay almost as soon as they're published. Reality is that the more productive employees are, the more additions and changes they will make to existing data. Documents are created, emails are sent, records are accessed, processed, updated, moved. Large organizations are changing their data constantly and at scale, and throttling the pace of activity to keep classifications fully up-to-date is wishful thinking at best. This is why data identification tools were created to assist with DLP. Even with modern scanning and tagging capabilities, however, there will still be opportunities for data to walk out the door.

IANS analysts outline some of those challenges:

Thinking any data loss prevention solution is a silver bullet and that all paths to data exfiltration are closed is a potential pitfall. There are always ways to exfiltrate data. From camera phones, hidden cameras, and narrative clips to simply memorizing it and writing it down later – there is always a way. The goal is to control access, segregate classification levels and maintain proper processes to handle generated or obtained data, its storage, processing, visualization, and disposal.

The problem is not a lack of enterprise strategies. Rather, it's the fact that much of an organization's data loss is not due to a lack of effort or security tools, but a gap in how most people manage users. If you ask your security team, they will likely tell you that their biggest fear is not a failed tool, but an accidental or malicious action taken by users. Mistakes happen, accounts get compromised, users make non-compliant choices sometimes – and those actions can be very difficult to detect. This is what keeps CISOs up at night.

DLP is a User Problem as Much as It is a Data Problem

The late Kevin Mitnick, a renowned malicious hacker turned successful cybersecurity awareness training mogul, based his life's work (on both sides of the ethical fence) on the foundational belief that cybersecurity issues are not a computer problem, but a human one. This brings us to the point that a critical examination of traditional methods of protecting data from loss and leakage, as well as a close look at emerging technologies rooted in a user-centric approach, will benefit organizations interested in more comprehensive DLP strategies.

How Thought Leaders View DLP

Business and Technology analyst Samuel Greengard offered an overview of the state of DLP, what's missing, and where innovation can lead in his article "Cybersecurity Gets Smart."

He notes plainly that the "traditional approach of using signature-based malware detection, heuristics, and tools such as firewalls and data loss prevention (DLP) simply is not getting the job done. … Traditional security methods aren't keeping up with cyberthieves. AI methods such as 'big data, pattern mapping and matching, cognitive computing, and deep learning methods that simulate the way the human mind works' are being explored by researchers as ways to defend information resources. 'The goal … is to better identify suspicious patterns and behavior'."

Additionally, Greengard makes the point that "manual approaches and signature-based approaches are no longer effective because of the large and increasing number of threats. Problems include the growing prevalence of zero-day attacks … polymorphous malware … viruses, Trojan horses … and graphics processing units. In addition, firewalls have become less effective as cloud computing and APIs string together data across enterprise boundaries. '[S]ecurity threats ranging from social engineering … to botnets … [are] more difficult to pinpoint and block because they use cloaking techniques and alias IP addresses'."

In other words, counting on knowing which IP address attackers are coming from, blocking websites, or trusting EDR (endpoint detection and response) tools to prevent malicious activity that can result in data theft is a fool's errand. It's not that those layers of security don't have a place in an enterprise security strategy, but a case where attackers can skirt those tools far too easily, because they are measuring and blocking based on data that is static at worst and not granularly user-centric at best.

Greengard goes on to point out that techniques being explored include “[c]ognitive computing … using … natural language processing to analyze code and data on a continuous basis.

As a result, it is better able to build, maintain, and update algorithms that better detect cyberattacks, including Advanced Persistent Threats (APTs) that rely on long, slow, continuous probing at an almost-imperceptible level in order to carry out a cyberattack.”

What Shifts Are Necessary?

McKinsey and Co analysts also weigh in on DLP challenges in The Future of Data Loss Prevention. Their perspective is that if user data is available, it should absolutely be leveraged in one’s DLP strategy:

> *Leading organizations with access to large data sets and strong capabilities in machine learning have begun using contextual heuristics (for example, log-in time, user behavior, and mouse movements) to identify, flag, and characterize potentially malicious activity. This approach entails collecting data from multiple endpoints, passing it through behavioral-analytics tools to identify anomalous behaviors, and inferring contextual information such as intent, secondary actors, and root causes…*
>
> *To implement this capability, organizations must have sufficient telemetry to collect data across the technical estate as well as advanced analytics. While many vendors offer nascent versions of these capabilities as part of their DLP tools, only high-tech organizations are seriously exploring this functionality today, mostly using custom-built solutions.*

The Problem With Users

A market is emerging around AI and behavioral heuristics, to be sure, as experts work to solve the user challenge. As these technologies evolve and the call for a new approach to DLP continues to grow, another question arises as central to protecting data in the context of users. How does one enable employees to be effective and highly productive with the data they must access to accomplish tasks, while addressing the challenge of securing data? This would involve understanding what normal looks like, and being able to see deviations before they become a serious incident.

So then, there seems to be a need to establish what one means by “normal” behavior, and what it looks like when someone becomes a risk to their organization. This effort, as security teams work to identify those threats early, means understanding what an at-risk user looks like, their psychosocial indicators.

How Do You Know When a User is Becoming a Threat?

In their article titled Psychosocial Modeling of Insider Threat Risk Based on Behavioral and Word Use Analysis, social scientists Frank Greitzer, Lars Kangas, Christine Noonan, Christopher Brown, and Thomas Ferryman outline key psychosocial signals associated with people likely to become a threat to their organization (see Table A).

Table A

Indicator	Description
Disregard for Authority	The employee disregards rules, authority or policies. Employee feels above the rules or that they only apply to others.
Disgruntlement	Employee is observed to be dissatisfied in current position; shows chronic indications of discontent, such as strong negative feelings about being passed over for a promotion or being underpaid or undervalued; may have a poor fit with current job.
Anger Management Issues	The employee often allows anger to get pent up inside; employee observed to have trouble managing lingering emotional feelings of anger or rage; hold strong grudges.
Confrontational Behavior	Employee exhibits argumentative or aggressive behavior or is involved in bullying or intimidation.
Disengagement	The employee keeps to self, is detached, withdrawn and tends not to interact with individuals or groups; avoids meetings.
Not Accepting Criticism	The employee is observed to have a difficult time accepting criticism, tends to take criticism personally or becomes defensive when message is delivered. Employee has been observed being unwilling to acknowledge errors; or admitting to mistakes; may attempt to cover up errors through lying or deceit.
Self-Centeredness	The employee disregards needs or wishes of others, concerned primarily with own interests and welfare.
Stress	The employee appears to be under physical, mental, or emotional strain or tension that he/she has difficulty handling.
Performance	The employee has received a corrective action (below expectation performance review, verbal warning, written reprimand, suspension, termination) based on poor performance.
Lack of Dependability	Employee is unable to keep commitments / promises; unworthy of trust.
Personal Issues	Employee has difficulty keeping personal issues separate from work, and these issues interfere with work.
Absenteeism	Employee has exhibited chronic unexplained absenteeism.

(e-Service Journal Volume 9 Issue 1)

The Challenge of Measuring Behavioral Baselines & Anomalies

Even with a list of attributes, how does one truly measure these outward signs of an internal struggle? In the days of primarily in-office work settings, supervisors, coworkers, and HR teams maintained closer physical contact with employees. Bumping into one another at the coffee pot, for example, people could read one another's facial expressions and body language. No matter how hard someone works to conceal changes in their outward appearance, people have a strong biological ability to read one another, especially those with high emotional intelligence.

Body Language Isn't Enough Anymore

In the modern workplace, many employees work from home all or part of the time, and even those who are in-office often spend most of their time on remote calls. This limits the information people can read to understand and support one another. Not only are people only partially visible, but the time exposure is limited, and some people attend calls without video enabled. A recent article in The Economist, Body language in the post-pandemic workplace, echos this sentiment.

> *If there is one thing for which online interactions are not suited, it is body language. That is partly because bodies themselves are largely hidden from view: whatever language they are speaking, it is hard to hear them. You will know the partners, pets and home-decor choices of new colleagues before you will know how tall they are. And although faces fill the video-conferencing screen, meaningful eye contact is impossible.*

The fact is that remote work in some form or another is probably here to stay, so it's time to shift the way we gather data. If employees interact primarily through digital means, that will be the best place and way to measure their interactions and behaviors.

Can Digital Tools Measure Behavior Accurately?

A recent paper published in Psychological Science added to a growing body of evidence that machine learning tools can accurately identify individual users over others based on their usage habits and language style. The key to these studies, they have found, is consistent usage of certain platforms and machines that allow for the gathering of data that becomes a behavioral baseline for each user.

Being able to identify a user's behavioral norms becomes extremely relevant in the context of DLP – especially in the scenario of an account compromise. If, for example, a savvy attacker does an account takeover but only views the data sets that authorized account has been granted access to, no alerts are likely to be triggered in standard EDR or SIEM tool, even with applied heuristics. That tool would need to be capable of measuring changes against a baseline for more complex behaviors than just access or usage of data. For example, how long did they view the data? Could that indicate an instance where data is being transcribed to another device, or other potentially malicious behavior? A more intelligent tool would be able to detect anomalous events or trend shifts that align with what behavioral psychologists have identified as key indicators of a threat (table A).

With the understanding that user behavior can be measured accurately by digital tools, then, we return to the subject of how to apply that to DLP.

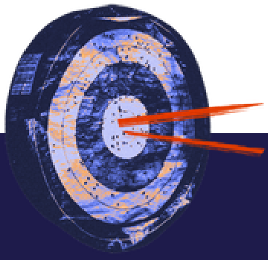
Best Practices for Implementation

Establish Baselines With User Activity Monitoring

Researchers at Eindoven University in the Netherlands were calling for improved strategies as far back as seven (7) years ago – an eternity in the world of tech. In A Hybrid Framework for Data Loss Prevention and Detection, the observation of existing technologies then still largely holds true: “DLP solutions usually aim either at detection, i.e. raising an alert when suspicious activities are observed, or at prevention, i.e. blocking malicious activities. In either cases a model distinguishing normal from suspicious activities is needed.” Understanding typical web browsing habits, what kinds of data an employee uses to accomplish their work, how long they look at that data, who they typically share it with, how often they use a USB drive, and more can improve the accuracy with which an anomaly can be detected. More than ever, today’s distributed workforce can increase the urgency with which mature security programs seek to build a more comprehensive approach to data loss protection.

Fortunately, next-gen productivity tracking software can enable organizations to understand their users’ typical productive, positive behaviors. This data helps security teams understand what has changed and may indicate a problem, the “normal” against which they are measuring and tracking anomalies. Without that information, there is no real way to know what’s happening, and actions or strategies will always be reactive, playing catch-up after the fact. Data loss, financial outcomes, and attack statistics would indicate that this strategy is not working.

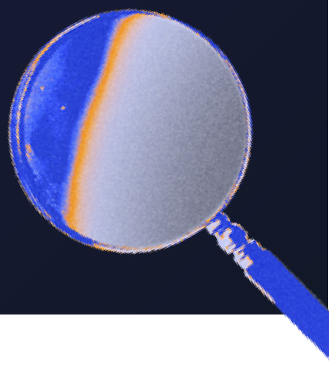
Without a clear and granular look at how someone accomplishes their work successfully as a productive employee, one can’t then measure changes in disposition that could indicate psychological states indicative of an insider threat, such as slow-downs in productivity, withdrawal, increases in aggressive language with coworkers, etc.



What Data Is Important?

While each industry and organization may have nuances in which verbiage or actions unique to their setting may indicate a problem, a behavioral data collection tool needs to be able to measure activities across keystrokes (to create a cache of searchable terminology), file transfer norms, types and sizes of typical email attachments, clipboard actions, time spent looking at sensitive assets to accomplish work, which digital assets are commonly accessed, and browser history.

What Should You Look for In a Tool?



Granular Data Collection Capabilities

To get a strong understanding of your users, you will want to be able to align the granularity of your data collection to the level of security you are working to achieve. For some users, you may only want to see red-flag actions like use of a USB drive or lingering in sensitive datasets. For others, you may only want to know which websites they visit as part of their normal routine. Either way, the ability to record and play-back screen activity can be a game changer for your security team. This, combined with time stamping, allows you to review events very quickly and gain a clear understanding of context and even intent by showing you what happened in the time leading up to an incident, as well as what happened immediately afterward. This can speed up investigations to minutes, as well as provide irrefutable evidence and forensic capabilities.

Access and Privacy Customizability

In the complex digital landscape of the modern enterprise, organizations need to consider compliance frameworks and laws with which they must comply in each geographical region they serve. So, ease of customizability and the ability to apply rules to whole groups of employees, as well as the ability to integrate with Active Directory for this purpose, will be important considerations. Can you limit who sees what data? Can you provide one view to some users and a wholly separate view to others? Can you house behavioral data on-premises or in your own cloud instance, or do you have to use your provider’s cloud?

Robust Forensic Capabilities

Sophisticated threat actors are adept at circumventing traditional DLP tools to evade detection. In these cases, organizations are typically unaware their data has been exposed until a crime is reported or researchers find their data on the darknet. In many of these situations DLP solutions were in place, but were successfully evaded. Optical character recognition (OCR) capabilities help prevent these situations, as well as enable deep forensics investigations to find any and all accounts used to commit data theft.

OCR technology scans user screens to quickly identify and alert when sensitive datasets appear on-screen. Additionally, investigations can be performed, using optical character recognition in conjunction with searches to trace leaked data back to every user who viewed that data on their device. This gives organizations the flexibility to maintain ongoing and automated alerting, as well as perform on-the-fly searches and investigations supported by irrefutable evidence.

Ease of Rollout

Your organization will differ from other organizations in which users need the most oversight, which data is most essential to protect, etc., so you can expect to spend some time working with support agents to customize agents out to test groups to ensure that your EDR or antivirus is not blocking user monitoring functionality. Those are basic steps you will want to consider. However, once you have settled the configurations you want and a few basic IT considerations, the time it takes to roll out the solution should be minimal. There is no reason you shouldn’t be able to accomplish this step relatively quickly.

Final Thoughts

As you work to create a more comprehensive DLP strategy for your organization, the most important thought to keep top-of-mind is the shifting nature of modern business. Whatever strategy you implement will need to be agile and capable of flexing from one year to the next, or even one quarter to the next, to enable you take your organization on a journey from present state to future state in data loss protection.

TERAMIND

Request Your Custom Demo Now

Get Demo