INFORMATION SECURITY

DATA PROTECTION

datadrain

Leaking sensitive information can pop the balloon on your company's reputation. DLP tools can mitigate incidents and offer insight into where data lives.

BY RICH MOGULL

IT'S THE CALL YOU'VE FEARED. The phone rings at 9 a.m. on a Sunday. You're the CISO of a medium-sized retailer, and weekend calls aren't all that unusual. But within 30 seconds of picking up the phone, you know your weekend, if not your job, is over. One of the customer service managers accidentally emailed an Excel file of all the clients acquired last quarter to an external distribution list while trying to send it to his personal Gmail account to work on over the weekend. Worse yet, the file contains full credit card and verification numbers.

The really bad news? You recently signed off on your self-assessment for your Payment Card Industry Data Security Standard audit and affirmed that you don't keep card numbers in an unencrypted format. No one told you about the nightly database extract the customer relations team runs with the credit card number as the primary key. Your external audit is scheduled for next month, making this about the worst time possible for an accidental disclosure. It's not like you can blame this one on evil hackers.

This situation is hypothetical, but it illustrates the pressures companies are under. Data protection grows more critical every day as our sensitive information faces increasing scrutiny from regulators and business partners. It's no longer just a matter of keeping the bad guys away from data. Businesses now are expected to handle it responsibly, often in accordance with contractual or legal requirements. Yet the average organization typically has little idea of where its sensitive data is, never mind how it's really being used.

Over the past five years, a new category of tools emerged to address this problem. Data loss prevention (DLP) products help companies understand where their sensitive data is located, where it's going, how it's being used, and can sometimes enforce protective policies. The technology may not always stop evil hackers, but it offers considerable help in protecting a business from internal mistakes and in cost-effectively managing compliance.

Knowing where sensitive content is located protects the organization and may reduce the time and cost of audits; a company can prove that its data is appropriately secured and show real-time controls to detect violations. By gaining considerable insight into how data is communicated internally and externally, odds are that an organization will identify a number of risky business processes—like the above nightly database dump and use of personal email accounts. It also gains the ability to prevent accidents and eliminate bad habits, like improper use of USB drives. DLP won't make you compliant, but its combination of risk reduction, insight and potential audit cost reduction is compelling.

Yet while DLP tools have significant potential to reduce an organization's risk of unapproved disclosures of sensitive information, they are among the least understood and most over-hyped security technologies on the market. Organizations that take the time to understand the technology, define their processes and set appropriate expectations will see significant value from their DLP investment, while those that make snap purchases or set their expectations inappropriately high will struggle with this powerful collection of tools.

DEFINING DLP

DLP is one of a dozen or so names for this market; others are information leak prevention and content monitoring and filtering. To further complicate matters, data loss prevention is so generic a term it could easily apply to any data protection technology; everything from encryption to port-blocking tools is hopping on the DLP bandwagon. While early tools were tightly focused on preventing data leaks on the network, the market is rapidly evolving toward robust solutions that protect data in motion on the network, at rest in storage and in use on the desktop, all based on deep content inspection and analysis.

So DLP is a class of products that, based on central policies, identify, monitor and protect data at rest, in motion and in use, through deep content analysis. Other defining characteristics are:

- Broad content coverage across multiple platforms and locations
- Central policy management
- Robust workflow for incident handling

It's important to recognize that DLP solutions are very effective at reducing the risk of accidental disclosures or data leakage through a bad business process, but offer minimal protection against malicious attacks. A smart internal or external attacker can easily circumvent most DLP tools, but the risk of inadvertent exposure is usually greater than that of a targeted attack.

GETTING STARTED

Long before contacting DLP vendors, set expectations and decide what content needs protection and how to protect it. Pull together a project team with representatives from major stakeholders including security, messaging, desktop management, networking, human resources and legal, and define protection goals, including content and enforcement actions. This is when you set expectations; educating project members on what's realistic with DLP can help avoid pitfalls that derail deployment.

These protection goals help determine required features. They'll establish needs for content analysis techniques, breadth of coverage (network/storage/endpoint), infrastructure integration, workflow, and enforcement requirements. You can decide if you need a full suite, dedicated DLP solution or just the DLP features of an existing product. Then, translate these requirements into an RFI or draft RFP and start contacting vendors.

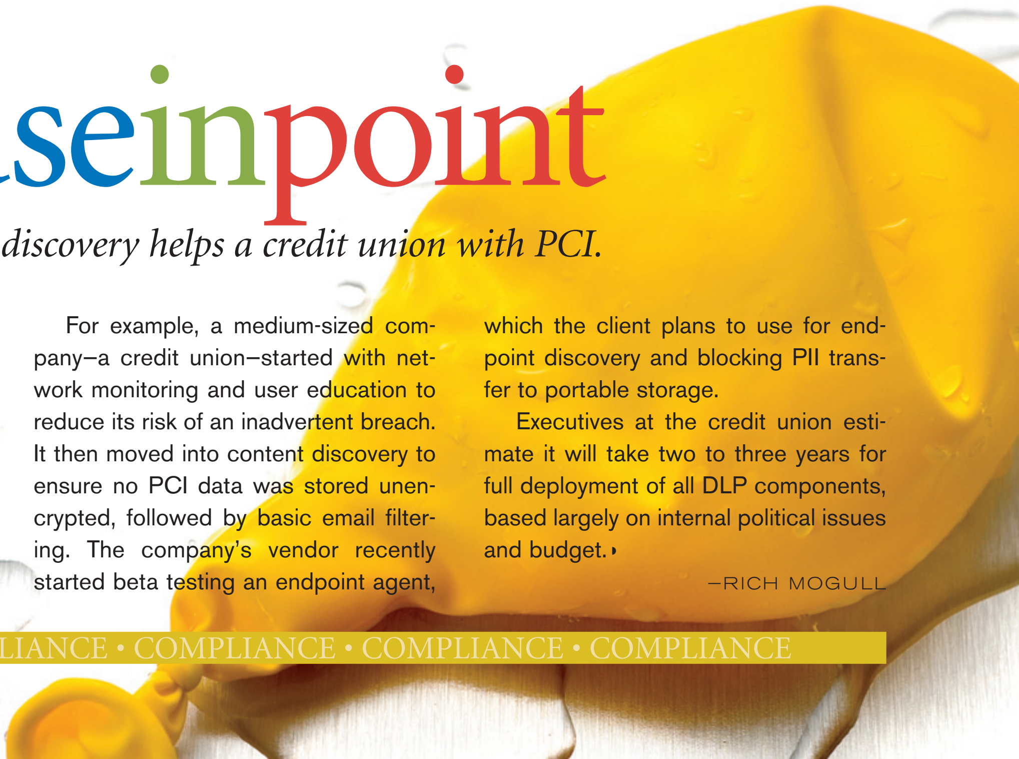
Most organizations find that content analysis techniques, architecture, infrastructure integration and workflow are the top priorities in selecting a product.

CONTENT ANALYSIS

The most important characteristic of DLP solutions is content analysis. This allows the tools to dig into network traffic and files, unwrap layers (like a spreadsheet embedded in a PDF in a .zip file) and identify content based on policies. While every product uses different content analysis techniques, they tend to fall into a few categories that also use contextual information, such as sender/recipient, location and destination.

Content description techniques use regular expressions, keywords, lexicons and other patterns to identify content. They include rules/regular expressions for pattern matching, conceptual analysis involving pre-set combinations of words and rules to match a specific concept like insider trading, and pre-set categories such as personally identifiable information (PII), HIPAA and PCI.

Content registration techniques rely on content you provide the system that then becomes a policy. They include full or partial document matching using hashes of files to identify content; database



caseinpoint

Content discovery helps a credit union with PCI.

The majority of organizations first deploy DLP for network data loss prevention since it's the quickest way to identify their risk exposure. But from a compliance standpoint, DLP for data at rest—or content discovery—is often more valuable since it helps quickly identify stored data in violation of policy, which is especially useful for PCI DSS.

For example, a medium-sized company—a credit union—started with network monitoring and user education to reduce its risk of an inadvertent breach. It then moved into content discovery to ensure no PCI data was stored unencrypted, followed by basic email filtering. The company's vendor recently started beta testing an endpoint agent, which the client plans to use for endpoint discovery and blocking PII transfer to portable storage.

Executives at the credit union estimate it will take two to three years for full deployment of all DLP components, based largely on internal political issues and budget.

—RICH MOGULL

fingerprinting by hashing live database content in combinations to identify matches; and statistical techniques that use a large repository of related content to identify consistencies and create policies.

All the leading products can combine different analysis techniques into a single policy to improve accuracy.

The content analysis technique will directly determine what products make the short list, but make sure to account for future needs. Although most of the market—90 percent by some estimates—is focused on protecting PII, about 30 to 40 percent of those organizations are also interested in protecting unstructured data. They start by using DLP to protect PII to reduce their compliance risk, and then slowly add other content, generally trade secrets and intellectual property, once they get comfortable with their tool.

ARCHITECTURE & INTEGRATION

DLP architectures are defined by where they protect the content: data-in-motion network monitoring, data-at-rest file storage scanning, and data-in-use monitoring of the endpoint. Full-suite solutions include components for each of these areas, while partial suite tools cover only a portion, such as an endpoint DLP tool with an email-only gateway *(see "DLP Vendors,").* There also are single-channel products and non-DLP tools that bundle some DLP features, like an email gateway that can block messages with credit card numbers. In the long run, most organizations—especially large enterprises—will prefer full-suite solutions, but partial-suite and DLP-as-a-feature tools often meet tactical needs where complete coverage isn't necessary.

The DLP market started with passive network monitoring tools focused on detecting information leakage over communications channels such as email, IM, FTP and HTTP. These simple monitoring and alerting tools evolved into more comprehensive solutions, adding email integration and gateway/proxy integration for Web, FTP and IM. This allows organizations to block traffic before the data escapes, rather than just being alerted when it's already gone. *(See "Network Monitoring Tips,").*

For email, DLP vendors embed an MTA (mail transport agent), which is then added as another hop in the email path to block, quarantine, encrypt or even bounce messages back to the user. Since email is a store-and-forward protocol, integration is fairly straightforward. A few tools support similar actions on internal mail by integrating with Exchange and other mail servers.

Other channels, such as Web, FTP and IM, are more difficult to block since that traffic uses synchronous protocols. By integrating with proxies, a session analysis can be performed to reconstruct and evaluate content before it's released. Few DLP tools provide proxies and instead partner with major gateway/proxy vendors, or use the Internet Content Adaptation Protocol (ICAP). When integrated with a tool that proxies SSL traffic, you gain the ability to sniff encrypted traffic.

DLP for data at rest is often equally if not more valuable than network monitoring. This is called content discovery; these tools scan enterprise repositories and file shares for sensitive content. Imagine knowing the identity of every server storing credit card information, and being alerted to unapproved ones.

Content discovery falls into three categories: network scanning, local agents and application integration. With network scanning, the DLP tool connects to file shares for analysis, which provides wide coverage but limited performance. A local agent may be available on major platforms to scan directly on the server rather than across the network, which is more effective for large repositories but requires more management. Some tools integrate directly with document management systems and other repositories to leverage native features.

The last major component of DLP solutions is endpoint agents to monitor use of data on the user's desktop. A "complete" agent theoretically monitors network, file and user activity such as cut and paste, but few real-world tools provide full coverage. Most products start with file monitoring for endpoint content discovery and to detect (and block) sensitive data transfers to portable storage. Rather than completely blocking USB thumb drives to protect data, an organization can use these tools to restrict file transfers based on content.

network monitoring tips

Performance requirements for monitoring outbound communications are less than expected.

When shopping for network monitoring tools for data loss prevention, don't get hung up on high performance. Since outbound communications traffic is the only concern, even if a company is running gigabit Ethernet, it will likely only monitor a fraction of that traffic.

Large enterprises typically need to monitor about 300 MB/s to 500 MB/s at most, while midsized enterprises fall below the 100 MB/s range, and small enterprises as low as 5 MB/s.

Also, make sure to determine if a product monitors all protocols, or just a subset, and if it requires hard-code port and protocol combinations or can detect traffic on non-standard ports. The stronger tools also detect tunneled traffic, like IM over HTTP.

—RICH MOGULL

CHECKLIST • CHECKLIST • CHECKLIST

Endpoint DLP tools are starting to add more advanced protection, such as limiting cut and paste, detecting sensitive content in unapproved applications such as certain encryption tools, and automatic encryption based on content. Over time, they will increase the type and number of policies they can enforce and integrate more deeply into common endpoint applications.

MANAGEMENT & WORKFLOW

DLP solutions are dedicated to the business problem of identifying and protecting sensitive information. Ideally, an enterprise wants to establish a single policy for data protection and apply it throughout its environment—a key advantage of a full-time DLP solution over security tools with a DLP feature. DLP suites centralize workflow for incident handling across the network, storage and endpoints, and provide user interfaces for technical and non-technical incident handlers. Many organizations find that compliance, legal and HR departments play just as large a role in policy enforcement as IT security.

Central policy management allows a user to define the content to protect—like a customer identification number—then apply different enforcement actions based on where the violation is triggered. You define the content once, and then build rules based on context. These policies are distributed throughout a DLP infrastructure, including the network, storage and endpoints. Policies apply differently to dif-

DLP vendors

Here is a representative list of some vendors offering data loss prevention products.

FULL-SUITE SOLUTIONS

EMC/RSA (acquired Tablus, Aug. '07) www.emc.com

Orchestria www.orchestria.com

Reconnex www.reconnex.net

Symantec (acquired Vontu, Nov. '07) www.symantec.com

Vericept www.vericept.com

Websense www.websense.com

PARTIAL-SUITE SOLUTIONS

Code Green Networks www.codegreennetworks.com

GTB Technologies www.gttb.com

McAfee www.mcafee.com

Workshare www.workshare.com

NETWORK-ONLY TOOLS

Clearswift www.clearswift.com

Fidelis Security Systems www.fidelissecurity.com

Palisade Systems www.palisadesys.com

Proofpoint www.proofpoint.com

ENDPOINT-ONLY TOOLS

NextSentry www.nextsentry.com

Trend Micro (acquired Provilla, Oct. '07) http://us.trendmicro.com

Verdasys www.verdasys.com

—COMPILED BY RICH MOGULL

PRODUCTS • PRODUCTS • PRODUCTS

ferent users, are rated at different sensitivity levels, have violation count thresholds, and are assigned to specific business units or incident handlers.

For example, a policy could be set that says: "The customer relations team is allowed to email a single account number to a recipient, but block account numbers in any other channels or by any user. Only customer team members can store account numbers on their laptops, but only if encrypted. Account numbers cannot be transferred to portable storage, and are only allowed on these servers."

Enforcing this kind of policy requires integration with enterprise directories and dynamic host configuration protocol (DHCP) servers to identify the user's location (system and IP address)—a critical feature to look for in the evaluation process. Role-based administration and hierarchical management ease management overhead and are particularly important in large deployments.

DLP policy violations are extremely sensitive and usually require dedicated workflow. Unlike virus infections or IDS alerts, these incidents lead to employee dismissal or legal actions. The heart of the DLP management system is the incident handling queue, where incident handlers see open violations assigned to them, take actions, and manage workflow for investigations. A good workflow interface eases identification of critical incidents and reduces incident handling time, management overhead and total cost of ownership.

Last year, a DLP customer chose its product ultimately on workflow. After narrowing the field to two vendors it considered equal in terms of technical features, the company selected the product with the workflow and interface its non-technical users (legal, HR and compliance) preferred.

Beyond policy management and incident handling, look for a tool that integrates well with existing infrastructure and includes robust management tools like incident archiving, backup, and performance monitoring. Since senior management and auditors might be interested in DLP activities, robust reports are needed for this non-technical audience and compliance support.

TESTING & DEPLOYMENT

After bringing in vendors for sales pitches and demonstrations, narrow the field to three or four and start a proof-of-concept trial. Preferably, place the tools side by side in passive monitoring mode on the network and test with representative policies. This allows a user to directly compare results for false positives and negatives, but is tougher to do with endpoint tools. Also test enforcement actions and integration into the infrastructure, especially directory integration. Finally, run the workflow past the business units involved with enforcement to ensure it meets their needs.

Organizations report that DLP deployments tend to go more smoothly than other security installations from a technical level, but it may take up to six months to tune policies and adjust workflow, depending on the complexity. Many find they only need part-time resources to manage incidents, but this varies based on the intricacy and granularity of policies. A 5,000-person organization, on average, only needs a half-time incident handler and administrator to manage incidents and keep the system running.

WHAT'S AHEAD

DLP tools are still fairly adolescent, which means they provide good value but are not as polished as more mature product categories. This shouldn't slow down deployments if you have data protection needs, but understand that the tools will evolve rapidly. Already, the market is transitioning from data loss prevention, focused on plugging leaks, to more-robust content monitoring and protection (CMP) designed to protect data throughout its lifecycle. CMP will eventually become one of the most important tools in the security arsenal.

Rich Mogull is founder of Securosis, and a former security analyst at Gartner. Send comments on this article to feedback@infosecuritymag.com.

Reprinted with permission from Information Security Magazine, February 2008.