WHITE PAPER

CDW PEOPLE WHO GET IT

DATA LOSS PREVENTION

Moving beyond perimeter security with a flexible, in-depth approach to protecting data

Executive Summary

Many organizations have come to recognize that their data is their most valuable resource. Unfortunately, many criminal elements have reached the same conclusion. The rise in high-profile security breaches where an organization's data was stolen reflects the growing need to secure data.

A security incident that results in the unintended exposure of sensitive information can have a dramatic effect on any organization. Data loss prevention (DLP) solutions can help reduce the risk of this type of exposure, preventing both malicious and unintentional data leaks.

For many years, enterprise security meant building a secure wall around IT assets and making sure only authorized individuals and devices crossed it. Firewalls, intrusion detection system (IDS) devices, virtual private networks (VPNs) and other security controls were the building blocks of that wall.

TWEET THIS!

Table of Contents

However, the days of perimeter-based security are numbered. Users are increasingly demanding remote access to enterprise data, and it is increasingly common to see organizations granting this access.

This shift to a "global" network requires a corresponding shift in the way organizations think about security. Security controls must now focus on data, rather than networks or devices. Instead of building virtual walls around geographic locations, enterprises should think about drawing walls around particular data elements. DLP solutions provide that data-driven security capability.

These technology tool sets give security administrators the ability to build inventories of sensitive information, monitor the flow of that information across networks, and prevent the leakage (intentional or unintentional) of that information to unauthorized locations. Strong DLP programs depend on a solid foundation — a data classification effort that accurately describes the types of information that must be protected.

Data Classification

Data classification schemes provide a mechanism for an organization's staff and partners to understand the sensitivity of different types of information. Every organization is free to choose a different set of classification labels, but many use something similar to the following three categories.

Highly sensitive data is the most critical information an organization possesses. Unauthorized disclosure of highly sensitive information could cause irreparable harm to the organization or its customers. Examples of information that commonly falls into this category include credit card numbers, Social Security numbers and sensitive business plans known only to a few top executives.

Confidential data is information that should not be disclosed outside of the organization without permission. This is the most common classification used and covers most business plans, financial records, personnel data and other sensitive information that does not meet the highly sensitive threshold.

Public data is information that may be freely shared with anyone. It typically includes marketing materials, information about the organization published on a website, newsletters and similar records.

Although many organizations use this scheme, it is by no means universal. Some enterprises choose to add a fourth level, sometimes referred to as sensitive data, which falls between the top two classifications.

One of the most important things an organization can do is clearly define the types of information that fall into the top category of its classification scheme. While it may be difficult, if not impossible, to classify all the records an organization creates, it is extremely important to articulate the highly sensitive information that requires the most stringent security controls. Accurately identifying this information is also critical to the successful deployment of a DLP system.

Government Classification Schemes

The federal government uses one of the most well-known and transposed classification schemes. It has four levels of sensitivity:

- **Top secret:** This refers to information that, if disclosed without authorization, would cause "exceptionally grave damage" to national security.
- **Secret:** This is information that would cause "grave damage" to national security if made available to the public.
- **Confidential:** This refers to information that would either "cause damage" or be "prejudicial" to national security if released.
- **Unclassified:** This is information that falls outside the classification scheme and may be publicly released.

Government agencies (and organizations that work with them) are required to follow this classification scheme. Decisions about how to classify data are made in a highly centralized fashion and are documented in classification guides that describe the appropriate classification label for different types of information. Individuals are left with very little discretion and must follow the guidance provided in those publications.

Organizations subject to the government's classification scheme can leverage it in their DLP environment. In fact, they're already a step ahead of most other enterprises because government regulations require the labeling of all documents with the classification level — a measure that is rarely used in corporate and nonprofit environments.

In cases where government classifications are used, DLP products can be configured to watch for the flow of classified information to unsecure networks. For example, a DLP solution might alert the proper parties and block transmission of a document labeled "Top Secret" across a network that is approved only as high as "Secret" data.

The Role of DLP Systems

Once an organization has clearly defined the types of sensitive information it possesses, DLP solutions can help manage that information. The first task is to build an inventory of the existing information, based on the newly defined levels of sensitivity. From there, DLP systems perform three important duties in an enterprise security infrastructure:

1. **DLP systems assist with inventorying the sensitive information possessed by an organization.** In many cases, an enterprise simply doesn't know where sensitive information exists. DLP tools can help ferret out hidden stores of data lurking on endpoint hard drives, in e-mail records and at other locations.
2. **DLP systems monitor the flow of information around an organization.** By monitoring both endpoints and network communications, DLP systems can help track the flow of sensitive information around an organization, potentially identifying unknown business processes that involve these records.
3. **DLP systems block potential data leaks before they occur.** One of the most common reasons organizations deploy DLP is to proactively block potential data leaks. When a DLP system detects an unauthorized flow of sensitive information, it can intervene, blocking the communication and preventing a potential data breach.

Data-detection Techniques

When organizations deploy DLP, they often use the system to assist in identifying sensitive information stores. This inventory-building process relies on sensitive data-detection techniques of DLP systems, which fall into two categories: pattern matching and document tagging.

Pattern matching: Pattern-matching DLP systems function in a manner similar to signature-based antivirus systems. They maintain a list of content patterns that commonly appear in sensitive information and then monitor the host or network for occurrences of those patterns.

The patterns may be specific keywords (such as "credit card" or "highly sensitive"), or they might contain wildcards where the system may substitute valid characters (such as "xxx-xx-xxxx" or "xxxx-xxxx-xxxx-xxxx" where x is any digit).

Pattern-matching systems often ship with a number of vendor-provided policies that are useful in certain settings. For example, DLP systems often contain rules allowing them to search for sensitive number formats, including Social Security numbers (and their non-U.S. equivalents), credit card numbers and driver's license numbers.

Other rules search for documents containing terms that commonly appear in healthcare or legal records, drawing upon a database stored in the DLP system. For example, the system might trigger an alert if it detects words and phrases such as "myocardial infarction," "lawsuit," "ulcer" or "liability."

DLP systems can add value to their pattern-matching rules by analyzing additional contextual information before triggering an alert. For example, a rule designed to detect Social Security numbers might trigger an alert on a nine-digit number only if it is properly hyphenated or if the document also contains the terms "SSN" or "Social Security number."

SSN Detection Tips

One of the most common frustrations with pattern matching is the high degree of false positives, sometimes generated by poorly written DLP rules.

For example, a DLP rule that attempts to detect Social Security numbers by searching for nine-digit numbers is bound to generate a large number of false-positive alerts. Fortunately, there are some simple rules that DLP administrators can use to help fine-tune SSN detection:

- No valid SSN begins with 000 or 666.
- No valid SSN begins with the number 9.
- No valid SSN contains 00 as the fourth and fifth digits.
- No valid SSN ends with 0000.
- SSNs with the first three digits ranging from 772 to 799 were not issued until June 2011.
- SSNs beginning with the number 8 were not issued until June 2011.

Until recently, it was possible to identify the geographic area where an SSN was issued by analyzing the first three digits of the SSN. In June 2011, the Social Security Administration began an SSN randomization process that eliminates the geographic significance of the first three digits.

In some cases, DLP products come with these rules predefined; in other cases, administrators may need to program them.

Document tagging and fingerprinting: Pattern-matching algorithms are not foolproof. Pattern-matching approaches are prone to false-positive alerts and are unable to detect sensitive information that may not fit one of the defined patterns. Document tagging offers an alternative approach that relies on indexing known stores of sensitive information and monitoring other environments for the presence of that information.

In addition to indexing known stores of sensitive data, some DLP systems are capable of monitoring the files accessed by a particular application and then automatically tagging those files as sensitive. For example, the DLP might automatically tag all files accessed or created by a health-records system as sensitive.

The tags applied by DLP systems may contain more information than simply the classification of the document. Tagging allows systems to retain fields such as the original location of the data, the user who originally created the file, the source application and other fields.

In addition to document tagging, DLP systems use document fingerprinting to help identify sensitive information. Systems that employ fingerprints use a hashing algorithm to create

a unique signature for each file that contains sensitive information. These short fingerprints are then stored in the DLP database.

Network-based DLP systems can then use the same hashing algorithm to analyze documents leaving the organization and compare the hash values of those documents to the DLP database. A match between the hash of a file leaving the organization and a record in the DLP database triggers a DLP policy alert.

Choosing the Right Approach

Organizations seeking a DLP system for their environments should thoroughly explore the detection mechanisms used by the candidate systems. In most cases, it is best to find a system that uses a combination of pattern matching and document tagging to provide a layered approach to DLP.

Pattern matching is useful for identifying previously unknown stores of sensitive information, while document tagging and fingerprinting ensures that known sensitive information doesn't slip through the cracks of a pattern-matching approach.

After building an inventory of sensitive information, DLP systems can be used to monitor networks and endpoints for the transmission or storage of sensitive information in violation of an organization's security policies. This requires building a sound monitoring strategy and conducting a user education campaign.

Organizations seeking to create a DLP monitoring strategy should first examine their existing security policies. Those broader policies should be used to guide DLP implementation efforts and dictate the appropriate responses when potential policy violations come to light. For example, an organization should use existing human resources disciplinary procedures to handle staff who commit first-time and/or repeat violations of the DLP policy.

Decisions made during the creation of a classification scheme should guide the creation of a DLP policy. One of the most effective approaches to DLP implementation is to start small, working first with the information that is most sensitive and easy to detect, then expanding the implementation outward as business needs and technology capabilities permit.

A DLP implementation focused on credit card numbers, for example, is more likely to be successful than a project that attempts to detect all forms of sensitive information from the start.

Another critical decision that organizations face during the DLP implementation process is selecting appropriate technical responses for when policy violations arise. The range of potential responses includes:

- Blocking the transmission of sensitive information;
- Quarantining files containing sensitive information detected in unauthorized locations;
- Automatically applying encryption controls;
- Displaying a notice to the end user that an action may violate the organization's security policy, but allowing the user to continue after acknowledging the warning;
- Taking no proactive remedy but logging the activity for administrator review (an approach commonly used as a first step in a DLP deployment).

The appropriate set of actions an organization may take will depend on a combination of its security policy, risk appetite and the data sensitivity.

Accounting for Compliance

Although no major security regulations specifically require the use of DLP technology, DLP can nevertheless play an important role in building a culture of compliance within an enterprise. It can also serve as a valuable control in security architectures that are built to satisfy compliance requirements.

Healthcare organizations commonly use DLP as a method to facilitate compliance with the Health Insurance Portability and Accountability Act (HIPAA). DLP systems can be used with a dictionary of terms that appear commonly in sensitive patient

DLP and PCI DSS Compliance

The Payment Card Industry Data Security Standard prescribes strict security controls that must be followed by organizations that store, process or transmit credit and/or debit card transactions. PCI DSS does not directly address data leak technology, but many organizations choose to deploy DLP in some related capacities:

- **DLP is deployed as a tool to verify the scope of the cardholder data environment.** When configured to detect plain-text storage or transmission of credit card numbers, DLP solutions can identify systems and business processes previously thought to be out of the scope of PCI DSS compliance that, in reality, contain cardholder data.
- **DLP is deployed as a backup control.** This lets the organization block and/or automatically encrypt any cardholder information that accidentally leaves the secured environment in unencrypted form.
- **DLP is deployed as a compensating control.** This is done with the approval of the organization's merchant bank in situations where it becomes difficult or impossible to meet one or more of the standard provisions of PCI DSS.

records to identify electronic protected health information that may leave the organization without the appropriate security controls. Any detected flow of information can be either blocked or automatically encrypted.

Similarly, financial services organizations can use DLP products to look for transmissions of sensitive customer information without security controls. Depending on the financial services company, the DLP solution can monitor broker/client communications, for example, for terms such as "guaranteed," "promise" or "avoid taxes" that may indicate a compliance violation.

User Education

Although user education is not often one of the intended purposes of a DLP deployment, many organizations that field DLP systems find that it is the most important outcome. Employees or other users of systems protected by DLP quickly become aware of security concerns after receiving messages such as, "The content you attempted to transfer has been blocked." Among the takeaways from this experience are:

- The organization is serious about data protection and is taking measures to ensure that sensitive information is not leaked without authorization.
- User activity really is monitored and action is taken when policy violations occur.
- Even authorized data transfers must use appropriate security controls, such as encrypting e-mail messages and attachments.
- If the user attempts to take shortcuts or violate the enterprise security policy, there is a high likelihood of detection.

In fact, organizations deploying DLP for the first time can often quantify this education process by monitoring the number of policy violations detected by the system. In many cases, the initial deployment of a DLP policy results in a large number of violations in the short term, followed by a rapid and sharp decrease as users begin to learn what data-sharing behaviors are acceptable.

The most successful DLP deployments include a strong user education program that goes beyond the DLP technology and provides users with a framework they can use to understand their role in data protection. It should also incorporate instructional material explaining the proper use of encryption and other controls to secure sensitive information.

DLP and Encryption

Encryption is a powerful tool in the hands of information security professionals. It provides a mechanism for obscuring the contents of communications in order to preserve the confidentiality of information. However, encryption can prove confounding to DLP systems by preventing them from analyzing content. Fortunately, most DLP systems now have the ability to overcome this limitation and even leverage encryption to enhance data security.

Although encryption technology intimidates many people, the basic concept is straightforward. Encryption uses mathematical algorithms to preserve the confidentiality of information by making it unintelligible to anyone other than an authorized recipient.

The process of converting plain-text information into an unintelligible format (encryption) is performed by the individual or system creating the information. The individual or system that requires access to that information then follows a similar process (decryption) to convert the unintelligible format back to plain text.

When two parties use encryption to communicate, they must agree on an encryption algorithm. The algorithm dictates the steps that each will follow to encrypt and decrypt the information. If the parties do not use the same encryption algorithm, the decryption will fail. The algorithm itself is not a secret, and the details of most encryption algorithms are publicly available.

Ensuring that only authorized parties have access to the applicable security keys preserves the secrecy of the information. The keys are long binary strings used in the encryption and decryption operations. Think of them as the passwords used to protect the message.

There are two basic forms of encryption algorithms that differ in the way they use these keys. In a symmetric encryption algorithm, both the party encrypting and the party decrypting the data use the same key for both operations. This is known as a shared secret key. In asymmetric algorithms, the sender and the receiver use different keys for the encryption/decryption operations.

Encryption is used to protect information in two different venues: at rest and in transit. The most common way to protect data at rest is through full-disk encryption, which protects the contents of notebooks and other mobile computing/storage devices in the event they are lost or stolen. Organizations may also use encryption to protect files stored on a system in the event that individuals not authorized to view the data gain access to the system.

Encryption technology also protects data in transit across a network, including local networks (such as an office wireless network) or across the Internet. Such encryption allows an organization to use otherwise unsecure networks to transfer sensitive information without exposing it to the risk of eavesdropping.

Common examples of encryption to protect data in transit are Hypertext Transfer Protocol Secure (HTTPS), which protects e-commerce transactions over the Internet, and virtual private networks (VPNs), which allow mobile workers to connect securely to an enterprise network.

Although encryption is a valuable tool in the security toolbox, it also poses a risk to DLP systems that seek to prevent the unauthorized export of sensitive information. Consider a scenario in which an individual uses an encrypted, web-based file storage service.

These services allow anyone with a web browser to open a free account and upload large quantities of information for retrieval from any location. An individual could use this type of service to transfer information from an organization and unwittingly (or intentionally) expose the data to unauthorized disclosure. Clearly, enterprises need to be aware of staff using this type of service.

Normally, a DLP system should have no difficulty detecting this type of information exchange by eavesdropping on communications between the end user and the remote web server as it passes through the enterprise network.

However, a secure HTTPS connection inhibits this eavesdropping by preventing both an outside party and the DLP system from seeing the actual content as it's transferred to the remote storage service. This prevents the DLP system from appropriately enforcing the organization's security policy.

Fortunately, DLP vendors have recognized the challenge that encryption poses to the accurate detection of confidential information. As a result, they've developed two work-arounds that allow the inspection of encrypted content as it is transferred across a network.

One approach uses network proxy servers to catch the information while it's in transit. The other approach leverages the unique capabilities of host-based DLP agents to inspect information before encryption or after decryption takes place.

In the proxy-server approach, the DLP system integrates with a web proxy that serves as an intermediary between clients and remote web servers. Instead of establishing connections directly to external servers, the client browsers pass those connections through a proxy server, which gains access to the unencrypted communications stream. The proxy server, in turn, provides the DLP system with access to this information and allows it to block unacceptable communications.

Alternatively, DLP systems that use host-based agents can inspect the data traffic sent from and received by a web browser in plain-text form outside of the encryption process. The agent can then analyze the information or pass it to a central DLP system for analysis.

Despite the challenge that encryption poses, DLP systems can also use encryption as a security control to protect an organization's sensitive information from unauthorized disclosure. The most common example of this is when a DLP system detects that someone inside an organization is attempting to send sensitive information, such as a Social Security number or credit card number, via unencrypted e-mail.

Many DLP systems integrate with secure, web-based e-mail gateways and can instruct the gateway to block the transmission of messages containing sensitive information. The message is then replaced with a note to the intended recipient indicating that a message awaits review in the organization's secure messaging system. The recipient then logs in to the web-based system (using a link provided in the e-mail) and retrieves the message over an encrypted web connection.

Monitoring Data

Data loss prevention systems offer organizations distinct advantages for building an inventory of sensitive information, monitoring the flows of that information across and out of enterprise networks, educating users about the importance of protecting sensitive information, and blocking unauthorized flows of information before they occur. The combination of network-based and host-based DLP provides a defense-in-depth approach to preventing unintended and/or unauthorized information leaks.

Network-based DLP systems are often deployed by organizations seeking an easy-to-implement, centralized approach to preventing data leakage. These systems sit at the network perimeter and monitor all communications that attempt to pass through the network border, watching for information that matches patterns or contains tags that violate the organization's security policies.

Administrators can configure rules that block unauthorized data traffic, notify the end user that the traffic may violate the organization's security policy, and/or alert the administrator that a policy violation may have occurred. In addition, some network-based DLP products integrate with other security controls, such as web-based e-mail gateways, allowing the seamless protection of outbound information in a manner transparent to the sender.

Network-based DLP sensors are typically capable of monitoring multiple application protocols, including HTTP and HTTPS, electronic mail, file transfer protocol (FTP), and instant messaging protocols. IT staffs should verify that the network sensors they consider deploying support the protocols commonly used in their organization and permit outbound access to the Internet via their network firewall.

Host-based DLP agents use software installed on user endpoints (desktops, notebooks, etc.) to monitor both stored and transmitted information for violations of the organization's security policies. These agents are able to detect DLP policy violations that might otherwise go undetected for two reasons.

First, the information is passed over an encrypted channel that is not visible to a network-based DLP system. And second, the information is stored on the system but never transmitted. The organization is still exposed to the risk of the loss, theft or hacking of the device, but this risk is invisible to a network-based DLP.

The agent, however, is not subject to these limitations because it has direct access to the operating system, web browser and file systems used to store, process and transmit data.

In some cases, using host-based DLP agents may be problematic for end users, especially when they've been permitted to use devices that are not owned by the organization. This situation is increasingly common given the rise in consumer electronics and bring-your-own-device (BYOD) programs.

Workers are bringing notebooks, tablets, smartphones and other devices from home and expect to use them on the enterprise network. Increasingly, organizations have begun to allow network access via personal devices. This causes several issues for host-based DLP solutions:

- License agreements may not permit the installation of agents on devices not owned by the organization.
- Users may object to installing monitoring software on their personal devices (although organizations can make the installation of the agent a condition of using the device on the network).
- Users may attribute unrelated system problems to the DLP agent and request that the enterprise IT staff resolve the problem.

Organizations that elect to deploy host-based agents should ensure they have clearly written policies that permit the use of the software to monitor user activity, and they should communicate to users the purpose and intent of the DLP system. This communication also serves to reinforce the educational role of DLP deployments.

The Selection Process

Once an organization makes the decision to implement a DLP system, it begins a process of evaluation and selection that typically leads through the offerings provided by a number of different vendors. At a high level, there are numerous questions that enterprises can ask in order to conduct a consistent and thorough evaluation process. These include:

Web-based E-mail and DLP

Although most enterprises often rely on large-scale e-mail and collaboration platforms, such as the popular combination of Microsoft Outlook and Microsoft Exchange, many consumers now use web-based e-mail.

The most popular of these, Google Gmail, now uses encryption to protect all communication between the client computer and Google servers. This poses a significant challenge to DLP efforts because unless IT takes additional measures, users can send messages to and from their Gmail accounts without the scrutiny of a DLP system. Users may then take advantage of encrypted web-based e-mail as a secure, easy-to-use channel for smuggling sensitive information out of the enterprise.

What's an enterprise to do? Some organizations take the heavy-handed approach of blocking access to nonwork e-mail accounts. Others choose to supplement basic DLP deployments with additional controls that allow the inspection of web-based e-mail, such as web proxies and endpoint DLP agents.

The choice an organization takes will likely hinge on the needs of the users as well as the security classifications in play. An office that must manage and transmit highly sensitive information regularly will more likely opt for more stringent controls.

- What capabilities does the proposed system have to inspect information? Does it provide integration with both host-based and network-based sensors?
- Does the system support proxies for the inspection of encrypted web communications? If the organization is already using a web proxy, is the DLP system compatible with the specific proxy in use?
- Does the system's host-based agent support the inspection of encrypted web communications? If so, is the agent compatible with the operating systems and web browsers in use by the enterprise?
- Is the system capable of searching for sensitive information on stand-alone data stores, such as enterprise file systems and collaboration systems? Is it compatible with the specific technologies in use in the enterprise?
- Do the system's reporting capabilities match the organization's security metrics and reporting requirements?
- Does the system support both pattern-matching and document-tagging approaches to detection? If the system includes predefined dictionaries, is there a dictionary available for the specific industry covered by the enterprise?

In addition to these questions, each enterprise will have organization-specific criteria that should be discussed during the vendor evaluation and selection phase of the project.

DLP in the Real World

Two main forces drive the adoption of DLP systems in modern enterprises: the consumerization of technology and the increasing sophistication of threats against an organization. Users have access to a wider range of technology devices, and unlike in the past, these devices are not always under administrative control of the organization.

At the same time, users are increasingly familiar with the many cloud-based services that allow them to easily export large quantities of information. They expect access to the same types of services that they use at home while in the office, often unaware of the potential security risks associated with this type of activity.

Attackers are also becoming increasingly sophisticated; security experts cite the evolution of the advanced persistent threat to enterprise data as a rising concern. Hackers no longer use the "target of opportunity" approach and now directly target organizations that may have lucrative stores of sensitive information.

They are willing to spend a significant amount of time analyzing and defeating the security controls used by those organizations. Even though perimeter security remains important, every organization must now also focus on securing what is within that perimeter.

Data loss prevention systems provide security administrators with a vehicle to combat the steadily increasing risk associated with these trends. DLP systems can thwart the determined attacker who manages to gain access to an organization's systems and attempts to export data.

They can also detect and block the unauthorized export of information by authorized users who may be simply seeking better and easier ways to do their work. Ultimately, the very presence of these systems increases users' awareness of the security risks their actions may pose.

TWEET THIS!