

# RISK ASSESSMENT AND DATA LOSS PREVENTION

Addressing security threats including those related to cloud computing, social networks, virtualization, mobility and peer-to-peer computing

## Table of Contents

- 
- 2 Executive Summary

---

  - 2 The Business Case for Security

---

  - 3 Risk/Benefit Analysis

---

  - 6 Data Loss Prevention

### Executive Summary

Cloud computing, virtualization, social networking, mobility, peer-to-peer computing: Although these terms spark innovative ideas in many parts of an organization, they may also simultaneously raise eyebrows among IT security professionals, many of whom may be unsure how new technology affects risk.

Indeed, these emerging technologies pose unique challenges to those charged with safeguarding the confidentiality, integrity and availability of information and IT systems.

Several important issues guide the evolution of security practices in response to these potential threats. First, IT security professionals must always keep the end goal in sight: protecting the operational interests of the organization. The second important driving force is risk assessment. Each security project should be based on due diligence associated with a risk assessment developed using a sound approach, either quantitative or qualitative.

Over the past decade, most security professionals focused on building a strong perimeter using firewalls, network segmentation, intrusion detection/prevention and similar technologies. Unfortunately, this approach doesn't work as well in the face of ubiquitous mobility. The expanded use of notebooks, tablets and other handhelds and smartphones has changed the security landscape forever.

Experts now recommend shifting security efforts from a network-centric approach of securing the perimeter to a focus on protecting data. Data loss prevention (DLP) technologies offer both network-based and host-based options for identifying and protecting an organization's sensitive information and eliminating breaches of the security guarding that information.

DLP tools play an important role in a defense-in-depth approach to information security, a strategy that uses multiple, overlapping controls to protect an organization's data and systems. In the event a single control fails, the multiple controls will compensate for that loss, continuing to protect the organization.

While nobody advocates abandoning the perimeter, conventional wisdom suggests that security professionals augment their existing perimeter-based defenses with new controls focused on preventing the intentional or accidental leaking of sensitive information.

This white paper will help organizations understand the business case for security, offer pointers on assessing risk, and fully explain how the leading DLP technologies can protect them from potential threats and vulnerabilities.

### The Business Case for Security

One of the most challenging issues security professionals face is bridging the gap in language and culture between the server room and the organization's top management. The most successful IT security leaders embrace these differences. IT professionals who can live with one foot in the world of servers, firewalls and encryption while the other remains firmly planted in the management environment are well positioned for success.

These individuals know how to forgo technical jargon and express risk in plain language easily understood by nontechnical managers. Similarly, they can translate the broad strategic objectives outlined by top management into tasks the technology staff can take action on.

#### Building the Case

Describing the value of security controls in the broader language of an organization's operations is an important challenge for security professionals. It's often necessary to summarize complex technical concepts, omitting important details to managers who lack the background to understand the technical nuances. Security professionals faced with this challenge can benefit from the following suggestions.

**KEEP IT SIMPLE.** Information security issues are often very complex. Faced with an overwhelming set of risks that include many unanswered questions, security professionals are often tempted to explain every contingency. This approach usually doesn't work well because it requires those with the least understanding of the issues to sort through the possibilities and make informed decisions.

Successful security staffers grow beyond the role of risk explainer and into the role of risk adviser. They recognize that it is acceptable to summarize issues and leave out many of the less critical details.

**EXPLAIN THE IMPACT ON THE ORGANIZATION.** Organization leaders typically understand basic security concepts such as confidentiality, integrity and availability. This makes it possible to start the conversation on a common footing when explaining the need for an investment in security technology. Compare the following approaches for obtaining funding:

*Proposal 1:* "Our school's web inspection system is unable to parse HTTPS encrypted traffic. Fixing this problem will cost \$3,000."

*Proposal 2:* "Students can bypass our web filtering system by simply adding an 's' to web addresses. Our existing system can't solve this problem, but a \$3,000 investment will let us close this loophole, preventing students from accessing prohibited websites at school."

Both proposals convey that the web content filter requires an upgrade, but the second approach translates the security problem into simple-to-understand English. School principals likely will not understand phrases like "HTTPS encrypted traffic" and are extremely unlikely to invest scarce resources on something they don't fully understand.

They will almost certainly agree, however, that students should not be allowed to access undesirable websites by simply adding an "s" to the address. This is a problem that requires immediate resolution and is worthy of a technology investment.

**TELL A STORY.** Illustrating complex security problems with stories is an extremely effective way to make a point to managers. Although they may not understand network-based DLP technologies, they will certainly understand the security concerns, for example, if a federal government worker accidentally posted sensitive documents on the web containing the names and locations of nuclear research facilities in the United States.

### MIT and the Sensitive Data Iceberg

Security professionals at the Massachusetts Institute of Technology recently discovered that large numbers of records containing sensitive information were collected over the years and neglected. These records were distributed across many campus locations and lines of responsibility.

Some were found in centralized admissions and human resources records, while others were in the hands of individual professors and administrators. Although school administrators knew the problem was large, they couldn't even develop a reasonable estimate of the number of sensitive records on campus.

Senior administrators could not understand why they were unable to arrive at an estimate. The security team came up with an excellent visual example to tell the story, depicting the risk landscape as an iceberg. The small portion of the iceberg located above the surface (the tip) reflected the known locations of sensitive data – admissions records, employment files and student financial aid records.

However, below the iceberg's surface was a much larger collection of sensitive data that might be stored without their knowledge. This included very old class lists maintained by professors, long-forgotten spreadsheets on desktop computers, and data in the hands of independent contractors who might not be subject to the same security policies as staff.

By using a familiar visual cue, the team at MIT conveyed the scope of the problem and the reason why the full scope was unknown. This helped them obtain funding for a campuswide initiative to secure sensitive data.

Stories like this can help top managers draw parallels to their own situations and imagine the impact of a similar incident on their organization. These basic communication principles can help bridge the gap between the server room and top management, opening the door for a two-way conversation on risk management.

## Risk/Benefit Analysis

One of the most common resources for directing a sound security strategy is a process known as risk/benefit analysis, performed using either a quantitative or qualitative approach.

This is similar to how financial professionals use quantitative risk/benefit analysis to assess the desirability of financial transactions: for example, expressing the risk of an investment as a 60 percent chance of achieving an 80 percent return on investment in two years, and a 40 percent chance of incurring a 10 percent loss over that same period. Managers can then determine whether the potential benefits outweigh the possible risks.

Expressing risks to data confidentiality, integrity and availability in clear quantitative terms has always been a challenge for IT security professionals. More often than not, information security risks call for a qualitative approach that shies away from concrete data in favor of easier-to-obtain, relative-risk evaluations. Risk/benefit analysis is an example of a very basic risk assessment approach. The next section discusses more comprehensive approaches to weighing risk.

### Assessing and Managing Risk

The techniques of risk/benefit analysis are formalized in two related categories: risk assessment and risk management. Risk assessment identifies and evaluates an organization's risk landscape, identifying those threats that pose the greatest risk to the organization. Security professionals may then manage those threats, using a variety of formal risk management practices.

Risk management uses common terminology designed to describe the components of risk, as shown in the sidebar *Risks, Threats and Vulnerabilities* on Page 4. Risk assessments break down each one of these components, identifying potential sources of risk to the organization.

Risk assessment methodologies use this language to guide the process. Risk assessors begin by developing lists of both the threats facing an organization and the known vulnerabilities in the organization's infrastructure. Cross-referencing those lists results in a tally of risks that the organization's leadership must manage.

### Risk Categories

When describing risks, it is often helpful to categorize them according to a common taxonomy. This approach assigns the responsibility for risk management and lets decision-makers view the organization's risk at a high level by analyzing the

## Risks, Threats and Vulnerabilities

The components of risk include the following items.

**Vulnerabilities:** This refers to weaknesses in an organization's defenses that might provide a foothold for an information security incident. Many of the vulnerabilities managed by security professionals are technical problems in information systems, such as a missing operating system patch or a misconfigured perimeter firewall.

Physical vulnerabilities, such as a faulty lock on a door or an inadequate fence around a secure facility, may also expose an organization's information to risk. Finally, personnel risks exist when staff or contractors are vulnerable to social engineering attacks, when attackers contact individuals and try to trick them into providing sensitive information, such as passwords or network configuration details.

**Threats:** This refers to external forces that undermine an organization's information security posture. Threats are either accidental, such as a hurricane bearing down on a data center, or malicious, such as a determined hacker seeking to gain access.

**Risks:** This refers to what occurs at the intersection of a vulnerability and a corresponding threat. Threats and vulnerabilities alone are not enough to jeopardize the security of an organization's information or systems. For example, the lack of a strong building shock-absorption system (a vulnerability) is not a risk in an area that is not prone to earthquakes (a threat).

Similarly, an attacker who knows how to exploit a buffer overflow in Windows Server 2008 (a threat) is not a risk to an organization with an infrastructure built upon Red Hat Linux, which is not vulnerable to that threat.

number and state of risks falling into each category. One commonly used taxonomy includes the five risk categories that follow.

**FINANCIAL RISKS:** These risks cause direct financial damage to an organization. Fraud, embezzlement and other forms of malfeasance are the most easily identifiable forms, although risks that jeopardize the organization's cash flow or investment strategies also fall into this area.

**STRATEGIC RISKS:** These risks jeopardize the organization's ability to fulfill its mission. For example, the loss of key faculty members is a strategic risk to an institution of higher education because it threatens the institution's ability to perform its core mission of teaching and research.

**OPERATIONAL RISKS:** These risks limit the ability of an organization to function daily. Children in colder climates

are very familiar with one such risk facing K–12 institutions: the risk of snow causing a closure because of difficult travel conditions. A school closing clearly jeopardizes the ability of the school to continue its daily operations.

**REPUTATIONAL RISKS:** These risks can potentially cause damage to an organization's goodwill among constituents, staff, vendor partners or other stakeholders. Government agencies that store sensitive personal information, such as Social Security numbers and financial data, may be immune from financial risks resulting from an inadvertent disclosure, but they are definitely subject to significant reputational risks.

**COMPLIANCE RISKS:** These risks result from laws or regulations that dictate certain aspects of an organization's information security practices. For example, organizations that process credit card information are subject to the compliance requirements of the Payment Card Industry Data Security Standard (PCI DSS). Similarly, agencies of the U.S. federal government must abide by the Privacy Act of 1974.

It is important to note that compliance with laws and regulations is usually the lowest common denominator of information security practice. Managing compliance is certainly necessary to stay in the bounds of the law. However, compliance alone is often not enough to provide a solid defense-in-depth approach to information security.

Although the categories present a useful way to classify and describe risks, most risks do not fit neatly into a single category. It is more common for risks to fall into multiple areas.

A security breach that jeopardizes a user's credit card information might be classified as a PCI DSS compliance risk, a reputational risk and a financial risk. In these cases, it makes sense to designate one primary category for the risk and designate the other categories as secondary risks.

## Risk Assessment Approaches

As with the less formal risk/benefit analysis technique discussed in the previous section, risk assessments make use of both quantitative and qualitative methodologies. Although quantitative approaches often make for easier decision-making, gathering all the data necessary to perform a rigorous quantitative risk assessment is a challenge.

Pinning down a quantitative assessment of reputational damage from a security breach is difficult. Quantitative risk management uses a formulaic approach to assessing the risks to an organization. Risk assessors gather or calculate each of the following measures as follows.

**ASSET VALUE:** AV is the underlying financial value of the asset subject to risk. Depending on the accounting practices and risk tolerance of the organization, this may be either the original cost of the asset, the depreciated book value or the replacement cost. For example, when assessing risks related to a data center, the replacement cost of that facility might be \$15 million.

**THE EXPOSURE FACTOR:** EF for a given asset/risk pair is the percentage of the asset expected to be lost if the risk materializes. In the assessment of flood risk to a data center, analysts might determine that approximately half of the facility would be destroyed, resulting in an exposure factor of 50 percent.

**THE SINGLE LOSS EXPECTANCY:** SLE is the financial loss incurred during one risk incident. It is calculated by multiplying the asset value by the exposure factor. In the data center flood example, the single loss expectancy would be 50 percent of \$15 million, or \$7.5 million.

**THE ANNUALIZED RATE OF OCCURRENCE:** ARO is the number of times that actuaries expect a risk to materialize in a given year. If the data center lies in a 100-year floodplain, actuaries expect that a flood will occur once every hundred years, making the annualized rate of occurrence 0.01.

**ANNUALIZED LOSS EXPECTANCY:** ALE ties together all of the above elements and provides the expected financial loss because of an action of a single risk upon a single asset in a given year. The annualized loss expectancy is calculated by multiplying the annualized rate of occurrence by the single loss expectancy. In the data center flood example, the annualized loss expectancy is \$75,000.

The ALE is an extremely useful measure for evaluating potential security controls. When faced with a decision regarding a control, analysts may simply compare the cost of the control with the ALE over the expected life of the control.

For instance, if a waterproofing company proposes a \$300,000 service to protect a government agency data center from floodwaters, the analyst should recommend the control if it is expected to last longer than four years, because the agency would expect to incur flood losses equal to four times the ALE (\$300,000) over those four years.

When it is not possible to perform a quantitative risk assessment, risk analysts often turn to qualitative techniques designed to describe relative risk to an organization. The most common qualitative risk assessment analyzes each risk on two dimensions: the probability that the risk will occur and the impact of the risk on the organization.

These are actually the same measures used in the quantitative approach, where the ARO is a measure of probability and the exposure factor is a measure of impact. The annualized loss expectancy combines probability and impact to derive a single risk measure.

Because qualitative risk assessment does not use numeric measures, analysts instead turn to categorical measures, often assessing risks as high, moderate or low on each dimension.

## Managing Risk

After completing a risk assessment, technology and organization leaders make decisions about the appropriate ways to manage the risks facing the organization. Although there are many possible variations of risk management techniques, they each fall into one of four general control strategies.

The first strategy, *risk avoidance*, is when an organization takes steps to remove itself from the situation that creates the risk in the first place. Returning to the risk model presented in the sidebar *Risks, Threats and Vulnerabilities*, the organization takes steps to remove the threat from the risk equation.

Risk avoidance is usually quite drastic, such as relocating a facility to avoid the risks associated with geographically specific natural disasters, or shutting down a risk-laden area of operation, such as a foreign campus of a university based in an unstable region of the world.

## Emerging Risks

Emerging technology often results in new areas of risk that security professionals must consider when designing a defense-in-depth strategy. There are three recent information technology trends of particular interest to risk managers:

**Cloud computing:** This approach to IT infrastructure lets organizations use servers located elsewhere on the Internet to store and/or process information. When an organization puts its data in the cloud, it must be careful to specify the terms in a written agreement.

**Virtualization:** This technology reduces data center costs by allowing multiple virtual computers to exist on a single hardware platform. Security analysts must be confident that the virtualization technology offers adequate isolation between guest operating systems, especially when several organizations share the same virtualization host.

**Remote access:** This innovation provides organizational resources to staff who are on the road. Virtual private networks (VPNs) should be deployed to provide secure remote connections by using encryption to tunnel private traffic over the Internet.

When evaluating any emerging technology for use in an organization, it is always helpful to return to the three categories of the information security triad – confidentiality, integrity and availability – and assess how the new technology might impact each of these principles.

*Risk transference* is the second approach to risk management. The primary tool organizations use to transfer risk is the purchase of an insurance policy that places the financial burden of a risk on a third-party insurance company. Unfortunately, it is difficult to use risk transference to manage some risk categories, such as strategic or reputational risk.

The third strategy, *risk acceptance*, is when an organization's leadership acknowledges a risk and decides that no further action is warranted. This often happens when management decides that the probability of the risk materializing is so low or the impact would be so marginal that it is not worth the cost of implementing another risk management approach. Risk acceptance should always be an active decision, rather than a passive practice that simply ignores a risk.

*Risk mitigation*, the final approach to risk management, is the most time-consuming because it requires the actual design and implementation of security controls. Risk mitigation installs controls that reduce the organization's vulnerability to a specific risk. For example, a data loss prevention system reduces the probability of a sensitive data breach, while the use of encryption reduces the impact of such a breach.

Progressive organizations treat risk management as a lifecycle process that never ends. Continuous risk assessments take place throughout the year, evaluating both new initiatives in the design phase and emerging threats, to maintain a comprehensive picture of risk.

## Data Loss Prevention

Technology professionals today have more data and applications to manage than ever before. The growth of cloud computing, remote and wireless access, removable media, and peer-to-peer computing creates unprecedented vulnerability to data loss. At the same time, identity thieves, organized crime and other hackers are growing increasingly sophisticated in their ability to illicitly acquire this information, substantially increasing the scope of the threat.

Referring back to the relationship shown in the *Risks, Threats and Vulnerabilities* sidebar, both the threat and vulnerability circles are growing larger, resulting in a proportionate increase in the total risk facing the organization.

Data loss prevention technology offers an additional tool for organizations to augment their defense-in-depth approaches to information security. DLP products monitor and protect sensitive data while it travels over the network or is stored either centrally or on an endpoint.

### Creating a Data Policy Framework

Before beginning a data loss prevention program, organizations should first adopt a supporting policy framework. As with any security control, the DLP product should be brought in to implement an existing policy designed

in collaboration with the organization's leaders. The data policy framework should incorporate the following elements:

**INFORMATION CLASSIFICATION SCHEME:** This is needed to describe the categories of information held by the organization. Typically, organizations create a scheme that has at least three levels: public information, internal information and sensitive information. Some schemes may add additional levels or use alternative names, but a three-level scheme is considered the most basic approach.

**DATA ACCESS POLICY:** This specifies how an individual gains permission to access certain types of information, calls for periodic reviews of access grants and revokes permissions that are no longer justified by operations requirements.

**DATA HANDLING POLICIES AND STANDARDS:** These denote the specific security controls that must be put in place for different levels of information classification. These controls might require that sensitive information leaving the organization must be encrypted.

### Another Way to Classify Data

Data loss prevention (DLP) works best when incorporated into a larger data management policy, according to research by the Aberdeen Group, an IT research company.

As part of that process, an organization must identify its high-value sets of information. Aberdeen suggests a different approach to classifying data than the three-tier scheme. It breaks data down into the following:

- Revenue generating (licensing fees or tuition, for example)
- Essential to future revenue (plans and intellectual property)
- Essential to operations (administrative or financial information)
- Protected by law (records containing personally identifiable information)

A DLP strategy must also account for data in various states, such as:

- At rest (in servers and storage systems within data centers)
- In use (when downloaded to a fixed-location PC or a mobile device)
- In transit (moving through the network infrastructure)

This data categorization approach identifies where mission- or operations-sensitive data resides and how it moves through the infrastructure. The infrastructure is considered the route to protecting information, not the object of protection itself.

Once an organization creates a data policy framework, DLP technology can assist with the technical controls supporting the framework. It might be configured to detect unencrypted sensitive information leaving the organization's network, for example, and either automatically encrypt it or block the communication.

## Data Loss Prevention Technology

DLP products are primarily network-based solutions, which typically target data in transit over the network, or host-based solutions, which target data stored on endpoints. Some products offer a hybrid approach that combines host- and network-centric DLP components under a single management interface.

Network-based DLP is typically deployed as a gateway that monitors all traffic entering or leaving the organization. A device identifies any sensitive information traveling in a manner that violates the organization's data handling policies.

For example, a government agency that routinely handles Social Security numbers might specify that this type of sensitive information should never leave the local network without the use of encryption. Network-based DLP might recognize SSN data and prevent it from leaving the organization's private network without appropriate encryption controls.

Host-based DLP is often used to identify sensitive information that has made its way to endpoints throughout the organization (as was the case at MIT). Host-based DLP products may recognize the presence of an organization's sensitive information on endpoints, referred to as inventory instances, and take appropriate action.

The detection techniques used by DLP products vary significantly in both effectiveness and ease of use. Common approaches to identifying the presence of sensitive information, either on the network or on the endpoint, include the following techniques.

**PATTERN MATCHING:** This technique takes advantage of known patterns found in sensitive data. It is most commonly used to detect instances of sensitive personally identifiable numbers, such as Social Security or credit card numbers. For example, a pattern-matching algorithm searching for SSNs would look for 9-digit numbers that are formatted in the pattern "xxx-xx-xxxx."

**KEYWORD MATCHING:** This technique monitors data for keywords indicative of sensitive information. For example, if an organization's data handling policies require that all sensitive information be labeled "Sensitive – Not For External Distribution," a keyword matching system can monitor for appearances of that phrase in documents leaving the organization's network.

**DOCUMENT FINGERPRINTING:** This approach requires that administrators develop an inventory of all sensitive information. The DLP then uses a cryptographic hash function to develop a fingerprint of each document known to contain sensitive information. When the DLP encounters an outbound document, it performs the same calculation and then compares the hash of the outbound document with the database of known sensitive documents. If there is a match, the document is treated as sensitive.

**PARTIAL DOCUMENT MATCHING:** This technique uses a more flexible but more computationally intensive approach. It checks the content of all outbound documents to see if any portions of it match content found in known sensitive documents. As with document fingerprinting, partial document matching requires an inventory of sensitive data as a starting point.

Each approach has both positive features and potential drawbacks. The pattern matching approach is easy to implement because it does not require an inventory of known sensitive information. However, it is likely to miss much sensitive data that is not formatted exactly according to the defined pattern.

Social Security numbers that omit the two hyphens would probably not be noticed by a pattern-matching DLP. Configuring the DLP to alert on any unformatted nine-digit number would result in an unacceptably high rate of false alarms.

On the other hand, document fingerprinting is extremely accurate at detecting the accidental or intentional loss of entire sensitive documents, but making even a minor modification to the document can easily foil it. Document fingerprinting also is not effective against sensitive information that has not been inventoried in advance.

Most commercial DLP systems balance the benefits and drawbacks of these detection techniques by using a hybrid approach that leverages each technique based on the circumstances.

DLP administrators might create an inventory of known sensitive information and configure the DLP to use document fingerprinting and partial document matching to detect the potential loss of documents stored in the inventory. Administrators might then supplement this approach with pattern matching that would identify unencrypted SSNs or credit card numbers in network traffic or on an endpoint.

Organizations seeking to evaluate and purchase DLP technology should ensure that they clearly understand the detection technologies implemented by systems under consideration. Organizations that have strong data inventory processes should favor systems that implement document

fingerprinting and partial document matching, while those with more unstructured data should emphasize pattern and keyword matching capabilities.

## DLP Configuration and Maintenance

Once a DLP system identifies a potential loss of sensitive information, there are many options available for handling the alert. DLP administrators may configure appropriate responses based on the organization's data handling policies.

The most passive DLP configuration simply monitors for potential violations and then triggers an alert when one occurs, without taking any action to prevent the data loss. This alert may be sent to a security administrator for investigation, or it might be sent to the end user to promote awareness. Alert-only mode is typically used during the testing and evaluation period to help assess the impact that DLP will have on an organization's daily activity.

More active DLP configurations implement aggressive responses when the system identifies a potential data loss. Potential actions might include:

- Blocking network traffic that contains sensitive data from leaving the organization;
- Quarantining files discovered on an endpoint and requiring user action to release them;
- Automatically encrypting files or network traffic containing sensitive information in a manner that complies with the data handling policy.

The appropriateness of each response will vary depending on the security policies of the organization and the sensitivity of the information discovered.

DLP systems should also offer reports that let administrators monitor all DLP activity from a centralized location. These reports offer a valuable source of information for the ongoing tuning and maintenance of DLP activity.

Administrators should routinely monitor system logs for evidence of false-positive alerts and tune the system to reduce those inaccurate detections as much as possible.

## Pattern Matching and the Luhn Algorithm

DLP systems tend to have much lower false-positive rates when searching for credit card numbers than when trying to identify Social Security numbers. This can be attributed to the Luhn algorithm, a formula that credit card processors use to screen out patently false numbers and flag typographical errors.

Here's how it works:

1. Starting with the first digit, sum together every odd-numbered digit in the candidate card number (the first, third, fifth digits and so on).
2. For each even-numbered digit, double the digit and add the two numbers of the result (for example, 9 doubles to 18 and  $1+8=9$ ). Then add those sums together for all of the even-numbered digits.
3. Add together the results of steps 1 and 2.
4. If the result from step 3 is evenly divisible by 10, the card number is potentially valid. If it is not evenly divisible by 10, it is definitely not a credit card number and can be eliminated as a false positive.

Of course, the Luhn algorithm identifies only whether a credit card number is feasible. It can't determine whether the account is actually valid, but it serves as a great check to reduce DLP false positives. Unfortunately, there is no similar algorithm to help with Social Security numbers.

High rates of false positives that block legitimate activity are one of the most common causes contributing to DLP project failures.

DLP systems play an important role in adapting perimeter-based defense-in-depth security approaches to threats from mobile and wireless computing, cloud computing and other evolving technologies. These systems let administrators shift attention from building walls around their networks to implementing varied levels of security control based on the sensitivity of the data in use.