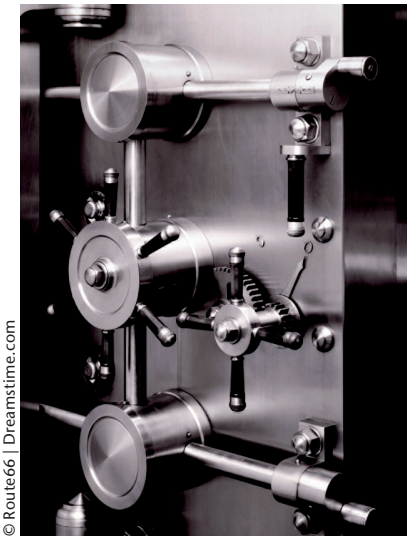

Data Loss Prevention

Simon Liu, *US National Agricultural Library*
Rick Kuhn, *US National Institute of Standards and Technology*

In today's digital economy, data enters and leaves cyberspace at record rates. A typical enterprise sends and receives millions of email messages and downloads, saves, and transfers thousands of files via various channels on a daily basis. Enterprises also hold sensitive data that customers, business partners, regulators, and shareholders expect them to protect. Unfortunately, companies constantly fall victim to massive data loss, and high-profile data leakages involving sensitive personal and corporate data continue to appear (http://opensecurityfoundation.org). Data loss could substantially harm a company's competitiveness and reputation and could also invite lawsuits or regulatory consequences for lax security. Therefore, organizations should take measures to understand the sensitive data they hold, how it's controlled, and how to prevent it from being leaked or compromised.

The Data Loss Problem

According to the Open Security Foundation, which tracks publicly reported incidents, 714 cases of data loss were reported in 2008, affecting a total of more than 86 million records (http://datalossdb.org/yearly_reports/dataloss-2008.pdf). Depending on the type of data loss experienced, an organization can suffer a variety of consequences, but in nearly all cases, it's both a financial and reputation cost.

Types of Loss

We can divide data loss into two sometimes overlapping categories:

- Leakage, in which sensitive data is no longer under the organization's control (in computer security parlance, this is a loss of confidentiality). This common form of data loss is often due to hacked customer databases, making its most common consequence identity theft. In the largest single attack of this type to date, hackers stole 130 million credit-card records from one of the US's largest payment processors (datalossdb.org). Another involved 94 million customer records held at a major retailer.
- Disappearance or damage, in which a correct data copy is no longer available to the organization (corresponding to a compromise of integrity or availability). An example occurred in 2009, when a major cell phone service provider suffered widespread loss of customer data that was supposed to be housed by a third-party cloud-based storage service. In normal operation, the smart phone would automatically sync its data at power-off with the central server, which stores it for use when the phone is on again. For reasons that still aren't fully known, a server crash at the storage service temporarily wiped out backups of memos, photos, and other data for more than a million smart phone customers (see "Sidekick Customers Can Recover Contacts," *The Wall Street J,* 20 Oct. 2009; http://blogs.wsj.com/digits/2009/10/20/sidekick-customers-can-recover-contacts/).

Clearly, if the last accurate data copy is physically stolen, the organization faces both problems. Alternatively, it might not be immediately clear which of these situations pertains in some cases. For example, a common problem for enterprises is laptop theft or loss. If an employee was updating or editing information on a laptop

 Published by the IEEE Computer Society

using multiple data sources, his or her copy might be the most current. Without the assurance of accurate records management, the firm might not be able to determine which records are correct.

Consequences of Loss

As with other security incidents, data loss incidents can result in significant cost, but the duration and magnitude vary with the type of data loss. Financial records can usually be reconstructed, and any fraud incurred might not result in a loss to the customer if national laws require financial institutions to bear this cost instead. Costs to the organization might be much more severe and could include liability costs that aren't always covered by corporate insurance policies.

Although loss of payment processing data might require years to repair, consumers generally are able to clear up problems and recover financial losses. Today's movement toward extensive use of electronic medical records can, however, present a new class of risk for both the consumer and the organization. For these records, the risk is to privacy, so if records become public, the damage to the individual could be permanent rather than temporary as with some fraudulent credit-card charges. Consequently, the organization could face increased litigation or regulatory consequences.

Why Data Loss Prevention?

Key drivers of establishing data loss prevention mechanisms include government or industry rules and intellectual property protection.

Government and Industry Requirements

Today, many companies fall under the oversight of government and industry rules that mandate controls on information in general and personal identifiable information in particular. Major US requirements include the following, and most nations have similarly strong rules:

- The Health Insurance Portability and Accountability Act of 1996 requires that to ensure privacy and confidentiality, all patient healthcare information must be protected when electronically stored, maintained, or transmitted.
- The Gramm-Leach-Bliley Act of 1999 mandates privacy and protection of customer records maintained by financial institutions.
- The Privacy Act of 1974 prohibits disclosure of information in personal records by any means of communication to any person or agency, except pursuant to certain statutory exceptions or to a written request by, or with the prior written consent of, the individual to whom the record pertains.
- The Federal Information Security Management Act of 2002 provides a comprehensive framework for ensuring the effectiveness of information security controls on information resources that supported federal operations and assets.
- The Payment Card Industry Data Security Standards helps organizations that process card payments prevent credit-card fraud through increased controls around data and its exposure to compromise.

Government and industry requirements are arguably the biggest drivers of data loss prevention. In addition, many states have passed data privacy or breach notification laws that require organizations to notify consumers when their information might have been exposed.

Intellectual Property Protection

According to the World Intellectual Property Organization (www.wipo.int/portal/index.html.en), intellectual property includes creations of the mind—inventions, literary and artistic works, symbols, names, images, and designs used in commerce. For many enterprises, intellectual property might be more valuable than its physical assets. Consequently, establishing policies and mechanisms for guarding against intellectual property loss or theft is critical to protect the brand and maintain competitiveness.

Today's movement toward extensive use of electronic medical records can present a new class of risk for both the consumer and the organization.

Data Loss Prevention Approach

Data loss prevention is an enterprise program targeted on stopping various sensitive data from leaving the corporation's private confines. With recent high-profile data loss incidents coming to light, data loss prevention technologies are emerging as important information security and privacy controls.

Loss Modes

Enterprise data generally exists in the following three major states:

- *data at rest,* meaning it resides in files systems, distributed

desktops and large centralized data stores, databases, or other storage centers;
- *data at the endpoint,* meaning it resides at network endpoints such as laptops, USB devices, external drives, CD/DVDs, archived tapes, MP3 players, iPhones, or other highly mobile devices; or
- *data in motion,* meaning it moves through the network to the outside world via email, instant messaging, peer-to-peer (P2P), FTP, or other communication mechanisms.

Data in each state often requires different techniques for loss prevention. For example, although deep content inspection is useful for data in motion, it doesn't help so much for data at rest. Therefore, an effective data loss prevention program should adopt appropriate techniques to cover all the organization's potential loss modes.

To work effectively, a data loss prevention solution must operate without diminishing system performance.

Solution Capabilities

An effective data loss prevention program should consist of the following essential capabilities:

- *Manage.* Define enterprise data usage policies, report data loss incidents, and establish incident response capability to enable corrective actions that remediate violations. Data loss prevention isn't just a technology issue—it's also a policy and policy management issue. Enterprise data usage policies should address issues such as how data access is determined and authenticated and how policies are enforced. Management functionalities should also include data loss reporting capability and incident remediation workflow management.
- *Discover.* Define the sensitivity of enterprise data, create an inventory of sensitive data, locate sensitive data wherever it's stored, and manage data cleanup. This includes discovering and inventorying sensitive data at rest or stored on the endpoint to inventory, secure, or relocate that data.
- *Monitor.* Monitor the use of sensitive data, understand sensitive data usage patterns, and gain enterprise visibility. This could include monitoring data in motion by inspecting network communications in violation of data security policies and monitoring data at the endpoints to see if it's downloaded to local drives, copied to USB or other removable media devices, burned to CD/DVDs, and printed or faxed electronically.
- *Protect.* Enforce security policies to proactively secure data and prevent it from leaving an enterprise. Automatic protection of sensitive data across endpoint, network, and storage systems should include protecting data at rest with automatic encryption, quarantine, and removal. Restrict printing, saving, copying, accessing, moving, and downloading sensitive data to removable media or other drives. Stop data in motion from being sent in violation of security policies or encrypt it for secure exchange.

When properly integrated, these four essential capabilities offer effective protection of valuable information assets.

Best Practices

Data loss prevention is a complex issue with no single effective solution. Organizations should leverage best practices and seek out a data loss prevention solution that best suits their particular needs.

Prioritize loss modes. Although a comprehensive program to address all relevant aspects of data loss is the ultimate goal, it makes far more tactical and financial sense to begin by protecting the data that represents the most danger to the enterprise. This means first identifying all the potential data loss modes and then prioritizing them based on criteria such as past breaches, communication volume, data volume, the likelihood of a breach, and the number of users with access to those modes. Focusing first on the most significant and highest impact areas makes it easier to justify solutions and get started on plugging the leaks.

Protect without disruption. Data loss prevention solutions shouldn't interrupt legitimate business activities. To work effectively, a data loss prevention solution must operate without diminishing system performance or preventing workers from doing their jobs. Solutions that don't scale can cause performance issues as companies grow. Solutions that aren't properly tested and tuned can also cause both false positives and false negatives that drain valuable resources.

Flexible and modular architecture. Solutions for data loss prevention are still evolving, with

no single option providing all the capabilities that most organizations require. Enterprises need to address the data loss problem by creating a flexible and modular architecture that lets them immediately and cost-effectively address their most pressing requirements while still being able to add new controls as those needs change. It also ensures speedy deployment, protects investments, and easily scales to accommodate expansion and growth.

Data loss prevention is a serious challenge for companies as the number of incidents continues to increase. There's no silver bullet, either—identifying and blocking all sensitive data is neither possible as an outcome nor wise as a goal. However, with a more focused goal of preventing the most damaging leaks and establishing better ways for users to exchange information securely, data loss prevention can be effective, practical, and successful.

Acknowledgments

We identify certain products in this document, but such identification doesn't imply recommendation by the US National Institute of Standards and Technology or other agencies of the US government, nor does it imply that the products identified are necessarily the best available for the purpose.

Simon Liu *is the director of the US National Agricultural Library. His research interests include IT architecture, cybersecurity, software engineering, and database and data mining. Liu has two doctoral degrees in computer science and higher education administration from George Washington University. Contact him at simonyliu@yahoo.com.*

Rick Kuhn *is a computer scientist at the US National Institute of Standards and Technology. His research interests include information security, software assurance, and empirical studies of software failure. Kuhn has an MS in computer science from the University of Maryland College Park and an MBA from William & Mary, Mason School of Business. Contact him at kuhn@nist.gov.*