Do You Know Where Your Data Are?

A useful new tool in the information security kit—outbound content monitoring and filtering—can help prevent sensitive data loss

By **Cedric Bennett**

Many of the information security appliances, devices, and techniques currently in use are designed to keep unwanted users and Internet traffic away from important information assets by denying unauthorized access to servers, databases, networks, storage media, and other underlying technology resources. These approaches employ firewalls, intrusion detection and prevention systems, reverse proxies, and event logging, for example, as well as processes designed to ensure that software does not contain known flaws that might allow miscreants to gain illicit access. Other techniques prevent unauthorized individuals from gaining system or data access through legitimate entry points. These techniques rely on passwords, multifactor logon, virtual private networks (VPN), identity management systems, and the like.

Protecting Sensitive Data

Axiomatically, no security system can be completely effective. Security professionals must assume that criminals will find a way past protections or that inappropriate behavior (accidental or otherwise) by trusted insiders will compromise critical data. To improve the safety of critical information assets, security professionals should consider adding additional defensive layers that focus on protecting the data directly. Best known among these technologies is encryption, which renders data unintelligible to anyone who accesses them except those who have the decryption key. Less well known but beginning to gain acceptance as an adept and versatile tool is outbound content monitoring and filtering, a technique that can prevent sensitive data from leaving authorized locations.

Encryption

Encryption has been used for many years to protect data during transmission, using techniques such as VPN connections. Nearly all business-oriented Web sites use encryption, at least when private or other sensitive information is transmitted, by applying secure socket layer (SSL) technologies. This use of SSL, which encrypts and decrypts data as they enter and leave the network, is useful for protecting data in transit. However, it does nothing to protect the data at rest at either end of the connection.

Encryption technology is also used to protect stored data. Using combinations of hardware and software, data are encrypted before storage and decrypted upon access. This technique has the additional benefit of protecting data transferred to another medium or device where they might not be as well protected as at the primary storage location.

For example, if backup media of encrypted data are lost, misdirected, or even stolen while in transit to a remote storage facility, the loss will rouse less concern because the data can't be read by anyone without the key. Similarly, if a laptop computer containing an encrypted copy of a database that includes sensitive information is lost or stolen, the data are still protected from whoever obtains that computer. To the extent that everyone in an institution follows information security policies, encryption is an effective security layer oriented specifically to protecting data.

Unfortunately, any of the vulnerability mediation methods discussed above are powerless against errors or malicious acts committed by a trusted insider—a person granted access to sensitive information. Once access has been permitted, damage can be done, even if unintended.

Because the Internet allows us to distribute work tasks and data to any place they are needed, it is ineffective to depend solely on information security policies to prevent problems—there are simply too many opportunities for data to end up in unexpected places. Over time, copies of sensitive information will end up stored in many different forms on many different devices and in many different geographic locations. The more widely the data are distributed, the more likely that one or more of those instances of the data will not be protected well.

For example, an analyst might have brought a large segment of sensitive information into a spreadsheet to explore the institution's business processes. Or, a department manager might have e-mailed sensitive information to someone in a central organization when asking for advice on a staffing issue. None of these data are likely to remain encrypted during such transactions, and they will reside on computers and in

locations where data protection, in the form of other security layers, is probably less rigorous than in the original location.

Outbound Content Monitoring and Filtering

As Internet use broadens further and local storage capacities grow larger and less expensive, the problem of widely distributed (and possibly forgotten) copies of sensitive data increases. "Data leakage" is the term often used to describe this particular problem. The data that might leak include not only very structured private information about people but also less-structured data that might harm the institution if released. In private enterprise, this could be trade secrets such as a formula for a new pharmaceutical, innovative software code implementing some new idea, or key customer information. In higher education, it might be correspondence with a major donor, unpublished research findings, courseware under development, information about controversial research, financial aid award letters, or the draft of a new book.

Outbound content monitoring and filtering data-protection technology has been available for only a few years and is just now showing signs of maturation. Basically, it acts something like a reverse firewall, examining packets of information just before they leave the institutional intranet to determine if they contain sensitive information. The device can be set to report a problem and let the information continue on its way or to stop the data and alert the sender and optionally the appropriate data owner. Table 1 shows vendors offering content monitoring and filtering products.

Unlike a firewall, however, this technology does not operate at just the level of protocols, ports, and IP addresses but also on the content. Effective products apply a variety of techniques for identifying sensitive data, from simple key words and pattern matching to complex signatures and linguistic analysis. Well-designed content monitoring and filtering products can find sensitive information in unstructured and structured data forms and identify fragments as well as complete segments of sensitive data.[1] Really effective systems can be set to detect when sensitive information is about to be printed, burned onto a CD, or copied to a flash drive. Whether a potential information compromise arises because someone has cracked into a computer and located the information or because someone is inadvertently file-transferring or e-mailing sensitive information outside the organization, outbound content monitoring and filtering can detect the data leakage and stop it.

Table 1

Outbound Content Monitoring and Filtering Vendors

Vendor	Network Filtering	Endpoint Scanning	Desktop/ Laptop Filtering
PortAuthority Technologies (http://www.portauthoritytech.com/)	X		
Reconnex Corporation (http://www.reconnex.net/)	X		
Tablus, Inc. (http://www.tablus.com/)	X	X	X
Vericept Corporation (http://www.vericept.com/)	X		X*
Vontu, Inc. (http://www.vontu.com/)	X	X	

* Capability acquired; no product yet available

Initial implementations of this technology have occurred primarily in the private sector, in pharmaceutical companies and software development shops, as a way to prevent accidental loss of key intellectual property. As the technology has matured, it has been used to protect against the accidental leakage of sensitive data covered by statute, such as personal health information or personal identity information. Higher education has just started to pay attention to it as another useful security layer.[2]

Content monitoring and filtering combined with data encryption deserve careful consideration when thinking about risk mitigation as well, particularly to prevent the exposure of critical data. They may be just the security tools needed to help reduce unauthorized access to sensitive information and to reduce the frequency of headlines about serious data breaches at our colleges and universities.

Endnotes

1. This means, for example, that it can protect against the loss of intellectual property that might be found in documents and electronic mail as well as sensitive information about individuals that might be found in databases or spreadsheets.
2. The Institute of Social Research at the University of Michigan has installed Content Alarm from Tablus, Inc., as a way to protect sensitive information such as personally identifiable information of survey participants and the premature leakage of consumer sentiment survey results.

Cedric Bennett (Ced.Bennett@Stanford.edu) is Emeritus Director, Information Security Services, at Stanford University in Stanford, California, and still spends considerable time helping Stanford and other institutions address their information security requirements through writing, teaching, presentations, and consultation.