



A Better Method to Contextually Protect Data

A new solution to data security is predicated on -- and anticipates -- human behavior by embedding corporate policies into documents and images.

. 11/17/2009

Protecting classified data from leaking -- intentionally or unintentionally -- into the wrong hands has always been a challenge. File sharing and collaboration software, such as SharePoint, host millions of documents in need of classification. However, the daunting task is to discover and properly group them, as well as secure data, in accordance with compliance regulations, while the information is still in use.

This on-going hazard to organizations is underscored by Gartner. According to Gartner's *Magic Quadrant Content-Aware Data Loss Prevention Report* (June, 2009), the data loss prevention (DLP) market will reach \$300 million in 2009. One of the top drivers propelling the \$300 million investment is the ever-present industry and government compliance regulation specter that hovers over most IT infrastructures. The increasing need to fulfill regulatory requirements for Sarbanes Oxley, HIPAA, Gramm-Leach-Bliley Act, and PCI, are perpetually growing. Gartner's report also underscores the growth of DLP within this recession stating, "The reasons for the continued strength of this market include the growing maturity of the available content-aware DLP technologies and buyer awareness of these technologies can help address regulatory compliance, which are actually increasing in the downturn."

"Data in use" (files that employees are editing in MS Word, Excel, and PowerPoint, etc.) has always presented a significant challenge to securing corporate information. As documents evolve from outline drafts into mature, final pieces, they often pass from one person to another. Each person contributes specific data or context -- transforming a simple draft outline into a potential compliance risk or security breach -- in a matter of seconds. Managing data-in-use relies on the assurance solution to be aware of its content, its context, and its usage patterns in real-time.

Today's DLP solution needs to follow the evolving data and automatically add stronger security features as information is added. This procedure is in contrast to legacy DLP solutions that focus on data-at-rest or data-in-motion. In these two examples, the data is stored on a server, laptop, or mobile phone or transported from one device to another via e-mail or instant message -- making it difficult to secure information with corporate policies. In essence, since the content of most corporate documents is dynamic in nature, the DLP security features must be, too. The security features must be administered before the end user has the ability to open, edit, or transfer the information.

Consider these two cases:

- March 2009: Classified data on the U.S. President's helicopter leaked from a computer belonging to a Bethesda, MD. military contractor and was discovered in a publicly available shared folder on a computer in Tehran, Iran. The location where the file was found included several other documents with classified and sensitive military information that were also leaked over file-sharing networks.
- May 2009: The U.S. government mistakenly made public a 266-page report, its pages marked "highly confidential," that gives detailed information about hundreds of the nation's civilian nuclear sites and programs, including maps showing the precise locations of stockpiles of fuel for nuclear weapons. The report was intended for the International Atomic Energy Agency for purposes of nuclear transparency but was not intended for consumption by the general public. The document appeared to have been published by the House Committee for Foreign Affairs, but spokeswoman Lynne Weil has denied that her committee was responsible.

Companies cannot be expected to police the key strokes of their employees to determine the line between benign documents and high-security-risk documents. There are just not enough eyeballs to watch it in real-time! Nor can employees be expected to review a phone book-sized manual of corporate or government regulations prior to altering, filing

or sending every document.

The perplexing question remains: How do you determine the exact moment when an employee transforms a document into a compliance or security risk if leaked to the wrong individuals or misfiled to the wrong server?

According to Forrester Research, an estimated 80 percent of information security breaches result from human errors as simple as accidentally posting a spreadsheet with customer Social Security numbers on a public Web site -- or a business traveler losing a portable data storage device with thousands of documents and hundreds of thousands of e-mail messages. One elegant solution to prevent security breaches is to let the document actually dictate all the corporate and compliance regulations to the employee -- based on the document's changing context —at the moment they try to access it.

Unlike traditional DLP or information rights management (IRM) solutions, there is a new breed of offerings that reside between the DLP side and the access-control logic from the identity management side. These products offer a new identity-aware approach to secure unstructured data while sensitive to the business-process *context*. For example, such a new solution to manage unstructured data securely offers:

- Visibility into all documents at the time they are used (context) without restricting or slowing down productivity
- A process to hold employees accountable without requiring them to do anything outside their normal course of action when retrieving or retaining data -- no end-user training required
- A means to cluster or group documents into specific categories to support enforcement of user security rules or compliance mandates and regulations, reducing management effort
- A means to tag files to follow the lifecycle of documents and ensure they don't fall into the wrong hands
- The ability to establish an audit trail to hold the appropriate person accountable for the security breach

The net result is effective data flow control and security within the context of a document and specific business processes.

A closer look into the core architecture finds a blended process of:

- E-discovery
- Real-time context-based classification
- Real-time policy decision
- Automated tag and sign with official and accurate "chain of custody"
- Coordination of enforcement
- Audit, monitor, warn, act, and report
- Identify the pedigree, or genetic, history of files

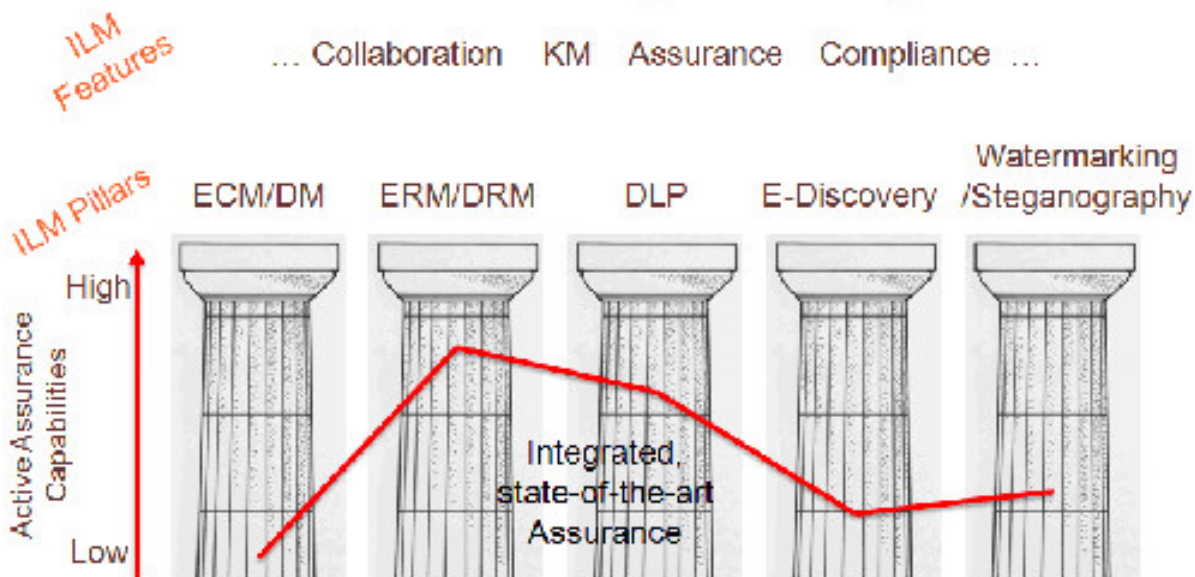
The following diagram describes the key facilities necessary to mitigate information risk by overlaying the new, integrated file assurance capabilities into today's existing IT infrastructures.

Figure 1:
Information
Lifecycle
Management

Information Lifecycle Management

... Collaboration KM Assurance Compliance ...

The new integrated process for securing unstructured data implements an e-discovery approach first. Financial, health care, HR/employee and product information, for example, is often scattered across the entire network and





takes many forms,

such as Excel spreadsheets, Microsoft Word documents, PDF files, PowerPoint presentations, CAD/CAM diagrams, JPEG images, MPEG video, and others. A scanning of all these items -- looking into the content -- takes place to automatically group or cluster all these items into one secure area.

Once grouped into business context, the files pass through the assurance system where they are assigned a tag (or sometimes referred to as a "barcode" or "watermark") that embeds selected corporate policies and file usage history -- aka "chain of custody" -- into a file's metadata. These embedded policies travel securely and safely with the file and are transparent to end users, who are presented with log-in panel where they must enter credentials to view, manipulate, or download any file, depending on their security credentials.

When the authorized users are finished with a file, it travels back to its server of origin, first passing through the assurance system, where the metadata containing the "who, what, when, where, and how" is stripped off, and the file is stored in its pristine, native format. As the file content or context changes, the metadata captures it and applies specific policies relative to keywords. Auditing and reporting is administered by accessing the metadata in the assurance system to view a document's usage lifecycle.

This process may easily apply to help strengthen Payment Card Industry Data Security Standard (PCI DSS) mandates. For example, the problem for any organization required to meet PCI DSS compliance standards is that all relevant data needed to manage documents does not simply, or neatly, fit within a single database systems. Compliance data can also reside in unstructured data such as forms, documents, and spreadsheets -- or even PDFs and JPEG images -- scattered throughout the organization, on multiple data stores and sites, a.k.a. server sprawl.

Also, credit card or client authentication information (such as driver's licenses, national health numbers, and passport numbers) within a document may trigger multiple compliance requirements, not to mention the value of the transaction might trigger a different workflow approval processes.

Failure to prove PCI DSS compliance translates into the possibility of losing the organization's ability to process credit card payments and/or be fined up to \$500,000 per incident in the United States. Although organizations with significant PCI transactions are subject to annual audits from PCI DSS qualified security assessors (QSAs), smaller companies (those processing fewer than 80,000 transactions a year) are allowed to perform a self-assessment questionnaire.

However, the PCI compliance example is only a piece of the overall data security puzzle that must be addressed. Consider pictures, videos, and medical images. Discovering, classifying, and securing media present an even greater challenge -- especially in the video game industry.

Recall that leaks were plentiful ahead of the Sony keynote at E3 in June, 2009. Before the show even kicked off, details and an image of the PSP Go showed up online and trailers for upcoming Sony games *Metal Gear Solid: Peace Walker* and *The Last Guardian* were leaked. Sony Computer Entertainment America head Jack Tretton was not happy about the leaks and feels that they stole some of the thunder from the Sony E3 announcements. Tretton told CNBC, "People don't respect confidentiality in this industry. It's tough enough to keep a secret within your own company, much less when you speak to third parties."

Protecting images such as video games and medical images are a particular challenge because they often seem harmless, but could produce devastating results -- such as loss of revenue, privacy, and national security issues -- if placed into the wrong hands. Unfinished or unapproved movies can have a negative impact on a consumer's decision to purchase a video game or attend a movie. These first drafts images often look unpolished but are taken as the final or corporate approved marketing material.

Enabling visibility into who accessed a picture or video, what they did with the file, when the file was accessed, where the file was saved, and how it was altered are facts that are just as important as information contained within a spreadsheet, a Word document, CAD drawing, or a PDF file.

To protect these corporate multimedia assets from security breaches, the same process as outlined in the PCI example applies, but this time the files are embedded with an invisible, or visible, digital watermark placed on the image. An organization is now able to view an image's metadata and search all file storage devices (such as WebDAV, Windows

file shares, SharePoint, NFS, Novell, and legacy systems) to locate and group any image according to predetermined criteria. Once the images are clustered together, additional compliance or corporate policies may be seamlessly embedded into the metadata -- in the same manner as applicable to Microsoft Word documents, PDF files, etc. Tags will follow the image where additional processes take over to automatically administer security policies prior to an end user opening the file. A complete audit trail is detailed, showing who viewed, manipulated, forwarded, or stored an image to a different server.

Utilizing these security features, Sony's security breaches may easily have been avoided. In the scenario of the federal government mistakenly publicizing a 266-page report showing maps and precise locations of stockpiles of fuel for nuclear weapons, these same procedures could have prevented the leak by securing the maps and images and automatically enforcing security policies before an end user forwards or even views the documents.

Summary

People often do not think twice before publicizing or filing documents. A new solution is available and is currently being deployed. Such a solution remains one that is predicated on and anticipates human behavior while counteracting its shortcomings by embedding documents and images with corporate policies that dictate the file's proper use according to corporate and government/industry policies and mandates. These files also contain the official history of usage on how the document arrived where it is stored.

In other words, sensitive data may attempt to leave the IT domain of an organization, but the tracking, policies and control never leave the file. IT nirvana is obtained when the end user simply conducts "business as usual" by (transparently) using a file's metadata without additional software installation on the users' desktop. All documents and images will inform the end user if they can be viewed, downloaded, forwarded, or re-filed to another location. All activity is neatly tracked and reported back to the administrator who can quickly run reports to see trends of possible security concerns or map corporate policies to government and industry compliance mandates. Knowledge workers continue their business as usual while the organization increases its ability to manage risk to its business introduced by spreadsheets, documents, presentations, media files, engineering drawings, medical images, and other electronic files used to run the business.

Rob Marano is president and chief executive officer of InDorse Technologies. His experience focuses on developing early-stage companies from start-up to IPO, having served several key roles in development, professional services, and sales. Mr. Marano holds a Bachelor of Engineering degree from The Cooper Union and a Master of Science degree from the University of Pennsylvania, both in electrical and computer engineering. He continues to lecture on computer engineering and collaborative, international business at The Cooper Union and the NYU Polytechnic Institute. You can contact the author at rob@cooper.edu