

## **Hidden data: You may be sharing more than you think**

By RONALD HACKETT

August 21, 2006

In April 2005, when the Defense Department's Multi-National Force — Iraq posted a redacted report on the death of Italian secret agent Nicola Calipari in Iraq, a group of Pentagon Web site visitors from Italy could copy and paste the classified portions from Adobe Acrobat Reader from the Web site into a Microsoft Word document, including the name of the U.S. soldier who accidentally killed her.

Last December, Web surfers found out from the posted White House policy document "Strategy for Victory in Iraq" who the report's author was, causing some embarrassment to the Bush administration.

Sensitive government procurement documents, such as requests for proposals, frequently have tracked changes, comments and other hidden data that could give savvy contractors an unfair advantage. In one case, an incumbent contractor was barred from participating in the new procurement after the electronic document's metadata revealed it had written the statement of work.

How did these things happen? For years, agencies have taken precautions to secure their physical and electronic repositories of confidential and classified information. But there's another, often overlooked vulnerability that federal agencies have only recently begun to recognize: hidden electronic data.

Sharing electronic files with outside parties has become a daily activity for many government employees, and in certain cases — such as in Freedom of Information Act requests — government employees need to make redactions to remove sensitive and classified information.

So what's at the root of the problem? Software applications provide tremendous user functionality. To retain ease of use, the application handles most of the details in the background. This usually requires the application to make the most functional assumption about the data. For example, inserting a chart into a document usually results in the original spreadsheet being embedded into the document.

A single user can create vast amounts of hidden data, but many documents have multiple authors. These authors have varying degrees of expertise, and some may intentionally create certain data structures unknown to the others. Inexperienced users may share sensitive data because they don't know how to look for it.

Many users start with an existing document and modify it for their needs. Unfortunately, the existing information may remain in the document. A properly trained expert can recover that "deleted" information.

Software developers are not helping. To induce customers to purchase new versions, they pack their applications with new features but rarely consider the security ramifications. Microsoft's Ad Hoc Review feature was added to Microsoft Office XP and 2003, for example. It uses the Excel, PowerPoint and Word's tracked changes feature to keep track of the document review process. To get customers to adopt it, Microsoft turned it on by default. E-mailing a Word, PowerPoint or Excel document using Outlook automatically enables the Ad Hoc Review.

Many security experts recommend using Adobe PDF documents, but they can contain deleted text, images and even entire pages. Annotations (similar to Microsoft's comments feature) can be hidden, and may contain metadata and hyperlinks that expose the document's origins.

The plan for victory in Iraq is a good example. Hidden data reveals that it was originally a Word document written on a classified system. There is an embedded hyperlink to the classified CIA Factbook via a classified network. The hidden data also reveals that the document was modified several hours after an unfavorable article appeared that revealed the document's author, which is also in the hidden data.

Many keyword scanners commonly used to screen information may not detect much hidden data. Compression and encoding techniques obscure a lot of the information from these keyword scanners.

Guard technology, used to screen electronic documents crossing security boundaries, relies on keyword scanners and may not screen the electronic documents.

Information-sharing has already created a vulnerability that needs to be addressed. Classified data spills are common, but the problem may be even worse. Because hidden data can be difficult to find, many classified data spills may go undetected.

Federal chief information officers should implement policies and procedures to ensure that documents with classified or confidential data are properly sanitized of all hidden data before being shared. Continuing to do business as usual is dangerous and could be giving our adversaries a tremendous advantage.

**Ronald Hackett**, a retired Air Force major, is program manager for SRS Technologies Inc.'s Systems Solutions Division.

<http://www.federaltimes.com/index.php?S=2044545>