National Security Cyberspace Institute

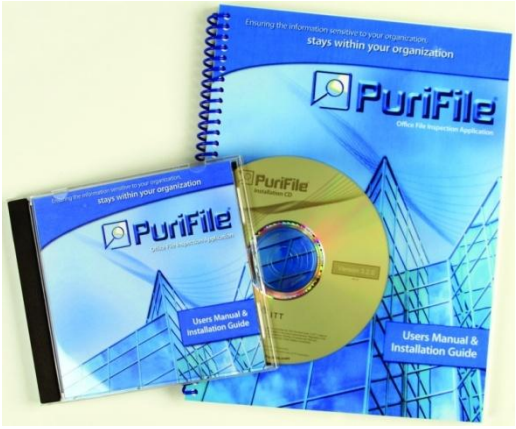
CyberPro

February 2009

Keeping Cyberspace Professionals Informed

CAPABILITY SPOTLIGHT: PURIFILE SOFTWARE EXPOSES HIDDEN DATA

BY JOHN IVORY, ITT



The efficient and secure exchange of information is the cornerstone of almost any successful operation. This truth is inclusive of both government organizations exchanging mission critical details and private sector commercial information exchange alike.

By far, the most common currency for these transactions is Microsoft Word, PowerPoint, Excel and Adobe PDF files. Unfortunately, each one is a virtual Pandora's box of possible dangers. There is a broad spectrum of ways in which data can be hidden in each of these file formats, making a reliable review, sanitation and release procedure almost impossible.

Fortunately, ITT's PuriFile software product is designed to directly address and solve this problem. Created with input and funding from the US government, PuriFile allows sites to define custom inspection rules and provides users with a simple mechanism for reviewing and correcting security issues.

The Issue at Hand

In rough terms, there are three ways in which extra information can come to exist inside a document. The first method is the most well documented – metadata. Microsoft Office and Adobe both store all sorts of extra information about a document within the file itself. Left un-monitored, details such as who document authors and reviewers were, or system information about how the document was created, linger within the file.

A broader category of security problems exists by way of unintentional inclusion. Information resides inside these file formats as a simple byproduct of normal use of the creating program. Some of these byproducts are well known, but many are not. Some examples include:

- Full images that remain in files, independent of how they are cropped
- Deleted content that continues to exist in Microsoft Office files after editing, as a result of either deliberate use of track changes or as a byproduct of how files are written
- Copy-and-pasted material from one document to another that brings along the entire document, not just the part expected (This is most easily demonstrated by copying a pie chart from Excel and pasting it into a PowerPoint presentation. Unless special steps are taken, the entire Excel spreadsheet gets included inside the resultant file.)

Final categories of concern are those situations when the user in some way created a document where a certain amount of content is not easily discovered. This can happen accidentally or deliberately and maliciously, and includes situations such as using font sizes too small for viewing, having objects off the side of the printable/viewable area, or having objects overlay and obscure others. However these situations come about, ITT's PuriFile software makes inspection and correction simple.

A Point Solution

PuriFile is typically installed on a single system at the customer site, allowing the inspection capabilities to be centrally administered and managed. Users access the inspection service through a number of possible routes.

Although the product supports inspection of files through a robust web interface, the most popular mechanism is through interfaces within Word, PowerPoint and Excel. Using an integrated "Assistant" plug-in, users can invoke an inspection by PuriFile directly from the Tools menu of each program.

Once complete, the results of the inspection are shown to the user in a pop-up window, with each line highlighting a different security discovery made in the file. Clicking on each discovery will cause the program to show more detail about the issue and, in most instances, will also result in the program driving directly to the affected area of the document and selecting the offending object or phrase.

For example, if it was discovered that an image had been heavily cropped, rather than just annotate the issue and give directions to where the image could be found, the Assistant would actually drive the document directly to the image and even select it for the user. This makes the review and edit process straightforward and reliable. The Assistant will even offer context sensitive buttons to help guide the user towards how to correct the problem (in this case, buttons to un-crop or delete the image).

The Assistant also offers the ability to perform some sweeping actions to cleanse the document of common security problems. These can be called into action independent of inspection, and provide a simple and repeatable means for fixing hidden data issues. If configured to do so, the tool will perform such actions as re-balancing image brightness and contrast, clearing off metadata, removing macros, and even removing embedded documents while leaving the appearance of the document intact; an action sometimes called "flattening."

When installed in the suggested client server mode, PuriFile easily connects to the existing Active Directory environment at the site to quickly grant users the ability to begin using the tool. It is even possible to stand up multiple servers in a load-balancing configuration to allow broader use within an enterprise.

For smaller operations, PuriFile can also be purchased and installed for use on a single machine.

Configurable and Powerful
There are score upon score of possible security issues which PuriFile can test for in each file format. Not all sites, however, share the same set of issues they are concerned about, and there are occasions where multiple policies may even be needed at any particular site. PuriFile is able to handle this easily.
PuriFile is completely configurable in this regard.

Administrators are able to define as many policies as they need. They can also establish rules for which users have access to certain inspection policies.

Any particular issue can be ignored altogether, or set to be identified as a note, concern or violation. Moreover, PuriFile supports a robust "dirty word" inspection capability which can be used to flag occurrences of offending text. Rather than relying on fixed phrases, the product takes full advantage of "regular expressions." These allow policy administrators to have PuriFile identify phrases that follow a fixed shape. For example, it would be possible to set a violation on any use of something that looked like a credit card number, a phone number from certain area codes, or latitude/longitude coordinates.

PuriFile actually performs the dirty word search multiple times and in different ways for Word, PowerPoint, Excel and PDF files to help ensure that every possible location is checked. In fact, every file passed through PuriFile is given a dirty word scan, even if it is not one of the formats that allow for deep inspection.

Industry Recognition
Winn Schwartau, a recognized expert in the security industry, selected PuriFile as a "Category Breaker" for Network World Magazine, saying "Microsoft doesn't disclose the vulnerabilities in a way that makes sense to non-techies. I became an instant fan of PuriFile."

Similarly, SC Magazine's Peter Stephenson gave the product a four-star rating when testing the product for their Data Leakage Prevention issue. He noted that PuriFile "is a first rate product for verifying both accidental and intentional data leakage."

The product's strong reputation was secured further by the recent purchase of an enterprise license by the Department of Defense. PuriFile is also integrated with the ISSE cross domain solution (also from ITT), as well as several others.

ITT is a leader in the development of information assurance technologies that enable secure networks for military, intelligence and law enforcement customers. The company specializes in cyber security, information assurance and other related computer intelligence services. ITT offers a wide range of custom-developed software to assist clients in defining their security needs and policies including cross-domain information sharing solutions, information systems security engineering, and cyber security.
To download ITT's PuriFile software, visit www.PuriFile.com/download.