THE FILE SCRUB TEAM

FILESCRUB

REVIEW, CLEANSING, AND TRUSTED TRANSFER TOOLS FOR FILES

Identifies and removes sensitive information from electronic documents and other binary files

Reduces file review and redaction time from days to minutes

Supports many electronic file formats and operating system platforms

Los Alamos
NATIONAL LABORATORY
EST.1943

File Scrub

Review, Cleansing, and Trusted Transfer Tools for Files

The File Scrub Team

ABOUT THE COVER

This cover is a representational view of what File Scrub tools do. The File Scrub and File Scrub Trusted Copy tools both "wash" several types of files, cleansing them of various kinds of hidden data. The File Scrub Trusted Copy tool provides the user with added functionality to transfer a TRUSTED copy of the cleansed files to external media for further use.

LA-UR-06-1253

Executive Summary

File Scrub

Review, Cleansing, and Trusted Transfer Tools for Files

Features

File Scrub and File Scrub Trusted Copy are security software applications designed to review files for the identification and elimination of sensitive information. Both File Scrub applications detect and remove hidden data in a review process designed to prevent the inadvertent or intentional release of sensitive materials, resulting in a cleansed file. File Scrub Trusted Copy (formerly known as Multi-Platform Trusted Copy, or MPTC) has additional features that provide the user with a consistent and regimented workflow for the transfer of the cleansed file to a removable medium for distribution outside classified and closed work environments.

Applications

- Review of government and corporate files for hidden, classified, and sensitive data before releasing the documents to other agencies, the public, and the news media.
- Protection against inadvertent release of personal private information such as medical, legal, and financial information.

Benefits

- Provides a thorough review of electronic documents, spreadsheets, Adobe Acrobat (PDF) files, and other file types. This review helps mitigate the threat of accidental or intentional release of hidden and sensitive information.
- Provides an audit trail by preserving the history of review actions, findings, and (in File Scrub Trusted Copy) transfers in encrypted logs.
- Minimizes the risk of inadvertent and deliberate release of confidential, classified, and sensitive information.
- Cleans files to ensure compliance with organizational protocols.
- Reviews many file types, including text files, PDF files, computer source code, some CAD files, and Microsoft Office files.
- Runs on a comprehensive set of operating system platforms (Solaris, Linux, Microsoft Windows, and Apple Macintosh OS X). The same intuitive interface is found on all platforms, simplifying training requirements and maximizing the use of existing resources.

Overview

The security of information of all types—organizational, personal, and sensitive or classified—is a major and growing concern of not just digital security professionals, but also of users at all levels.

All electronic files contain some information invisible to the reader and author. In most cases this information, such as author name, dates, organization name, comments, and editing history, is created by software applications as part of the file management process. Users send and receive files electronically, not knowing that hidden information is being transmitted. On the other hand, users can also intentionally hide information in files and transmit the data to someone wanting to cause personal or financial harm or even to compromise national security. Though many private sector organizations already have information security policies and system controls in place, File Scrub offers protection at a file level, thus filling a gap that many computer users do not know exists.

File Scrub is file review software that identifies, logs, and removes many kinds of hidden information in a file. File Scrub is written in Java for ease of cross-platform development and for support of object-oriented design. File Scrub abstracts out commonalities between file formats as parent or super classes, so that new file formats can leverage work already performed and easily be added to the review framework. Where appropriate, File Scrub utilizes third-party Java libraries like JIntegra and PDFBox to provide functionality necessary for querying and modifying documents via their generating application or by directly manipulating the document format. File Scrub runs on Windows, Linux, Solaris, and Macintosh OS X platforms. It reviews many popular file formats including Microsoft Office applications, Adobe PDF files, WordPerfect files, text files, binary files, and some CAD DXF files—virtually any binary file containing textual and graphic content.

During review, File Scrub examines every file for metadata, macros, embedded objects, executables (intelligent agent/malware detection), hidden data streams, data hidden in JPEG images using steganography, and links to objects or other files. Any items detected are logged and, if possible, removed from the resulting cleansed file. Additionally, in Microsoft Office documents, any tracked changes are accepted and comments are deleted so that potentially embarrassing, if not sensitive, text is removed. (For a more detailed description of these elements and how they can contain damaging information, see “Elements That Can Hide Information” in the Appendix.)

File Scrub also searches for keywords, phrases, or regular expressions provided by the user in a keyword file and/or the user interface. This feature allows customization while increasing

efficiency—a core of keywords may be used with every file, with additional words or phrases manually entered for a specific document or type of document. File Scrub does more than the typical ASCII word search; it also searches in Unicode, a 16-bit character set that includes non-English characters.

Embedded objects such as graphics and pictures can also hide or obscure information: objects can be hidden in layers or images can be cropped, but when expanded, they reveal hidden data. File Scrub uses either the generating application or directly examines the file under review to discover such objects and report their number and type.

When objects are detected during a review that cannot be automatically removed, File Scrub will fail the file and log the results. The user can then use File Scrub's External Tool Launch feature to open the failed file in its original application, find each keyword, embedded object, image, or other issue, and review its usage. If the usage in context does not reveal any sensitive information, the user can override the review failure.

To maintain the integrity of the review process, File Scrub records all findings and user actions in encrypted log files. The logging capability is a critical feature of File Scrub. Log file encryption prevents tampering, and the log files can be opened and read only from within File Scrub. The review log is the user's primary tool: it records all the review findings, plus time of review, username, file name, and review status.

As an additional safeguard against inadvertent release of information, File Scrub's sister application, File Scrub Trusted Copy, allows the user to transfer the cleansed file to another medium—floppy, Jaz or Zip drive, memory stick, and CD ROM. File Scrub Trusted Copy's transfer log records the time of transfer, username, transfer destination, and transfer status. It also provides a comments area where reviewers can record how they resolved any issues found in the review. File Scrub Trusted Copy ensures that only the file is transferred; any space remaining between the end of the file and the end of the sector block transferred is padded with zeros.

Competition

The number of competitors has increased from three to six since 2002, indicating that the private sector sees a need for such a product as File Scrub. The following matrix compares File Scrub with the most mature of those products.

Metadata Assistant—Payne Consulting Group

Workshare—Workshare Software

ezClean—KKL Software

Comparison matrix

Parameters	File Scrub, File Scrub Trusted Copy	Workshare	Metadata Assistant	ezClean
Web Address	filescrub.lanl.gov	www.workshare.com/products/	www.payneconsulting.com/products/	www.kklsoftware.com
Platforms	Windows 2000; Windows XP; Red Hat Enterprise Linux Workstation 3 and 4; Fedora Core 2, 3, and 4; Solaris 8 and 9; Mac OS X 10.4 (Tiger)	Windows	Windows	Windows XP (only)
File Types	Word, Excel, PowerPoint, Access, WordPerfect, PDF, any ASCII text file, most any binary file, JPEG files, executable files	Word, WordPerfect. Future update to include Excel, PowerPoint, WordPerfect, PDF.	Word, Excel, PowerPoint	Word, Excel, PowerPoint
Removes	metadata, macros, embedded objects, intelligent agents, hidden data streams, links to objects or other files, and keywords, including those embedded in ASCII or Unicode, steganography (in JPEG files), compressed or encrypted files, MS Word fields, and cropped pictures	metadata, macros, hidden data streams, links to objects or other files, MS Word fields, white text, small text, “smart tags”	metadata, macros, hidden data streams, links to objects or other files, MS Word fields, white text, small text	metadata
Searches	Searches for user-supplied keywords and phrases, in lists or manually entered	NONE	NONE	NONE
Results	Produces a cleansed version of the file. Records all file review and transfer activities in encrypted logs.	Creates Document Audit Report and Risk Report.	Analysis results available in RTF and XML format	Produces a report
Benefits	Reviews up to 250 files at a time. Identical interface across all platforms, including a window with tabs for selecting, reviewing, and transferring files. Opens file in its original application for closer examination of the keywords and phrases, pictures, and other nontext objects identified in the review.	Centralized installation. Remote access support.		
Cost	Currently available free for government use. Commercialization partner being sought.	Workshare DeltaView ($1795 for 5-pack licenses with 1st year upgrades), Workshare DeltaView PE (Single seat: $119, yearly renewal, $239 for lifetime, upgrades cost more), Workshare Protect (Single seat: $29.95, yearly renewal, $59.95 for lifetime, upgrades cost more), and Trace.	Retail Version - 1 standalone license with no upgrades or patches for $79.00. Enterprise Version for larger purchases.	Minimum 25 seats. Pricing $20 per seat (25–100), $16 per seat (101–500), $12 (501+).

Advantages

File Scrub reviews multiple platforms and file types: File Scrub goes well beyond the competing products, which focus primarily on cleansing metadata (hidden data about the file and file content) from files created by Microsoft Office. File Scrub also reviews files from other applications, such as Mathematica, text editors such as Notepad or emacs, and Adobe Acrobat. It can also review any other plain text file such as XML, HTML, and some CAD files. There is simply no other product that reviews non-Microsoft Office files or files created/run on Linux, Solaris, or Macintosh OS X platforms. File Scrub offers a single solution to file review and protection for multiapplication, multiplatform organizations.

File Scrub performs a more comprehensive review: In addition to removing metadata, hidden data streams, macros, and links to other files, as its competitors do, File Scrub reviews files for steganography (data hidden in JPEG files), intelligent agents, and malware. File Scrub inspects file contents to ensure the proper review regardless of the file extension (i.e., an executable file disguised with a .doc extension will be reviewed as an executable). File Scrub also performs a signature-based inspection of certain executable files looking for evidence of malware or intelligent agent code. File Scrub can also detect when an executable has been packed or encrypted, indicating the possible presence of malware. Any of these hidden elements could corrupt or destroy sensitive information on a private network or could send information secretly out of the organization. File Scrub also reports numbers of embedded objects, pictures, and other elements in which data can be hidden. The type of hidden data found can suggest whether its presence was intentional or inadvertent and could help identify the person who hid the data.

Each File Scrub review automatically includes the above-mentioned tests. In addition, the user can perform a keyword search to evaluate the presence or usage of specific sensitive terms. File Scrub allows users to submit a keyword file and/or enter words or phrases manually. File Scrub searches for these words in two character sets: ASCII and Unicode. No other tool provides this functionality. The only other alternative is to use the Find tool within an application—one word or phrase at a time—a time-consuming and impractical task that File Scrub **makes fast and simple.**

File Scrub provides detailed logs: Metadata Assistant provides a summary or a detailed review log in XML or RTF format. Workshare creates a Document Audit Report and a Risk Report. File Scrub, however, provides detailed review and session logs for each review session. The session log records the time and date when File Scrub was opened, username, opening of any external applications used to view the file for the review log findings, and all review and transfer actions. The review log lists results of every test performed.

It also provides guidance to the reviewer, based on the findings. For example, if Track Changes is ON or View Hidden Text is OFF in a Word document, the review fails, and the log provides instructions to change those settings in the original file so the user can re-review the file. See an example of a review log in the Appendix.

File Scrub transfers files and retains their formats. Metadata Assistant and Workshare convert a file to PDF. Conversion to PDF can generate its own metadata and should not be considered a foolproof way to cleanse files. Only File Scrub can review a PDF to ensure that nothing gets out. And, by using File Scrub Trusted Copy, the file can then be transferred to external media. The transfer log records all transfers: file names, time and date, file size, file review status, and file destination. A comments field allows the reviewer to state how review findings were resolved, e.g., whether all graphics were accounted for and visually reviewed, or whether all keywords found were removed or deemed safe in the context they were used. These logs provide evidence that a file was reviewed and document what the findings were, in case there is ever a problem. All logs are encrypted to prevent tampering and must be viewed from within File Scrub.

Primary applications

The current primary application for File Scrub Trusted Copy is in federal government review and declassification of documents for public release. Certain federal government agencies, particularly agencies dealing with national security, must review every document intended for public distribution to ensure no information that could compromise national security is released. Before Multi-Platform Trusted Copy, File Scrub's predecessor, document reviewers usually examined paper copies. Now, using File Scrub, reviewers perform comprehensive tests that go beyond a visual paper inspection to ensure no classified or sensitive information is released.

File Scrub Trusted Copy has over a thousand computer security users in dozens of agencies, including the Department of Energy (DOE), Department of State, National Nuclear Security Administration, branches of the US military, and the National Security Agency.

The creation of File Scrub (without the Trusted Copy function necessary to transfer files from the classified environment) opens an additional large market, that of corporate users. Examples include legal firms seeking to prevent any inadvertent or deliberate release of sensitive information and medical organizations seeking to protect patient privacy. Financial institutions, who already consider information security and privacy a very high priority, could add File Scrub as an additional "smart" layer of protection. Use of File Scrub would reduce the risk of inadvertent release of

damaging information and thus prevent the sometimes devastating consequences of the misuse of sensitive information (e.g., financial damage, personal defamation, expensive lawsuits).

With the release of File Scrub, industry and other private sector users will be able to use the same "government tested" software to prevent the release and transfer of proprietary, personal, and security-related information. This is of particular importance given the recent emphasis on personal privacy, corporate espionage by competitors and foreign nations, as well as the legal requirements of legislation including the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and the Sarbanes-Oxley Act of 2002 (Sarbox or SOX).

Other applications

A commercialization partner, currently being sought by Los Alamos, could readily take the core File Scrub features and create automated cleansing tools for e-mail, digital faxes, and file transmission applications (such as instant messaging).

In addition to adding to the features of File Scrub, a partner could make File Scrub available to individuals. With increasing emphasis on personal privacy and the security of personal information, individuals could add File Scrub to their arsenal of computer security software.

Summary

The use of electronically generated documents and other types of information has brought new opportunities and new dangers. No longer can sensitive information be cut out with scissors or covered by bold black ink, as was done with paper. An inadvertent slip of a few bytes of information could provide the missing data needed by someone wishing to harm a person, company, or government.

There is no other file review product currently available in the government or private sector that provides as much file protection as File Scrub and File Scrub Trusted Copy. The File Scrub software family reviews more software applications for more vulnerabilities than any of its competitors. There is no other tool that will review and cleanse files on Mac OS X, Linux, or Solaris platforms.

Government use of File Scrub has made a difference in preparing files for public access, making reviews much faster than individuals reading paper files, and providing the assurance that documents have been reviewed and contain nothing that would harm national security. This contribution of File Scrub is invisible to the public, but it is an important tool in protecting our country.

When available commercially, File Scrub will offer protection to companies and individuals who want assurances that their sensitive information is secure.