



Cyberspies Exploit Microsoft Office

Byron Acohido writes in *USA Today* that, "cyberspies have a new secret weapon: tainted Microsoft Office files."

The bad guys have been taking advantage of the "zero day" security flaws in Microsoft Office to attack computers in search of information. "Zero day" is an exploit or vulnerability that is announced publicly before a patch is available. These attacks are usually focused at a limited target or audience.

Hidden data is another way for cyberspies to exploit Microsoft Office files to obtain information about a specific target. The only real difference is that there is no discernable signature for hidden data exploitation. Sending tainted files to a target has a very definite and detectable signature. Hidden data exploitations involve a one-way transfer of information from the victim to the person harvesting hidden data, so there is no feedback to the victim indicating their data has been exploited. If cyberspies are taking a proactive approach to targeting information, then we should assume they are also using passive exploitation to harvest the hidden data offered to them freely by potential victims?

Human nature responds quickly to visible threats, such as using tainted Office documents to perpetrate information attacks. Just because a threat is not visible, does not mean that it is not real or eminent. Hidden data in electronic documents is a very real threat and can be used to exploit information just like the tainted files threat mentioned in this article. Be alert and proactive in preventing data leaks caused by hidden data. Read this article for yourself at:

http://www.usatoday.com/tech/news/computersecurity/2007-04-22-cyberspies-microsoft-office_N.htm

The Truth About Redaction

The push to digitize and publish public records is creating new vulnerabilities and hidden data. In an article published on the *News for Public Officials...* website, David Bloys discusses specific cases in New York and New Jersey where hand written social security numbers and other personal information on public records have been scanned to image formats and published electronically on the Internet. This type of information is extremely difficult to detect using computer algorithms, but in many cases there has been no attempt to screen information before it is published. Once published, the information is mined, mirrored,

aggregated, and sold, so the information can never be redacted. Read the full story at:

<http://www.davickservices.com/REDACTION.htm>

SRS Technologies recognized the problem of finding sensitive information in the images embedded in electronic documents, and we accounted for it in our 4-step review process. In step three, the reviewer can step through and view all of the images embedded in a Microsoft Office document. The human reviewer is capable of recognizing sensitive content in images that can not be reliably detected using computer algorithms. Integrating the human ability to process very complex data with a rigorous computer-controlled review process provides the best of both capabilities and ensures the most thorough review possible when sensitive information needs to be protected.

The same website that published this article about redaction maintains a list of information security web breaches that is worth reading. You will find it at the following link:

http://www.davickservices.com/web_breaches.htm

Census Bureau Exposed Personal Data

In recent months, the Census Bureau has posted over 63,000 social security numbers on the Internet according to a recent article in the *Washington Post* by Ellen Nakashima. The problem was discovered when an Illinois farmer found his personal information while surfing the Internet. The information had already been mirrored to "at least a half-dozen sites," before the breach had been reported to the Census Bureau. There is no way to know how much of this data was mined and harvested before it could be cleaned up.

<http://www.washingtonpost.com/wp-dyn/content/article/2007/04/20/AR2007042002208.html>

<http://www.fcw.com/article97859-03-08-07-Web>

A computer aided review of the website before it was published could have prevented this situation, but many organizations still rely on the human to catch this kind of error. Social security numbers are highly structured, so they can be easily detected using a regular expression keyword search like the one done by Document Detective. Document Detective digs deep into the Microsoft Office file formats to extract both visible and hidden text, and then it scans all text with a regular expression keyword scanner. Even information hidden in plain sight is easily detected and flagged for the

human reviewer's consideration. Without tools like Document Detective, such security breaches will continue to be commonplace.

Metadata Among Colleagues

Our last article is a bit dated, but it is still relevant. Last year, the President-elect of the Florida Bar brought the issue of metadata to the Board of Governors attention because of an incident that had occurred at his law firm. A legal brief sent to another firm contained hidden data allowing the other firm to reconstruct every change made to the document, including emails between the client and his attorney. This was clearly a violation of attorney-client privilege. The board's initial response was to make metadata analysis unethical, but it is really no different than fingerprinting a physical document to determine who has handled it. This article talks about using Adobe PDF documents as a partial solution, and even goes so far as to describe a PDF as an image of the document. Neither is true, and PDF documents can also be mined for hidden data. Document Detective is the only tool we know about that actually exposes the hidden data in a PDF document. For more information about the Florida case, please visit:

<http://www.law.com/jsp/legaltechnology/pubArticleLTN.jsp?id=1145538533635>

Document Detective Finds Leaks in Security Presentations

The presentation on Personal Security Clearance Investigations given by the Federal Investigative Services Division at the National Security Institute's 22nd Annual National Security Forum (IMPACT 2007) in Falls Church, Virginia in April 2007 had a message from the President, George W. Bush and the Director of the Office of Personnel Management, Kay Coles James, hidden in the Tracked Changes. A Space and Missile Defense Command presentation on Information Assurance contained an embedded copy of a presentation on the legal aspects of proposed regulation changes hidden in the Tracked Changes. A Defense Intelligence Agency presentation on Spear Phishing contained unrelated information on human engineering, a reference to the IAPC Data spill webpage "for incidents, FAQ, Data Transfer and other DIA Policies in the IA Library" hidden in the Notes Pages. All of these are common examples of hidden data that are becoming epidemic as we share more information in the form of electronic documents.

The DIA presentation on Spear Phishing had actually been cleansed using a product called IC Clear. When we examined this presentation using Document Detective, we found several cropped and resized images, and one dangerous data construction in addition

to the unrelated Notes Page content. This begs the question, what did IC Clear really cleanse?

SRS Technologies also evaluated a product called Purifile, which is an electronic document review and sanitization tool being developed for the Intelligence Community. Purifile did not perform well against the sample documents distributed as Document Detective training aids. Purifile did detect all of the embedded objects in our sample Word document, but the information provided would not help the average user to clean up this document. It took our expert several hours to relate the Purifile findings to the data in our sample document. Most of the embedded objects were identified by comparing height and width properties to the Purifile report. To get that information, our expert already had to know where all of the objects were located. Purifile reported content on non-existent paragraphs, and keywords "near" some unknown offset. Purifile also failed to locate some of the hidden data content in the sample documents, such as improperly redacted paragraphs, keywords in Custom Document Properties, and text and images positioned off the viewable page. The Purifile Assistant that was supposed to "flatten" this document failed to remove cropped areas from images and failed to flatten four of the five embedded OLE objects in the document.

SRS Technologies is committed to continuing research and product development in the area of electronic document security. While no process or procedure is perfect, Document Detective is the best available solution for reviewing and sanitizing electronic documents, especially when National Security Information is at stake.

Please feel free to forward this message to anyone you believe would benefit from this information.

This newsletter is dedicated to raising awareness about the Desktop Publishing Threat to Information Security and to disseminating information that will help mitigate these risks. Our newsletter is published aperiodically when there is news to report. We will not clutter your inbox with idle chatter when there is nothing significant or useful to say. This newsletter is available free of charge. Subscriptions are at the discretion of SRS Technologies.

Suggestions and comments are always welcome. If you do not wish to receive this newsletter, please e-mail us and we will gladly remove you from the list. All EDS-related correspondence should be sent to rhackett@srs.com

For information about SRS Technologies' Document Detective electronic document review and sanitization toolkit, please visit: <http://www.docdet.com/>

For a list of articles and information regarding EDS, including previous editions of this newsletter, please visit the Document Detective website and click on "The Threat."

To subscribe or unsubscribe from this newsletter, send email to: rhackett@srs.com

SRS Technologies
Systems Solutions Division
500 Discovery Drive
Huntsville, AL 35806
Copyright 2007 SRS Technologies