Tactical Fingerprinting using metadata, hidden info and flown lost data using FOCA
Microsoft Word bytes Tony Blair in the butt

Home > Privacy > Blair’s Iraq Dossier

Richard M. Smith (rms@computerbytesman.com)
June 30, 2003

Microsoft Word documents are notorious for containing private information in file headers which people would sometimes rather not share. The hard way.

Back in February 2003, 10 Downing Street published a dossier on Iraq’s security and intelligence organizations. This dossier was cited by Colin Dr. Glen Rangwala, a lecturer in politics at Cambridge University, quickly discovered that much of the material in the dossier was actually plagiarized.

You can read Dr. Rangwala’s original analysis of the dossier from Feb. 5, 2003 at this URL:

http://www.casi.org.uk/discuss/2003/msp00457.html

Blair’s government made one additional mistake: they published the dossier as a Microsoft Word file on their Web site. When I first heard from I had worked on the document. I downloaded the Word file containing the dossier from the 10 Downing Street Web site (http://www.number-10.

Rev. #1: "doc22" edited file "C:\DOCUME~1\phamill\LOCALS~1\Temp\AutoRecovery save of Iraq - security.doc"
Rev. #2: "doc22" edited file "C:\DOCUME~1\phamill\LOCALS~1\Temp\AutoRecovery save of Iraq - security.doc"
Rev. #3: "doc22" edited file "C:\DOCUME~1\phamill\LOCALS~1\Temp\AutoRecovery save of Iraq - security.doc"
Rev. #4: "JPratt" edited file "C:\TEMP\Iraq - security.doc"
Rev. #5: "JPratt" edited file "A:\Iraq - security.doc"
Rev. #6: "ablackshaw" edited file "C:\Ablackshaw\Iraq - security.doc"
Rev. #7: "ablackshaw" edited file "C:\Ablackshaw\Iraq - security.doc"
Rev. #8: "ablackshaw" edited file "A:\Iraq - security.doc"
Rev. #9: "MKhan" edited file "C:\TEMP\Iraq - security.doc"
Rev. #10: "MKhan" edited file "C:\WINNT\Profiles\mkhan\Desktop\Iraq.doc"
What kind of data can be found?

- **Metadata:**
  - Information stored to give information about the document.
    - For example: Creator, Organization, etc..

- **Hidden information:**
  - Information internally stored by programs and not editable.
    - For example: Template paths, Printers, db structure, etc...

- **Lost data:**
  - Information which is in documents due to human mistakes or negligence, because it was not intended to be there.
    - For example: Links to internal servers, data hidden by format, etc...
Wrong management
Bad format conversion
Unsecure options

New apps
or program
versions

Search engines
Spiders
Databases

Embedded files

Wrong management
Bad format conversion
Unsecure options

Embedded files

Lost Data

Hidden info
The answer is NOT.
Almost nobody is cleaning documents.
Companies publish thousand of documents without cleaning them before:
  - Metadata.
  - Hidden Info.
  - Lost data.
Sample: FBI.gov

Total: 4841 files
Demo: Looking for info in public FBI files
Are they clean?

<table>
<thead>
<tr>
<th>Real Name</th>
<th>Username</th>
<th>Internal Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>.. And more...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Image of software interface showing file scan results]
Sample: ec.europa.eu

<table>
<thead>
<tr>
<th>Search Query</th>
<th>Results</th>
<th>Files Found</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>site:ec.europa.eu filetype: xls</td>
<td>10</td>
<td>399</td>
<td>0.28 seconds</td>
</tr>
<tr>
<td>site:ec.europa.eu filetype: doc</td>
<td>10</td>
<td>103</td>
<td>0.13 seconds</td>
</tr>
<tr>
<td>site:ec.europa.eu filetype: ppt</td>
<td>10</td>
<td>191</td>
<td>0.12 seconds</td>
</tr>
<tr>
<td>site:ec.europa.eu filetype: pdf</td>
<td>10</td>
<td>152</td>
<td>0.12 seconds</td>
</tr>
<tr>
<td>site:ec.europa.eu filetype: ods</td>
<td>10</td>
<td>11</td>
<td>0.16 seconds</td>
</tr>
<tr>
<td>site:ec.europa.eu extension:eps</td>
<td>1</td>
<td>1</td>
<td>0.37 seconds</td>
</tr>
</tbody>
</table>

Total: 896 files
Demo: Looking for info in public ec.europa.eu files
The Midcourse Phase

The Midcourse Phase - During midcourse the warhead separates from the missile body and proceeds through space.

Quick Links
- 2009 MDA Math Challenge Week
- Proposed European Assets
- 2009 BMDS Booklet (pdf)
- Budget Information
- Technology Transfer
- Environmental Information

Latest News

NAVY COMPLETES AIR AND BALLISTIC MISSILE EXERCISE
Commander, U.S. 3rd Fleet, Vice Adm. Samuel J. Locklear announced the completion of the fleet operational exercise, Stellar Daggers to test the Aegis system’s ability to detect, track, engage and destroy multiple incoming air and ballistic missile threats during terminal or final phase of flight.

27 MAR 09
Total: 1075 files
Are they clean?

User
Software Version
Internal Server NetBIOS name
Remote Printer Name
Local Printer
How many files is my company publishing?
Not only Microsoft Office documents.
Sample: Printer info found in odf files returned by Google

$ grep Impresora resultados.txt | grep \\\n10.odt Impresora \svrprint\VISA28 329LC3
12.odt Impresora \servidor\HP 2000C
19.odt Impresora \10.177.1.126\EPSON EPL-6200 Advanced
23.odt Impresora \1-215-01\hp LaserJet 1000
34.odt Impresora \srv2.rm.oasi-servizi.it\RICOH roma corriodio nord
3.odt Impresora \filserver\4100
63.odt Impresora \gamma\HP Deskjet F300 series
27.ods Impresora \Silegio\HP psc 2350 series
33.ods Impresora \ABB\BigGrayUglyThing
40.ods Impresora \\MUSTAVA\Samsung ML-1710 Series
46.ods Impresora \http://10.0.0.42:631\Chaucer
80.ods Impresora \192.168.250.1\HP OfficeJet G55
86.ods Impresora \\WLGPRINT\L10 GA T632
89.ods Impresora \\B-THOMAS\hp LaserJet 1300 PCL 6
17.odp Impresora \tatra\Minolta Di152 PCL6
22.odp Impresora \\10.177.1.126\EPSON EPL-6200 Advanced
4.odp Impresora \\Verra\LEMURIA
55.odp Impresora \\CHEGUEVARA\lp_serv
57.odp Impresora \\sm-ubor01-01.east.sun.com\bur116
69.odp Impresora \\ipp://iprint.innerweb.novell.com\prv-mktg-hp8000dn-2
86.odp Impresora \\EFOITDP02OS4\hp4100dtn
$
$ grep iMac resultados.txt
49.ods Impresora hp psc 2500 series @ Luisa’s iMac
$
Sample: Info found in a PDF file
What files store Metadata, hidden info or lost data?

- Office documents:
  - Open Office documents.
  - MS Office documents.
  - PDF Documents.
    - XMP.
  - EPS Documents.
  - Graphic documents.
    - EXIFF.
    - XMP.
  - And almost everything....
Pictures with GPS info..

EXIFREADER
http://www.takenet.or.jp/~ryuuji/
Demo: Looking for EXIF information in ODF file
Even Videos with users...

And of course, printed txt
Lost data everywhere

<filename>init.sh</filename>
<interpreter>shell</interpreter>
<source><![CDATA[
#!/bin/sh

echo 'Registering to ZLM'

rug shutdown -f

rm /var/lib/zmd/services

/etc/init.d/novell-zmd start

rug set require-verified-certs false

rug sa https://IP-ZLM server

echo 'Finished Registration'
]]><![CDATA[
</script>
</init-scripts>

Note: For Non-Code 10 machines, replace line 9 (rm /var/lib/zmd/

“rm /var/opt/novell/zenworks/lib/zmd/services”

5> Start OS installation using autoyast

Copy the edited control file to NFS server (preferably)

When the system boots from the CD, the Grub installation screen

---

Página 4 / 6 Predeterminado Inglés (Estados Unidos) INSERT STD
Public server

Not Found

The requested URL /administrator/ was not found on this server.

Apache/2.0.52 (Win32) DAV/2 Server at 164.99.164.7 Port 80
What can be found? (I)

- Users:
  - Creators.
  - Modifiers.
  - Users in paths.
    - \C:Documents and settings\jfoo\myfile
    - \home/johnnyf

- History of use.
- Operating systems.
- Software versions.
- Paths.
  - Local and remote.
- Network info.
  - Shared Printers.
  - Shared Folders.
  - ACLS.
What can be found? (II)

- Printers.
  - Local and remote.
- Internal Servers.
  - NetBIOS Name.
  - Domain Name.
  - IP Address.
- Database structures.
  - Table names.
  - Column names.
- Devices info.
  - Mobiles.
  - Photo cameras.
- Private Info.
  - Personal data.
How can metadata be extracted?

- Info is in the file in raw format:
  - Binary.
  - ASCII.
- Therefore Hex or ASCII editors can be used:
  - HexEdit.
  - Notepad++.
  - Bintext
- Special tools can be used:
  - Exif redaer
  - ExifTool
  - Libextractor.
  - Metagoofil.
  - ...
- ...or just open the file!
Tools: Libextractor

D:\> extract -l libextractor_ole2 C-24-45.doc
mimetype: application/msword
revision history - Revision #1: Author 'mark' worked on 'C:\Documents and Settings\mark\Mis documentos\Top Secret.doc'
revision history - Revision #0: Author 'mark' worked on 'C:\Documents and Settings\mark\Mis documentos\Top Secret.doc'
revision history - Revision #0: Author 'mark' worked on 'C:\Documents and Settings\mark\Mis documentos\Top Secret.doc'
revision history - Revision #0: Author 'mark' worked on 'C:\Documents and Settings\mark\Mis documentos\Top Secret.doc'
company: Corporation
paragraph count - 1
line count - 1
last saved by - charlie
character count - 10
template: Normal
creation date - 2008-11-24T18:55:0
title: Top Secret
word count - 1
page count - 1
creator: mark
date - 2008-11-24T18:55:0
generator: Microsoft Word 8.0
D:\>
Tools: MetaGoofil

http://www.edge-security.com/metagoofil.php
Yes, also Google....
Your FBI user
Your Scotland Yard user

Google

site:police.uk intitle:"c:\documents and" filetype:pdf

Buscar en: ◯ la Web ◯ páginas en español ◯ páginas de España ◯ Per...
Your WhiteHouse user
Yes, we can!
Of course, your Novell user
Drawbacks

- These tools only extract metadata.
- Not looking for Hidden Info.
- Not looking for lost data.
- Not post-analysis.
Fingerprinting Organizations with Collected Archives.

- Search for documents
- Automatic file downloading
- Capable of extracting Metadata, hidden info and lost data.
- Cluster information
- Analyzes the info to fingerprint the network.
Demo: FOCA
FOCA: Gathering Files
Esta aplicación no almacena los ficheros subidos ni el contenido de los mismos. La única información almacenada es con fines estadísticos.


Examinar   Analizar Fichero!

Ver estadísticas de la foca.
Solutions?
First: Clean all public documents
Clean your documents:
MSOffice 2k7
Clean your documents: MSOffice 2k3 & XP


Office 2003/XP Add-in: Remove Hidden Data

Brief Description
With this add-in you can permanently remove hidden data and collaboration data, such as change tracking, from PowerPoint files.

On This Page
- Quick Details
- System Requirements
- Related Resources
- Overview
- Instructions
- What Others Are Downloading

Download

Quick Details
- File Name: rhdttool.exe
- Version: 1.1
- Date Published: 7/8/2008
- Language: English
- Download Size: 250 KB
- Estimated Download Time: Dial-up (56K) 1 min

Change Language: English
Demo: OpenOffice “Security” Options...
Cleaning documents

- OOMetaExtractor

http://www.codeplex.org/oometaextractor
Are you safe relying on your users?
IIS MetaShield Protector

http://www.metashieldprotector.com
MetaShield Protector
Demo: Cleaning Documents
Second: Beg Google to delete all the cached files
Don’t trust your users!!!
Don’t complain about your job!!
Questions?

- Authors
  - Chema Alonso
    - chema@informatica64.com
  - Enrique Rando
    - Enrique.rando@juntadeandalucia.es
  - Alejandro Martín
    - amartin@informatica64.com
  - Francisco Oca
    - froca@informatica64.com
  - Antonio Guzmán
    - antonio.guzman@urjc.es
### File Information

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>URL</strong></td>
<td>C:\Universidad\BH 209\Fingerprinting_networks.pptx</td>
</tr>
<tr>
<td><strong>Local path</strong></td>
<td>C:\Universidad\BH 209\Fingerprinting_networks.pptx</td>
</tr>
<tr>
<td><strong>Download</strong></td>
<td>True</td>
</tr>
<tr>
<td><strong>Analyzed</strong></td>
<td>True</td>
</tr>
<tr>
<td><strong>Download date</strong></td>
<td>16/04/2009 16:10:56</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td>5.65 MB</td>
</tr>
</tbody>
</table>

### Dates

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Creation date</strong></td>
<td>10/03/2009 13:00:07</td>
</tr>
<tr>
<td><strong>Modified date</strong></td>
<td>16/04/2009 14:08:35</td>
</tr>
</tbody>
</table>

### Other Metadata

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Application</strong></td>
<td>Microsoft Office PowerPoint - 12.0000</td>
</tr>
<tr>
<td><strong>Revisions</strong></td>
<td>1</td>
</tr>
</tbody>
</table>

### Software

| Software | Microsoft Office 2007 |