Information leakage in proprietary documents

Patrick CHAMBET
patrick@chambet.com

Eric DETOISIEN
valgasu@rstack.org
Planning

• General points
• Some examples
  • Acrobat
  • Microsoft Word
  • Misc
• Recommendations
• Conclusion
General points (1/2)

- Proprietary documents use more and more complex formats
  - Elaborate object model
  - Not documented
  - Partial reverse engineering

- Some MS Word headers
  - DC A5 65 00
  - DC A5 68 00
  - 97 A6 68 00
  - EC A5 C1 00
General points (2/2)

• The trend is to include information of diverse kinds, without the user knowing it
  • Personal information

• Marketing information
  • Use time
  • Use habits
  • Relations with other documents, applications, network resources (including the Internet)

• Active content
  • Can modify the appearance of documents depending on the environment in which they are open
  • Problem for active documents signing
Planning

• General points
• Some examples
  • Acrobat
  • Microsoft Word
  • Misc
• Recommendations
• Conclusion
Adobe Acrobat (1/3)

- PDF documents generation
  - Tools
    - Acrobat Distiller
    - PDF Maker
    - Other tools (Fineprint PDF Factory, etc.)
  - Methods
    - Document conversion (through PostScript)
    - Direct scan of the document
Two examples of information disclosure

- Un-crop a document
  - Gives access to zones that should have been deleted
    - Who would have doubted that, except hackers?
  - In Document / Crop pages, click on Reset

- Deletion of opaque forms
  - Reveals intentionally hidden information

DEMONSTRATIONS
After trying to convince you that it cannot be trusted, I would like to mention that it was not possible to prevent the leak in the handling of the “tunabex” file. The idea was to delete the file containing the text and not to save any copies. Unfortunately, it was not possible to completely delete the file and the text was still available. This raises some serious questions about the security of the system and the handling of sensitive information.

There is an interesting situation concerning. On the one hand, the press, television, and politicians are making headlines by selling them to the public. On the other hand, newspapers are not available by any means.

When I was a child living before Congress, it was common for us to be aware of the sanctions of these acts. There was obviously a shared goal. The notion of bringing the computer back to the same social status as breaking into a neighbor’s house, would not make it the neighbor’s house is unlocked. The press must be aware of the misuse of computers in national and international affairs.

Acknowledgment

For their invaluable assistance in this project, I would like to thank R. B. Murphy, T.J. L. and A. T. for their support. I am grateful to them for their assistance.

References

1. R. B. Murphy, T.J. L. and A. T. for their support. I am grateful to them for their assistance.

2. R. B. Murphy, T.J. L. and A. T. for their support. I am grateful to them for their assistance.

3. R. B. Murphy, T.J. L. and A. T. for their support. I am grateful to them for their assistance.

4. R. B. Murphy, T.J. L. and A. T. for their support. I am grateful to them for their assistance.
- **Document properties**

![Document properties](image)

**Test.doc Properties**

- **General**
  - **Description**: Infos confidentielles
  - **Title**: Infos confidentielles
  - **Subject**: Secret
  - **Category**: Notes internes
  - **Keywords**: Business Plan
  - **Template**: Document Générique GIE.doc
  - **Page Count**: 1
  - **Word Count**: 500
  - **Character Count**: 2654
  - **Lines**: 23
  - **Paragraphs**: 5
  - **Scale**: Yes
  - **Links Dirty?**: False
  - **Comments**: Draft

- **Orbita**
  - **Author**: Baobab
  - **Last Saved By**: n/a
  - **Revision Number**: 11
  - **Application Name**: Microsoft Word 18.0
  - **Company Name**: Edalweb
  - **Date of Creation**: 08/07/2002
  - **Date Last Saved**: 15/08/2002
  - **Last Printed**: 19/07/2000
Some directly readable information
- Author name
- Author company
- Creation date and time
- Edit time
- Print date and time
- Etc.

Some deductible information
- If a 100 page document has been edited in 5 minutes, it’s a simple copy-paste!
- Be careful with the « track changes » option: one can access previous versions!
  - Example: the Alcatel case
• Editing the document with an hexa editor
  • Names of the **successive** authors
  • Machine name
  • Complete path of the document on each author’s disk
    C:\Documents and Settings\Student Smith\Confidential\Customer X\Contract.doc

• Complete path of the document template
  \\FILE_SRV_NT\PUBLIC\WORD_TMPL\Generic Banking Contract.dot
  => You can deduce the file server name in the company and its kind of customers

• Print server and printers
  \\SRV_NT_PDC\HPPCL5MS LaserJet 4 Plus
  => You can also deduce the name of the NT PDC
• Names of files included in the document
  • Ex: image files

• GUID (Global Unique Identifier)
  • Look for « _PID_GUID » :
    \{F165CB92-D166-12D5-AB67-0010A41432AF}\)
  • The last 12 numbers are the network adapter MAC address !
  • Included in Office documents but also Visual C++, some ActiveX, etc.
Some analysis tools
• Information embedding
  • The INCLUDETEXT directive can be used to automatically include a whole document form the hard disk into the current document

```plaintext
{ IF { INCLUDETEXT { IF { DATE } = { DATE } }
  "C:\\confidential.txt"
  "C:\\confidential.txt" } /* MERGEFORMAT */
  "" "" /* MERGEFORMAT */
```

DEMONSTRATION
Microsoft Word (6/8)

- Active content
  - Stealth modification of the document content

- Test:
  ```
  { IF { FILENAME /* MERGEFORMAT { DATE } } = "contract.doc" "white" "black" /* MERGEFORMAT }
  ```
Microsoft Word (7/8): Digital signature

- Problem when digitally signing a document
  - Alice and Bob are setting up a contract
  - Both of them sign it
  - Later, Bob opens the contract again, that shows totally different clauses
  - Nevertheless, the digital signature is still valid

- The digital signature of such a document is like a blank check

- The screen appearance and the print appearance can be different!
• Word bugs

• The author can get information about readers of his document (reverse process)
  • Document opening time
  • Document opening place (IP address)
  • Some information about the reader identity and environment (OS and software used, language, etc.) and about his network connection
Other examples

- WordPerfect saves every stage of the edition in the documents
  - You can get any previous state by undoing the last operations one by one!

- MS Outlook et MS Exchange
  - Under certain conditions, sends a *winmail.dat* file containing the full path of the sender’s mail box (.PST file)

- Spywares
  - Ex: MediaPlayer and RealPlayer, Windows XP, ...
  - Cf TCPA/Palladium and Ross Anderson’s FAQ
Planning

- General points
- Some examples
  - Acrobat
  - Microsoft Word
  - Misc
- Recommendations
- Conclusion
Recommendations (1/3)

- Be careful before releasing a document that has been modified many times
  - If possible, regenerate documents before public release (very constraining!)

- Choose an open source word editor, compatible with the market leaders
  - StarOffice
  - OpenOffice

- Use a personal firewall to prevent some applications to open outbound connections
Recommendations (2/3)

• With MS Word
  • Disable “Fast saves”
  • Disable “Track changes”
  • Disable every kind of macros (including signed ones)
  • Set template files (.dot) as read-only

• With Word XP/2002/2003
  • Check “Remove personal information from this file on save”
  • Check “Warn before printing, saving or sending a file that contains tracked changes or comments”
Recommendations (3/3)

• Digital signature
  • Do not sign a potentially dynamic file: DOC, XLS, MDB, ...
  • Prefer less complex formats: RTF, ...
  • Carefully investigate unknown formats before accepting to sign them or to acknowledge their signature

• Improve the user awareness about information leakage

• Include complex and/or proprietary documents management within your corporate security policy
Conclusion

• Proprietary document formats trend is to get more and more complex (in spite of XML lack of success)

• Spywares and marketing data collecting are growing

• The cases are sometimes very serious because they also occur in critical environments (enterprises, administrations)

• Every organization has to evaluate its own risks depending on the confidentiality level of its information
Links (1/2)

• Alcatel case
  • http://www.landfield.com/isn/mail-archive/2001/Apr/0096.html
  • Alcatel Word document http://web.morons.org/external/CPE_statement.doc

• Exchange
  • http://support.microsoft.com/default.aspx?scid=kb;en-us;298917
  • http://support.microsoft.com/default.aspx?scid=kb;en-us;259037
  • http://support.microsoft.com/default.aspx?scid=kb;en-us;138053
• TCPA
  • http://www.cl.cam.ac.uk/~rja14/tcpa-faq.html

• MISC Magazine
  • http://www.miscmag.com