ph-neutral 2004

Information leakage

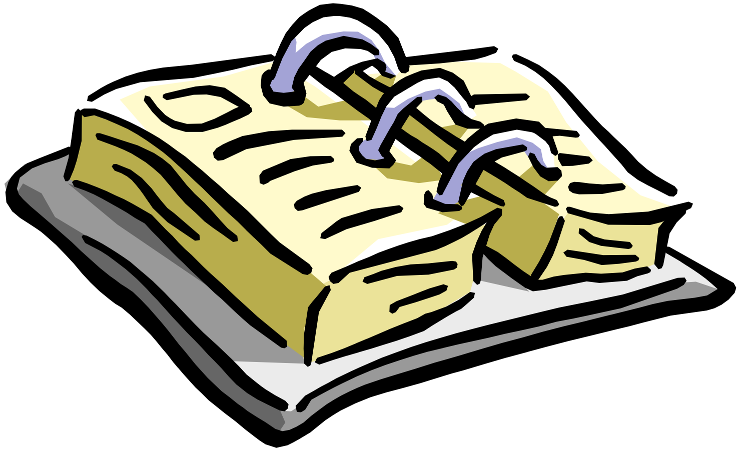
in proprietary documents

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

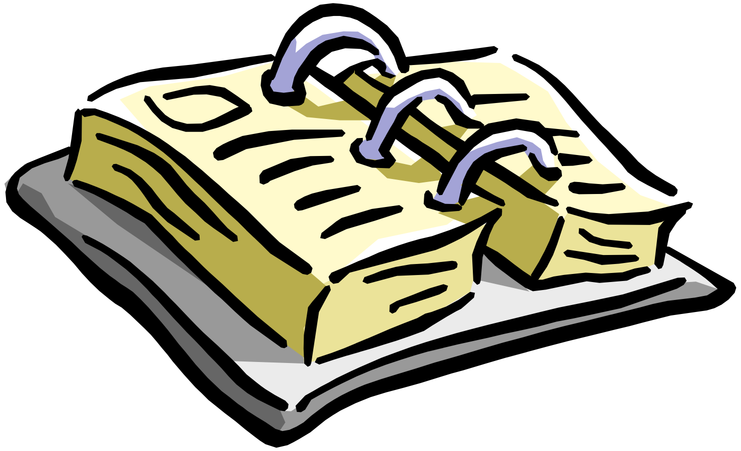
- Some MS Word headers
  - `DC A5 65 00`
  - `DC A5 68 00`
  - `97 A6 68 00`
  - `EC A5 C1 00`

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)

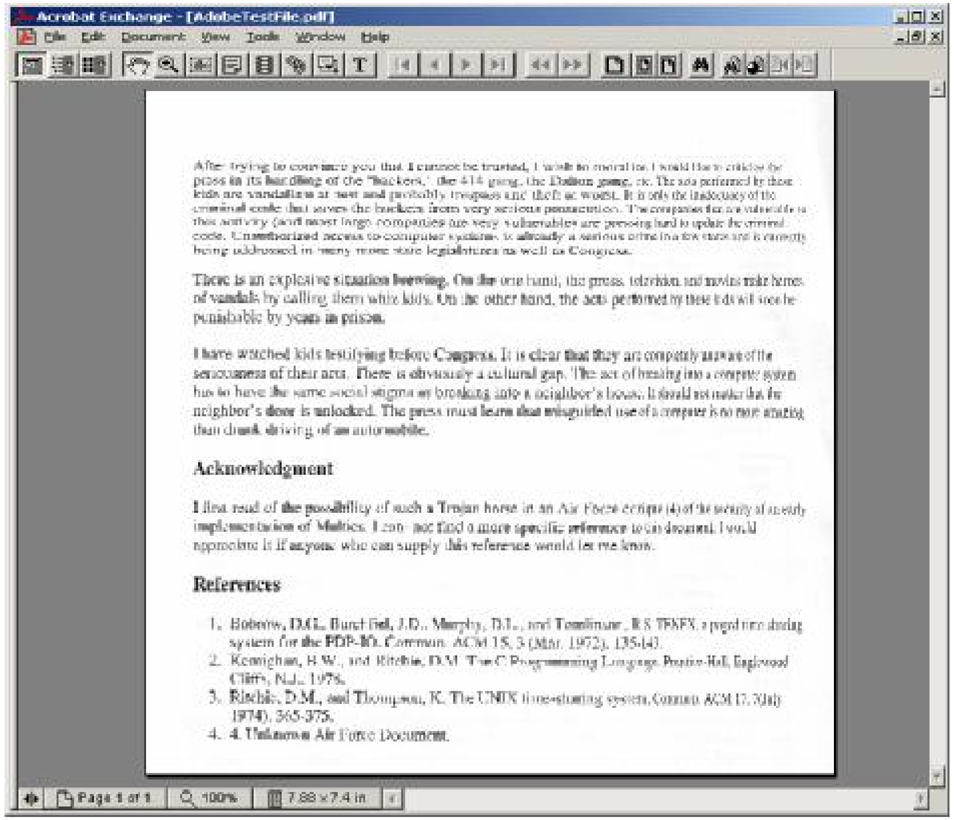
- PDF documents generation
  - Tools
    - Acrobat Distiller
    - PDF Maker
    - Other tools (Fineprint PDF Factory, etc.)
  - Methods
    - Document conversion (through PostScript)
    - Direct scan of the document

Adobe Acrobat (2/3)

- 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 ? J
    - In *Document / Crop pages*, click on *Reset*
  - Deletion of opaque forms
    - Reveals intentionally hidden information

DEMONSTRATIONS

Adobe Acrobat (3/3)


Acrobat Exchange - [AdobeTestFile.pdf]
File Edit Document View Tools Window Help
Page 1 of 1
100%
7.88 x 7.4 in

After trying to convince you that I cannot be trusted, I wish to moralize. I would like to criticize the press in its handling of the "hackers," the 414 gang, the Dalton gang, etc. The acts performed by these kids are vandalism at best and probably trespass and theft at worst. It is only the inadequacy of the criminal code that saves the hackers from very serious prosecution. The companies that are vulnerable to this activity (and most large companies are very vulnerable) are pressing hard to update the criminal code. Unauthorized access to computer systems is already a serious crime in a few states and is currently being addressed in many more state legislatures as well as Congress.

There is an explosive situation brewing. On the one hand, the press, television, and movies make heroes of vandals by calling them whiz kids. On the other hand, the acts performed by these kids will soon be punishable by years in prison.

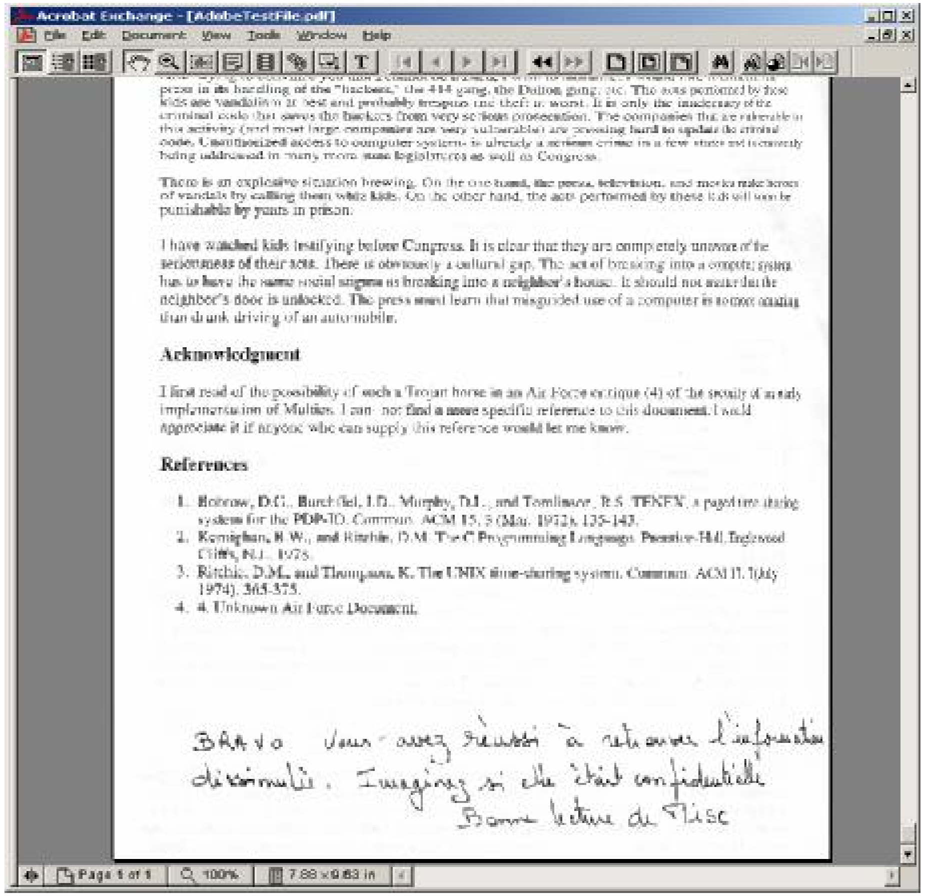
I have watched kids testifying before Congress. It is clear that they are completely unaware of the seriousness of their acts. There is obviously a cultural gap. The act of breaking into a computer system has to have the same social stigma as breaking into a neighbor's house. It should not matter that the neighbor's door is unlocked. The press must learn that misguided use of a computer is no more amazing than drunk driving of an automobile.

Acknowledgment

I first read of the possibility of such a Trojan horse in an Air Force critique (4) of the security of an early implementation of Multics. I can- not find a more specific reference to this document. I would appreciate it if anyone who can supply this reference would let me know.

References

1. Bobrow, D.G., Burchfiel, J.D., Murphy, D.L., and Tomlinson, R.S. TENEX, a paged time sharing system for the PDP-10. Commun. ACM 15, 3 (Mar. 1972), 135-143.
2. Kernighan, B.W., and Ritchie, D.M. The C Programming Language. Prentice-Hall, Englewood Cliffs, N.J., 1978.
3. Ritchie, D.M., and Thompson, K. The UNIX time-sharing system. Commun. ACM 17, 7(July 1974), 365-375.
4. 4. Unknown Air Force Document.


Acrobat Exchange - [AdobeTestFile.pdf]
File Edit Document View Tools Window Help
Page 1 of 1
100%
7.88 x 9.83 in

press in its handling of the "hackers," the 414 gang, the Dalton gang, etc. The acts performed by these kids are vandalism at best and probably trespass and theft at worst. It is only the inadequacy of the criminal code that saves the hackers from very serious prosecution. The companies that are vulnerable to this activity (and most large companies are very vulnerable) are pressing hard to update the criminal code. Unauthorized access to computer systems is already a serious crime in a few states and is currently being addressed in many more state legislatures as well as Congress.

There is an explosive situation brewing. On the one hand, the press, television, and movies make heroes of vandals by calling them whiz kids. On the other hand, the acts performed by these kids will soon be punishable by years in prison.

I have watched kids testifying before Congress. It is clear that they are completely unaware of the seriousness of their acts. There is obviously a cultural gap. The act of breaking into a computer system has to have the same social stigma as breaking into a neighbor's house. It should not matter that the neighbor's door is unlocked. The press must learn that misguided use of a computer is no more amazing than drunk driving of an automobile.

Acknowledgment

I first read of the possibility of such a Trojan horse in an Air Force critique (4) of the security of an early implementation of Multics. I can- not find a more specific reference to this document. I would appreciate it if anyone who can supply this reference would let me know.

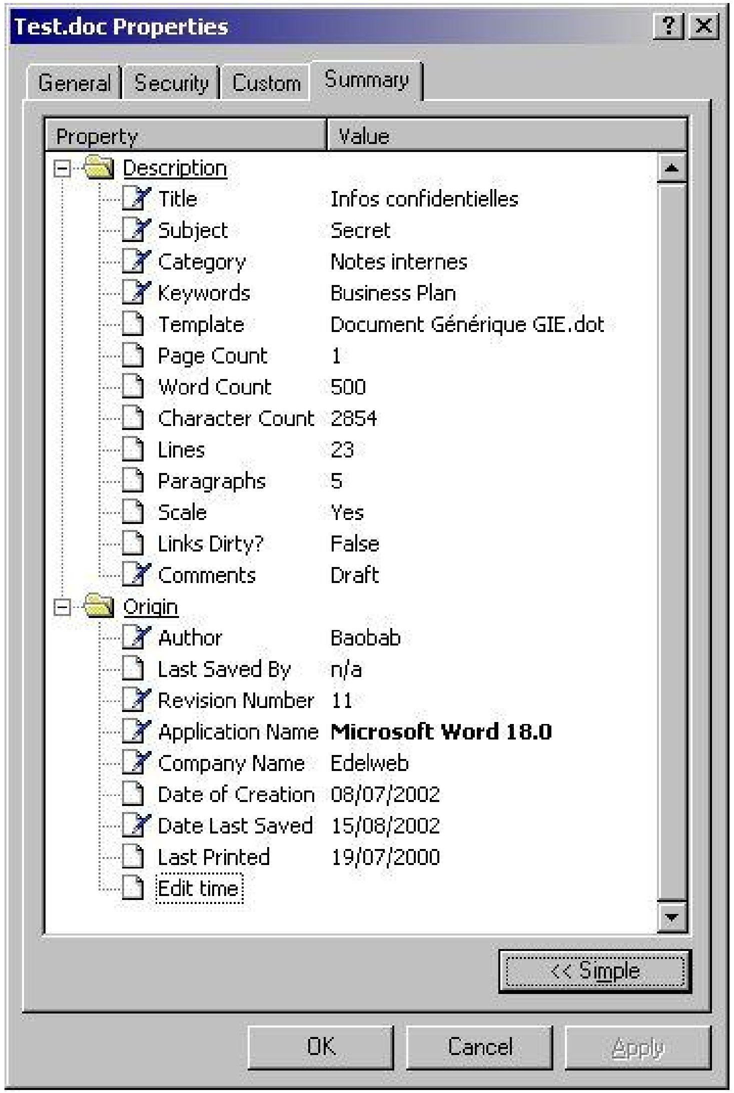
References

1. Bobrow, D.G., Burchfiel, J.D., Murphy, D.L., and Tomlinson, R.S. TENEX, a paged time sharing system for the PDP-10. Commun. ACM 15, 3 (Mar. 1972), 135-143.
2. Kernighan, B.W., and Ritchie, D.M. The C Programming Language. Prentice-Hall, Englewood Cliffs, N.J., 1978.
3. Ritchie, D.M., and Thompson, K. The UNIX time-sharing system. Commun. ACM 17, 7(July 1974), 365-375.
4. 4. Unknown Air Force Document.

BRAVO Vous avez réussi à retrouver l'information dissimulée. Imaginez si elle était confidentielle
Bonne lecture de Misc

Microsoft Word (1/8)

- **Document properties**



Property	Value
Description	
Title	Infos confidentielles
Subject	Secret
Category	Notes internes
Keywords	Business Plan
Template	Document Générique GIE.dot
Page Count	1
Word Count	500
Character Count	2854
Lines	23
Paragraphs	5
Scale	Yes
Links Dirty?	False
Comments	Draft
Origin	
Author	Baobab
Last Saved By	n/a
Revision Number	11
Application Name	**Microsoft Word 18.0**
Company Name	Edelweb
Date of Creation	08/07/2002
Date Last Saved	15/08/2002
Last Printed	19/07/2000
Edit time	

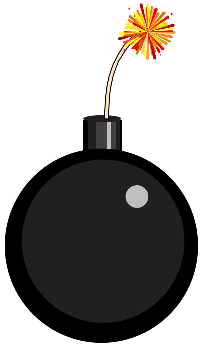
Microsoft Word (2/8)

- 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

DEMONSTRATION

Microsoft Word (3/8)



- 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`

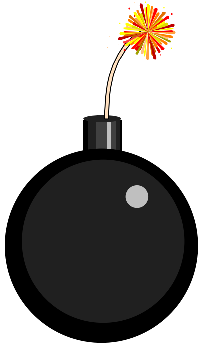
    =>You can deduce the file server name in the company and its kind of customers
  - Print server and printers
    `\\SRV_NT_PDC\HPPCL5MS LaserJet 4 Plus`

    => You can also deduce the name of the NT PDC

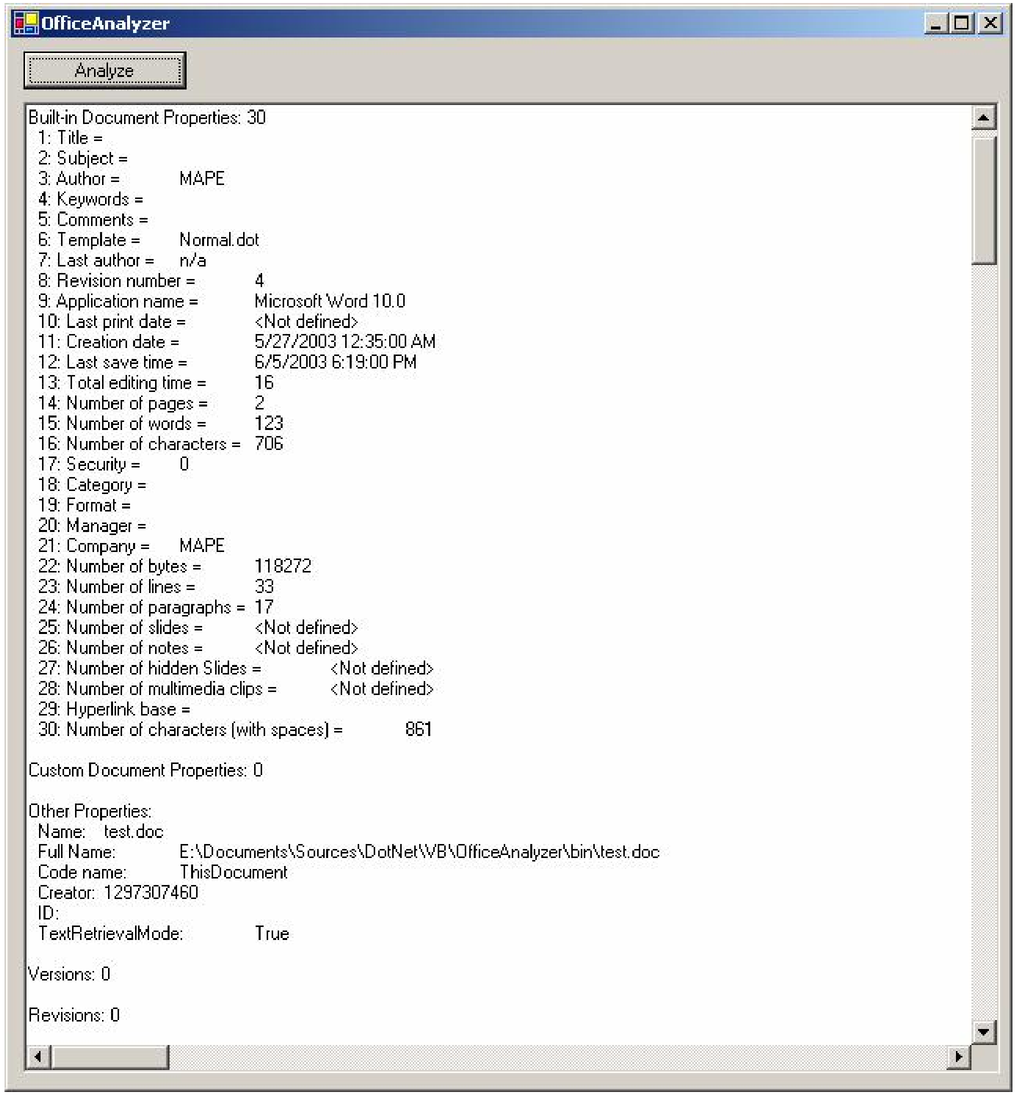
Microsoft Word (4/8)

- 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



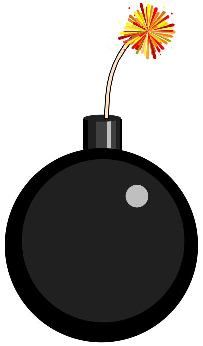
```
Built-in Document Properties: 30
  1: Title =
  2: Subject =
  3: Author =          MAPE
  4: Keywords =
  5: Comments =
  6: Template =        Normal.dot
  7: Last author =     n/a
  8: Revision number =          4
  9: Application name =         Microsoft Word 10.0
  10: Last print date =         <Not defined>
  11: Creation date =           5/27/2003 12:35:00 AM
  12: Last save time =          6/5/2003 6:19:00 PM
  13: Total editing time =      16
  14: Number of pages =         2
  15: Number of words =         123
  16: Number of characters =    706
  17: Security =       0
  18: Category =
  19: Format =
  20: Manager =
  21: Company =        MAPE
  22: Number of bytes =         118272
  23: Number of lines =         33
  24: Number of paragraphs =    17
  25: Number of slides =        <Not defined>
  26: Number of notes =         <Not defined>
  27: Number of hidden Slides =          <Not defined>
  28: Number of multimedia clips =       <Not defined>
  29: Hyperlink base =
  30: Number of characters (with spaces) =        861

Custom Document Properties: 0

Other Properties:
  Name:    test.doc
  Full Name:         E:\Documents\Sources\DotNet\VB\OfficeAnalyzer\bin\test.doc
  Code name:         ThisDocument
  Creator:  1297307460
  ID:
  TextRetrievalMode:            True

Versions: 0

Revisions: 0
```

DEMONSTRATION

Microsoft Word (5/8)

- Information embedding
  - The INCLUDETEXT directive can be used to automatically include a whole document form the hard disk into the current document

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

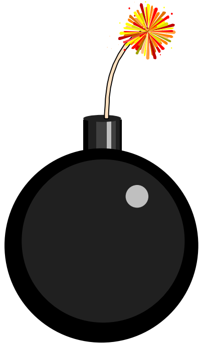


DEMONSTRATION

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

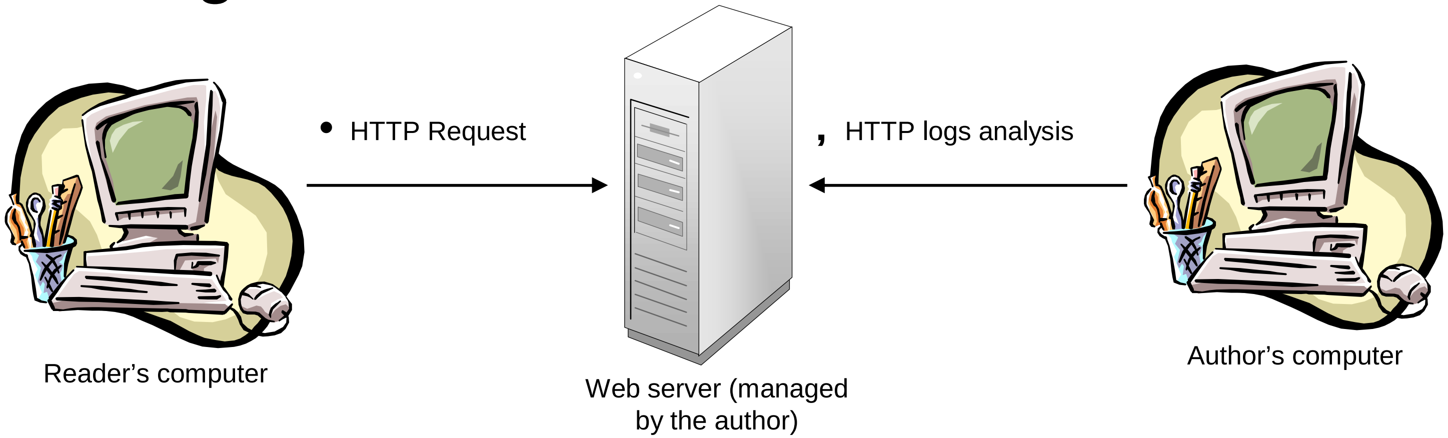
- Problem when digitally signing a document
  - Alice and Bob are setting up a contract
  - Both of them sign it
  - Later, Bob opens the contract again, that shows totally different clauses
  - Nevertheless, the digital signature is still valid

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

- The screen appearance and the print appearance can be different !

Microsoft Word (8/8)

- Word bugs



• HTTP Request

‚ HTTP logs analysis

Reader’s computer

Web server (managed by the author)

Author’s computer

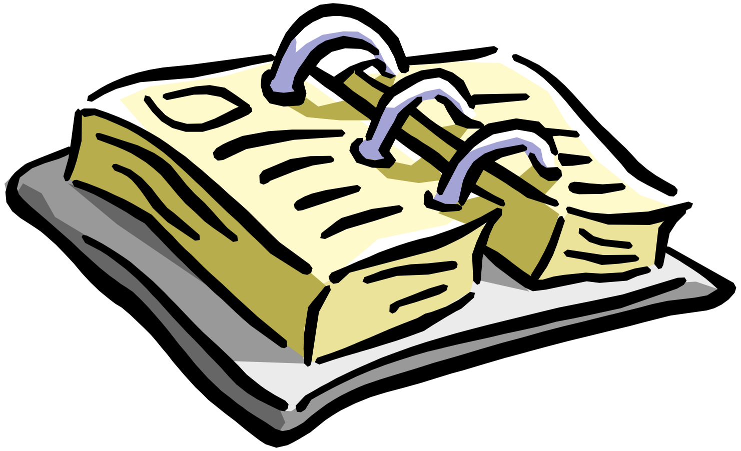
- The author can get information about *readers* of his document (reverse process)
  - Document opening time
  - Document opening place (IP address)
  - Some information about the reader identity and environment (OS and software used, language, etc.) and about his network connection

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)

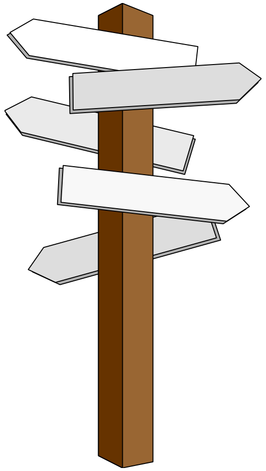
- Digital signature
  - Do not sign a potentially dynamic file: DOC, XLS, MDB, …
  - Prefer less complex formats: RTF, …
  - Carefully investigate unknown formats before accepting to sign them or to acknowledge their signature
- Improve the user awareness about information leakage
- Include complex and/or 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

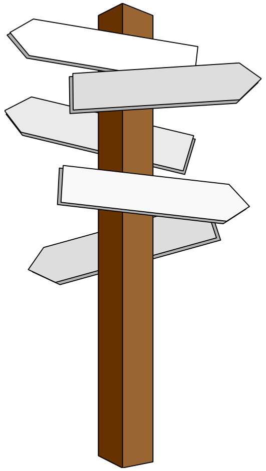
Links (1/2)

- 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

Links (2/2)



- TCPA
  - http://www.cl.cam.ac.uk/~rja14/tcpa-faq.html

- *MISC Magazine*

  - http://www.miscmag.com