Advice

An unknown data leakage risk with MS Office

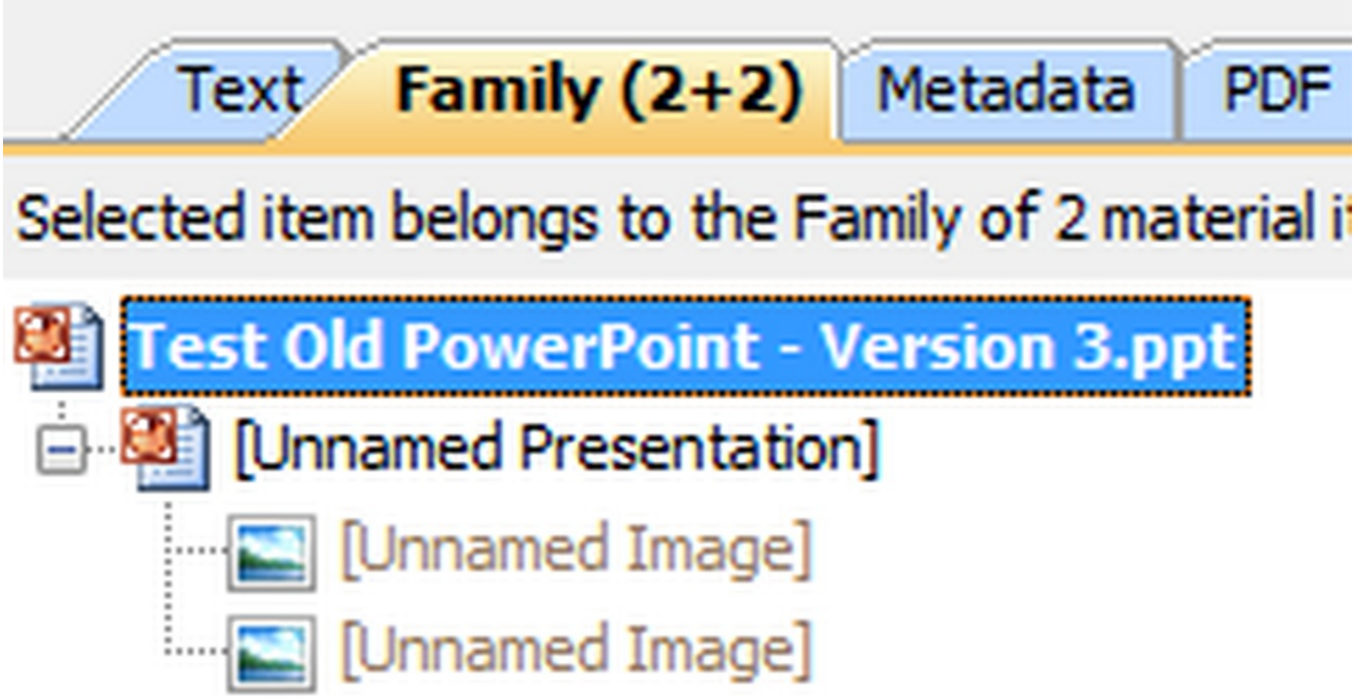
10. June 2015, by Irène Wilson

Introduction

During an investigation earlier this year, we located a document with some peculiar properties: Nuix revealed an embedded PowerPoint presentation within another PowerPoint file. This seemed rather mundane, until we discovered that the embedded data could not be accessed through Microsoft Office using any common method. Swiss FTS were asked to investigate the potential origin of this embedded content. Through this specific case, we discovered an obscure Microsoft Office feature which poses a risk for potential data leakage. We believe that it is our duty to share the results of this investigation with our clients and the wider forensics community.

A child of unknown origin…

During our investigation, Nuix located a unique PowerPoint presentation (the child presentation) within an apparently unrelated PowerPoint presentation (the parent presentation). The contents and subjects of both presentations were completely independent from each other, and the child couldn't be accessed directly through any well-known Microsoft Office feature or method.

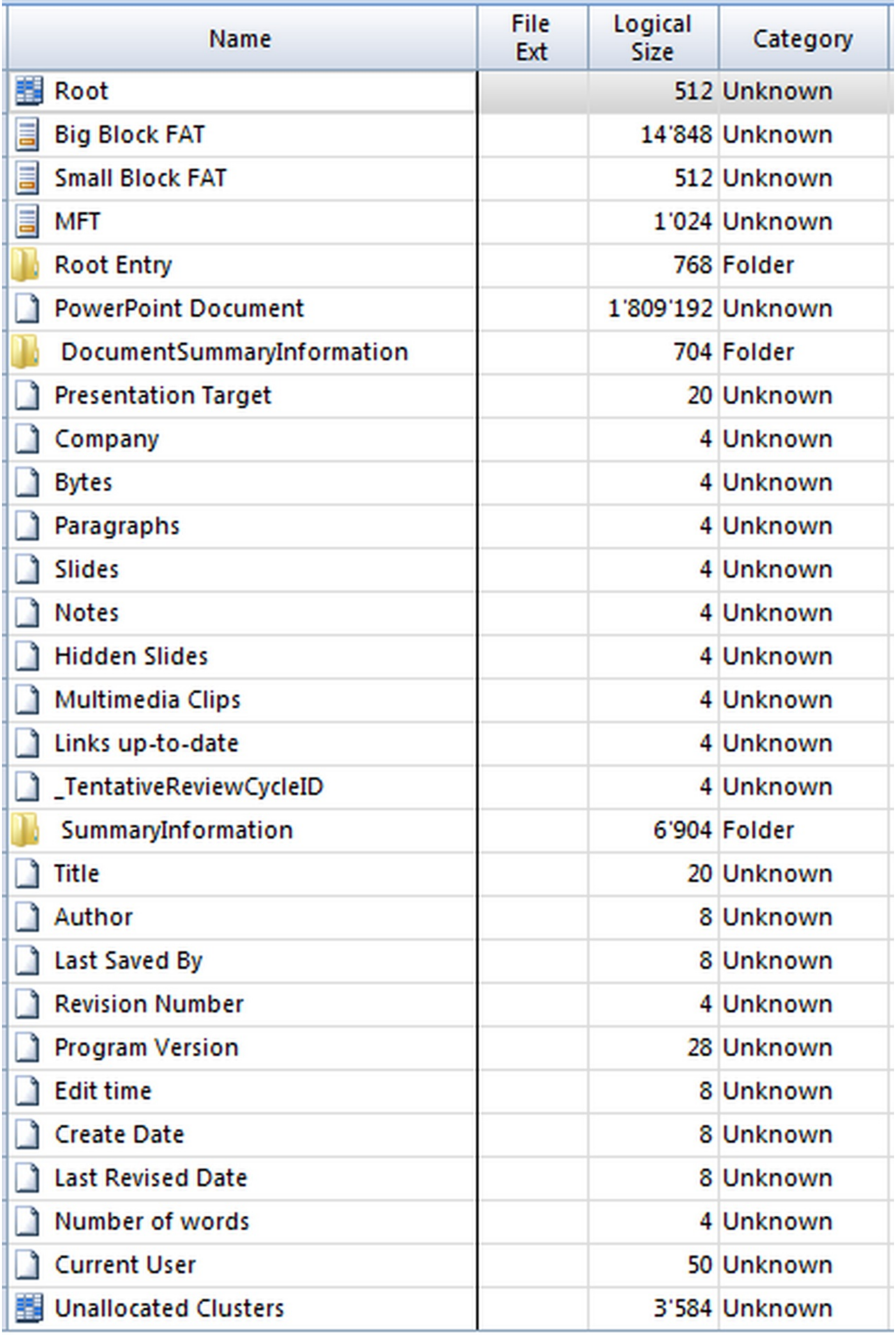


Selected item belongs to the Family of 2 material i

- Test Old PowerPoint - Version 3.ppt
  - [Unnamed Presentation]
    - [Unnamed Image]
    - [Unnamed Image]

EnCase 7 was used to visualize the file structure and try shed some light on where the embedded content was located. Encase was unable to offer any additional insight into this issue, and if we had relied solely on EnCase for our investigation then the child might have remained unnoticed.

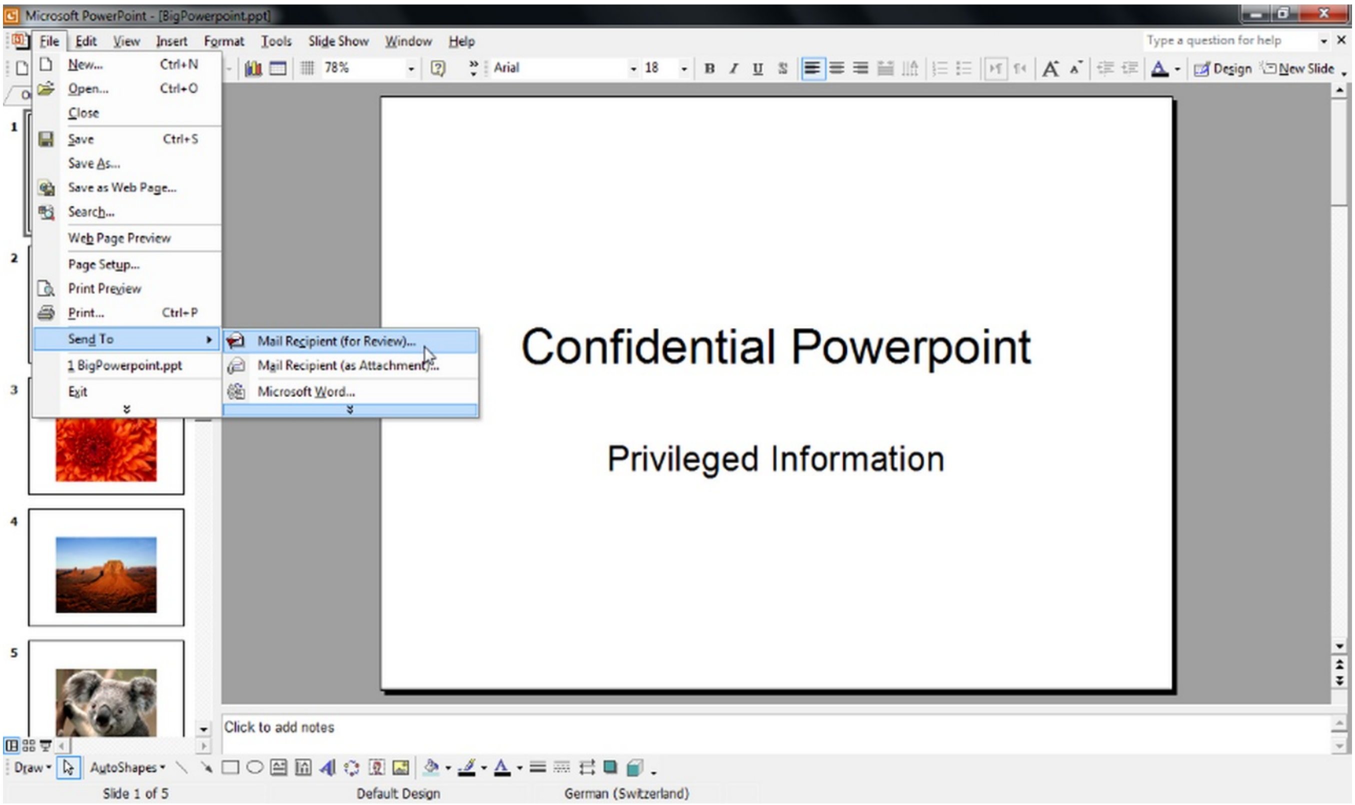
Another way to access embedded content within modern Microsoft Office documents is to change the file extension to "ZIP" and browse through the file's contents like a regular compressed ZIP file. This method has its limitations forensically, but it can be useful to provide more information on the structure and contents of a Microsoft Office file.



Name	File Ext	Logical Size	Category
Root		512	Unknown
Big Block FAT		14'848	Unknown
Small Block FAT		512	Unknown
MFT		1'024	Unknown
Root Entry		768	Folder
PowerPoint Document		1'809'192	Unknown
DocumentSummaryInformation		704	Folder
Presentation Target		20	Unknown
Company		4	Unknown
Bytes		4	Unknown
Paragraphs		4	Unknown
Slides		4	Unknown
Notes		4	Unknown
Hidden Slides		4	Unknown
Multimedia Clips		4	Unknown
Links up-to-date		4	Unknown
_TentativeReviewCycleID		4	Unknown
SummaryInformation		6'904	Folder
Title		20	Unknown
Author		8	Unknown
Last Saved By		8	Unknown
Revision Number		4	Unknown
Program Version		28	Unknown
Edit time		8	Unknown
Create Date		8	Unknown
Last Revised Date		8	Unknown
Number of words		4	Unknown
Current User		50	Unknown
Unallocated Clusters		3'584	Unknown

In our particular case, the parent presentation was in the older PPT format, which is incompatible with this method. A copy of the PPT file was opened with PowerPoint and then saved in the modern PPTX format, enabling the process described above. This revealed the inner structure of the file as expected, but it also had the interesting side effect of removing the child presentation. Neither the child presentation, nor its embedded images, could be found in the ZIP version, and Nuix didn't locate any child presentation when processing the PPTX version of the parent file. The child presentation had been completely removed during the conversion to the newer PPTX format.

After additional research and testing, we eventually discovered the culprit: the Microsoft Office "Send for review" feature. This feature causes the specific behavior noticed during this investigation with the 2002 and 2003 versions of Microsoft Office. Let's dig a bit further into this feature and the consequences of its use.

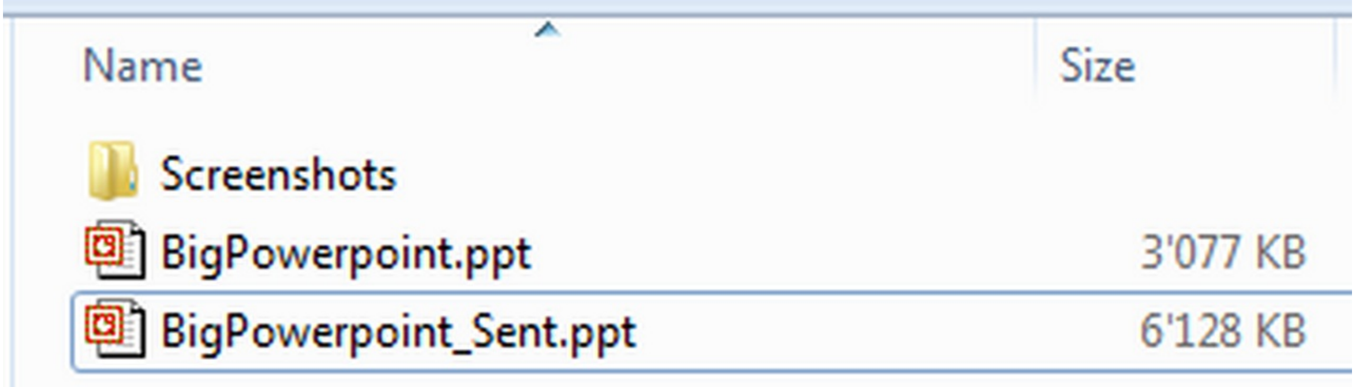


Microsoft Office "Send for review" feature

The website PPTools describes this feature as follows:

> "In PowerPoint 2002 and later [this behavior was removed with Office 2007], you can choose File, Send to, Mail Recipient (for Review). When you do this, the PPT file retains all the original information AND any changes or new information that gets added by the recipient(s). The file will grow every time it's changed in any way, even if the change is deleting material or whole slides. When you enable Review, PowerPoint stores a copy of the original presentation as a hidden OLE object – this is the baseline for comparisons with the presentation itself as it's edited later. It's how PowerPoint knows what's changed and what hasn't."

With these older versions of Microsoft Office, if a PowerPoint presentation is mailed for review using the built-in "Send To" feature, a template email is automatically generated with a modified version of the PowerPoint presentation attached to it. This modified version appears to be almost twice the file size of the original, and actually contains a hidden copy of the original PowerPoint presentation. This hidden copy is designed to help the original author to reconcile and review any changes made to their original PowerPoint presentation. This hidden copy cannot be accessed by anyone through Microsoft Office without also having a copy of the original file.



Name	Size
Screenshots	
BigPowerpoint.ppt	3'077 KB
BigPowerpoint_Sent.ppt	6'128 KB

Data leakage risk

In the best-case scenario, overlooking the use of this feature could pose a risk to your investigation by missing older and non-final versions of PowerPoint files, which could have a varying level of impact depending on the nature of your investigation.

It can be easy to overlook the presence of the child presentation as it is not directly accessible or visible to Microsoft Office users, although the same could be said for any form of hidden information. It is already common knowledge that PowerPoint presentations often contain more information than what is directly visible to the user, for example the source data used to create graphs, and we recommend adding a check for this “send for review” feature to your work process for revealing hidden data.

The core problem with this feature, other than the obvious intentional leaking of data, is the common habit of users reusing old presentations as templates. If the “template” PowerPoint presentation contains a hidden copy of another presentation, the user will not necessarily be aware of it. A copy of the original PowerPoint presentation will remain within any new presentation made using the “sent for review” version. Even if it’s not directly available to standard users, it is there and might contain confidential or sensitive data. Presentations could be sent to clients or shared over the Internet, etc., without the author suspecting the presence of the child presentation and the risks it bears.

Unlike the known issue of embedded tables being used to create graphs in presentations, here the child presentation is completely independent from your final presentation. Even if you delete the entire content of your PowerPoint file, the hidden presentation will remain. Ultimately, a one-slide presentation containing only text can hide an important confidential report without the user even suspecting it.

Conclusion

Nowadays, old formats of Microsoft Office are less common, and therefore the risk of such data leakage seems less likely to occur. Still, a lot of companies are not running the latest versions of Office or prefer old formats for compatibility reasons. It is also worth considering the impact of date ranges in eDiscovery investigations. The scope of investigations often goes back for multiple years, and this increases the probability of encountering this type of hidden data.

In terms of information governance, the same questions remain: do you really know what’s hidden in your data? To which extent does your information governance plan provide the necessary knowledge and control of your data?

If you wish to test your forensic tools as we have regarding this issue, we will be happy to support you and share our test data. Please let us know your tools’ behavior, as knowledge of our tools weaknesses and limits is what makes us experts in our field. For us, we were really appreciative of Nuix's performance with discovering and handling this issue. Even though this might look more like a computer forensic issue than an eDiscovery one, Nuix definitely met our expectations.

Irène Wilson

Irene Wilson specializes in Digital Forensics and eDiscovery, with many years of experience in a wide range of industries across Europe. Her qualifications include the illustrious Master titles for Nuix Workstation and Nuix Discover.