

# KEYSCORE

## Finding and Querying on Document Metadata

Booz|Allen|Hamilton

Sigint Development Support / SIGINT Technical Analysis (SDS/STA)

April 2009

Derived From: NSA/CSSM 1-52

Dated: 20070108

Declassify On: 20291123

DERIVED FROM: NSA/CSSM 1-52

# Document Metadata Agenda



- Why to Query on Document Metadata
- How to Find Document Metadata
  - e.g. File - > Properties
  - Google
- How to Create Queries in XKS
  - XKEYSCORE Document Metadata and PDF Metadata



# Document Metadata Analysis

- **What?:** Use *non-traditional* selectors to find and track targets sending/receiving documents of interest
- **How?** It targets documents by Author, Organization, or embedded images (logos)
- **Why?** We don't always know WHO is sending the documents, but they are "guilty-by-association" if they send/receive the document. So, who are THEY?



# Finding Document Metadata

- We find “Document Metadata” in File Properties

The screenshot shows a Microsoft Word window titled "XKEYSCORE\_Terms.doc - Microsoft Word". The "File" menu is open, and the "Properties" option is highlighted. A red circle highlights the "Summary" tab in the "XKEYSCORE\_Terms.doc Properties" dialog box. The "Summary" tab contains the following fields:

- Title: XKeyscore Terms
- Subject: (empty)
- Author: Joe BaggaDonuts
- Manager: (empty)
- Company: ZMFA Zendian MFA
- Category: (empty)
- Keywords: (empty)
- Comments: (empty)
- Hyperlink base: (empty)
- Template: Normal.dot
- Save preview picture

A red arrow points from the "Properties" menu item to the "Summary" tab. A callout box with a black border and white background contains the text: "If unique, these Document Properties can be targeted".



# Document Metadata Analysis

- How do you find document metadata?
  - Passive Collection: Collected Documents already contain data
  - Active Collection: CNE “Categorized Collection” from TUNINGFORK Data or Pinwale Queries on “US-3101”
  - Open Source: Google Hacking



# Finding Document Metadata

ID	To	From	Cc	Bc	Date	Subject	Size(K)	Type
48	[REDACTED]	"DESTOCKPRO" <des@...>			5/15/2008 6:31:35 PM	ARRIVAGE G STAR DOLCE&GABBANA DIESEL	16	text/html
49	[REDACTED]	"Target Center" <...>			5/16/2007 9:36:13 PM	Confirmation: Target Card	2	text/html
51	[REDACTED]	[REDACTED]			5/15/2008 11:14:32 PM		1	text/html
41	[REDACTED]	[REDACTED]			5/15/2008 3:05:25 PM	Las Villas de Dubai	3898	application/octet-stre
32	[REDACTED]	[REDACTED]			5/15/2008 11:15:21 PM	skriv ut	452	application/msword
							50	application/msword
51		"C. Kerlin" <...>			5/15/2008 11:14:32 PM		452	application/msword
							50	application/msword

Display original Raw SMTP header **Properties** Control C2C Trailer Collected Doc Search Kwd Dictionary Kwd SRI

## Document Properties

Category	
Company	
HiddenSlideCount	0
LineCount	29
LinksUpToDate	False
Manager	
MMClipCount	29
NoteCount	29
ParagraphCount	8
PresentationTarget	
ScaleCrop	False
SlideCount	8
AppName	Microsoft Word 10.0
Author	GoGo
CharacterCount	3582
Comments	
DateCreated	5/12/2008 3:13:00 AM
SecurityLevel	none
Keywords	
LastAuthor	[REDACTED]

3898	application/octet-stream	bas
452	application/msword	bas
50	application/msword	bas



# Finding Document Metadata

- Active Collection: CNE “Categorized Collection” from TUNINGFORK Data



## Collection ?

No EP user information found.

[Raw Project Details](#)[s3115 only]

[Mailbox Collection](#)

Last Collection [limit 3 dates listed]: [2008-08-29](#)  
[2008-08-27](#)  
[2008-07-19](#)

[List All](#)

[Collection](#)

[Categorized Collection](#)

Cipher (8)	MicroSoft (277)	Multimedia (17)	Mail (35)	Inst Msgr (9)	VOIP (1642)	HTM
Cipher	Access (0)					
<input type="checkbox"/> Show Pat	Excel (2)	Filename	Extension	Collected	Size	
	Execs (4)	▲▼	▲▼	▲▼	▲▼	
<a href="#">81e0bd</a>	Ini files (2)	21-4a68af648ec5		2008-07-19	388	
<a href="#">b850ea</a>	Other Office (5)			2008-07-19	388	
<a href="#">0c6527</a>	Powerpoint (0)	31d-3dff4d38926				
	Thumbs.db (12)					
	Word (252)	30-c1ed1756266f		2008-03-13	388	
<a href="#">5443e72e17e4278e09b50950d454e5f</a>				2008-03-13	388	

To find Document Metadata in TUNINGFORK, you must view each Document in Categorized Collection (manual intensive)



# Finding Document Metadata

- Using XKEYSCORE to query on CNE data

Fields ▾ Advanced Features ▾ Show Hidden Search Fields Clear Search Values Reload Last Search Values

Search: Document Metadata

Input Source:

Selected implant exfils from active collection (xks-cne.corp.nsa.ic.gov:xs\_web\_db)

This query in XKS

Filename	Extension	Author	Title	Input Source
System Admin CV.doc	doc	Authorised User		XXXXXXXXXXE4
C:\Downloads\Physical_Layer_in_RPR_0402.pdf	pdf			XXXXXXXXXXER12
C:\Documents and Settings\Guest\Desktop\05070807_excelbook.pdf	pdf	Center For Excelence	(Microsoft Word - 130713411343133613	XXXXXXXXXXER12
C:\Documents and Settings\user\Desktop\desktop icons\yementrack.doc	doc	user		XXXXXXXXXXER12
C:\Downloads\ins-3-overview.pdf	pdf			XXXXXXXXXXER12
C:\Documents and Settings\user\Desktop\desktop icons\servers_exp.doc	doc	results	الأخ/ مدير عام المشتريات و المخازن المطرم	XXXXXXXXXXER12

Produced these results





# Finding Document Metadata

## ■ Open Source: Google Hacking

The Google logo in its multi-colored font.

[Advanced Search](#) [Search Tip](#) [Help](#)

site:comsats.net.pk filetype:doc

Google Search

Or....

filetype:xls

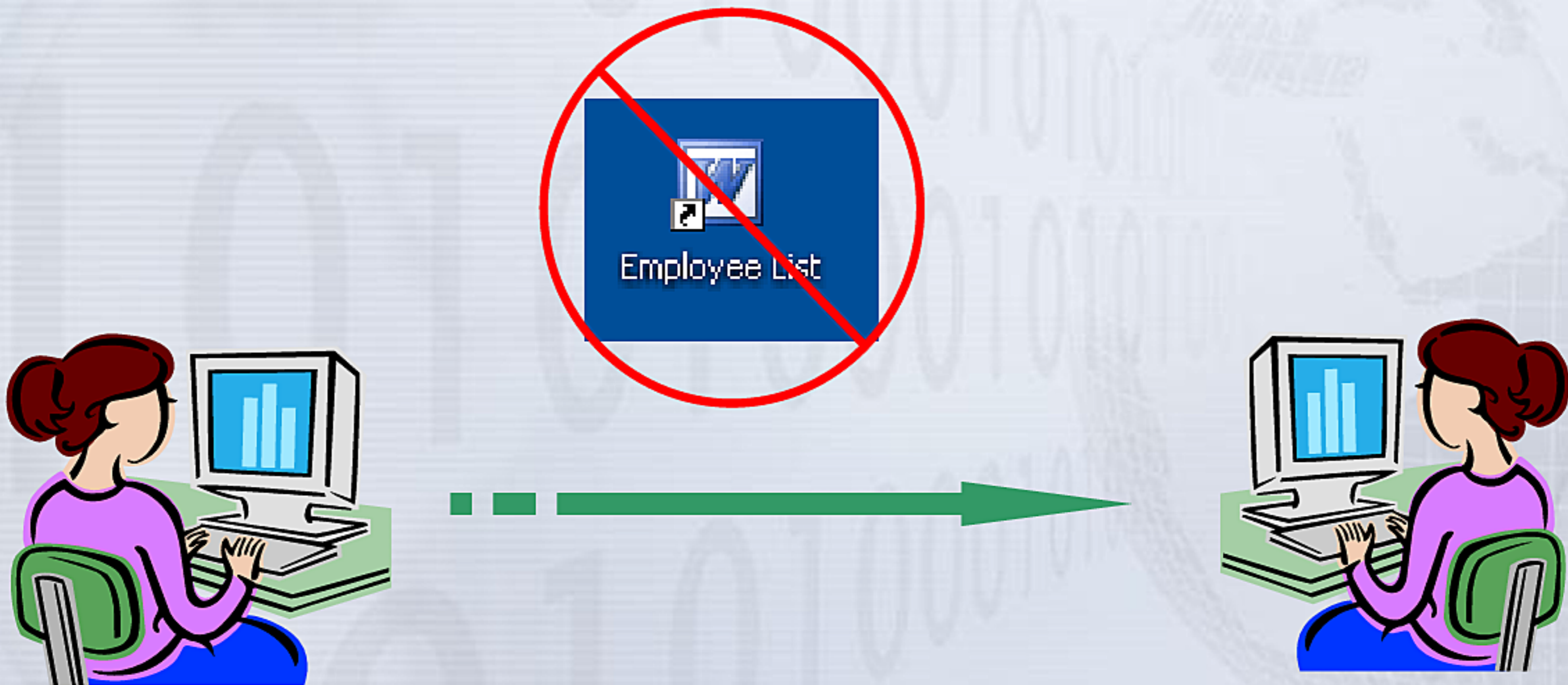
filetype:pdf

- Search by domains
  - “site:comsats.net.pk”
- Search by file types
  - “filetype:pdf” or “filetype:doc”



# Document Metadata Analysis

How to find Document Metadata when you have NEVER collected a document







# Targeting Document Metadata

- Use XKEYSCORE to Find Who Else is sending the files?



# Document Metadata Analysis

Take "File Properties" information and fill-in query

**XKEYSCORE\_Terms.doc Properties**

General Summary Statistics Contents Custom

Title: XKeyscore Terms

Subject:

Author: Joe BaggaDonuts

Manager:

Company: ZMFA Zendian MFA

Category:

Keywords:

Comments:

Hyperlink base:

Template: Normal.dot

Save preview picture

OK Cancel

## Search: Document Metadata

Document Type:

Encrypted?:

Corrupted?:

Filename:

Extension:

\*Subject\*:

\*Creation Time\*:

\*Last Modified Time\*:

\*Unique ID\* [fulltext]:

Author: Joe BaggaDonuts

Last Author:

Organization: ZMFA Zendian MFA

Title:

Language:

\*Comment\* [fulltext]:

File/Embedded Image Hash [fulltext]:

Metadata Name:

Metadata Value [fulltext]:



# Document Metadata Analysis

## Sample Query

Sample Query:

Organization = PTCL

To/From Country = Pakistan

Search: Document Metadata

Organization:

Title:

Language:

\*Comment\* [\[fulltext\]](#):

File/Embedded Image  
Hash [\[fulltext\]](#):

Metadata Name:

Metadata Value [\[fulltext\]](#):

IP Address:

From

IP Address:

To

Port:

From

Port:

To

Country:

Either



# Document Metadata Analysis

## Sample Query (Results)

Previous Slide produces these results

Filename	Organization
<b>Instructions to Kunar province bidders community midwifery.doc</b>	<b>PTCL</b>
<b>Instructions to Kunar province bidders community midwifery.doc</b>	<b>PTCL</b>



# Embedded Images

- Turn a logo into a selector



= SIGINT VALUE





# Embedded Images

- XKEYSCORE parses out logos from within documents (PDFs, DOCs, Outlook Emails, etc) embedded as images

The screenshot shows a document viewer interface. At the top, there is a toolbar with icons for printing, saving, navigation, and zooming (set to 80.6%). Below the toolbar, the document content is displayed. A logo for "Yemenet" is circled in red. To its right, a 32-character hash is highlighted in a blue box and also circled in red. A red arrow points from the hash to the logo. A text box on the right explains that the hash can be parsed and queried. Below the logo, there is Arabic text: "المحترم" and "مخازن". The main heading of the document is "الموضوع : طلب مراسلة للشركات المقدمة في مناقصة توسعة سرفرات خدمات الانترنت". Below this, there is a table with three columns: "الصفحة", "اسم الشركة", and "الملاحظات". The table contains one row of data.

Logo/Image 32-character hash can be parsed out and queried.

المحترم

مخازن

**الموضوع : طلب مراسلة للشركات المقدمة في مناقصة توسعة سرفرات خدمات الانترنت**







بالإشارة الى الموضوع أعلاه نرجوا مخاطبة الشركات المذكورة بالملاحظات الخاصة بكل منها :

الصفحة	اسم الشركة	الملاحظات
	MDS	لم تذكر: MB L2 On Chip Cache per Processor 2• System Controller Card 1• Solaris 10 03/05 HW1 Operating System Preinstalled•



# Embedded Images

Files often contain embedded images, such as company logos.

	Summary	GPRS TLLI
تروات خدمات الإنترنت بالإنترنت إلى الموضوع إعلان بريد		
نرجو التوجه إلى من يلزم بتفعيل خدمة ADSL الأخر		
		
بالأرقام ١٥٢٥١٦ للإعدادات والآراء ١١ و ١٥٢٥١٦		
VPN Clients Configuration Example I		
١٨ 12/1/2008 2007 السنوي لإدارة الإنترنت للعام		
by ITE. 316) & Abdullah Mohammed		

Step 1: Identify if a document HAS an image in it



# Embedded Images

Datetime	Case Notation	From IP
2009-03-26 15:54:50	YM.PGQXXXABDDTC	

[Session](#) | [Header \(3\)](#) | [Attachments \(6\)](#) | [Meta \(3\)](#)

Formatter: [AUTO](#) | Send to: [Download Session](#) | **Mode: [Full Session](#)** | [Options](#)

**Quick Clicks** ⏪ ⚙️ Retrieving Attachment...

- Session
- Attachments
  - sigint
    - image\_summary\_mont
    - image\_summary\_m
    - document\_meta
    - c:\documents and s
  - unknown
    - text
      - document\_body.رسید
      - document\_body.ی (ره)
  - image
    - jpeg
      - b3d7853e4bfde7087
  - office
    - pdf
      - C:\Documents and S
- One-Click Searches**
  - Find opposite side of sessi
    - :0 ->
    - :0
  - Find More Docs with Same
    - 635ed0657cfe25b7790f
    - b3d7853e4bfde70874cf

Step 2: Open Document and click on "Full Session"



# Embedded Images

The screenshot shows a network analysis tool interface. On the left, a 'Quick Clicks' menu lists various file types, with 'image' and 'jpeg' sub-categories circled in red. A specific file, 'b3d7853e4bfde7087', is selected under the 'jpeg' category. The main pane displays details for this file, including the filename 'b3d7853e4bfde70874cf402a3d6cfe10.jpg' circled in red, and the text 'Virus scan results'. Below this, a blue box highlights the 32-character name 'b3d7853e4bfde70874cf402a3d6cfe10.jpg'. The background of the main pane features the logo of the General Telecommunication Institute (GTI) with Arabic text 'المعهد العام للاتصالات'.

Step 3: In left-side menu bar, select an image and copy/paste the 32-character name (without the extension)



# Embedded Images

Step 4: Paste the 32-character name into the "File/Embedded Image Hash" Field in the Document Metadata query

A screenshot of the KEYScore search interface. On the left is a navigation pane with a tree view under "Classic A-M" containing items like "ASF and WMV Metadata", "Alert", "BlackBerry", "CNE", "Call Logs", "Category DNI", "Cellular DNI", "Cisco Passwords", "Document Metadata", and "Document Tagging". The "Document Metadata" item is selected. The main area has a search bar with "Search: Document Metadata" and a "File/Embedded Image Hash [fulltext]:" field containing the value "b3d7853e4bfde70874cf402a3d6cfe10". At the top of the main area are links for "Fields", "Advanced Features", "Show Hidden Search Fields", "Clear Search Values", and "Reload Last Search Values".

Step 5: Select all of your good collection sites + SUBMIT!

 A screenshot of the KEYScore search database selection screen. On the left, under "Search Databases", are "Clear Checks" and "Reset Checks" buttons. On the right, there is a list of search databases with checkboxes:
 

- (xks-central.corp.nsa.ic.gov:qsummary)
- Australian sites** (xkcentral2.dsd:xs\_web\_db)
- CARBOY** (carboy-proxy.r1.r.nsa:carboy\_web\_db)
- CARDAMON** (xkey-dsd.r1.r.nsa:xs\_web\_db)

 At the bottom are "Submit" and "Cancel" buttons.



# Embedded Images

Session | Header (3) | Attachments (6) | Meta (4)

Formatter: AUTO | Send to: Down

Quick Clicks

- image\_summary\_montage.jpeg
- document\_meta
  - c:\documents and settings\usuaric
- unknown
- text
  - document\_body.SOLICITANTE .txt
- office
  - word
    - C:\Documents and Settings\usuari
- One-Click Searches**
  - Find opposite side of session
    - :0 ->
    - :0
  - Find More Docs with Same hash
    - a97d82d06aaa9017cacbe5fe4b12f15c
    - f4c6353ebd01ba02b7c087a91bdf29c4
- Find email address
  - zakimoussa@hotmail.com

Or... You can one-click query to create a new query

## Search: Document Metadata

Query Name:

Justification:

Additional Justification:

Miranda Number:

Language:

\*Comment\* [fulltext]:

File/Embedded Image Hash [fulltext]:





# Embedded Images

- Stand-alone files can be uploaded into XKS and images parsed out
  - Useful for TAO collection that didn't get into XKS (non United Rake)
  - [https://xks-central.corp.nsa.ic.gov/general/view\\_file.php](https://xks-central.corp.nsa.ic.gov/general/view_file.php)



This system is au  
CLASSIFICATION: TOI

**XKEYSC**

You can upload SOTF and D-124 files, as well as just random files (.doc, .ppt, etc.)

Upload File

This system is au  
CLASSIFICATION: TOI

- To task the hex values for images in CADENCE or Query in PINWALE, contact The Xtreme Target Pursuit Team [REDACTED] [REDACTED] S217 and [REDACTED] S3114



# Embedded Images

- Questions on any of these tools or techniques, contact:
  - [REDACTED]
  - [REDACTED]