



Finding and Querying on Document Metadata

[REDACTED]
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

XKEYSCORE_Terms.doc - Microsoft Word

File Edit View Insert Format Tools Table Window Help SecretAgent Classif

- New...
- Open... Ctrl+O
- Close
- Save Ctrl+S
- Save As...
- Save as Web Page...
- Remove Hidden Data...
- File Search...
- Versions...
- Web Page Preview
- Page Setup...
- Print Preview
- Print... Ctrl+P
- Send To
- Properties
- 1 U:\private\andrew.dnr
- 2 U:\...\XKEYSCORE Tips and Tricks 4 ap...
- 3 U:\private\Présentation.doc
- 4 U:\private\...\NIA Cross Training.doc
- 5 U:\private\NIA\MSRs\April 2009.doc
- 6 C:\Documents and Settings\...\IMEI.doc
- 7 U:\...\NIA\RITCHIE_DNI_IWPS(revised).doc
- 8 U:\...\XKS_kmkeith_tips7April.doc
- 9 U:\...\Zwakenberg, Garrit Trey IPWS.doc
- Exit

STM4, STM16, STM64, multip

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
<input type="checkbox"/> Save preview picture	

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	Bcc	Date	Subject	Size(K)	Type
48	cezad...@y...com" <cezad...@y...com>	"DESTOCKPRO" <destockpro...@y...com>			5/16/2008 6:31:35 PM	ARRIVAGE G STAR DOLCE&GABBANA DIESEL	16	text/html
49	...@y...com" <...@y...ahoo.com>	"Target Center" <target...@m.../3.../y...com>			5/16/2007 9:38:13 PM	Confirmation: Target Card	2	text/html
51		"...@y...com" <...@y...com>			5/16/2008 11:14:32 PM		1	text/html
41	...@y...com" <...@y...mail.com>, "abu zubeer alyamagi" <...@y...com>	"...@y...com" <...@y...com>			5/16/2008 3:05:26 PM	Las Villas de Dubai	3898	application/octet-stream
32	"...@y...com" <...@y...com>	"...@y...com" <...@y...com>			5/16/2008 11:15:21 PM	skriv ut	452	application/msword
							60	application/msword
51		"...@y...com" <...@y...com>			5/16/2008 11:14:32 PM		452	application/msword
							60	application/msword

Document Properties	
Category	
Company	
HiddenSlideCount	0
LineCount	29
LinksUpToDate	False
Manager	
MMCipCount	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]

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)	Filename	Extension	Collected	Size	
<input type="checkbox"/> Show Pat	Excel (2)					
	Execs (4)					
81e0bd	Ini files (2)	21-4a68af648ec5		2008-07-19	388	
	Other Office (5)					
b850ea	Powerpoint (0)	31d-3dffb4d38926		2008-07-19	388	
	Thumbs.db (12)					
0c6527	Word (252)	30-c1ed1756266f		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

This query in XKS

Input Source: FOXACID* or FOXBASE*

Selected implant exfils from active collection (xks-cne.corp.nsa.ic.gov:xs_web_db)

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 Excellence	(Microsoft Word - 130713411343133613)	
C:\Documents and Settings\user\Desktop\desktop icons\yementracker.doc		user		XXXXXXXXXXER12
C:\Downloads\ins-3-overview.pdf	pdf			XXXXXXXXXXER12
C:\Documents and Settings\user\Desktop\desktop icons\servers_explor.doc		results	الباحث / مدير عام المشتريات و المخازن المحترف	XXXXXXXXXXER12

Produced these results



Finding Document Metadata

■ Open Source: Google Hacking



[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”

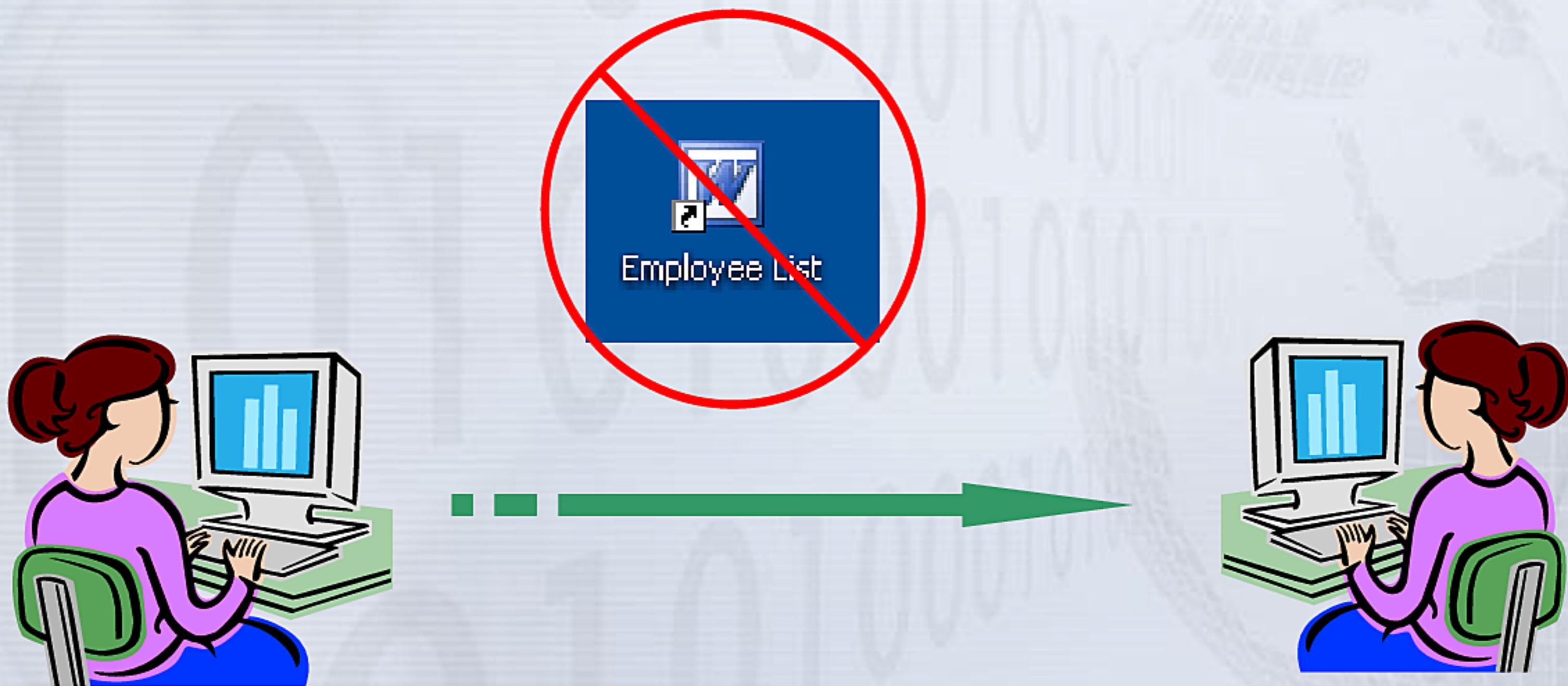
■ Search by file types

- “filetype:pdf” or “filetype:doc”

Document Metadata Analysis



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



Document Metadata Analysis



Take Client's (Active User) IP address and query on it in XKEYSCORE



Active User:
[REDACTED]@yahoo.com

ACTIVE_USER	ACTIVE_USER_IP
[REDACTED]<yahoo>	89.[REDACTED]
[REDACTED]<yahoo>	89.[REDACTED]

Search: Document Metadata

Extension: ppt or doc or pdf or xls

IP Address: 89.[REDACTED] Either

ACTIVE_USER ACTIVE_USER_IP
[REDACTED]<yahoo> 89.[REDACTED]
[REDACTED]<yahoo> 89.[REDACTED]

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				
<input type="checkbox"/> Save preview picture					
<input type="button" value="OK"/> <input type="button" value="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
Instructions to Kunar province bidders community midwifery.doc	PTCL
Instructions to Kunar province bidders community midwifery.doc	PTCL



Embedded Images

- Turn a logo into a selector



Embedded Images



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

**الموضوع : طلب مراسلة للشركات المقدمة في مناقصة توسيعه
سرفات خدمات الانترنت**

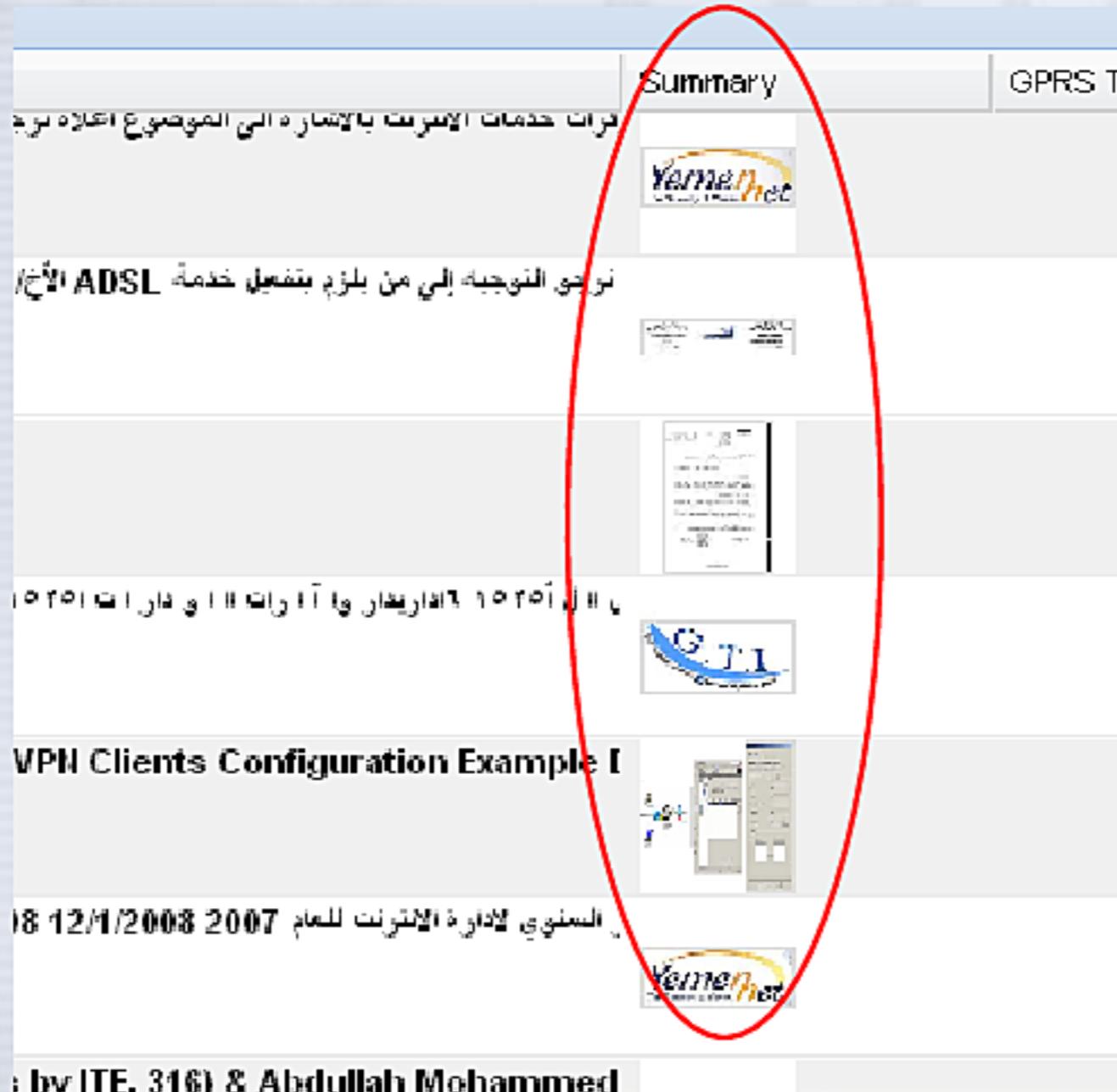
بالإشارة الى الموضوع أعلاه نرجوا مخاطبة الشركات المذكورة بالملحوظات الخاصة بكل منها :

الملحوظات	اسم الشركة	الصنف
لم تذكر: MB L2 On Chip Cache per Processor 2+ System Controller Card 1+ Solaris 10 03/05 HW1 Operating System Preinstalled+	MDS	Sun Fire V490 Server



Embedded Images

Files often contain embedded images, such as company logos.



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_month
- image_summary_month

document_meta

- c:_documents and s

unknown

text

- document_body.wor
- document_body.wor

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 on "Full Session"

Step 2: Open Document and click on “Full Session



Embedded Images

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

Formatter: AUTO | Send to: Download Session | Mode: Full Session | Options | Search Con

Quick Clicks

- Session
- Attachments
- sigint
 - image_summary_monthly
 - image_summary_monthly
 - document_meta
 - c:_documents and s
- unknown
- text
 - ? document_body.دокумент
 - ? document_body.و
- image
 - jpeg
 - b3d7853e4bfde70874cf402a3d6cfe10.jpg
- office
 - pdf
 - C:\Documents and S
- One-Click Searches
 - Find opposite side of sessi

:0 ->

Virus scan results

Using IMAGE format

b3d7853e4bfde70874cf402a3d6cfe10.jpg

General Telecommunication Institute جهاز الاتصالات العامة

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

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

Search: Document Metadata

File/Embedded Image Hash [fulltext]: b3d7853e4bfde70874cf402a3d6cfe10

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

Search Databases <input type="button" value="Clear Checks"/> <input type="button" value="Reset Checks"/>	<input checked="" type="checkbox"/> (xks-central.corp.nsa.ic.gov:qsummary) <input checked="" type="checkbox"/> Australian sites (xkcentral2.dsdxs_web_db) <input checked="" type="checkbox"/> CARBOY (carboy-proxy.r1.r.nsa:carboy_web_db) <input checked="" type="checkbox"/> CARDAMON (xkey-dsd.r1.r.nsa:xs_web_db)
<input type="button" value="Submit"/> <input type="button" value="Cancel"/>	



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\usuari
- > 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: One-click search on document hash: f4c6353

Justification: One-click search to find more documents with

Additional Justification:

Miranda Number:

Language:

Comment [fulltext]:

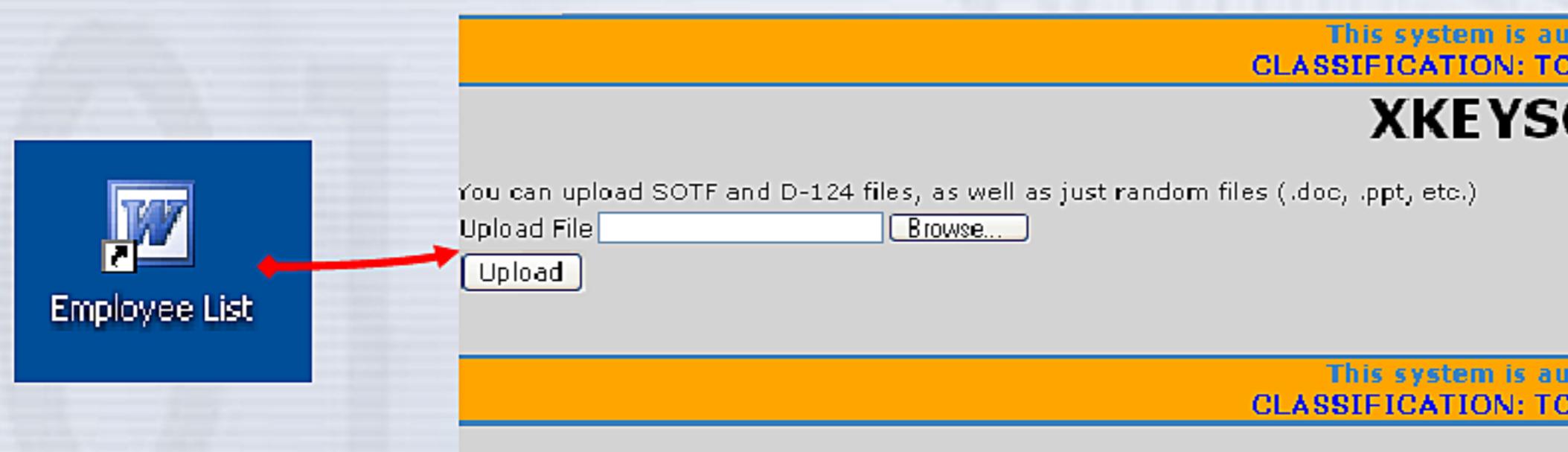
File/Embedded Image Hash [fulltext]:

f4c6353ebd01ba02b7c087a91bdf29c4



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



A screenshot of a web-based file upload interface. At the top, there's a yellow bar with the text "This system is audited" and "CLASSIFICATION: TOP SECRET". Below that is a grey header with the "XKEYSCORE" logo. The main area has a light grey background. On the left, there's a blue button labeled "Employee List" with a white "W" icon. A red arrow points from this button to the "Upload File" input field. To the right of the input field is a "Browse..." button. Below these is a "Upload" button. At the bottom, there's another yellow bar with the same audit and classification information.

- To task the hex values for images in CADENCE or Query in PINWALE, contact The Xtreme Target Pursuit Team [REDACTED] S2I7 and [REDACTED] S3114



Embedded Images

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