



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

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

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" <[redacted]>			5/15/2008 6:31:35 PM	ARRIVAGE G STAR DOLCE&GABBANA DIESEL	16	text/html
49	[redacted]	"Target Card" <[redacted]>			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]	"abu zubeer alyamagi" <[redacted]>			5/15/2008 3:05:25 PM	Las Villas de Dubai	3898	application/octet-stream
32	[redacted]	[redacted]			5/15/2008 11:15:21 PM	skriv ut	452	application/msword
61	[redacted]	[redacted]			5/15/2008 11:14:32 PM		50	application/msword

Display original Raw SMTP header **Properties** Control C2C Trailer Collected Doc Search Kwd Dist 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
<input type="checkbox"/> Show Pat	Excel (2)	Filename	Extension	Collected	Size		
	Execs (4)						
<input type="checkbox"/> 81e0bc	Ini files (2)	21-4a68af648ec5		2008-07-19	388		
<input type="checkbox"/> b850ea	Other Office (5)	1d-3dff4d38926		2008-07-19	388		
<input type="checkbox"/> Oc6527	Powerpoint (0)	B0-c1ed1756266f		2008-03-13	388		
	Timings.db (12)						
	Word (252)						

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 exfls from active collection (xks-cne.corp.nsa.ic.gov:xs web db)

This query in XKS

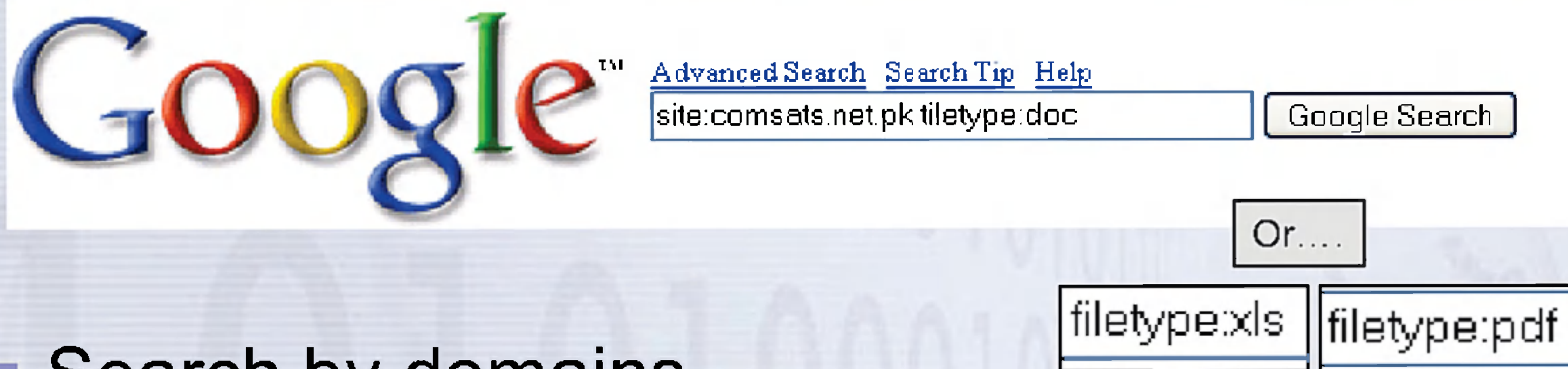
Filename	Extension	Author	Title	Input Source
.System Admin CV.doc	doc	Authorised User		XXXXXXXXXXE4
C:\Downloads\Physical_Layer_in_RPR_0402.pdf	pdf			XXXXXXXXXXER12
C:\Documents and Settings\Guest\Desktop\05070807_excelbook.pdf	pdf	Center For Excellence	{Microsoft Word - 130713411343133613}	XXXXXXXXXXER12
C:\Documents and Settings\user\Desktop\desktop icons\yenientrack.doc	doc	user		XXXXXXXXXXER12
C:\Downloads\ns-3-overview.pdf	pdf			XXXXXXXXXXER12
C:\Documents and Settings\user\Desktop\desktop icons\servers_exp.doc	doc	results	الاج مدير عام المستويات و المخازن المحتزم	XXXXXXXXXXER12

Produced these results

Finding Document Metadata



- Open Source: Google Hacking

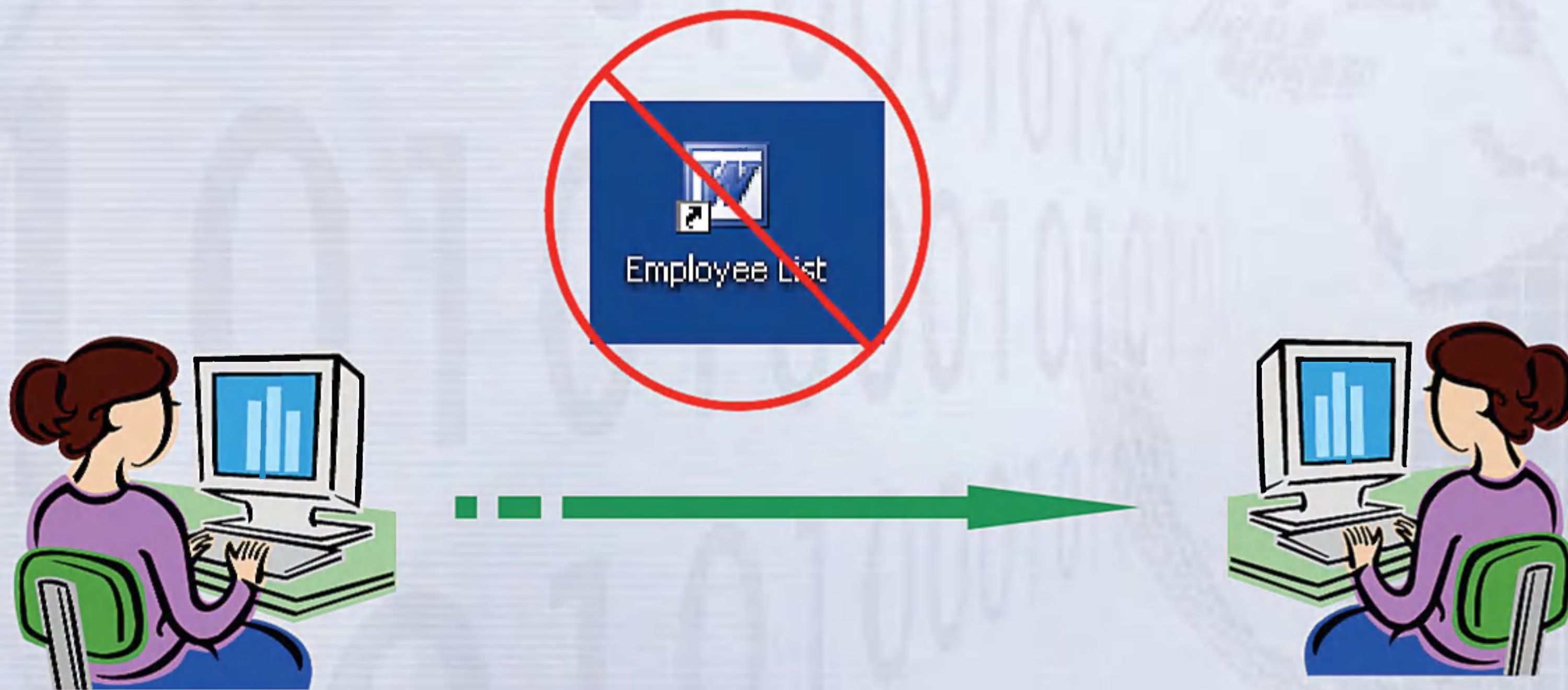


- Search by domains
 - "site:comsats.net.pk"
- Search by file types
 - "filetype:pdf" or "filetype:doc"

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:

██████████@yahoo.com

ACTIVE_USER	ACTIVE_USER_IP
██████████<yahoo>	89 ██████████
██████████<yahoo>	89 ██████████

Search: Document Metadata

Extension:

IP Address:

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
Instructions to Kumar province bidders community midwifery.floc	PTCL
Instructions to Kumar province bidders community midwifery.doc	PTCL



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


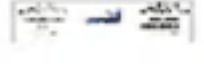



The screenshot shows a document viewer interface. At the top, there is a toolbar with icons for printing, saving, navigation, and a search box labeled 'Find'. Below the toolbar, the document content is displayed. On the left side, there is a sidebar with a document icon and a green question mark icon. The main content area features the 'Yemen.net' logo, which is circled in red. To the right of the logo, a blue box contains a 32-character hash: `image6b71557c23ce00a03a662d6dd1eb1c1b.jpeg`. A red arrow points from this hash to the logo. A text box on the right contains the text: 'Logo/Image 32-character hash can be parsed out and queried.' Below the logo, there is Arabic text: 'المحترم' and 'مخازن'. The main heading in Arabic is: 'الموضوع : طلب مراسلة الشركات المقدمة في مناقصة توسعة سرفرات خدمات الانترنت'. Below this, there is a sub-heading: 'بالإشارة الى، الموضوع أعلاه ترحباً مخاطبة الشركات المذكورة بالملاحظات الخاصة بكل منها :'. At the bottom, there is a table with three columns: 'الصفحة', 'اسم الشركة', and 'الملاحظات'. The table contains one row with the following data:

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
خدمات الإنترنت بالإنجاز التي المصنوع أعلاه يوجد		
برامج التوجيه التي من يلزم بفتح خدمة ADSL الأخر		
الخدمات التي تقدمها شركة الاتصالات		
الخدمات التي تقدمها شركة الاتصالات		
VPII Clients Configuration Example I		
المعنى لتأثير الإنترنت للعام 2007 12/1/2008		

: by (TE. 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.رسید
 - image
 - jpeg
 - b3d7853e4bfde7087
 - office
 - pdf
 - C:\Documents and S
- One-Click Searches
 - Find opposite side of sess
 - :0 ->
 - :0
 - Find More Docs with Same
 - 635ed0657cfe25b7790f
 - b3d7853e4bfde70874cf

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_mont
 - image_summary_m
 - document_meta
 - c:\documents and s
 - unknown
 - text
 - document_body.رسید
 - document_body.ره
 - image
 - jpeg
 - b3d7853e4bfde7087**
 - office
 - pdf
 - C:\Documents and S
- One-Click Searches
 - Find opposite side of sessi

b3d7853e4bfde70874cf402a3d6cfe10.jpg
Virus scan results

b3d7853e4bfde70874cf402a3d6cfe10.jpg

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

Classic A-M

- ASF and VMV Metadata
- Alert
- BlackBerry
- CNE
- Call Logs
- Category DNI
- Cellular DNI
- Cisco Passwords
- Document Metadata
- Document Tagging

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

Search: Document Metadata

File/Embedded Image Hash [fulltext]:

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

Search Databases

(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)

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:

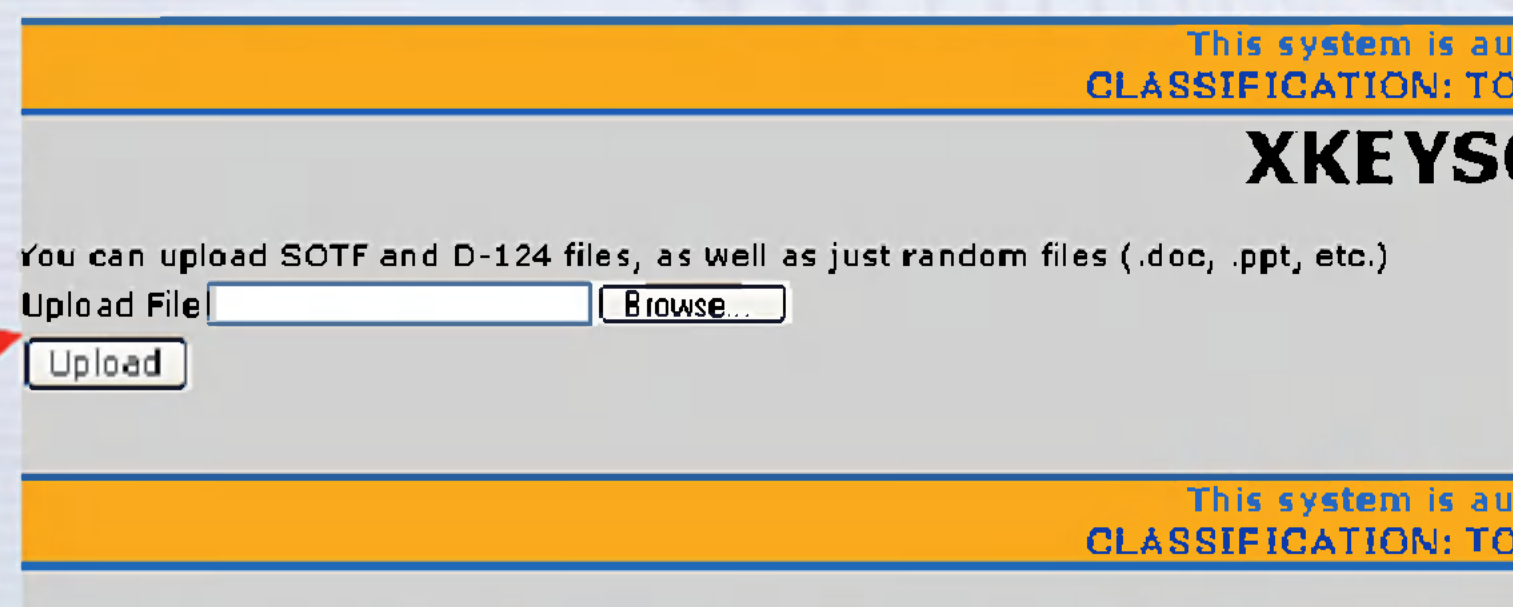
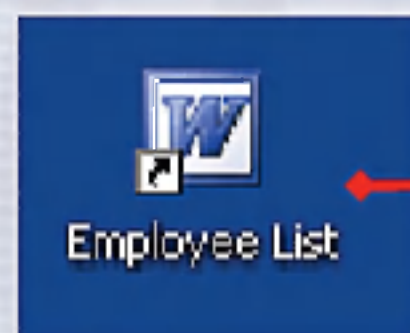
* Comment* [fulltext]:

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



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



Embedded Images

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