

**TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL**



# XKEYSCORE

25 Feb 2008  
xkeyscore@nsa

**TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL**

**DERIVED FROM: NSA/CSSM 1-52  
DATED: 20070108  
DECLASSIFY ON: 20320108**





# What is XKEYSCORE?

1. DNI Exploitation System/Analytic Framework
2. Performs strong (e.g. email) and soft (content) selection
3. Provides real-time target activity (tipping)
4. "Rolling Buffer" of ~3 days of ALL unfiltered data seen by XKEYSCORE:
  - Stores full-take data at the collection site – indexed by meta-data
  - Provides a series of viewers for common data types
5. Federated Query system – one query scans all sites
  - Performing full-take allows analysts to find targets that were previously unknown by mining the meta-data



# Methodology



- Small, focused team
- Work closely with the analysts
- Evolutionary development cycle (deploy early, deploy often)
- React to mission requirements
- Support staff integrated with developers
- Sometimes a delicate balance of mission and research



# System Details

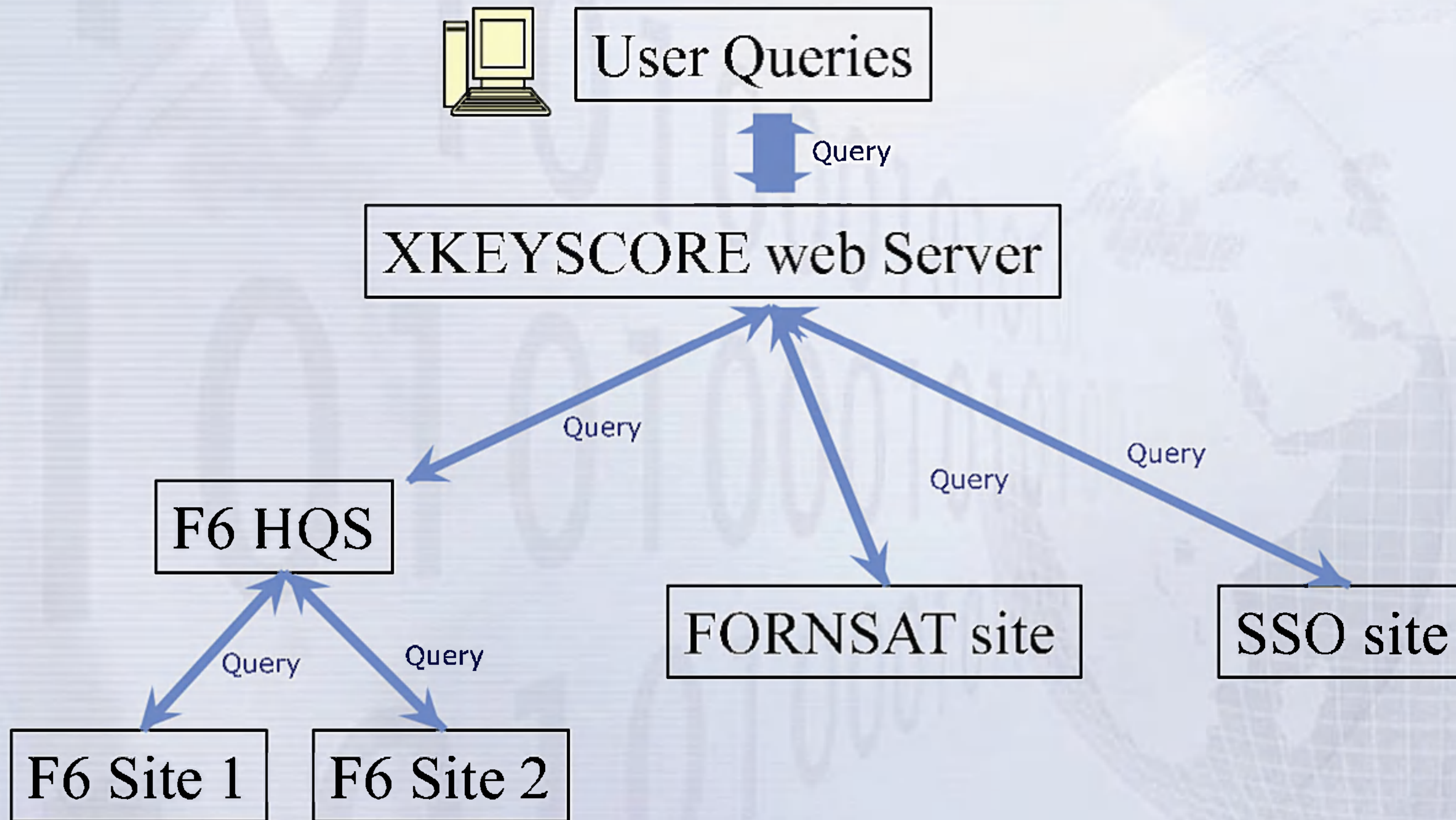


- Massive distributed Linux cluster
- Over 500 servers distributed around the world
- System can scale linearly – simply add a new server to the cluster
- Federated Query Mechanism





# Query Hierarchy

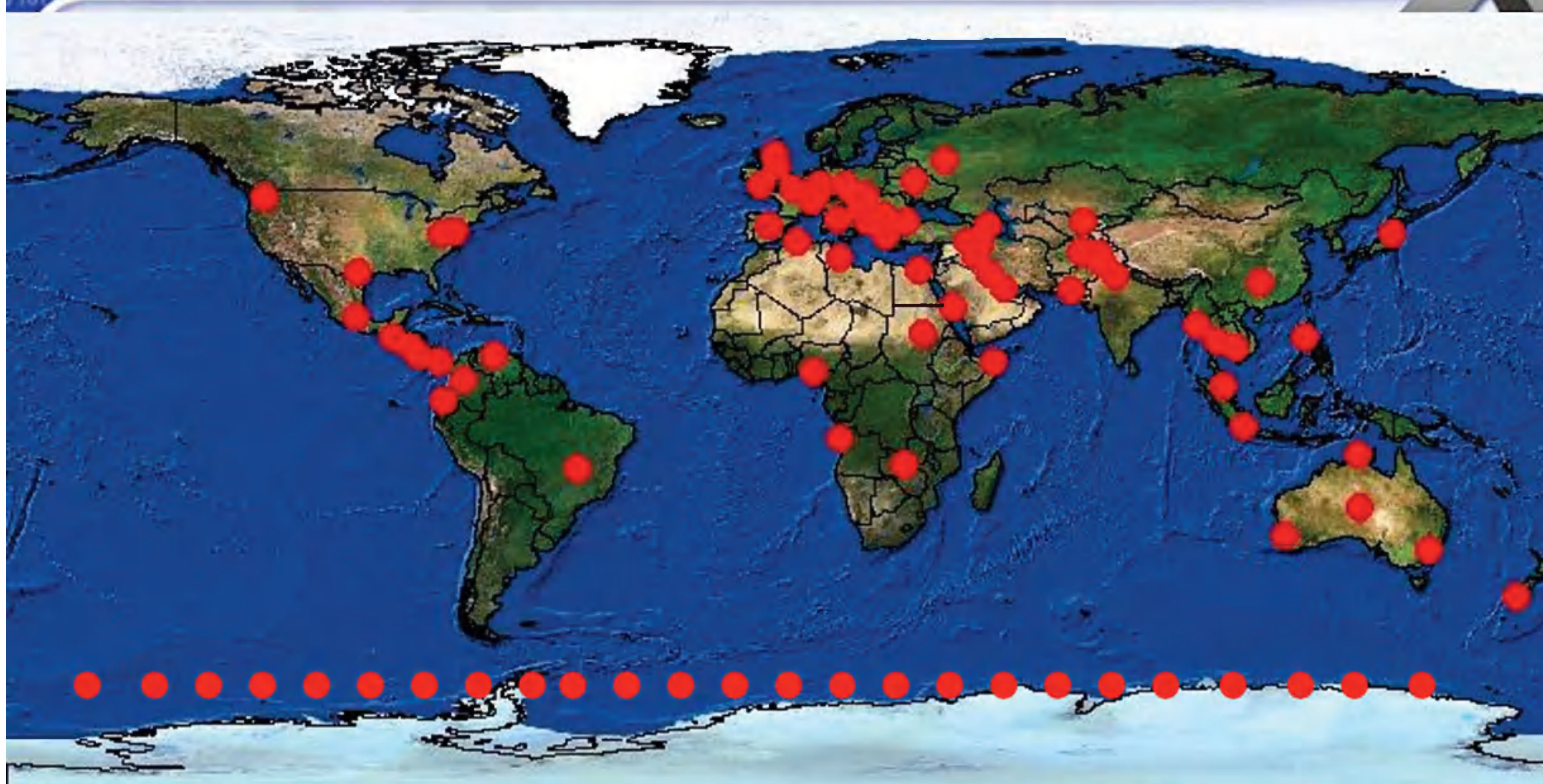




TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL



# Where is X-KEYSCORE?



Approximately 150 sites

Over 700 servers

TOP SECRET/ COMINT//REL TO USA, AUS, CAN, GBR NZL



**TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL**



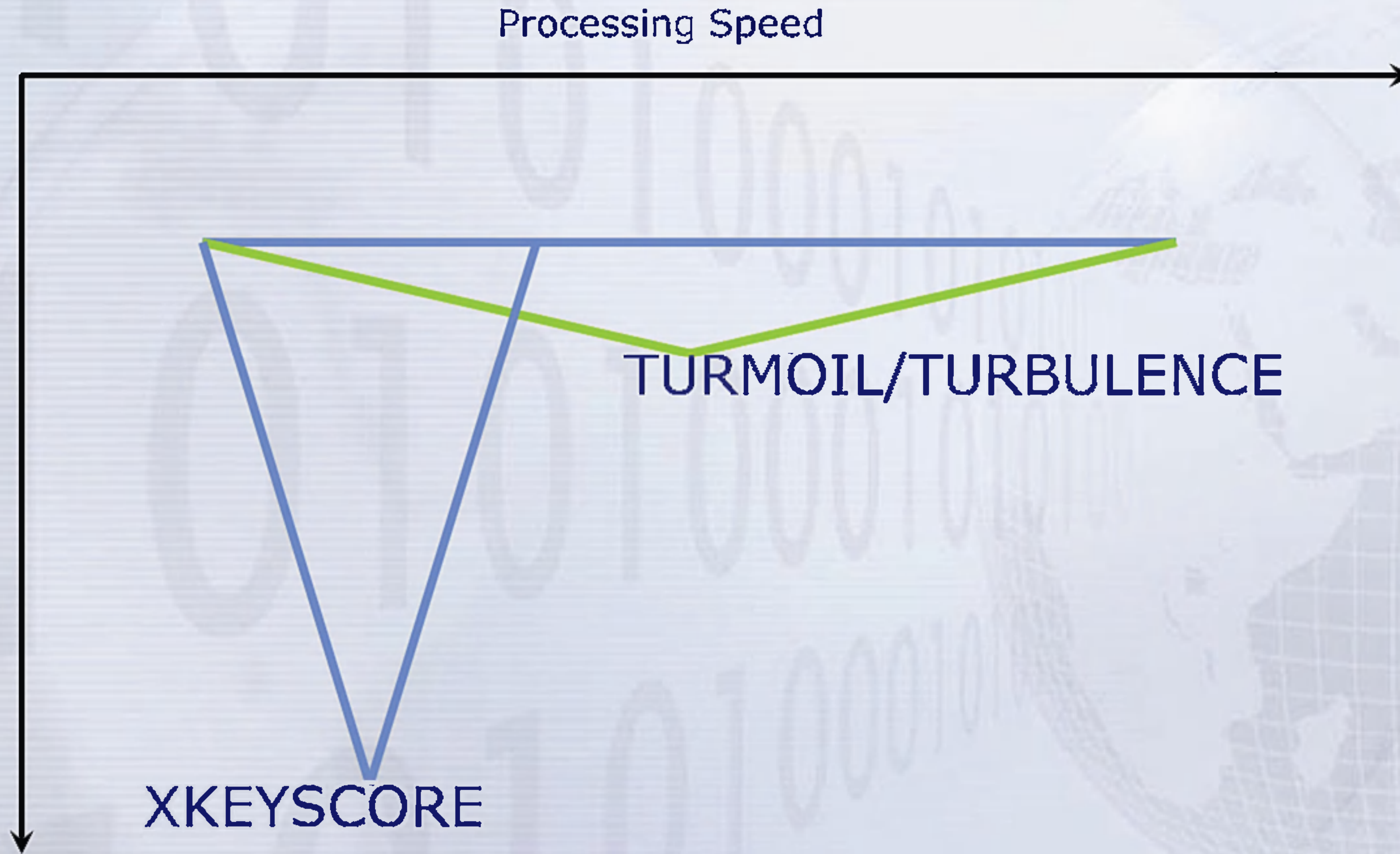
What is unique about  
XKEYSCORE?

**TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL**





# General Capability







# Why do shallow

- Can look at more data
- XKEYSCORE can also be configured to go shallow if the data rate is too high





# Why go deep

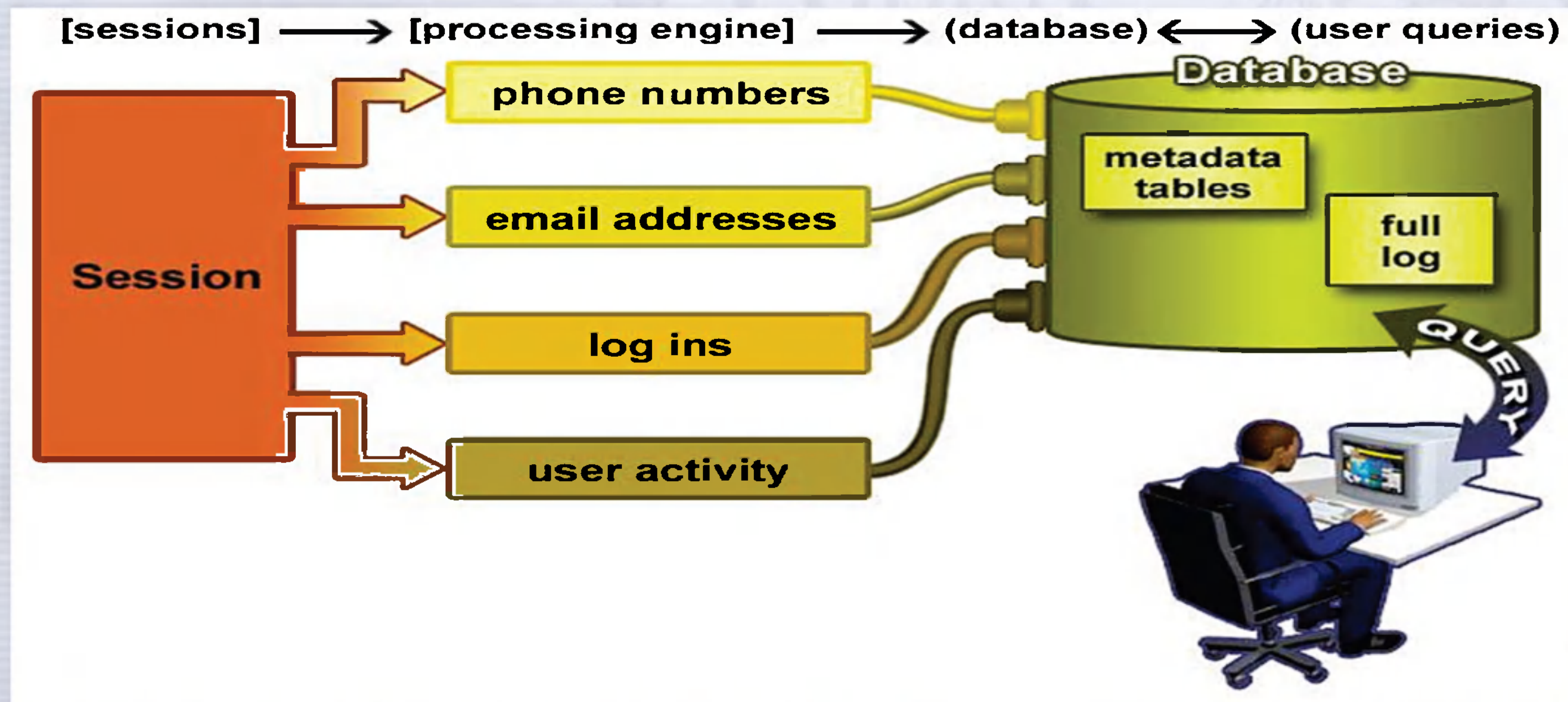
- Strong Selection itself give us only a very limited capability
- A large amount of time spent on the web is performing actions that are anonymous
- We can use this traffic to detect anomalies which can lead us to intelligence by itself, or strong selectors for traditional tasking





# What XKS does with the Sessions

Plug-ins extract and index metadata into tables







# Plug-ins

<b>Plug-in</b>	<b>DESCRIPTION</b>
E-mail Addresses	Indexes every E-mail address seen in a session by both username and domain
Extracted Files	Indexes every file seen in a session by both filename and extension
Full Log	Indexes every DNI session collected. Data is indexed by the standard N-tuple (IP, Port, Casenotation etc.)
HTTP Parser	Indexes the client-side HTTP traffic (examples to follow)
Phone Number	Indexes every phone number seen in a session (e.g. address book entries or signature block)
User Activity	Indexes the Webmail and Chat activity to include username, buddylist, machine specific cookies etc.





# What Can Be Stored?

- Anything you wish to extract
  - Choose your metadata
  - Customizable storage times
  - Ex: HTTP Parser

```
FM IP 58. [REDACTED] TO IP 64. [REDACTED]
GET /search?hl=en&q=islamabad&meta= HTTP/1.0
Accept: image/gif, image/x-bitmap, image/jpeg, image/pjpeg, application/vnd.ms-
application/msword, application/x-shockwave-flash, */*
Referer: http://www.google.com.pk/
Accept-Language: en-us
User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)
Host: www.google.com.pk
Cookie: PREF=ID=678100a34384e2f6:TU=1168503483:LM=1168503483:S=KKzZb3kPcw4vNxGt
Via: 1.0 proxy.[REDACTED]:8080 (squid/2.5.STABLE13)
X-Forwarded-For: 58.[REDACTED]
Cache-Control: max-age=259200
Connection: keep-alive
```

No username/strong selector



**TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL**



What can you do with  
XKEYSCORE?

**TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL**



# Finding Targets



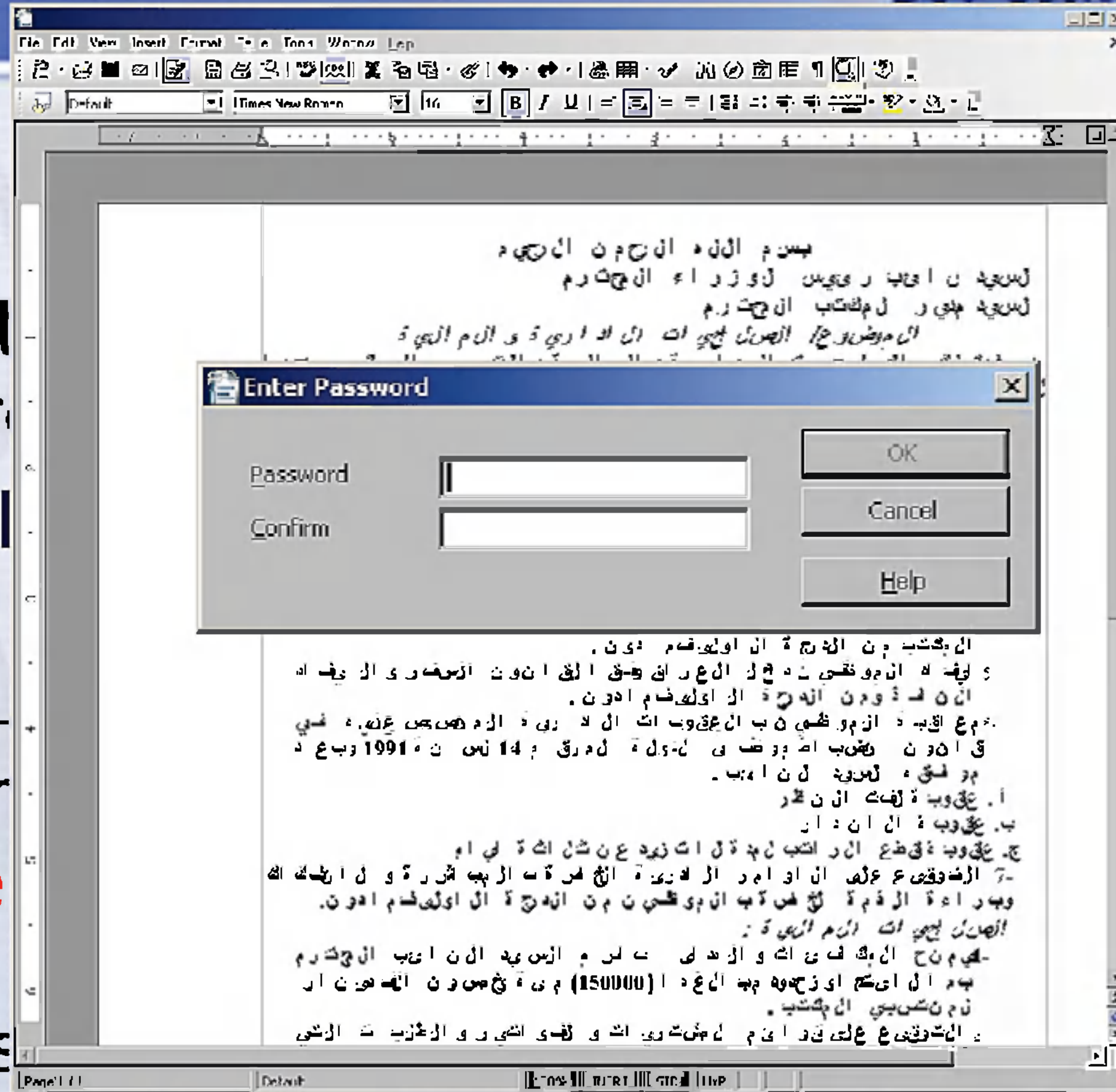
- How do I find a strong-selector for a known target?
- How do I find a cell of terrorists that has no connection to known strong-selectors?
- Answer: Look for anomalous events
  - E.g. Someone whose language is out of place for the region they are in
  - Someone who is using encryption
  - Someone searching the web for suspicious stuff



# Encryption



- Show me all documents from [redacted]
- Show me all [redacted]
  
- Once again – forwarding the [redacted]
- **No strong-security**
- Can perform query, then search [redacted] from site as required







# Technology Detection

- Show me all the VPN startups in country X, and give me the data so I can decrypt and discover the users
  - These events are easily browsable in XKEYSCORE
    - **No strong-selector**
  - XKEYSCORE extracts and stores authoring information for many major document types – can perform a retrospective survey to trace the document origin since metadata is typically kept for up to 30 days
  - **No other system** performs this on raw unselected bulk traffic, **data volumes prohibit forwarding**



# Persona Session Collection



- Traditionally triggered by a strong-selector event, but it doesn't have to be this way
- Reverse PSC – from anomalous event back to a strong selector. You cannot perform this kind of analysis when the data has first been strong selected.
- Tie in with Marina – allow PSC collection after the event





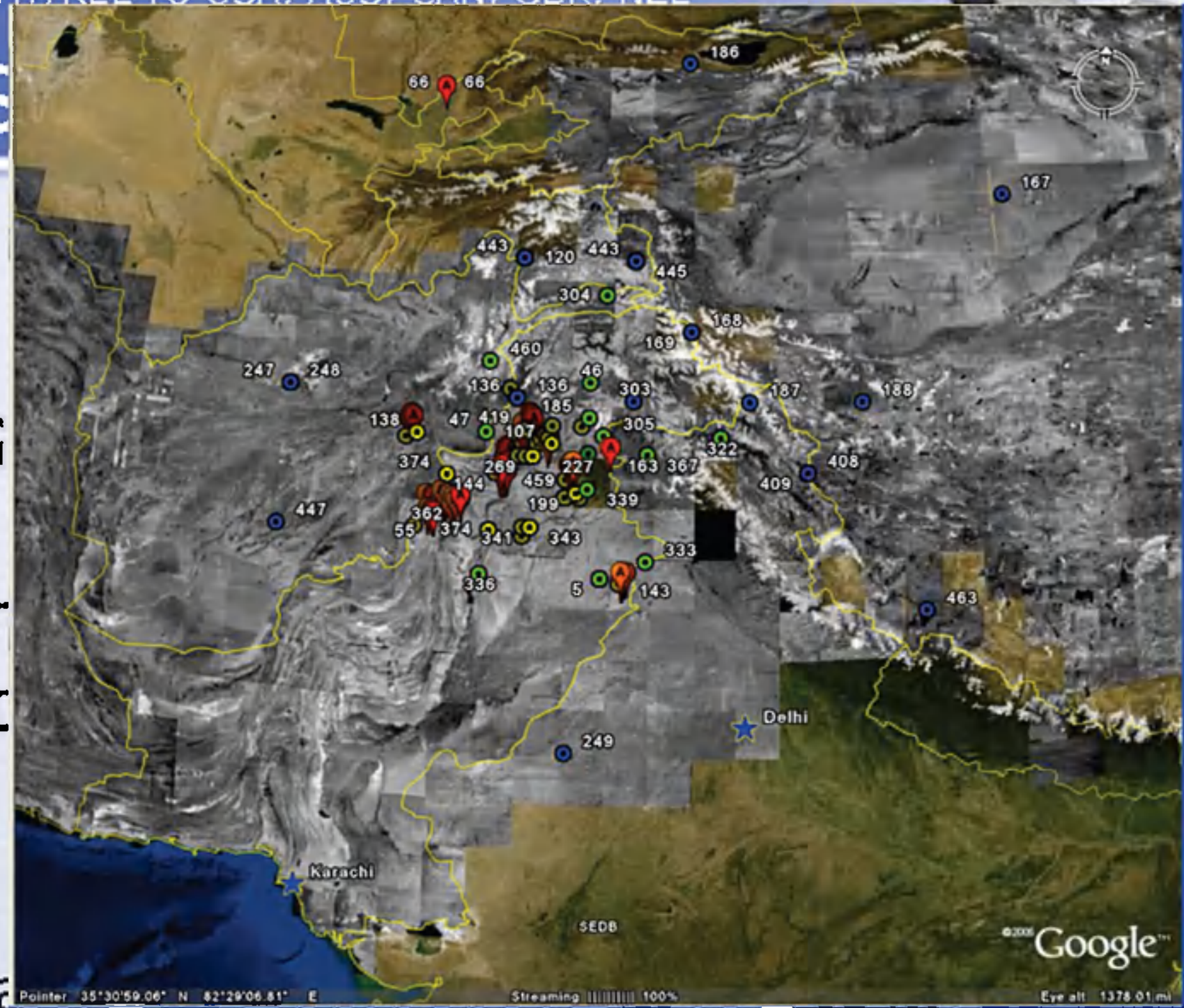
# Language Tracking

- My target speaks German but is in Pakistan – how can I find him?
  - XKEYSCORE's HTTP Activity plugin extracts and stores all HTML language tags which can then be searched
  - Not possible in any other system but XKEYSCORE, nor could it be –
    - **volumes are too great to forward**
    - **No strong-selector**



# Google Maps

- My target uses Google locations – can I determine his email web-searches – could be suspicious?
- XKEYSCORE extracts data including all web-based searches which can be **retrospectively** queried
- **No strong-selector**
- **Data volume too high to forward**





# Document Tracking



- I have had this document tracked and...

The screenshot displays a Microsoft Word document with a tracking window open. The document's title is "الصراف ايجي ك ال ادا فة و اليم الخبة doc". The tracking window shows the following metadata:

- Type:** MS Word document
- Location:** [Redacted]
- Size:** [Redacted]
- Created:** [Redacted]
- Modified:** [Redacted]
- Digitally signed:** [Redacted]
- Last printed:** [Redacted]
- Total editing time:** [Redacted]
- Revision number:** [Redacted]
- Apply user:**
- Template:** [Redacted]

The tracking window also includes a "Description" tab with the following fields:

- Title:** بصر الكبة ال صر ن ال صر
- Subject:** [Redacted]
- Keywords:** [Redacted]
- Comments:** [Redacted]

The document content is in Arabic and includes the following text:

عليه اسي  
199 وسع د  
وال اخصه الك  
م ادون.  
اي م ن ح ال زمك فة و ال د ل ا ب ل م ال سيه ال ن ايب ال ج ك ر م  
به ال ا ي س ا و ز ج و د ب س ن غ د (150000) م ية ب ص د : ن ف ه ر  
ل م ن س ب ي ال ب ك ت ب  
ع ال ت و ق ي ع ا ل ي ق و ا ي م ال م س ت ر ي ا ت و ا ل ف و ا ت ي ر و ن ط ر ب ت ن س ي





# Document Tracking

- All images are hashed in the metadata so that you can search for anyone who has received or transmitted this document.
- This is really useful for company logos.

*Biographical Data Sheet* / *استمارة معلومات شخصية*

Note: Fill the blocks in English & Type - Submit Electronically via e-mail / ملاحظة: ملاءمة الأسماء الانكليزية باستخدام برنامج الورد وارسلها بالبريد الالكتروني

First Name الاسم الأول	1 <sup>st</sup> Middle (father's) Name اسم الأب	2 <sup>nd</sup> Middle (Grandfather's) Name اسم الجد				
[Redacted]	[Redacted]	[Redacted]	Mother's Name اسم الام	Mother's Tribal Name لقب الام		
[Redacted]	[Redacted]	[Redacted]	[Redacted]	[Redacted]		
Birth Date تاريخ لميلاد		City/Country of Birth مكان لولادة الدولة - (المدينة للبلدة)	Nationality الجنسية	Race القومية	Religion الديانة	
DD/يوم	MM/شهر	YY/سنة	Iraq-kut-dor alnoalmean	Iraqi	Arab	Muslim
17	3	[Redacted]				
Gender (Male/Female) الجنس (ذكر/انثى)	Marital Status (Single/Married/Divorced) الحالة الزوجية (عزب/متزوج/مطلق)	Number of children عدد اطفال				
male	single	non				

Metadata XML snippets:

```

<lastauthor>[Redacted]</lastauthor>
<title>Biographical Data Sheet</title>
<language>Arabic</language>
<comment>
  <METADATA_ROW...>
  <Language>Arabic</Language>
  </DataItem>
  <Processing>
  <Name>[Redacted]</Name>
  </METADATA_ROW...>
  </document_metadata>
</xks_meta>

```

from [Redacted]



# Interests

- Show contents so I can
- New document dictionary info
- No
- Data
- Multiple dictionaries targeted at specific data types

Microsoft Excel - Najaf Civil Defense

PROACTIVE Communications, Inc. IC2N Closeout Document

Site Information

Site Name/City: [REDACTED] Engineer(s) Name: [REDACTED]

Function (FIS, PJOC, POL...): [REDACTED] Equipment Shipment Date: [REDACTED]

GPS Coordinates: [REDACTED] Equipment Delivered Date: [REDACTED]

Site Pre-Commissioned Date: [REDACTED] CLIN: 8

Site Commissioned Date: [REDACTED] Quantity of Power Strips Used: 4

Meters of CAT 5 Used: 300 Meters Container: Harding Case

PCI Delivered Equipment

Serial Number	NetModem Model
[REDACTED]	[REDACTED]

Ethernet Switch Serial Number	# of Ports / Make / Model
[REDACTED]	[REDACTED]

Serial Number	Model
[REDACTED]	[REDACTED]

LNH: [REDACTED]

DUC: [REDACTED]

Feed Horn Assembly: [REDACTED] POL: X

Voip Telephone MAC Address	Model	Telephone Number
[REDACTED]	[REDACTED]	[REDACTED]

Government Furnished Equipment

RouterGuard Serial Number	Model
[REDACTED]	[REDACTED]

Dell Model type / Serial Number	Monitor type / Serial Number
[REDACTED]	[REDACTED]

Customer Receipt

Keach



# TAO



- Show me all the exploitable machines in country X
  - Fingerprints from TAO are loaded into XKEYSCORE's application/fingerprintID engine
  - Data is tagged and databased
  - No strong-selector
  - Complex boolean tasking and regular expressions required



TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL



# XKEYSCORE Success Stories

TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL



**TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL**



Over 300 terrorists  
captured using  
intelligence generated  
from XKEYSCORE

**TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL**

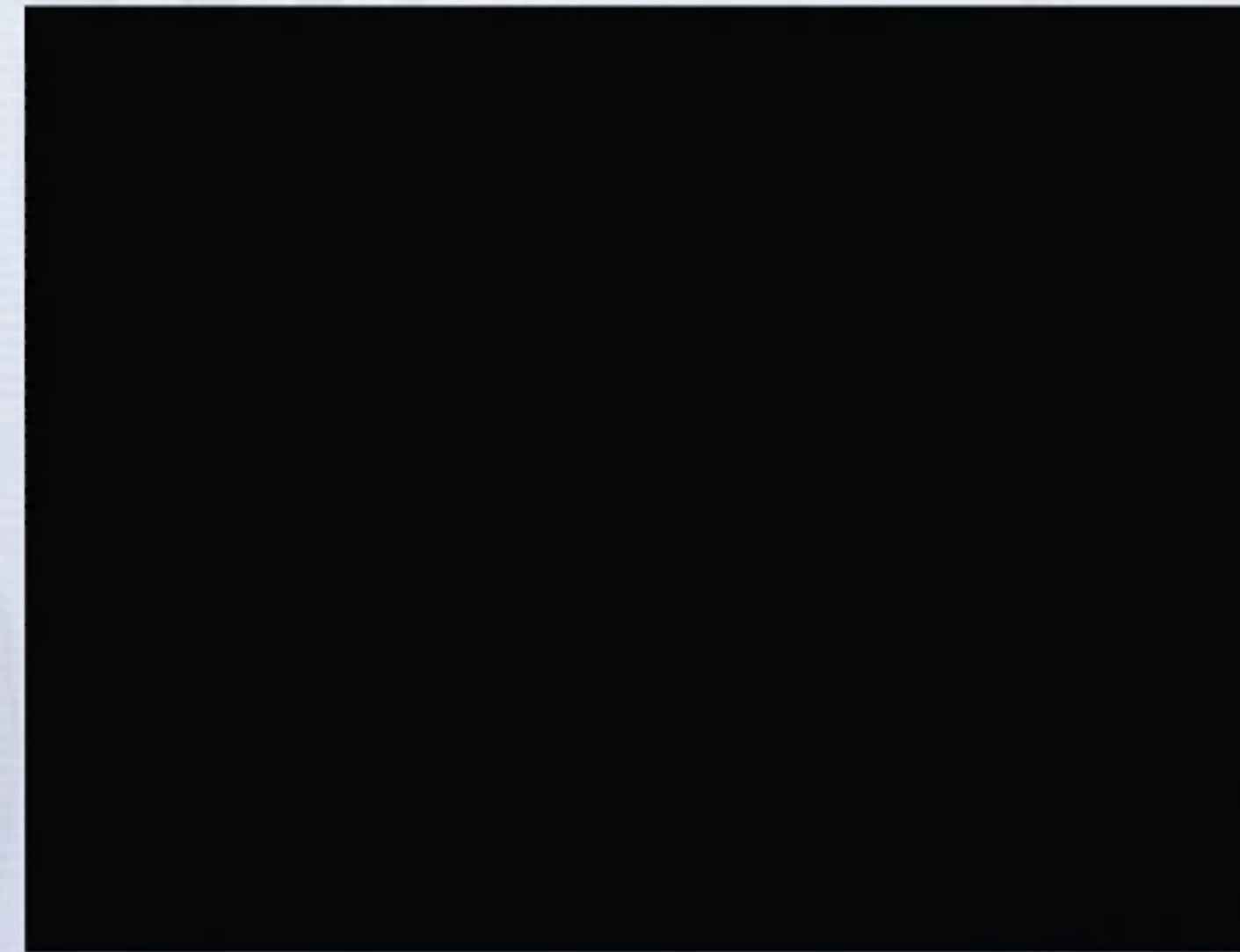
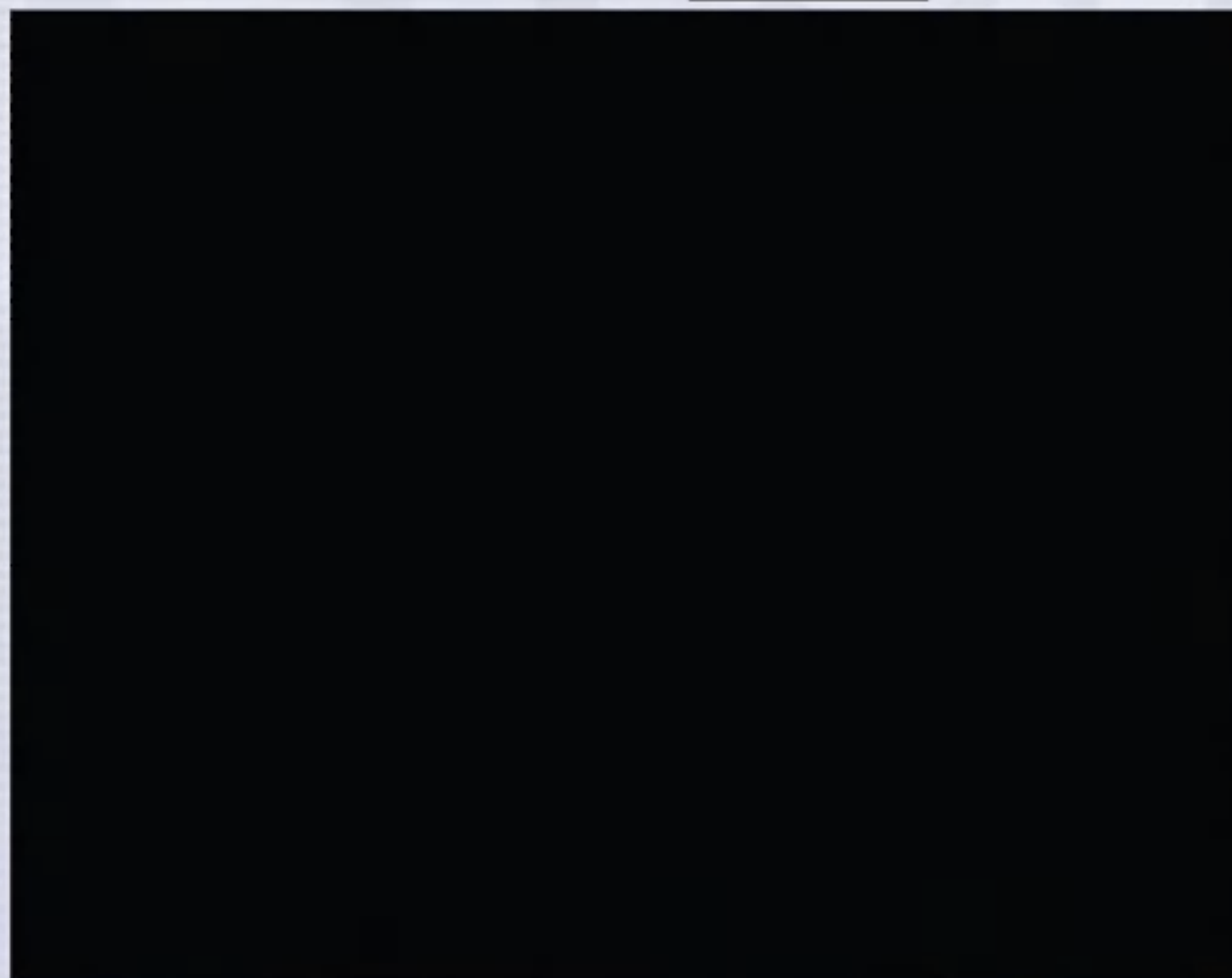




# XKEYSCORE and TRAFFICTHIEF

- Customer: CounterTerrorism (CT)
- Provides near real-time tips to TRAFFICTHIEF server in operations in coordination with coalition forces in Iraq 24 hours a day
- Currently producing hundreds of confirmed alerts per day on over 3000 user accounts

Afternoon of [REDACTED] 2004 – coalition detained individuals below:





# XKEYSCORE Success



## May 2006, WealthyCluster2 and X-KEYSCORE Installed at [REDACTED]

- Connected to Moonshine
- Enabled processing of wireless collection
- Enabled near-real-time tipping
- Enabled full-take SIGDEV

## Un-locatable cafés were geolocated:

- [REDACTED] – “A Goldmine”
- Four Other Cafés Being Developed

## Acquired important targets:

- NSA/Georgia Tips With Precise Locations
- JSOC Tools In New [REDACTED]
- Reacquired [REDACTED] Lost When Zarkanet Went Down

## Terrorists were captured:

- Members of the [REDACTED]
- Members of the [REDACTED]





# Innovation

- High Speed Selection
- Toolbar
- Integration with Marina
- GPRS, WLAN integration
- SSO CRDB
- Workflows
- Multi-level Dictionaries





# Future

- High speeds yet again (algorithmic and Cell Processor (R4))
- Better presentation
- Entity Extraction
- VoIP
- More networking protocols
- Additional metadata
  - Expand on google-earth capability
  - EXIF tags
  - Integration of all CES-AppProcs
- Easier to install/maintain/upgrade