

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



XKEYSCORE

25 Feb 2008

xkeyscore@nsa

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

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



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



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
1. 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

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



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

110 01 101 1001 1001 1001
10011 001 1001 11010 101 101 101
1001 1001 1001 11010 101 101 101
010 01 101 1001 1001 1001
0 10 01 101 1001 1001 1001
01010001 1001 1001 1001 1001 1001
0 101010 10 01 10011 001 1001

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



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

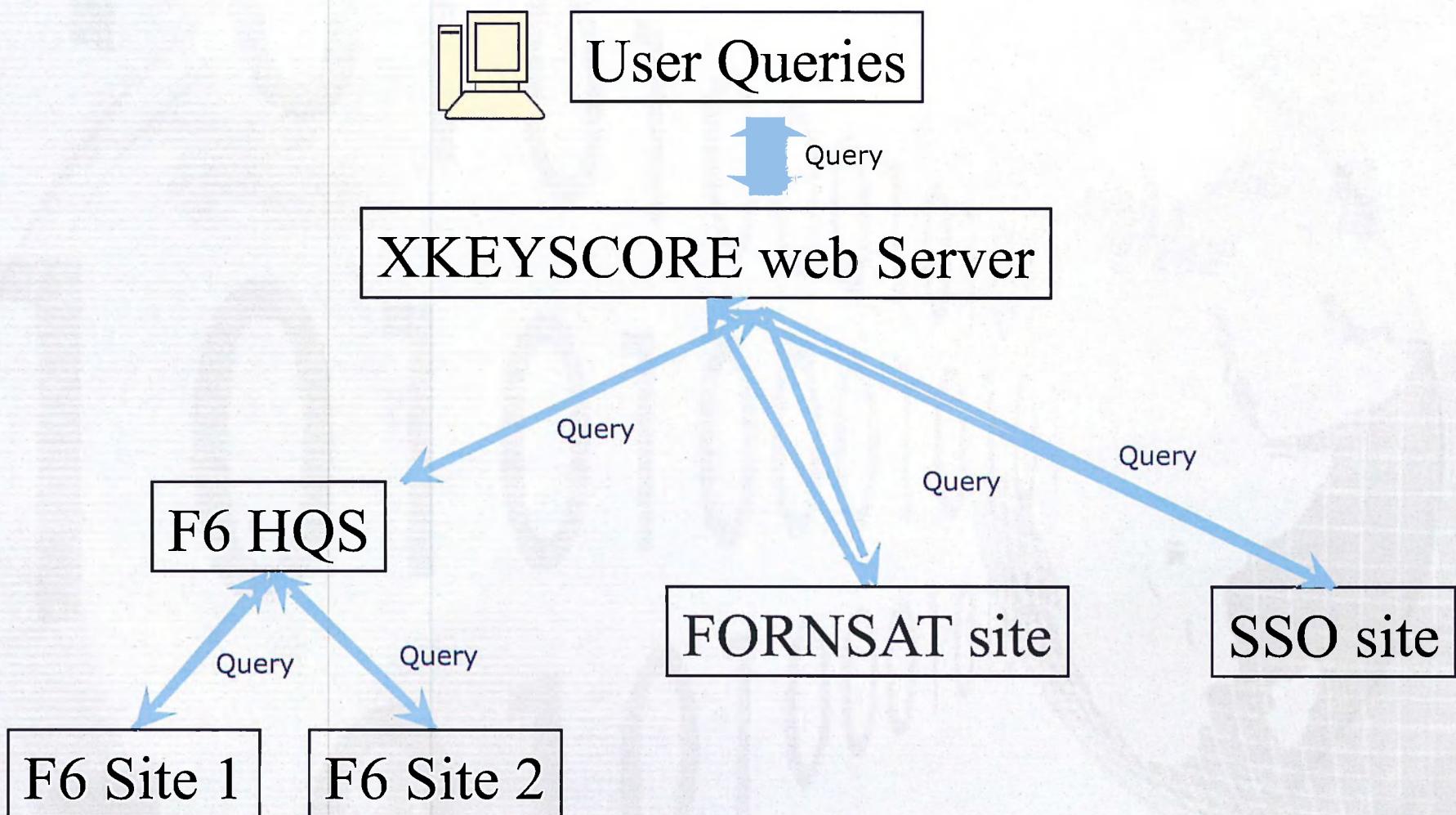
TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL

100 01 101 1001 1001 1001
1001 001 1001 101 101 101
100 1001 1001 11010101 1010101
010 01 101 1001 1010 101
0 10 01 101 1001 1 001 1010101
101 010001 10101 10 10 101 01010101
0 01 101010 10 0 101 1001

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



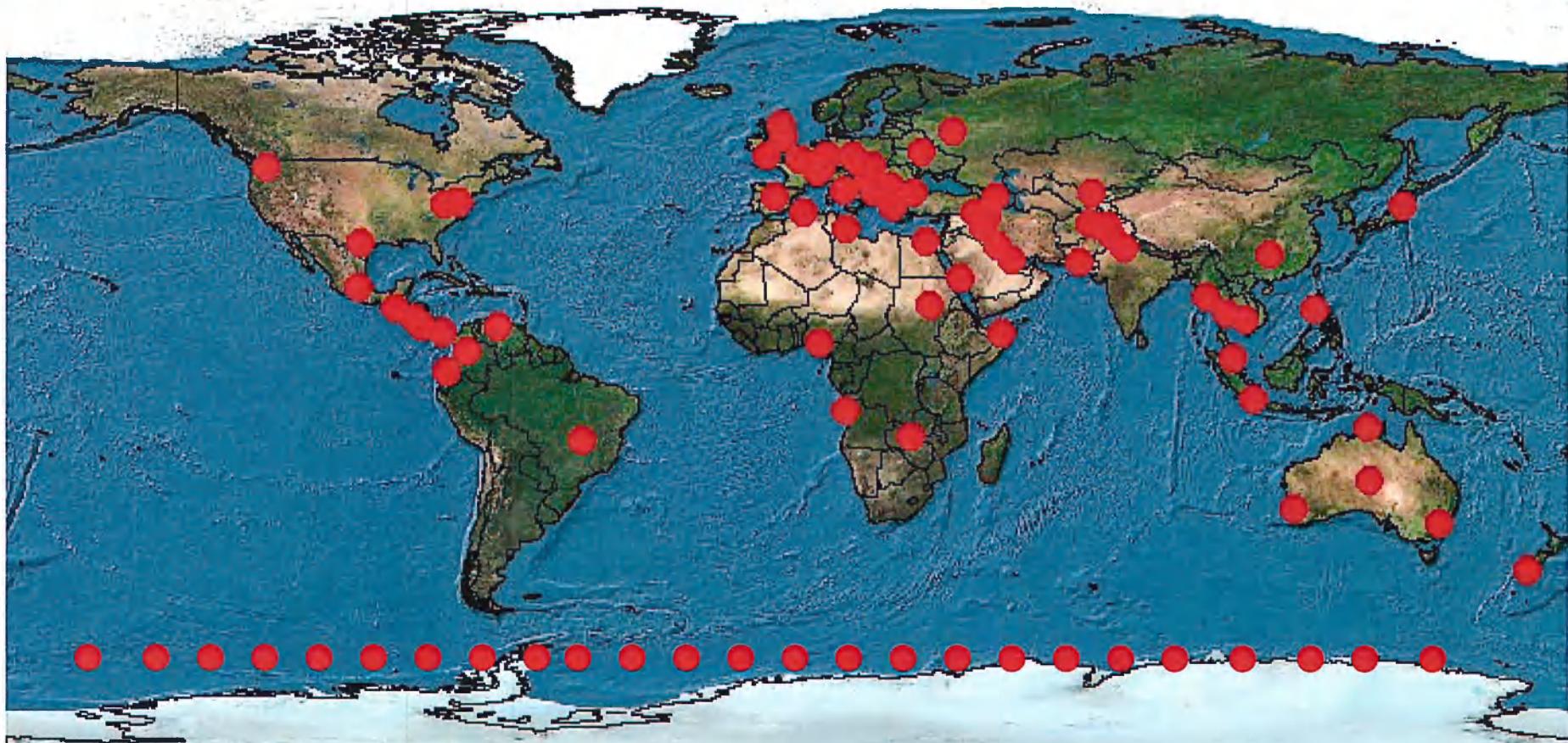
Query Hierarchy



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

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

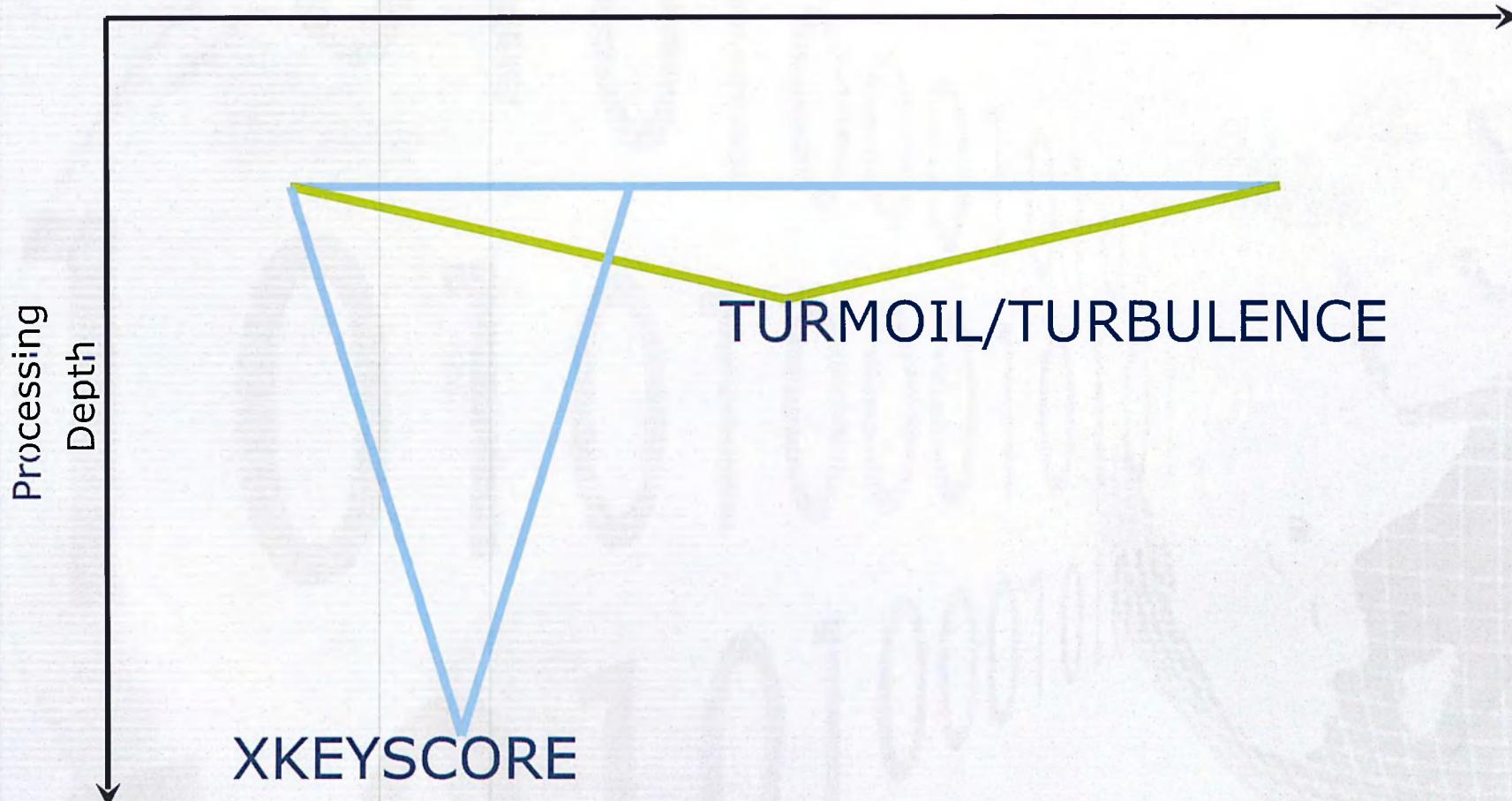
1010 01 101 1001 1001 1001
1001 1 001 1001 11010101
1001 1001 1001 11010101
010 01 101 1001 101010101
010 01 101 101 101010101
0101010001 101010101010101
01 101010 10 101010101010101

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



General Capability

Processing Speed



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

Why do shallow



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



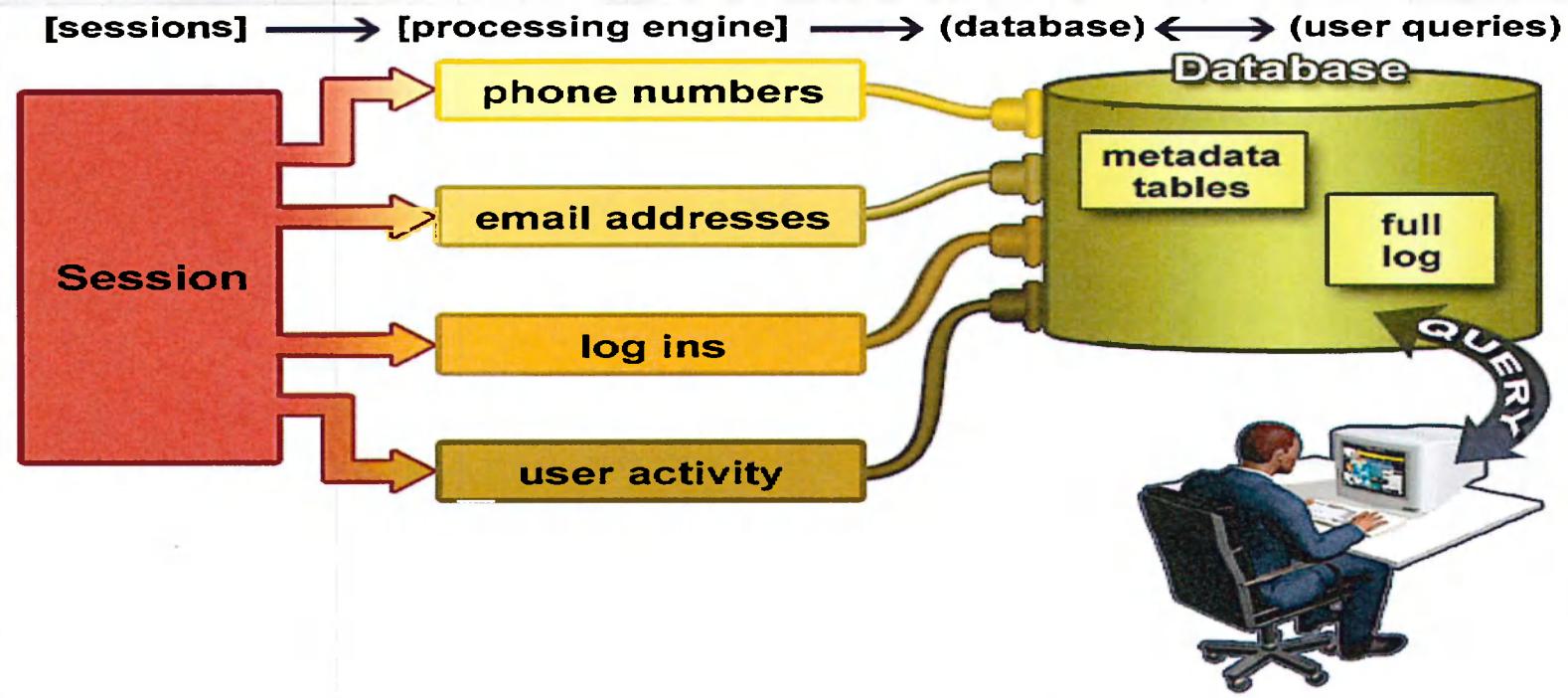
Why go deep

- 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

Plug-in	DESCRIPTION
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

[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
[REDACTED]
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

1001 101 1001 1001 1001
10011 001 1001 1001 1001
1001 1001 1001 1001 1001
01001 1001 1001 1001 1001
0100 1001 1001 1001 1001
0101 010001 1001 1001 1001
01 101010 10 01 101 1001 1001

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

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



Encryption

- Show me all the encrypted word documents from Iran
- Show me all PGP usage in Iran
 - Once again – **data volume too high** so forwarding these back is not possible
 - **No strong-selector**
 - Can perform this kind of retrospective query, then simply pull content of interest from site as required

10 01 101 001 1001 1001
1001 1 001 1001 101010101
1001 1001 1001 11010101
010 01 101 11010101
0 10 01 101 101010101
0101 010001 1010 101010101
0 01 101010 101010101

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



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

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

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

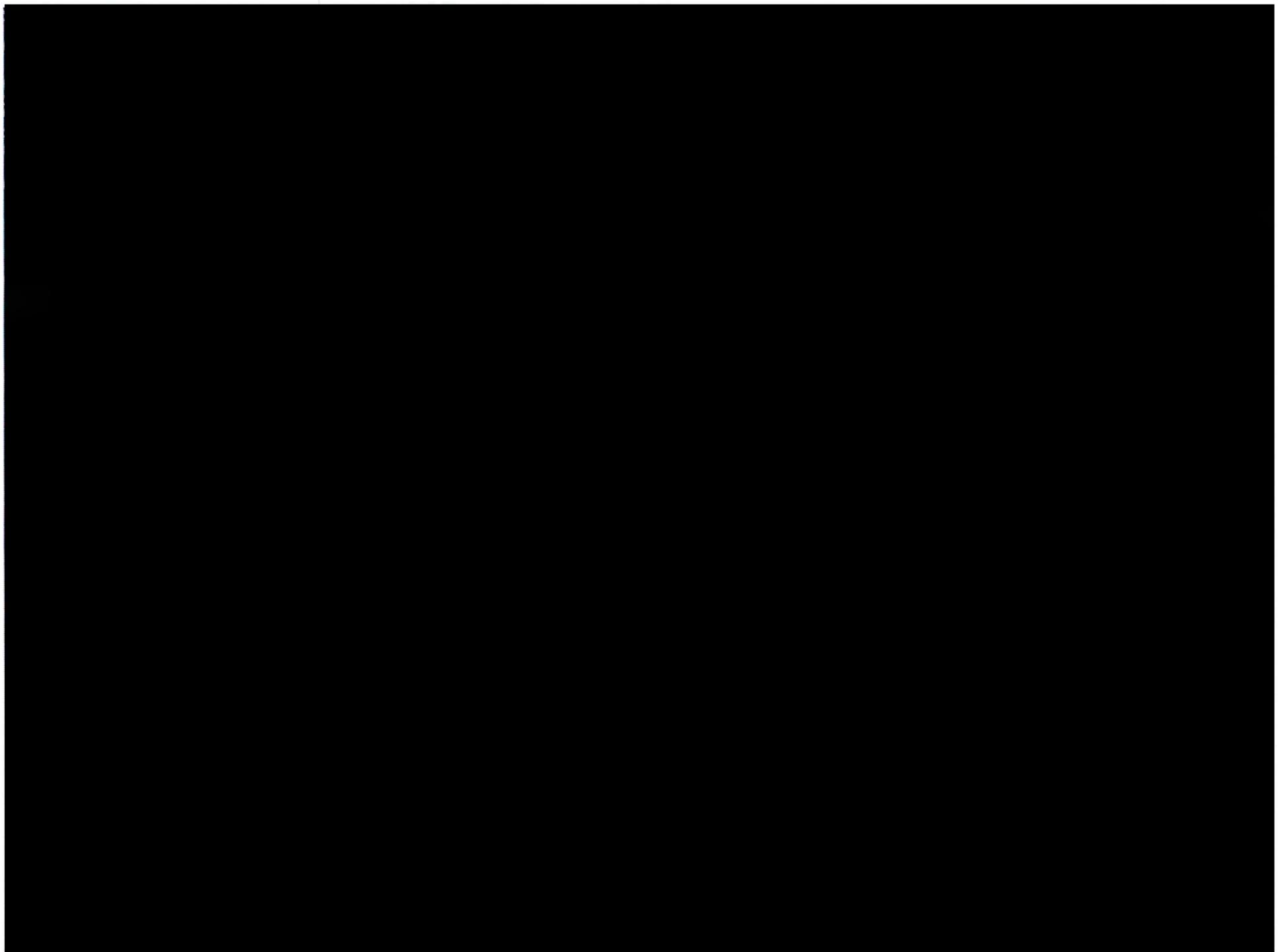


- My target uses Google Maps to scope target locations – can I use this information to determine his email address? What about the web-searches – do any stand out and look suspicious?
 - XKEYSCORE extracts and databases these events including all web-based searches which can be **retrospectively** queried
 - No strong-selector
 - Data volume too high to forward



Document Tracking

- I have a Jihadist document that has been passed around through numerous people, who wrote this and where were they?





Interesting Document Discovery

- Show me all the Microsoft Excel spreadsheets containing MAC addresses coming out of Iraq so I can perform network mapping
 - New extractor allows different dictionaries to run on document/email bodies – these more complex dictionaries can generate and database this information
 - No strong-selector
 - Data volume is high
 - Multiple dictionaries targeted at specific data types



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



- 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

Discovery of new target web services



- New web services every day
 - Scanning content for the userid rather than performing strong selection means we may detect activity for applications we previously had no idea about

Entity Extraction



- Have technology (thanks to R6) – for English, Arabic and Chinese
- Allow queries like:
- Show me all the word documents with references to IAE
- Show me all documents that reference Osama Bin Laden
- Will allow a 'show me more like this' capability

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



XKEYSCORE Success Stories

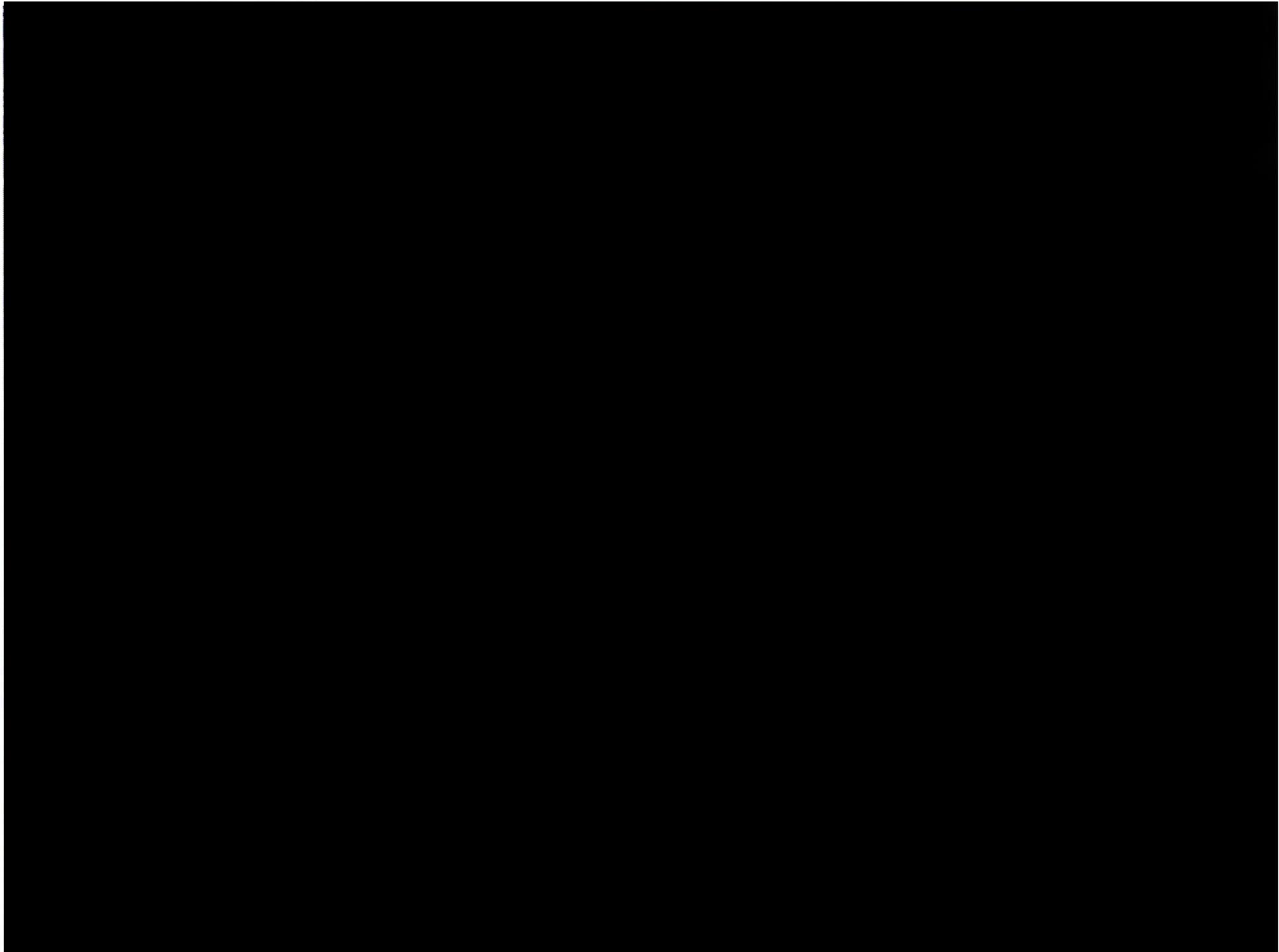
TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL

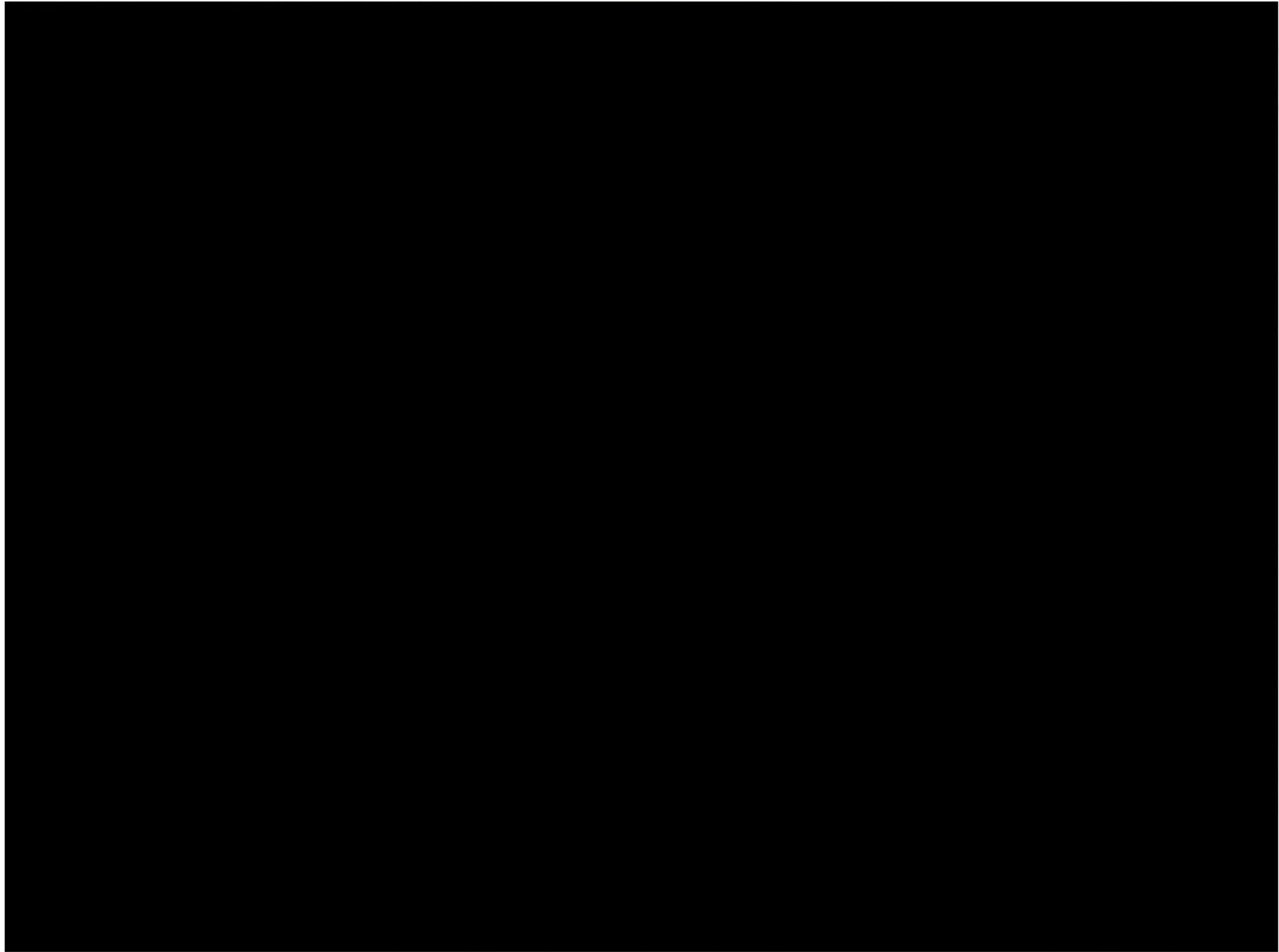
TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL



Over 300 terrorists
captured using
intelligence generated
from XKEYSCORE

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







Innovation

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

Future

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



- 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

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