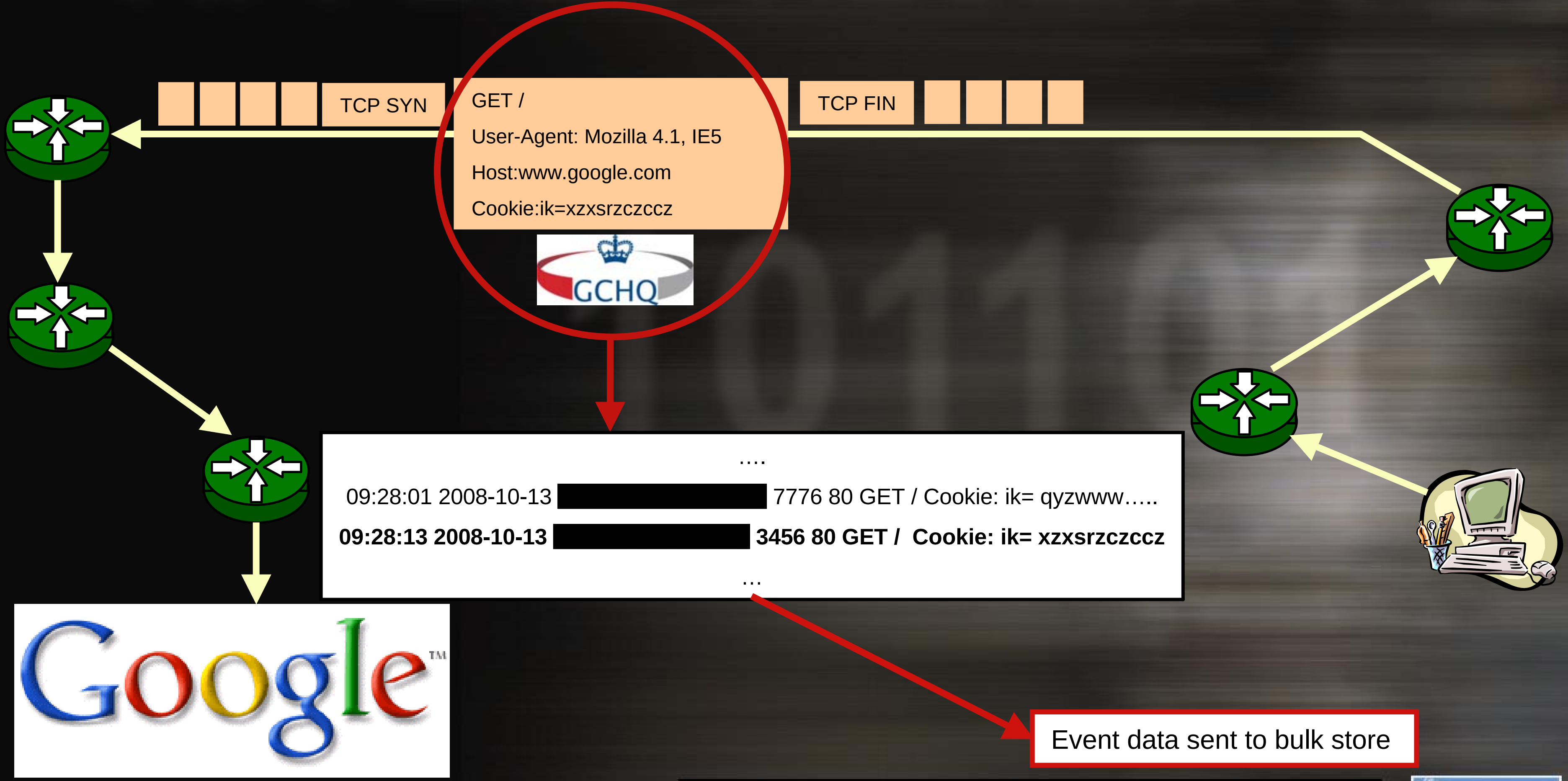
Target Detection Identifiers

[redacted]

March 2009

High-Speed Internet Processing



....

09:28:01 2008-10-13 [redacted] 7776 80 GET / Cookie: ik= qyzwww…..

09:28:13 2008-10-13 [redacted] 3456 80 GET / Cookie: ik= xzxsrzczccz

...

Event data sent to bulk store

High-Speed Internet Processing

- Bulk events key to SIGINT success on Internet
- Event types that are valuable for Intelligence change (quickly)
  - 2000 SMTP/POP3
  - 2001 Webmail
  - …
  - 2007 vBulletin
  - 2008 Social Networks,…,?
- GCHQ’s Applied Research are pioneering ways of dealing with this:
  - Presence Events (TDI)
  - Very large scale high speed flat file storage to bulk store TDIs
  - Just enough data marts

IP Packet Information

- Many possible types of information
- Many techniques available
- HTTP Get requests dominate cutting edge techniques
- To get Intelligence value Information must relate to a person or device… a TDI

TDI ...?

;

TDI ...?

TDI

Target

Detection

Identifier

;

TDI

Target

Detection

Identifier

;

Who

When

Where

(doing) What

TDI

Target

Detection

Identifier

;

Who

When

Where

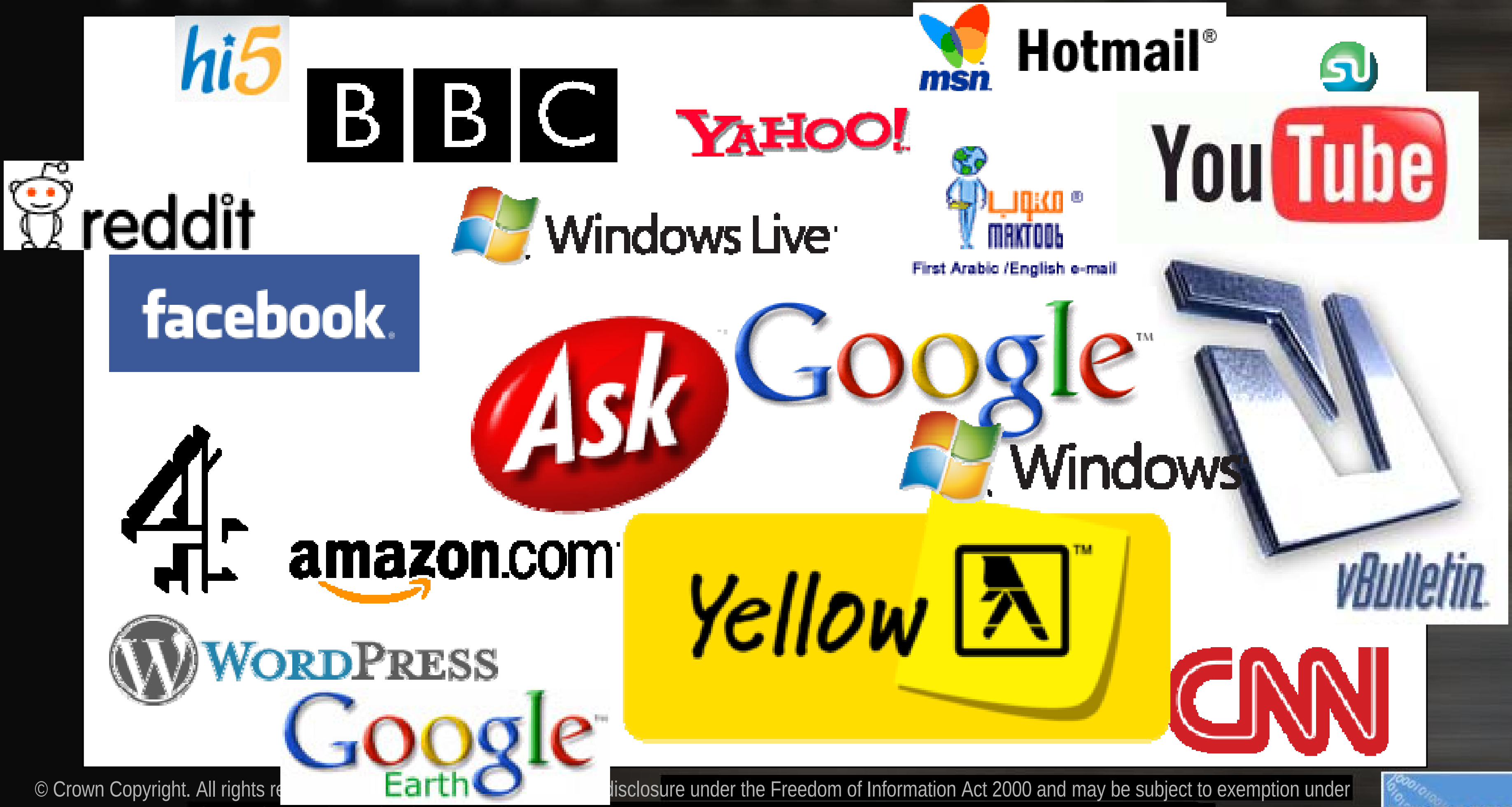
(doing) What

Fundamental atom of the Internet age.

Target Detection Identifiers

- DEFINITION
  - TDIs are **definite** indicators of presence, that are **unique** and **persistent** for a user/machine.
- Built on the familiar
  - Telephony +44 [redacted] – international phone code
  - Signalling tells us this phone user is ‘online’
- Target Detection Identifiers
  - Started with the Internet, mobile networks too.
  - TDI is a ‘SIGINT standardised code’.
  - Not a standard managed by the ITU/ETSI.
  - Extraction from packets much more complex.

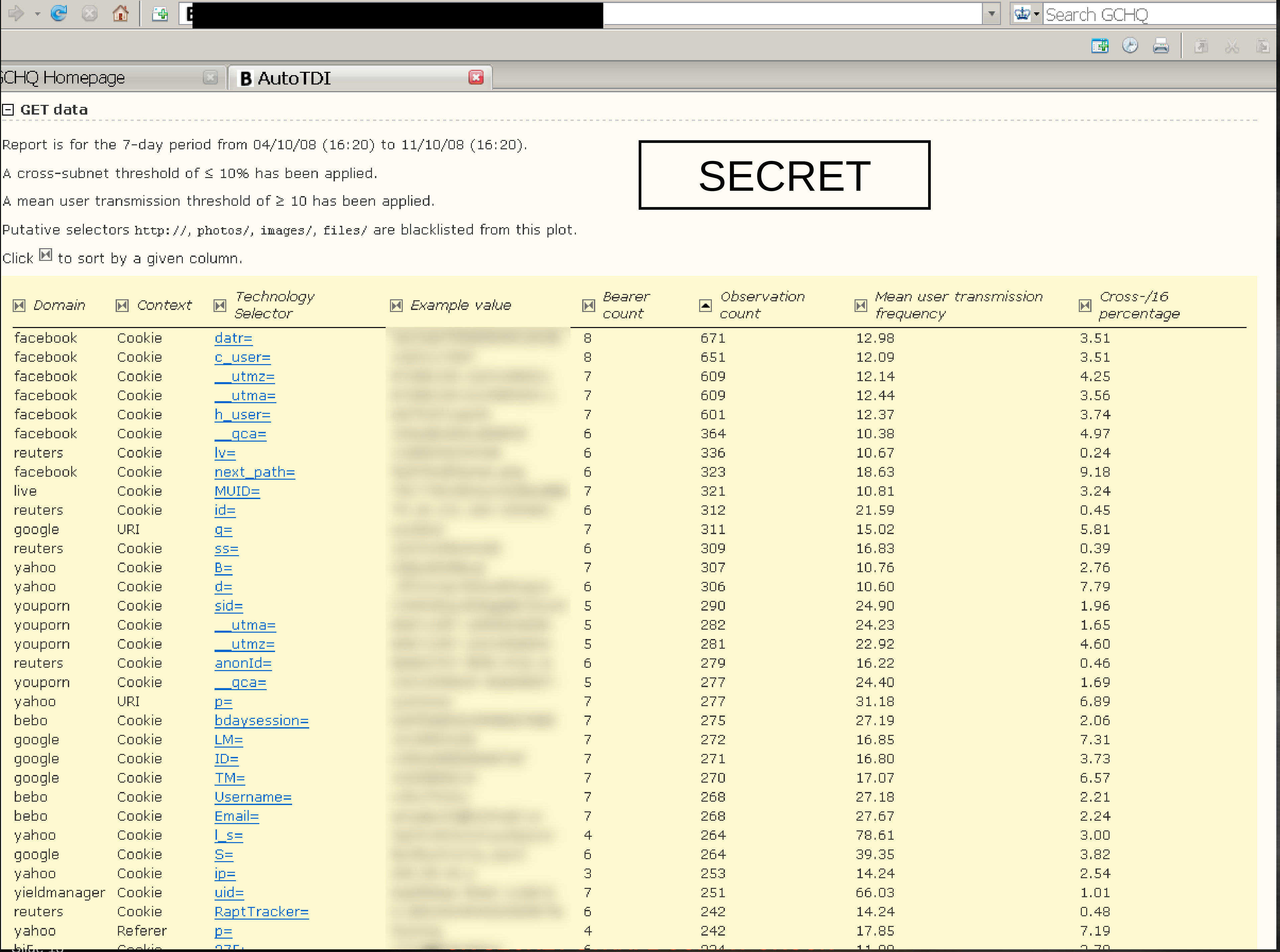
TDI sources


hi5
BBC
YAHOO!
msn
Hotmail®
YouTube
reddit
Windows Live
مكتوب
MAKTOOB
First Arabic /English e-mail
facebook
Ask
Google™
Windows
vBulletin
4
amazon.com
Yellow
CNN
WordPress
Google Earth

Target Detection Identifiers

- 70 distinct TDI types discovered.
- 2500 TDIs/sec (GET, de-duplicated)
- => 200 Million per day per 10Gbps
- De-dupe rate ???
- Cost – 250 hours per TDI
- Automated discovery prototype

TDI Type	TDI Location	User/Machine
Yahoo-Y-Cookie	Cookie	User
Yahoo-B-Cookie	Coookie	Machine
Google-IK	Request-URI	User
Paltalk-Nickname	Request-URI	User
MS-MUID-Cookie	Cookie	Machine
Google-SID-Cookie	Cookie	Machine
Maktoob-MEUser-Cookie	Cookie	User
Orkut-PREFID-Cookie	Cookie	User
Cloob-Username	Cookie	User



SECRET

⊟ GET data

Report is for the 7-day period from 04/10/08 (16:20) to 11/10/08 (16:20).

A cross-subnet threshold of ≤ 10% has been applied.

A mean user transmission threshold of ≥ 10 has been applied.

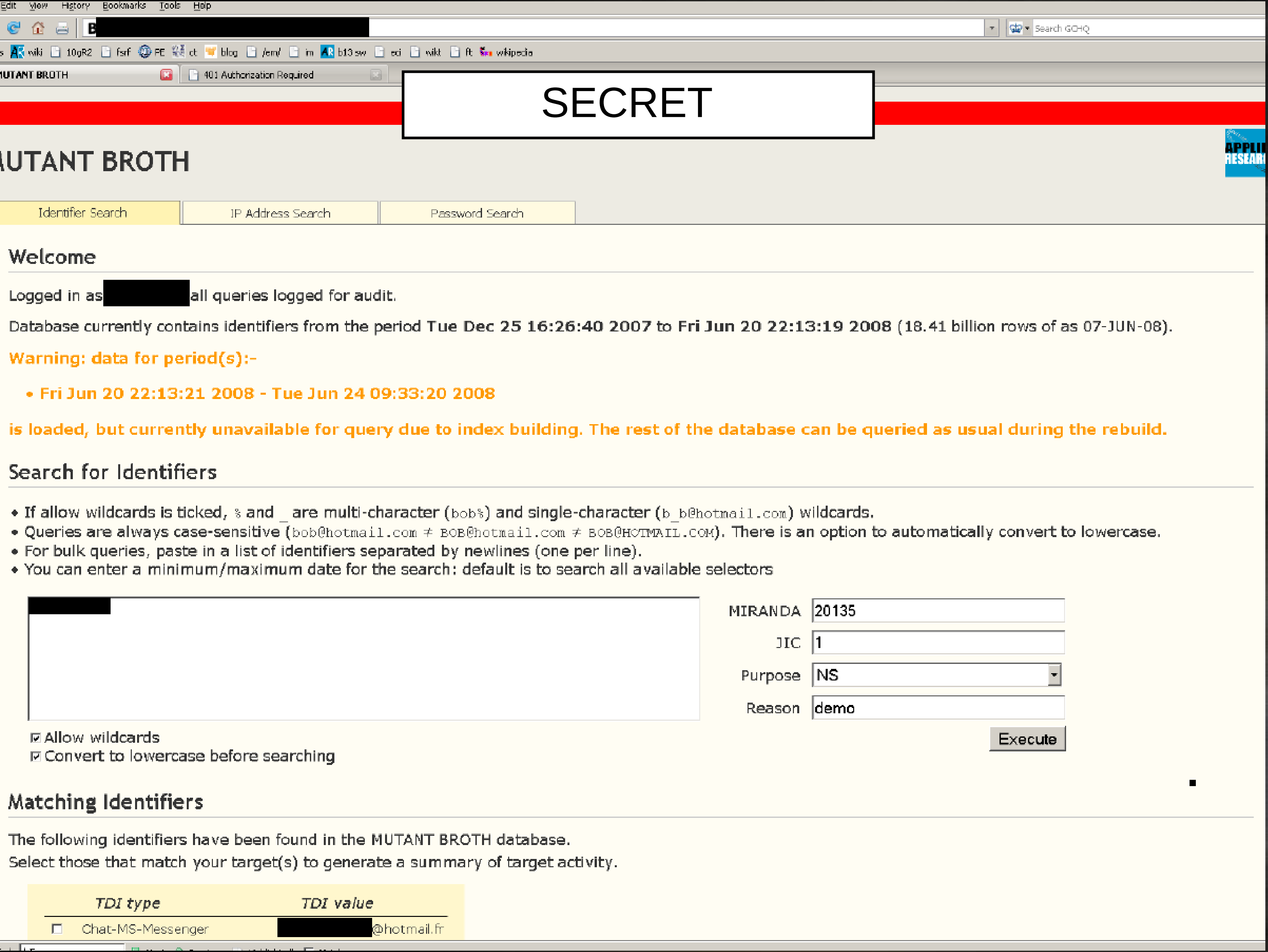
Putative selectors `http://`, `photos/`, `images/`, `files/` are blacklisted from this plot.

Click ⧓ to sort by a given column.

⧓ Domain	⧓ Context	⧓ Technology Selector	⧓ Example value	⧓ Bearer count	▲ Observation count	⧓ Mean user transmission frequency	⧓ Cross-/16 percentage
facebook	Cookie	datr=	[redacted]	8	671	12.98	3.51
facebook	Cookie	c_user=	[redacted]	8	651	12.09	3.51
facebook	Cookie	__utmz=	[redacted]	7	609	12.14	4.25
facebook	Cookie	__utma=	[redacted]	7	609	12.44	3.56
facebook	Cookie	h_user=	[redacted]	7	601	12.37	3.74
facebook	Cookie	__qca=	[redacted]	6	364	10.38	4.97
reuters	Cookie	lv=	[redacted]	6	336	10.67	0.24
facebook	Cookie	next_path=	[redacted]	6	323	18.63	9.18
live	Cookie	MUID=	[redacted]	7	321	10.81	3.24
reuters	Cookie	id=	[redacted]	6	312	21.59	0.45
google	URI	q=	[redacted]	7	311	15.02	5.81
reuters	Cookie	ss=	[redacted]	6	309	16.83	0.39
yahoo	Cookie	B=	[redacted]	7	307	10.76	2.76
yahoo	Cookie	d=	[redacted]	6	306	10.60	7.79
youporn	Cookie	sid=	[redacted]	5	290	24.90	1.96
youporn	Cookie	__utma=	[redacted]	5	282	24.23	1.65
youporn	Cookie	__utmz=	[redacted]	5	281	22.92	4.60
reuters	Cookie	anonId=	[redacted]	6	279	16.22	0.46
youporn	Cookie	__qca=	[redacted]	5	277	24.40	1.69
yahoo	URI	p=	[redacted]	7	277	31.18	6.89
bebo	Cookie	bdaysession=	[redacted]	7	275	27.19	2.06
google	Cookie	LM=	[redacted]	7	272	16.85	7.31
google	Cookie	ID=	[redacted]	7	271	16.80	3.73
google	Cookie	TM=	[redacted]	7	270	17.07	6.57
bebo	Cookie	Username=	[redacted]	7	268	27.18	2.21
bebo	Cookie	Email=	[redacted]	7	268	27.67	2.24
yahoo	Cookie	l_s=	[redacted]	4	264	78.61	3.00
google	Cookie	S=	[redacted]	6	264	39.35	3.82
yahoo	Cookie	ip=	[redacted]	3	253	14.24	2.54
yieldmanager	Cookie	uid=	[redacted]	7	251	66.03	1.01
reuters	Cookie	RaptTracker=	[redacted]	6	242	14.24	0.48
yahoo	Referer	p=	[redacted]	4	242	17.85	7.19

TDI Applications

- Bulk store of all TDIs seen in last 6 months [MUTANT BROTH]
- Bulk store TDI correlations (6 months) [AUTO ASSOC]
- Bulk store TDI <-> website correlations (6 months) [KARMA POLICE]
- Bulk store TDI vBulletin activity [INFINITE MONKEYS]
- Bulk store TDI Social Networking Site activity [SOCIAL ANIMAL]
- Bulk store web search requests [MEMORY HOLE]
- Bulk store Google Earth requests [MARBLED GECKO]
- Bulk store of Host-Referer references [HRMAP]


Edit View History Bookmarks Tools Help
B[redacted]
Search GCHQ
wiki 10gR2 fsrf PE ct blog /em/ im b13 sw edi wikt ft wikipedia
IUTANT BROTH
401 Authorization Required
APPLIE
RESEAR
Identifier Search
IP Address Search
Password Search

SECRET

IUTANT BROTH

Welcome

Logged in as [redacted] all queries logged for audit.

Database currently contains identifiers from the period **Tue Dec 25 16:26:40 2007** to **Fri Jun 20 22:13:19 2008** (18.41 billion rows of as 07-JUN-08).

Warning: data for period(s):-

- **Fri Jun 20 22:13:21 2008 - Tue Jun 24 09:33:20 2008**

is loaded, but currently unavailable for query due to index building. The rest of the database can be queried as usual during the rebuild.

Search for Identifiers

- If allow wildcards is ticked, `%` and `_` are multi-character (`bob%`) and single-character (`b_b@hotmail.com`) wildcards.
- Queries are always case-sensitive (`bob@hotmail.com` ≠ `BOB@hotmail.com` ≠ `BOB@HOTMAIL.COM`). There is an option to automatically convert to lowercase.
- For bulk queries, paste in a list of identifiers separated by newlines (one per line).
- You can enter a minimum/maximum date for the search: default is to search all available selectors

[redacted]

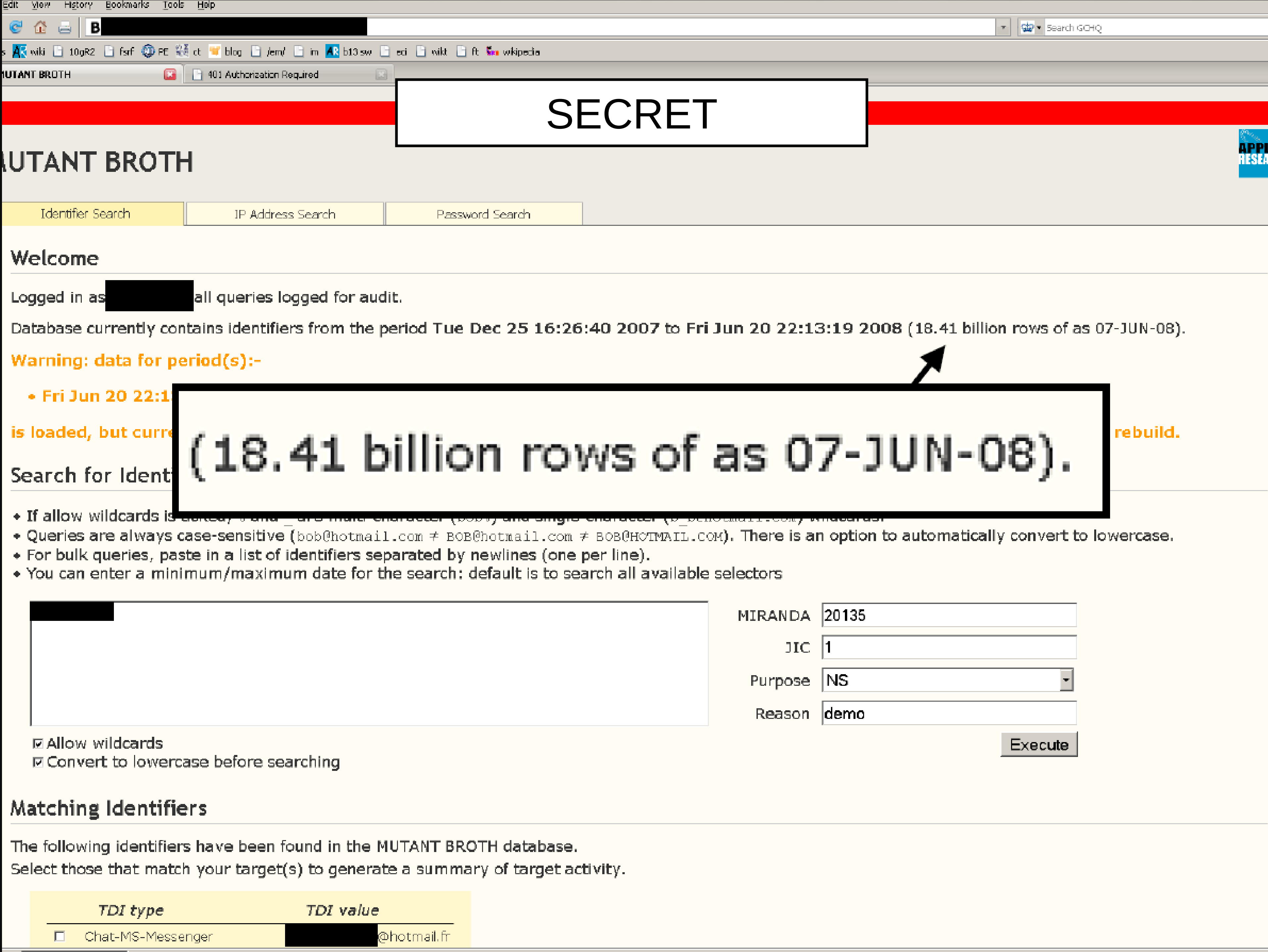
MIRANDA	20135
JIC	1
Purpose	NS
Reason	demo

☑ Allow wildcards
☑ Convert to lowercase before searching

Matching Identifiers

The following identifiers have been found in the MUTANT BROTH database.
Select those that match your target(s) to generate a summary of target activity.

	TDI type	*TDI value*
☐	Chat-MS-Messenger	[redacted]@hotmail.fr


Edit View History Bookmarks Tools Help
B[redacted]
Search GCHQ
wiki 10gR2 fsrf PE ct blog /em/ im b13 sw eci wikt ft wikipedia
UTANT BROTH
401 Authorization Required

SECRET

UTANT BROTH

Identifier Search
IP Address Search
Password Search

Welcome

Logged in as [redacted] all queries logged for audit.

Database currently contains identifiers from the period Tue Dec 25 16:26:40 2007 to Fri Jun 20 22:13:19 2008 (18.41 billion rows of as 07-JUN-08).

Warning: data for period(s):-

- Fri Jun 20 22:1

is loaded, but curr rebuild.

(18.41 billion rows of as 07-JUN-08).

Search for Ident

- If allow wildcards is
- Queries are always case-sensitive (bob@hotmail.com ≠ BOB@hotmail.com ≠ BOB@HOTMAIL.COM). There is an option to automatically convert to lowercase.
- For bulk queries, paste in a list of identifiers separated by newlines (one per line).
- You can enter a minimum/maximum date for the search: default is to search all available selectors

[redacted]

MIRANDA	20135
JIC	1
Purpose	NS
Reason	demo

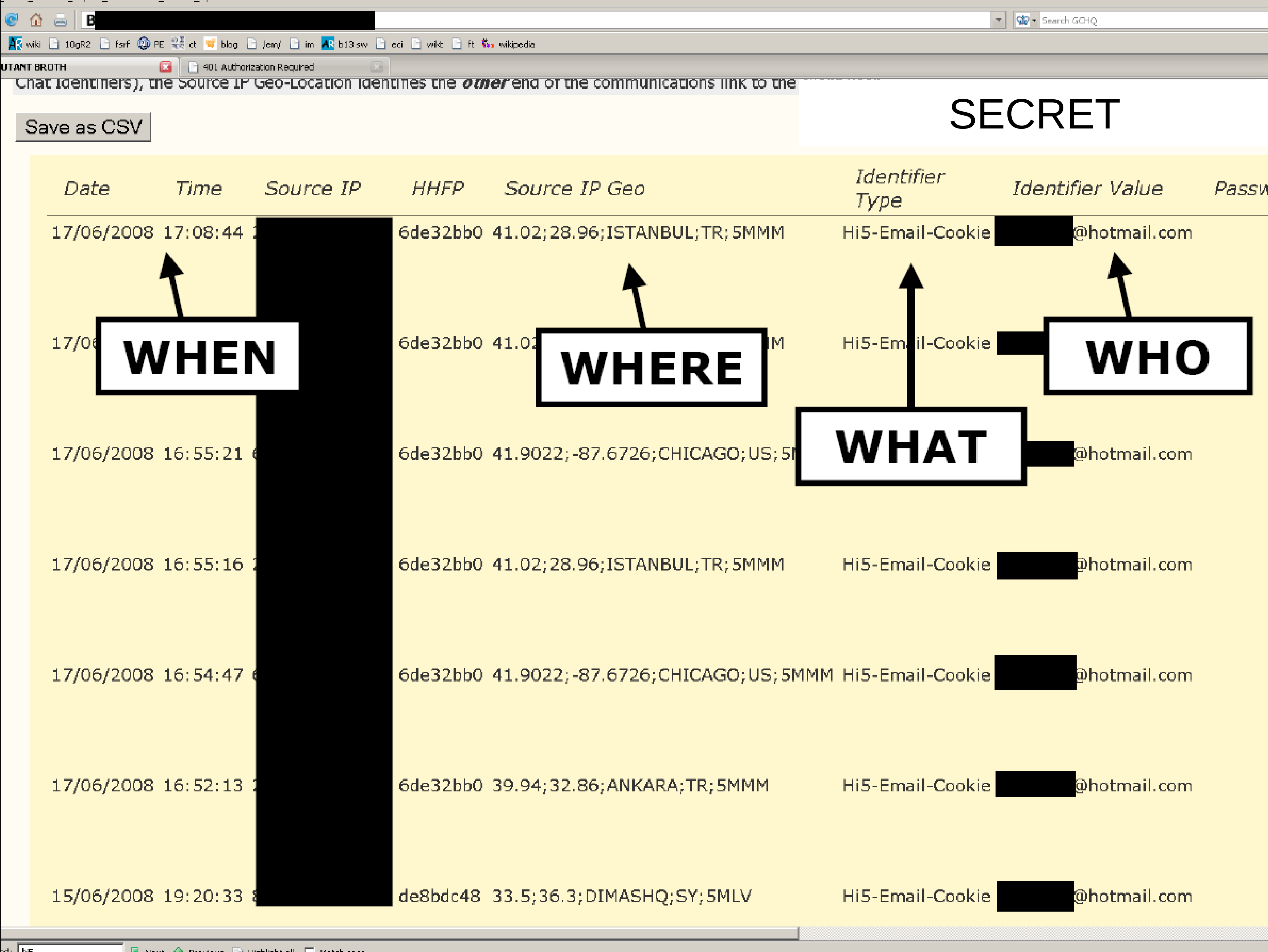
☑ Allow wildcards
☑ Convert to lowercase before searching

Matching Identifiers

The following identifiers have been found in the MUTANT BROTH database.

Select those that match your target(s) to generate a summary of target activity.

	TDI type	*TDI value*
☐	Chat-MS-Messenger	[redacted]@hotmail.fr



SECRET

Chat Identifiers), the Source IP Geo-Location identifies the *other* end of the communications link to the

Save as CSV

Date	Time	Source IP	HHFP	Source IP Geo	Identifier Type	Identifier Value	Passw
17/06/2008	17:08:44	[redacted]	6de32bb0	41.02;28.96;ISTANBUL;TR;5MMM	Hi5-Email-Cookie	[redacted]@hotmail.com	
17/06		[redacted]	6de32bb0	41.02 M	Hi5-Em il-Cookie	[redacted]	
17/06/2008	16:55:21	[redacted]	6de32bb0	41.9022;-87.6726;CHICAGO;US;5		[redacted]@hotmail.com	
17/06/2008	16:55:16	[redacted]	6de32bb0	41.02;28.96;ISTANBUL;TR;5MMM	Hi5-Email-Cookie	[redacted]@hotmail.com	
17/06/2008	16:54:47	[redacted]	6de32bb0	41.9022;-87.6726;CHICAGO;US;5MMM	Hi5-Email-Cookie	[redacted]@hotmail.com	
17/06/2008	16:52:13	[redacted]	6de32bb0	39.94;32.86;ANKARA;TR;5MMM	Hi5-Email-Cookie	[redacted]@hotmail.com	
15/06/2008	19:20:33	[redacted]	de8bdc48	33.5;36.3;DIMASHQ;SY;5MLV	Hi5-Email-Cookie	[redacted]@hotmail.com	

WHEN

WHERE

WHAT

WHO

Other Bulk Event Applications

- Most events that can be associated back to TDIs:
- File Transfer Signature (eg proof of life videos)
- Detection by Internet profile – eg ‘Dead Letter Drop’.
- Yahoo webcam images
- Airline reservation confirmation emails