Full Log vs. HTTP

11 June 2009

Derived From: NSA/CSSM 1-52
Dated: 20070108
Declassify On: 20291123

DERIVED FROM: NSA/CSSM 1-52

Full Log - Pros

- The Full Log search gives you access to all DNI sessions collected by X-KEYSCORE
- Data is indexed by the basic meta-data like IP Address, Country Codes Port, Casenotation, Application ID/Fingerprints etc
- If you're only interested in content, Full Log will give you access to everything

Full Log – Cons

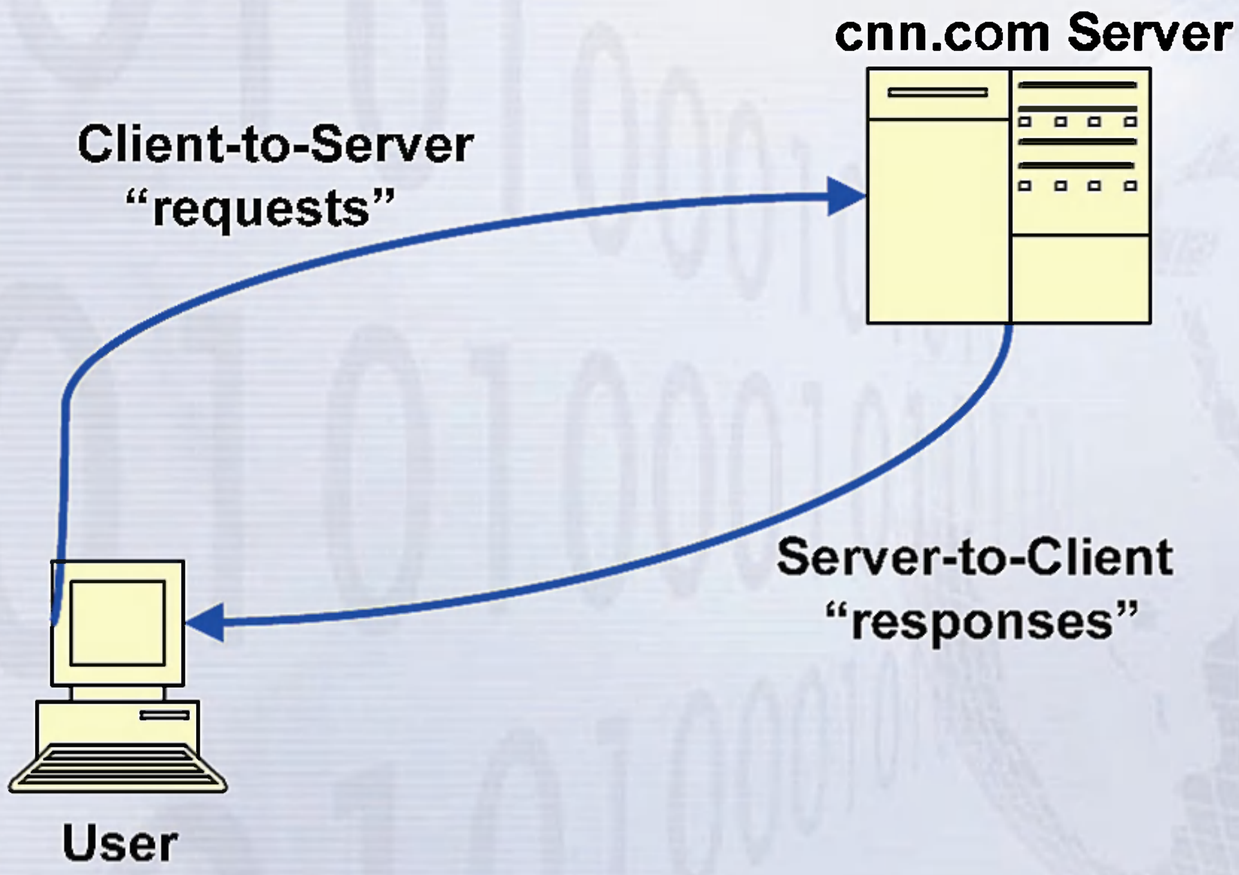
- However, in most cases there will be too many results in XKS to look through every piece of content by hand
- To be more efficient, it's important to utilize the meta-data contained in the other search forms (E-mail Addresses, HTTP Activity, Extracted Files, Document Meta-data etc.)

HTTP Activity

- HTTP Activity is essentially all web-based activity from a user's internet browser (with some exceptions)
- It includes, web-surfing, Internet Searching (like Google), Mapping Website (Google Earth/Maps) etc.
- Most of this data will not contain a strong selector like E-mail address

HTTP Activity

- HTTP activity comes in two types:



HTTP Activity Client-to-Server


```
GET /search?tab=urdu&order=sortboth&q=musharraf&start=3&scope=urdu&link=next HTTP/1.1
Accept: */*
Referer: http://search.bbc.co.uk/search?tab=urdu&order=sortboth&q=musharraf&start=2&scope=urdu
Accept-Language: en-us
Accept-Encoding: gzip, deflate
User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)
Host: search.bbc.co.uk
Cookie: BBC-UID=b479a5f4ad230a53063d513630203acb22684634a0e0b164c45f96efc054cf950Mozilla%2f4%2e0%20%28co
Cache-Control: max-stale=0
Connection: Keep-Alive
X-BlueCoat-Via: 66808702E9A98546
```

Host	URL Path	URL Args
search.bbc.co.uk	/search	tab=urdu&order=sortboth&q=musharraf&start=3&scope=urdu&link=next

Search Terms	Language	Browser	Via
musharraf	en	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)	66808702E9A98546

Referer
http://search.bbc.co.uk/search?tab=urdu&order=sortboth&q=musharraf&start=2&scope=urdu

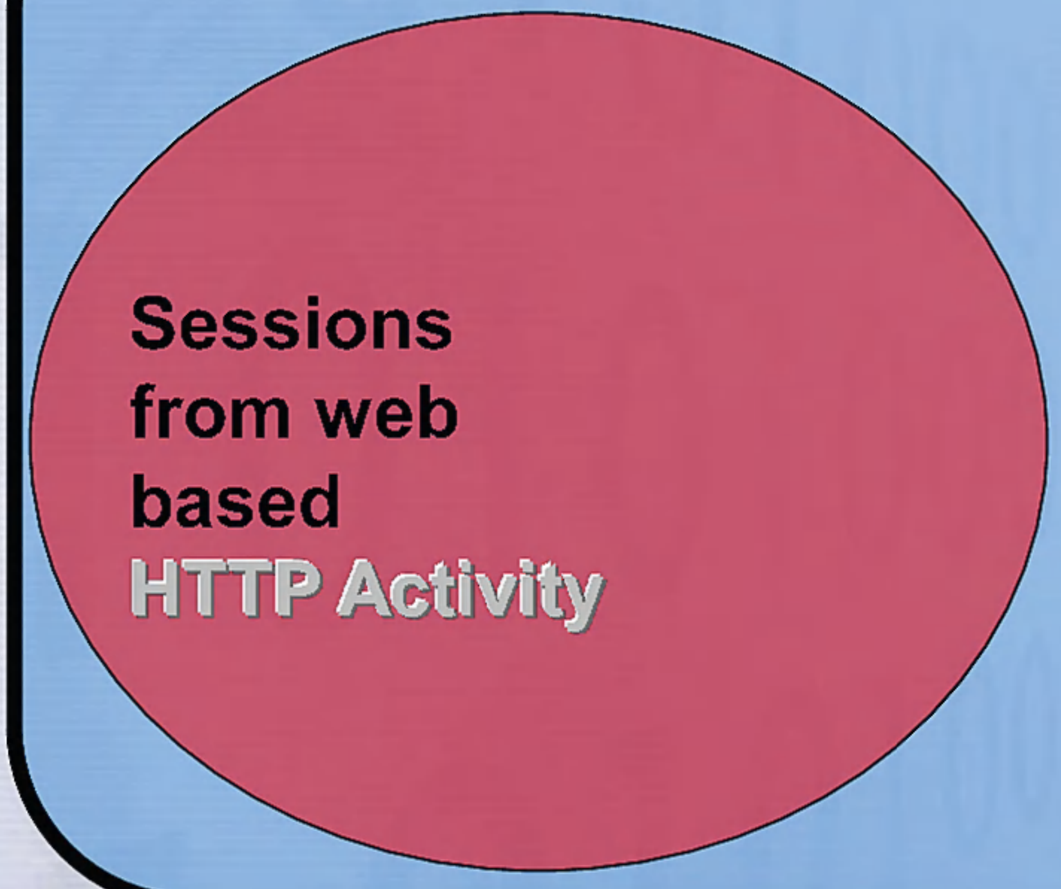
Cookie
BBC-UID=b479a5f4ad230a53063d513630203acb22684634a0e0b164c45f96efc054cf950Mozilla%2f4%2e0%20%28com

Overlap

- Full log contains basic information on every single DNI session XKS processes.
- HTTP activity contains more detailed information on the subset of that data which is web-based (aka port 80 “internet browser” traffic)

How the Search Forms Fit Together

Full Log of all DNI sessions collected



Example #1

KEYSCORE

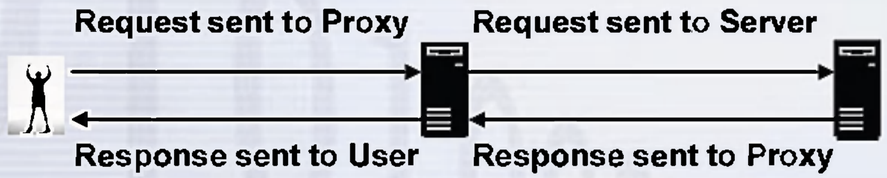
- Analysis of 14 May Internet session of PK based target started in MARINA

TS ▲	USERID	PHONE	USER_A	*ACTIVITY*	USER_B
20090514 132353Z			[redacted]:yahoo>	logged in (im)	119.[redacted]
20090514 132416Z			[redacted]@hotmail.com<msnpassport>	logged in (im)	119.[redacted]
20090514 132419Z			[redacted]@hotmail.com<msnpassport>	logged in (im)	119.[redacted]
20090514 132834Z			[redacted]@hotmail.com<msnpassport>	logged in (im)	119.[redacted]
20090514 132843Z			[redacted]@hotmail.com<msnpassport>	logged in (im)	119.[redacted]
20090514 133517Z			[redacted]@hotmail.com<msnpassport>	logged in (im)	119.[redacted]
20090514 133522Z			[redacted]@hotmail.com<msnpassport>	logged in (im)	119.[redacted]

Understand what is behind the IP

- Ensure Activity on IP can be associated with Target
- Understand IP usage Dynamic/Static
- Research IP using Foxtrail/NKB
- Is it a Proxy, DVBLAN, Dial-Up, DSL, etc
- Is it Client to Server or Server to Client
- Still not sure? User Activity pull for 5 minute period on Foreign IP

HTTP with a Proxy

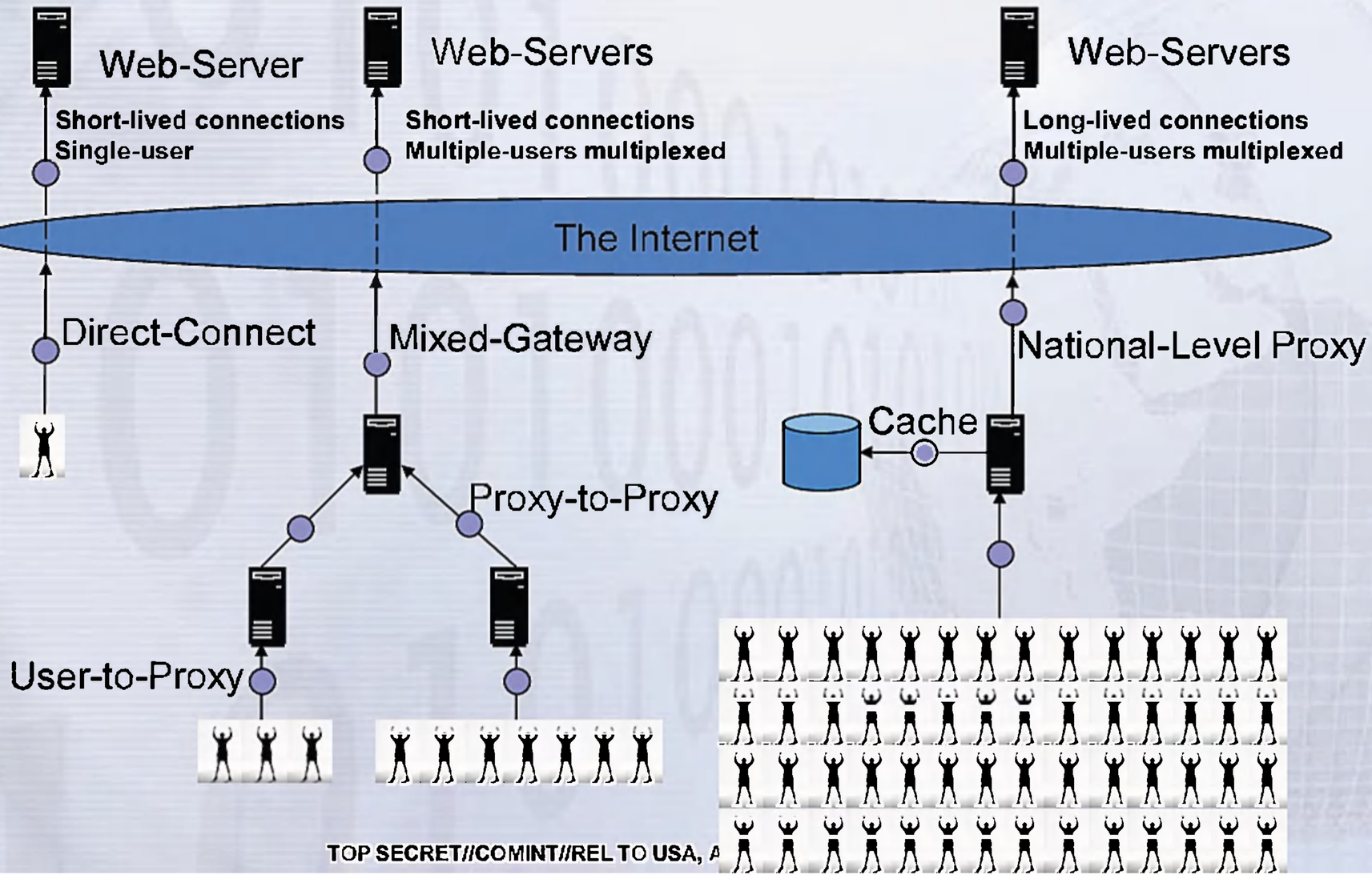


- **Performance**: Proxy can cache responses for static pages
- **Censorship**: Proxy can filter traffic
- **Security**: Proxy can look for malware
- **Access-Control**: Proxy can control access to restricted content

Proxy can be run by

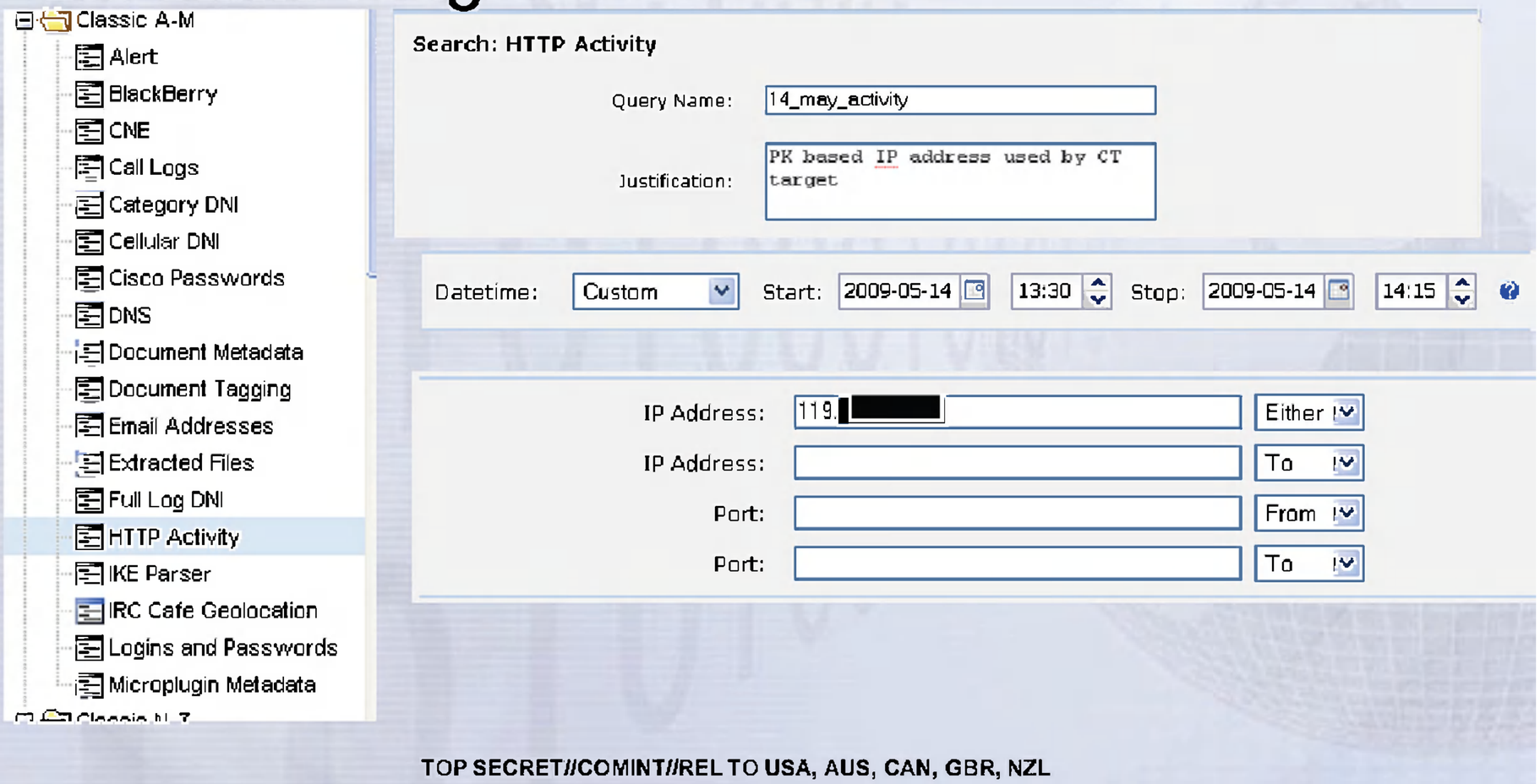
- a user
- an ISP
- a web-hosting company
- a content-delivery network (i.e. Akamai)

Proxies on the Internet



Example #1

- The analyst then did an HTTP activity query to find all web surfing from that IP address within the same rough timeframe.



14 May – Strange HTTP Activity

- HTTP meta-data indicated strange web-based activity

Host	URL Path
infoservice.inf.tu-dresden.de	/mixcascadestatus/F30905FCD73B6B30CB5FEFD3250FD66EF4B32591
infoservice.inf.tu-dresden.de	/infoservices

Browser
RPT-HTTPClient/0.4-dev
RPT-HTTPClient/0.4-dev

GET /mixcascadestatus/F30905FCD73B6B30CB5FEFD3250FD66EF4B32591 HTTP/1.1	
Host:	infoservice.inf.tu-dresden.de
Connection:	Keep-Alive, TE
TE:	trailers, deflate, gzip, compress
User-Agent:	RPT-HTTPClient/0.4-dev
Cache-Control:	no-cache
Pragma:	no-cache

14 May – Strange HTTP Activity

- Indications from the HTTP activity

Browser
RPT-HTTPClient/0.4-dev
RPT-HTTPClient/0.4-dev

Note the strange User Agent/Browser

GET /mixcascadestatus/F30905FCD73B6B30CB5FEFD3250FD66EF4B32591 HTTP/1.1	
Host:	infoservice.inf.tu-dresden.de
Connection:	Keep-Alive, TE
TE:	trailers, deflate, gzip, compress
User-Agent:	RPT'-HTTPClient/0.4-dev
Cache-Control:	no-cache
Pragma:	no-cache

Tip off to possible anonymizer

- Open Source research indicated that this user agent was indicative of multi-cast traffic. A likely tip off that this was some type of anonymizer

Browser
RPT-HTTPClient/0.4-dev
RPT-HTTPClient/0.4-dev

GET /mixcascadestatus/F30905FCD73B6B30CB5FEFD3250FD66EF4B32591 HTTP/1.1	
Host:	infoservice.inf.tu-dresden.de
Connection:	Keep-Alive, TE
TE:	trailers, deflate, gzip, compress
User-Agent:	RPT-HTTPClient/0.4-dev
Cache-Control:	no-cache
Pragma:	no-cache

HTTP Results led to Full Log Query

- The two tu-dresden.de requests were the only HTTP activity seen within that timeframe, but given the open source research suggesting that the user agent was an anoynmizing proxy a full log query was ran to identify all other traffic originating from that same IP address during the same time

Datetime End ▲	Fm IP
2009-05-14 13:47:33	119.[redacted]
2009-05-14 13:48:07	119.[redacted]

Host	URL Path
infoservice.inf.tu-dresden.de	/mixcascadestatus/F30905FCD73B6B30CB5FEFD3250FD66EF4B32591
infoservice.inf.tu-dresden.de	/infoservices

Full Log Results

Datetime End	Fm IP	To IP	Fm Por	To Pc	Application Type	Application
2009-05-14 13:47:33	119.[redacted]	141.[redacted]	1495	80	web	http/get
2009-05-14 13:48:07	119.[redacted]	65.5[redacted]	1502	443	network_encryption	network_encryption/https
2009-05-14 13:48:07	119.[redacted]	141.[redacted]	1494	80	web	http/get
2009-05-14 13:48:35	119.[redacted]	141.[redacted]	1134	6544	unknown	unknown/tcp
2009-05-14 13:49:02	119.[redacted]	141.[redacted]	1134	6544	unknown	unknown/tcp

Note the two HTTP activity (port 80) sessions were seen, but in addition there was one SSL (port 443) session and two unknown port 6544 sessions

Full Log Results

Fm IP	To IP	Fm Por	To Pc	Application Type	Application	Data Length	Session Length
119.[redacted]	141.[redacted]	1134	6544	unknown	unknown/tcp	6986	7803
119.[redacted]	141.[redacted]	1134	6544	unknown	unknown/tcp	93088	93901

Of the unknown port 6544 traffic, the data length of the sessions indicated that a significant amount of data was leaving the Pakistan IP used by our target

Full Log Results

The content appeared unreadable. Further analysis by CES and open source research showed that the content was encrypted

DNI Display | Raw Data | DNI Format

Services

No presentation available for this type of data. Try sending it to another service for a better view. Below is an attempt to display the document as plain text.

%þW>��|�����-7��'+0M|��j��w����Ki0@@f0\���x0Sg0f��/0#-��^90C0`0%���i��|ò[kF"0:
l0iI��Elg?��Kv0p��`�� ���L0\H��
P0U���$v0a0tdVq2`0Ì090f00U00U000VZ0}0 v{0A001��@00?0~X0;00bbnU000%0>0c8a0*0NC0r
"0]0
z.ò00FBnÀ000ßÿ00<00m0h0ò0E0Ag/0\00.50i6\?0piZ000/,0f00 0,04ºp000Fj0)00000>0wD0
;0Gj500K0000000 0/4~0U0N00ªz^0GXµj00J0Q,0H0l4000
:0G00‹00000f0v"R0>H0000000XG 000000y0JY000/000>030?m0l0~E0[l0U0"000Z000k0i0
0œm0W0g/0000300E[80kzN?0
X¹.X0kcYs0 0H0vi_030,030w01000/Uv0000C05*X!00Yp00#@
0N0000 M6K¾0l0000l,śX0U20X0000ZG0I00000
0|000Z0Y0"f0q0:0d0R?0:0g0Ry0002G0"c00000_0+)t?0D00UWX·0J0x4['3c0A|0I/$01080000:
ct00M000[04+60-0[00oswg0*_m00Zh"0#+000=0x000+Ÿ*0$"z8&00000µ00||)0000yu*0000s00:
*国000#0f00V0k0000Oey0d 00$0O0000_05=0000<0`0/00NXd^00ñD000k00QH.0050>(j+006
00-Ts0|0O000
00S"0C%0@l0r0035+bf0o000u00XD%]-
000h_d000u00:0j+00
0ţ000y00s00:0k0?0*000"bz004 <00j600i40G.00v$v00Z00 0R000990501000l0{00>0
0t+gV0t0P0000?iO/0Ō̈X-02le000500DN0i0%=p0Y00000D0%o00000`N0$^0000??(v0E1X-06
R

HTTP Results led to Full Log Query

- While we were ultimately unable to identify what was underneath the 150K of encrypted traffic, we were at the least able to identify that our target was using an anonymizing service to mask a portion of his Internet activity

SERI SERIAL: *3/00/513109-09

German Anonymizing Proxy (TS//SI/REL TO USA, FVEY)

(TS//SI//OC/REL TO USA, FVEY) During the 14, 16, and 18 May sessions on the [redacted] telephone, the user(s) were using a free German-based anonymizing proxy (http://infoservice.inf.tu-dresden.de), presumably to mask the source of Internet traffic. Use of the anonymizing proxy occurred primarily during the times in which [redacted]'s [redacted]@hotmail.com and [redacted]@yahoo.com accounts were accessed.

COMMENT: (TS//SI//OC/REL TO USA, FVEY) The German-based proxy is a Java Anonymous Proxy (JAP), which was developed by the Technical University of Dresden as a free and open source anonymity tool. The proxy functions in a manner similar to The Onion Router (TOR) network. Given his background in computer science and networking, it is not surprising that [redacted] would use an anonymizing proxy to secure his Internet activity.

Why not only use Full Log?

If the Full Log query gave us the HTTP traffic in addition to the other non web based traffic, why don't we only use the Full Log query?

- Because the meta-data options in the full log table are limited

Datetime End	Fm IP	To IP	Fm Por	To Pc	Application Type	Application
2009-05-14 13:47:33	119.[redacted]	141.[redacted]	1495	80	web	http/get
2009-05-14 13:48:07	119.[redacted]	65.5[redacted]	1502	443	network_encryption	network_encryption/https
2009-05-14 13:48:07	119.[redacted]	141.[redacted]	1494	80	web	http/get
2009-05-14 13:48:35	119.[redacted]	141.[redacted]	1134	6544	unknown	unknown/tcp
2009-05-14 13:49:02	119.[redacted]	141.[redacted]	1134	6544	unknown	unknown/tcp

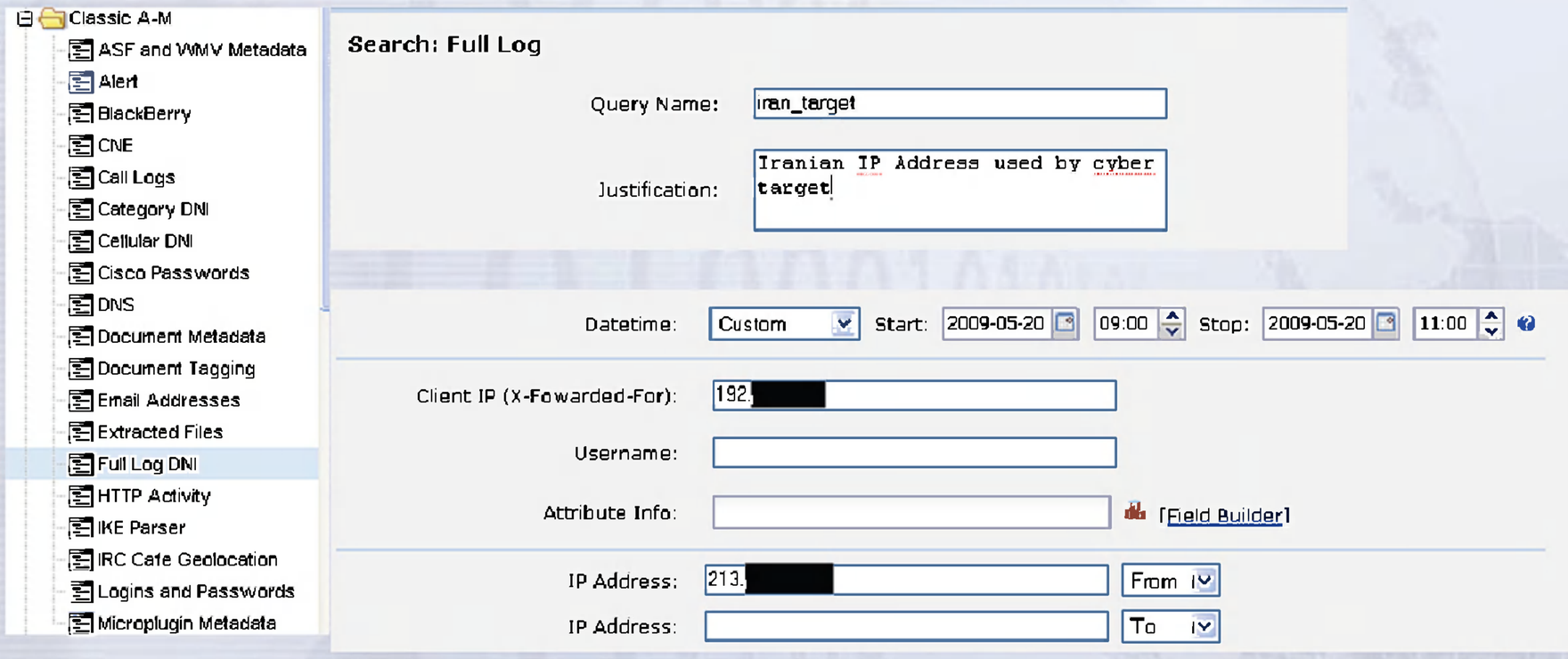
Example #2

- Starting with MARINA results of a 20 May Internet session of an Iran based target

TS ▼	USERID	PHONE	*USER_A*	*ACTIVITY*	USER_B	COOKIE
20090520 092139Z			[redacted]<yahoo>	logged in (email)	213.[redacted]	fuq8af14q5kjt<yahooBcookie>
20090520 092139Z			[redacted]<yahoo>	used xforip	192.[redacted]	fuq8af14q5kjt<yahooBcookie>
20090520 092130Z			[redacted]<yahoo>	logged in (email)	213.[redacted]	fuq8af14q5kjt<yahooBcookie>
20090520 092130Z			[redacted]<yahoo>	used xforip	192.[redacted]	fuq8af14q5kjt<yahooBcookie>

Example #2

- The analyst then did a full log query based off the IP & X-Forwarded-IP pair



Example #2

- Full Log table contains the standard DNI meta-data with *some but not all* information from other plug-ins included (ie. Username from User Activity and Application Info contains some HTTP activity)

Application Info	Username	Fm	Fm City (	To (	To City (IP)	Datetime	Datetime End	Fm IP	To IP	Fm Por	To Pc
http://update.nai.com/Products/Cc		IR	TEHRAN	US	NEWYORK	2009-05-20 10:05:16	2009-05-20 10:09:16	213.[redacted]	72.[redacted]	34847	80
http://platform.ak.facebook.com/v	[redacted]@gmail.co	IR	TEHRAN	DE	FRANKFURT	2009-05-20 10:06:54	2009-05-20 10:10:16	213.[redacted]	212.[redacted]	42806	80
http://platform.ak.facebook.com/v	[redacted]@gmail.co	IR	TEHRAN	DE	FRANKFURT	2009-05-20 10:06:54	2009-05-20 10:10:16	213.[redacted]	212.[redacted]	42806	80
http://platform.ak.facebook.com/v	[redacted]@gmail.co	IR	TEHRAN	DE	FRANKFURT	2009-05-20 10:06:54	2009-05-20 10:10:16	213.[redacted]	212.[redacted]	42806	80
http://newsrss.bbc.co.uk/rss/new		IR	TEHRAN	GB	LONDON	2009-05-20 10:07:43	2009-05-20 10:07:54	213.[redacted]	212.[redacted]	37459	80
http://b.static.ak.fbcdn.net/inbox.u		IR	TEHRAN	DE	FRANKFURT	2009-05-20 10:08:31	2009-05-20 10:11:33	213.[redacted]	212.[redacted]	41092	80
http://b.static.ak.fbcdn.net/rsrc.ph		IR	TEHRAN	DE	FRANKFURT	2009-05-20 10:08:31	2009-05-20 10:09:57	213.[redacted]	212.[redacted]	41092	80
http://platform.ak.facebook.com/v	[redacted]@gmail.co	IR	TEHRAN	DE	FRANKFURT	2009-05-20 10:09:39	2009-05-20 10:12:12	213.[redacted]	195.[redacted]	49648	80
http://photos-d.ak.fbcdn.net/phot		IR	TEHRAN	NL	AMSTERDAM	2009-05-20 10:09:56	2009-05-20 10:12:05	213.[redacted]	195.[redacted]	41696	80
http://photos-d.ak.fbcdn.net/phot		IR	TEHRAN	NL	AMSTERDAM	2009-05-20 10:09:56	2009-05-20 10:12:05	213.[redacted]	195.[redacted]	41696	80
http://b.static.ak.fbcdn.net/inbox/i		IR	TEHRAN	DE	FRANKFURT	2009-05-20 10:09:57	2009-05-20 10:10:09	213.[redacted]	195.[redacted]	34400	80
http://b.static.ak.fbcdn.net/images		IR	TEHRAN	DE	FRANKFURT	2009-05-20 10:09:57	2009-05-20 10:10:09	213.[redacted]	195.[redacted]	34400	80
http://b.static.ak.fbcdn.net/images		IR	TEHRAN	DE	FRANKFURT	2009-05-20 10:09:57	2009-05-20 10:10:09	213.[redacted]	195.[redacted]	34400	80
http://b.static.ak.fbcdn.net/images		IR	TEHRAN	DE	FRANKFURT	2009-05-20 10:10:00	2009-05-20 10:12:09	213.[redacted]	195.[redacted]	44898	80
http://b.static.ak.fbcdn.net/inbox.u		IR	TEHRAN	DE	FRANKFURT	2009-05-20 10:10:00	2009-05-20 10:12:09	213.[redacted]	195.[redacted]	44898	80
http://newsrss.bbc.co.uk/rss/new		IR	TEHRAN	GB	LONDON	2009-05-20 10:16:12	2009-05-20 10:16:24	213.[redacted]	212.[redacted]	48804	80

Example #2

- The analyst wanted to know if the From IP & X-Forwarded-For IP pair was representing a single computer or if there were multiple users on multiple computers in this data.
- Full log only provides the bare minimum meta-data to make this determination

ID	Datetime ▲	Application Info
70	2009-05-20 10:21:45	http://us.mg1.mail.yahoo.com/dc/rs?log=ActivityMaxIdleTime:High&&.gx=1/login_webmail
61	**2009-05-20 10:22:17**	**http://0.channel31.facebook.com/x/3023227576/false/p_1406565350=0/login_webmail**

Fm IP	To IP	Fm Port	To Pc
213.[redacted]	209.191.106.109	55828	80
213.[redacted]	**69.63.176.213**	**55435**	**80**

Application Type	Application	Data Length	Session Length
mail	mail/webmail/yahoo	1446	1968
social	**social/facebook**	**3402**	**3922**

Example #2

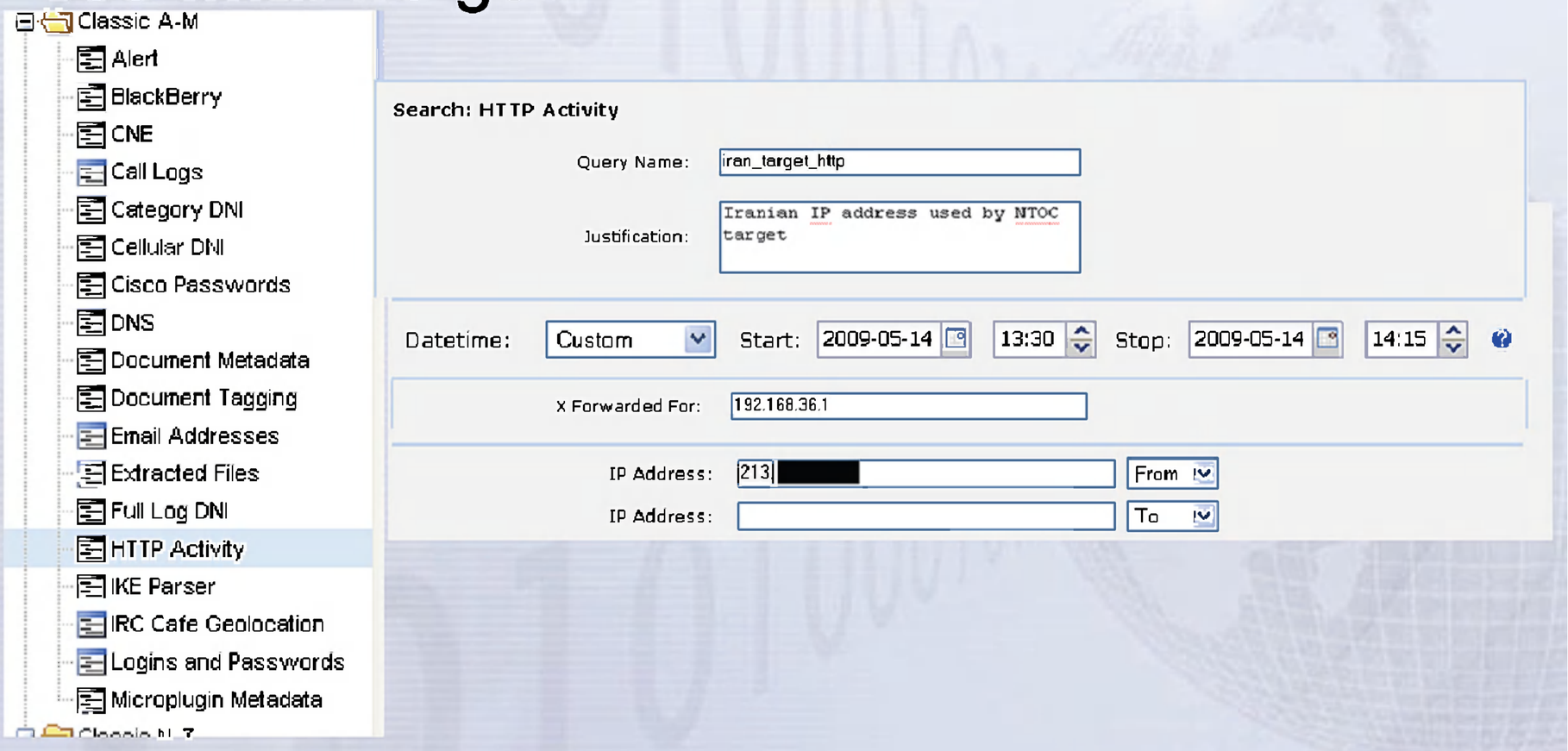
- MARINA provided this information:

TS ▲	USERID	PHONE	USER_A	ACTIVITY	USER_B	COOKIE
20090520 102145Z			[redacted]<yahoo>	previously logged in	213.[redacted]	fuq8af14q5kjt<yahooBcookie>
20090520 102145Z			[redacted]<yahoo>	previously logged in	213.[redacted]	fuq8af14q5kjt<yahooBcookie>
20090520 102145Z			213.[redacted]	used xforip	192.168.36.1	fuq8af14q5kjt<yahooBcookie>
20090520 102217Z			[redacted]<facebook>	used xforip	192.168.36.1	
20090520 102217Z			[redacted]<facebook>	registered with	[redacted]@gmail.com<google>	
20090520 102217Z			[redacted]<facebook>	logged in (forum)	213.[redacted]	

- The Yahoo and Facebook activity came from the same proxy IP and the same X-Forwarded-For-IP and around the same time but was it from the same computer?

HTTP Activity Query

- Let's query that same date time range and IP and XFF IP pair in the HTTP Activity query to see what we get



Now view the HTTP Activity results

- We saw this meta-data in the Full Log results:

ID	Datetime ▲	Application Info
70	2009-05-20 10:21:45	http://us.mg1.mail.yahoo.com/dc/rs?log=ActivityMaxIdleTime:High&&.gx=1/login_webmail
61	**2009-05-20 10:22:17**	**http://0.channel31.facebook.com/x/3023227576/false/p_1406565350=0/login_webmail**

Fm IP	To IP	Fm Port	To Pc
213.[redacted]	209.[redacted]	55828	80
213.[redacted]	**69.[redacted]**	**55435**	**80**

Application Type	Application	Data Length	Session Length
mail	mail/webmail/yahoo	1446	1968
social	**social/facebook**	**3402**	**3922**

- And then these three fields are among the unique (and valuable) fields only found in the HTTP activity table:

Cookie	Referer	Browser
YM[redacted]=fpwidth=165&suc=3	http://us.mg1.mail.yahoo.com/dc/launc	Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.10
datr=1242109330-dc7046de296a31363cbe218f6	http://0.channel31.facebook.com/ifram	Mozilla/5.0 (Windows; U; Windows NT 5.2; en-US; rv:1.9.0.10

Now view the HTTP Activity

- Of interest, note the differences between the two user agents

Browser
Mozilia/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.10) Gecko/2009042316 Firefox/3.0.10
Mozilia/5.0 (Windows; U; Windows NT 5.2; en-US; rv:1.9.0.10) Gecko/2009042316 Firefox/3.0.10

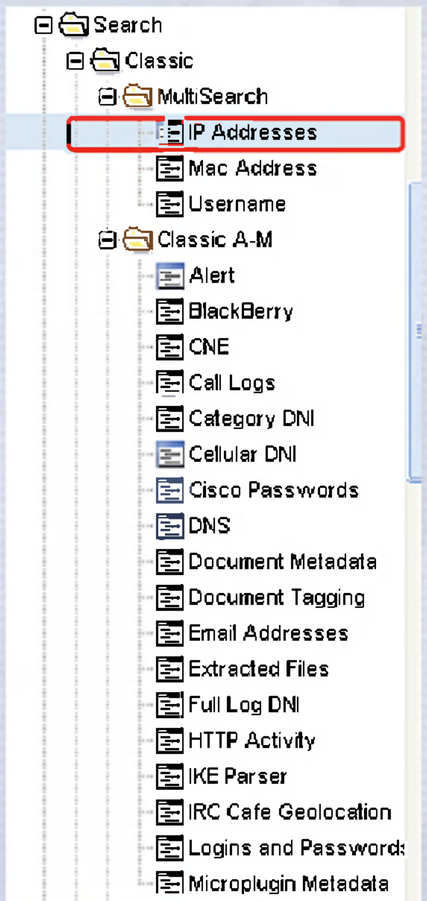
- This indicates different versions of Windows, so unless they did an upgrade within the 1 minute difference of activity, there were at least two different computers behind that Proxy and XFF IP pair

Moral of the story

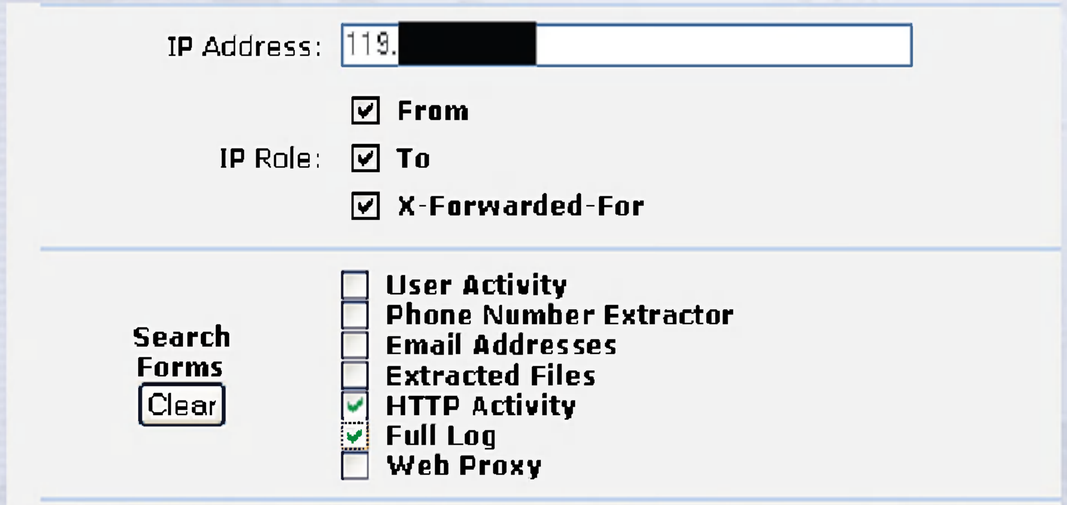
- You should be use both the HTTP activity and Full Log queries to help discover everything your target does when he's online
- HTTP Activity will give you great meta-data for quick analysis of "web-based" (port 80) activity
- But not all DNI is done through an Internet Browsers, so it's important to look at the Full Log query results for indications of the use of other applications

Moral of the story

- The Multi-Search page gives you the ability to search full log and HTTP activity based on an IP address at the same time



Simply enter in an IP address, choose any or all "roles" (ie. from/to/xff) and then choose what search forms you want.



Moral of the story

- It will submit the multiple searches at the same time, you can either view the results separately or view them as a merged table

My Recent Results

Help Actions View FILTERS:

	Datetime Submitted	Query Name	Status	Num Results	Query Type
	2009-06-08 14:33:40	pk_ip_16_may	finished	605	full_log
	2009-06-08 14:33:40	pk_ip_16_may	finished	475	http_parser