Writing XKS Fingerprints

[redacted]

November 2010

Agenda

- Naming Fingerprints
- Simple Keywords
- Boolean Logic
- Variables
- Context-Sensitive

Fingerprints 101

- What's in a name?

- The XKS Fingerprint naming convention can help organize fingerprints and make searching easier so its important to make sure you name your fingerprint inline with the existing convention

What's in a name

- For example, fingerprint names look like this:

- encryption/archive/rar
- encryption/archive/pkzip
- encryption/archive/pkzip

- Notice the directory-like structure so that all encryption fingerprints are within the same "folder" and all encryption/archive fingerprints are within the same "folder"

What's in a name

- This allows for smarter searching because you could look for all encryption fingerprints by searching for encryption/* or search for all encryption/archive fingerprints by searching for encryption/archive/* and etc.

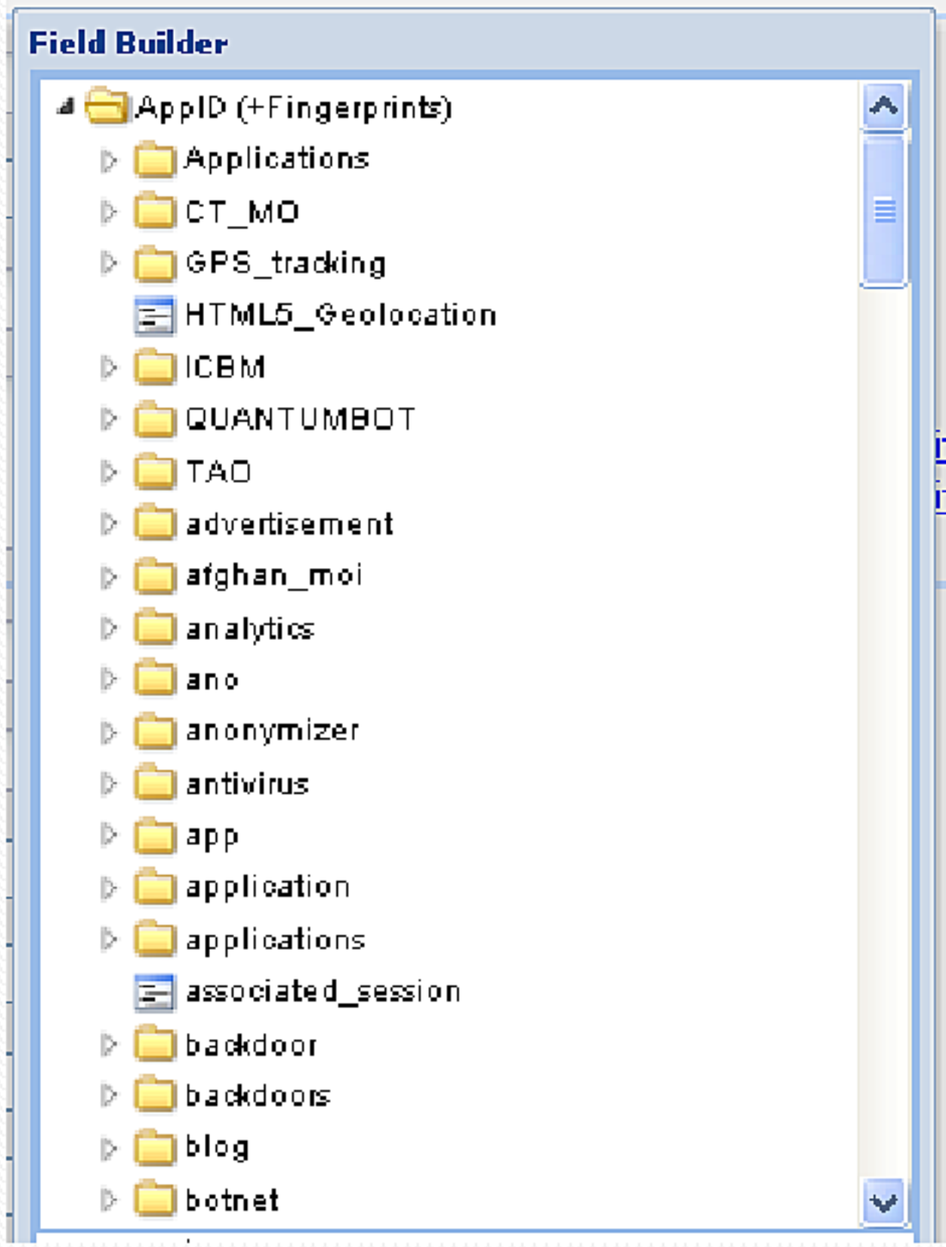
What's in a name

- When you want to submit a new fingerprint, look to see if it would fit into any existing fingerprint folders.

- Best way to do this is to use either the "Field Builder" or "Tree Field Builder" next to the AppID+Fingeprints field in the search forms

What's in a name

- The field builders allow you to browse existing fingerprint directories to see if one already exists for your new fingerprint



Fingerprint directories

Field Builder

AppID (+Fingerprints)

topic/wmd/iran/irisl

- topic/wmd/iran/irisl/edi1/chat_body
- topic/wmd/iran/irisl/edi1/document_body
- topic/wmd/iran/irisl/edi1/email_body
- topic/wmd/iran/irisl/edi1/filename
- topic/wmd/iran/irisl/edi1/url_path
- topic/wmd/iran/irisl/edi2
- topic/wmd/iran/irisl/edi3

Field Builder

AppID (+Fingerprints)

mojahe

- encryption/mojaheden2
- encryption/mojaheden2/encodedheader
- encryption/mojaheden2/hidden
- encryption/mojaheden2/hidden2
- encryption/mojaheden2/hidden44
- encryption/mojaheden2/secure_file_encoded
- encryption/mojaheden2/securefile

Field Builder

AppID (+Fingerprints)

botnet/black

- botnet/blackenergybot/command/die
- botnet/blackenergybot/command/flood
- botnet/blackenergybot/command/icmp
- botnet/blackenergybot/command/stop
- botnet/blackenergybot/command/syn
- botnet/blackenergybot/command/wait

What's in a name

- If no existing directory makes sense for your fingerprint, you can always create a new one.

Fingerprints 101: Getting Started

- The first step is to define the name of the fingerprint.
- To do that, follow the syntax below:

fingerprint('encryption/archive/test_new') =

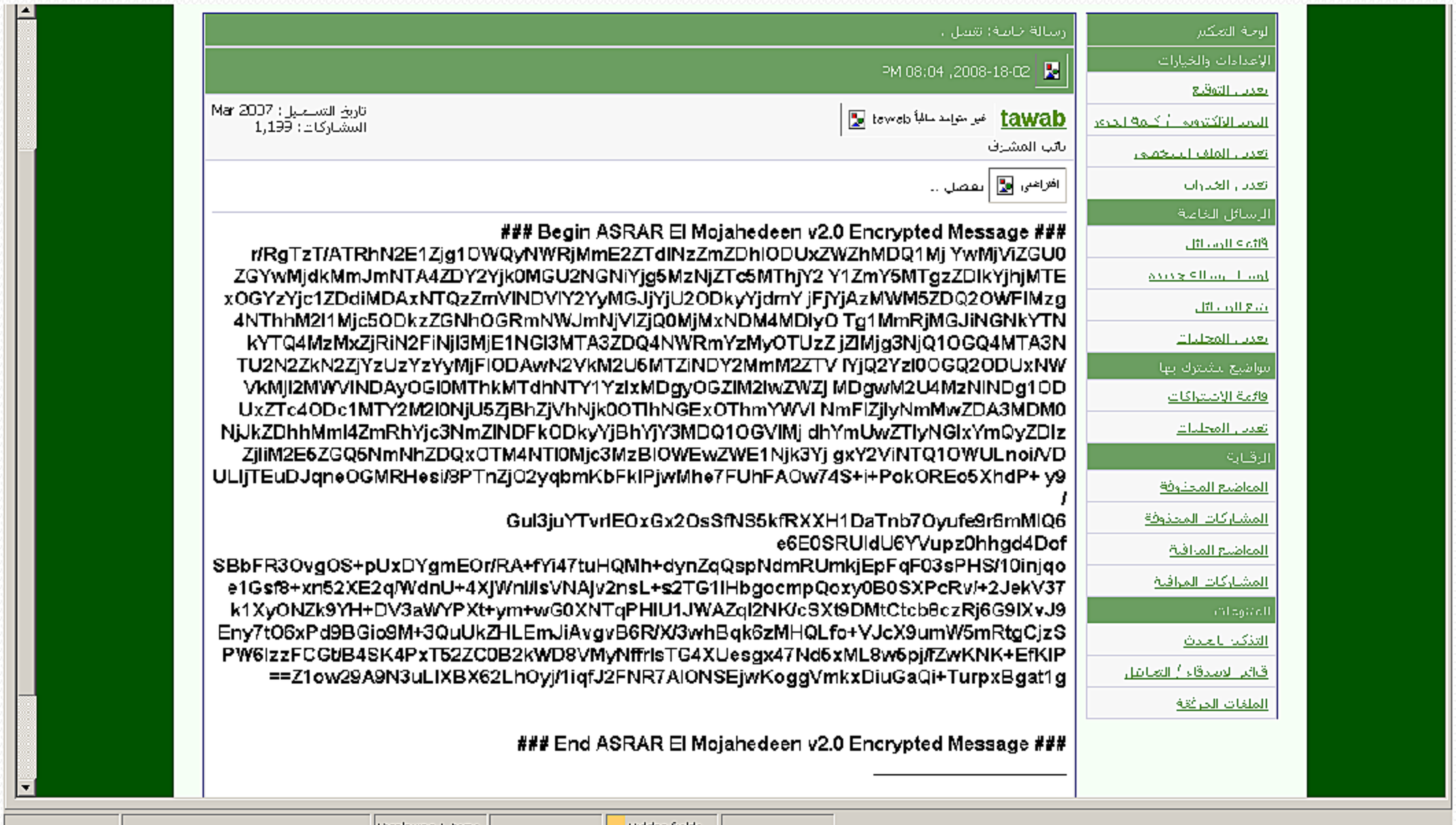
Fingerprints 101: Getting Started

- Note that fingerprint names can not have spaces or any other punctuation other than / which denote directories and _ which can be used in the place of spaces to make fingerprint names easier to read

- fingerprint('encryption/archive/test_new') =

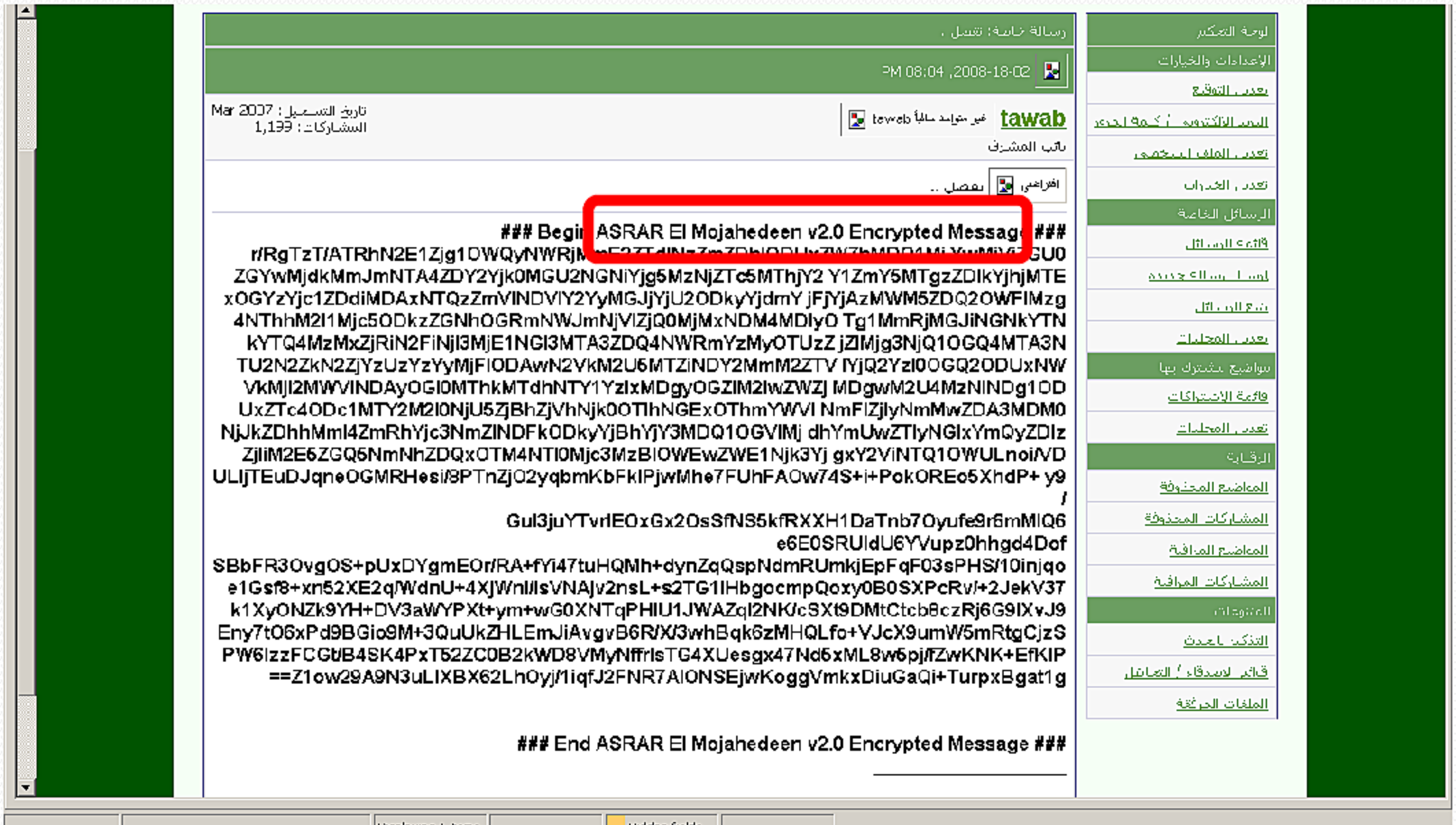
Fingerprints 101

- As an example, let's say we want to fingerprint traffic like this:



Fingerprints 101

- One thing that could be used to find data like this is the string ASRAR El Mojahdeen v2.0 Encrypted Message



Fingerprints 101: Keywords

- So let's create a fingerprint to tag any data that contains that string

ASRAR El Mojahdeen v2.0 Encrypted Message

Fingerprints 101: Keywords

- First we'd define the fingerprint with a name:

fingerprint('encryption/mojahdeen2') =

Fingerprints 101: Keywords

- Then, simply put the string in single quotes to denote that XKS needs to look for it as a keyword:

fingerprint('encryption/mojahdeen2') =

'ASRAR El Mojahdeen v2.0 Encrypted Message'

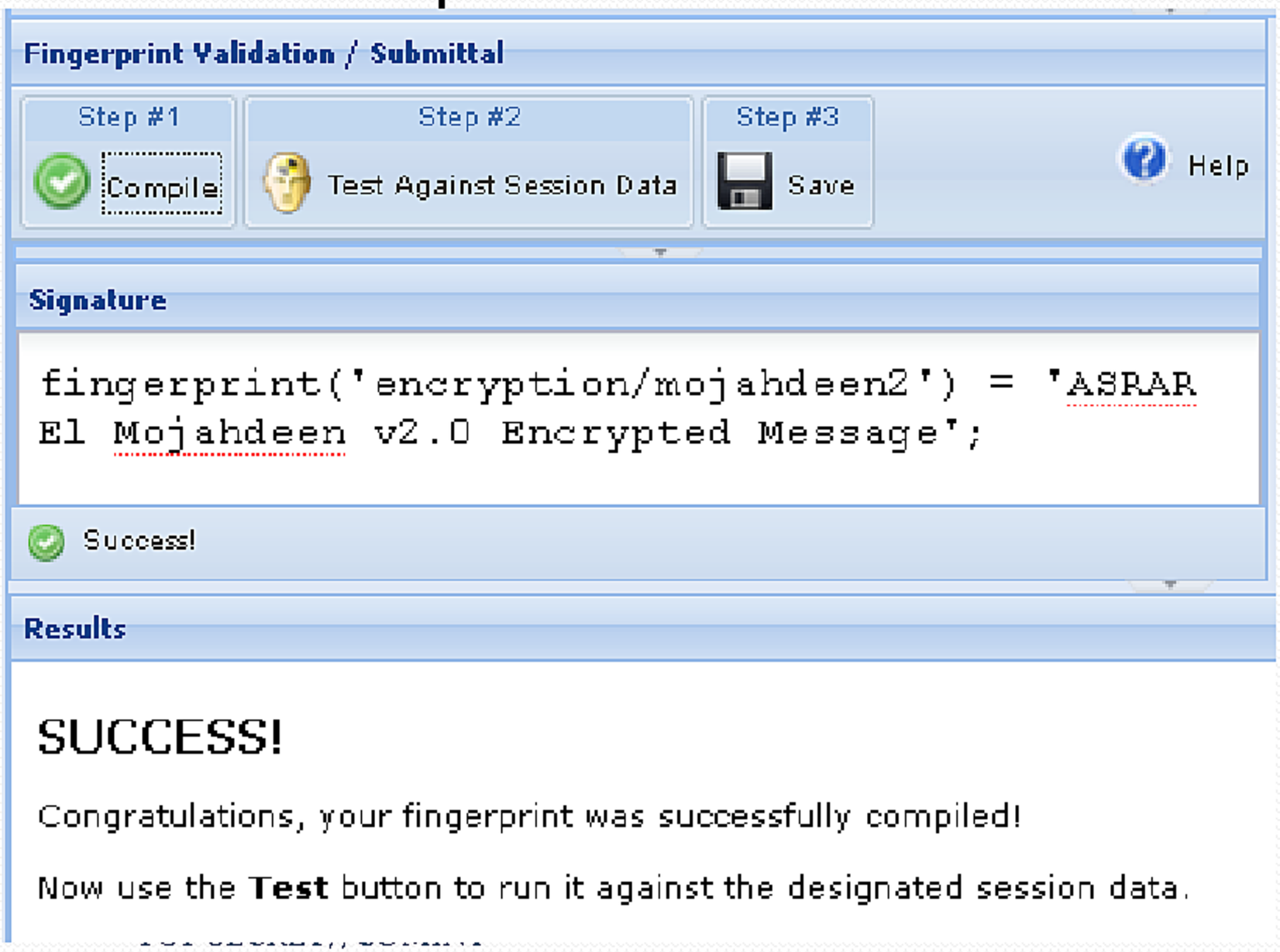
Fingerprints 101: Keywords

- Finally, all fingerprint definitions need to end with a semi colon to tell XKS that the definition is finished

fingerprint('encryption/mojahdeen2') =
'ASRAR El Mojahdeen v2.0 Encrypted Message' ;

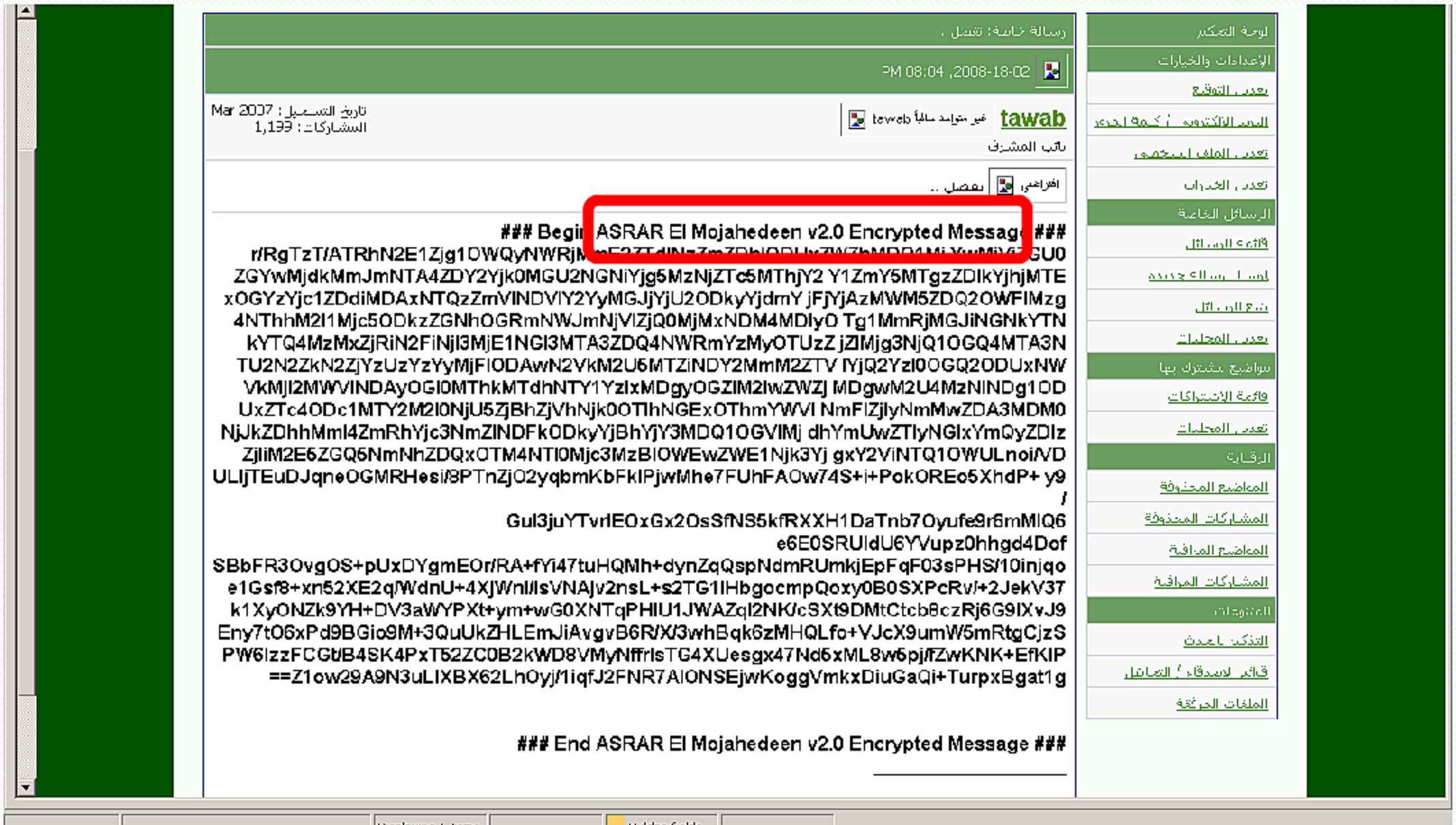
Fingerprints 101: Keywords

- Using the fingerprint GUI on XKS Central, we can test to see if this compiles:



Fingerprints 101

- Once checked in, the fingerprint will hit on data like this:



Fingerprints 101

- As a second example, let's say we want to find data like this:

Using TXT formatter

```
Ref: June 07, 201000803/Q-02135 Islamabad:
National Development Complex
Plot No: [redacted] Street No: [redacted]
Sector: [redacted]
Islamabad.
Attn: [redacted]
AM Purchase
SUBJECT : QUOTATION AGAINST YOUR ENQUIRY REF:Purchase of RTV Silicon DATED:
18/05/2010
Dear Sir,
With reference to your subject enquiry, we are pleased to enclose our Quotation No: Q-02135-05-567
dated: 07/06/2010, for your perusal.
Please see the 'Terms of Sale' attached with our quote for any further details.
We hope our offer suits your requirements and we look forward to your valuable purchase order in due
```

Fingerprints 101

- Look for keywords that could be used to find traffic like this in the future.

Using TXT formatter

```
Ref: June 07, 201000803/Q-02135 Islamabad:
National Development Complex
Plot No: [redacted] Street No: [redacted]
Sector: [redacted]
Islamabad.
Attn: [redacted]
AM Purchase
SUBJECT : QUOTATION AGAINST YOUR ENQUIRY REF:Purchase of RTV Silicon DATED:
18/05/2010
Dear Sir,
With reference to your subject enquiry, we are pleased to enclose our Quotation No: Q-02135-05-567
dated: 07/06/2010, for your perusal.
Please see the 'Terms of Sale' attached with our quote for any further details.
We hope our offer suits your requirements and we look forward to your valuable purchase order in due
```

Fingerprints 101

- What if we looked for “National Development Complex” and “Quotation”

Using TXT formatter

```
Ref: June 07, 2010/NDC/Q-02135 Islamabad:
National Development Complex
Plot No: [redacted] Contact No: [redacted]
Sector: [redacted]
Islamabad.
Attn: [redacted]
AM Purchase
SUBJECT : QUOTATION AGAINST YOUR ENQUIRY REF:Purchase of RTV Silicon DATED:
18/05/2010
Dear Sir,
With reference to your subject enquiry, we are pleased to enclose our Quotation No: Q-02135-05-567
dated: 07/06/2010, for your perusal.
Please see the 'Terms of Sale' attached with our quote for any further details.
We hope our offer suits your requirements and we look forward to your valuable purchase order in due
```

Fingerprints 101: Boolean Logic

- Starting with these two keywords, we'd like to use Boolean Logic to create our new fingerprint
  - national development complex
  - quotation

Fingerprints 101: Boolean Logic

- Again, step one think of a name:

fingerprint('cp/pakistan/agencies/ndc') =

Fingerprints 101: Boolean Logic

- Step two, put single quotes around all keywords:

fingerprint('cp/pakistan/agencies/ndc') =

'National Development Complex'

'quotation'

Fingerprints 101: Boolean Logic

- Use the Boolean operator *and*

fingerprint('cp/pakistan/agencies/ndc') = 'National Development Complex' and 'quotation'

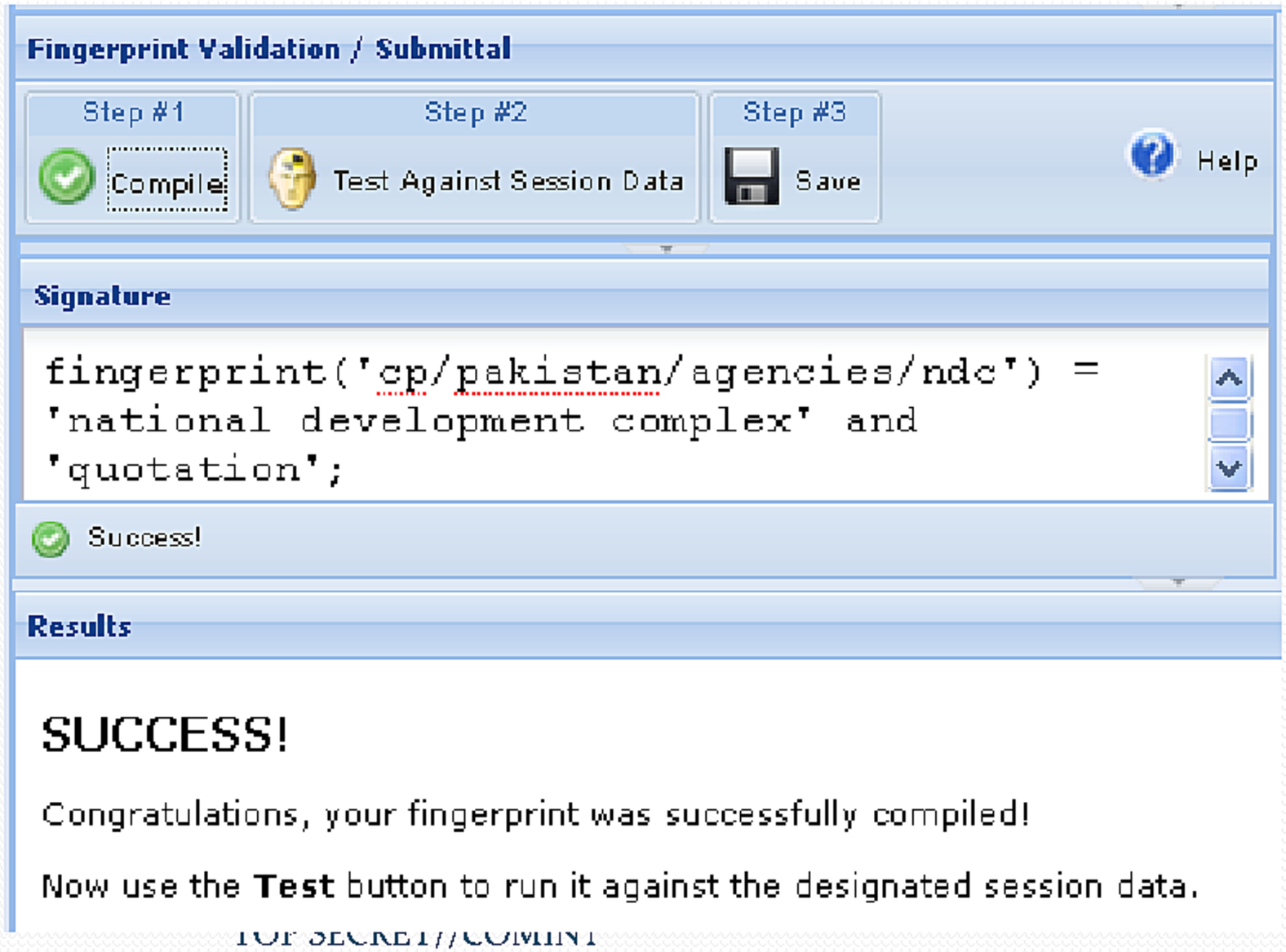
Fingerprints 101: Boolean Logic

- Finish the expression with the semi-colon.

fingerprint('cp/pakistan/agencies/ndc') = 'National Development Complex' and 'quotation' ;

Fingerprints 101: Boolean Logic

- Use the fingerprint GUI to confirm the fingerprint definition compiles



Fingerprints 101

- This fingerprint will now successfully find all sessions like this in the future!

Using TXT formatter

```
Ref: June 07, 201000803/Q-02135 Islamabad:
National Development Complex
Plot No: [redacted]  Street No: [redacted]
Sector: [redacted]
Islamabad.
Attn: [redacted]
AM Purchase
SUBJECT : QUOTATION AGAINST YOUR ENQUIRY REF:Purchase of RTV Silicon DATED:
18/05/2010
Dear Sir,
With reference to your subject enquiry, we are pleased to enclose our Quotation No: Q-02135-05-567
dated: 07/06/2010, for your perusal.
Please see the 'Terms of Sale' attached with our quote for any further details.
We hope our offer suits your requirements and we look forward to your valuable purchase order in due
```

Fingerprints 101

- However, how can we account for variations of how the traffic might be seen? Maybe "National Development Complex" will be listed as "NDC". Or maybe instead of a "Quotation" it will be a "Invoice" and etc.

Using TXT formatter

```
Ref: June 07, 201000803/Q-02135 Islamabad:
National Development Complex
Plot No: [redacted] Street No: [redacted]
Sector: [redacted]
Islamabad.
Attn: [redacted]
AM Purchase
SUBJECT : QUOTATION AGAINST YOUR ENQUIRY REF:Purchase of RTV Silicon DATED:
18/05/2010
Dear Sir,
With reference to your subject enquiry, we are pleased to enclose our Quotation No: Q-02135-05-567
dated: 07/06/2010, for your perusal.
Please see the 'Terms of Sale' attached with our quote for any further details.
We hope our offer suits your requirements and we look forward to your valuable purchase order in due
```

Fingerprints 101: Boolean Logic

- Keywords can also be grouped together by parentheses to form more complex Boolean logic:

Fingerprints 101: Boolean Logic

- For example, we can expand on our previous fingerprint like so

fingerprint('cp/pakistan/agencies/ndc') =
('National Development Complex' or 'NDC')
and ('quotation' or 'invoice') ;

Quick Aside 1: Context Sensitivity

- All keywords in X-KEYSCORE are case-insensitive by default.

- So in the previous fingerprint 'NDC' will match on *ndc*, *NdC*, *nDC* etc.

Quick Aside 1: Context Sensitivity

- If you want to force a keyword to be case sensitive, simply append a *c* after the single quotes.
- Ex: 'NDC'c will only hit when *NDC* is found in all caps, or 'ndc'c will hit only when *ndc* is found in all lower case and etc.

Quick Aside 2: Keyword Scanning

- By default keywords in fingerprints can hit in substrings since for example 'ndc' is found within grandchildren.
- So this fingerprint

fingerprint('cp/pakistan/agencies/ndc') =
'NDC';

Will hit on terms like:

- grandchildren
- handcard
- handcuffs
- etc.

Quick Aside 2: Keyword Scanning

- In specific cases to avoid false hits you can use the 'word' context.
- Or force there to be a space on either or both ends of the term by including them inside the single quotes
- So this fingerprint becomes:

fingerprint('cp/pakistan/agencies/ndc') =
' NDC ';

OR:

fingerprint('cp/pakistan/agencies/ndc') =
word('NDC');

Fingerprints 101: Boolean Logic

- Let's say that this fingerprint is producing good hits, but it also hitting on spam E-mails.

fingerprint('cp/pakistan/agencies/ndc') =
('National Development Complex' or 'NDC')
and ('quotation' or 'invoice') ;

Fingerprints 101: Boolean Logic

- We can use the Boolean *and not* to defeat unwanted traffic like below:

fingerprint('cp/pakistan/agencies/ndc') =
(('National Development Complex' or 'NDC')
and ('quotation' or 'invoice')) and not
('viagra' or 'herbal supplement');

Fingerprints 101: Variables

- Variables allow you to link to a list of keywords.
- For example, working with this fingerprint, we could create variables to each grouping of terms.

fingerprint('cp/pakistan/agencies/ndc') =

(('National Development Complex' or 'NDC') and ('quotation' or 'invoice')) and not ('viagra' or 'herbal supplement');

Fingerprints 101: Variables

Variables use the same syntax as fingerprints

$NDC_terms = 'National Development Complex' or 'NDC';

$procurement_terms = 'quotation' or 'invoice';
$spam_defeats = 'viagra' or 'herbal supplement';

fingerprint('cp/pakistan/agencies/ndc') = ($NDC_terms and $procurement_terms) and not $spam_defeats;

Fingerprints 101: Variables

- Variables can be re-used in multiple fingerprints.
- For example, we could have:

fingerprint('cp/pakistan/agencies/ndc') = ($NDC_terms and $procurement_terms) and not $spam_defeats;

fingerprint('cp/pakistan/angencies/ndc/testing') = $NDC_terms and ('missile launch' or 'tactical radio');

Fingerprints 101: Variables

- In the future, you can modify the variable $NDC_terms and it will automatically affect both fingerprints since they use that variable in their definition.

When that's not enough…

- For example, take the first scenario:

"I want to look for documents from Iran that mention a banned item"

- Just using keywords with Boolean equations, how could we restrict the term to only a document body and only coming from Iran?

Context Sensitive Scanning

- X-KEYSCORE's context sensitive scanning engine allows you to explicitly say where you want a term to hit.

- As an early example, the Tech Strings in Documents capability allowed analysts to restrict terms to only Email, Chat or Documents Bodies

- The full XKS Context Sensitive Scanning engine allows for over 70 unique contexts to be used as part of an fingerprint

Context Sensitive Scanning

- For example, take the first scenario:

"I want to look for documents from Iran that mention a banned item"

- Using the XKS context for Country Code (based on NKB information) and the XKS context for Document Bodies, this easily becomes:

fingerprint('demo/scenario1') =
cc('ir') and doc_body('banned item')

Context Sensitive Scanning

- As another example, let's say we want to tag all Iphone usage

- Using the XKS context for User Agent this easily becomes

fingerprint('demo/scenario2') =
user_agent('iphone');

USSID18/HRA Considerations

- XKS Fingerprints may not be USSID18 or HRA compliant if they are queried on by themselves

- For example, we may want to fingerprint the use of mobile web devices like the IPhone, so that attribute could be used as part of a more complex query.

- But querying for the IPhone fingerprint itself would be a USSID18 and HRA violation.

USSID18/HRA Considerations

- **But if you want to look for an IPhone user from an Iranian Proxy accessing his Mail.ru account:**

IP Address: 78.[redacted] Either

AppID (+Fingerprints) [fulltext]:

Field Builder

AppID (+Fingerprints)

browser/cellphone/iphone

Add to Field Close

Field Builder

AppID (+Fingerprints)

mail/webmail/mailru

mail/webmail/mailru

mail/webmail/mailru/attachment

mail/webmail/mailru/post

Context Sensitive Scanning

What contexts are available for use in XKS Fingerprints?

HTTP Activity Contexts (1 of 2)

html_title(expr)	The normalized extracted text web page titles html_title('how to' and 'bomb')
http_host(expr)	The "Host:" name given in the http header. http_host('yahoo.com')
http_url(expr)	Every URL from HTTP GET and POST commands. http_url('/mail/inbox?action=delete')
http_url_args(expr)	All arguments given as part of a URL (ie. all text following the '?' in a URL string) http_url('action=delete')
http_referer(expr)	The "Referer:" URL given in the HTTP header http_referer('http://badwebsite/cp?action=show')
http_language(expr)	The normalized two letter iso-6393 language code as inferred from any http and or html header info http_language('fa' or 'de')

HTTP Activity Contexts (2 of 2)

http_cookie(expr)	The "Cookie:" field given in the http header. http_cookie(/PREF=\d\d[a-z]/)
http_server(expr)	The "Server:" type name in the http header. http_server('GWS/2.1' or 'Apache')
http_user_agent(expr)	The "User-Agent:" field given in the http header. http_user_agent(/Mozilla\/[45]/ or 'Chrome')
web_search(expr)	The normalized extracted text from web searches web_search('ricin' or 'plague')
x_forwarded_for(expr)	The X-Forwarded For IP address from the HTTP Header x_forwarded_for('1.2.3.4')

Protocol Contexts 1 of 2

ip(expr)	The source or destination IP address of the session ip('127.0.0.1')
from_ip(expr)	The source IP address of the session from_ip('127.0.0.1')
to_ip(expr)	Every URL from HTTP GET and POST commands. to_ip('127.0.0.1')
ip_subnet(expr)	IP subnet in CIDR notation. ip_subnet('7.211.143.148/24')
port(expr)	The source or destination TCP or UDP port number. port('22')
from_port(expr)	The source TCP or UDP port number. from_port('22')
to_port(expr)	The destination TCP or UDP port number. to_port('22')

Protocol Contexts 1 of 2

cc(expr)	The country (either to OR from) based on IP address cc('ir' or 'pk')
from_cc(expr)	The source country based on IP address from_cc('ir' or 'pk')
to_cc(expr)	The destination country based on IP address to_cc('ir' or 'pk')
protocol(expr)	The textual form of the IP next protocol. protocol('TCP')
next_protocol(expr)	The textual form of the IP next protocol. ip_next_protocol('17')
mac_address(expr)	The MAC address of the target network device. mac_address('00:16:3E:3F:BD:EF')

Communication Based Contexts

email_body(expr)	The UTF-8 normalized text of all email bodies. email_body('how to' and 'build' and ('bomb' or 'weapon'))
chat_body(expr)	The UTF-8 normalized text of all chat bodies. chat_body('how to' and 'build' and ('bomb' or 'weapon'))
document_body(expr)	The UTF-8 normalized text of the Office document. – Office documents include (but are not limited to) Microsoft Office, Open Office, Google Docs and Spreadsheets. document_body('how to' and 'build' and ('bomb' or 'weapon'))
calendar_body(expr)	The UTF-8 normalized text of all calendars. An example is Google Calendar. calendar_body('wedding')
archive_files(expr)	Matches a list of files from within an archive. For example is a ZIP file is transmitted, all names of files within are passed to this context. archive_files('bad.dll' or 'virus.doc')
http_post_body(expr)	The UTF-8 normalized text HTTP url-encoded POSTs. http_post_body('action=send' and 'badguy@yahoo')

Communication Based Contexts Aliases

doc_email_body(expr)	This covers the email_body and document_body contexts doc_email_body('how to' and 'build' and ('bomb' or 'weapon'))
communication_body(expr)	This covers the email_body, document_body and chat_body contexts chat_body('how to' and 'build' and ('bomb' or 'weapon'))

A guide to XKS contexts can be found here

Context sensitivity

Why use context-sensitive scanning?

- More intuitive - you can say what you mean
- More accurate - if 'maps.google.com' is mentioned in a blog post, you don't want to try processing it as a Google Maps session
- Better performance for XKEYSCORE

Examples

- "I want to look for people doing web searches on Jihad from Kabul"

- Using the from_city() and web_search() context this becomes

fingerprint('demo/scenario3') =
    from_city('kabul') and web_search('jihad');

Examples

- "I want to look for people using Mojahedeen Secrets encryption from an IPhone"

- You can even use existing fingerprints in a fingerprint definition! So this becomes:

fingerprint('demo/scenario4') =
fingerprint('encryption/mojahdeen2' and
fingerprint('browser/cellphone/iphone')

Example 4

- "I want to look for E-mails that mention words from various categories of interest to CP"

- You can use multiple variables in an equation like this:

```
topic('wmd/acw/govtorgs') =
    email_body($acwitems and $acwpositions and
($acwcountries or $acwbrokers or $acwports));
```

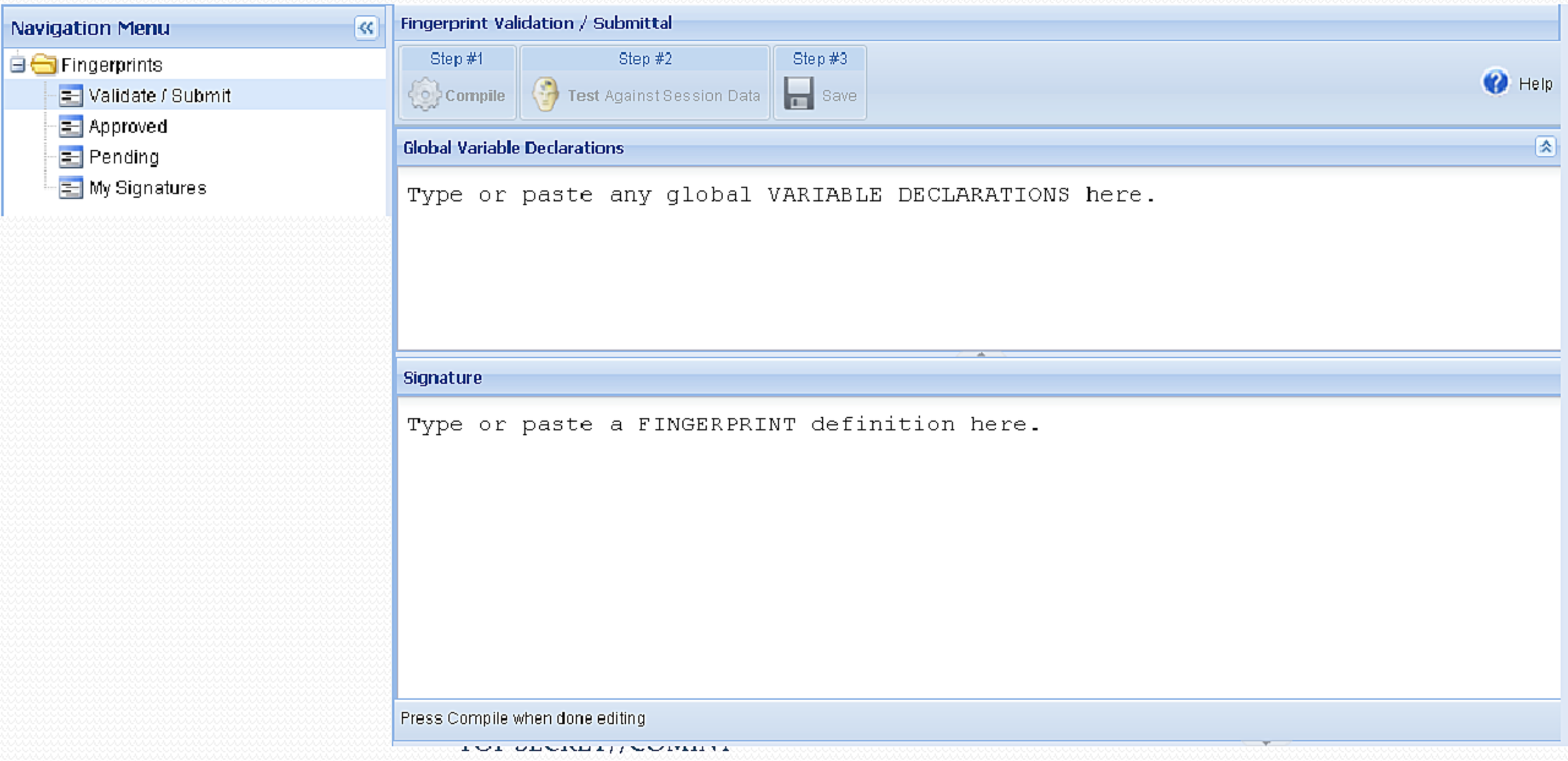
Example 4

- $acwitems = ‘machine gun’ or ‘grenade’ or ‘AK 47’
- $acwpositions = ‘minister of defence’ or ‘defense minister’
- $acwcountries = ‘somalia’ or ‘liberia’ or ‘sudan’
- $acwbrokers = ‘south africa’ or ‘serbia’ or ‘bulgaria’
- $acwports = ‘rangood’ or ‘albasra’ or ‘dar es salam’

```
topic('wmd/acw/govtorgs') =
    email_body($acwitems and $acwpositions and
  ($acwcountries or $acwbrokers or $acwports));
```

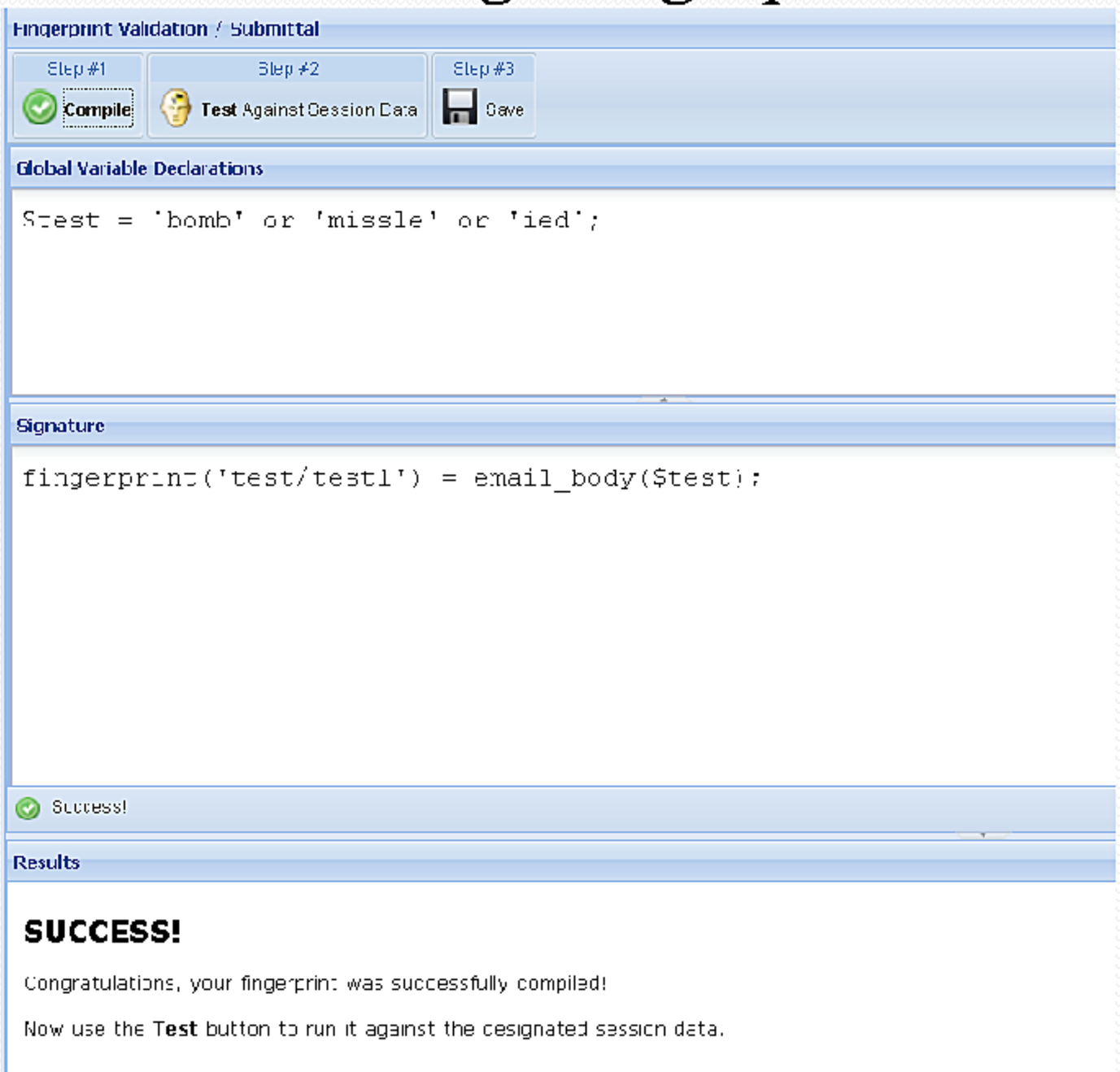
New Fingerprint GUI

- New XKS Fingerprint GUI allows analysts to directly test, submit and manage fingerprints through the web



New Fingerprint GUI

- New XKS Fingerprint GUI allows analysts to directly test, submit and manage fingerprints through the web



Questions?

Syntax Rules

- The definition of the fingerprint will look like this:

fingerprint('test/blah/something', owner = '[redacted]') =

Note the single quotes needed for the fingerprint name and owner

Syntax Rules

- Secondly every fingerprint definition must be completed by a semi-colon.

fingerprint('test/blah/something', owner = '[redacted]') =
'badguy' ;

Syntax Rules

- Variables also must be completed by a semi-colon.

```
$badguy =
  'bomb' or 'gun' or 'weapon' ;
fingerprint('test/blah/something', owner = '[redacted]') =
                $badguy;
```

Syntax Rules

- Definitions and Variables can span multiple lines

```
$badguy =
  'bomb' or
   'gun' or
   'weapon' ;
fingerprint('test/blah/something', owner = '[redacted]') =
                $badguy;
```