Hey folks,

There have been a few scenarios in which the following pcre has been applied to a rule -> "/(?:^|\.)google.com$/" and I was curious to know if we can make this a keyword for DNS signatures.

The idea behind this is that when we have a domain, ex. ‘google.com’, we want to match on either ‘google.com’ or ‘subdomain.google.com’ but not something such as ‘agoogle.com’.

Example rule structure with the new keyword:

dns_query; content:"google.com"; full_domain; pcre:"/(?:^|\.)google.com$/";

google.com - Match
hello.google.com - Match
hey.agoogle.com - No match
agoogle.com - No match

I think this can be easily implemented using the transform API.

Jeff, could you implement 2 transforms:

- **tld** -> this will transform an input buffer of example.com to just com
- **domain** -> this will transform an input buffer of subdomain.example.com to just example.com

The rule would look like:

```
alert dns any any -> any any (dns.query; tld; content:"com"; sid:1;)
alert http any any -> any any (http.host; domain; content:"example.com"; startswith; sid:2;)
```

Have a look at the exist transforms for how to use this API: strip_whitespace, to_md5, etc.
Country code TLDs consist of a two letter country abbreviation, e.g., nl and are usually of the form .nl.co but not always (wit www.google.com.au).

The second level domain resides to the left of the TLD (generic or cc): foo.bar.co.jp would produce a domain of bar.co.jp

I suggest that the tld and domain transformation operate as follows:
- Recognize the TLD as the rightmost label.
- Amend the TLD to include the next most rightmost label if it consists of 2 characters.
- The domain is the label to the left of the TLD after these steps.

Examples:
1. example.com -> tld = com, domain = example.com
3. google.co.uk -> tld = co.uk, domain = google.co.uk
4. windows.update.microsoft.co.es -> tld = co.es, domain = microsoft.co.es

It was suggested that the TLD could be referred to as the domain_suffix or even dns.suffix.

#5 - 07/24/2019 06:03 AM - Victor Julien
I don't know why I didn't remember this yesterday, but in the NL the most popular news site is 'nu.nl'. In this case the tld is 'nl' and the 'domain' would be 'nu.nl'. So I'm afraid this simple heuristic approach is not going to be sufficient.

#6 - 07/24/2019 11:43 AM - Jeff Lucovsky
I think it would work with nu.nl with this modification
- Amend the TLD to include the next most rightmost label if it consists of 2 characters and there is another label to the left

#7 - 07/24/2019 01:06 PM - Victor Julien
But that would fail if there is a subdomain, right? E.g. video.nu.nl would become domain: video.nu.nl, tld: 'nu.nl'.

#8 - 07/28/2019 11:49 PM - Jeff Lucovsky
Regarding domains only, in a separate conversation, Victor suggested the following:

```
alert dns any any -> any any {dns_query; domain; content:".google.com"; sid:1;}
1. hello.google.com --> match
2. hey.agoogle.com --> no match
3. agoogle.com --> no match
4. something.google.com.au --> match
```

To restrict matches to a domain:

```
alert dns any any -> any any {dns_query; domain; content:".google.com"; endswith; sid:1;}
1. hello.google.com --> match
2. hey.agoogle.com --> no match
3. agoogle.com --> no match
4. something.google.com.au --> no match
```

#9 - 07/29/2019 10:52 AM - Peter Manev
I like it.

#10 - 07/29/2019 09:54 PM - Jeff Lucovsky
Regarding TLDs: there are many cases that cannot be covered reliably by heuristics and require stored data to validate/check whether something is a tld or not.

Perhaps a rule like alert dns any any -> any any {dns_query; content:".co.uk"; endswith; sid:1;} is sufficient in place of a non-data augmented TLD transformation?

#11 - 07/30/2019 01:19 PM - Peter Manev
- File Screenshot from 2019-07-30 14-17-37.png added

Would that work in such similar cases? (please see attached)

#12 - 08/17/2019 03:13 PM - Jeff Lucovsky
The buffer ‘time.windows.com’ would be replaced with `.time.windows.com`
#13 - 09/16/2019 08:26 AM - Victor Julien
- Status changed from Assigned to Closed
- Difficulty deleted (low)

https://github.com/OISF/suricata/pull/4190

Files

<table>
<thead>
<tr>
<th>File Name</th>
<th>Size</th>
<th>Date</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screenshot from 2019-07-30 14-17-37.png</td>
<td>80.5 KB</td>
<td>07/30/2019</td>
<td>Peter Manev</td>
</tr>
</tbody>
</table>