When a signature contains a semicolon ";" inside a PCRE keyword match, an error occurs.

```
alert ip any any -> any any (msg:"semicolon in pcre"; content:"value:"; pcre:"/value:[0-9]+; test/" sid:1; rev:1;)
```

History

#1 - 11/23/2016 12:50 PM - Andreas Moe

I see, that this also occurs when having a semicolon in the "msg" field, and a "content" field. As far as documentation i have come across (both with regard to PCRE and Suricata) ";" is not a special character that should need to be escaped. Correct me if i'm wrong.

#2 - 11/23/2016 02:24 PM - Jason Ish

- Assignee set to Jason Ish

First note that your rule is missing a ';' between the pcre option and sid.

Anyways, I think this is a documentation issue. Semi-colons must be escaped in msg and content like they do in Snort, and it looks like this applies to pcre's as well. I'll try to come up with something generic in the documentation, and then explicitly mention it in the msg, content and pcre options like it is done in the Snort manual.

#3 - 11/23/2016 03:50 PM - Andreas Moe

Ah, yes missed that missing semicolon, the provided rule was just to quickly test / provide documentation for a previously encountered scenario. When running the same rule with a semicolon after the PCRE, it gives the same error.

Would it not be possible to check if this semicolon is confined to within quotes for a keyword and or within a PCRE? Seems that there is something within the splitting of keywords and parsing of the signature that prematurely splits the signature into different "tokens" / substrings.

I haven't tested, but does this f.ex also affect other keywords that contain quoted strings: "replace", "filename", "fileext", "filenmagic", "tls.subject", "tls.issuerdn". And possibly more.

#4 - 11/23/2016 03:59 PM - Jason Ish

Andreas Moe wrote:
Ah, yes missed that missing semicolon, the provided rule was just to quickly test / provide documentation for a previously encountered scenario. When running the same rule with a semicolon after the PCRE, it gives the same error.

[...]

Would it not be possible to check if this semicolon is confined to within quotes for a keyword and or within a PCRE? Seems that there is something within the splitting of keywords and parsing of the signature that prematurely splits the signature into different "tokens" / substrings.

I haven’t tested, but does this f.ex also affect other keywords that contain quoted strings: "replace", "filename", "fileext", "filemagic", "tls.subject", "tls.issuerdn". And possibly more.

This will affect all those options. Honestly I'm not sure if we could handle this case without breaking all the rules out there that already do proper escaping.

#5 - 11/28/2016 03:55 PM - Jason Ish
- Status changed from New to Closed
- Target version set to 3.2

Merged into master. See https://github.com/inliniac/suricata/pull/2422