Suricata - Bug #2555
Ensure strings in eve-log are json-encodable
07/26/2018 12:52 PM - Maurizio Abba

<table>
<thead>
<tr>
<th>Status:</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority:</td>
<td>Normal</td>
</tr>
<tr>
<td>Assignee:</td>
<td>Maurizio Abba</td>
</tr>
<tr>
<td>Category:</td>
<td></td>
</tr>
<tr>
<td>Target version:</td>
<td>4.1rc2</td>
</tr>
<tr>
<td>Affected Versions:</td>
<td>4.0beta1</td>
</tr>
<tr>
<td>Difficulty:</td>
<td>low</td>
</tr>
<tr>
<td>Effort:</td>
<td>medium</td>
</tr>
</tbody>
</table>

Description
Halfway between bug and feature.

In our current code any binary data would make function json_string fail. In case of failure, this function return NULL.
It would be cool to add a check before trying to dump data that, in case of binary data being sent, would just hex-encode the result before sending it to json_string.

The idea would be to create a SCJsonString that takes as parameter a const char * and check the return value of json_string, returning it if != NULL.
Otherwise, for every character belonging to the string it would apply the following transformation:
if the character is printable -> keep it as it is
if the character is NOT printable -> transform it in its hex-encoded version.

Note that this transformation would force the encoding of any non json-encodable string to ASCII.

There is one more drawback: for sake of saving dynamic allocation, we will need to hardcode a max boundary on the size of the final buffer after the transformation: the buffer can be limited to 2048 (a string bigger than 2048) would be suspicious. Note that this limitation DOES NOT apply to any string that can't have binary data (such as base64 encoding of TLS certificate).

Attached, an example pcap with a file transfer with binary data in the user agent.
The resulting http.log eve line is

```
{
    "timestamp": "2015-03-25T15:40:41.558505+0000",
    "flow_id": 149850720655109,
    "pcap_cnt": 7,
    "event_type": "http",
    "src_ip": "192.168.1.44",
    "src_port": 58068,
    "dest_ip": "127.0.0.1",
    "dest_port": 80,
    "proto": "TCP",
    "tx_id": 0,
    "http": {
        "hostname": "ev0lve.org",
        "url": "/a/expl/2008.mp4",
        "http_content_type": "video/mp4",
        "http_method": "GET",
        "protocol": "HTTP/1.1",
        "status": 200,
        "length": 195
    }
}
```

with the modification it would become

```
{
    "timestamp": "2015-03-25T15:40:41.558505+0000",
    "flow_id": 67447478114053,
}
```
There is also a small change in the rust code. It's actually not definitive, and something where I would like to have some inputs. In the current code the data would just be returned as ASCII, losing any utf-8 encoding capability provided by json.
I think we should change this behavior, but I don't have any good idea other than just calling the SCJsonString function from json.rs deleting the current to_cstring code.

What do you guys think?

---

**History**

#1 - 08/03/2018 08:48 AM - Maurizio Abba
- Status changed from New to Resolved

#2 - 08/03/2018 08:50 AM - Victor Julien
- Status changed from Resolved to Closed
- Target version set to 4.1rc2

[https://github.com/OISF/suricata/pull/3437](https://github.com/OISF/suricata/pull/3437)

**Files**

<table>
<thead>
<tr>
<th>File Name</th>
<th>Size</th>
<th>Created Date</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008.mp4-binary-ua.pcap</td>
<td>1.48 KB</td>
<td>07/25/2018</td>
<td>Maurizio Abba</td>
</tr>
</tbody>
</table>