

Identifying Validator Requests

Mark Taylor (Bristol)

based on ideas of Alberto Micol (ESO)

IVOA Interop
College Park MD, USA

November 2018

`$Id: valid-id.tex,v 1.4 2018/11/08 13:40:45 mbt Exp $`

The Problem

- Validators are good!
- But they can mess up your usage statistics

Date: Thu, 4 Oct 2018 16:20:36 +0000

From: Alberto Micol <amicol@eso.org>

To: Mark Taylor <M.B.Taylor@bristol.ac.uk>, Renaud Savalle <renaud.savalle@obspm.fr>

Subject: Re: User agent for validators - Re: V0 Paris validator of ESO V0 services

...

At the moment I see 4 V0 Validators regularly checking the ESO TAP.

Now, for the time being those 4 IPs account for 95% of the entire TAP traffic...

So, it is very important to remove them from the stats.

But I need a handy way to recognise them.

The Solution

HTTP User-Agent header

- The User-Agent field in HTTP requests can contain information identifying the client
- RFC 2616 Section 14.43:

14.43 User-Agent

The User-Agent request-header field contains information about the user agent originating the request. This is for statistical purposes, the tracing of protocol violations, and automated recognition of user agents for the sake of tailoring responses to avoid particular user agent limitations. User agents SHOULD include this field with requests.

- Several VO clients already do this
- This can help, but doesn't actually tell you whether a client is a validator or a real user
 - ▶ E.g. STILTS identifies itself like
"User-Agent: STILTS/3.1-4 Java/1.8.0_181"
— but that could be a user TAP query or the taplint validator
- Can we have some standard way to identify validators from User-Agent?

Details

User-Agent rules (RFC 2616 sections 4.43, 3.8, 2.2):

The field can contain multiple product tokens (section 3.8) and comments identifying the agent and any subproducts which form a significant part of the user agent. By convention, the product tokens are listed in order of their significance for identifying the application.

```
User-Agent      = "User-Agent" ":" 1*( product | comment )

product        = token [ "/" product-version ]
product-version = token

comment        = "(" *( ctext | quoted-pair | comment ) ")"
ctext          = <any TEXT excluding "(" and ">
```

Product tokens SHOULD be short and to the point. They MUST NOT be used for advertising or other non-essential information. Although any token character MAY appear in a product-version, this token SHOULD only be used for a version identifier (i.e., successive versions of the same product SHOULD only differ in the product-version portion of the product value).

- i.e. you can add (anything in brackets) as long as it's *“short and to the point”*
 - ▷ e.g. STILTS/3.1-4 (I am a validator) Java/1.8.0_181”

Proposals?

Propose some standardised content of User-Agent:

- Anything we propose should probably be alongside, not instead of, existing content
- Standard token?
(`IVOA-Validator`)
- Additional information like URL of validation results?
(`IVOA-Validator https://heasarc.gsfc.nasa.gov/vo/validation/vresults.pl?show=details&sid=34403&runid=1177476&switch=no`)
- Other ideas?

Next steps

- Informal agreement?
- IVOA Note?