Sec Bug #77423 FILTER_VALIDATE_URL accepts URLs with invalid userinfo

Submitted: 2019-01-07 10:16 UTC **Modified:** 2021-02-15 10:28 UTC
From: jifan dot jf at alibaba-inc dot com **Assigned:** stas (profile)
Status: Closed **Package:** *URL Functions
PHP Version: 5.6.39 **OS:** linux
Private report: No **CVE-ID:** 2020-7071

View | Add Comment | Developer | Edit

[2019-01-07 10:16 UTC] jifan dot jf at alibaba-inc dot com

```
Description:
------------
---
From manual page: https://php.net/datetime.formats.time
---


Test script:
---------------
<?php
        $a = "http://php.net\@aliyun.com/aaa.do";
        $b = parse_url($a);
        var_dump($b);
?>

use "str_replace("\","/",$b);" to fix it

Expected result:
----------------
for security, we will filter the host before using redirect($url) .
we'll test if the host of the url to redirect is in our whitelist(e.g. aliyun.com).
how ever, a attacker will use "http://php.net\@aliyun.com/" to bypass the filter.
-------
parse_url("http://php.net\@aliyun.com/aaa.do");
["host"]=> string(10) "aliyun.com"

-------
parse_url("http://php.net/@aliyun.com/aaa.do");
["host"]=> string(7) "php.net"

----------------
then we'll export the url to the webpage as "<a href="http://php.net\@aliyun.com/">" aaa</a>

then click the button, browser will open the "php.net".
```

Patches

Add a Patch

Pull Requests

Add a Pull Request

History

All | Comments | Changes | Git/SVN commits | Related reports

[2019-03-13 09:53 UTC] cmb@php.net

Related To: Bug #77730

[2020-02-13 14:37 UTC] cmb@php.net

```
-Status: Open
+Status: Duplicate
```

[2020-02-13 14:44 UTC] cmb@php.net

```
-Status: Duplicate
+Status: Open
```

[2020-02-13 14:50 UTC] cmb@php.net

Related To: Bug #77730

[2020-02-13 16:06 UTC] cmb@php.net

```
-Assigned To:
+Assigned To: stas
```

[2020-02-13 16:06 UTC] cmb@php.net

Given that the documentation of parse_url()[1] states

```
| This function is not meant to validate the given URL, it only
| breaks it up into the above listed parts. Partial URLs are also
| accepted, parse_url() tries its best to parse them correctly.
```

I don't think this qualifies as security issue. However, given that filter_var($a, FILTER_VALIDATE_URL) returns the URL (instead of false), I think there is a security issue in php_url_parse_ex().

A possible fix would be recognize the userinfo only if it is valid

according to RFC 3986[2], which could look like
<https://gist.github.com/cmb69/180b63eea1c5163cb29440421007ba4a>
(this has been developed against PHP-7.4).

Stas, what do you think about this?

[1] <https://www.php.net/parse-url>
[2] <https://tools.ietf.org/html/rfc3986#appendix-A>

[2020-05-11 20:44 UTC] stas@php.net

```
-Summary: parse_url() will deliver a wrong host to user
+Summary: FILTER_VALIDATE_URL accepts URLs with invalid userinfo
```

[2020-05-11 20:44 UTC] stas@php.net

I think if userinfo is invalid FILTER_VALIDATE_URL definitely should not accept it.

[2020-05-11 20:45 UTC] stas@php.net

We'd probably need a fix for this from 7.1 up.

[2020-05-11 20:47 UTC] stas@php.net

Oops, I meant 7.2 of course not 7.1

[2020-05-13 07:46 UTC] cmb@php.net

Ah, right! Patch for PHP-7.2:
<https://gist.github.com/cmb69/449e29fa609cf2fd6b617cf5f41638e3>.

[2021-01-03 01:48 UTC] stas@php.net

```
-CVE-ID:
+CVE-ID: 2020-7071
```

[2021-01-04 09:14 UTC] stas@php.net

Automatic comment on behalf of cmbecker69@gmx.de
Revision: http://git.php.net/?p=php-src.git;a=commit;h=2d3d72412a6734e19a38ed10f385227a6238e4a6
Log: Fix #77423: parse_url() will deliver a wrong host to user

[2021-01-04 09:14 UTC] stas@php.net

```
-Status: Assigned
+Status: Closed
```

[2021-01-04 09:15 UTC] stas@php.net

Automatic comment on behalf of cmbecker69@gmx.de
Revision: http://git.php.net/?p=php-src.git;a=commit;h=2d3d72412a6734e19a38ed10f385227a6238e4a6
Log: Fix #77423: parse_url() will deliver a wrong host to user

[2021-01-04 09:15 UTC] stas@php.net

Automatic comment on behalf of cmbecker69@gmx.de
Revision: http://git.php.net/?p=php-src.git;a=commit;h=2d3d72412a6734e19a38ed10f385227a6238e4a6
Log: Fix #77423: parse_url() will deliver a wrong host to user

[2021-01-04 09:19 UTC] stas@php.net

Automatic comment on behalf of cmbecker69@gmx.de
Revision: http://git.php.net/?p=php-src.git;a=commit;h=b132da7f9df39c1774997f21016c522b676a6ab0
Log: Fix #77423: parse_url() will deliver a wrong host to user

[2021-01-04 09:20 UTC] stas@php.net

Automatic comment on behalf of cmbecker69@gmx.de
Revision: http://git.php.net/?p=php-src.git;a=commit;h=5174de7cd33c3d4fa591c9c93859ff9989b07e8c
Log: Fix #77423: parse_url() will deliver a wrong host to user

[2021-01-04 09:48 UTC] stas@php.net

Automatic comment on behalf of cmbecker69@gmx.de
Revision: http://git.php.net/?p=php-src.git;a=commit;h=2d3d72412a6734e19a38ed10f385227a6238e4a6
Log: Fix #77423: parse_url() will deliver a wrong host to user

[2021-01-04 09:54 UTC] stas@php.net

Automatic comment on behalf of cmbecker69@gmx.de
Revision: http://git.php.net/?p=php-src.git;a=commit;h=b132da7f9df39c1774997f21016c522b676a6ab0
Log: Fix #77423: parse_url() will deliver a wrong host to user

[2021-01-04 09:54 UTC] stas@php.net

Automatic comment on behalf of cmbecker69@gmx.de
Revision: http://git.php.net/?p=php-src.git;a=commit;h=2d3d72412a6734e19a38ed10f385227a6238e4a6
Log: Fix #77423: parse_url() will deliver a wrong host to user

[2021-01-04 10:49 UTC] cmb@php.net

Automatic comment on behalf of stas
Revision: http://git.php.net/?p=php-src.git;a=commit;h=d4f5aed22193106271510efd643ba8f349b7d85f
Log: Fix #77423: parse_url() will deliver a wrong host to user

[2021-01-05 13:31 UTC] carusogabriel@php.net

Automatic comment on behalf of cmbecker69@gmx.de
Revision: http://git.php.net/?p=php-src.git;a=commit;h=b7f837381ef642d7fb369bfd0069e7525d4c22ea
Log: Fix #77423: parse_url() will deliver a wrong host to user

[2021-01-27 07:52 UTC] gkamathe at redhat dot com

Hello,

Have we considered a case where the input has only "@" within the URL instead of "\@". A sample use case is below on latest version of PHP, where the hostname being returned is still incorrect (malicious).

```
$ php -v
PHP 7.4.14 (cli) (built: Jan  5 2021 10:45:06) ( NTS )
Copyright (c) The PHP Group
Zend Engine v3.4.0, Copyright (c) Zend Technologies
    with Zend OPcache v7.4.14, Copyright (c), by Zend Technologies
$
$ rpm -qa | grep -i php-7
php-7.4.14-1.fc33.x86_64
$
$ cat test.php
<?php
        #$a = "http://php.net\@aliyun.com/aaa.do";
        $a = "http://php.net@aliyun.com/aaa.do";
        var_dump($a);
        $b = parse_url($a);
        var_dump($b);
?>
$
$ php -f test.php
string(32) "http://php.net@aliyun.com/aaa.do"
array(4) {
  ["scheme"]=>
  string(4) "http"
  ["host"]=>
  string(10) "aliyun.com"     <<<<<<<<
  ["user"]=>
  string(7) "php.net"
  ["path"]=>
  string(7) "/aaa.do"
}
$
```

[2021-01-27 08:05 UTC] stas@php.net

The hostname returned in this case is correct. Also, please don't comment on the issues' page about different issues, it makes it very hard to track things.

[2021-02-01 17:25 UTC] ryan at association dot drupal dot org

Was this intended to be merged into the 7.2 branch? I thought that was out of support now?

[2021-02-15 09:29 UTC] tomasnorre at gmail dot com

It looks like there is a regression regarding this fix.

https://3v4l.org/hGkOj

The fix is in: 7.3.26, 7.4.14 and 8.0.1 as wanted, but it's gone again in: 7.3.27, 7.4.15 and 8.0.2

[2021-02-15 10:28 UTC] stas@php.net

The change to parse_url was intentionally reversed. The change in filter_url still there.

[2021-02-15 10:31 UTC] tomasnorre at gmail dot com

Thanks for the info.