» Issue36216



This issue tracker **has been migrated to GitHub**, and is currently **read-only**.
For more information, **see the GitHub FAQs in the Python's Developer Guide.**



This issue has been migrated to GitHub:
https://github.com/python/cpython/issues/80397

classification

Title:	CVE-2019-9636: urlsplit does not handle NFKC normalization		
Type:	security	**Stage**:	resolved
Components:	Unicode	**Versions**:	Python 3.8, Python 3.7, Python 3.6

process

Status:	closed	**Resolution**:	fixed
Dependencies:		**Superseder**:	
Assigned To:	steve.dower	**Nosy List**:	benjamin.peterson, ezio.melotti, jeremy.kloth, jkloth, koobs, martin.panter, mcepl, ned.deily, steve.dower, vstinner, xtreak
Priority:	normal	**Keywords**:	patch, security_issue

Created on **2019-03-06 17:37** by **steve.dower**, last changed **2022-04-11 14:59** by **admin**. This issue is now **closed**.

Pull Requests

URL	Status	Linked	Edit
PR 12201	closed	steve.dower, 2019-03-06 17:45	
PR 12213	merged	steve.dower, 2019-03-07 16:07	
PR 12215	merged	steve.dower, 2019-03-07 16:11	
PR 12216	merged	steve.dower, 2019-03-07 16:29	
PR 12223	merged	steve.dower, 2019-03-07 18:31	
PR 12224	merged	steve.dower, 2019-03-07 18:34	
PR 12291	merged	steve.dower, 2019-03-12 15:27	

Messages (22)

msg337336 - (view) **Author: Steve Dower (steve.dower) *** **Date: 2019-03-06 17:37**

```
URLs encoded with Punycode/IDNA use NFKC normalization to decompose characters
[1]. This can result in some characters introducing new segments into a URL.

For example, \uFF03 is not equal to '#' under direct comparison, but normalizes
```

to '#' which changes the fragment part of the URL. Similarly \u2100 normalizes to 'a/c' which introduces a path segment.

Currently, urlsplit() does not normalize, which may result in it returning a different netloc from what a browser would

```
>>> u = "https://example.com\uFF03@bing.com"
>>> urlsplit(u).netloc.rpartition("@")[2]
bing.com

>>> # Simulate
>>> u = "https://example.com\uFF03@bing.com".encode("idna").decode("ascii")
>>> urlsplit(u).netloc.rpartition("@")[2]
example.com
```

(Note that .netloc includes user/pass and .rpartition("@") is often used to remove it.)

This may be used to steal cookies or authentication data from applications that use the netloc to cache or retrieve this information.

The preferred fix for the urllib module is to detect and raise ValueError if NFKC-normalization of the netloc introduce any of '/?#@:'. Applications that want to avoid this error should perform their own decomposition using unicodedata or transcode to ASCII via IDNA.

```
>>> # New behavior
>>> u = "https://example.com\uFF03@bing.com"
>>> urlsplit(u)
...
ValueError: netloc 'example.com#@bing.com' contains invalid characters under NFKC normalization

>>> # Workaround 1
>>> u2 = unicodedata.normalize("NFKC", u)
>>> urlsplit(u2)
SplitResult(scheme='https', netloc='example.com', path='', query='', fragment='@bing.com')

>>> # Workaround 2
>>> u3 = u.encode("idna").decode("ascii")
>>> urlsplit(u3)
SplitResult(scheme='https', netloc='example.com', path='', query='', fragment='@bing.com')
```

Note that we do not address other characters, such as those that convert into period. The error is only raised for changes that affect how urlsplit() locates the netloc and the very common next step of removing credentials from the netloc.

This vulnerability was reported by Jonathan Birch of Microsoft Corporation and Panayiotis Panayiotou (p.panayiotou2@gmail.com) via the Python Security Response Team. A CVE number has been requested.

[1]: https://unicode.org/reports/tr46/

msg337383 - (view) **Author: STINNER Victor (vstinner) *** **Date: 2019-03-07 10:35**

FYI I added this vulnerability to:
https://python-security.readthedocs.io/vuln/urlsplit-nfkc-normalization.html

msg337398 - (view) **Author: Steve Dower (steve.dower) *** **Date: 2019-03-07 16:02**

New changeset **16e6f7dee7f02bb81aa6b385b982dcdda5b99286** by Steve Dower in branch 'master':
~~**bpo-36216**~~: Add check for characters in netloc that normalize to separators (~~**GH-12201**~~)
https://github.com/python/cpython/commit/16e6f7dee7f02bb81aa6b385b982dcdda5b99286

msg337411 - (view) **Author: Steve Dower (steve.dower) *** **Date: 2019-03-07 17:08**

New changeset **daad2c482c91de32d8305abbccc76a5de8b3a8be** by Steve Dower in branch '3.7':
~~**bpo-36216**~~: Add check for characters in netloc that normalize to separators (~~**GH-12201**~~)
https://github.com/python/cpython/commit/daad2c482c91de32d8305abbccc76a5de8b3a8be

msg337412 - (view) **Author: Steve Dower (steve.dower) *** **Date: 2019-03-07 17:08**

New changeset **e37ef41289b77e0f0bb9a6aedb0360664c55bdd5** by Steve Dower in branch '2.7':
~~**bpo-36216**~~: Add check for characters in netloc that normalize to separators (~~**GH-12201**~~)
https://github.com/python/cpython/commit/e37ef41289b77e0f0bb9a6aedb0360664c55bdd5

msg337532 - (view) **Author: Steve Dower (steve.dower) *** **Date: 2019-03-08 21:29**

This issue is now assigned CVE-2019-9636

https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2019-9636

msg337566 - (view) **Author: Jeremy Kloth (jkloth) *** **Date: 2019-03-09 09:49**

A missed print statement in the 2.7 patch is causing the tests to fail.

Line 647 of **Lib/test/test_urlparse.py**

msg337645 - (view) **Author: Larry Hastings (larry) *** **Date: 2019-03-11 04:58**

New changeset **62d36547f97210a26cc6051da78714fd078e158c** by larryhastings (Steve Dower) in branch '3.4':
~~**bpo-36216**~~: Add check for characters in netloc that normalize to separators (~~**GH-12201**~~) (~~**#12224**~~)
https://github.com/python/cpython/commit/62d36547f97210a26cc6051da78714fd078e158c

msg337646 - (view) **Author: Larry Hastings (larry) *** **Date: 2019-03-11 04:59**

New changeset **c0d95113b070799679bcb9dc49d4960d82e8bb08** by larryhastings (Steve Dower) in branch '3.5':
~~**bpo-36216**~~: Add check for characters in netloc that normalize to separators (~~**GH-**~~

~~12201~~) (~~#12223~~)
https://github.com/python/cpython/commit/c0d95113b070799679bcb9dc49d4960d82e8bb08

msg337704 - (view) **Author: Steve Dower (steve.dower) *** **Date: 2019-03-11 22:23**

> A missed print statement in the 2.7 patch is causing the tests to fail.

How does that cause tests to fail? Is it going to stderr? Or just causing an error.

msg337711 - (view) **Author: Jeremy Kloth (jeremy.kloth)** **Date: 2019-03-12 02:20**

>
> How does that cause tests to fail? Is it going to stderr? Or just causing
> an error.
>

It is causing an "unexpected output error". When the test is re-run at the end, it is run in verbose mode so the extra output is ignored and thus passes at that point.

>

msg337720 - (view) **Author: Ned Deily (ned.deily) *** **Date: 2019-03-12 04:34**

New changeset 23fc0416454c4ad5b9b23d520fbe6d89be3efc24 by Ned Deily (Steve Dower) in branch '3.6':
[3.6] ~~bpo-36216~~: Add check for characters in netloc that normalize to separators (~~GH-12201~~) (~~GH-12215~~)
https://github.com/python/cpython/commit/23fc0416454c4ad5b9b23d520fbe6d89be3efc24

msg337725 - (view) **Author: Karthikeyan Singaravelan (xtreak) *** **Date: 2019-03-12 08:17**

Sample buildbot log of print statement in testcase causing rerun of test :
https://buildbot.python.org/all/#/builders/101/builds/364/steps/4/logs/stdio

msg337753 - (view) **Author: Steve Dower (steve.dower) *** **Date: 2019-03-12 15:34**

Thanks. Just posted a PR that puts the print behind a verbose flag (on Python 3 it uses subtest so that we see which character caused the failure).

If that doesn't work for whatever reason, I'll just leave it out and hope that we never have to debug it on Python 2 :)

msg337771 - (view) **Author: STINNER Victor (vstinner) *** **Date: 2019-03-12 16:09**

Commit e37ef41289b77e0f0bb9a6aedb0360664c55bdd5 introduced a regression on AMD64 Windows8.1 Refleaks 2.7:

test_urlparse leaked [114, 114, 114] references, sum=342

https://buildbot.python.org/all/#/builders/33/builds/532

It is fixed by ~~PR 12291~~.

You can test:

python -m test -R 3:3 test_urlparse

msg337773 - (view) **Author: Steve Dower (steve.dower) *** **Date: 2019-03-12 16:16**

If it's fixed by not printing to the console, then it must be a refleak in print. Or perhaps in the test failure/repeat.

That PR is going to be merged as soon as AppVeyor finishes, so nothing to worry about here.

msg337806 - (view) **Author: Steve Dower (steve.dower) *** **Date: 2019-03-12 20:52**

New changeset **507bd8cde60ced74d13a1ffa883bb9b0e73c38be** by Steve Dower in branch '2.7':
[3.7] ~~**bpo-36216**~~: Only print test messages when verbose (~~**GH-12291**~~)
https://github.com/python/cpython/commit/507bd8cde60ced74d13a1ffa883bb9b0e73c38be

msg337811 - (view) **Author: Steve Dower (steve.dower) *** **Date: 2019-03-12 21:25**

Hah, I typo'd the commit message and marked it as 3.7 instead of 2.7. Oh well.

msg339391 - (view) **Author: Matej Cepl (mcepl) *** **Date: 2019-04-03 13:35**

I am trying to investigate the impact of this bug on Python 2.6 (yes, it is for SLE), and I have hard to replicate the steps in the description even on 2.7:

```
~$ ipython2
Python 2.7.15 (default, May 21 2018, 17:53:03) [GCC]
Type "copyright", "credits" or "license" for more information.

IPython 5.8.0 -- An enhanced Interactive Python.
?         -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help      -> Python's own help system.
object?   -> Details about 'object', use 'object??' for extra details.

In [1]: from urlparse import urlsplit

In [2]: u = "https://example.com\uFF03@bing.com".encode("idna").decode("ascii")

In [3]: u
Out[3]: u'https://example.com\\uFF03@bing.com'

In [4]: urlsplit(u).netloc.rpartition('@')[2]
Out[4]: u'bing.com'

In [5]: u = "https://example.com\uFF03@bing.com"

In [6]: urlsplit(u).netloc.rpartition('@')[2]
Out[6]: 'bing.com'
```

```
In [7]: u = u.encode("idna").decode("ascii")

In [8]: urlsplit(u).netloc.rpartition('@')[2]
Out[8]: u'bing.com'

In [9]: import unicodedata

In [10]: u2 = unicodedata.normalize('NFKC', u)

In [11]: u2
Out[11]: u'https://example.com\\uFF03@bing.com'

In [12]: urlsplit(u2)
Out[12]: SplitResult(scheme=u'https', netloc=u'example.com\\uFF03@bing.com',
path=u'', query='', fragment='')

In [13]:

Yes, the results are weird, and most likely they would break any software relying
on them, but I am not sure that it is a security issue.

vstinner ? steve.dower ? What do you think?
```

msg339428 - (view) **Author: Steve Dower (steve.dower) *** **Date: 2019-04-04 03:42**

```
You need a "u" prefix on some of your strings or they're probably being
immediately decomposed. The result of urlsplit should be unicode on Python 2 for
a Unicode input, and yours are not.
```

msg339433 - (view) **Author: Matej Cepl (mcepl) *** **Date: 2019-04-04 05:31**

```
You are right. Thank you.
```

msg393450 - (view) **Author: Kubilay Kocak (koobs)** **Date: 2021-05-11 09:57**

```
Fix meta (not incl 2.7 which is no longer available to select).
```

History

Date	User	Action	Args
2022-04-11 14:59:12	admin	set	github: 80397
2021-05-11 14:28:00	larry	set	nosy: - **larry**
2021-05-11 09:57:21	koobs	set	nosy: **vstinner**, **larry**, **benjamin.peterson**, **jkloth**, **ned.deily**, **mcepl**, **ezio.melotti**, **jeremy.kloth**, **martin.panter**, **koobs**, **steve.dower**, **xtreak** messages: + **msg393450** components: + Unicode, - FreeBSD versions: + Python 3.7, Python 3.8
2021-05-11 05:51:48	dcockcn	set	nosy: + **koobs** components: + FreeBSD, - Unicode
2021-05-11 05:51:11	dcockcn	set	versions: - Python 2.7, Python 3.4, Python 3.5, Python 3.7, Python 3.8

2019-05-10 18:05:52	ned.deily	set	messages: - **msg342088**
2019-05-10 17:36:37	ned.deily	set	messages: + **msg342088**
2019-04-04 05:31:15	mcepl	set	messages: + **msg339433**
2019-04-04 03:42:50	steve.dower	set	messages: + **msg339428**
2019-04-03 13:35:37	mcepl	set	nosy: + **mcepl** messages: + **msg339391**
2019-03-14 12:11:56	vstinner	set	title: urlsplit does not handle NFKC normalization -> CVE-2019-9636: urlsplit does not handle NFKC normalization
2019-03-12 21:25:47	steve.dower	set	status: open -> closed resolution: fixed messages: + **msg337811** stage: patch review -> resolved
2019-03-12 20:52:03	steve.dower	set	messages: + **msg337806**
2019-03-12 16:16:04	steve.dower	set	messages: + **msg337773**
2019-03-12 16:09:04	vstinner	set	messages: + **msg337771**
2019-03-12 15:34:51	steve.dower	set	messages: + **msg337753**
2019-03-12 15:27:07	steve.dower	set	pull_requests: + **pull_request12272**
2019-03-12 08:17:20	xtreak	set	messages: + **msg337725**
2019-03-12 04:34:05	ned.deily	set	messages: + **msg337720**
2019-03-12 02:20:00	jeremy.kloth	set	nosy: + **jeremy.kloth** messages: + **msg337711**
2019-03-11 22:23:12	steve.dower	set	messages: + **msg337704**
2019-03-11 04:59:26	larry	set	messages: + **msg337646**
2019-03-11 04:58:50	larry	set	messages: + **msg337645**
2019-03-09 09:49:04	jkloth	set	nosy: + **jkloth** messages: + **msg337566**
2019-03-08 21:29:47	steve.dower	set	messages: + **msg337532**
2019-03-07 18:34:38	steve.dower	set	pull_requests: + **pull_request12214**
2019-03-07 18:31:25	steve.dower	set	pull_requests: + **pull_request12213**
2019-03-07 17:08:47	steve.dower	set	messages: + **msg337412**
2019-03-07 17:08:33	steve.dower	set	messages: + **msg337411**

2019-03-07 16:29:21	steve.dower	set	pull_requests: + **pull_request12207**
2019-03-07 16:11:00	steve.dower	set	pull_requests: + **pull_request12205**
2019-03-07 16:07:01	steve.dower	set	pull_requests: + **pull_request12203**
2019-03-07 16:02:31	steve.dower	set	messages: + **msg337398**
2019-03-07 10:35:26	vstinner	set	messages: + **msg337383**
2019-03-06 23:25:00	xtreak	set	nosy: + **martin.panter**, **xtreak**
2019-03-06 17:45:19	steve.dower	set	keywords: + **patch** stage: needs patch -> patch review pull_requests: + **pull_request12196**
2019-03-06 17:37:20	steve.dower	create	