New issue

[CVE-2025-0938] urlparse does not flag hostname *containing* [or] as incorrect #105704

Closed

Assignees

Labels stdlib type-bug

mjpieters opened on Jun 12, 2023 · edited by bedevere-app Edits Contributor

#103848 updated the URL parsing algorithm to handle IPv6 and IPvFuture addresses when parsing URLs.

However, the algorithm is incomplete. `[` and `]` are only permitted in the hostname portion **if they are the first and last characters** and only if they then contain an IPv6 or IPvFuture address. The current implementation ignores everything before the first `[` and everything after the first `]` found in the `netloc` portion.

The WhatWG URL standard states that `[` and `]` are forbidden characters in a hostname, and the host parser only looks for IPv6 or IPvFuture if the `[` and `]` characters are the first and last characters of the section, respectively.

The current implementation thus accepts such bizarre hostnames as:

- `http://prefix.[v1.example]/`
- `http://[v1.example].postfix/`

but then only reports the portion between the brackets as the hostname:

```
>>> urlparse('http://prefix.[v1.example]/').hostname
'v1.example'
>>> urlparse('http://[v1.example].postfix/').hostname
'v1.example'
```

The `.netloc` attribute, in both cases, contains the whole string.

Both URLs should have been rejected instead.

Your environment

- CPython versions tested on: 3.12.0b1
- Operating system and architecture: Darwin M1

Linked PRs

- gh-105704: Disallow IPv6 URLs with invalid prefix/suffix #111261
- gh-105704: Disallow square brackets (`[` and `]`) in domain names for parsed URLs #129418
- [3.13] gh-105704: Disallow square brackets (`[` and `]`) in domain names for parsed URLs (GH-129418) #129526
- [3.12] gh-105704: Disallow square brackets (`[` and `]`) in domain names for parsed URLs (GH-129418) #129527
- [3.11] gh-105704: Disallow square brackets (`[` and `]`) in domain names for parsed URLs (GH-129418) #129528
- [3.10] gh-105704: Disallow square brackets (`[` and `]`) in domain names for parsed URLs (GH-129418) #129529
- [3.9] gh-105704: Disallow square brackets (`[` and `]`) in domain names for parsed URLs (GH-129418) #129530

👍 6

mjpieters added **type-bug** on Jun 12, 2023

arhadthedev added **stdlib** on Jun 12, 2023

arhadthedev on Jun 12, 2023 Member

@**orsenthil** (as an urllib module expert)

csreddy98 on Jun 12, 2023 · edited by csreddy98 Edits

I think this should return a valid object only if the hostname starts with a `[` and ends with `]`. With the current logic any string with `[` and `]` inside the hostname is being considered as a valid IPv6 or IPvFutire hostnames. I believe we must verify the start and end characters of a hostname as mentioned here for IPv6 and optionally IPvFuture.

👍 1

orsenthil self-assigned this on Jun 13, 2023

lscottod on Jun 29, 2023 · edited by lscottod

Edits

Just to add to this thread,

I'm using django, django-environ and postgres. This issue is quite significant since I, unfortunately, happen to have brackets inside the postgres passwords in several prod/staging servers. This password is given inside a URL to django-environ that parses it (using urllib) then sends it back to django.

As far as Python 3.11.3, everything was going well, but since 3.11.4 it's all broken now.

It is related to what is stated above. urlsplit now spots '[' and ']' inside the netloc and what's inside theses brackets (a fragment of the password) is wrongfully considered a hostname. It then throws an exception since it tries to convert it to an ip address.

The lines that seem to be of importance :
in urllib/parse.py -- urlsplit()

```
if '[' in netloc and ']' in netloc:
      bracketed_host = netloc.partition('[')[2].partition(']')[0]
      _check_bracketed_host(bracketed_host)
```

It feels like an important breaking change/regression that doesn't seem documented

👍 5

pschoen-itsc on Jul 11, 2023

> Just to add to this thread,
>
> I'm using django, django-environ and postgres. This issue is quite significant since I, unfortunately, happen to have brackets inside the postgres passwords in several prod/staging servers. This password is given inside a URL to django-environ that parses it (using urllib) then sends it back to django.
>
> As far as Python 3.11.3, everything was going well, but since 3.11.4 it's all broken now.
>
> It is related to what is stated above. urlsplit now spots '[' and ']' inside the netloc and what's inside theses brackets (a fragment of the password) is wrongfully considered a hostname. It then throws an exception since it tries to convert it to an ip address.
>
> The lines that seem to be of importance : in urllib/parse.py -- urlsplit()

```
if '[' in netloc and ']' in netloc:
      bracketed_host = netloc.partition('[')[2].partition(']')[0]
      _check_bracketed_host(bracketed_host)
```

> It feels like an important breaking change/regression that doesn't seem documented

I experienced the same problem. Just to be sure I also checked the URL spec referenced in the code and it states that username and password can be an ASCII string, so this is clearly a bug.

bcail on Oct 23, 2023

@**orsenthil** Is it OK if I open a PR for this issue for you to look at?

orsenthil on Oct 23, 2023 Member

@**bcail** , yes please.

bcail added a commit that references this issue on Oct 24, 2023

`pythongh-105704: Disallow IPv6 URLs with invalid prefix/suffix` `c3bc170`

bedevere-app mentioned this on Oct 24, 2023

gh-105704: Disallow IPv6 URLs with invalid prefix/suffix #111261

bcail on Oct 24, 2023

@**orsenthil** I opened a PR.

bcail on Nov 29, 2023

Hi @**orsenthil**. Are you going to be able to take a look at my PR? If not, in a week or so I can post on Discourse and see if someone else can take a look. Thanks.

bcail added a commit that references this issue on Jan 25, 2024

`pythongh-105704: Disallow IPv6 URLs with invalid prefix/suffix` `ac55176`

bcail added a commit that references this issue on Nov 22, 2024

`pythongh-105704`: Disallow IPv6 URLs with invalid prefix/suffix `23e23d9`

66 remaining items

Load more

Metadata

Assignees

gpshead

orsenthil

Labels

stdlib **type-bug**

Projects

No projects

Milestone

No milestone

Relationships

None yet

Development

No branches or pull requests

Participants

+7