Issue43075

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/87241

classification

Title:	CVE-2021-3733: ReDoS in urllib.request		
Type:	security	**Stage**:	resolved
Components:	Library (Lib)	**Versions**:	Python 3.10, Python 3.9, Python 3.8, Python 3.7, Python 3.6

process

Status:	closed	**Resolution**:	fixed
Dependencies:		**Superseder**:	
Assigned To:		**Nosy List**:	crazybyte, lukasz.langa, miss-islington, ned.deily, orsenthil, serhiy.storchaka, vstinner, yetingli
Priority:	normal	**Keywords**:	easy, newcomer friendly, patch

Created on **2021-01-30 08:11** by **yetingli**, last changed **2022-04-11 14:59** by **admin**. This issue is now closed.

Files

File name	Uploaded	Description	Edit
redos_python.py	yetingli, 2021-01-30 08:11		
redos_python2.py	vstinner, 2021-04-07 11:25		

Pull Requests

URL	Status	Linked	Edit
PR 24391	merged	yetingli, 2021-01-31 05:12	
PR 25247	merged	miss-islington, 2021-04-07 11:28	
PR 25248	merged	miss-islington, 2021-04-07 11:28	
PR 25249	merged	miss-islington, 2021-04-07 11:29	
PR 25250	merged	miss-islington, 2021-04-07 11:29	

Messages (16)

msg385974 - (view) **Author: yeting li (yetingli) *** **Date: 2021-01-30 08:11**

Hi,

I find this regex '(?:^|,)[\t]*([^ \t]+)[\t]+' may be stucked by input.

The vulnerable regex is located in
https://github.com/python/cpython/blob/5c5a938573ce665f00e362c7766912d9b3f3b44e/Lib/urllib/request.py#L946

The ReDOS vulnerability of the regex is mainly due to the sub-pattern ',([^ \t]+)' and can be exploited with the following string
attack_str = "," * 10000

You can execute redos_python.py to reproduce the ReDos vulnerability.

I am willing to suggest that you replace '(?:^|,)[\t]*([^ \t]+)[\t]+' with '(?:^|,)[\t]*([^ \t,]+)[\t]+'

Looking forward for your response!

Best,
Yeting Li

msg385987 - (view) **Author: Serhiy Storchaka (serhiy.storchaka) *** **Date: 2021-01-30 16:32**

I agree. There is no catastrophic backtracking here (it was fixed in ~~issue39503~~), but the complexity of matching the regular expression is linear. Searching the pattern in a sequence of commas has quadratic complexity, because every step has linear complexity and we advance only one character at every attempt.

The proposed solution looks correct to me and fixes the issue. Yeting Li, do you mind to create a pull request for it? I can do it myself, but since you have found the problem and the solution, it would be better if the commit be attributed to you.

msg385993 - (view) **Author: Senthil Kumaran (orsenthil) *** **Date: 2021-01-30 18:59**

+1. The suggested fix looks good to me.

msg386009 - (view) **Author: yeting li (yetingli) *** **Date: 2021-01-31 05:44**

Thank you for your quick reply!

I agree. Catastrophic backtracking is typically regarded as a regex with exponential worst-case matching time. Besides regexes with exponential worst-case time complexity, ReDoS also includes ones with other super-linear (e.g., quadratic) worst-case time complexity.

Thanks again for your reply, I'm trying to create a pull request for it.

msg388358 - (view) **Author: STINNER Victor (vstinner) *** **Date: 2021-03-09 11:51**

I see that you attached a redos_python.py benchmark (which looks like a benchmark that I wrote recently ;-)) but you didn't give results. Can you please show that your fix is effective to avoid catastrophic performances?

Is this issue related to the CVE-2020-8492? Is it the same issue or is it different?
https://python-security.readthedocs.io/vuln/urllib-basic-auth-regex.html

msg388665 - (view) **Author: yeting li (yetingli) *** **Date: 2021-03-14 08:50**

Sorry for the delay. I analyzed the performance of the current version '(?:^|,)[\t]*([^ \t]+)[\t]+' and the fixed version '(?:^|,)[\t]*([^ \t,]+)[\t]+'. I ran the following HTTP header ten times:

```
header = '' + ',' * (10 ** 5)
```

The current version takes about 139.178s-140.946s, while the repaired version takes about 0.006s.

You can analyze them with the code below.

```
    from time import perf_counter
    for _ in range(0, 10):
        BEGIN = perf_counter()
        header = repeat_10_5_simple
        headers = Headers(header)
        handler.http_error_auth_reqed("WWW-Authenticate", host, req, Headers(header))
        DURATION = perf_counter() - BEGIN
        print(f"took {DURATION} seconds!")
```

For CVE-2020-8492, it is the backtracking performance caused by some ambiguity during the matching, and this issue is caused by the regex engine constantly moves the matching regex across the malicious string that does not have a match for the regex.

Because the locations of the vulnerabilities are the same, so I refer to your code. Thanks for the code ;-)!

msg390415 - (view) **Author: STINNER Victor (vstinner) *** **Date: 2021-04-07 10:59**

> header = '' + ',' * (10 ** 5)

I guess that a more generic protection against future attacks would be to limit the maximum length of a HTTP header. 100,000 characters for a HTTP Basic authentification does not sound reasonable.

But for now, let's fix the regex.

msg390417 - (view) **Author: yeting li (yetingli) *** **Date: 2021-04-07 11:14**

For a regex has polynomial worst-case complexity, limiting the maximum input length is indeed a very effective method.

As shown below, as the input length becomes smaller, the matching time becomes significantly smaller.

```
header = '' + ',' * (10 ** 4)    1.617s
header = '' + ',' * (10 ** 3)    0.014s
header = '' + ',' * (10 ** 2)    0.00017s
```

msg390419 - (view) **Author: STINNER Victor (vstinner) *** **Date: 2021-04-07 11:25**

redos_python2.py: Updated benchmark.

I confirm that **~~PR 24391~~** fix a worst case performance, starting with 100 characters.

Since the complexity is quadratic, strings longer 10^4 characters are likely to hang a client for several minutes.

```
== Reference (vulnerable) ==

simple: Mean +- std dev: 2.10 us +- 0.05 us
repeat 10: Mean +- std dev: 3.85 us +- 0.13 us
repeat 10^2: Mean +- std dev: 133 us +- 3 us
repeat 10^4: Mean +- std dev: 1.23 sec +- 0.05 sec
```

== With the **~~PR 24391~~** fix ==

```
simple: Mean +- std dev: 2.15 us +- 0.15 us
repeat 10: Mean +- std dev: 2.44 us +- 0.04 us
repeat 10^2: Mean +- std dev: 7.45 us +- 0.17 us
repeat 10^4: Mean +- std dev: 574 us +- 28 us

== Comparison ==

simple: Mean +- std dev: [ref] 2.10 us +- 0.05 us -> [fix] 2.15 us +- 0.15 us: 1.02x slower
repeat 10: Mean +- std dev: [ref] 3.85 us +- 0.13 us -> [fix] 2.44 us +- 0.04 us: 1.58x faster
repeat 10^2: Mean +- std dev: [ref] 133 us +- 3 us -> [fix] 7.45 us +- 0.17 us: 17.80x faster
repeat 10^4: Mean +- std dev: [ref] 1.23 sec +- 0.05 sec -> [fix] 574 us +- 28 us: 2152.36x faster

Geometric mean: 15.59x faster
```

msg390420 - (view) **Author: STINNER Victor (vstinner) *** **Date: 2021-04-07 11:27**

New changeset **7215d1ae25525c92b026166f9d5cac85fb1defe1** by Yeting Li in branch 'master':
~~bpo-43075~~: Fix ReDoS in urllib AbstractBasicAuthHandler (**~~GH-24391~~**)
https://github.com/python/cpython/commit/7215d1ae25525c92b026166f9d5cac85fb1defe1

msg390425 - (view) **Author: miss-islington (miss-islington)** **Date: 2021-04-07 11:45**

New changeset **e7654b6046090914a8323931ed759a94a5f85d60** by Miss Islington (bot) in branch '3.8':
~~bpo-43075~~: Fix ReDoS in urllib AbstractBasicAuthHandler (**~~GH-24391~~**)
https://github.com/python/cpython/commit/e7654b6046090914a8323931ed759a94a5f85d60

msg390441 - (view) Author: STINNER Victor (vstinner) * Date: 2021-04-07 15:58

New changeset a21d4fbd549ec9685068a113660553d7f80d9b09 by Miss Islington (bot) in branch '3.9':
~~bpo-43075~~: Fix ReDoS in urllib AbstractBasicAuthHandler (~~GH-24391~~) (~~GH-25247~~)
https://github.com/python/cpython/commit/a21d4fbd549ec9685068a113660553d7f80d9b09

msg392885 - (view) Author: Łukasz Langa (lukasz.langa) * Date: 2021-05-04 12:46

New changeset ada14995870abddc277addf57dd690a2af04c2da by Miss Islington (bot) in branch '3.7':
~~bpo-43075~~: Fix ReDoS in urllib AbstractBasicAuthHandler (~~GH-24391~~) (~~#25249~~)
https://github.com/python/cpython/commit/ada14995870abddc277addf57dd690a2af04c2da

msg393109 - (view) Author: Ned Deily (ned.deily) * Date: 2021-05-06 17:00

New changeset 3fbe96123aeb66664fa547a8f6022efa2dc8788f by Miss Islington (bot) in branch '3.6':
~~bpo-43075~~: Fix ReDoS in urllib AbstractBasicAuthHandler (~~GH-24391~~) (~~GH-25250~~)
https://github.com/python/cpython/commit/3fbe96123aeb66664fa547a8f6022efa2dc8788f

msg400130 - (view) Author: Gianluca Gabrielli (crazybyte) Date: 2021-08-23 11:13

RedHat has now assigned CVE-2021-3733 to this security bug.

msg401341 - (view) Author: STINNER Victor (vstinner) * Date: 2021-09-07 20:12

I created https://python-security.readthedocs.io/vuln/urllib-basic-auth-regex2.html to track this vulnerability.

History

Date	User	Action	Args
2022-04-11 14:59:40	admin	set	github: 87241
2021-09-07 20:12:23	vstinner	set	messages: + msg401341
2021-09-07 19:52:26	vstinner	set	title: ReDoS in urllib.request -> CVE-2021-3733: ReDoS in urllib.request
2021-08-23 11:13:04	crazybyte	set	nosy: + crazybyte messages: + msg400130
2021-05-06 17:00:47	ned.deily	set	status: open -> closed resolution: fixed stage: patch review -> resolved
2021-05-06 17:00:15	ned.deily	set	nosy: + ned.deily messages: + msg393109
2021-05-04 12:46:44	lukasz.langa	set	nosy: + lukasz.langa messages: + msg392885
2021-04-07 15:58:07	vstinner	set	messages: + msg390441
2021-04-07 11:45:13	miss-islington	set	messages: + msg390425
2021-04-07 11:29:19	miss-islington	set	pull_requests: + pull_request23989
2021-04-07 11:29:11	miss-islington	set	pull_requests: + pull_request23988
2021-04-07 11:28:20	miss-islington	set	pull_requests: + pull_request23987
2021-04-07 11:28:09	miss-islington	set	nosy: + miss-islington pull_requests: + pull_request23986
2021-04-07 11:27:50	vstinner	set	messages: + msg390420
2021-04-07 11:25:55	vstinner	set	files: + redos_python2.py messages: + msg390419
2021-04-07 11:14:28	yetingli	set	messages: + msg390417
2021-04-07 10:59:29	vstinner	set	messages: + msg390415
2021-03-14 08:50:54	yetingli	set	messages: + msg388665
2021-03-09 11:51:39	vstinner	set	nosy: + vstinner messages: + msg388358
2021-03-03 13:42:33	taleinat	set	keywords: + newcomer friendly
2021-03-03 03:02:02	zach.ware	set	keywords: + patch
2021-03-03 03:00:23	zach.ware	set	keywords: + easy, - patch, easy (C)
2021-03-03 02:45:12	eric.araujo	set	title: ReDoS in request -> ReDoS in urllib.request
2021-03-03 00:50:17	orsenthil	set	keywords: + easy (C)
2021-01-31 05:44:12	yetingli	set	messages: + msg386009
2021-01-31 05:12:49	yetingli	set	keywords: + patch stage: needs patch -> patch review pull_requests: + pull_request23205
2021-01-30 18:59:10	orsenthil	set	messages: + msg385993
2021-01-30 16:32:26	serhiy.storchaka	set	stage: needs patch versions: + Python 3.10
2021-01-30 16:32:06	serhiy.storchaka	set	messages: + msg385987
2021-01-30 09:21:50	xtreak	set	nosy: + orsenthil, serhiy.storchaka type: security
2021-01-30 08:11:46	yetingli	create	