

python / cpython Public

<> Code Issues 5k+ Pull requests 2.2k Actions Projects Security and q

New issue



Incorrect parsing of TarInfo header when GNU long name and type AREGTYPE are combined #141707

Open

Labels

stdlib

type-bug

type-security



e-nomem opened on Nov 18, 2025 · edited by bedevere-app

Edits ▾

Contributor



Bug report

Bug description:

When an entry uses GNU long name encoding the `tarfile` module reads in the data blocks for the name and then [calls `self.fromtarfile\(\)` again](#) to get the 'actual' header. This second header is the source of truth for everything *except* the name which is just garbage data.

The problem is that `fromtarfile()` eventually calls `frombuf()` where [this logic](#) incorrectly uses the garbage data and overrides the entry type to directory, corrupting the entry.

Because the entry is detected as a directory, the offset is not updated properly and the next call to read a TarInfo entry will usually result in an exception. However, the exception lands up in [this block](#) where neither of the `if` conditions are met, so the exception is silently discarded. `tarinfo` remains `None` and the code [eventually decides](#) that there are no more entries in the tar file.

I initially ran into this issue due to reports of invalid sdsts being generated by maturin.

See: [PyO3/maturin#2855](#)

CPython versions tested on:

3.9, 3.10, 3.11, 3.12, 3.13, 3.14

Operating systems tested on:

macOS, Linux

Linked PRs

- [gh-141707: Fix tarfile type corruption with GNU long names #143157](#)
- [gh-141707: Skip TarInfo DIRTYE normalization during GNU long name handling #143934](#)
- [\[3.11\] gh-141707: Skip TarInfo DIRTYE normalization during GNU long name handling #145815](#)
- [\[3.10\] gh-141707: Skip TarInfo DIRTYE normalization during GNU long name handling #145816](#)
- [\[3.12\] gh-141707: Skip TarInfo DIRTYE normalization during GNU long name handling #145817](#)
- [\[3.13\] gh-141707: Skip TarInfo DIRTYE normalization during GNU long name handling #145818](#)
- [\[3.14\] gh-141707: Skip TarInfo DIRTYE normalization during GNU long name handling #145819](#)

  **e-nomem** added **type-bug** on Nov 18, 2025

  **e-nomem** mentioned this on Nov 18, 2025

[Set entry type when adding to the tar file PyO3/maturin#2859](#)

  **messense** added a commit that references this issue on Nov 18, 2025

Set entry type when adding to the tar file ([#2859](#)) 

Verified 5e42853

  **e-nomem** mentioned this on Nov 18, 2025

[Initialize Header::entry_type to EntryType::Regular alexcrichon/tar-rs#422](#)

  **StanFromIreland** added **stdlib** on Nov 18, 2025

  **StanFromIreland** added this to  Tarfile issues on Nov 18, 2025


  **messense** added a commit that references this issue on Nov 19, 2025

Set entry type when adding to the tar file ([#2859](#)) 

Verified 10c977d









  **woodruffw** mentioned this on Dec 9, 2025

[Explicitly set EntryType for file entries in tar astral-sh/uv#17043](#)

 **konstin** on Dec 11, 2025

Contributor 

uv is affected by this too: [astral-sh/uv#17043](#)

- 
heathdutton added a commit that references this issue [on Dec 24, 2025](#)
[pythongh-141707](#): Fix tarfile type corruption with GNU long names 8e61d0b
- 
heathdutton mentioned this [on Dec 24, 2025](#)
 [gh-141707: Fix tarfile type corruption with GNU long names #143157](#)
- 
heathdutton added a commit that references this issue [on Dec 24, 2025](#)
[pythongh-141707](#): Fix tarfile type corruption with GNU long names 5186da2
- 
sethmlarson mentioned this [on Jan 16](#)
 [Skip TarInfo DIRTYE normalization during GNU long name handling #143933](#)
- 
sethmlarson added type-security [on Jan 16](#)
- 
sethmlarson added a commit that references this issue [on Jan 16](#)
[pythongh-141707](#): Skip TarInfo DIRTYE normalization during GNU long n. 861656d
Verified

27 remaining items

Load more

[Sign up for free](#) to join this conversation on GitHub. Already have an account? [Sign in to comment](#)

Metadata

Assignees


No one assigned

Labels

stdlib type-bug type-security

Projects

 Tarfile issues

Status No status 



Milestone


No milestone

Relationships


None yet

Development

 Code with agent mode 

 **gh-141707: Fix tarfile type corruption with GNU long names**
python/cpython

Participants

 +5