New issue

Jump to bottom

While processing, division by zero causes an arithmetic exception #3498

Closed **tank0123** opened this issue on Jul 15, 2021 · 13 comments

Labels leptonica

tank0123 commented on Jul 15, 2021

System Configuration

- tesseract version: 5.0.0-alpha-20210401
- linked library version:
  leptonica-1.79.0
  libgif 5.1.4 : libjpeg 8d (libjpeg-turbo 2.0.3) : libpng 1.6.37 : libtiff 4.1.0 : zlib 1.2.11 : libwebp 0.6.1 : libopenjp2 2.3.1
  Found AVX512BW
  Found AVX512F
  Found AVX2
  Found AVX
  Found FMA
  Found SSE
  Found OpenMP 201511
  Found libcurl/7.68.0 GnuTLS/3.6.13 zlib/1.2.11 brotli/1.0.7 libidn2/2.2.0 libpsl/0.21.0 (+libidn2/2.2.0) libssh/0.9.3/openssl/zlib nghttp2/1.40.0 librtmp/2.3
- Environment (Operating system, version and so on): Ubuntu 20.04.2 64bit

Program received signal SIGFPE, Arithmetic exception.

#0 0x00007ffff7dbe24a in pixBlockconvGray () from /lib/x86_64-linux-gnu/liblept.so.5
#1 0x00007ffff7dbeadd in pixBlockconv () from /lib/x86_64-linux-gnu/liblept.so.5
#2 0x00005555556b2d9b in tesseract::TextlineProjection::ConstructProjection (this=0x55555586e230,
input_block=input_block@entry=0x55555587e110, rotation=..., nontext_map=...) at ./src/ccstruct/image.h:34
#3 0x00005555556980a7 in tesseract::StrokeWidth::GradeBlobsIntoPartitions (this=0x55555586ac20,
pageseg_mode=pageseg_mode@entry=tesseract::PSM_AUTO, rerotation=...,
block=block@entry=0x55555587e110, nontext_pix=...,
denorm=, cjk_script=0x0, projection=0x55555586e230, diacritic_blobs=0x7fffffffd018,
part_grid=0x55555586e1d8, big_parts=0x55555586e208) at src/textord/strokewidth.cpp:371
#4 0x00005555566b4e5 in tesseract::ColumnFinder::FindBlocks (this=this@entry=0x55555586e0a0,
pageseg_mode=pageseg_mode@entry=tesseract::PSM_AUTO, scaled_color=..., scaled_factor=,
input_block=input_block@entry=0x55555587e110, photo_mask_pix=..., thresholds_pix=..., grey_pix=...,
pixa_debug=0x7ffff624cbd0, blocks=0x7fffffffcf78, diacritic_blobs=0x7fffffffd018, to_blocks=0x7fffffffd020)
at src/textord/colfind.cpp:295
#5 0x00005555555b1c8f in tesseract::Tesseract::AutoPageSeg (this=0x7ffff6229010,
pageseg_mode=tesseract::PSM_AUTO, blocks=0x555556f3f830, to_blocks=0x7fffffffd020,
diacritic_blobs=0x7fffffffd018, osd_tess=,
osr=0x7fffffffd3d0) at src/ccmain/pagesegmain.cpp:226
#6 0x00005555555b214d in tesseract::Tesseract::SegmentPage (this=0x7ffff6229010, input_file=,
blocks=0x555556f3f830, osd_tess=osd_tess@entry=0x0, osr=osr@entry=0x7fffffffd3d0) at
./src/ccutil/params.h:202
#7 0x0000555555580e17 in tesseract::TessBaseAPI::FindLines (this=0x7fffffffe100) at
/usr/include/c++/9/bits/basic_string.h:2300
#8 0x0000555555583608 in tesseract::TessBaseAPI::Recognize (this=0x7fffffffe100, monitor=0x0) at
src/api/baseapi.cpp:838
#9 0x0000555555583c0a in tesseract::TessBaseAPI::ProcessPage (this=this@entry=0x7fffffffe100,
pix=0x55555587a110, page_index=page_index@entry=0x0,
filename=filename@entry=0x7fffffffe77a "/home/ubuntu/Aws-Results/orcheFuzz-
newbug/output_tesseract_of/initial_crashes/2021-05-06-03:01:41_0x7b7d0fd6_0xb1c1261c",
retry_config=retry_config@entry=0x0,
timeout_millisec=timeout_millisec@entry=0x0, renderer=0x55555586e710) at src/api/baseapi.cpp:1259
#10 0x0000555555584888 in tesseract::TessBaseAPI::ProcessPagesInternal (this=0x7fffffffe100, filename=,
retry_config=0x0, timeout_millisec=0x0, renderer=0x55555586e710) at
/usr/include/c++/9/bits/basic_string.h:2300
#11 0x0000555555584e33 in tesseract::TessBaseAPI::ProcessPages (this=0x7fffffffe100, filename=,
retry_config=, timeout_millisec=, renderer=) at src/api/baseapi.cpp:1071
#12 0x0000555555575ba5 in main (argc=argc@entry=0x3, argv=argv@entry=0x7fffffffe528) at
/usr/include/c++/9/bits/unique_ptr.h:360
#13 0x00007ffff771f0b3 in __libc_start_main (main=0x555555574ee0 <main(int, char**)>, argc=0x3,
argv=0x7fffffffe528, init=, fini=, rtld_fini=, stack_end=0x7fffffffe518)
at ../csu/libc-start.c:308
#14 0x00005555557d1be in _start () at /usr/include/x86_64-linux-gnu/bits/stdio2.h:100

I've attached the file. Please download and check the file.
2021-05-06-03_01_41_0x7b7d0fd6_0xb1c1261c.zip

stweil commented on Jul 15, 2021 • edited ▾ Contributor

The exception happens in Leptonica code (function `pixBlockconvGray`), see the stack listed in the report.

stweil commented on Jul 15, 2021 • edited ▾ Contributor

Leptonica prints a hint:

```
Corrupt JPEG data: 2 extraneous bytes before marker 0xd9
```

So the JPEG image is invalid. With latest Leptonica code Tesseract no longer crashes, but delivers an empty text result (which is fine for this image).

amitdo added the leptonica label on Jul 16, 2021

stweil commented on Sep 5, 2021 Contributor

I close this issue because it only happens with an invalid JPEG image, and it is fixed in latest Leptonica.

stweil closed this as completed on Sep 5, 2021

---

ajakk commented on Sep 9

This is referenced in CVE-2022-38266. Do we what Leptonica patch fixes this?

zdenop commented on Sep 10 Contributor

First of all: provide an image for replicating the problem.
Next: As reported by stweil: the current version of leptonica reports a problem (Corrupt JPEG data) and does not crash.
So what do you want to fix?

stweil commented on Sep 10 Contributor

Citing the CVE: "An issue in the Leptonica linked library (v1.79.0) in Tesseract v5.0.0 allows attackers to cause an arithmetic exception leading to a Denial of Service (DoS) via a crafted JPEG file."

This sounds rather strange. Attackers who are able to provide a JPEG for Tesseract OCR can simply run a lot of Tesseract processes with normal JPEG files. That is also some kind of DoS.

zdenop commented on Sep 10 Contributor

I find jpeg in 2021-05-06-03_01_41_0x7b7d0fd6_0xb1c1261c.zip.
On Windows tesseract finished without crash/exception with empty results as stweil reported:

```
>tesseract 2021-05-06-03_01_41_0x7b7d0fd6_0xb1c1261c.jpeg -
Corrupt JPEG data: 2 extraneous bytes before marker 0xd9
Estimating resolution as 1625
```

```
>tesseract -v
tesseract 5.2.0-8-ge589b
 leptonica-1.83.0 (Aug  1 2022, 13:11:36) [MSC v.1929 LIB Release x64]
  libgif 5.2.1 : libjpeg 6b (libjpeg-turbo 2.0.91) : libpng 1.6.37 : libtiff 4.4.0 : zlib 1.2.12 : li
 Found AVX2
 Found AVX
 Found FMA
 Found SSE4.1
 Found OpenMP 2019
 Found libarchive 3.5.1 zlib/1.2.11 liblzma/5.2.4 bz2lib/1.0.6 libzstd/1.4.9
 Found libcurl/7.75.0 zlib/1.2.12 libssh2/1.10.1_DEV
```

So I do not see what we can do with it in tesseract.

ajakk commented on Sep 10

> First of all: provide an image for replicating the problem. Next: As reported by stweil: the current version of leptonica reports a problem (Corrupt JPEG data) and does not crash. So what do you want to fix?

I want to make the CVE clearer. There's obviously an issue in Leptonica, but it doesn't contain any information about remediation other than "use latest Leptonica". That's not an option for certain distributions, so it's important to be able to identify the patches that fix it so those distributions can apply those patches to their distributions.

Leptonica also has many CVEs. If this issue has already been assigned a CVE, then CVE-2022-38266 is a duplicate and should be rejected. Very hard to tell without knowing what the fix is.

👍 2

stweil commented on Sep 10 Contributor

Leptonica > 1.80.0 should be fine. Older releases require patch DanBloomberg/leptonica@ `f062b42` .

choonginlee commented 5 days ago

Many thanks to the developers of tesseract for their hard work. I am part of the team that reported this issue. When I found this issue, I was only using the library provided by default in the ubuntu environment mentioned above (after apt-get update), so this bug was found even though it was fixed in the latest leptonica. As stweil said, with the above bug in mind, it might be better to specify the library version for secure run . It is a recent and frequent debate that an outdated library can cause problems in the product itself that uses it. We know that it imposes a lot of burden on developers. Thanks again for the action.

stweil commented 5 days ago Contributor

Users with Linux distributions which still have that problem can report it as an issue to their distribution. It can be fixed in old releases with the patch mentioned above.

choonginlee commented 5 days ago

@stweil Let me contact the distribution so that it requires Leptonica > 1.80.0.
@ajakk I could not find a CVE regarding divide by zero vulnerability of Leptonica in any version. I guess it is not a duplicate.

ajakk commented 5 days ago

Thanks. I'll request that MITRE update the CVE accordingly.

Assignees

No one assigned

Labels

leptonica

Projects

None yet

Milestone

No milestone

Development

No branches or pull requests

---

6 participants