Stack overflow via recursive loop in Gawk

Description

GAWK is a widely used domain-specific language designed for text processing. The "check_dst_limits_calc_pos_1" function in "support/regexec.c" may trigger stack overflow via recursive loop. The vulnerability exists in latest stable release (gawk-5.1.1) and latest master branch (12f0f4f27872cd4df6c63977fd0c6ff63d29e424, updated on July 29, 2022).The vulnerable code (located at support/regexec.c) and the bug's basic introduction are highlighted as follows:

```
// support/regexec.c
static int
check_dst_limits_calc_pos_1 (const re_match_context_t *mctx, int boundaries, Idx
subexp_idx, Idx from_node, Idx bkref_idx)
{
    // line 1915
    // under this poc, the function has recursive loop which leads to stack
overflow
    cpos =  check_dst_limits_calc_pos_1 (mctx, boundaries, subexp_idx, dst,
bkref_idx);
```

Proof of Concept

Build the gawk (5.1.1 or latest commit 12f0f4f27872cd4df6c63977fd0c6ff63d29e424) and run it using the input POC.

```
# build the gawk with address sanitizer
export CFLAGS=" -fsanitize=address "; export CXXFLAGS=" -fsanitize=address ";
export LDFLAGS=" -fsanitize=address ";
CFGARG=" --enable-shared=no "; configure

AFL_USE_ASAN=1 LD=afl-gcc--disable-shared make

# the INPUT_FILE is not decesive, which can be any valid input.
./gawk -f POC_FILE INPUT_FILE
```

The stack dump is:

```
Disassembly:
Stack Head (1000 entries):
   check_dst_limits_calc_pos @ 0x00005555556dcab3: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
```

```
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
   check_dst_limits_calc_pos @ 0x00005555556dce86: in /src/AWKs/gawk-5.1.1/gawk
Registers:
rax=0x0000555555964850 rbx=0x00007ffff7770010 rcx=0x0000000000000023
rdx=0x0000000000000230
rsi=0x0000000000000003 rdi=0x00007fffffffdee0 rbp=0xffffffffffffff90
rsp=0x00007fffff7fefd0
 r8=0x0000000000000003  r9=0x00007ffff777fdba r10=0x000055555595fde0
r11=0x0000000000000003
r12=0x0000000000000220 r13=0x000055555595fde0 r14=0x0000555555980940
r15=0x0000000000000022
rip=0x00005555556dcab3 efl=0x0000000000010202  cs=0x0000000000000033
 ss=0x000000000000002b
 ds=0x0000000000000000  es=0x0000000000000000  fs=0x0000000000000000
 gs=0x0000000000000000
 k0=0x00000000feffff00  k1=0x000000000000ffff  k2=0x00000000010000ff
 k3=0x0000000000000000
 k4=0x0000000000000000  k5=0x0000000000000000  k6=0x0000000000000000
 k7=0x0000000000000000

#0  0x0000555555922ff9 in check_dst_limits_calc_pos_1 (
    mctx=0x7fffffffaed0, boundaries=3, subexp_idx=3, from_node=34,
    bkref_idx=3) at ./regexec.c:1871
#1  0x00005555559235b5 in check_dst_limits_calc_pos_1 (
    mctx=0x7fffffffaed0, boundaries=3, subexp_idx=3, from_node=35,
    bkref_idx=3) at ./regexec.c:1915
#2  0x00005555559235b5 in check_dst_limits_calc_pos_1 (
    mctx=0x7fffffffaed0, boundaries=3, subexp_idx=3, from_node=34,
    bkref_idx=3) at ./regexec.c:1915
#3  0x00005555559235b5 in check_dst_limits_calc_pos_1 (
    mctx=0x7fffffffaed0, boundaries=3, subexp_idx=3, from_node=35,
    bkref_idx=3) at ./regexec.c:1915
......
#24901 0x00005555559235b5 in check_dst_limits_calc_pos_1 (mctx=0x7fffffffaed0,
boundaries=3, subexp_idx=3, from_node=17, bkref_idx=3) at ./regexec.c:1915
#24902 0x00005555559235b5 in check_dst_limits_calc_pos_1 (mctx=0x7fffffffaed0,
boundaries=3, subexp_idx=3, from_node=16, bkref_idx=3) at ./regexec.c:1915
#24903 0x0000555555922f7f in check_dst_limits_calc_pos (mctx=0x7fffffffaed0,
limit=8, subexp_idx=3, from_node=16, str_idx=1, bkref_idx=3) at ./regexec.c:1970
#24904 0x0000555555922a4d in check_dst_limits (mctx=0x7fffffffaed0,
limits=0x7fffffffa558, dst_node=16, dst_idx=1, src_node=17, src_idx=1) at
./regexec.c:1848
#24905 0x000055555591f4a5 in sift_states_bkref (mctx=0x7fffffffaed0,
sctx=0x7fffffffa540, str_idx=1, candidates=0x608000002a28) at ./regexec.c:2113
#24906 0x000055555591bf3d in update_cur_sifted_state (mctx=0x7fffffffaed0,
sctx=0x7fffffffa540, str_idx=1, dest_nodes=0x7fffffffa3a0) at ./regexec.c:1748
#24907 0x000055555591a62c in sift_states_backward (mctx=0x7fffffffaed0,
sctx=0x7fffffffa540) at ./regexec.c:1559
```

```
#24908 0x000055555591f7fe in sift_states_bkref (mctx=0x7fffffffaed0,
sctx=0x7fffffffac80, str_idx=1, candidates=0x608000002a28) at ./regexec.c:2133
#24909 0x000055555591bf3d in update_cur_sifted_state (mctx=0x7fffffffaed0,
sctx=0x7fffffffac80, str_idx=1, dest_nodes=0x7fffffffaae0) at ./regexec.c:1748
#24910 0x000055555591a62c in sift_states_backward (mctx=0x7fffffffaed0,
sctx=0x7fffffffac80) at ./regexec.c:1559
#24911 0x00005555558fc2bc in prune_impossible_nodes (mctx=0x7fffffffaed0) at
./regexec.c:931
#24912 0x00005555558ba1e4 in re_search_internal (preg=0x60b0000007d0,
string=0x615000001980 "drwxr-xr-x 22 1000 1000    4096 Jul 31 15:39 .",
length=46, start=0, last_start=46, stop=46, nmatch=1, pmatch=0x602000008a30,
eflags=0) at ./regexec.c:805
#24913 0x00005555558bc742 in re_search_stub (bufp=0x60b0000007d0,
string=0x615000001980 "drwxr-xr-x 22 1000 1000    4096 Jul 31 15:39 .",
length=46, start=0, range=46, stop=46, regs=0x0, ret_len=false) at
./regexec.c:420
#24914 0x00005555558bcc70 in re_search (bufp=0x60b0000007d0,
string=0x615000001980 "drwxr-xr-x 22 1000 1000    4096 Jul 31 15:39 .",
length=46, start=0, range=46, regs=0x0) at ./regexec.c:284
#24915 0x0000555555862406 in research (rp=0x60b0000007d0, str=0x615000001980
"drwxr-xr-x 22 1000 1000    4096 Jul 31 15:39 .", start=0, len=46, flags=0) at
re.c:354
#24916 0x000055555579c2b4 in r_interpret (code=0x622000001908) at
./interpret.h:1151
#24917 0x000055555581cddd in main (argc=4, argv=0x7fffffffe3c8) at main.c:553
```

Impact

This vulnerability can be used for causing the crash or long-term loop of the software which would leads to denial of service(DoS). Besides, the attacker can exploit weak points to launch remote code execution.

Reference

POC_FILE