0x27 Automated Testing Exercise Session





Recap

Fuzzing

- Automatic and dynamic software testing technique
- Key concept: fastest way to test a program is to run it
 - Runs on bare metal => "efficient"
- Analogy: obstacle course
 - Programs are riddled with control-flow decisions
 - Only one path can be taken at a time
 - o path[-1] may depend on path[0:-1]



In this Lab, we use AFL++^[1], current State-of-the-Art Greybox Fuzzer

Environment Setup

- Download the Dockerfile https://github.com/kdsjZh/COM402-Testing-Lab/blob/master/docker/Dockerfile
- docker build -t com402/testing:2024 /path/to/dockerfile
- docker run -it com402/testing:2024 bash

Program Under Test

```
unsigned int size;
if (read(fd, &size, 4) != 4) {
    perror("Failed to read size");
    close(fd);
   return 1;
// Vulnerable buffer: fixed-size but could overflow based on file contents
char buffer[MAX_BUFFER_SIZE];
// Read 'size' bytes into the buffer,
if (read(fd, buffer, size) < 0) {</pre>
    perror("Failed to read file content");
    close(fd);
    return 1;
// Null-terminate the buffer
buffer[MAX_BUFFER_SIZE - 1] = '\0';
```

Program Under Test

buffer[MAX BUFFER SIZE - 1] = '\0';

```
unsigned int size;
if (read(fd, &size, 4) != 4) {
   perror("Failed to read size");
   close (fd);
   return 1:
// Vulnerable buffer: fixed-size but could overflow based on file contents
char buffer[MAX_BUFFER_SIZE];
// Read 'size' bytes into the buffer,
                                                  if size > MAX_BUFFER_SIZE,
if (read(fd, buffer, size) < 0) {</pre>
                                                  stack buffer overflows!
   perror("Failed to read file content");
   close (fd);
   return 1;
// Null-terminate the buffer
```

[AFL++ 225c9e640908] /AFLplusplus # cd /COM402-Testing-Lab/demo/ex1

demo/ex1 # make clean && make

demo/ex1 # cat run.sh

demo/ex1 # ./run.sh

```
american fuzzy lop ++4.22a {} (./ex1) [explore]
                                                       overall results
 process timing
       run time : 0 days, 0 hrs, 0 min, 3 sec
                                                       cycles done : 26
  last new find : n/a (non-instrumented mode)
                                                      corpus count : 1
last saved crash : 0 days, 0 hrs, 0 min, 0 sec
                                                     saved crashes: 59
 last saved hang : none seen yet
                                                       saved hangs: 0
 cycle progress
                                        map coverage
 now processing: 0*79 (0.0%)
                                          map density : 0.00% / 0.00%
 runs timed out : 0 (0.00%)
                                       count coverage: 0.00 bits/tuple
 stage progress
 now trying : havoc
                                       favored items: 0 (0.00%)
 stage execs : 57/100 (57.00%)
                                        new edges on : 0 (0.00%)
 total execs : 7864
                                       total crashes : 59 (59 saved)
 exec speed: 2526/sec
                                        total tmouts : 0 (0 saved)
 fuzzing strategy yields
                                                      item geometry
  bit flips : 0/0, 0/0, 0/0
                                                        levels: 1
 byte flips: 0/0, 0/0, 0/0
                                                       pending: 0
 arithmetics : 0/0, 0/0, 0/0
                                                      pend fav : 0
 known ints: 0/0, 0/0, 0/0
                                                     own finds: 0
 dictionary: 0/0, 0/0, 0/0, 0/0
                                                      imported : n/a
havoc/splice : 53/7800, 0/0
                                                     stability : n/a
py/custom/rg : unused, unused, unused, unused
   trim/eff : n/a. n/a
                                                              [cpu000: 15%]
 strategy: explore — state: in progress — ^C
```

For Simple Programs, Blackbox fuzzers works pretty well.

But what if code becomes complex?

Ex2. Coverage Feedback

Program Under Test

```
unsigned char *size bytes = (unsigned char *)&size;
unsigned char magic_bytes[2] = { 0xef, 0xbe};
// Check and print messages for each matching byte
if (size bytes[0] == magic bytes[0]) {
   printf("Byte 1 is match!\n");
} else {
   printf("Byte 1 does not match.\n");
   return 1;
if (size bytes[1] == magic bytes[1]) {
   printf("Byte 2 is match!\n");
} else {
   printf("Byte 2 does not match.\n");
   return 1;
```

Program Under Test

```
unsigned char *size bytes = (unsigned char *)&size;
unsigned char magic bytes[2] = { 0xef, 0xbe};
   Check and print messages for each matching byte
if (size bytes[0] == magic bytes[0]) {
    printf("Ryte 1 is match!\n"):
} else {
    printf("Byte 1 does not match.\n");
    return 1;
if (size bytes[1] == magic bytes[1]) {
    printf("Byte 2 is match!\n");
} else {
    printf("Byte 2 does not match.\n");
    return 1;
```

First two bytes of size have to match magic bytes, else program exit

Ex2. Coverage-Guided Fuzzing

[AFL++ 225c9e640908] /AFLplusplus # cd /COM402-Testing-Lab/demo/ex2 demo/ex2 # make clean && make

demo/ex2 # \$AFL_PATH/afl-fuzz -i in/ -o out/ -n -- ./ex2 @@

Try Blackbox fuzzing for 5 minutes.

Does Blackbox fuzzing works? Why/Why not?

Ex2. Coverage-Guided Fuzzing

demo/ex2 # make clean && CC=afl-cc make

demo/ex2 # cat run.sh

We build with afl-cc, a clang wrapper that compile program and **instrument coverage**

Ex2. Converge-Guided Fuzzing

```
american fuzzy lop ++4.22a {} (./ex2) [explore]
       run time : 0 days, 0 hrs, 0 min, 30 sec
                                                       cycles done : 244
   last new find : n/a (non-instrumented mode)
                                                       corpus count : 1
last saved crash : none seen yet
                                                      saved crashes : 0
                                                       saved hangs : 0
 last saved hang : none seen yet
 cycle progress -
                                        map coverage
  now processing: 0*733 (0.0%)
                                          map density : 0.00% / 0.00%
  runs timed out : 0 (0.00%)
                                       count coverage : 0.00 bits/tuple
  stage progress ---
  now trying : havoc
                                        favored items: 0 (0.00%)
 stage execs : 94/100 (94.00%)
                                        new edges on : 0 (0.00%)
 total execs : 73.3k
                                        total crashes: 0 (0 saved)
  exec speed : 2392/sec
                                        total tmouts : 0 (0 saved)
 fuzzing strategy yields
                                                      item geometry
  bit flips : 0/0, 0/0, 0/0
                                                        levels : 1
  byte flips: 0/0, 0/0, 0/0
                                                       pending: 0
 arithmetics : 0/0, 0/0, 0/0
                                                       pend fav : 0
  known ints: 0/0, 0/0, 0/0
                                                      own finds : 0
  dictionary: 0/0, 0/0, 0/0, 0/0
                                                      imported : n/a
havoc/splice : 0/73.2k, 0/0
                                                      stability : n/a
py/custom/rg : unused, unused, unused, unused
                                                               [cpu000: 10%]
    trim/eff : n/a, n/a
                           — state: in progress — ^C
  strategy: explore ———
```

```
american fuzzy lop ++4.22a {default} (./ex2) [explore]
       run time : 0 days, 0 hrs, 0 min, 14 sec
  last new find: 0 days, 0 hrs, 0 min, 14 sec
                                                      corpus count : 4
last saved crash : 0 days, 0 hrs, 0 min, 1 sec
                                                     saved crashes : 1
last saved hang : none seen yet
                                                      saved hangs : 0
cycle progress —
                                        map coverage
 now processing : 2.388 (50.0%)
                                          map density: 14.29% / 35.71%
 runs timed out : 0 (0.00%)
                                       count coverage : 77.80 bits/tuple
                                       findings in depth —
 stage progress —
 now trying : havoc
                                       favored items : 4 (100.00%)
stage execs : 98/100 (98.00%)
                                        new edges on: 4 (100.00%)
total execs : 120k
                                       total crashes : 1 (1 saved)
 exec speed: 7991/sec
                                       total tmouts : 0 (0 saved)
 fuzzing strategy vields
                                                     item geometry
  bit flips : 0/0, 0/0, 0/0
                                                        levels : 3
 byte flips: 0/0, 0/0, 0/0
                                                       pending: 0
arithmetics : 0/0, 0/0, 0/0
                                                      pend fav : 0
 known ints: 0/0, 0/0, 0/0
 dictionary: 0/0, 0/0, 0/0, 0/0
                                                     imported: 0
havoc/splice : 4/120k, 0/0
                                                     stability: 100.00%
py/custom/rg : unused, unused, unused, unused
                                                              [cpu000: 10%]
   trim/eff : n/a, n/a
 strategy: explore ----
                            — state: started :-) — ^C
```

Without Coverage

With Coverage

Ex2. Coverage-Guided Fuzzing

If the magic bytes are validated byte-by-byte, coverage feedback works well,

But what if program compares 4 bytes at once?

Program Under Test

```
// Check and print messages for each matching byte
if (size == MAGIC_BYTES) {
    printf("Magic Bytes match!\n");
} else {
    printf("Magic Bytes not match.\n");
    return 1;
}
```

Program Under Test

```
// Check and print messages for each matching byte
if (size == MAGIC_BYTES) {
    printf("Magis Bytes match!\n");
} else {
    printf("Magic Bytes not match.\n");
    return 1;
}
```

Check 4 Bytes at once

```
demo/ex2 # cd /COM402-Testing-Lab/demo/ex3/
demo/ex3 # make clean && CC=afl-cc make
demo/ex3 # $AFL_PATH/afl-fuzz -i in/ -o out/ -- ./ex3 @@
```

Try fuzzing for 5 minutes

Does coverage guided fuzzing still work for magic bytes comparison?

CmpLog^[1] is a technique that extract the compared value from register and fill it back to input.

It's enabled in AFL++ if we compile program with AFL_LLVM_CMPLOG=1

[1] Aschermann, Cornelius, et al. "REDQUEEN: Fuzzing with Input-to-State Correspondence." *NDSS*. Vol. 19. 2019.

```
demo/ex3 # CC=afl-cc make ex3 && mv ex3 ex3.afl

demo/ex3 # CC=afl-cc AFL_LLVM_CMPLOG=1 make ex3 && mv ex3 ex3.cmplog

demo/ex3 # $AFL_PATH/afl-fuzz -i in/ -o out/ -c ./ex3.cmplog -1 3 -- ./ex3.afl @@
```

We build one binary with cmplog instrumentation and run AFL++ with cmplog mutator enabled

```
american fuzzy lop ++4.22a {default} (./ex3.afl) [explore]
       run time : 0 days, 0 hrs, 0 min, 40 sec
                                                      cycles done : 830
  last new find: 0 days, 0 hrs, 0 min, 40 sec
                                                      corpus count : 2
                                                     saved crashes: 0
last saved crash : none seen vet
 last saved hang : none seen yet
                                                      saved hangs : 0
 cycle progress —
                                        map coverage
 now processing: 0.1685 (0.0%)
                                          map density: 15.38% / 23.08%
 runs timed out : 0 (0.00%)
                                       count coverage: 129.00 bits/tuple
 stage progress ---
                                       findings in depth ———
 now trying : havoc
                                       favored items : 2 (100.00%)
 stage execs: 72/100 (72.00%)
                                       new edges on : 2 (100.00%)
 total execs : 332k
                                       total crashes : 0 (0 saved)
 exec speed: 8180/sec
                                        total tmouts : 0 (0 saved)
 fuzzing strategy vields -

    item geometry

  bit flips : 0/0, 0/0, 0/0
                                                       levels: 2
 byte flips: 0/0, 0/0, 0/0
                                                      pending: 0
 arithmetics : 0/0, 0/0, 0/0
                                                     pend fav : 0
 known ints : 0/0, 0/0, 0/0
                                                     own finds: 1
 dictionary: 0/0, 0/0, 0/0, 0/0
                                                     imported: 0
havoc/splice : 1/332k, 0/0
                                                     stability: 100.00%
py/custom/rq : unused, unused, unused, unused
   trim/eff : 22.22%/2, n/a
                                                             [cpu000: 10%]
 strategy: explore — state: started :-) — ^C
```

```
american fuzzy lop ++4.22a {default} (./ex3.afl) [explore]
       run time : 0 days, 0 hrs, 0 min, 4 sec
                                                      cvcles done : 34
   last new find: 0 days, 0 hrs, 0 min, 4 sec
                                                      corpus count : 3
last saved crash : 0 days, 0 hrs, 0 min, 1 sec
                                                     saved crashes : 1
last saved hang : none seen yet
                                                      saved hangs : 0
 cycle progress ———
                                        map coverage
 now processing : 2.74 (66.7%)
                                          map density: 15.38% / 30.77%
 runs timed out : 0 (0.00%)
                                       count coverage : 97.00 bits/tuple
 stage progress ----
                                       findings in depth ——
 now trying : havoc
                                       favored items : 3 (100.00%)
stage execs : 55/100 (55.00%)
                                        new edges on: 3 (100.00%)
 total execs : 35.6k
                                       total crashes : 2 (1 saved)
                                        total tmouts : 0 (0 saved)
 exec speed : 7184/sec

    fuzzing strategy yields

    item geometry

  bit flips : 0/0, 0/0, 0/0
 byte flips : 0/0, 0/0, 0/0
                                                      pending: 0
arithmetics : 0/0. 0/0. 0/0
                                                      pend fav : 0
 known ints : 0/0, 0/0, 0/0
                                                     own finds: 2
 dictionary: 0/0, 0/0, 0/0, 0/0
                                                     imported: 0
havoc/splice : 1/17.4k, 1/18.0k
                                                    stability: 100.00%
py/custom/rg : unused, unused, 0/22, 1/49
                                                              [cpu000: 10%]
   trim/eff : disabled, n/a
 strategy: explore — state: started :-) — ^C
```

Without Cmplog

With Cmplog

All our assumption is based on that vulnerability will crash the program, e.g., overwriting the return address in the stack

But what if the vulnerability does not crash the program?

Program Under Test

```
char *buffer = (char *)malloc(MAX_BUFFER_SIZE);
```

Now buffer is on heap, overflow does not directly crash the program

```
demo/ex3 # cd /COM402-Testing-Lab/demo/ex4

demo/ex4 # make clean && CC=afl-cc make

demo/ex4 # $AFL_PATH/afl-fuzz -i in/ -o out/ -- ./ex4 @@
```

Fuzz for 5 minutes, any finding?

Is the bug being triggered?

AddressSanitizer^[1] is a technique that stops the program execution if the memory safety is violated, i.e. don't wait until program cannot execute, stop as soon as its wrong.

It's integrated both in gcc and clang, enabled in AFL++ if we compile with AFL_USE_ASAN=1

Tip: set `ulimit -c unlimited`

[1] Serebryany, Konstantin, et al. "{AddressSanitizer}: A fast address sanity checker." 2012 USENIX annual technical conference (USENIX ATC 12). 2012.

```
american fuzzy lop ++4.22a {default} (./ex4) [explore]
       run time : 0 days, 0 hrs, 0 min, 30 sec
  last new find : 0 days, 0 hrs, 0 min, 29 sec
                                                      corpus count: 4
last saved crash : none seen vet
                                                     saved crashes : 0
last saved hang : none seen yet
                                                      saved hangs: 0
 cycle progress ---
                                        map coverage
 now processing : 2.369 (50.0%)
                                          map density: 14.29% / 35.71%
 runs timed out : 0 (0.00%)
                                       count coverage : 77.80 bits/tuple
 stage progress -
                                        findings in depth -
 now trying : havoc
                                       favored items : 4 (100.00%)
 stage execs : 2/100 (2.00%)
                                        new edges on: 4 (100.00%)
 total execs : 249k
                                       total crashes : 0 (0 saved)
                                        total tmouts : 0 (0 saved)
 exec speed: 8169/sec
 fuzzing strategy yields
                                                     item geometry
  bit flips: 0/0, 0/0, 0/0
                                                       levels : 3
 byte flips: 0/0, 0/0, 0/0
                                                      pending: 0
arithmetics : 0/0, 0/0, 0/0
                                                      pend fav : 0
 known ints: 0/0, 0/0, 0/0
 dictionary: 0/0, 0/0, 0/0, 0/0
                                                     imported: 0
havoc/splice : 3/249k. 0/0
                                                     stability: 100.00%
py/custom/rq : unused, unused, unused, unused
   trim/eff : 23.53%/3, n/a
                                                             [cpu000: 10%]
 strategy: explore — state: started :-) — ^C
```

```
american fuzzy lop ++4.22a {default} (./ex4) [explore]
       run time : 0 days, 0 hrs, 0 min, 4 sec
                                                      cvcles done : 15
  last new find: 0 days, 0 hrs, 0 min, 1 sec
                                                      corpus count: 4
last saved crash : 0 days, 0 hrs, 0 min, 0 sec
                                                     saved crashes: 1
 last saved hang : none seen yet
                                                      saved hangs : 0
 cycle progress ---
 now processing : 3.7 (75.0%)
                                          map density: 14.29% / 35.71%
 runs timed out : 0 (0.00%)
                                       count coverage : 77.80 bits/tuple
 stage progress -
                                       favored items : 4 (100.00%)
  now trying : havoc
 stage execs: 37/400 (9.25%)
                                        new edges on: 4 (100.00%)
 total execs : 14.8k
                                       total crashes : 2 (1 saved)
 exec speed: 2988/sec
                                        total tmouts : 0 (0 saved)
 fuzzing strategy yields -
                                                     item geometry
  bit flips : 0/0, 0/0, 0/0
                                                       levels : 3
 byte flips: 0/0, 0/0, 0/0
                                                      pending: 0
 arithmetics : 0/0, 0/0, 0/0
                                                      pend fav : 0
 known ints: 0/0, 0/0, 0/0
                                                     own finds: 3
 dictionary: 0/0, 0/0, 0/0, 0/0
                                                      imported: 0
havoc/splice : 4/14.7k. 0/0
                                                     stability: 100.00%
py/custom/rg : unused, unused, unused, unused
   trim/eff: 77.94%/16, n/a
                                                              [cpu000: 10%]
 strategy: explore — state: started :-) — ^C
```

Without ASan

With ASan

Crashes are not equal to bugs!

We still need to manually analyze the crashes to verify if its exploitable.

Take ex4 as example, run # ./ex4 /path/to/crash

```
5920==ERROR: AddressSanitizer: heap-buffer-overflow
[Detaching after fork from child process 25923]
   #0 0x5a34619237ae in interceptor read (/COM402-Testing-Lab/demo/ex4/ex4+0x3c7ae) (BuildId: 8d20f5e6139030f47c7563
31ffb6dbadee64791e)
   #1 0x5a34619db064 in main /COM402-Testing-Lab/demo/ex4/ex4.c:52:9
   #2 0x/2c35ad94d8f (/lib/x86 64-linux-gnu/libc.so.6+0x29d8f) (BuildId: 490fef8403240c91833978d494d39e537409b92e)
   #3 0x72c35ad94e3f in libc start main (/lib/x86 64-linux-gnu/libc.so.6+0x29e3f) (BuildId: 490fef8403240c91833978d4
94d39e537409b92e)
   #4 0x5a34619063c4 in start (/COM402-Testing-Lab/demo/ex4/ex4+0x1f3c4) (BuildId: 8d20f5e6139030f47c756331ffb6dbadee
64791e)
0x60b000000a4 is located 0 bytes after 100-byte region [0x60b000000040,0x60b000000a4)
allocated by thread TO here:
   #0 0x5a34619a01ee in interceptor malloc (/COM402-Testing-Lab/demo/ex4/ex4+0xb91ee) (BuildId: 8d20f5e6139030f47c75
6331ffb6dbadee64791e)
   #1 0x5a34619dafed in main /COM402-Testing-Lab/demo/ex4/ex4.c:49:28
   #2 0x72c35ad94d8f (/lib/x86 64-linux-gnu/libc.so.6+0x29d8f) (BuildId: 490fef8403240c91833978d494d39e537409b92e)
SUMMARY: AddressSanitizer: heap-buffer-overflow (/COM402-Testing-Lab/demo/ex4/ex4+0x3c7ae) (BuildId: 8d20f5e6139030f47c
756331ffb6dbadee64791e) in interceptor read
```

```
=25920==ERROR: AddressSanitizer: heap-buffer-overflow
[Detaching after fork from child process 25923]
   #0 0x5a34619237ae in interceptor read (/COM402-Testing-Lab/demo/ex
31ffb6dbadee64791e)
   #1 0x5a34619db064 in main /COM402-Testing-Lab/demo/ex4/ex4.c:52:9
   #2 0x/2c35ad94d8† (/lib/x86 64-linux-gnu/libc.so.6+0x29d8†) (BuildI
   #3 0x72c35ad94e3f in libc start main (/lib/x86 64-linux-qnu/libc.s
94d39e537409b92e)
   #4 0x5a34619063c4 in start (/COM402-Testing-Lab/demo/ex4/ex4+0x1f3c
64791e)
0x60b0000000a4 is located 0 bytes after 100-byte region [0x60b000000040,
allocated by thread TO here:
   #0 0x5a34619a01ee in interceptor malloc (/COM402-Testing-Lab/demo/
6331ffb6dbadee64791e)
   #1 0x5a34619dafed in main /COM402-Testing-Lab/demo/ex4/ex4.c:49:28
   \#2\ 0\times72c35ad94d8f\ (/lib/x86\ 64-linux-qnu/libc.so.6+0x29d8f)\ (BuildI
SUMMARY: AddressSanitizer: heap-buffer-overflow (/COM402-Testing-Lab/dem
756331ffb6dbadee64791e) in interceptor read
```

```
// Read 'size' bytes into the buffer,
if (read(fd, buffer, size) < 0) {
    perror("Failed to read file
content");
    close(fd);
    return 1;
}</pre>
```

```
(qdb) set args /path/to/crash
(qdb) b ex4.c:52
Breakpoint 1 at 0xf3fff: file ex4.c, line 52.
(adb) r
Starting program: /COM402-Testing-Lab/demo/ex4/ex4
out/default/crashes/id\:000000\,sig\:06\,src\:000003\,time\:6471\,execs\:19701\,op\:havoc\,rep
\:16
Byte 1 is match!
Byte 2 is match!
Breakpoint 1, main (argc=2, argv=0x7ffe2567e6e8) at ex4.c:52
52
         if (read(fd, buffer, size) < 0) {</pre>
(qdb) display size
                                               Size larger than buffer, Overflow!
1: size = 2763964143
```

```
#0 0x5a34619237ae in interceptor read (/COM402-Testing-Lab/demo/ex4/ex4+0x3c7ae)
31ffb6dbadee64791e)
   #1 0x5a34619db064 in main /COM402-Testing-Lab/demo/ex4/ex4.c:52:9
   #2 0x72c35ad94d8f (/lib/x86 64-linux-gnu/libc.so.6+0x29d8f) (BuildId: 490fef8403240
   #3 0x72c35ad94e3f in libc start main (/lib/x86 64-linux-gnu/libc.so.6+0x29e3f) (Bu:
94d39e537409b92e)
   #4 0x5a34619063c4 in start (/COM402-Testing-Lab/demo/ex4/ex4+0x1f3c4) (BuildId: 8d20
64791e)
 x60b0000000a4 is located 0 bytes after 100-byte region [0x60b000000040,0x60b000000a4)
allocated by thread 10 here:
   #0 0x5a34619a01ee in interceptor malloc (/COM402-Testing-Lab/demo/ex4/ex4+0xb91ee)
6331ffb6dbadee64791e)
   #1 0x5a34619dafed in main /COM402-Testing-Lab/demo/ex4/ex4.c:49:28
   #2 0x72c35ad94d8f (/lib/x86 64-linux-gnu/libc.so.6+0x29d8f) (BuildId: 490fef8403240d
SUMMARY: AddressSanitizer: heap-buffer-overflow (/COM402-Testing-Lab/demo/ex4/ex4+0x3c7a
756331ffb6dbadee64791e) in interceptor read
```

Why the size is 2763964143, but the actual write size is 108?

```
demo/ex4 # ls -lh
```

out/default/crashes/id\:000000\,sig\:06\,src\:000003\,time\:6471\,execs\:19701\,op\:havoc\,rep\:16 -rw----- 1 root root 112 Nov 2 14:09

out/default/crashes/id:000000,sig:06,src:000003,time:6471,execs:19701,op:havoc,rep:16