Basic - Debugging

EECS 201 Fall 2024

Submission Instructions

This assignment will be submitted as a repository on the EECS GitLab server (see Basic - Git 1). Create a private, blank, **README-less (uncheck that box!)** Project on it with the name/path/URL `eecs201-basic-debug` and add `brng` as a Reporter. The submission branch will be `trunk`. If this branch is not already the default initial branch, you initialize the local repo with an additional argument: `git init --initial-branch=trunk` or `git init -b trunk` if your version of Git is recent enough. Otherwise you can create a branch with this name after your first commit.

To finish the submission process and get your grade, you will need to run the autograder. You will need to SSH into the course server (see Basic - Intro):
`local$ ssh uniqname@peritia.eecs.umich.edu`
(the `local$` referring to a shell prompt on your system)

and then run `eecs201-test`:
`peritia$ eecs201-test basic-debug`
(the `peritia$` referring to a shell prompt on the server)

Preface

In this homework you'll be provided yet another zipped archive containing some starter files.

`https://www.eecs.umich.edu/courses/eecs201/fa2024/files/assignments/basic-debug.tar.gz`

Note that this assignment is to be submitted on a remote branch called `trunk`. Initialize a Git repository **inside of the extracted** `basic-debug` **directory**; as noted above `git init -b <branch name>` or `git init --initial-branch <branch name>` can initialize the repo with a different branch name (e.g. `trunk` as per the submission instructions). If your version of Git is too old for these options, you could create the `trunk` branch afterwards after your first commit, or you can set the local branch's tracking information manually. Create a file called `report.txt` in this directory.

Add all of the present files and commit them.

Create a **private** project named `eecs201-basic-debug` on the EECS GitLab (`gitlab.eecs.umich.edu`) and add the instructor `brng` as a **Reporter**. Set this EECS GitLab project as your remote: you'll be pushing to it in order to submit.

For this assignment we'll be utilizing an interactive command-line debugger, your preference of GDB (associated with the GNU project that brings us GCC) or LLDB (associated with the LLVM project that brings us Clang), and Valgrind. If you are on macOS, you might want to look at LLDB since it's better integrated with macOS. Unfortunately, Valgrind can be a bit of a hassle for macOS: there's ways to install it (e.g. with Homebrew and the https://github.com/LouisBrunner/valgrind-macos), but if your Mac has an M1 chip you might be out of luck, and as of this writing the patched Valgrind don't support macOS 11 and later when it comes to leak detection due to changes in macOS internals. However, in any case you are free to use the course server (peritia.eecs.umich.edu) to complete this assignment: it has GDB, LLDB, and Valgrind already installed, where these tools will work as expected.
The package names for Ubuntu are `gdb`, `lldb`, and `valgrind`. If you haven't already installed `g++`, the GNU C++ compiler, you can install the `g++` package.

1 Debugging with GDB (or LLDB)

1. `cd` into the `1` directory.
2. Use `make` to build the `seqmean` program. What this program does is take the average of a sequence of integers from 1 to *n*, where *n* is an argument.
3. Run `$ ./seqmean 3`. Note that a segfault occurs.
4. Run `$ gdb seqmean` or `$ lldb seqmean`. This will run GDB or LLDB and it will load the `seqmean` executable.
5. Look at the line above the `(gdb)` (if you're running GDB): notice how it mentions that there are no debugging symbols. Debugging symbols are what GDB and LLDB uses to be able to have a sense of what code "lines" and other things are.
6. Let's get some debugging symbols: exit GDB with `(gdb) quit` or `(gdb) q`, or exit LLDB with `(lldb) quit` or `(lldb) q`. From now on, the `(gdb)` prompt will indicate commands you can enter for GDB and the `(lldb)` prompt will indicate commands you can enter for LLDB.
7. Open the `Makefile` in an editor. Set the `CFLAGS` variable so that it contains `-g`. The `-g` flag will cause `gcc` to compile with debugging symbols.
8. Use `make` to clean and rebuild the program, and then open the program in GDB or LLDB again.
9. To run the program, run `(gdb) run 3`, `(gdb) r 3`, `(lldb) run 3`, or `(lldb) r 3`. This will run the loaded program with the given arguments: in this case `3`.
10. GDB/LLDB will catch the program at the segfault, allowing us to investigate the state of the program. It's kind enough to tell us what line of what file it happened at.
11. With `(gdb) backtrace`, `(gdb) bt` (or `(gdb) where`), or `(lldb) bt` we can get a backtrace, which shows the function calls and their stack frames that lead up to the current place. Frame #0 represents the current function and frame #1 represents the caller of that function (and frame #2 would be its caller and so on).
12. `(gdb) frame`, `(gdb) f`, or `(lldb) f` will show the current frame and the line of code it's at.
13. Segfaults ("segmentation faults") come about due to accessing memory that doesn't "exist" (or unmapped for you OS folks ;)) or if there is a permission access violation (e.g. writing to a read-only spot). In this line, the only operation that performs a risky memory operation is the struct pointer member access operation (`->`).
14. That operation looks risky: let's print out the value of the pointer variable (which is the argument of the function): `(gdb) print d`, `(gdb) p d`, `(lldb) print d`, or `(lldb) p d`. (Note: `(gdb) frame` and `(lldb) f` also shows the arguments and their values that passed into the function).
15. Note that the value of `d` is `0x0` / `NULL`, a null pointer. Null pointer dereferences are some common ways a segfault can happen. But where did it come from? Look at your backtrace again.
16. Let's switch to the frame of the calling function: run `(gdb) frame 1` or `(lldb) f 1` to investigate it.
17. `frame` / `f` conveniently prints out the line of code: we can see that at line 30 of `main.c`, a `NULL` (i.e. null pointer) is passed into the function instead of a pointer to a `data` structure.
18. Now that we've identified the bug, let's fix it. Exit GDB or LLDB and open `main.c` for editing. Fix line 30 so that `data_calc_stats` takes in a pointer to the `d` variable, which is an instance of the `data` structure.
19. Rebuild the program and run `$ ./seqmean 3` again. There should not be a segfault this time.
20. Add and commit the `main.c` file.

Note how the result is incorrect: $\frac{1+2+3}{3}$ should equal 2, not 1. Let's figure out what's wrong and fix the bug.

1. Open the program in GDB or LLDB again, and run `(gdb) run 3`, `(gdb) r 3`, `(lldb) run 3`, `(lldb) r 3`.

2. You'll see that GDB/LLDB will run the program to completion. GDB/LLDB will remember the arguments you pass along with the run: `(gdb) run` / `(lldb) run` will cause it to run with `3` as the argument again.

3. It's not particularly interesting if it runs to completion. Let's set a breakpoint at `main()` to have it stop there: Run `(gdb) break main`, `(gdb) b main`, `(lldb) breakpoint set -name main`, `(lldb) breakpoint set -n main`, or `(lldb) b main`.
   You can list out the breakpoints with `(gdb) info breakpoint`, `(gdb) info b`, `(lldb) breakpoint list`.

4. Run the program again. It should stop at `main()`.

5. Let step through the code: run `(gdb) next`, `(gdb) n`, `(lldb) next`, `(lldb) n`, or `(lldb) thread step-over` to step over a line.

6. Hit Enter/Return without typing out any command. Entering an empty command will run the previously entered command.

7. Place a breakpoint at line 30 with `(gdb) break 30`, `(gdb) b 30`, `(lldb) b 30`, `(lldb) breakpoint set -line 30`, `(lldb) breakpoint set -l 30`. Implicitly, without specifying a file, the current file that the current line is in will be referred to when it comes to line numbers and function names. `(gdb) break main.c:30`, `(gdb) break main.c:main`, `(lldb) b main.c:30`, and `(lldb) b main.c:main` are examples of explicitly referring to a file.

8. Continue execution with `(gdb) continue`, `(gdb) c`, `(lldb) continue`, or `(lldb) c`.

9. This is line that calls the function that calculates the mean. Step *into* the function with `(gdb) step`, `(gdb) s`, `(lldb) step`, `(lldb) s`, `(lldb) thread step-in`.

10. Let's see if the sum is calculated properly: place a breakpoint at line 24. Continue so that this breakpoint gets triggered.

11. Print out the `sum` variable. It's 3! $1 + 2 + 3 \neq 3$!

12. That's curious: maybe the data was bad. Let's investigate the `buffer` member of `d`: try `(gdb) p d->buffer` or `(lldb) p d->buffer`.

13. Unsurprisingly, GDB/LLDB prints out the actual pointer value that `buffer` holds. We can dereference it: `(gdb) p *d->buffer` or `(lldb) p *d->buffer`.

14. Well, that gives us only the 0th element. We can get GDB/LLDB to give us elements after it as well: `(gdb) p *d->buffer@3` or `(lldb) parray 3 d->buffer` (or if we're being really fancy `(gdb) p *d->buffer@d->size` or `(lldb) parray `d->size` d->buffer`).

15. Now we can see the issue. The buffer should be a sequence of 3 integers starting at 1, not 0: the buffer's contents should be `{1, 2, 3}`.

16. I'll let you figure out the fix for this :) Investigate the C source and header files, and take advantage of whatever debugging knowledge you have.

17. When you're done with the fix, add and commit the files associated with the fix. Make sure to remove any debug prints that you make!

2 Finding memory issues with Valgrind

1. `cd` into the `2` directory.

2. Build and run the program. You should see `n = 5` and `0.2 0.4 0.6 0.4 0.2`. (Some people's systems may result in a segfault; we'll deal with it soon :)).

3. While it may be functionally correct, there may be bad memory usage lurking about which may cause issues if this were a more complicated program.

4. We can use Valgrind (more specifically it's Memcheck utility) to analyze the memory usage and access of your program. It's easy: just run `$ valgrind ./conv`.
5. You'll see a lot printed out (which by default goes to `stderr`). There is a lot of good info here: we get info about bad writes and reads and there's a memory leak summary as well.
6. First, let's deal with the memory leaks since those can be fairly easy to deal with. Take a note of the message that says "Rerun with --leak-check=full to see details of leaked memory", and do that.
7. The resulting heap summary will list out where memory is being lost and is nice enough to tell us the function call stack that associated with the allocations. We can see that the buffer allocated by Signal constructor is never getting deallocated.
8. Fix this memory leak by adding the necessary `delete[]` to `Signal`'s destructor.
9. Rebuild and rerun the program with Valgrind. The memory leak should be gone. Now onto the invalid write and read.
10. First, let's look at the invalid write which occurs in `main.cpp`. Note how in the `for` loop, `x[3]` is written to due to the `for` loop condition, writing past the end of the allocated buffer. Change the `for` loop's condition to fix this.
11. Rebuild and rerun with Valgrind. The invalid write should be gone now. The invalid read is still there, however. Let's look at the associated line in `conv.cpp`.
12. The line points to an accumulation: `sum += x[xI] * h[hI]`. The culprits of this invalid read would either be the `x[xI]` or `h[hI]`, as we could be indexing out of the bounds of the buffers for these `Signal`s.
13. If we investigate the code a bit, the `for` loop that immediately contains the `if` statement checks that `xI` is less than `xN`, the size of `x`'s buffer. Since in C/C++ indexing is 0-indexed the indexing is correct for `x` since the last element is at index `xN - 1`.
14. If we look at the `if` that controls the accumulation, we can see that it's checking if `hI` is in the bounds of `h`. Notice anything wrong here?
15. Fix that issue in the `if` condition. When you rebuild and rerun, you should get a clean report from Valgrind.
16. Now that it's leak and memory error free, add and commit the files that you made fixes for.

3 Conclusion

1. Add and commit any changes you intend to submit.
2. Fill out the `report.txt` file in the following steps:
3. On the first line provide an integer time in minutes of how long it took for you to complete this assignment. It should just be an integer: no letters or words.
4. On the second line and beyond, write down what you learned while doing this assignment. If you already knew how to do all of this, put down "N/A".
5. Commit your `report.txt` file and push your commits to your remote.s
6. Run the autograder!