Modern C++ Programming

Debugging

Table of Context

- 1 Debugging
- 2 Assertions
- **3** Execution Debugging

Table of Context

4 Memory Debugging

- valgrind
- Stack Protection

5 Sanitizers

- Address Sanitizer
- Leak Sanitizer
- Memory Sanitizers
- Undefined Behavior Sanitizer

6 Debugging Summary

Table of Context

7 Code Checking and Analysis

- Compiler Warnings
- Static Analyzers

8 Code Testing

- Unit Test
- Code Coverage
- Fuzz Testing

9 Code Quality

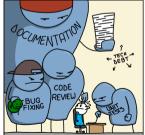
clang-tidy

Feature Complete









Debugging

Is this a bug?

"Software developers spend 35-50 percent of their time validating and debugging software. The cost of debugging, testing, and verification is estimated to account for 50-75 percent of the total budget of software development projects"

from: John Regehr (on Twitter)
The Debugging Mindset

Program Errors

A **program error** is a set of conditions that produce an *incorrect result* or *unexpected behavior*, including performance regression, memory consumption, early termination, etc.

We can distinguish between two kind of errors:

Recoverable Conditions that are not under the control of the program. They indicates "exceptional" run-time conditions. e.g. file not found, bad allocation, wrong user input, etc.

Unrecoverable *It is a synonym of a bug.* The program must terminate. e.g. out-of-bound, division by zero, etc.

Sometimes a *recoverable* error is considered *unrecoverable* if it is extremely rare and difficult to handle, e.g. bad allocation due to out-of-memory error

Dealing with Program Errors and Bugs

Software defects can be identifies by:

Dynamic Analysis A *mitigation* strategy that acts on the runtime state of a program.

Techniques: Print, run-time debugging, sanitizers, fuzzing, unit test support Limitations: Infeasible to cover all program states

Static Analysis A *proactive* strategy that examines the source code for (potential) errors.

Techniques: Warnings, static analysis tool, compile-time checks

Limitations: Turing's undecidability theorem, exponential code paths

Assertions

Unrecoverable Errors and Assertions

<u>Unrecoverable</u> errors cannot be handled. They should be prevented by using *assertion* for ensuring *pre-conditions* and *post-conditions*

An **assertion** is a statement to detect a violated assumption. An assertion represents an *invariant* in the code

It can happen both at run-time (assert) and compile-time (static_assert). Run-time assertion failures should never be exposed in the normal program execution (e.g. release/public)

Assertion

```
#include <cassert> // <-- needed for "assert"</pre>
#include <cmath> // std::is finite
#include <type traits> // std::is arithmetic v
template<typename T>
T sqrt(T value) {
    static_assert(std::is_arithmetic_v<T>,  // precondition
                 "T must be an arithmetic type");
    assert(std::is_finite(value) && value >= 0); // precondition
    int ret = ...
                                                // sart computation
    assert(std::is_finite(value) && ret >= 0 && // postcondition
          (ret == 0 || ret == 1 || ret < value)):
   return ret:
```

Assertions may slow down the execution. They can be disable by define the NDEBUG macro

```
#define NDEBUG // or with the flag "-DNDEBUG"
```

Execution

Lxecution

Debugging

Execution Debugging (gdb)

How to compile and run for debugging:

- stores the symbol table information in the executable (mapping between assembly

- -00 Disable any code optimization for helping the debugger. It is implicit for most compilers
 - -g Enable debugging
 - and source code lines)
 - for some compilers, it may disable certain optimizations
 - slow down the compilation phase and the execution
- -g3 Produces enhanced debugging information, e.g. macro definitions. Available for most compilers. Suggested instead of -g

gdb - Breakpoints/Watchpoints

Command	Abbr.	Description
breakpoint <file>:<line></line></file>	Ъ	insert a breakpoint in a specific line
${\tt breakpoint} < \!\! function_name \!\! >$	ъ	insert a breakpoint in a specific function
breakpoint < ref > if < condition >	b	insert a breakpoint with a conditional statement
delete	d	delete all breakpoints or watchpoints
${\tt delete} < breakpoint_number>$		delete a specific breakpoint
<pre>clear [function_name/line_number]</pre>		delete a specific breakpoint
enable/disable < breakpoint_number >		enable/disable a specific breakpoint
watch < expression >		stop execution when the value of expression changes (variable, comparison, etc.)

gdb - Control Flow

Command	Abbr.	Description
run [args]	r	run the program
continue	С	continue the execution
finish	f	continue until the end of the current function
step	s	execute next line of code (follow function calls)
next	n	execute next line of code
until <pre>cprogram_point></pre>		continue until reach line number, function name, address, etc.
CTRL+C		stop the execution (not quit)
quit	q	exit
help [<command/>]	h	show help about command

gdb - Stack and Info

Command	Abbr.	Description
list	1	print code
<pre>list <function #start,#end="" or=""></function></pre>	1	print function/range code
up	u	move up in the call stack
down	d	move down in the call stack
backtrace	bt	prints stack backtrace (call stack)
$\mathtt{backtrace} < \mathit{full} >$	bt	print values of local variables
info <args locals="" variables=""></args>		print current function arguments/local variables/all variables
<pre>info <bre> <bre>breakpoints/watchpoints/registers</bre></bre></pre>		show information about program breakpoints/watchpoints/registers

gdb - Print

Command	Abbr.	Description
print <variable></variable>	p	print variable
print/h < variable>	p/h	print variable in hex
print/nb <variable></variable>	p/ <mark>n</mark> b	print variable in binary (n bytes)
print/w <address></address>	p/w	print address in binary
p /s <char address="" array=""></char>		print char array
p *array_var@n		print n array elements
p (int[4]) < address>		print four elements of type int
p *(char**)& <std::string></std::string>		print std::string

gdb - Disassemble

Command	Description
disasseble <function_name></function_name>	disassemble a specified function
${\tt disasseble} < {\tt 0xStart,0xEnd} \ {\tt addr} >$	disassemble function range
nexti <variable></variable>	execute next line of code (follow function calls)
stepi < <i>variable</i> >	execute next line of code
x/nfu <address></address>	examine address n number of elements, f format (d: int, f: float, etc.), u data size (b: byte, w: word, etc.)

The debugger automatically stops when:

- breakpoint (by using the debugger)
- assertion fail
- segmentation fault
- trigger software breakpoint (e.g. SIGTRAP on Linux) github.com/scottt/debugbreak

Full story: www.yolinux.com/TUTORIALS/GDB-Commands.html (it also contains a script to de-referencing STL Containers)

gdb reference card V5 link

Memory Debugging

issues"

Matt Miller, Microsoft Security Engineer

Chromium Security Report

"you can expect at least 65% of your security vulnerabilities to be

"Chrome: 70% of all security bugs are memory safety issues"

"70% of all the vulnerabilities in Microsoft products are memory safety

caused by memory unsafety"

What science can tell us about C and C++'s security

Terms like buffer overflow, race condition, page fault, null pointer, stack exhaustion, heap exhaustion/corruption, use-after-free, or double free — all describe **memory** safety vulnerabilities

Solutions:

- Run-time check
- Static analysis
- Avoid unsafe language constructs



<u>valgrind</u> is a tool suite to automatically detect many memory management and threading bugs

How to install the last version:

```
$ wget ftp://sourceware.org/pub/valgrind/valgrind-3.20.tar.bz2
$ tar xf valgrind-3.20.tar.bz2
$ cd valgrind-3.20
$ ./configure --enable-lto
$ make -j 12
$ sudo make install
$ sudo apt install libc6-dbg #if needed
```

some linux distributions provide the package through apt install valgrid, but it could be an old version

Basic usage:

compile with -g

```
$ valgrind ./program <args...>
```

Output example 1:

Output example 2:

```
!!memory leak
==19182== 40 bytes in 1 blocks are definitely lost in loss record 1 of 1
==19182==
           at 0x1B8FF5CD: malloc (vg replace malloc.c:130)
==19182==
           by 0x8048385: f (main.cpp:5)
==19182==
           by 0x80483AB: main (main.cpp:11)
==60127== HEAP SUMMARY:
==60127==
           in use at exit: 4.184 bytes in 2 blocks
==60127== total heap usage: 3 allocs, 1 frees, 4,224 bytes allocated
==60127==
==60127== LEAK SUMMARY:
==60127==
            definitely lost: 128 bytes in 1 blocks
                                                    !!memorv leak
==60127==
           indirectly lost: 0 bytes in 0 blocks
==60127==
              possibly lost: 0 bytes in 0 blocks
==60127==
            still reachable: 4.184 bytes in 2 blocks !!not deallocated
==60127==
                 suppressed: 0 bytes in 0 blocks
```

Memory leaks are divided into four categories:

- Definitely lost
- Indirectly lost
- Still reachable
- Possibly lost

When a program terminates, it releases all heap memory allocations. Despite this, leaving memory leaks is considered a *bad practice* and *makes the program unsafe* with respect to multiple internal iterations of a functionality. If a program has memory leaks for a single iteration, is it safe for multiple iterations?

A robust program prevents any memory leak even when abnormal conditions occur

Definitely lost indicates blocks that are *not deleted at the end of the program* (return from the main() function). The common case is local variables pointing to newly allocated heap memory

```
void f() {
    int* y = new int[3]; // 12 bytes definitely lost
}
int main() {
    int* x = new int[10]; // 40 bytes definitely lost
    f();
}
```

Indirectly lost indicates blocks pointed by other heap variables that are not deleted. The common case is global variables pointing to newly allocated heap memory

```
struct A {
    int* array;
};

int main() {
    A* x = new A;  // 8 bytes definitely lost
    x->array = new int[4]; // 16 bytes indirectly lost
}
```

Still reachable indicates blocks that are *not deleted but they are still reachable at the end of the program*

```
int* array;
int main() {
    array = new int[3];
}
// 12 bytes still reachable (global static class could delete it)
```

Possibly lost indicates blocks that are still reachable but pointer arithmetic makes the deletion more complex, or even not possible

Advanced flags:

- --leak-check=full print details for each "definitely lost" or "possibly lost" block, including where it was allocated
- --show-leak-kinds=all to combine with --leak-check=full. Print all leak kinds
- --track-fds=yes list open file descriptors on exit (not closed)
- --track-origins=yes tracks the origin of uninitialized values (very slow execution)

Track stack usage:

```
valgrind --tool=drd --show-stack-usage=yes ./program <args...>
```

Stack size check:

- -Wstack-usage=<byte-size> Warn if the stack usage of a function might exceed byte-size. The computation done to determine the stack usage is conservative (no VLA)
- -fstack-usage Makes the compiler output stack usage information for the program, on a per-function basis
- -Wvla Warn if a variable-length array is used in the code
- -Wvla-larger-than=<byte-size> Warn for declarations of variable-length arrays whose size is either unbounded, or bounded by an argument that allows the array size to exceed byte-size bytes

Adding _FORTIFY_SOURCE define, the compiler provides buffer overflow checks for the following functions:

```
memcpy, mempcpy, memmove, memset, strcpy, stpcpy, strncpy, strcat, strncat, sprintf,
vsprintf, snprintf, vsnprintf, gets.

#include <cstring> // std::memset
#include <string> // std::stoi
int main(int argc, char** argv) {
   int size = std::stoi(argv[1]);
   char buffer[24];
   std::memset(buffer, 0xFF, size);
}
```

```
$ gcc -01 -D_FORTIFY_SOURCE program.cpp -o program
$ ./program 12 # 0K
$ ./program 32 # Wrong
$ *** buffer overflow detected ***: ./program terminated
29/52
```

Sanitizers

Address Sanitizer

Sanitizers are compiler-based instrumentation components to perform *dynamic* analysis

Sanitizer are used during development and testing to discover and diagnose memory misuse bugs and potentially dangerous undefined behavior

Sanitizer are implemented in Clang (from 3.1), gcc (from 4.8) and Xcode

Project using Sanitizers:

- Chromium
- Firefox
- Linux kernel
- Android

Address Sanitizer

Address Sanitizer is a memory error detector

- heap/stack/global out-of-bounds
- memory leaks
- use-after-free, use-after-return, use-after-scope
- double-free, invalid free
- initialization order bugs
- * Similar to valgrind but faster (50X slowdown)

```
clang++ -01 -g -fsanitize=address -fno-omit-frame-pointer cprogram>
```

- -01 disable inlining
 - -g generate symbol table
 - clang.llvm.org/docs/AddressSanitizer.html
 - github.com/google/sanitizers/wiki/AddressSanitizer
 - gcc.gnu.org/onlinedocs/gcc/Instrumentation-Options.html

Leak Sanitizer

LeakSanitizer is a run-time memory leak detector

- integrated into AddressSanitizer, can be used as standalone tool
- * almost no performance overhead until the very end of the process

```
g++ -01 -g -fsanitize=address -fno-omit-frame-pointer clang++ -01 -g -fsanitize=leak -fno-omit-frame-pointer cprogram>
```

- clang.llvm.org/docs/LeakSanitizer.html
- github.com/google/sanitizers/wiki/AddressSanitizerLeakSanitizer
- gcc.gnu.org/onlinedocs/gcc/Instrumentation-Options.html

Memory Sanitizers

Memory Sanitizer is detector of uninitialized reads

- stack/heap-allocated memory read before it is written
- * Similar to valgrind but faster (3X slowdown)

```
clang++ -01 -g -fsanitize=memory -fno-omit-frame-pointer program>
```

```
-fsanitize-memory-track-origins=2 track origins of uninitialized values
```

Note: not compatible with Address Sanitizer

- clang.llvm.org/docs/MemorySanitizer.html
- github.com/google/sanitizers/wiki/MemorySanitizer
 - gcc.gnu.org/onlinedocs/gcc/Instrumentation-Options.html

Undefined Behavior Sanitizer

UndefinedBehaviorSanitizer is a undefined behavior detector

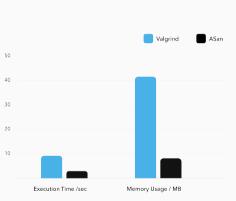
- signed integer overflow, floating-point types overflow, enumerated not in range
- out-of-bounds array indexing, misaligned address
- divide by zero
- etc.
- * Not included in valgrind

```
clang++ -01 -g -fsanitize=undefined -fno-omit-frame-pointer program>
```

- -fsanitize=integer Checks for undefined or suspicious integer behavior (e.g. unsigned integer overflow)
- -fsanitize=nullability Checks passing null as a function parameter, assigning null to an Ivalue, and returning null from a function
 - clang.llvm.org/docs/UndefinedBehaviorSanitizer.html
 - gcc.gnu.org/onlinedocs/gcc/Instrumentation-Options.html

Sanitizers vs. Valgrind

Bug	Valgrind detection	ASan detection	
Uninitialized memory read	Yes	No *	
Write overflow on heap	Yes	Yes	
Write overflow on stack	No	Yes	
Read overflow on heap	Yes	Yes	50
Read underflow on heap	Yes	Yes	40
Read overflow on stack	No	Yes	30
Use-after-free	Yes	Yes	
Use-after-return	No	Yes	20
Double-free	Yes	Yes	10
Memory leak	Yes	Yes	
Undefined behavior	No	No **	



Debugging Summary

How to Debug Common Errors

Segmentation fault

- gdb, valgrind, sanitizers
- ullet Segmentation fault when just entered in a function o stack overflow

Double free or corruption

• gdb, valgrind, sanitizers

Infinite execution

■ gdb + (CTRL + C)

Incorrect results

lacktriangledown valgrind + assertion + gdb + sanitizers

Code Checking and

Analysis

Compiler Warnings

Enable specific warnings:

```
g++ -W<warning> <args...>
```

Disable specific warnings:

```
g++ -Wno-<warning> <args...>
```

Common warning flags to minimize accidental mismatches:

- **-Wall** Enables many standard warnings (\sim 50 warnings)
- -Wextra Enables some extra warning flags that are not enabled by -Wall (\sim 15 warnings)
- -Wpedantic Issue all the warnings demanded by strict ISO C/C++

Enable <u>ALL</u> warnings (only clang) -Weverything

Static Analyzers - clang static analyzer



The <u>Clang Static Analyzer</u> is a source code analysis tool that finds bugs in C/C++ programs at compile-time

It find bugs by reasoning about the semantics of code (may produce false positives) Example:

```
void test() {
   int i, a[10];
   int x = a[i]; // warning: array subscript is undefined
}
```

How to use:

```
scan-build make
```

scan-build is included in the LLVM suite

Static Analyzers - cppcheck



The <u>GCC Static Analyzer</u> can diagnose various kinds of problems in C/C++ code at compile-time (e.g. double-free, use-after-free, stdio related, etc) -fanalyzer

<u>cppcheck</u> provides code analysis to detect bugs, undefined behavior and dangerous coding construct. The goal is to detect only real errors in the code (i.e. have very few false positives)

```
cmake -DCMAKE_EXPORT_COMPILE_COMMANDS=ON .
cppcheck --enable=<enable_flags> --project=compile_commands.json
```

Static Analyzers - PVS-Studio, FBInfer



Customers: IBM, Intel, Adobe, Microsoft, Nvidia, Bosh, IdGames, EpicGames, etc.



<u>FBInfer</u> is a static analysis tool (also available online) to checks for null pointer deferencing, memory leak, coding conventions, unavailable APIs, etc.

Customers: Amazon AWS, Facebook/Ocolus, Instagram, Whatapp, Mozilla, Spotify, Uber, Sky, etc.

Static Analyzers - DeepCode, SonarSource

deepCode is an Al-powered code review system, with DEEP, ODE machine learning systems trained on billions of lines of code from open-source projects

Available for Visual Studio Code, Sublime, IntelliJ IDEA, and Atom

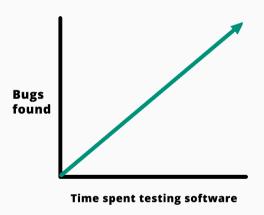


<u>SonarSource</u> is a static analyzer which inspects source code for bugs, code smells, and security vulnerabilities for multiple languages (C++, Java, etc.)

SonarLint plugin is available for Visual Code, Visual Studio Code, Eclipse, and IntelliJ IDEA

Code Testing

Code Testing



see Case Study 4: The \$440 Million Software Error at Knight Capital

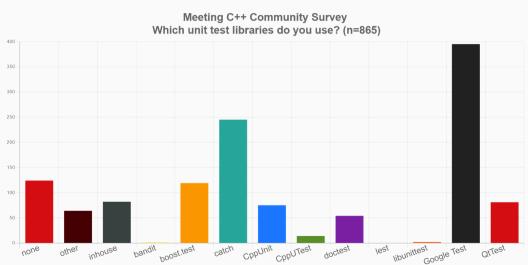
Unit testing involves breaking your program into pieces, and subjecting each piece to a series of tests

Unit Testing Benefits:

- Increases confidence in changing/ maintaining code
- The cost of fixing a defect detected during unit testing is lesser in comparison to that of defects detected at higher levels
- Debugging is easy. When a test fails, only the latest changes need to be debugged

C++ Unit testing frameworks:

- catch
- doctest
- Google Test
- CppUnit
- Boost.Test 43/52



$\underline{\mathtt{Catch2}}$ is a multi-paradigm test framework for C++

Catch2 features

- Header only and no external dependencies
- Assertion macro
- Floating point tolerance comparisons

Basic usage:

- Create the test program
- Run the test

```
$ ./test_program [<TestName>]
```

- github.com/catchorg/Catch2
- The Little Things: Testing with Catch2

```
#define CATCH CONFIG MAIN // This tells Catch to provide a main()
#include "catch.hpp" // only do this in one cpp file
unsigned Factorial(unsigned number) {
   return number <= 1 ? number : Factorial(number - 1) * number;</pre>
"Test description and tag name"
TEST_CASE( "Factorials are computed", "[Factorial]" ) {
    REQUIRE( Factorial(1) == 1 );
    REQUIRE( Factorial(2) == 2 );
    REQUIRE( Factorial(3) == 6 );
    REQUIRE( Factorial(10) == 3628800 );
float floatComputation() { ... }
TEST_CASE( "floatCmp computed", "[floatComputation]" ) {
    REQUIRE( floatComputation() == Approx( 2.1 ) );
```

Code coverage is a measure used to describe the degree to which the source code of a program is executed when a particular test suite runs

gcov is a tool you can use in conjunction with GCC to test code coverage in programs
gcovr is a utility for managing gcov and generating code coverage results

Step for code coverage:

- compile with --coverage flag (objects + linking)
- run the test
- visualize the results with gcovr

```
program.cpp:
#include <iostream>
#include <string>

int main(int argc, char* argv[]) {
    int value = std::stoi(argv[1]);
    if (value % 3 == 0)
        std::cout << "first\n";
    if (value % 2 == 0)
        std::cout << "second\n";
}</pre>
```

```
$ gcc -g --coverage program.cpp -o program
$ ./program 9
first
$ gcovr -r --html --html-details <path_to_cover>
# generate coverage.html
```

```
1: 4:int main(int argc, char* argv[]) {
1: 5: int value = std::stoi(argv[i]);
1: 6: if (value % 3 == 0)
1: 7: std::cout << "first\n";
1: 8: if (value % 2 == 0)

#####: 9: std::cout << "second\n";
4: 10:}
```



Coverage-Guided Fuzz Testing

A **fuzzer** is a specialized tool that tracks which areas of the code are reached, and generates *mutations* on the corpus of input data in order to *maximize* the code coverage

<u>LibFuzzer</u> is the library provided by LLVM and feeds fuzzed inputs to the library via a specific fuzzing entrypoint

The *fuzz target function* accepts an array of bytes and does something interesting with these bytes using the API under test:

Code Quality

lint: The term was derived from the name of the undesirable bits of fiber
clang-tidy provides an extensible framework for diagnosing and fixing typical programming errors, like style violations, interface misuse, or bugs that can be deduced via static analysis

```
$ cmake -DCMAKE_EXPORT_COMPILE_COMMANDS=ON .
$ clang-tidy -p .
```

clang-tidy searches the configuration file $\underline{.clang-tidy}$ file located in the closest parent directory of the input file

clang-tidy is included in the LLVM suite

Coding Guidelines:

- CERT Secure Coding Guidelines
- C++ Core Guidelines
- High Integrity C++ Coding Standard

Supported Code Conventions:

- Fuchsia
- Google
- LLVM

.clang-tidy

Bug Related:

- Android related
- Boost library related
- Misc
- Modernize
- Performance
- Readability
- clang-analyzer checks
- bugprone code constructors

```
Checks: 'android-*,boost-*,bugprone-*,cert-*,cppcoreguidelines-*, clang-analyzer-*,fuchsia-*,google-*,hicpp-*,llvm-*,misc-*,modernize-*, performance-*,readability-*'
```