



# Using Clang's source-based code coverage at scale

## Code coverage in Fuchsia

We collect incremental coverage at pre-submit testing and surface it in the code review tool.

```
4
5 #include "backtrace.h"
6
7 #include "threads_impl.h"
8
9 namespace __libc_sanitizer {
10
11 size_t BacktraceByFramePointer(cpp20::span<uintptr_t> pcs) {
12     struct FramePointer {
13         const FramePointer* fp;
14         uintptr_t pc;
15     };
16
17     auto on_stack = [&stack = __pthread_self()->safe_stack](const FramePointer* fp) -> bool {
18         uintptr_t address = reinterpret_cast<uintptr_t>(fp);
19         return address >= reinterpret_cast<uintptr_t>(stack.iov_base) &&
20             address < reinterpret_cast<uintptr_t>(stack.iov_base) + stack.iov_len;
21     };
22
23     uintptr_t ra = reinterpret_cast<uintptr_t>(__builtin_return_address(0));
24     auto fp = reinterpret_cast<const FramePointer*>(__builtin_frame_address(0));
25     size_t i = 0;
26     while (i < pcs.size() && on_stack(fp) && fp->pc != 0) {
27         if (i == 0 && fp->pc != ra) {
28             pcs[i++] = ra;
29         } else {
30             pcs[i++] = fp->pc;
31             fp = fp->fp;
32         }
33     }
34     if (i == 0 && i < pcs.size()) {
35         pcs[i++] = ra;
36     }
37
38     return i;
39 }
40
41 #if __has_feature(shadow_call_stack)
42
43 namespace {
44
```

## Code coverage in Fuchsia

We collect absolute coverage in continuous integration and surface it in the code search tool.

- Repository root
  - ▶ inttypes
  - ▶ libc++-stub
  - ▶ math
  - ▶ pthread
  - ▼ sanitizers
    - ▣ BUILD.gn
    - ▣ \_\_asan\_early\_init.c
    - ▣ \_\_sanitizer\_fast\_backtrac
    - ▣ asan-stubs.c
    - ▣ backtrace-tests.cc
    - ▣ **backtrace.cc**
    - ▣ backtrace.h
    - ▣ debugdata.cc
    - ▣ fuchsia-io-constants.h
    - ▣ hooks.c
    - ▣ hwasan-stubs.cc
    - ▣ hwasan-stubs.h
    - ▣ log.c
    - ▣ memory-snapshot.cc
    - ▣ sancov-stubs.cc
    - ▣ sancov-stubs.h
    - ▣ sanitizer-stubs.h
    - ▣ ubsan-stubs.cc
  - ▶ scudo
  - ▶ setjmp
  - ▶ stdio
  - ▶ stdlib
  - ▶ string
  - ▶ stubs
  - ▶ test
  - ▶ zircon

```
1 // Copyright 2021 The Fuchsia Authors. All rights reserved.
2 // Use of this source code is governed by a BSD-style license that can be
3 // found in the LICENSE file.
4
5 #include "backtrace.h"
6
7 #include <lib/arch/backtrace.h>
8
9 #include "threads_impl.h"
10
11 namespace __libc_sanitizer {
12
13 size_t BacktraceByFramePointer(cpp20::span<uintptr_t> pcs) {
14     struct IsOnStack {
15         bool operator()(const arch::CallFrame* fp) const {
16             const iovec& stack = __pthread_self()->safe_stack;
17             if (stack.iov_len < sizeof(*fp)) [[unlikely]] {
18                 // This should be impossible, but assume nothing in a critical
19                 // error-reporting path since this might be used after clobberation.
20                 return false;
21             }
22             const uintptr_t base = reinterpret_cast<uintptr_t>(stack.iov_base);
23             const uintptr_t frame = reinterpret_cast<uintptr_t>(fp);
24             return frame >= base && frame - base <= stack.iov_len - sizeof(*fp);
25         }
26     };
27     using FpBacktrace = arch::FramePointerBacktrace<IsOnStack>;
28
29     return arch::StoreBacktrace(FpBacktrace::BackTrace(), pcs, __builtin_return_address(0));
30 }
31
32 #if __has_feature(shadow_call_stack)
33
34 size_t BacktraceByShadowCallStack(cpp20::span<uintptr_t> pcs) {
35     const iovec& shadow_call_stack_block = __pthread_self()->shadow_call_stack;
36     return arch::StoreBacktrace(
37         arch::ShadowCallStackBacktrace{
38             {static_cast<const uintptr_t*>(shadow_call_stack_block.iov_base),
39              shadow_call_stack_block.iov_len / sizeof(uintptr_t)},
40             arch::GetShadowCallStackPointer(),
41             pcs, __builtin_return_address(0)});
42 }
43
44 #endif // __has_feature(shadow_call_stack)
45
46 } // namespace __libc_sanitizer
```

## Clang source-based code coverage

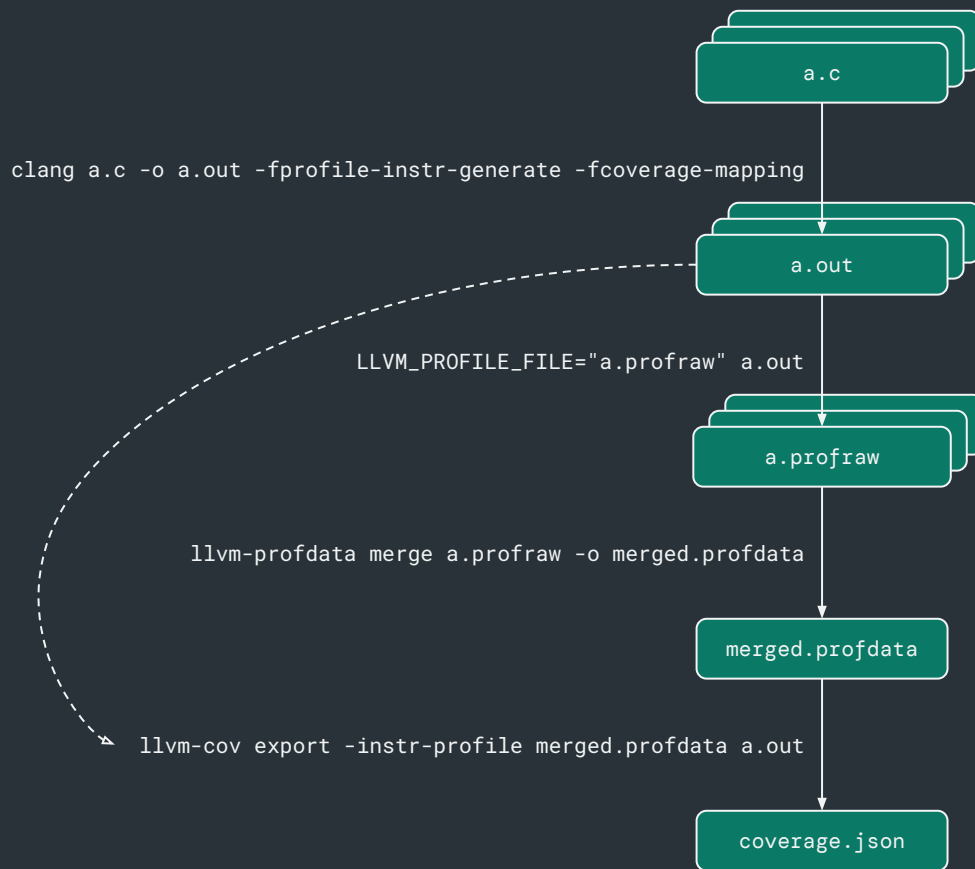
Combines profiling (PGO) instrumentation with mapping derived from AST and preprocessor information.

The instrumentation is applied early, before optimizations to avoid negative impact on coverage report quality.

This generates precise coverage data, but with significant performance overhead.

## In Fuchsia

14,000 sources  
4,500 tests  
5,000 binaries  
7,000 raw profiles  
1GB indexed profile



## Emitting profiles with abnormal termination

The profile runtime uses an `atexit()` hook to write out the raw profile to disk.

If the process terminates abnormally, `atexit()` hooks may not be executed resulting in missing coverage.

This is a problem for tests that spawn subprocesses such as "death tests".

# Emitting profiles during abnormal termination

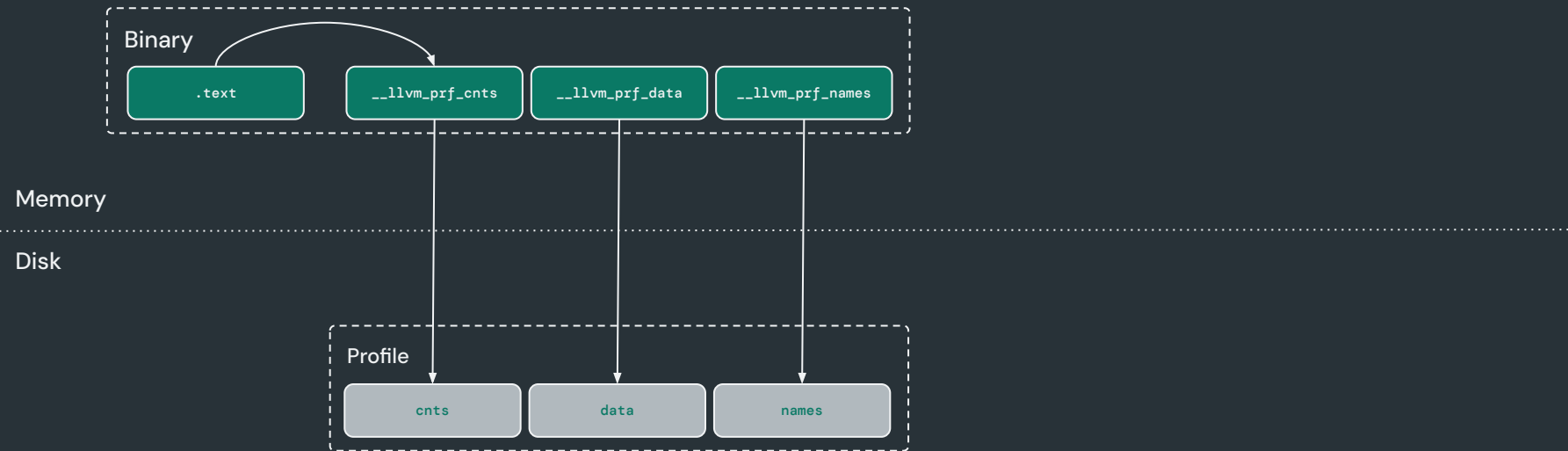
During abnormal termination, an empty profile is generated.

```
0 int main(int argc, char** argv) {  
0     if (argc != 1) {  
0         abort();  
0     }  
0     return 0;  
0 }
```

a.c

```
$ clang a.c -o a.out \  
    -fprofile-instr-generate -fcoverage-mapping  
$ LLVM_PROFILE_FILE="a.profraw" ./a.out LLVM DevMtg  
Aborted  
$ llvm-profdata merge -sparse a.profraw -o a.profdata  
$ llvm-cov show ./a.out -instr-profile=a.profdata
```

```
counter update:  
c = __profc[idx];  
c++;  
__profc[idx] = c
```

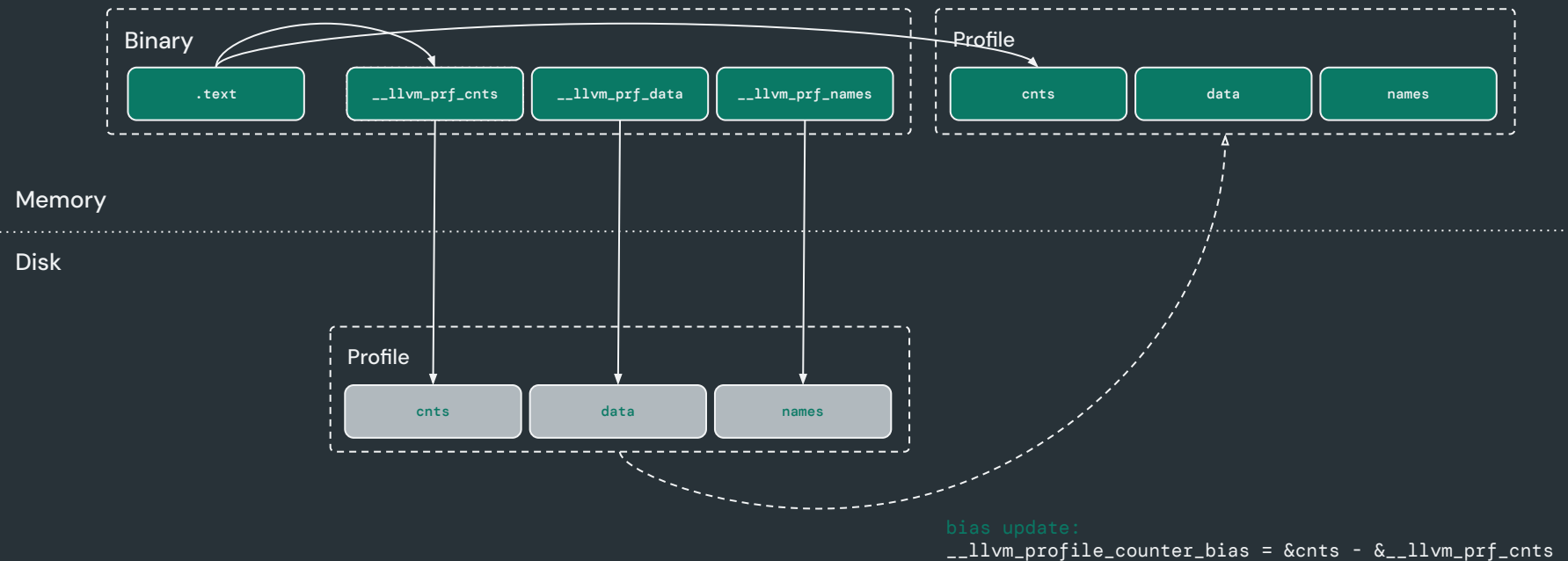




```

counter update:
c = *(&__profc[idx] + __llvm_profile_counter_bias)
c++
*(&__profc[idx] + __llvm_profile_counter_bias) = c

```



# Using runtime counter relocation

During abnormal termination, profile is written out as expected.

```
1 int main(int argc, char** argv) {
1   if (argc != 1) {
1     abort();
0   }
0   return 0;
1 }
```

a.c

Runtime counter relocation can be enabled by a backend option and requires %c flag.

```
$ clang a.c -o a.out \
    -fprofile-instr-generate -fcoverage-mapping \
    -mllvm -runtime-counter-relocation
$ LLVM_PROFILE_FILE="a%c.profraw" ./a.out LLVM DevMtg
Aborted
$ llvm-profdata merge -sparse a.profraw -o a.profdata
$ llvm-cov show ./a.out -instr-profile=a.profdata
```

## Writing counters on-the-fly

In Fuchsia, we use runtime counter relocation by default. Since the profile is emitted at the start of the program, and the counters are updated on-the-fly, abnormal termination is no longer an issue.

Runtime counter relocation introduces a level of indirection which results in runtime overhead and increased binary size.

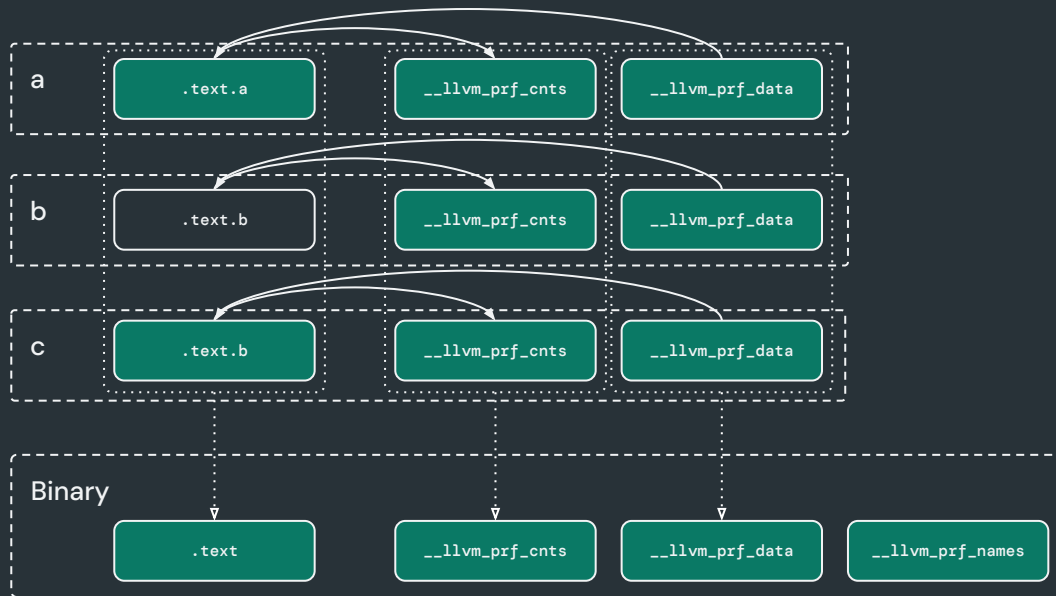
*Note that macOS uses a different approach called "continuous mode" which relies on overmapping.*

## Reducing the size of instrumented binaries

In a typical C/C++ and Rust binary, there is large number of unused functions.

In an uninstrumented build, these would be stripped by the linker `--gc-sections` feature (in ELF).

That was not possible for instrumented binaries as the metadata sections had references to the `.text` section which prevented the linker from discarding these.



## Support for ELF zero-flag section groups

We explored several potential solutions, we ended up introducing a new Comdat selection kind in LLVM IR: `nodeduplicate`

This is lowered to ELF zero-flag section group which is now supported by LLVM and LLD.

Addressing this issue reduced the size of instrumented binaries and generated profiles by 50% in Fuchsia.

*This approach could be used for other kinds of instrumentation that generates metadata sections.*

## Selective instrumentation

A patch typically only modifies a small subset of files.

We can significantly reduce the coverage overhead by only instrumenting the modified files.

We reuse the sanitizer special case list format to specify files/functions to allow/skip/forbid instrumentation for.

At the LLVM IR level, this translates to `noprofile` and `skipprofile` function attributes.

# Using selective instrumentation

Specify which functions and sources to allow/skip/forbid instrumentation for using the sanitizer special case list format.

`-fprofile-list` flag is used to pass the list to compiler.

```
# Only apply to frontend instrumentation.  
[clang]  
  
# Instrument function named foo.  
function:foo=allow  
  
# Instrument all source files in lib/foo.  
source:lib/foo/*.c=allow  
  
# Otherwise skip instrumentation.  
default:skip
```

cov.list

```
$ clang a.c -o a.out -fprofile-list=cov.list \  
-fprofile-instr-generate -fcoverage-mapping
```



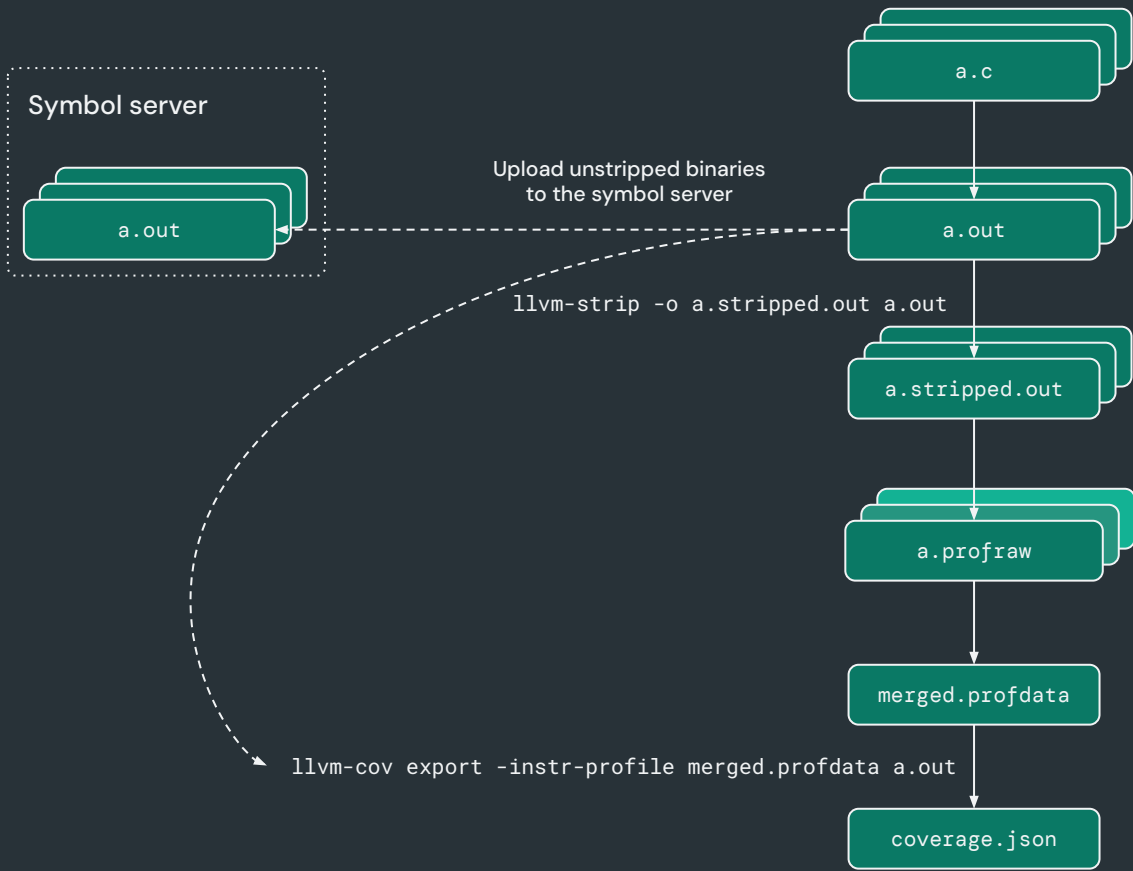
# Fuchsia coverage pipeline

We use different machines for building, running tests and coverage post-processing.

We always strip binaries to reduce their size, and upload the unstripped binaries to symbol server.

We need to use unstripped binaries for coverage post-processing.

We need a way to associate the collected profiles with unstripped binaries during post-processing.



# Embedding binary ID in profiles

Binary ID refers to the unique identifiers for binaries in different file formats.

- Build ID as a unique identifier in ELF
- LC\_UUID as an identifier in Mach-O
- GUID used in COFF

Binary ID embedded inside the profile can be used to map the profile back to the binary that produced it.

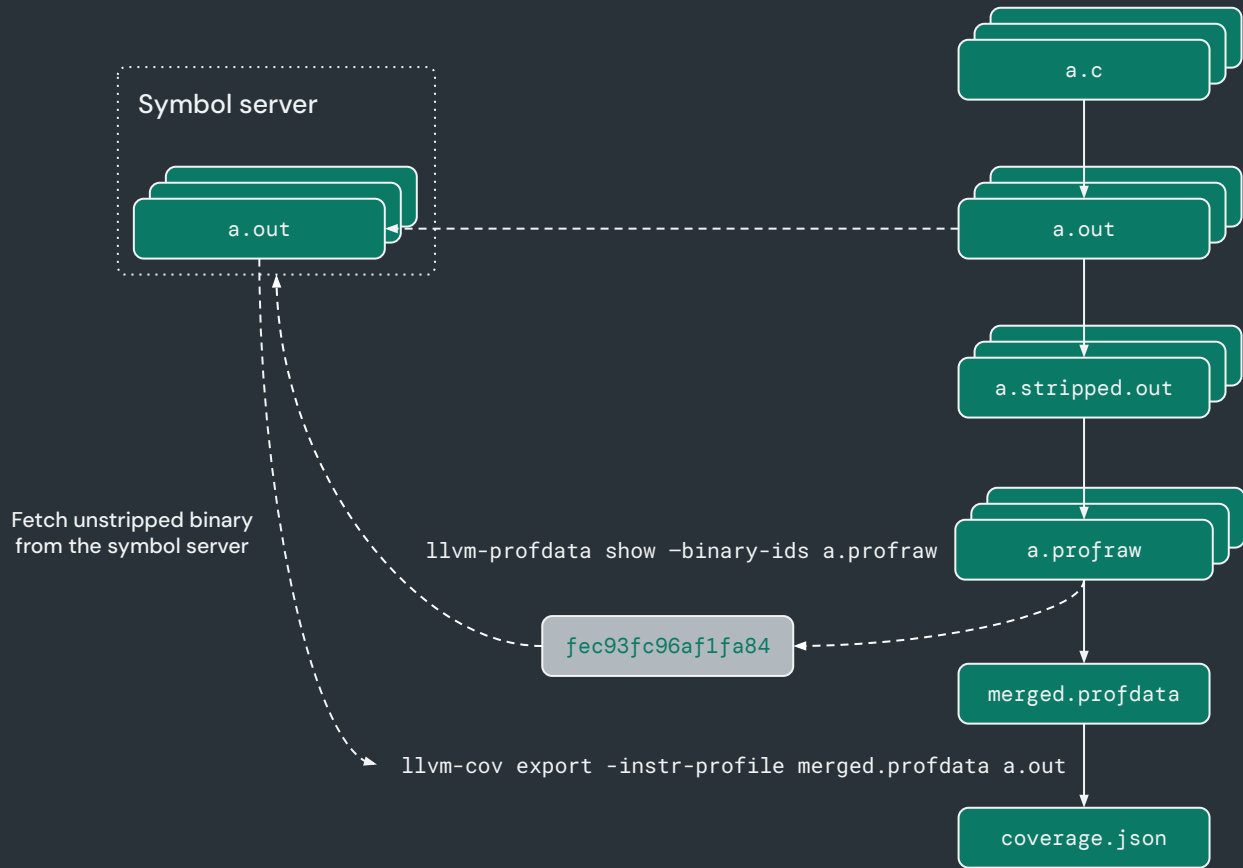
*Note that GCC generates Build ID by default, Clang can be opted in by the `ENABLE_LINKER_BUILD_ID` CMake flag, you can also use the `--build-id` linker flag.*

## Support for binary ID in coverage

The profile runtime writes binary ID into the profile as an optional field.

`llvm-profdata` can be used to display it.

```
$ clang a.c -o a.out -Wl,--build-id \  
    -fprofile-instr-generate -fcoverage-mapping  
$ LLVM_PROFILE_FILE="a.profraw" ./a.out  
$ llvm-profdata show --binary-ids a.profraw  
Binary IDs:  
02274a7974e4593e65b37d81ce602dba1b54edee
```



## Using binary ID in profiles

In Fuchsia, using binary ID simplified the pipeline, increased the reliability and reduced coverage post-processing time by 25%.

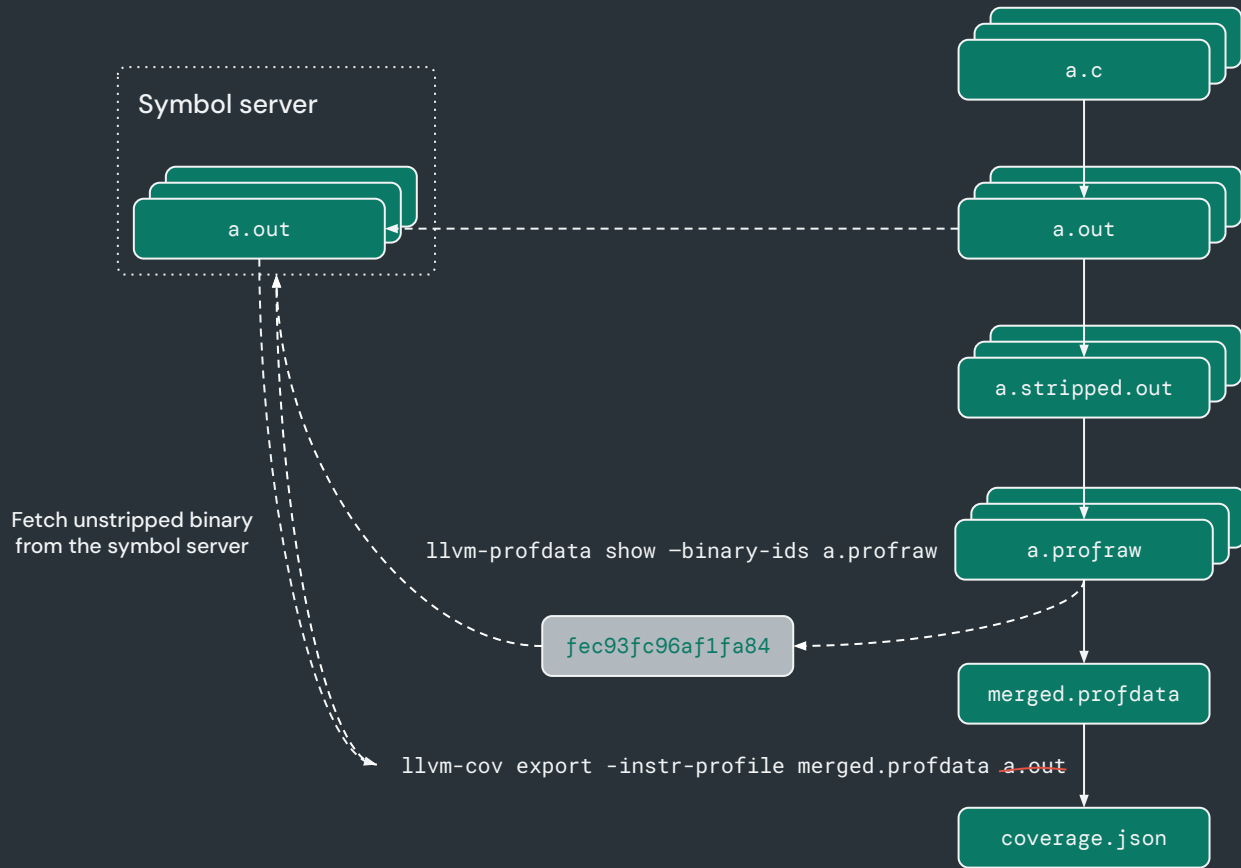
# Debuginfod support in LLVM

debuginfod is a simple HTTP API that can be used to fetch unstripped binaries by build ID.

LLVM libDebuginfod is a client/server implementation which can be easily integrated into LLVM tools.

debuginfod is already supported by `llvm-symbolize` and `llvm-objdump`.

[Introducing debuginfod, the elfutils debuginfo server](#)





## Using debuginfod for coverage

We added debuginfod support to `llvm-cov` to fetch binaries using the binary IDs embedded in the indexed profile.

This further simplified our infrastructure which already uses debuginfod for symbolization.

## Per-directory index in coverage reports

Co-mentored **Yuhao Gu** who participated in the [Google Summer of Code Program](#) (GSoC) with LLVM organization this summer.

Improved the readability of textual and HTML coverage reports by enhancing `llvm-cov`.

[Enhancing llvm-cov to Generate Hierarchical Coverage Reports](#)



## Per-directory index in coverage reports

llvm-cov can generate per-directory index with `-show-directory-coverage`

[Support directory layout in coverage reports](#)

### Coverage Report ([https://r/r/x/w/fuchsia/](#))

Created: 2023-10-05 22:15

Click [here](#) for information about interpreting this report.

Filename	Function Coverage	Line Coverage	Region Coverage	Branch Coverage
<a href="#">build/</a>	68.27% (426/624)	90.46% (2087/2305)	77.29% (1674/2166)	56.45% (78/124)
<a href="#">example/</a>	84.83% (1381/1628)	87.45% (12854/12812)	70.23% (12555/17851)	49.17% (1760/354)
<a href="#">out/llvm-objdump/</a>	37.93% (7788/20536)	45.88% (16430/35845)	41.53% (44906/110457)	30.18% (6861/22384)
<a href="#">pegbuild/clang_opts/</a>	4.77% (439/921)	9.10% (3684/4052)	13.86% (1527/1087)	40.28% (405/149)
<a href="#">obj/</a>	83.89% (1562/1398)	85.44% (7447/8718)	71.01% (20317/28543)	34.16% (13328/6288)
<a href="#">obj/</a>	76.38% (22160/29015)	81.88% (22617/27671)	60.41% (27333/38812)	48.12% (4040/8374)
<a href="#">clang_opts/</a>	24.85% (5921/2388)	24.58% (5262/2148)	29.36% (4293/1465)	19.43% (1497/762)
<a href="#">obj/</a>	81.42% (1827/1927)	88.29% (10431/12095)	50.26% (20354/40437)	42.19% (1017/152)
<a href="#">obj/</a>	71.44% (12630/1768)	82.21% (12950/1565)	50.05% (18356/10518)	42.38% (8770/2054)
<b>Totals</b>	<b>39.92%</b> (13912/3484)	<b>52.57%</b> (41798/79328)	<b>53.65%</b> (47518/88564)	<b>34.79%</b> (18214/22376)

Generated by llvm-cov - llvm version 18.0.0git

## Per-directory index in coverage reports

This feature is also enabled for LLVM coverage reports.

- Site Map:
  - Overview
  - Features
  - Documentation
  - Command Guide
  - FAQ
  - Publications
  - LLVM Projects
  - Open Projects
  - LLVM Users
  - Bug Tracker
  - LLVM Logo
  - Blog
  - Meetings
  - LLVM Foundation
- Download!
  - Download now:
  - LLVM 16.0.6
  - All Releases
  - API Packages
  - Feature Snapshot
  - Packages
  - Pre-releases
- View the open-source license
- Search this Site
  - Search
- Useful Links
  - Forums
  - LLVM Discussion
- Mailing Lists:
  - Commits List
- Discord (Real-time Chat):
  - Discord
- JRC Channel:
  - irc-offic-net.zulip
- Calendar:
  - LLVM Community Calendar
- Dev. Resources:
  - doxygen
  - Source (GitHub)
  - Code Review
  - Blog
  - Bug Tracker
  - Buildbot
  - Green Dragon
  - LLVM
  - Source build
  - llvm-con

LLVM Coverage bot

## LLVM Overview

The LLVM Project is a collection of modular and reusable compiler and toolchain technologies. Despite its name, LLVM has little to do with traditional virtual machines. The name "LLVM" itself is not an acronym; it is the full name of the project.

LLVM began as a [research project](#) at the [University of Illinois](#), with the goal of providing a modern, SSA-based compilation strategy capable of supporting both static and dynamic compilation of arbitrary programming languages. Since then, LLVM has grown to be an umbrella project consisting of a number of subprojects, many of which are being used in production by a wide variety of [commercial](#) and [open source](#) projects as well as being widely used in [academic research](#). Code in the LLVM project is licensed under the [Apache 2.0 License with LLVM exceptions](#).

The primary sub-projects of LLVM are:

1. The **LLVM Core** libraries provide a modern source- and target-independent **optimizer**, along with **code generation support** for many popular CPUs (as well as some less common ones). These libraries are built around a **well specified** code representation known as the LLVM intermediate representation ("LLVM IR"). The LLVM Core libraries are **well documented**, and it is particularly easy to invent your own language (or port an existing compiler) to use LLVM as an **optimizer and code generator**.
2. **Clang** is an "LLVM native" C/C++/Objective-C compiler, which aims to deliver amazingly fast compilers, extremely useful **error and warning messages** and to provide a platform for building great source level tools. The **Clang Static Analyzer** and **Clang-tidy** are tools that automatically find bugs in your code, and are great examples of the sort of tools that can be built using the Clang frontend as a library to parse C/C++ code.
3. The **LLDB** project builds on libraries provided by LLVM and Clang to provide a great native debugger. It uses the Clang ASTs and expression parser, LLVM JIT, LLVM disassembler, etc so that it provides an experience that "just works". It is also blazing fast and much more memory efficient than GDB at loading symbols.
4. The **libc++** and **libc++ABI** projects provide a standard conformant and high-performance implementation of the C++ Standard Library, including full support for C++11 and C++14.
5. The **compiler-rt** project provides highly tuned implementations of the low-level code generator support routines like `__floatsisd` and other calls generated when a target doesn't have a short sequence of native instructions to implement a core IR operation. It also provides implementations of run-time libraries for dynamic testing tools such as **AddressSanitizer**, **ThreadSanitizer**, **MemorySanitizer**, and **DataFlowSanitizer**.
6. The **MLIR** subproject is a novel approach to building reusable and extensible compiler infrastructure. MLIR aims to address software fragmentation, improve compilation for heterogeneous hardware, significantly reduce the cost of building domain specific compilers, and aid in connecting existing compilers together.
7. The **OpenMP** subproject provides an **OpenMP** runtime for use with the OpenMP implementation in Clang.
8. The **polly** project implements a suite of cache-locality optimizations as well as auto-parallelism and vectorization using a polyhedral model.
9. The **libclc** project aims to implement the OpenCL standard library.
10. The **klee** project implements a "symbolic virtual machine" which uses a theorem prover to try to evaluate all dynamic paths through a program in an effort to find bugs and to prove properties of functions. A major feature of klee is that it can produce a testcase in the event that it detects a bug.
11. The **LLD** project is a new linker. That is a drop-in replacement for system linkers and runs much faster.
12. The **BOLT** project is a post-link optimizer. It achieves the improvements by optimizing application's code layout based on execution profile gathered by sampling profiler.

In addition to official subprojects of LLVM, there are a broad variety of other projects that [use components of LLVM for various tasks](#). Through these external projects you can use LLVM to compile Ruby, Python, Haskell, Rust, D, PHP, Pure, Lua, Julia, and a number of other languages. A major strength of LLVM is its versatility, flexibility, and reusability, which is why it is being used for such a wide variety of different tasks: everything from doing light-weight JIT compiles of embedded languages like Lua to compiling Fortran code for massive super computers.

As with everything else, LLVM has a broad and friendly community of people who are interested in building great low-level tools. If you are interested in [getting involved](#), a good first place is to skim the [LLVM Blog](#) and join [LLVM Discussion](#). For information on how to send in a patch, get commit access, and copyright and license topics, please see the [LLVM Developer Policy](#).

## Latest LLVM Release!

**3 Oct 2023:** LLVM 17.0.2 is now [available for download!](#) LLVM is publicly available under an [MIT license](#). If you would like to contribute to LLVM, please use the [LLVM GitHub](#). If you would like to check out the [new features](#) in Git that will appear in the next LLVM release, if you are using LLVM through anonymous Git.

## Upcoming Events

October 10-11, 2023 - LLVM Dev Mtg

## ACM Software System Award!

LLVM has been awarded the 2012 ACM Software System Award! This award is given by ACM every year. LLVM is a [highly distributed company](#). Click on any of the individual recipient's citation describing the award.

## Upcoming Releases

- LLVM Release Schedule:**
- 17.0.x
    - + Jul 28th: 17.0.0-rc1 was released
    - + Aug 8th: 17.0.0-rc2 was released
    - + Aug 22nd: 17.0.0-rc3 was released
    - + Sep 5th: 17.0.0-rc4 was released
    - + Sep 19th: 17.0.1 was released
    - + Oct 3rd: 17.0.2 was released
    - + Oct 17th: 17.0.3
    - + Oct 31st: 17.0.4
    - + Nov 14th: 17.0.5

## Developer Meetings

- Upcoming:
- [October 10-12, 2023](#)

## Proceedings from past meetings:

- [May 16-11, 2021](#)
- [November 8-9, 2022](#)
- [May 10-11, 2022](#)
- [November 16-15, 2021](#)
- [October 6-8, 2020](#)
- [October 22-23, 2019](#)
- [April 8-9, 2019](#)
- [October 2-18, 2018](#)
- [April 16-17, 2018](#)
- [October 18-19, 2017](#)
- [March 27-28, 2017](#)
- [November 1-4, 2016](#)
- [March 17-18, 2016](#)
- [October 29-30, 2015](#)
- [April 13-14, 2015](#)
- [October 28-29, 2014](#)
- [April 7-8, 2014](#)
- [Nov 6-7, 2013](#)

## Coverage Report (/Users/buildslave/jenkins/workspace/coverage/llvm-project/)

Created: 2023-10-07 09:23

Click [here](#) for information about interpreting this report.

Filename	Function Coverage	Line Coverage	Region Coverage	Branch Coverage
<a href="#">clang/</a>	89.25% (56326/63113)	88.28% (704033/797518)	76.05% (578466/760649)	76.29% (387845/508360)
<a href="#">lldb/</a>	70.43% (18308/25995)	59.53% (185200/311089)	58.65% (118152/201447)	47.19% (62080/131558)
<a href="#">llvm/</a>	86.83% (91053/104860)	86.41% (1214804/1405799)	81.07% (900153/1110362)	78.71% (629281/799514)
<b>Totals</b>	<b>85.42%</b> (165688/193972)	<b>83.68%</b> (2104050/2514436)	<b>77.05%</b> (1596772/2072464)	<b>74.97%</b> (1079206/1439436)

Generated by llvm-cov -- llvm version 18.0.0git

# Q&A

## Support for boolean counters to reduce runtime overhead

For coverage, we only need to know if region was executed

## Omit the unnecessary sections from binaries and profiles

Sections other than counters can be stripped for coverage

## Support string merging for `__llvm_prf_names`

This would enable deduplication resulting in 10x reduction

## Make `__llvm_prf_data` position independent

Avoid per-function dynamic relocation and allow sharing

## Avoid the use of indirection for runtime counter relocation

This requires assistance from linker and dynamic linker

There are of opportunities for further improvements that we would like to explore in the future.

If you're interested in collaborating on the ideas listed on this slide, please reach out.

