Debloating Address Sanitizer

Yuchen Zhang† Chengbin Pang†‡∗ Georgios Portokalidis† Nikos Triandopoulos† Jun Xu†‡
†Stevens Institute of Technology ‡Nanjing University

Abstract

Address Sanitizer (ASan) is a powerful memory error detector. It can detect various errors ranging from spatial issues like out-of-bound accesses to temporal issues like use-after-free. However, ASan has the major drawback of high runtime overhead. With every functionality enabled, ASan incurs an overhead of more than 1x.

This paper first presents a study to dissect the operations of ASan and inspects the primary sources of its runtime overhead. The study unveils (or confirms) that the high overhead is mainly caused by the extensive sanitizer checks on memory accesses. Inspired by the study, the paper proposes ASan--, a tool assembling a group of optimizations to reduce (or “debloat”) sanitizer checks and improve ASan’s efficiency. Unlike existing tools that remove sanitizer checks with harm to the capability, scalability, or usability of ASan, ASan-- fully maintains those decent properties of ASan.

Our evaluation shows that ASan-- presents high promise. It reduces the overhead of ASan by 41.7% on SPEC CPU2006 and by 35.7% on Chromium. If only considering the overhead incurred by sanitizer checks, the reduction rates increase to 51.6% on SPEC CPU2006 and 69.6% on Chromium. In the context of fuzzing, ASan-- increases the execution speed of AFL by over 40% and the branch coverage by 5%. Combined with orthogonal, fuzzing-tailored optimizations, ASan-- can speed up AFL by 60% and increase the branch coverage by 9%. Running in Chromium to support our daily work for four weeks, ASan-- did not present major usability issues or significant slowdown and it detected all the bugs we reproduced from previous reports.

1 Introduction

Programs developed in low-level languages, such as C and C++, often contain an abundance of memory error bugs. When exploited, the bugs can lead to severe security issues like data breach and hijacked execution. To help detect memory errors, various runtime tools [10, 14, 27, 36, 37, 42, 44, 47, 48, 50, 53] have been created. These tools vary in many aspects such as scope and capability. In recent years, address sanitizer (ASan) [42] stands out from the rest, thanks to its superior capability (detection of a wide spectrum of spatial errors and temporal errors), scalability (ability to support industry-grade programs like operating system kernels and web browsers), and usability (nearly zero configuration and seamless integration into mainstream compilers).

Technically, ASan allocates shadow memory to track the status (e.g., addressable or not) of the application memory. When the application issues an operation to a memory location, ASan checks the shadow byte mapped to the location and reports any embedded errors. The use of shadow memory and runtime checks leads to ASan’s two major drawbacks: high memory overhead and runtime overhead. The drawbacks lock many potentials of ASan, such as whole system sanitization.

Past research has extended many efforts towards addressing ASan’s drawbacks. In particular, hardware-assisted address sanitizer (HWASan) [25] utilizes Address Tagging [1], an AArch64 hardware feature, to reduce the use of shadow memory, decreasing the memory overhead of ASan from 1.5x-3x to 10%-35% [43]. However, the runtime overhead of ASan remains largely unsolved. Even HWASan still incurs similar overhead to ASan. Past research commonly blames sanitizer checks inserted by ASan for its high overhead. However, ASan has many runtime activities besides sanitizer checks, such as the preparation of metadata and logging. There has not been a systematic understanding of the overhead introduced by different activities. That means we are not even aware of how much overhead the sanitizer checks truly bring.

Study In this paper, we dissect ASan’s runtime activities into five groups: sanitizer checks, library function interception, poisoning and redzone (i.e., metadata management), logging, and heap management. Overhead brought by each group of activities to ASan is individually measured on SPEC CPU2006. Details of the study are presented in §3.1.

Our study confirms that sanitizer checks are the dominant source of overhead. It causes about 80.8% of the overhead (86.5% out of 107.8%). The study also unveils other sources of the overhead. In particular, the customized heap management of ASan incurs an overhead of 9.6%, which is less known before. More details can be found in Figure 3.

Literature Inspired by the study, this paper proposes to reduce ASan’s overhead by reducing sanitizer checks. There are many similar attempts [16,29,46,49,55] in the past. However, the techniques they proposed harm ASan’s descent properties (capability, scalability, usability). GWP-ASan [16] randomly picks a sub-set of heap objects to add guard pages and ignores the others. ASAP [49] removes sanitizer checks to satisfy a performance budget. Its criterion is to remove sanitizer checks on “hot” code that is more often executed, offering no safety assurance of the removing operations. Similar to ASAP, partiSan [29] follows a performance-driven metric to remove sanitizer checks without ensuring safety. GWP-ASan, ASAP, and partiSan all degrade the capability of ASan. Alternatively, SANRAZOR [55] combines static patterns and dynamic patterns to identify and remove redundant sanitizer (i.e., checks.
whose safety property is covered by others). SANRAZOR is more safety-driven but still cannot ensure the removed checks are indeed redundant because the patterns used are not sound. Also, SANRAZOR needs user inputs for profiling. In short, SANRAZOR harms both the capability and usability of ASan.

**Optimization** In this paper, we assemble a group of optimizations to reduce ASan checks. All the optimizations are purely static. They only involve sound, lightweight, and configuration-free analysis, which maintains the capability, scalability, and usability of ASan.

- **Optimization 1** identifies stack/global accesses that can be proven in-bound and removes their sanitizer checks. The optimization has been explored by many tools [13,18,19,38,39,46]. However, these tools often rely on heavy static analysis, which may not scale. We use an approach involving only control flow traverse and basic constant propagation to pinpoint the desired memory accesses (§4.1).
- **Optimization 2** finds sanitizer checks that are dominated (or post-dominated) and meanwhile covered by another check. Such checks are redundant and removed. Previous research [42] has briefly discussed this optimization but did not provide full solutions. We develop an algorithm that combines domination analysis and basic, flow-insensitive alias analysis to achieve this optimization (§4.2).
- **Optimization 3** determines memory accesses that are neighbors in space and merge their checks into one. We propose, design, and implement the optimization (§4.3).
- **Optimization 4** focuses on memory accesses in loops. It locates memory accesses with an invariant address and moves the sanitizer checks outside the loop. It also locates memory accesses that use a constantly increasing/decreasing address and groups the sanitizer checks across iterations into one. We propose, design, and implement the optimization (§4.4).

**Evaluation** We have integrated the optimizations into LLVM and created a tool called ASan--. We evaluate ASan-- from multiple dimensions. First, we run ASan-- on the full Juliet Test Suite [40] and 34 vulnerabilities from the Linux Flaw project [34]. ASan-- achieves an identical error detection rate to ASan. Second, we build SPEC CPU2006 and Chromium with ASan--. ASan-- imposes a moderate increase to the compilation time and reduces the binary size by 20%. Binaries built for the programs can support all the benchmark tests. The first and second evaluations show that ASan-- inherits the capability, scalability, and usability of ASan. Third, we compare the performance of ASan-- and ASan on both SPEC CPU2006 and Chromium. The evaluation shows that ASan-- can reduce the overhead of ASan by 41.7% on SPEC CPU2006 and by 35.7% on Chromium. If only considering the overhead incurred by sanitizer checks, the reduction rates increase to 51.6% on SPEC CPU2006 and 69.6% on Chromium. Fourth, we apply ASan-- in the application of fuzzing. ASan-- can increase the execution speed of AFL by over 40% and the branch coverage by 5%. Combined with fuzzing-tailored optimization techniques, ASan-- can speed up AFL by 60% and increase the branch coverage by 9%. Finally, we run Chromium built with ASan-- to support our daily work for four weeks. We did not experience major usability issues or significant slowdown, and ASan-- detected all the bugs we reproduced from previous reports.

**Contribution** We make the following main contributions.

- We present a study to dissect ASan’s runtime activities and unveil the sources of overhead. The study confirms existing understandings but also brings new findings.
- We design and implement a group of optimizations to reduce sanitizer checks of ASan. To our knowledge, two of the optimizations are brought up by us for the first time.
- We implement the optimizations and integrate them into LLVM. Source code is publicly available at https://github.com/junxzm/ASAN--.git.
- We conduct an extensive evaluation on our optimizations. The results show that our optimizations can significantly reduce the overhead of ASan without compromising its properties. The results also show that our optimizations can benefit the applications of ASan.

## 2 Technical Background

This section covers the technical background of ASan, based on the implementation in LLVM [42].

### 2.1 Shadow Memory

ASan uses a shadow memory model illustrated in Figure 1 to facilitate efficient runtime checks. It spares one-eighth of the virtual address space as the shadow memory where each byte records the status, addressable or not, of eight bytes used by the application. Given a memory byte at address Addr, ASan places its shadow byte at \((\text{Addr}+3)\times\text{Offset}\), where Offset is a constant determined at the compiling time. To ensure the region for shadow memory is always available, Offset needs to be chosen such that the memory area \([\text{Offset}, \text{Offset}+\text{Max_Addr}/8]\) is not occupied before the shadow memory is allocated.

The value of a shadow byte encodes the addressable status of the corresponding eight application bytes. 0 means all the eight bytes are addressable, while a value K ranging from 1 to 7 means only the first K bytes are addressable. A negative value indicates all eight bytes are not addressable, and different negative values unveil different types of non-addressable memory (out-of-bound heap, out-of-bound stack, out-of-bound global, freed memory, etc.).

### 2.2 Redzone

ASan places a redzone before and after each data object, as shown in Figure 2. The redzones are poisoned. Namely,
their shadow bytes mark them non-addressable. The object, in contrast, is unpoisoned. The creation and poisoning of redzones slightly differ when the object is in different regions.

**Heap** ASan replaces the standard heap management functions (e.g., malloc, calloc, free, etc.) with customized versions. Once a heap buffer is allocated, ASan places a redzone both before and after the buffer. Size of the redzone, a power of two, increases with the buffer size and can range from 16 to 2,048. By default, the end of the buffer is 8-byte aligned. If a gap of fewer than 8-bytes exists between the aligned end and the actual end, ASan poisons the gap via partial poisoning. For efficiency, ASan allocates multiple buffers of the same size as a group, positioning them next to each other. The left redzone of one buffer works as the right redzone of the previous buffer. This way ASan only needs to allocate one redzone (the left redzone) for buffers allocated from such a group. Once a buffer is freed, ASan poisons the entire buffer and places it into quarantine of a certain size, for the sake of detecting use-after-free.

**Stack** Before and after each array on the stack, ASan inserts two neighboring arrays to work as the left redzone and the right redzone, respectively. The left redzone has 32 bytes, and the right redzone has 32 bytes plus up to 31 bytes to align the original array. Both redzones are poisoned when the host function is entered at execution time.

**Global** ASan replaces each global object with a new one that contains a trailing redzone. The size of the redzone is 32 or one-fourth of the size of the object, whichever is larger. The size of the object plus the redzone is rounded up to a multiple of 32 bytes and partial poisoning is used if the end of the object is not 8-byte aligned. All the redzones for global objects are poisoned at the initialization of the process.

### 2.3 Runtime Checks

ASan instruments every memory access, load or store, with a check on its shadow memory. Depending on the size of the memory access, the check works differently. For an 8-byte memory access, the check loads its shadow byte and inspects whether the shadow byte is zero:

```c
if (ShadowAddr != 0)
    ReportAndCrash(Addr);
```

For a \(N\)-byte (\(N = 1, 2, \) or 4) memory access, the check examines whether its shadow byte indicates that the first \(N\) bytes are addressable:

```c
ShadowAddr = (Addr >> 3) + Offset;
V = *ShadowAddr;
if (V != 0 && ((Addr & 7) + N > V))
    ReportAndCrash(Addr);
```

Both checks only need one memory read and a few arithmetic/comparison instructions, presenting high efficiency. ASan also intercepts C library functions that often cause memory errors (e.g., memcpy). When such a function is entered, ASan checks whether the source buffer and the destination buffer are poisoned, using an optimized implementation.

### 3 Motivating Study

A major drawback of ASan is its high runtime overhead. As reported in [42], the implementation of ASan in LLVM incurs an average overhead of 73%. To gain insights into reducing the overhead of ASan, we run a study to understand the origins of the overhead.

#### 3.1 Study Methodology

Following the setup of [42], our study focuses on ASan implemented in LLVM and considers SPEC CPU2006 as the benchmark. SPEC CPU2006 [6] is an industry-standardized, CPU-intensive benchmark suite, consisting of 12 integer programs and 7 floating point programs. It comes with three workloads called train, test, and reference. We use the reference workload in our study as it represents a real dataset and activates long-duration execution.

The first step of our study measures the overhead of ASan with default settings. LLVM (or Clang) fails to compile omnetpp. Hence, we skip this program. Problems also occur on h264ref and perlbench because the two programs trigger ASan errors given the reference inputs. To continue their execution at the errors, we compile them with `-fsanitize-recover=address` and run them after setting the ASAN_OPTIONS environment variable to “halt_on_error=0”. The follow-up steps in turn make the
following modifications to ASan and separately measure the overhead after each.

- **Sanitizer Checks**: disabling the sanitizer checks on memory accesses (both loads and stores).
- **Interceptor**: disabling the interception of standard C libraries (memcpy, memset, and memmove, etc.).
- **Poisoning**: skipping the poisoning of redzones.
- **Logging**: removing intermediate logging, in particular the logging of the stack trace during malloc and free.
- **Heap Management**: revoking ASan’s heap management to the one from the standard C library.
- **Redzone**: removing the remaining redzones and the related operations.

### 3.2 Result Analysis

Figure 3 shows the average results of 30 repeated tests. Stock ASan brings an overhead of 107.8%. This is higher than what is reported in [42], but it is expected since the evaluation in [42] disabled all the logging and removed the quarantine for freed objects. Checks on individual memory accesses are the primary source of ASan overhead, which is attributed to an overhead of 86.5% (80.8% of all the overhead). This matches the common understanding [49, 55] since memory accesses are prevalent. A further breakdown unveils that the checks in sequential code take about 55% of the 80.8% overhead, and the checks in loops take the remaining. Each other activity of ASan also brings overhead, leading to a total overhead of 20.5%. A previously less-known finding is that ASan’s heap management averages an overhead of 9.6%. On the programs of perlbench, it incurs an overhead of over 39%. The remaining activities are less significant, mostly leading to an overhead of 2%-3%.

### 3.3 An Extension

Past research [24] has done a related study to understand the time spent by ASan. The study brings a unique finding that complements ours. They find that modern computers broadly use a 64-bit address space, where ASan uses a 4TB heap and needs a 16TB shadow memory, as shown in Figure 1. When the application’s memory is scattered in the address space, so will be the shadow memory in the 16TB space. This can result in sparse page table entries (PTEs) and increased page faults during page table management, bringing extra memory management costs. By restricting the application to run in a “smaller” address space, ASan can also use a smaller shadow memory. Both help compact the page table and reduce page faults, thus benefiting various ASan activities (e.g., redzone poisoning and shadow memory checks) that access the application memory and the shadow memory.

The original study in [24] focuses on smaller programs under the context of fuzzing. We replicate the study on SPEC CPU2006 in normal execution environments. As we will explain in §5.2, the public version of the tool presented in [24] fails on many CPU2006 programs due to its strict restriction of memory usage. Hence, we use an approximating approach. The idea is to only allocate one shadow memory page and map all the application memory to the same memory page. Technically, we replace (Addr \(\gg\) 3) + Shadowoffset with (Addr \(\gg\) 36) + Shadowoffset when mapping Addr to its shadow byte. We also set all the shadow bytes to zero when poisoning redzones, preserving the operations but disabling all the errors. This way, we simulate the execution with a 4KB shadow memory. Table 6 in the Appendix shows the results averaged on 30 runs. The results match the observation in [24]. Using a 4-KB shadow memory can reduce about 28.6% of...
the page faults, bringing the overhead of ASan from 107% down to 95.5%.

4 Debloating ASan

To reduce the overhead of ASan, one strategy is to shrink the shadow memory [24]. It has demonstrated effectiveness but still leaves behind high overhead (95.5% per our study). To further cut down ASan overhead, an orthogonal, promising strategy is to remove sanitizer checks [16, 46, 49, 55], since they constitute the primary source of overhead.

When removing ASan checks, a desirable principle is to maintain the decent characteristics of ASan, instead of sacrificing them for better efficiency. The principle can break down into the following properties.

- **Capability**: ASan offers the capability of detecting most cases of both spatial and temporal memory errors. Removal of checks should not reduce the detecting capability. Formally, the removal of a check should only proceed after it is proven safe to do so.

- **Scalability**: ASan can scale to support large-size programs like Chromium and Firefox browsers, enabling high practicality and broad adoption. Preferably, the removal of checks should not degrade the scalability of ASan.

- **Usability**: ASan has been integrated into mainstream compilers like LLVM and GCC. Enabling ASan simply requires a compiler flag (e.g., `-fsanitize=address`) on LLVM (disabled by default). The analysis focuses on removing sanitizer checks with LLVM (disabled by default). The analysis focuses on removing sanitizer checks with LLVM.

In this work, we assemble a set of techniques satisfying the three properties to reduce (or “debloat”) ASan checks, with a focus on the ASan in LLVM. Our contribution to different techniques varies, which is indicated by the number of ∆. Three ∆ mean that we are the first ones proposing, designing, and implementing the technique. In contrast, two ∆ indicate that others have informally mentioned the technique while we deepen, design, and implement the technique. Finally, a single ∆ suggests that the technique has been explored, and we implement the technique in LLVM or improve the existing implementation. Our efforts on the single ∆ techniques are less scientific but are deemed necessary as they will enable an understanding of the overall improvement by all the techniques. In the rest of this section, we elaborate on the designs of our techniques. Properties of the techniques are described and validated in Appendix A, due to the limited space.

4.1 Removing Unsatisfiable Checks ∆

The first technique focuses on removing ASan checks on global/stack access. The key insight is to remove checks where the error condition can never be satisfied. Below shows such an example on the stack (code in red is inserted by ASan).

```c
int foo(){
    char buf[20];
    unsigned int i = 10;
    i++; // i = 11
    ASan(buf+i);// ASan check 1; unsatisfiable
    buf[i] = 0;
}
```

The access to buf, an array on the stack, has a constant index of 11, which cannot overflow. Thus, no ASan check is needed there. Formally, given an array buf of n bytes on the stack or in the global, a K-byte access buf[OFFSET] needs no ASan check if on any execution path:

**Condition-1**: OFFSET ≥ 0 & & N − OFFSET ≥ K

Past research has discussed similar issues several times and has explored various solutions [13, 18, 19, 39, 46] to identify stack/global accesses that meet the above condition. A systematic approach [52] is to symbolically interpret the execution paths reaching a target access and verify whether every path satisfies the condition. However, the approach may not scale as it involves complex constraint solving.

**Our Choice** We opt to adopt a lightweight analysis shipped with LLVM (disabled by default). The analysis focuses on direct stack/global accesses whose target object is propagated from an `alloca` instruction or a global object. Given a direct stack/global access, the analysis backward traces whether the index propagates from constants. Consider buf[i] at Line 6 in the above code as an example. It traces i backward from Line 6 to Line 4 where it represents i@L6 as i@L4 + 1, followed by another step jumping to Line 3 to further represent i@L4 + 1 as 10 + 1. This way, it learns the constant value, 11, of i at Line 6 and can compare i with the array size. The backward tracing stops wherever uncertainty appears (e.g., an intermediate value is loaded from memory or a store happens to an intermediate value) and returns an unknown result, thus offering soundness.

We slightly extend the above analysis. When backward tracing the index of an access, we concurrently collect comparisons that dominate the access and operate between the index and a constant. Doing so enables us to accumulate loose boundaries of the index and compare the index with the object size even the index is not a constant. Our extension helps identify removable accesses like the one below. Our extension inherits the soundness from the basic version. Further, whenever there is a store on the index, we abandon the comparisons affected by the store.

```c
int foo(){
    char buf[20];
    unsigned int i = input();// i is not a constant
    if(i < = 10) // the comparison will be captured
        ASan(buf+i);// ASan check 1; still unsatisfiable
    buf[i] = 0;
    }
```
4.2 Removing Recurring Checks

The second technique aims to remove recurring checks or checks guaranteed to have been done. The code below is an example of recurring checks.

```c
int *p;
ASan(p); // ASan check 1

int *p = 0;
ASan(p); // ASan check 2 (a recurring check)
*p = 1;
```

In the example, ASan places two checks on the dereferences at Line 1 and Line 3. However, the two dereferences access the same location with the same size, meaning that the check at Line 4 is a replica of that at Line 2 and thus, can be removed. Past research [42] has demoed recurring checks but did not provide full solutions. We first formalize recurring checks. Given two ASan checks, denoted as ASan(ptr1) and ASan(ptr2), on memory accesses *ptr1 and *ptr2. ASan(ptr2) is a recurring check of ASan(ptr1) if:

- **Condition-1**: ptr1 == ptr2 || alias(ptr1, ptr2) && sizeof(*ptr1) >= sizeof(*ptr2) && dominate(*ptr1, *ptr2) || post-dominate(*ptr1, *ptr2));

Following the above formalization, Algorithm 1 is designed to identify recurring ASan checks in individual functions. The first step aggregates memory accesses to the same location into one group (Line 12–19). When determining whether two memory accesses have the same location, aliases are considered but in a sound way. Specifically, we reuse the basic, flow-insensitive alias analysis in LLVM and only accept the must-alias results. The internal algorithm of LLVM ensures that two memory objects with must-alias results always start at exactly the same location [3], protecting the safety of our optimization. The second step processes each memory group separately. Given a group, the algorithm inspects each pair of memory accesses therein and identifies recurring ASan checks according to the above conditions (Line 20–37). In this step, we reuse the analysis offered by LLVM to understand the domination/post-domination relation [4].

**Discussion**: The soundness of Algorithm 1 can fail in certain cases. For instance, *ptr1 dominates *ptr2 and satisfies the above condition. Thus, the ASan check on ptr2 will be removed. However, if the memory is freed between *ptr1 and *ptr2, *ptr2 will be use-after-free, and we will miss it. We use a conservative approach to avoid such issues: we skip the removing if any function call exists on any execution path from *ptr1 to *ptr2. We do not do this if *ptr1 post-dominates *ptr2 as *ptr1 shall always execute after *ptr2.

4.3 Optimizing Neighbor Checks

This technique explores optimizing ASan checks on neighboring memory accesses. Code below illustrates the ideas.

```c
struct{
  int a; int b; int c;
}test;
int foo(){
  struct test *ptr1, *ptr2;
  ASan(ptr1->a); // ASan check 1
```
ASan(ptr2->c); //ASan check 5
ASan(ptr2->a); //ASan check 3

//slow path
if (second_chk(ptr1)
Shadow = (ptr1
short *Shadow //2 bytes
Shadov = (ptr1->a >> 3) + Offset;
if (*Shadow != 0){/Load and check shadow bytes for both ptr1->a and ptr1->b in a single operation
//slow path
second_chk(ptr1->a);
//check whether ptr1->a is addressable
second_chk(ptr1->b);
//check whether ptr1->b is addressable
}

Since the majority of the accesses should be error-free, the check will likely exit at line 3 instead of routing to the slow path. Therefore, the efficiency of the check should approximate a single ASan check. We call the checks on ptr1->a and ptr1->b mergeable neighbor checks.

Second, ptr2->b locates in the gap between ptr2->a and ptr2->c, and the gap is 4 bytes in size. If ptr2->b is in a redzone, then at least one of ptr2->a and ptr2->c must also be in the redzone (maybe partially) since a redzone has at least 16 bytes. As such, any error on ptr2->b will always be captured by the checks on ptr2->a (Line 14) and ptr2->c (Line 18). Thus, the check on ptr2->b (Line 16) can be removed. We call such cases removable neighbor checks.

Merging Neighbor Checks Formally, given two memory accesses *ptr1 and *ptr2, ASAN(ptr2) can be merged to ASAN(ptr1) if

Given:
base1 = base(ptr1); base2 = base(ptr2);
/*base(p) gets the base address of the object that p points to*/
offset1 = offset(ptr1); offset2 = offset(ptr2);
/*offset(p) gets the offset of p in the object*/
offset2 > offset1;
MaxBitRead = the maximal number of bits a single memory-read can load;

Then Condition-3:
base1 == base2 &&
RoundUpTo(offset2 - offset1 + sizeof(*ptr2), 8) <= MaxBitRead &&
(dominate(*ptr1,*ptr2) &&
post-dominate(*ptr2,*ptr1)) ||
(dominate(*ptr2, *ptr1) &&
post-dominate(*ptr1, *ptr2))

At the high level, the condition ensures that (i) ASan(ptr1) and ASan(ptr2) either both happen or both not happen and (ii) the shadow bytes for the region between the first byte of *ptr1 and the last byte of *ptr2 can be loaded in one memory-read. These ensure that ASan(ptr1) and ASan(ptr2) can be merged to the above form, after adjusting the type of Shadow based on the number of shadow bytes.

We use an intra-procedural analysis to locate mergeable neighbor checks. Given function F, memory accesses therein are grouped based on their base addresses and sorted by their offsets. The challenge here is to identify the base address and offset of a given access. In LLVM IR, a memory access is either a load or a store, and the address can be extracted. We backward trace the propagation of the address until a getelementptr instruction [5], where the base address and the offset can be separately obtained. In the backward tracing, we only allow bitcast instructions because other instructions may change the address. A getelementptr instruction often includes multiple-dimensional offsets, all following the format of constants or variables. If all the offsets are constants, we flatten them into a single one based on the scale of each dimension. Otherwise, we mark the offset unknown.

Given a group of memory accesses with the same base address (denoted \( M = \{m_1,m_2,...,m_n\} \)), all the mergeable pairs are gathered and then reorganized as \( \tilde{m}_i \rightarrow \{m_{i1},m_{i2},...,m_{in}\} \) \((1 \leq i \leq n)\), meaning that the checks on \( m_{i1},m_{i2},...,m_{in} \) can be merged to the check on \( \tilde{m}_i \). Algorithm 2, a greedy algorithm, is then used to pick pairs to merge. The algorithm, by itself, is straightforward to understand, but its properties are worth mentioning. First, the algorithm prioritizes the handling of \( \tilde{m}_i \) with a larger size, potentially helping identify more mergeable pairs. Second, the algorithm supports recursive merging. For instance, after \( m_i \) is merged to \( m_{i1}, m_{i1} \) can be later merged to \( m_k \), if \( m_k \) can also be merged to \( m_k \).

For a pair of mergeable memory accesses \( (m_1,m_2) \), where ASan(m_2) can be merged to ASan(m_1), a new check will be created and inserted under the post-dominating access. Both the original checks will be removed. All checks that are previously merged to m_2 will be migrated to the new check.

Removing Neighbor Checks Removable neighbor checks can be formalized as follows. Given three memory accesses *ptr1, *ptr2, and *ptr3, ASAN(ptr2) can be removed if

Given:
base1 = base(ptr1); base2 = base(ptr2);
base3 = base(ptr3);
off1 = offset(ptr1); off2 = offset(ptr2);
off3 = offset(ptr3);
Algorithm 2: Picking mergeable neighbor checks.

Input: A list of mergeable relations:
\[ M = \{ \vec{m}_1, \vec{m}_2, \ldots, \vec{m}_n \} \]
\[ /\ast \vec{m}_i = \{m_{i1}, m_{i2}, \ldots, m_{in} \} \] represents the list of memory accesses that can be merged to \( \vec{m}_i \)\n
Output: A list of mergeable memory access pairs
\[ \vec{P} = \{ (m_{11}, m_{12}), (m_{21}, m_{22}), \ldots, (m_{nk}, m_{jk}) \} \]

Initialization:
1. Initialize: \( \vec{P} = \emptyset \);
2. \( \vec{M} = \text{sorted}(M) \);
3. \( /\ast \text{sort by } |\vec{m}| \text{ in descending order} \)
4. for each list \( \vec{m}_i \) in \( \vec{M} \), do
5. for each member \( m \) in \( \vec{m}_i \), do
6. if \( M_r, \text{merged}_\text{others}(m) \) then
7. /* \( m \) has merged to another memory access */
8. Continue;
9. /* avoid repeated merging */
10. end
11. \( \text{tmp} = \vec{M}_r, \text{find}_\text{merged_by}(m); \)
12. /* find accesses that are merged to \( m \) */
13. if \( \text{tmp} \not\in \vec{m}_i \) then
14. /* not all accesses merged to \( m \) can be merged to \( m_i \), skip */
15. end
16. \( \vec{P}.\text{add}(m, m_i); \)
17. /* merge \( m \) to \( m_i \) */
18. end
19. return \( \vec{P} \);

off1 < off2 < off3;
size1 = sizeof(*ptr1); size2 = sizeof(*ptr2);
size3 = sizeof(*ptr3);
MinRdSz = the minimal size of a redzone;

Then Condition-4:
base1 == base2 == base3 \&\& off3 - off1 < MinRdSz
\&\& off2 + size2 \leq off3 + size3 \&\&
(dominate(*ptr1, *ptr2) ||
post-dominate(*ptr1, *ptr3)) \&\&
(dominate(*ptr3, *ptr2) ||
post-dominate(*ptr3, *ptr2))

In essence, the above condition guarantees two properties:
First, ptr2 being in a redzone implies ptr1 and/or ptr2 are in the same redzone, fully or partially. Second, ASAN(ptr2) being reached means both ASAN(ptr1) and ASAN(ptr3) are reached. As such, if the condition holds, any errors on *ptr2 will be captured by ASAN(ptr1) or ASAN(ptr3). Thus, it is safe to remove ASAN(ptr2).

We also use an intra-procedural analysis to collect removable neighbor checks. Memory accesses in each function are first grouped based on their base addresses. From each group, tuples satisfying Condition-4 are then gathered. The tuples follow the format of \([m_i, m_j, m_k]\), meaning that ASAN\((m_k)\) can be removed because it is covered by ASAN\((m_i)\) and ASAN\((m_j)\). In the follow-up step, the analysis aggregates the tuples and produces a removable list for each \((m_i, m_j)\):
\[ \{m_{ij1}, m_{ij2}, \ldots, m_{i,j_{k}}\} \]. The final step is to go over each list and collect removable memory accesses. Simply including all the members in each list will cause mistakes. Consider the example where \((m_e, m_h) = \{m_{e}, m_{h}\}\) and \((m_e, m_j) = \{m_{e}\}\). After handling \((m_e, m_h)\), we remove \(m_{e}\). If we continue with \((m_e, m_d)\) to remove \(m_{e}\), mistakes will arise because \(m_{e}\) has been removed and cannot help cover \(m_{e}\). To avoid such mistakes, after including \(m_i\) to the removable set, we abandon all the lists matching \((m_i, \ast)\) and \((\ast, m_i)\). To maximize the number of removable cases, we sort all the lists based on their size in descending order and then handle each of them in turn.

4.4 Optimizing Checks in Loops

ASan checks in loops are expensive. According to our study in §3.2, checks in loops account for 45% of the overhead introduced by all ASan checks. Modern compilers run many optimizations to reduce operations in loops, which, however, still cannot handle many optimizeable ASan checks. The code below shows two types of optimizeable ASan checks in loops.

```c
char *ptr1;
char *ptr2;

for(int index = 0; index < limit; index += 2){
    ASan(ptr1+index);
    /*ASan check 1; can be merged*/
    ptr1[index] = getchar();
    ASan(ptr2);
    /*ASan check 2; can be moved out of loop*/
    *ptr2 = getchar();
}
```

First, ptr2 is dereferenced (Line 9) and ASan-checked (Line 7) in each loop iteration. However, ptr2 never changes. Thus, one ASan check on ptr2 out of the loop is sufficient. The compiler cannot optimize the memory access because it is a write. We call checks like ASan(ptr2) invariant checks.

Second, ptr1 is accessed (Line 6) with an offset increased by 2 in each iteration (Line 3). Thus, any two consecutive accesses to ptr1 are 2-byte apart, and the checks on them resemble the mergeable neighbor checks we discussed in §4.3 (Condition-3). Merging the checks together will reduce the cost. Considering that the address of the access (e.g., index at Line 6) monotonically increases/decreases, we call the check on it (e.g., Line 6) a monotonic check.

Re locating Invariant Checks

Formally, given a memory access *ptr in a loop, ASan(ptr) is an invariant check if
Condition-5: ptr is a loop invariant

To identify invariant checks, we gather all the loops from a function then process each loop in turn. Given a loop, each memory access inside is visited to inspect whether the address
is a loop invariant. If so, we mark the ASan check as an invariant check. LLVM offers a built-in interface to determine loop invariant. It is, however, very preliminary. Only when the address is a value created outside the loop, the interface considers it a loop variant. This misses invariant addresses (e.g., an address composed of two other invariant values). We design a new algorithm, Algorithm 3, to more systematically identify loop invariant. Due to the space limit, Algorithm 3 is presented in the Appendix. The algorithm backward traces the generation of an address until outside of the loop. If all the ingredients involved in generating the address are invariants, the algorithm reports the address also as an invariant.

To optimize an invariant check, an intuitive idea is to move the check after the exit of the loop. This works when the invariant check dominates the loop exit. However, many invariant checks are guarded by conditional statements inside a loop. Their execution is not guaranteed when the loop is entered. Consider the code below as an example.

```c
bool asan_ptr = false;
for(condition1){ //outer loop
  for(condition2){ //inner loop
    if(condition3){ //guardian condition
      asan_ptr = true;
      //original ASan(ptr) is here
      *
    }
  }
  if (asan_ptr){
    asan_ptr = false; //if not outermost loop
    ASan(ptr);
  }
}
//simply putting ASan(ptr) here is erroneous
if (asan_ptr){
  asan_ptr = false; //if not outermost loop
  ASan(ptr);
}
```

At Line 7, `ptr` is a loop invariant and is dereferenced. An ASan check on it originally exists at Line 6. The dereference and the ASan-check are guarded by condition3 at Line 4. Moving the ASan check to the exit of the loop (Line 10) can cause mistakes. The dereference and the original ASan check may not always execute when the loop (Line 3–9) is entered, while the new check at Line 10 always does. That means we may introduce checks that should not happen.

To optimize a conditional invariant check, we introduce a supportive local variable to track whether the memory access is ever executed. In the code above, `asan_ptr` is introduced to trace the execution of Line 8. `asan_ptr` is initialized as `false` and is set to `true` at the memory access. At the loop exit, `asan_ptr` being `true` will trigger the relocated ASan check. Otherwise, the ASan check is skipped. Further, if the memory access is not in an outermost loop, we reset the local variable to `false` at the loop exit (Line 12–15 in the above code) because, otherwise, the `true` status will carry over when the loop is re-entered in the next iteration of the outer loop. **Discussion:** The optimization of invariant checks can run into another problem. The code from an invariant check to the loop exit can free the memory. A check at the loop exit may, thus, report false use-after-free. Our backward analysis of invariant already considers the problem. If any value involved in the generation of the address is passed to a function, the analysis marks the address as a variant (see Algorithm 3).

**Grouping Monotonic Checks**
Given a memory access `*ptr` in a loop, ASan(ptr) is a monotonic check if

**Condition-6:** ptr in any two consecutive iterations has a constant distance.

To identify monotonic checks, a systematic approach is to use the scalar evolution (SCEV) [7], which is an analysis to represent variables with complicated behavior in a more straightforward way. In particular, SCEV can represent monotonically increasing/decreasing loop variables as *add recurrences* [11]. An add recurrence has the format of \{Init, +/-, Step\} (Step is a constant), meaning that the variable has an initial value of Init and increases by Step in an iteration.

Consider `ptr1[index]` at Line 8 in the code presented at the beginning of this subsection as an example. The original ASan check, executed in each loop iteration, is:

```c
V = * ((ptr1 + index >> 3) + Offset);
if (V != 0 && ((ptr1 + index & 7) + 1 > V))
  ReportAndCrash(Addr);
```

In our optimization, we first run SCEV to represent the address as \{Init, +/-, Step\} (i.e., \{ptr1, +, 2\}) and then confirm that Init (i.e., ptr1) is an invariant. Finally we replace the ASan check in the loop with a new one of the following format (the address is represented as Addr):

```c
intN *Shadow; //hold shadow memory for MinRdSz
application bytes (MinRdSz represents the minimal
size of a redzone); N >= MinRdSz
2 if((Addr - Init) % MinRdSz < Step){
3     Shadow = (Addr >> 3) + Offset;
4     if (*Shadow != 0) //check the shadow memory for
5         MinRdSz bytes
6         ASan(Addr);
```

At the loop exit, we further insert another piece of code. The code inspects whether the final Addr differs from Init. If so, it performs an ASan check on Addr.

The idea of our optimization is to check the shadow memory of a chunk of MinRdSz bytes at once, using only one operation. If the shadow memory is all zero, we skip the checks on follow-up access to the same chunk. Otherwise a regular ASan check is performed. This way we group multiple checks into one. The number of checks grouped depends on Step (MinRdSz/Step checks will be grouped into one). By default, the maximal Step we allow is MinRdSz/4.

**Discussion:** A memory access may cross the boundary of a MinRdSz chunk, but its beginning address does not meet the condition of Line 6. Thus, we will miss checking the second half of the access. This is OK as the second half will be checked in the next chunk thanks to the alignment in the mapping to shadow memory.
Table 1: Detection capability of ASan and ASan-- on the Juliet Test Suite. Good tests measure false positives. Bad tests measure false negatives. ASan and ASan-- achieve identical results shown in this table.

<table>
<thead>
<tr>
<th>CWD (number)</th>
<th>Good Test (Pass/Total)</th>
<th>Bad Test (Pass/Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stack-based Buffer Overflow (121)</td>
<td>2348/2348</td>
<td>1111/2348</td>
</tr>
<tr>
<td>Heap-based Buffer Overflow (122)</td>
<td>1677/1677</td>
<td>1595/1677</td>
</tr>
<tr>
<td>Buffer Under-write (124)</td>
<td>584/584</td>
<td>571/584</td>
</tr>
<tr>
<td>Buffer Over-read (126)</td>
<td>445/445</td>
<td>420/445</td>
</tr>
<tr>
<td>Buffer Under-read (127)</td>
<td>590/590</td>
<td>565/590</td>
</tr>
<tr>
<td>Total</td>
<td>5644/5644</td>
<td>5262/5644</td>
</tr>
</tbody>
</table>

5 Implementation and Evaluation

We have implemented the optimization techniques in a tool called ASan--. ASan-- is built on top LLVM-4.0, with around 2.5K lines of C++ added. We have ported a preliminary version of ASan-- to LLVM-12.0. ASan-- on both LLVM-4.0 and LLVM-12.0 is publicly available at https://github.com/junxzm1990/ASAN--.git. The use of ASan-- is identical to ASan. The optimizations are enabled by default but can be disabled through a customized environment variable. The rest of this section presents our evaluation of ASan--, centering around three questions:

- Can ASan-- maintain the capability, scalability, and usability of ASan?
- Can ASan-- reduce the runtime overhead of ASan?
- Can ASan-- benefit the applications of ASan?

5.1 Capability, Scalability, and Usability

Capability To measure the detection capability of ASan--, we run two experiments. In the first experiment, we run both ASan-- and ASan on the memory errors from the Juliet Test Suite (version 1.3) [40]. As shown in Table 1, ASan-- and ASan achieve identical results on both good and bad tests. For comparison, we further run SANRAZOR and ASAP on the Juliet Test Suite. SANRAZOR failed to run because of an error in the profiling component. The error has been confirmed by the developers and is pending fix. ASAP runs with LLVM-3.7 which does not support continuous execution at ASan errors. Thus, ASAP could not finish the test as Juliet triggers all the bugs in a single execution. To address this issue, we extend LLVM-3.7 to avoid halting at ASan errors. When running ASAP, we use the good tests of Juliet for profiling and the bad tests for evaluation. Further, we configure ASAP to use ASan--’s overhead as the performance budget (i.e., setting the cost level to 51.6%). At the end, we successfully run ASAP on 1,524 Juliet bugs. ASAP detects 985 of them, missing about 35% of the bugs. This shows that ASan-- detects more bugs than ASAP when presenting similar performance.

In the second experiment, we collect vulnerabilities from the Linux Flaw Project [34]. After visiting all the vulnerabilities in the database, we reproduce 34 of them, including 23 used in the evaluation of SANRAZOR [55]. Details of the vulnerabilities are presented in Table 2. Both ASan-- and ASan can detect all the 34 vulnerabilities. In contrast, SANRAZOR, removing ASan checks based on static/dynamic patterns, misses two vulnerabilities (CVE-2017-12858 and CVE-2017-14406). The experiment demonstrates the soundness of ASan-- and also its advantages over SANRAZOR.

Scalability ASan-- introduces extra analyses to ASan. Although the analyses are not complex, they may still affect the scalability of ASan. To this end, we apply ASan-- to SPEC CPU2006 and Chromium (58.0.3003.0). Chromium and many programs in CPU2006 (e.g., perl and gcc) should be large enough to stress test the scalability. ASan-- can successfully compile and build CPU2006 and Chromium. The binaries can pass all the benchmark tests (see §5.2). We have also been using the Chromium built with ASan-- for four weeks and have not encountered major issues (see §5.3).

We also measure the time cost of compilation. We believe it is also scalability-related since extremely high compilation time will make the tool impractical for large programs. Table 7 shows the results. On average, ASan-- increases the compilation time by 1x+, but the overall compilation time should still be acceptable. Precisely, ASan-- can finish the compilation of most programs in seconds or minutes. Even in the case of Chromium, ASan-- only needs several hours. The length of the time is tolerable in practice.

Usability ASan-- is implemented to share the same assumption, same requirement, and same interfaces with ASan. Usage-wise, it is identical to ASan. ASan-- also helps reduce the binary size because it removes many checks. On average, ASan-- can shrink the binary size by 20.4% (see Table 7 in Appendix), which helps improve the usability of ASan, in particular in contexts where the storage space is limited.
Table 3 presents the comparison between ASan-- and SANRAZOR. The Average data in this table differs from Figure 3 and Figure 4 because only 11 programs from SPEC CPU2006 are included.

Table 3: Performance comparison with SANRAZOR. The Average data in this table differs from Figure 3 and Figure 4 because only 11 programs from SPEC CPU2006 are included.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>ASan</th>
<th>UNSAT_Chk</th>
<th>Rec_Chk</th>
<th>Merge_Chk</th>
<th>Rem_Neighbor</th>
<th>Loop_Inv</th>
<th>Loop_Grp</th>
<th>No_Sanitizer_Chk</th>
<th>ASan--</th>
</tr>
</thead>
<tbody>
<tr>
<td>401 bzip2</td>
<td>81.90%</td>
<td>58.55%</td>
<td>45.06%</td>
<td>50.76%</td>
<td>49.39%</td>
<td></td>
<td></td>
<td></td>
<td>40.81%</td>
</tr>
<tr>
<td>429 mcf</td>
<td>54.10%</td>
<td>35.11%</td>
<td>9.03%</td>
<td>2.67%</td>
<td>26.80%</td>
<td></td>
<td></td>
<td></td>
<td>52.89%</td>
</tr>
<tr>
<td>445 gobmk</td>
<td>76.90%</td>
<td>62.81%</td>
<td>47.57%</td>
<td>30.65%</td>
<td>44.83%</td>
<td></td>
<td></td>
<td></td>
<td>41.92%</td>
</tr>
<tr>
<td>456 hammer</td>
<td>17.90%</td>
<td>133.8%</td>
<td>31.38%</td>
<td>39.52%</td>
<td>103.3%</td>
<td></td>
<td></td>
<td></td>
<td>49.72%</td>
</tr>
<tr>
<td>458 sphing</td>
<td>202.9%</td>
<td>108.3%</td>
<td>95.71%</td>
<td>97.74%</td>
<td>52.53%</td>
<td></td>
<td></td>
<td></td>
<td>62.81%</td>
</tr>
<tr>
<td>462 libquantum</td>
<td>108.0%</td>
<td>48.31%</td>
<td>53.51%</td>
<td>58.96%</td>
<td>52.05%</td>
<td></td>
<td></td>
<td></td>
<td>68.21%</td>
</tr>
<tr>
<td>433 milc</td>
<td>89.10%</td>
<td>44.97%</td>
<td>30.19%</td>
<td>43.25%</td>
<td>49.72%</td>
<td></td>
<td></td>
<td></td>
<td>54.94%</td>
</tr>
<tr>
<td>444 amdmd</td>
<td>86.10%</td>
<td>41.92%</td>
<td>34.19%</td>
<td>26.00%</td>
<td>39.22%</td>
<td></td>
<td></td>
<td></td>
<td>39.52%</td>
</tr>
<tr>
<td>453 povray</td>
<td>101.0%</td>
<td>64.81%</td>
<td>54.94%</td>
<td>50.62%</td>
<td>43.95%</td>
<td></td>
<td></td>
<td></td>
<td>53.51%</td>
</tr>
<tr>
<td>470 hm</td>
<td>62.40%</td>
<td>31.23%</td>
<td>6.62%</td>
<td>18.61%</td>
<td>33.08%</td>
<td></td>
<td></td>
<td></td>
<td>54.94%</td>
</tr>
<tr>
<td>482 sphinx3</td>
<td>103.0%</td>
<td>80.57%</td>
<td>43.59%</td>
<td>50.19%</td>
<td>68.21%</td>
<td></td>
<td></td>
<td></td>
<td>54.94%</td>
</tr>
<tr>
<td>Average</td>
<td>94.88%</td>
<td>64.31%</td>
<td>40.81%</td>
<td>42.08%</td>
<td>52.89%</td>
<td></td>
<td></td>
<td></td>
<td>54.94%</td>
</tr>
</tbody>
</table>

5.2 Runtime Overhead

SPEC CPU2006 We run both ASan-- and ASan on SPEC CPU2006, using the reference workload as testing data. In the experiment, both SANRAZOR [54] and FuZZan [24] are included as baselines. Following the paper presenting SANRAZOR [55], we perform the profiling for SANRAZOR using the training workload and measure its performance using the reference workload. Currently, the public version of SANRAZOR only supports 11 programs of CPU2006, as listed in Table 3. FuZZan offers different heap sizes (1G, 4G, 8G, and 16G). For each program, we start with the minimal heap and increase it until the test can succeed. In total, FuZZan can run 10 programs of CPU2006, as shown in Table 9. All the tests are repeated 30 times, and average results are collected.

Figure 4 compares the overhead of ASan and ASan--. On average, ASan-- reduces the overhead of ASan from 107.8% to 63.3%, producing a reduction rate of 41.7%. Only considering the overhead of ASan checks, the reduction rate is 51.6%. Figure 4 also unveils that every of our optimizations contributes. The amount of overhead reduced by most optimizations is over 5%. In particular, removing recurring checks reduces an overhead of 16.7%. We further measure the number of ASan checks reduced by ASan--. The results, shown in Table 8, are consistent with ASan--’s impacts on the overhead. Averagely, ASan-- removes 38.7% of the ASan checks and every optimization brings a meaningful contribution.

Table 3 presents the comparison between ASan-- and SANRAZOR. Both ASan-- and SANRAZOR can bring the overhead down. When SANRAZOR runs in L1 and L2 modes, it more aggressively removes ASan checks. Thus, it incurs less overhead (40.81% and 42.63%) than ASan-- (52.89%). However, the two modes can miss detecting memory errors, such as the two reported in Table 2. In contrast, when SANRAZOR runs in the more conservative L0 mode, it removes fewer checks and brings an overhead higher than ASan-- (64.31% v.s. 52.89%), although it still cannot offer the soundness of ASan--.

Table 9 in the Appendix shows the improvement of FuZZan to ASan. Regardless of which program and what heap size we test, FuZZan does not bring evident improvement to ASan. In summary, ASan-- and FuZZan provide similar bug detection but ASan-- is more scalable and runs faster.
An Extension: Considering that SPEC CPU2006 was years old, we repeated the evaluation of ASan-- on the new version of CPU2017 [2]. CPU2017 includes two modes, SPECrate and SPECspeed, to separately measure time-based performance and throughput performance. Due to issues of compiler compatibility, we could only run 10 SPECrate programs and 5 SPECspeed programs SPECspeed. The results, summarized in Table 11 in the Appendix, show that ASan-- similarly benefits CPU2017. The reduction rates of overhead are 40.6% and 39.1% respectively on SPECrate and SPECspeed.

Chromium In this experiment, we measure the performance of Chromium under ASan and ASan--. Six popular web-browser benchmarks are used, including both Time-based ones (Sunspider [9], Kraken [33], Lite Brite [32]) and Score-based ones (Octane [21], Basemark [12], WebXPRT [41]). We repeat the experiment 30 times and report the average results in Figure 5. Other tools were omitted as they could not run Chromium. On average, ASan-- reduces the overhead of ASan from 117.8% to 75.8%, presenting a reduction rate of 35.7%. If only counting the overhead of ASan checks, the reduction rate is 69.6%. Considering the scale of Chromium, such reduction rates are meaningful in general.

5.3 Applications

Fuzzing One of the most popular applications of ASan is to help bug detection in fuzzing. Our evaluation of this application focuses on how much ASan-- can improve the efficiency of ASan-enhanced fuzzing. We also include FuZZan (https://github.com/HexHive/FuZZan) as a baseline. To properly compare with FuZZan, we consider AFL-2.52b as the fuzzing tool and reuse the configurations presented in [24] (see Table 10 in the Appendix). For consistency, we run all the tests on Amazon EC2 instances with Intel XeonE5 Broadwell 16 cores, 64GB RAM, and Ubuntu 18.04 LTS. To avoid interference, different tests are run separately for 24 hours. All the tests are repeated 10 times with average results gathered. We also perform the evaluation twice, separately using seeds shipped with AFL and seeds shipped with FuZZan.

Table 4 shows the evaluation results. On average, ASan-- increases the execution speed of AFL by 42.1% and 39.8%, on AFL’s seeds and FuZZan’s seeds, respectively. The increase of execution speed leads to a growth rate of 5.0% and 5.1% in branch coverage. In comparison to FuZZan, ASan-- presents

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SPECrate</td>
<td>35.7%</td>
<td>54.2%</td>
<td>39.1%</td>
<td>55.1%</td>
<td>35.8%</td>
<td>55.1%</td>
<td>35.8%</td>
<td>55.1%</td>
<td>35.8%</td>
<td>55.1%</td>
<td>35.8%</td>
<td>55.1%</td>
</tr>
<tr>
<td>SPECspeed</td>
<td>35.7%</td>
<td>54.2%</td>
<td>39.1%</td>
<td>55.1%</td>
<td>35.8%</td>
<td>55.1%</td>
<td>35.8%</td>
<td>55.1%</td>
<td>35.8%</td>
<td>55.1%</td>
<td>35.8%</td>
<td>55.1%</td>
</tr>
</tbody>
</table>

Figure 6: Increase of branch coverage brought by ASan-- with ASan as the baseline. 1/2/4 instance represents fuzzing with one sanitized instance and 0/1/3 non-sanitized instances. A comparable, but slightly smaller, improvement. Execution speed wise, FuZZan brings an increase of 50.2% and 45.7%, given AFL’s seeds and FuZZan’s seeds. Branch coverage wise, FuZZan brings a growth of 6.24% and 6.58%, depending on the seeds. The results are understandable, because FuZZan is tailored for fuzzing while ASan-- is more general.

As we pointed out in §3, ASan-- and FuZZan use orthogonal techniques. By intuition, combining techniques of both tools can produce better improvement. We, thus, migrate FuZZan’s approach of reducing memory into ASan--. We limit the heap size to 16GB and relocate the heap plus stack [26] to below the shadow memory. This way, we shrink the address space to 19GB (16GB heap + 1GB stack + 2GB BSS/DATA/TEXT) and shadow memory to 19/8GB. Similar to FuZZan, we also disable the logging and relocate the poisoning of global data to run before AFL’s fork server. As shown in Table 4, combining ASan-- and FuZZan outperforms both tools alone. It increases the execution speed of AFL by 59.4% and 59.3% on AFL’s seeds and FuZZan’s seeds. This leads to 8.99% and 9.73% more branch coverage. An outlier is tcpdump, where FuZZan runs the fastest. The primary reason is FuZZan also uses a red-black-tree based check to replace ASan check in certain cases, bringing extra optimization.

Our evaluation above runs only one fuzzing instance. In practice, we often need to run multiple instances in parallel and the community suggests using only one sanitized instance [8]. By intuition, the improvement of ASan-- will decrease with the number of non-sanitized instances. Accordingly, we perform an extra experiment where we increase the number of non-sanitized instances from 0 to 1 and 3, using AFL’s seeds. Figure 6 shows the results. Not surprisingly, the improvement of ASan-- reduces when more non-sanitized in-
Selective Deployment In general, ASan-- can still be considered unsuitable for deployed software because of its remaining runtime overhead and high memory cost. However, modern computers usually provide high computation power and large memory. Thus, it makes sense to selectively run ASan-- on critical deployed software (e.g., web browsers). Accordingly, we have been using the Chromium browser compiled with ASan-- for four weeks in our daily work. None of us experienced usability issues or significant slowdown.

To test ASan--’s capability of bug detection in the deployed environment, we collected 89 memory bugs from bugs.chromium.org and the CVE database. Replaying the bugs in the Chromium (same version as ours) pre-built with ASan by Google, we reproduced 4 bugs as listed in Table 5. Running the bugs in our Chromium, ASan-- detected all of them. We also tested the bugs with SANRazor and ASAP. However, both tools could not run Chromium. Alternatively, we performed static reasoning and observed that SANRazor and ASAP may miss 1 and 3 of the bugs (see Appendix C).

6 Related Work

6.1 Memory Error Detection

Spatial Error Detectors StackGuard [48] and P-SSP [50] insert random values before the return address and detect whether the random value is overwritten before the function returns. They introduce nearly zero runtime overhead but could only detect stack buffer overflows. Safe-C [10], CCured [37], SoftBound [36], Low-fat Pointers [27], and Gregory Jet al. [17] encode pointers with boundary information. At runtime, the information is used to check if memory accesses are within the boundaries. They can detect various spatial errors, but they have many problems (e.g., high runtime overhead and ABI incompatibility) that hinder their adoption.

Temporal Error Detectors DangNull [28], FreeSentry [53], and DangSan [47] keep track of pointers of allocated objects and invalidate pointers once the object is freed. CETs [35] maintains a unique identifier with each object and assigns a corresponding key to its pointers. When the object is freed, the lock is reset. Any access after the free can be detected by checking if the key encoded in the pointer matches the lock of the object. Undangle [14] taints pointers at the malloc site and propagates the pointers dynamically. It detects temporal errors by checking if the source of a pointer has been freed. These methods usually have high runtime overhead.

Sanitizers ASan [42] and memory sanitizer (MSan) [44] use shadow memory to record the sanity of application memory. They further place checks on memory accesses to inspect the shadow bytes and detect errors at runtime. ASan can detect both spatial ones and temporal ones. MSan complements ASan to detect uninitialized memory read. Compared to other memory error detectors, ASan and MSan offer better detection, higher efficiency, and broader generality.

6.2 Address Sanitizer Optimization

Reducing Address Space Techniques in this line reduce the address space used by the application and the shadow memory. This reduces cost of memory management and the overall overhead. A technique is FuZZan [24]. As we have discussed FuZZan, we omit the details here. According to our study in 3, using the approach of FuZZan helps reduce runtime overhead but still cannot handle 90%+ of it.

Reducing Sanitizer Checks Techniques in this category adopt various strategies to reduce sanitizer checks. ASAP [49], PartiSan [29], and Bunshin [51] enforce sanitizer checks to a subset of the code. ASAP profiles the program to identify “hot” code that is more often executed. It removes checks in the hot code and retains the remaining. As such, fewer, cheaper checks are inserted. PartiSan creates different versions of the same code segment, making some versions more sanitized and the others less. At runtime, PartiSan adjusts the versions per the performance budget. Both ASAP and PartiSan are performance-driven, thus missing memory errors. In contrast, Bunshin runs multiple variants of the same software. It distributes the sanitizer checks into different variants. Bunshin does not lose checks, but it incurs several times more resource consumption. Similar to ASAP, SANRazor [55] also profiles the program, but the goal is to identify and remove redundant checks. In principle, SANRazor offers higher safety than ASAP but still no guarantee.

7 Conclusion

This paper first presents a study on dissecting the overhead of ASan, bringing knowledge about the primary sources of its overhead. Inspired by the study, the paper develops ASan--, a tool assembling a set of four optimizations to reduce the runtime overhead of ASan. Unlike existing techniques that negatively hurt the capability, scalability, or usability of ASan, ASan--’s techniques well maintain these decent properties. Our evaluation shows high utility of ASan--. It can reduce about 40% and 36% of the overhead introduced by ASan into SPEC CPU2006 and Chromium. Applying to fuzzing, ASan-- can meaningfully increase the execution speed and benefit the branch coverage. Deployed in Chromium for daily activities, ASan-- can detect bugs discovered by the community.

Acknowledgments

We thank our shepherd Wenke Lee and the anonymous reviewers for their feedback. This project was supported by National Science Foundation (Grant#: CNS-2031377; CNS-1718782), Office of Naval Research (Grant#: N00014-17-1-2787;
Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agency.

References


Thus, false negatives are impossible.

This optimization only removes ASan checks. Hence, it will not introduce false positives. In the following, we focus on the discussion of false negatives.

Spatial Errors Our optimization ensures that the access falls into the legitimate range of the target object, thus avoiding false negatives.

Proof by Contradiction: Given a data object buf of N bytes on the stack or in the global, we assume the ASan check on a K-byte access buf[offset] is removed but the access goes out of bound. That means at least one path exists where offset < 0 || offset+K > N. This conflicts with Condition-1 presented in §4.1, and, thus, is infeasible.

Temporal Errors Global objects are always alive. As such, it has no temporal errors. Stack objects can have use-after-return and use-of-out-scope issues. However, we only target direct stack access, where allocation of the object can be identified by LLVM. If the access is either use-after-return or use-of-out-scope, LLVM will throw a compilation error. Thus, false negatives are impossible.
A.2 Removing Recurring Checks

This optimization also only removes ASan checks, so there are no false positives. Therefore, we again focus on the discussion of false negatives.

Spatial Errors When removing ASan checks on *ptr1, we ensure there is another access, *ptr2, to the same memory location, and *ptr2 exists in any path visiting *ptr1. This guarantees that any spatial error happening to *ptr1 will also happen to *ptr2. Thus, the error will be captured by the check on *ptr2.

Proof by Contradiction: There are two situations of false negatives. First, *ptr1 incurs a spatial error but *ptr2 does not. That means *ptr1 is in a redzone but *ptr2 is not. This conflicts with that *ptr1 and *ptr2 are in the same memory location. Second, *ptr1 is executed but *ptr2 is not. This conflicts with that *ptr2 exists in any path visiting *ptr1. In summary, false negatives are impossible.

Temporal Errors False negatives of temporal errors can only happen when the lifetime of the target memory changes. Specifically, given that *ptr1 dominates *ptr2 and the check on *ptr2 is removed, false negative of use-after-free can happen when the memory is freed between *ptr1 and *ptr2. This can happen in two scenarios. First, the memory is freed by the code between *ptr1 and *ptr2 through a heap deallocation interface. This is impossible because we will avoid removing the check in such a case (recall §4.2). Second, the memory is freed by another concurrent thread. This is possible, but it shall be a race condition issue instead of memory error, which in principle goes out of ASan’s scope. We also want to note that, when *ptr1 post-dominates *ptr2 and the check on *ptr2 is removed, no action is needed. The reason is that *ptr1 always runs later than *ptr2 and, hence, the use-after-free will be safely captured.

A.3 Merging Neighbor Checks

To simplify our discussion, we assume that the checks on *ptr1 and *ptr2 are merged into one by our optimization.

Spatial Errors False positives can never happen because whenever the merged check indicates an error, the original ASan checks will be performed again. Next we validate that false negatives cannot happen neither.

Proof by Contradiction: False negatives can happen only when *ptr1 and/or *ptr2 fall into a redzone but our merged check does not capture it. Formally, assuming that *ptr1 *ptr2 respectively fall into the memory ranges of [addr_1l, addr_1h] and [addr_2l, addr_2h]. False negatives mean our merged check does not cover the range of [addr_1l, addr_2h] (assuming addr_1l < addr_2h). This will conflict with Condition-3 in §4.3. Condition-3 ensures that our check covers the range [addr_1l, addr_1l + MaxBitRead] and MaxBitRead > addr_2h – addr_1l. False negatives are, thus, excluded.

Temporal Errors False positives are impossible. Our optimization ensures that *ptr1 and *ptr2 access the same data object, and modern heap allocator mandates that an data object cannot be partially freed. Thus, whenever use-after-free is detected, that certainly happens to *ptr1 and/or *ptr2. False negatives can only happen when the object is freed between *ptr1 and *ptr2. But this is OK since we always insert the merged check at the post-dominating access. The free will be safely captured.

A.4 Removing Neighbor Checks

This optimization again only removes ASan checks, so no false positives will be introduced. Below we discuss the case of false negatives. For simplicity, we assume *ptr2 falls into the range between *ptr1 and *ptr3 and the check on *ptr2 is removed. Further, we assume (i) *ptr1, *ptr2, and *ptr3 fall into the memory ranges of [addr_1l, addr_1h], [addr_2l, addr_2h], and [addr_3l, addr_3h] and (ii) the location of *ptr1, *ptr2, and *ptr3 ranges from low memory address to high memory address.

Spatial Errors False negatives are impossible.

Proof by Contradiction: False negatives can only happen when *ptr2 is in a redzone but both *ptr1 and *ptr3 are not. That means (addr_21 – addr_1h) + (addr_2h – addr_21) + (addr_3l – addr_2h) >= MinRdSz. This will conflict with Condition-4.

Temporal Errors False negatives of temporal errors are impossible. Our optimization relocates a check to be executed later than the original one. Thus, no false negatives will be introduced because a freed object will not be reallocated. False positives can only happen when the object is freed in the code between the original access and the loop exit. This is possible, but our algorithm to identify invariant address (Algorithm 3) has considered such cases. If the target is possibly freed inside the loop, we will skill relocating the ASan check. Note that, similar to removing recurring checks, this optimization may report false temporal errors caused by de-allocation in another thread.

A.5 Relocating Invariant Checks

Spatial Errors Since the address of invariant checks never changes, no false negative or false positive will be introduced.

Temporal Errors This optimization relocates a check to be executed later than the original one. Thus, no false negatives will be introduced because a freed object will not be reallocated. False positives can only happen when the object is freed in the code between the original access and the loop exit. This is possible, but our algorithm to identify invariant address (Algorithm 3) has considered such cases. If the target is possibly freed inside the loop, we will skill relocating the ASan check. Note that, similar to removing recurring checks, this optimization may report false temporal errors caused by de-allocation in another thread.
A.6 Grouping Monotonic Checks

This optimization only reduces checks, so it will not introduce false positives. We discuss false negatives in the following.

**Spatial Errors** Given a chunk of memory ranging from addr\_l to addr\_h (addr\_h - addr\_l \(\leq\) MinRdSz), our optimization only checks the beginning bytes. False negatives may happen when the beginning bytes fall out of a redzone but the remaining bytes fall in. This is possible but will not bring false negatives. In general, there are two situations. First, the loop continues iterating and enters the next chunk. In this case, the beginning bytes of the new chunk will overlap with the redzone and the error will be detected. Second, the loop exits without entering the next chunk. In this case, the last access will touch the redzone and our check at the exit of the loop will detect the redzone.

**Temporal Errors** If the object is free when the loop iterate over the same chunk, false negative can happen. But it can only happen when the chunk is the last chunk, because otherwise the free will be detected when the next chunk is accessed. Similar to spatial errors, such cases shall be captured by our check at the exit of the loop.

B Other Applications of Optimizations

Our optimizations are designed to improve ASan as it draws more attention. However, they can also be applied to both Undefined Behavior Sanitizer (UBSan) [31] and Memory Sanitizer (MSan) [44]. Specifically, removing unsatisfiable checks and relocating invariant checks can be directly applied to UBSan. In fact, removing unsatisfiable checks has been applied on UBSan before and presented promising results [15, 30]. Removing recurring checks is also applicable, and it can be even further extended by defining recurring checks on non-memory objects (e.g., integers). Other optimizations may not be applied as UBSan uses no shadow memory.

All our optimizations, except removing neighbor checks, are applicable to MSan. Removing neighbor checks cannot be applied because MSan uses no redzones. When applied to MSan, our optimizations need the following customization:

- **Removing unsatisfiable checks**: In the context of MSan, the unsatisfiable condition should be redefined as that a memory read is dominated by a write to the same location.
- **Removing recurring checks**: This optimization is directly applicable. One thing worth noting is that when \(*ptr2\) post-dominates \(*ptr1\), we need to make sure there is no write to the same location between \(*ptr1\) and \(*ptr2\), before we remove the check on \(*ptr1\).

- **Merging/removing neighbor checks**: The two optimizations are also applicable. When merging the checks on \(*ptr1\) and \(*ptr2\) (assuming \(*ptr1\) dominates \(*ptr2\)), we will place the merged check on \(*ptr1\) under the requirement that no write can happen to \(*ptr2\) after \(*ptr1\). When removing neighbor checks, we also need to ensure no write happens between the removed one and dominating others.

- **Optimization loop checks**: The two optimizations for loops are in principle applicable. But for safety, we will have to make sure the target memory location is not initialized inside the loop. Statically guaranteeing this can lead to skipping of too many cases, and thus, the optimizations can present limited effectiveness.

C Discussion of Chromium Bugs

This section discusses unsound tools, ASAP and SANRAZOR, can detect the Chromium bugs and identify the reasons if not.

**Issue 116869 (SANRAZOR \(\checkmark\); ASAP \(\times\))**: The bug, shown in the code below, happens because the index m\_clipIndex gets excessively decreased to become negative (line 4) and thus, the access to m\_commands at line 5 goes out of bound. Since no memory access shares the same data dependency as m\_commands[m\_clipIndex], SANRAZOR will not remove the check and can detect the bug. In contrast, ASAP will consider m\_commands[m\_clipIndex], which happens in a loop, as hot code. If the while loop is executed in profiling with a low overhead budget, ASAP may remove its sanitizer check, and in result, miss detecting the bug.

```c
void AdjustClip(int limit) {
    while (m_clipIndex > limit) {
        --m_clipIndex;
    }
}
```

**Issue 1099446 (SANRAZOR \(\checkmark\); ASAP \(\checkmark\))**: This bug happens due to unsafe use of memcpy at line 5. The number of bytes operated is larger than size of the source buffer. Both SANRAZOR and ASAP should be able to detect this bug as they do not optimize the use of library functions.

<table>
<thead>
<tr>
<th>Module</th>
<th>Type</th>
<th>ASan</th>
<th>ASan--</th>
<th>Issue (with PoC)</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDFium</td>
<td>heap-buffer-underflow</td>
<td>✓</td>
<td>✓</td>
<td><a href="https://bugs.chromium.org/p/chromium/issues/detail?id=116869">https://bugs.chromium.org/p/chromium/issues/detail?id=116869</a></td>
<td>Compiling with pdf_use_skia=true</td>
</tr>
<tr>
<td>PDFium</td>
<td>heap-buffer-overflow</td>
<td>✓</td>
<td>✓</td>
<td><a href="https://bugs.chromium.org/p/chromium/issues/detail?id=1099446">https://bugs.chromium.org/p/chromium/issues/detail?id=1099446</a></td>
<td>Compiling with pdf_use_skia=true</td>
</tr>
<tr>
<td>FreeType</td>
<td>heap-buffer-overflow</td>
<td>✓</td>
<td>✓</td>
<td><a href="https://bugs.chromium.org/p/chromium/issues/detail?id=1039953">https://bugs.chromium.org/p/chromium/issues/detail?id=1039953</a></td>
<td>Linking lib_freetype_harfbuzz</td>
</tr>
</tbody>
</table>

Table 5: Memory errors replayed with the deployed Chromium. Both ASan and ASan-- can detect all of them.
sk_sp<SkData> SkData::PrivateNewWithCopy(...) {
    ... if (srcOrNull) {
        // heap-buffer-overflow here; The size of
        "srcOrNull" is 68 bytes, but "length" is 256
        memcpy(data->writable_data(), srcOrNull, length);
    }
    return data;
}

function do {...}

if (dp && sp && ... ) {
    void png_combine_row(png_const_structrp png_ptr,
        png_bytep dp, int display){
        ... if (dp & sp & ...) {
            png_uint_32p dp32 = dp, sp32 = sp;
            do{ 
                *dp32++ = *sp32++; //dp32 points to the
                buffer and overflows
            ...}
            while (c > 0);
            }while(bytes_to_copy <= row_width)
    }
}

D Supplementary Algorithms and Data

Table 6: Page fault and runtime overhead before and after we shrink the shadow memory of ASan.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th># of Page Faults (K)</th>
<th>Runtime Overhead</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
<td>After</td>
</tr>
<tr>
<td>400.perlibench</td>
<td>187,718</td>
<td>138,214</td>
</tr>
<tr>
<td>401.bzip2</td>
<td>1,159</td>
<td>933</td>
</tr>
<tr>
<td>403.gcc</td>
<td>136,262</td>
<td>91,466</td>
</tr>
<tr>
<td>429.mcf</td>
<td>626</td>
<td>463</td>
</tr>
<tr>
<td>445.gobmk</td>
<td>432</td>
<td>390</td>
</tr>
<tr>
<td>456.hmmer</td>
<td>373</td>
<td>333</td>
</tr>
<tr>
<td>458.sjeng</td>
<td>89</td>
<td>78</td>
</tr>
<tr>
<td>462.libquantum</td>
<td>555</td>
<td>413</td>
</tr>
<tr>
<td>464.h264ref</td>
<td>390</td>
<td>315</td>
</tr>
<tr>
<td>473.astar</td>
<td>607</td>
<td>526</td>
</tr>
<tr>
<td>483.xalancbmk</td>
<td>648</td>
<td>593</td>
</tr>
<tr>
<td>433.milc</td>
<td>29,400</td>
<td>21,380</td>
</tr>
<tr>
<td>444.namd</td>
<td>50</td>
<td>46</td>
</tr>
<tr>
<td>447.dealII</td>
<td>1,694</td>
<td>1,424</td>
</tr>
<tr>
<td>450.soplex</td>
<td>12,863</td>
<td>9,294</td>
</tr>
<tr>
<td>451.povray</td>
<td>121</td>
<td>117</td>
</tr>
<tr>
<td>470.lbm</td>
<td>117</td>
<td>138</td>
</tr>
<tr>
<td>482.sphinx3</td>
<td>12,863</td>
<td>9,293</td>
</tr>
<tr>
<td>Average</td>
<td>21,442</td>
<td>15,301</td>
</tr>
</tbody>
</table>

Table 7: Comparison of compilation time and binary size

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Compilation Time (Sec)</th>
<th>Binary Size (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Allen</td>
<td>Allen--</td>
</tr>
<tr>
<td>400.perlibench</td>
<td>65.2</td>
<td>397</td>
</tr>
<tr>
<td>401.bzip2</td>
<td>12.1</td>
<td>21.5</td>
</tr>
<tr>
<td>403.gcc</td>
<td>25.2</td>
<td>85.5</td>
</tr>
<tr>
<td>429.mcf</td>
<td>7.01</td>
<td>11.1</td>
</tr>
<tr>
<td>445.gobmk</td>
<td>20.3</td>
<td>32.5</td>
</tr>
<tr>
<td>456.hmmer</td>
<td>18.8</td>
<td>36.7</td>
</tr>
<tr>
<td>458.sjeng</td>
<td>7.11</td>
<td>14.1</td>
</tr>
<tr>
<td>462.libquantum</td>
<td>7.12</td>
<td>14.1</td>
</tr>
<tr>
<td>473.astar</td>
<td>31.4</td>
<td>113</td>
</tr>
<tr>
<td>483.xalancbmk</td>
<td>10.8</td>
<td>15.3</td>
</tr>
<tr>
<td>445.soplex</td>
<td>50.3</td>
<td>412</td>
</tr>
<tr>
<td>453.milc</td>
<td>12.1</td>
<td>21.9</td>
</tr>
<tr>
<td>444.namd</td>
<td>17.3</td>
<td>57.1</td>
</tr>
<tr>
<td>447.dealII</td>
<td>44.7</td>
<td>125</td>
</tr>
<tr>
<td>450.sphinx3</td>
<td>6.81</td>
<td>19.7</td>
</tr>
<tr>
<td>453.povray</td>
<td>9.81</td>
<td>45.1</td>
</tr>
<tr>
<td>470.lbm</td>
<td>4.57</td>
<td>7.76</td>
</tr>
<tr>
<td>482.sphinx3</td>
<td>6.17</td>
<td>9.34</td>
</tr>
<tr>
<td>Average</td>
<td>19.2</td>
<td>29.2</td>
</tr>
</tbody>
</table>

Issue 1139963 (SANRAZOR X; ASAP X): This bug is caused by an integer truncation. At line 4 and line 6, imgWidth and imgHeight are truncated from 32 bits to 16 bits. The truncated values are then used to calculate the size of a bitmap at line 9. However, the bitmap is still accessed using a 32-bit index at line 20. When imgWidth and/or imgHeight is larger than 65535 (2\(^16\) - 1), the allocated buffer cannot fit the bitmap and thus, overflow can happen at line 22.

According to SANRAZOR’s definition, dp at line 16 and dp32 at line 20 share similar data dependency as they “flow” from the same source. Further, SANRAZOR deems line 17 as a “user-check”. As such, SANRAZOR considers that ASAP check on *dp32 at 20 is a redundancy of line 17 and will remove it. This way SANRAZOR shall miss the memory error. Similar to Issue 116869, *dp32 at 22 happens in a loop, and ASAP may miss it when a low overhead budget is given.

Issue 848914 (SANRAZOR ✓; ASAP X): Code below demonstrates the bug. In this issue, id_states_ is a C++ standard vector. When an id object is created, it will be inserted to id_states_ and the index of the object will be saved in ids. However, when objects are popped out from id_states_, the index saved to ids is not updated. As a consequence, heap over流传 can happen when id_states_ is accessed using the old index (id_states_[id - 1] at line 5). This case is similar to Issue 116869. SANRAZOR will not remove the check and can detect the overflow. However, if the loop is executed in profiling with a low overhead budget, ASAP may remove the check and miss detecting the bug.

bool FreeIds(...) override {
    ... for(GLsize i = 0; i < n; ++i) {
        GLuint id = ids[i];
        if (id != 0) id_states_[id - 1] = kIdPendingFree;
    }
    ...
Table 8: The amount of ASan checks after the adoption of ASan—Baseline is ASan with the default setting. From left to right, the columns show the amount of ASan checks left after enabling different optimizations in ASan—, one after another.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Ratio of ASan Checks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1G</td>
</tr>
<tr>
<td>400.getdsk</td>
<td>64.64%</td>
</tr>
<tr>
<td>401.tomp2</td>
<td>88.29%</td>
</tr>
<tr>
<td>401.tomp3</td>
<td>70.95%</td>
</tr>
<tr>
<td>420.mcf</td>
<td>83.06%</td>
</tr>
<tr>
<td>440.mcf</td>
<td>82.96%</td>
</tr>
<tr>
<td>450.mcf</td>
<td>72.66%</td>
</tr>
<tr>
<td>450.mttm</td>
<td>88.95%</td>
</tr>
<tr>
<td>462.mttm</td>
<td>64.82%</td>
</tr>
<tr>
<td>462.mttm</td>
<td>93.59%</td>
</tr>
<tr>
<td>471.asan</td>
<td>92.76%</td>
</tr>
<tr>
<td>471.asan</td>
<td>99.99%</td>
</tr>
<tr>
<td>472.milc</td>
<td>89.93%</td>
</tr>
<tr>
<td>472.milc</td>
<td>90.55%</td>
</tr>
<tr>
<td>474.milc</td>
<td>95.83%</td>
</tr>
<tr>
<td>475.milc</td>
<td>94.83%</td>
</tr>
<tr>
<td>476.milc</td>
<td>93.19%</td>
</tr>
<tr>
<td>477.milc</td>
<td>92.96%</td>
</tr>
<tr>
<td>Average</td>
<td>87.78%</td>
</tr>
</tbody>
</table>

Table 9: Performance improvement brought by FuZZan to ASan on SPEC CPU2006. Cells with "-" indicate no results.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Performance Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1G</td>
</tr>
<tr>
<td>401.hzip2</td>
<td>—</td>
</tr>
<tr>
<td>429.mcf</td>
<td>—</td>
</tr>
<tr>
<td>456.milmer</td>
<td>—</td>
</tr>
<tr>
<td>458.sjeng</td>
<td>—</td>
</tr>
<tr>
<td>462.libquantum</td>
<td>—</td>
</tr>
<tr>
<td>433.milc</td>
<td>—</td>
</tr>
<tr>
<td>444.namh</td>
<td>1.89%</td>
</tr>
<tr>
<td>453.povray</td>
<td>—</td>
</tr>
<tr>
<td>470.lbm</td>
<td>—</td>
</tr>
<tr>
<td>482.sphinx3</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 10: Setup of fuzzing evaluation.

<table>
<thead>
<tr>
<th>Programs</th>
<th>Name</th>
<th>Version</th>
<th>Driver</th>
<th>Source</th>
<th>Options</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>BINUTLIS</td>
<td>2.31</td>
<td>OBJUMP</td>
<td>[20]</td>
<td>-a @ @</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BINUTLIS</td>
<td>2.31</td>
<td>SIZE</td>
<td>[20]</td>
<td>@ @</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BINUTLIS</td>
<td>2.31</td>
<td>C++</td>
<td>[20]</td>
<td>-n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BINUTLIS</td>
<td>2.31</td>
<td>NM</td>
<td>[20]</td>
<td>@ @</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIBPCAP</td>
<td>5.00</td>
<td>TCPDUMP</td>
<td>[45]</td>
<td>-n e - r @ @</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIBPCAP</td>
<td>5.00</td>
<td>FNGFX</td>
<td>[45]</td>
<td>@ @</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FILE</td>
<td>1.62</td>
<td>FILE</td>
<td>[23]</td>
<td>-m magic.mgc @ @</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11: Runtime overhead on SPEC CPU2017. **_r represents SPECrate and **_s represents SPECSpeed.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Performance Overhead</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1G</td>
</tr>
<tr>
<td>864.fft</td>
<td>83.70%</td>
</tr>
<tr>
<td>865.fft</td>
<td>83.70%</td>
</tr>
<tr>
<td>866.fft</td>
<td>83.70%</td>
</tr>
<tr>
<td>867.fft</td>
<td>83.70%</td>
</tr>
<tr>
<td>Average</td>
<td>83.70%</td>
</tr>
<tr>
<td>864.inst</td>
<td>83.70%</td>
</tr>
<tr>
<td>865.inst</td>
<td>83.70%</td>
</tr>
<tr>
<td>866.inst</td>
<td>83.70%</td>
</tr>
<tr>
<td>867.inst</td>
<td>83.70%</td>
</tr>
<tr>
<td>Average</td>
<td>83.70%</td>
</tr>
</tbody>
</table>

Algorithm 3: Identifying Invariant in Loop

```
Procedure backward_slicing_inLoop(\mathcal{A}, \mathcal{L}, \mathcal{V}):
1. if \neg L.contains(A) \lor \mathcal{V}.find(A) then
2. return;
3. end
4. \mathcal{V}.push_back(A);
5. if Typelnst P = dyn_cast<Typelnst>(A) then
6. for each op in P.getOperands() do
7. backward_slicing_inLoop(op, \mathcal{L}, \mathcal{V});
8. end
9. end
10. return;
11. Procedure check_addr_type(\mathcal{A}, \mathcal{V}, \mathcal{P}):
12. if \mathcal{P}.find(A) then
13. return NON_INV;
14. end
15. for each user in A.users() do
16. if isa<CallInst>(user) || isa<LifeTimeEndIntrinsic>(user) then
17. return NON_INV;
18. end
19. if isa<StoreInst>(user) then
20. return INV;
21. end
22. end
23. if \neg \mathcal{V}.find(A) then
24. return INV;
25. end
26. \mathcal{P}.push_back(A);
27. if Typelnst K = dyn_cast<Typelnst>(A) then
28. for each op in K.getOperands() do
29. opTy = check_addr_type(op, \mathcal{V}, \mathcal{P});
30. if opTy == NON_INV then
31. return NON_INV;
32. end
33. end
34. end
35. end
36. return INV;
37. Algorithm 4: Identifying Memory Invariant in Loop

Input : Address \mathcal{A}, Loop \mathcal{L}
Output : INV or NON_INV
Initialize: \mathcal{P} = vector();
backward_slicing_inLoop(\mathcal{A}, \mathcal{L}, \mathcal{V});
return check_addr_type(\mathcal{A}, \mathcal{V}, \mathcal{P});
```