

# HardRace: A Dynamic Data Race Monitor for Production Use

XUDONG SUN, Nanjing University, China  
 ZHUO CHEN, Nanjing University, China  
 JINGYANG SHI, Nanjing University, China  
 YIYU ZHANG, Nanjing University, China  
 PENG DI, Ant Group, China  
 XUANDONG LI, Nanjing University, China  
 ZHIQIANG ZUO, Nanjing University, China

Data races are critical issues in multithreaded program, leading to unpredictable, catastrophic and difficult-to-diagnose problems. Despite the extensive in-house testing, data races often escape to deployed software and manifest in production runs. Existing approaches suffer from either prohibitively high runtime overhead or incomplete detection capability. In this paper, we introduce HardRace, a data race monitor to detect races on-the-fly while with sufficiently low runtime overhead and high detection capability. HardRace firstly employs sound static analysis to determine a minimal set of essential memory accesses relevant to data races. It then leverages hardware trace instruction, *i.e.*, Intel PTWRITE, to selectively record only these memory accesses and thread synchronization events during execution with negligible runtime overhead. Given the tracing data, HardRace performs standard data race detection algorithms to timely report potential races occurred in production runs. The experimental evaluations show that HardRace outperforms state-of-the-art tools like ProRace and Kard in terms of both runtime overhead and detection capability – HardRace can detect all kinds of data races in read-world applications while maintaining a negligible overhead, less than 2% on average.

CCS Concepts: • **Do Not Use This Code** → **Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; *Generate the Correct Terms for Your Paper*; *Generate the Correct Terms for Your Paper*.

Additional Key Words and Phrases: data race, dynamic detection, hardware, static analysis

## ACM Reference Format:

Xudong Sun, Zhuo Chen, Jingyang Shi, Yiyu Zhang, Peng Di, Xuandong Li, and Zhiqiang Zuo. . HardRace: A Dynamic Data Race Monitor for Production Use. In . ACM, New York, NY, USA, 20 pages. <https://doi.org/XXXXXXX.XXXXXXX>

## 1 INTRODUCTION

In the post-Moore era, multithreaded software become prevalent; and concurrency errors become more and more common in multithreaded programs. Such errors cause critical issues such as program crashes [19], security vulnerabilities [13], and incorrect computations [22], leading to serious real-world social and economic hazards, *e.g.*, the Northeast blackout, and mismatched Nasdaq Facebook share prices.

In spite of extensive in-house testing, data races often escape to deployed software and manifest in production runs [20, 24]. This is because data races are highly sensitive to the execution states, including program inputs, thread interleavings, platform configurations, and other execution environments [20]. Such huge execution space can hardly be completely covered by testing. For the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Conference acronym 'XX,

© Copyright held by the owner/author(s). Publication rights licensed to ACM.  
<https://doi.org/XXXXXXX.XXXXXXX>

same reason, production-run data races are difficult to reproduce and fix offline [12]. As a result, it is highly desirable to propose an online data race detector which is able to discover data races in production runs practically (with sufficiently low overhead) and effectively (with high accuracy).

**Prior Work.** Over the past decades, extensive studies have been conducted, primarily classified as static and dynamic approaches. Static data race detectors, such as RacerD [5] and CHES [21], analyze the code statically without executing it. It can report the potential races before deployment, thus being free of runtime overhead. However, due to the inherent limitation of over-approximation, the static approaches inevitably report many false positive warnings, severely affecting its practicability.

On the other hand, dynamic data race detectors [10, 23, 24], monitor the program during execution to identify actual data races. Compared to static approaches [5, 21] which suffer from high false positives, dynamic detectors have the advantage of high precision. However, as they demand to collect and analyze massive execution data online, the program execution is dramatically slowed down. For example, Google's ThreadSanitizer (a.k.a., TSan) [24], the most mature and widely used tool in industry, can result in an average 7-12x slowdown [1]. FastTrack also incurs the runtime overhead of a similar magnitude as reported [10]. Such high overheads severely inhibit the practical usage of these dynamic detectors – they nowadays only work in debugging/testing mode but not production-run environment.

Recently, several attempts have been performed to lower the runtime overhead of dynamic race detectors [1, 11, 14, 30, 31]. RaceMob [14] adopts crowdsourcing mechanism to lower the runtime overhead of each individual user. ProRace [30] samples memory accesses using hardware PMU (in particular, Intel's Precise Event Based Sampling), thus achieving low runtime overhead. However, like all sampling approaches [6, 25, 30], finding the right sampling rate is often challenging. Even worse, the detection accuracy is usually not guaranteed. Kard[1] leverages Intel Memory Protection Keys (MPK) to achieve low detection overhead. However, due to the limitations of MPK, Kard can only detect a specific type of races, namely Inconsistent Lock Usage (ILU). Other common races remain undetectable.

**Our Work.** This paper introduces HardRace, a novel data race detector which leverages modern hardware tracing module (in particular, Intel Processor Trace instructions PTWRITE) to monitor data races in production runs, with sufficiently low overhead and high detection capability. However, naively employing hardware tracing to dynamically detect data races faces two problems.

First, naively tracing all the memory accesses via hardware still yields prohibitively high runtime overhead. The reason is that data race detection requires tracking extensive runtime information, including memory accesses and thread synchronization events. To trace such dynamic information, a massive number of hardware tracing instructions (*i.e.*, *ptwrite*) have to be instrumented and executed, leading to non-negligible overhead. Our empirical experiments show that the overhead of naive hardware tracing for data race detection reaches 19.8% on average (see §7.2).

Second, naively recording the intensive memory access and thread synchronization information via hardware results in severe data loss, greatly diminishing detection capability. This is because the hardware generates traces much faster than memory can keep up. As a result, certain traces would be lost especially if the hardware trace instructions (*i.e.*, *PTWRITE*) is too dense [32]. In our experiments, naive hardware tracing leads to frequent data loss, 37% on average, which significantly affects the detection capability (see §7.4).

To tackle the above problems, HardRace firstly employs effective static analysis to safely eliminate most memory accesses that are unlikely to be involved in data races. It then selectively instruments and traces only the remaining accesses via hardware.

**Results.** We implemented HardRace and evaluated it over a common set of benchmarks, including the widely used PARSEC/SPLASH-2x benchmark suite and a set of real-world applications with known data races [29]. The experimental results show that the runtime overhead of HardRace is only around 1.6% on average, which is significantly lower than that of the state-of-the-art approaches. Moreover, HardRace can detect the data races in all the experimental subjects without any false negatives, whereas the existing dynamic detection tools like ProRace and Kard miss them a lot.

This paper makes the following contributions:

- HardRace presents an effective data race detector with minimal overhead, that can be deployed to monitor production runs.
- HardRace firstly employs binary static analysis to safely prune away unnecessary memory accesses, and then leverage modern hardware tracing module (*i.e.*, Intel PTWRITE) to realize selective tracing, achieving sufficiently low overhead and high detection precision. To the best of our knowledge, HardRace is the first to utilize Intel PTWRITE for precise and low-overhead data race detection.
- The experimental evaluations show that HardRace outperforms state-of-the-art tools like ProRace and Kard in terms of both runtime overhead and detection capability – HardRace can detect all kinds of data races in read-world applications while maintaining a negligible overhead compared to the existing solutions.

**Outline.** The rest of the paper is organized as follows. §2 gives the necessary background of hardware tracing and dynamic data race detection. §3 provides the overview of HardRace. §4 and §5 describe the key components we proposed, followed by the implementation in §6. We present the empirical evaluations in §7. We talk about the related work in §8. Finally, §9 concludes.

## 2 BACKGROUND

### 2.1 Intel PTWRITE

Intel PTWRITE is an extended hardware tracing feature, which is available at the commodity PCs with the 12th generation (Alder Lake) desktop processors. It provides the PTWRITE instruction to efficiently and flexibly record data values from registers or memory. If a register value of interest needs to be recorded, users can insert a PTWRITE instruction with the register as the operand. When the instruction is executed as the program runs, the hardware module writes the dynamic value of the register into a specific system buffer, which can be then read for usage.

Intel PTWRITE is particularly advantageous for tracing multithreaded programs, where traditional software instrumentation often relies on expensive locking operations to determine the order of memory events across different threads. Such heavy intervention not only leads to a dramatic slowdown of the execution, but also significantly interferes with the original thread interleaving. In contrast, Intel PTWRITE can efficiently and precisely record traces containing timestamp information (like TSC packets) thanks to dedicated hardware. When combined with timestamped thread-switching events recorded via OS tools (such as perf events), it becomes possible to associate the hardware trace packets (*i.e.*, PTW packet) with each thread, thereby allowing for the reconstruction of a timeline of events across all threads.

Despite that Intel PTWRITE is of much lower overhead than traditional software techniques, it still encounters bottlenecks when recording the sheer volume of data. In particular, high-frequency recording scenarios can result in performance degradation and/or severe data loss [7]. Thus, naively applying PTWRITE does not fully address the issue of multi-threaded recording. In this paper, we propose dedicated static analysis to eliminate the unnecessary tracing points, finally ensuring low-overhead and avoiding data loss.

## 2.2 Data Race Detection

Given the memory access traces, a race detection algorithm can be employed to determine if a race potentially happens – two memory accesses may occur simultaneously. Over the past years, various existing data race detection algorithms are proposed, including lockset [23], happens-before relation [16], causally-precedes relation [27], weak-causally-precedes relation [15], etc.

The lockset algorithm, as its name suggests, focuses on lock/unlock accesses in multithreaded programs [23]. Its primary principle is that shared variables must be protected by the same lock; otherwise, it is assumed that simultaneous access exists. However, not all multithreaded programs use locks to ensure logical non-concurrency. For instance, operations like signal and wait can also establish a temporal ordering. A sequence like `<access A -> signal -> wait -> access A>` clearly does not constitute a data race. Therefore, the lockset algorithm is sometimes overly strict, leading to a significant number of false positives.

The happens-before algorithm considers whether there is a sequential execution relationship between events across different threads (named as happens-before relationship) [16]. It avoids the false positives of lockset algorithm. However, its detection result is sensitive to the particular thread interleaving, meaning that even if one execution does not exhibit a race, it does not guarantee that other interleavings are also free of races. Thus, exploring a sufficient number of interleavings is necessary to reduce the risk of false negatives. There are also other algorithms as the extension of happens-before relation, such as causally-precedes relation [27] and weak-causally-precedes relation [15]. Basically, by relaxing the relation constraints, more races can be detected.

In this paper, we propose an online data race monitor HardRace, which focuses on reducing the runtime overhead without sacrificing detection capability. In brief, HardRace first collects the necessary data access events at runtime, which are then fed into an existing detection algorithm to generate the report. For the sake of fair comparison, the offline analysis of HardRace adopts the combination of happens-before and lockset algorithms which are used by the state-of-the-art dynamic race detectors [1, 10, 24, 30]. Note that the contribution of this paper is to propose a low-overhead and high-precision data access monitoring approach for multithreaded programs via hardware tracing, which is orthogonal to the race detection algorithm. The race detection algorithms can benefit from our lightweight monitoring; and HardRace can also adopt any other detection algorithms in the offline analysis.

## 3 OVERVIEW

HardRace is a data race monitor which can on-the-fly detect races happened in production runs. Figure 1 demonstrates the workflow of HardRace, consisting of three main stages: static selective instrumentation, runtime trace collection, and offline trace analysis.

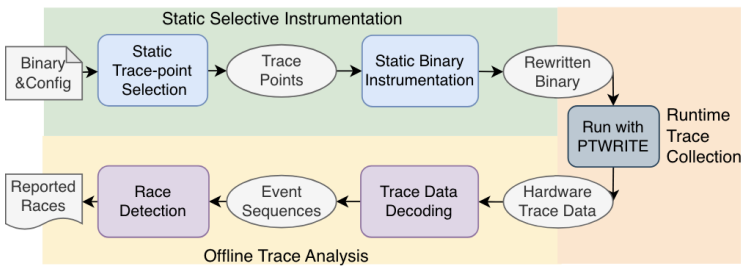


Fig. 1. Workflow of HardRace

**Static Selective Instrumentation.** The static selective instrumentation stage(see §4) is the core of our contribution. It is designed to statically identify a minimum set of memory accesses that are involved in data races and selectively instrument PTWRITE instructions to record them at runtime. It contains two key components: static trace-point selection (§4.1–§4.3) and static binary instrumentation (§4.4). Specifically, HardRace takes the target binary file and some configuration settings as input. Static trace-point selection utilizes a series of sound static analysis algorithms to determine a set of memory accesses that must be not involved in data races and excludes them from the instrumentation points. Given a minimum set of memory access points to be recorded, the binary instrumentation module is responsible to insert PTWRITE instructions to record the identified data accesses and thread synchronization events.

**Runtime Trace Collection.** At this stage, the instrumented binary is executed on the CPUs with Intel PTWRITE supported in production runs. With the appropriate setting, hardware trace packets are continuously generated. We will give more implementation details in §6.

**Offline Trace Analysis.** The offline analysis takes the hardware traces as input, and employs race detection algorithm to produce the final report. In particular, the tracked hardware traces are first decoded into per-thread memory access and synchronization event sequences (see §5.1). These sequences are then passed to race detection algorithm for efficient data race detection (see §5.2). The details about multithreaded-trace decoding and race detection will be elaborated in §5 shortly.

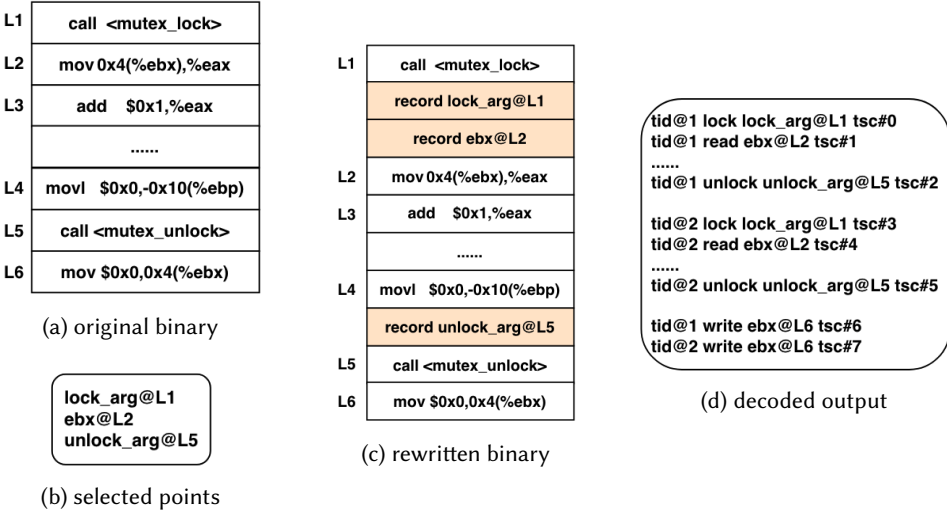


Fig. 2. A toy example

**Example.** Let’s use a toy example to illustrate the entire workflow of HardRace. Suppose there is an assembly fragment in the original binary as shown in Figure 2a. There are five instructions that involving memory accesses: *L1*, *L2*, *L4*, *L5*, and *L6*. Naively, the values we need to record are *lock\_arg@L1*, *ebx@L2*, *ebp@L4*, *unlock\_arg@L5* and *ebx@L6*. After filtering by the selection module (§4), we can narrow down the trace points to *lock\_arg@L1*, *ebx@L2*, and *unlock\_arg@L5*, as shown in Figure 2b. In this example, the two points *ebp@L4* and *ebx@L6* can be safely eliminated since *ebp@L4* is a stack address so we can easily notice it is race-free. The value of register *ebx@L6* can be statically deduced according to that of *ebx@L2*. During the static binary instrumentation, we insert three PTWRITE instructions into the binary, as illustrated in Figure 2c. This rewritten

binary is then executed during runtime trace collection, which generates trace data including thread information and PTWRITE packets. By decoding trace data, we get the sequence of memory accesses and synchronization events distinguished by threads, as shown in Figure 2d. Here we assume there are two threads in the output, and each event records thread id, event type, access address and timestamp. Having the decoded traces, we can directly employ the existing data race detection algorithm to obtain the final report. In this example, it is evident that the instructions at *L2* and *L6* access the same memory address in different threads *tid@1* and *tid@2*, and not be protected by same lock, which leading to a data race.

#### 4 STATIC SELECTIVE INSTRUMENTATION

As mentioned above, the purpose of static selective instrumentation is to identify a minimum set of trace points including memory accesses and synchronization events. It filters out unnecessary trace points through a series of static analyses, thus reducing the runtime overhead. In the following, we will elaborate the static trace-point selection analysis in §4.1 to §4.3. Once having a subset of trace points, we then insert Intel PTWRITE instructions into the subject binary so as to record the necessary data at runtime. §4.4 gives the detailed description about static binary instrumentation.

---

##### Algorithm 1: Static Selective Instrumentation

---

**Input:** Binary *prog*

**Output:** Rewritten binary *prog<sub>re</sub>*

---

- 1  $ICFG \leftarrow$  construct the interprocedural control flow graph of *prog*
  - 2  $valueSetResult \leftarrow$  MULTIVSA( $ICFG$ ) /\*inter-procedural value set analysis for multithreaded programs (§4.1)\*/
  - 3  $T_{shared} \leftarrow$  FINDALLSHAREDADDRESSES( $valueSetResult$ ) /\*find all the race-relevant trace points involving shared variable addresses (§4.1)\*/
  - 4  $T_{raceFree} \leftarrow$  MUSTRACEFREE( $T_{shared}, valueSetResult$ ) /\*identify the trace points which must not be relevant to data races (§4.2)\*/
  - 5  $T_{redundant} \leftarrow$  REDUNDANTANALYSIS( $T_{shared} - T_{raceFree}$ ) /\*identify the redundant trace points whose values can be statically deduced (§4.3)\*/
  - 6  $T_{trace} \leftarrow T_{shared} - T_{raceFree} - T_{redundant}$
  - 7  $prog_{re} \leftarrow$  INSTRUMENT(*prog*,  $T_{trace}$ ) /\*instrument PTWRITE instructions (§4.4)\*/
- 

Algorithm 1 provides a high-level description of how we progressively narrow down the set of trace points  $T_{trace}$  that require instrumentation. At first, we construct the interprocedural control flow graph (ICFG) of the input program (Line 1). Next, we perform our inter-procedural value set analysis tailored for multithreaded programs to produce the binary-level alias results (Line 2, see §4.1). Based on the value set results  $valueSetResult$ , we identify a set of trace points  $T_{shared}$  that involve shared variable addresses, i.e., global and heap addresses (Line 3). Memory access points that only access stack addresses are filtered out at this stage, as they cannot cause data races. Given the value set results  $valueSetResult$  and all the potential trace points  $T_{shared}$ , a must race-free analysis is then performed to identify the set of trace points  $T_{raceFree}$  that are impossible to cause races (Line 4, see §4.2). Having the remaining trace points in  $T_{shared} - T_{raceFree}$ , we further identify the redundancies among them where the accessed addresses exhibit derivable relationships (Line 5, see §4.3). These redundant trace points are not necessary to be instrumented and traced at runtime since their values can be deduced during offline analysis. With the three instruction sets identified, we compute  $T_{trace} \leftarrow T_{shared} - T_{raceFree} - T_{redundant}$  (Line 6), which represents the set of trace



points that actually need to be instrumented and tracked online. The static binary instrumentation module takes  $T_{trace}$  as input and produces the instrumented program  $prog_{re}$  (Line 7, see §4.4).

#### 4.1 Value-Set Analysis for Multithreaded Programs

As mentioned above, in order to identify the set of trace points involving shared variable addresses  $T_{shared}$ , we need to determine from which memory region the operand (register) of an instruction originates, (i.e., stack region, heap region, or global region). We treat the registers stemming from a heap or global variable as shared addresses, which are likely involved in data races. The registers that are only relevant to stack addresses are excluded safely as they cannot cause races. To this end, we need a value-set analysis to identify the relevant region (stack, heap, or global) for each register. Moreover, in the following must race-free analysis (§4.2), we also need the value-set analysis results to determine if two different registers may reference to the same memory location.

Value-set analysis [3] is a static binary analysis technique which over-approximates a set of values each register can take on at each program point. It can be considered as a binary-level alias analysis. Generally, value-set analysis involves two basic terms: abstract location and value set. An abstract location, or a-loc, is a variable-like entity, which can represent a register or an address in global, stack, and heap regions. For instance, for the instruction `mov 0x4,%eax`, both global address `0x4` and the register `%eax` correspond to an a-loc. A value-set represents a set of a-locs, and it is usually divided into three separate sets: stack, heap, and global. The value set of an a-loc is a collection of addresses and registers that can be accessed by referencing that a-loc. For the instruction `mov 0x4,%eax`, the value set of `%eax` would be  $\langle global \mapsto \{0x4\}, stack \mapsto \{\}, heap \mapsto \{\} \rangle$ , meaning that the register `%eax` holds the value `0x4` from the global memory region after the instruction executes.

Although multiple value-set analyses have been proposed [3, 8, 17], none of them supports multithreaded programs well. In other words, existing value-set analysis cannot ensure the soundness for multithreaded programs. To be specific, the existing value-set analysis maintains a value-set for a register at each instruction of a single thread. It does not consider the effects of shared (heap or global) accesses by multiple threads. Therefore, for a given register of an instruction, its value-set may not be a safe over-approximation of the actual values at runtime. As a consequence, we may erroneously exclude certain trace-points (registers) which are actually related to shared accesses, leading to loss of detection capability. To guarantee soundness of value-set analysis, one way is to extend the control-flow graph by adding all the possible edges because of interleavings. Unfortunately, such approach could cause the control-flow graph to be amplified dramatically considering the large number of interleavings, thus leading to poor analysis scalability.

In this paper, we devise a dedicated value-set analysis ensuring both soundness and scalability. Similar to the existing analysis, we record the value set of a register at each instruction along the traditional control-flow graph. However, for global and heap locations, we maintain a shared summary across the entire program. In this way, the value-set of shared accesses can be guaranteed to be an over-approximation of the actual values. thus ensuring soundness. Moreover, the analysis can scale well to large programs, since the analysis is performed along the original control-flow graph and the value-set relevant to global and heap locations are flow-insensitive.

Algorithm algorithm 2 shows the algorithm in details. At the beginning, `localValueSet` and `sharedValueSet` are initialized as empty (Lines 1-2), which represent the value-set results for stack and shared (global and heap) locations, respectively. The worklist algorithm processes each instruction, starting from the entry instruction of the program (Line 5). The transfer function is performed to update `localValueSet[i]` and `sharedValueSet` (Line 7), followed by propagating the value-set to all the successors (Line 8), until all `localValueSet` entries remain unchanged. Since the value-set updates in the transfer function are performed using *union* operation, the algorithm is guaranteed to converge and terminate. In the transfer function, we update `localValueSet` and `sharedValueSet`

**Algorithm 2:** Value-set analysis for multithreaded programs

**Data:**  $localValueSet[i][x]$ , the value set of a-loc  $x$  for instruction  $i$  where  $x$  is a register or stack a-loc;  $sharedValueSet[y]$ , the value set of a-loc  $y$  which is a global or heap a-loc

```

1   $localValueSet \leftarrow \emptyset$ 
2   $sharedValueSet \leftarrow \emptyset$ 
3   $worklist \leftarrow$  put all the entry instructions into worklist
4  while  $worklist \neq \emptyset$  do
5       $i \leftarrow worklist.pop()$ 
6       $oldLocalValueSet[i] \leftarrow localValueSet[i]$ 
7       $transfer(i, localValueSet, sharedValueSet)$ 
8       $propagate(i, succs(i))$ 
9      if  $oldLocalValueSet[i] \neq localValueSet[i]$  then
10          $worklist.push(succs(i))$ 

11 Function  $transfer(i, localValueSet, sharedValueSet)$ 
12     if  $i$  is mov then
13          $src\_alocs \leftarrow getAliasAlocs(i, src\_op, localValueSet, sharedValueSet)$ 
14          $dst\_alocs \leftarrow getAliasAlocs(i, dst\_op, localValueSet, sharedValueSet)$ 
15         foreach  $dst\_a-loc \in dst\_alocs$  do
16             if  $dst\_a-loc$  is register or stack a-loc then
17                  $localValueSet[i][dst\_a-loc].union(src\_alocs)$ 
18             else
19                 /* $dst\_a-loc$  is global or heap.*/
20                  $sharedValueSet[dst\_a-loc].union(src\_alocs)$ 
21     else if ... then
22         ... /*The propagation of other instructions is omitted here.*/

```

according to the specific semantics of each type of instruction. Here, we only give the transfer logic of *mov* instruction due to space limit. Generally, it first retrieve the value-sets (*i.e.*,  $localValueSet$  and  $sharedValueSet$ ) to acquire the corresponding alias a-locs of the source and destination operands of instruction  $i$ . Then, based on the type of destination a-loc, it decides to update  $localValueSet$  or  $sharedValueSet$  accordingly.

Having the value-set results, we can obtain the set of trace points (*i.e.*, registers)  $T_{shared}$ , which are relevant to shared (heap or global) memory regions. In particular, given a register of an instruction, if its value-set contains any heap or global a-locs, we include it into  $T_{shared}$ . Otherwise, if the value-set of a register only contains stack addresses, we can safely exclude it from tracing. Specifically, for an instruction  $i = mov\ 0x18(\%eax), \%edx$ , we check  $localValueSet[i][\%eax]$ . If it contains a-locs belonging to global or heap regions, then  $\%eax$  could represent the address of a shared variable. We thus put  $\%eax@i$  into  $T_{shared}$ . Otherwise,  $\%eax@i$  can be safely eliminated.

#### 4.2 Must Race-free Analysis

In this section, we would like to employ static analysis to further prune away the trace points irrelevant to data races. As is well known, data races occur only if three conditions are satisfied: 1)



two memory accesses target the same address; 2) they are accessed concurrently; 3) at least one of them is a write operation. Based on this, we propose a static must race-free analysis. In brief, for each memory trace-point in  $T_{shared}$ , we statically check if at least one of the above three conditions must be violated. If so, we consider it as race-free access. We can thus safely avoid it from tracing.

The core logic of the algorithm is shown in Algorithm 3. We take the set of trace-points  $T_{shared}$  involving global/heap memory as input and iterate over all pairs of registers  $x, y \in T_{shared}$ . At Line 6, we check if any one of the three conditions are violated. A trace point  $x$  is considered to belong to  $T_{raceFree}$  if and only if it does not form a data race with any possible point  $y$ . The rest trace points of  $T_{shared}$  would be treated as  $T_{mayRace}$ .

In the following, we elaborate how to check the three conditions. *notWrite* can be checked easily by simply determining whether  $x$  or  $y$  is a *write* operation. The determination of *notAlias*( $x, y$ ) is done on the basis of value-set analysis (§4.1). The result of value-set analysis provides a set of potential values for a specific register at each program point, which essentially identifies the possible may-alias relationships between the register and different a-locs (i.e., other registers and addresses). As long as the intersection of the value-sets for  $x$  and  $y$  is empty, it can be concluded that  $x$  and  $y$  access different addresses, i.e., *notAlias*( $x, y$ ) returns true. To determine whether the concurrent access condition is met, the core logic of the *notConcurrent* function is shown in Lines 16-22. Basically, it follows the logic of lockset algorithm. If  $x$  or  $y$  is an allocated heap object intra-procedurally and does not escape, we can reach that  $x$  and  $y$  cannot be executed concurrently. The detailed logic for determining *isOwned* is shown as Lines 23-34. More importantly, it also considers the accesses within critical sections enclosed by locks and unlocks. The *getLockSet* function analyzes and returns the lock variables involved. If the intersection of *getLockSet*( $x$ ) and *getLockSet*( $y$ ) is non-empty, then  $x$  and  $y$  are protected by the same lock. Thus, true is returned.

### 4.3 Redundant Register Elimination

After the trace-point elimination discussed above, the set  $T_{shared} - T_{raceFree}$  may still have potential for further reduction. The rationale is that the addresses accessed in two instructions may have a static relationship. In other words, the value of a register in one instruction can be statically derived from that of another in other instruction. In such cases, we only need to record the deriving register. The values of subsequent registers can be deduced offline.

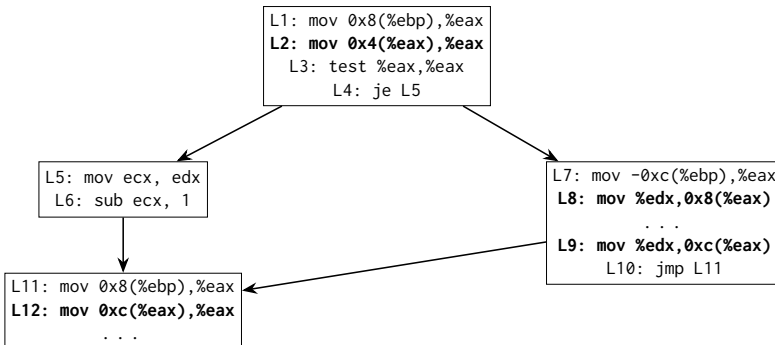


Fig. 3. A toy example for redundant register elimination

For example, Figure 3 has four basic blocks, and the memory instructions that need to be tracked are highlighted in bold (i.e., L2, L8, L9, and L12). For L8 and L9, there are two memory accesses,  $0x8(\%eax)$ @L8 and  $0xc(\%eax)$ @L9. Naively, we need to instrument and trace the value of *eax* at

**Algorithm 3:** Must Race-free Analysis**Input:** a set of shared trace-points  $T_{shared}$ **Output:** a set of trace-points that must be irrelevant to races  $T_{raceFree}$ 


---

```

1   $T_{raceFree} \leftarrow \emptyset, T_{mayRace} \leftarrow \emptyset$ 
2  foreach  $x \in T_{shared}$  do
3       $flag_{raceFree} \leftarrow true$ 
4      if  $x \notin T_{mayRace}$  then
5          foreach  $y \in T_{shared}$  do
6              if  $notAlias(x, y)$  or  $notConcurrent(x, y)$  or  $notWrite(x, y)$  then
7                  continue
8              else
9                   $flag_{raceFree} \leftarrow false$ 
10                  $T_{mayRace}.push(y)$ 
11                 break
12         if  $flag_{raceFree}$  then
13              $T_{raceFree}.push(x)$ 
14         else
15              $T_{mayRace}.push(x)$ 
16 Function  $notConcurrent(x, y)$ 
17     if  $isOwned(x)$  or  $isOwned(y)$  then
18         /*isOwned means a heap object is allocated intra-procedurally and does not escape.*/
19         return true
20     if  $getLockSet(x) \cap getLockSet(y) \neq \emptyset$  then
21         return true
22     return false
23 Function  $IsOwned(x)$ 
24      $i \leftarrow$  the instruction of  $x$ 
25      $alocs \leftarrow localValueSet[i][x]$ 
26     foreach  $aloc \in alocs$  do
27         if  $aloc$  is heap and is allocated within the function of  $x$  then
28             foreach  $arg\_aloc$ : the argument of callsites within the function of  $x$  do
29                 if  $aloc \in localValueSet[i][arg\_aloc]$  then
30                     return false /*escape to other function, not intra-procedural*/
31             if  $aloc \in sharedValueSet$  then
32                 return false /*escape to memory*/
33     return false
34 return true

```

---

both L8 and L9. In fact, *eax* at both L8 and L9 has the same value, which equals to  $-0xc(\%ebp)$  where *ebp* is a local address. Thus, it is sufficient to record the value of *%eax* only at L8, rather than both.

Beyond the elimination within a basic block, we also consider the situations across basic blocks. For instance, the *eax* at *L2* and *L12* are also identical no matter which branch is taken, which equals to  $0x8(\%ebp)@L1$ . Therefore, only one *eax* needs to be traced.

Technically, given a memory access register, we perform an intra-procedural backward symbolic propagation from it until the entry of function. If the symbolic expressions of two registers are identical or have statically fixed relation, then we only keep one as the trace-point. The value of another will be statically deduced.

#### 4.4 Static Binary Instrumentation

Based on §4.1 to §4.3, we identified a minimal set of trace points to be recorded at runtime. Here, we exploit hardware tracing module to achieve low runtime overhead. In particular, we do this by inserting a PTWRITE instruction with the operand being the specific register. Taking Figure 2 as an example, for the instruction  $\text{<mov } 0x4(\%ebx), \%eax\text{>}$  at location *L2*, the value to be recorded is *ebx@L2*, so a  $\text{<ptwrite } \%ebx\text{>}$  instruction is inserted right before *L2*. Each register may be instrumented multiple times at different locations. Therefore, we need to distinguish which instruction each PTWRITE packet corresponds to. To this end, during instrumentation, we manually maintain the mapping between each PTWRITE instruction and the instruction to be traced. This allows us to further determine the exact register corresponding to the PTWRITE packet. For redundant register elimination in §4.3, we maintain the arithmetic relationship between two registers. In the decoding phase, the eliminated memory accesses are reconstructed based on the recorded register value.

### 5 OFFLINE TRACE ANALYSIS

#### 5.1 Hardware Trace Decoding

The hardware traces generated by Intel PTWRITE is stored in a compact packet format. Before analyzing them, we need to firstly decode these packets into the memory read/write events and thread synchronization events required by the race detection algorithm on a per-thread basis.

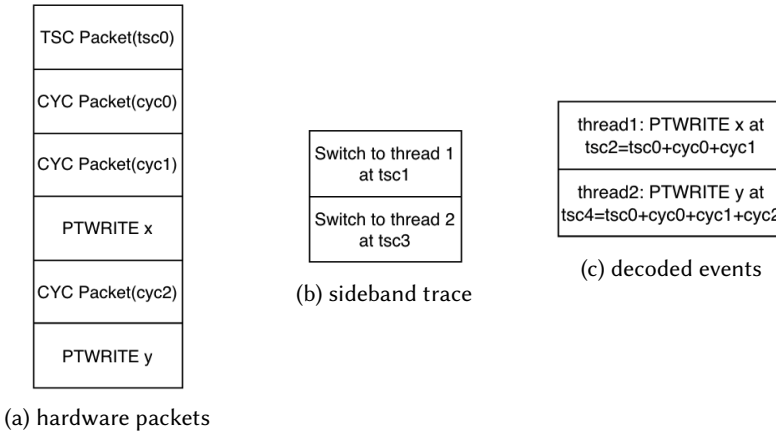


Fig. 4. A trace example

The trace data is stored separately for each CPU. We first consider the case of a single CPU. For simplicity, we represent the PTWRITE trace packets as shown in Figure 4a, which includes TSC, CYC, and PTWRITE packets. The TSC packet contains a specific timestamp. Similarly, each CYC packet indicates the number of clock cycles elapsed since the previous CYC or TSC packet. Based on

these TSC and CYC packets, we can compute the timestamp for each PTWRITE packet. For example, the timestamp of  $\langle PTWRITE\ x \rangle$  in Figure 4a can be calculated as  $tsc2 = tsc0 + cyc0 + cyc1$ , while the timestamp of  $\langle PTWRITE\ y \rangle$  is  $tsc4 = tsc0 + cyc0 + cyc1 + cyc2$ . Meanwhile, the runtime data also includes a sideband trace obtained using perf events, which contains thread switch information with timestamps, shown as Figure 4b. Suppose  $tsc1 < tsc2 < tsc3 < tsc4$ , thus  $\langle PTWRITE\ x \rangle$  is executed after  $tsc1$  but before  $tsc3$ . Therefore, we can derive that  $\langle PTWRITE\ x \rangle$  belongs to thread 1. Similarly,  $\langle PTWRITE\ y \rangle$  can be determined to belong to thread 2. As such, we obtain the final memory access events as shown in Figure 4c. For multiple CPU cores, the event sequence from each CPU can be integrated according to their timestamps, resulting in a complete sequence of accesses.

At this stage, we essentially know the thread ID (tid) and the timestamp corresponding to each PTWRITE. Combined with the mapping information between each PTWRITE and the original instruction, we can determine the specific type of memory event (e.g., read, write, lock, or unlock). This ultimately allows us to reconstruct a trace similar to that shown in Figure 2d.

## 5.2 FastTrack-based Offline Detector

With the per-thread memory access events and thread synchronization events decoded in the previous stage, we can fully leverage dynamic data race detection algorithms for offline race detection. For the sake of fair comparison, HardRace adopts the combination of happens-before and lockset algorithms which are utilized by the state-of-the-art dynamic race detectors [1, 10, 24, 30]. In brief, the offline detector reads and processes memory access events and thread synchronization events in the recorded order. It simulates dynamic data race detection by using information such as the recorded thread ID (TID), memory access addresses, lock variable addresses, and event types. Note that again the contribution of this paper is to propose a low-overhead and high-precision data access monitoring approach for multithreaded programs via hardware tracing, which is orthogonal to the race detection algorithm. The race detection algorithms can benefit from our lightweight monitoring; and HardRace can also adopt any other detection algorithms in the offline analysis.

## 6 IMPLEMENTATION

In the static selective instrumentation module, we use Capstone [2] and Angr [26] to construct an inter-procedural control flow graph (i.e., ICFG), and then perform our inter-procedural value set analysis and must race-free analysis over it. The instrumentation part is implemented based on Dyninst [4], a well-known binary instrumentation framework. The runtime trace collection module mainly utilizes the `perf_event_open` function to configure the CPU buffer, and controls the relevant registers to enable Intel PTWRITE and IP filtering. By default, we allocate a 128MB memory buffer to maintain the hardware traces for each CPU core. At the offline trace analysis part, the decoder is further developed based on the source code of the libipt library [9] provided by Intel. We directly reuse the implementation of FastTrack-based detector in ProRace [30] as our detection module.

## 7 EVALUATIONS

In this section, we evaluate HardRace over a comprehensive set of benchmarks to answer the following research questions.

- How well does HardRace perform in terms of runtime overhead, static analysis scalability, instrumentation cost, and offline analysis efficiency? And how about comparison with the state-of-the-arts with respect to runtime overhead? (§7.2)
- What about the detection capability of HardRace compared with the state-of-the-arts? (§7.3)

- How effective is our static selective instrumentation mechanism in reducing the runtime overhead and data loss? (§7.4)

Table 1. The performance of HardRace in terms of the time of static analysis (§4.1, §4.2 and §4.3), static binary instrumentation (§4.4), offline decoding (§5.1), and offline race detection (§5.2). #Inst and #Func represent the number of instructions and functions in the program, respectively.

| subject           | #Inst   | #Func | static analysis | instrument | decoding | race detection |
|-------------------|---------|-------|-----------------|------------|----------|----------------|
| streamcluster     | 4278    | 95    | 2.1s            | 2.3s       | 192.3s   | 255.8s         |
| x264              | 178647  | 1173  | 122.2s          | 10.8s      | 70.9s    | 73.0s          |
| vips              | 815185  | 7333  | 462.3s          | 59.3s      | 32.2s    | 20.8s          |
| bodytrack         | 107186  | 3417  | 32.7s           | 3.2s       | 43.7s    | 47.1s          |
| fluidanimate      | 7220    | 108   | 2.8s            | 2.5s       | 117.6s   | 122.0s         |
| ocean_cp          | 24601   | 60    | 12.5s           | 2.7s       | 0.2s     | 0.0s           |
| ocean_ncp         | 15376   | 50    | 6.9s            | 2.6s       | 0.2s     | 0.0s           |
| raytrace          | 19947   | 191   | 7.1s            | 2.6s       | 109.4s   | 107.4s         |
| water_nsquared    | 7754    | 61    | 3.3s            | 2.4s       | 0.7s     | 0.6s           |
| water_spatial     | 8152    | 62    | 3.3s            | 2.4s       | 0.2s     | 0.0s           |
| radix             | 3314    | 48    | 2.2s            | 2.4s       | 0.8s     | 0.5s           |
| lu_ncb            | 3098    | 56    | 1.9s            | 2.3s       | 0.2s     | 0.0s           |
| lu_cb             | 3725    | 58    | 2.1s            | 2.4s       | 0.1s     | 0.0s           |
| barnes            | 8385    | 107   | 3.5s            | 2.4s       | 114.7s   | 131.9s         |
| fft               | 3908    | 57    | 2.2s            | 2.5s       | 0.1s     | 0.0s           |
| (Arithmetic Mean) | 80718.4 | 858.4 | 44.5s           | 6.9s       | 45.6s    | 50.6s          |

## 7.1 Experimental Setup

**Benchmarks.** We primarily choose two representative sets of subject programs as our evaluation benchmarks. At first, to understand the performance of HardRace, we take PARSEC/SPLASH-2x benchmark suite (version 3.0beta-20150206), which is commonly used to measure the performance in related works, such as ProRace [30] and Kard [1]. The first three columns in Table 1 list the detailed characteristics about each subject program in terms of subject name, the number of instructions, the number of functions.

Moreover, to compare with the state-of-the-arts with respect to detection capability, we borrow the benchmark from ProRace [30], which is a set of real-world applications with known data races [29]. We follow ProRace’s measurement and test 11 buggy program versions, each with one different race triggered, including *apache*, *mysql*, *cherokee*, *pbzip* and *aget*. Note that ProRace originally tested 12 program versions. However, since *pfscan* is not publicly available in the repository, we can only test 11 of them. Table 3 shows all the buggy versions and their manifest information.

**Comparison Tools.** We select Kard [1], ProRace [30] and naive hardware tracing approach as the comparison approaches. Kard leverages Intel MPK to allocate keys for inter-thread memory access protection, thus achieving low runtime overhead. ProRace, on the other hand, samples memory accesses using hardware PMU (in particular, Intel’s Precise Event Based Sampling). We select these two as comparison since they are the most recent work targeting low-overhead dynamic data race detection. The naive hardware tracing approach is a variant of HardRace by disabling the static selective instrumentation module. In other words, we naively treat all the registers involving shared

memory accesses as the potential trace-points, and instrument PTWRITE instructions to record all of them at runtime.

**Environments.** We conduct all the experiments on a workstation with an Intel Core i9-14900K processor with 32 logical cores supporting Intel PTWRITE. The workstation has 64 GB memory and is equipped with a solid-state drive (SSD). It runs Ubuntu 22.04 LTS with a kernel version of 5.15.

## 7.2 Overall Performance

We run the subjects five times and report the average time for each metric. Table 1 gives the detailed performance with respect to the time of static analysis (*i.e.*, value-set analysis in §4.1, must race-free analysis in §4.2 and redundant register elimination in §4.3), static binary instrumentation (§4.4), offline trace decoding (§5.1), and offline race detection (§5.2).

As can be seen, the series of static selection analysis are efficient enough, which can be done with a couple of seconds for most subjects. The average time cost is only about 14 seconds. This is reasonable since we treat the shared value-set results in our analysis as flow-insensitive so as to achieve soundness and efficiency, which is elaborated in §4.1. Moreover, the static binary instrumentation is also efficient enough as its complexity is linear to the program size. For the offline part, both trace decoding and race detection can be easily finished with seconds on average. All of these validate that HardRace is efficient enough, and can scale well in practice.

Table 2. The runtime overhead (TO) of HardRace, compared with Kard, ProRace and Naive hardware tracing approach. The three columns under ProRace indicate the overhead with different sampling rate.  $S_{inst}$  and  $D_{inst}$  represent the number of instructions to be instrumented statically and to be traced dynamically, respectively.

| Bench Name        | HardRace |            |            | Kard  | ProRace (TO) |        |         | Naive |            |            |
|-------------------|----------|------------|------------|-------|--------------|--------|---------|-------|------------|------------|
|                   | TO       | $S_{inst}$ | $D_{inst}$ |       | 1/100        | 1/1000 | 1/10000 | TO    | $S_{inst}$ | $D_{inst}$ |
| streamcluster     | 0.2%     | 125        | 173.7M     | 0.3%  | >30%         | >5%    | >5%     | 8.6%  | 487        | 464.2M     |
| x264              | 3.5%     | 5994       | 52.1M      | 3.0%  | -            | -      | -       | 65.2% | 20665      | 319.7M     |
| vips              | 0.1%     | 29426      | 23.8M      | 1.3%  | -            | -      | -       | 34.7% | 99703      | 70.4M      |
| bodytrack         | 1.1%     | 234        | 37.0M      | 10.4% | >350%        | >5%    | >1%     | 36.3% | 5855       | 0.0        |
| fluidanimate      | 9.6%     | 73         | 96.5M      | 61.9% | >600%        | >90%   | >5%     | 24.3% | 620        | 104.1M     |
| ocean_cp          | 4.9%     | 132        | 6.7K       | -5.9% | -            | -      | -       | 10.4% | 4266       | 151.2M     |
| ocean_ncp         | 4.0%     | 70         | 6.5K       | 0.0%  | -            | -      | -       | 57.8% | 2247       | 271.5M     |
| raytrace          | 4.3%     | 444        | 89.5M      | 3.7%  | >290%        | >30%   | >1%     | 30.7% | 1726       | 57.3M      |
| water_nsquared    | 0.3%     | 118        | 320.7K     | 18.0% | -            | -      | -       | 4.1%  | 685        | 255.9M     |
| water_spatial     | 0.2%     | 123        | 21.5K      | 5.6%  | -            | -      | -       | 3.7%  | 799        | 92.4M      |
| radix             | -0.0%    | 129        | 356.9K     | -1.0% | -            | -      | -       | 7.7%  | 531        | 141.1M     |
| lu_ncb            | -8.1%    | 91         | 6.5K       | -5.2% | -            | -      | -       | 5.3%  | 393        | 27.8M      |
| lu_cb             | -2.1%    | 94         | 6.5K       | -4.7% | -            | -      | -       | 9.3%  | 531        | 33.9M      |
| barnes            | 3.6%     | 229        | 96.0M      | 34.1% | -            | -      | -       | -5.7% | 827        | 192.2M     |
| fft               | 2.5%     | 115        | 201.0      | 1.0%  | -            | -      | -       | 4.6%  | 556        | 21.1M      |
| (Arithmetic Mean) | 1.6%     | 2493.1     | 38.0M      | 8.2%  | >317.5%      | >32.5% | >3.0%   | 19.8% | 9326.1     | 146.9M     |

In addition, we also compare the runtime overhead of HardRace with the state-of-the-arts, Kard [1], ProRace [30], and the naive hardware tracing approach. Table 2 shows the detailed results. ProRace is a sampling-based approach, whose overhead relies on the exact sampling rate. As such, we measure the overheads of ProRace under three different sampling rates (*i.e.*, 1/100, 1/1000,



1/10000). To calculate the overhead, we run the baseline subjects and the instrumented subjects each five times and compute the average. The overhead is computed as the execution time of instrumented program divided by the execution time of original program minus 100%. Note that due to nondeterministic execution of the subjects and the tiny execution time, it is possible that the overhead is negative. To understand further why the overheads differ greatly between HardRace and naive tracing, we calculate the number of PTWRITE instructions to be instrumented statically and to be traced dynamically, denoted as  $S_{inst}$  and  $D_{inst}$  in Table 2, respectively. Furthermore, the subjects evaluated by ProRace and Kard are not totally identical. We chose to align them with Kard since it is more recent than ProRace. The symbol “-” of Table 2 indicates that the subject is not evaluated and no data is available for ProRace.

We can observe that HardRace achieves a negligible overhead of 1.6% on average. Kard suffers from an overhead of 8.2%. But importantly, it only supports a limited types of races. We will discuss them shortly in §7.3. For ProRace, the overheads under different sampling rates vary significantly. It reaches beyond 300% with 1/100 sampling rate. It also has around 3% overhead under 1/10000. Again as is well known, the low sampling rate usually corresponds to low detection capability (see Table 3). In contrast, HardRace performs selective tracing, which achieves low overhead while not sacrificing detection capability. Additionally, the average overhead of naive hardware tracing approach is 19.8%. Even worse, an immense amount of data loss happens – nearly 40% of traces are lost shown as Table 4. We also compared the number of static instrumentation points ( $S_{inst}$ ) and dynamically decoded instrumentation points ( $D_{inst}$ ) between HardRace and the naive method. To understand the overhead differences between HardRace and Naive hardware tracing, the data of  $S_{inst}$  and  $D_{inst}$  provides clues. The numbers of PTWRITE instructions to be instrumented statically (*i.e.*,  $S_{inst}$ ) and to be executed dynamically (*i.e.*,  $D_{inst}$ ) by HardRace are about 2400+ and 38M, which are dramatically smaller than that of naive hardware tracing (*i.e.*, 9300+ and 146M), respectively. This also indicates that our static selective instrumentation module significant prunes away unnecessary trace points, thus reducing the overhead.

Table 3. Detection probability for each approach.

|                     | Bug manifestation   | Type    | ProRace/% |        |         | HardRace/% | Kard/% |
|---------------------|---------------------|---------|-----------|--------|---------|------------|--------|
|                     |                     |         | 1/100     | 1/1000 | 1/10000 |            |        |
| apache-21287        | double free         | non-ILU | 50        | 3      | 0       | 100        | 0      |
| apache-25520        | corrupted log       | non-ILU | 57        | 52     | 15      | 100        | 0      |
| apache-45605        | assertion           | non-ILU | 60        | 11     | 1       | 100        | 0      |
| mysql-3596          | crash               | ILU     | 5         | 1      | 0       | 100        | 100    |
| mysql-644           | crash               | ILU     | 21        | 6      | 1       | 100        | 100    |
| mysql-791           | missing output      | ILU     | 59        | 2      | 0       | 100        | 100    |
| cherokee-0.9.2      | corrupted log       | non-ILU | 63        | 29     | 8       | 100        | 0      |
| cherokee-bug1       | corrupted log       | non-ILU | 57        | 19     | 5       | 100        | 0      |
| pbzip2-0.9.4-crash  | crash               | ILU     | 0         | 0      | 0       | 100        | 100    |
| pbzip2-0.9.4-benign | -                   | ILU     | 100       | 100    | 100     | 100        | 100    |
| aget-bug2           | wrong record in log | ILU     | 100       | 100    | 100     | 100        | 100    |
| (Arithmetic Mean)   |                     |         | 52.0      | 29.4   | 20.9    | 100.0      | 54.5   |

### 7.3 Detection Capability

In this section, we measure the detection capability of HardRace, and compare it with the state-of-the-arts. Each subject presented in Table 3 is reported to contain a hard-to-trigger data race.

And the reporters provided corresponding patches to control thread interleaving, enabling the data race to be triggered within a short period of time. We evaluate the detection capability in terms of detection probability which is introduced by ProRace [30]. To be specific, we run each buggy program 100 times and to count how many runs each detection tool can report the race. A probability of 50% indicates that, among 100 runs, the data race is detected in 50 of them. A probability of 100% means that the tool can detect the race every time, indicating that there are no false negatives for that subject.

Table 3 lists the detection probability data of various tools. HardRace is able to detect the bugs in all the runs without any false negatives. This is because HardRace only prunes away unnecessary memory accesses in a safe manner. Technically, the memory accesses that can trigger data races should all be recorded in hardware traces, allowing for detection when a data race occurs. In contrast, ProRace shows average detection probabilities of 52.0%, 29.4%, and 20.9% for sampling rates 1/100, 1/1K, and 1/10K, respectively. This means that even with very dense sampling, the detection probability only reaches about a half, while the overhead is prohibitively high (as mentioned earlier in Table 2). For Kard, due to the lack of relevant experimental data and the absence of source code, a direct comparison can hardly be conducted. However, based on our understanding of its approach, we can reason if each race can be detected by Kard. Specifically, Kard states that it can only detect Inconsistent Lock Usage (ILU) races. Upon manual checking of the data race types of the subjects, we find that five of the eleven programs exhibit non-ILU type of data races. Thus, they cannot be detected by Kard.

Table 4. The data loss times and percentage due to buffer overflow

| subject           | HardRace     |            | Naive Hardware Tracing |            |
|-------------------|--------------|------------|------------------------|------------|
|                   | loss percent | loss times | loss percent           | loss times |
| streamcluster     | 0.0%         | 0          | 8.7%                   | 7          |
| x264              | 0.0%         | 0          | 84.0%                  | 317        |
| vips              | 0.0%         | 0          | 33.5%                  | 5          |
| bodytrack         | 0.0%         | 0          | 60.6%                  | 36         |
| fluidanimate      | 0.0%         | 0          | 22.0%                  | 14         |
| ocean_cp          | 0.0%         | 0          | 0.0%                   | 0          |
| ocean_ncp         | 0.0%         | 0          | 76.5%                  | 80         |
| raytrace          | 0.0%         | 0          | 61.3%                  | 10         |
| water_nsquared    | 0.0%         | 0          | 59.6%                  | 16         |
| water_spatial     | 0.0%         | 0          | 0.0%                   | 0          |
| radix             | 0.0%         | 0          | 51.6%                  | 12         |
| lu_ncb            | 0.0%         | 0          | 13.1%                  | 1          |
| lu_cb             | 0.0%         | 0          | 44.7%                  | 1          |
| barnes            | 0.0%         | 0          | 43.2%                  | 17         |
| fft               | 0.0%         | 0          | 0.0%                   | 0          |
| (Arithmetic Mean) | 0.0%         | 0          | 37.3%                  | 34         |

#### 7.4 Effectiveness of Selective Instrumentation

As mentioned before, our static selective instrumentation module plays the crucial role in reducing runtime overhead and hardware data loss. In this section, we would like to validate the significance of

selective instrumentation (§4) empirically. First, regarding overhead reduction, Table 2 provides the experimental data about the runtime overhead, the number of PTWRITE instructions instrumented statically and traced dynamically by HardRace and naive hardware tracing. Apparently, the overhead with selective instrumentation enabled (*i.e.*, HardRace) is much smaller than that of naive hardware tracing. Second, as for data loss, we measure the times of loss happened and the total percentage of data loss by HardRace and naive hardware tracing. As shown in Table 4, HardRace incurs no data loss for all subject programs under stress testing. In contrast, the naive hardware tracing has an average 37.3% of data loss, with an average of 34 times of data loss events per subject. It is noteworthy that while the naive approach performs well on certain subjects such as *streamcluster*, *ocean\_cp*, and *water\_spatial*, it still experiences significant data loss in the majority of cases. For *x264* and *ocean\_ncp*, it successfully collects less than 30% of the data. The reason is that for the subjects like *x264* and *ocean\_ncp*, intensive memory accesses occur frequently in the program, leading to severe hardware data loss. All in all, we conclude that the selective analysis in HardRace is highly effective in reducing data loss, thus ensuring detection capability.

## 8 RELATED WORK

Over the past years, various approaches for data race detection have been proposed, including static, dynamic, and hybrid approaches.

**Dynamic Race Detection.** The lockset algorithm, introduced by Eraser [23], checks whether shared memory is consistently protected by locks to detect data races. Although prone to false positives, its efficiency makes it a reference point for many subsequent studies. For example, the well-known ThreadSanitizer (TSAN) [24] employs a combination of happens-before and lockset algorithms to balance detection accuracy and performance. Another prominent dynamic tool, FastTrack [10], optimizes the original happens-before approach for better performance. Other dynamic detectors like ProRace [30], Kard[1], and TxRace [31], integrate hardware-based techniques into their detection processes. ProRace uses Intel Processor Trace (PT) to log the program’s control flow and PEBS to sample memory accesses, then applies the FastTrack algorithm offline to detect races. Kard and TxRace, on the other hand, rely entirely on hardware-based methods. Kard uses Memory Protection Keys (MPK) to ensure that a shared object is accessible by only one thread within a critical section. TxRace utilizes hardware transactional memory (HTM) to detect data races dynamically, treating critical regions as atomic transactions and checking for conflicts. At the same time, some studies optimize the happens-before (HB) relationship at the algorithmic level to mitigate the shortcomings of dynamic data race detection. For instance, [27] introduces the causally-precedes (CP) relationship, which is a subset of the HB relationship, allowing for the observation of more races without sacrificing robustness. Meanwhile, WCP[15] further weakens the CP relationship and enables race detection within linear time.

**Static Race Detection.** Static detectors often exhibit high false positives and low false negatives. For instance, RacerD [5] is classified as a lockset-based detector, allowing it to avoid dealing with the vast number of interleavings typical in static analysis. Conversely, O2 [18], which combines lockset and happens-before analysis, must construct a static happens-before graph (SHB) during static analysis and explore as many interleavings as possible to minimize false negatives.

**Hybrid Race Detection.** Although relatively uncommon, some work has explored the combination of static analysis followed by dynamic execution. For instance, RaceMob [14] integrates both static and dynamic techniques: it first uses the static analysis tool RELAY [28] to detect potential data races, then breaks down these potential race conditions into tasks that are crowdsourced to users, ensuring minimal overhead for individual users.

## 9 CONCLUSION

HardRace is a data race monitor which can on-the-fly detect races happened in production runs. Its core technical contribution lies on a series of sound static analysis which are unitized to prune away unnecessary memory accesses significantly, thus achieving super-low runtime overhead. To the best of our knowledge, HardRace is the first work to leverage Intel PTWRITE for production-run data race detection. The experimental evaluations validate that HardRace can achieve sufficiently low overhead while ensuring good detection capability. It to some extent proves that the holistic design combining static analysis and hardware tracing is promising for multithreaded program monitoring. We are looking forward to more attempts along this direction for multithreaded programs.

## DATA-AVAILABILITY STATEMENT

We would like to provide the artifact later and submit it for Artifact Evaluation. It would contain the source code of HardRace, the benchmarks used, as well as all the experimental results.

## REFERENCES

- [1] Adil Ahmad, Sangho Lee, Pedro Fonseca, and Byoungyoung Lee. 2021. Kard: lightweight data race detection with per-thread memory protection. In *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (Virtual, USA) (ASPLOS '21)*. Association for Computing Machinery, New York, NY, USA, 647–660. <https://doi.org/10.1145/3445814.3446727>
- [2] Quynh Nguyen Anh. 2014. Capstone: Next generation disassembly framework. *Proceedings of the 2014 Black Hat USA, Black Hat USA 14* (2014).
- [3] Gogul Balakrishnan and Thomas Reps. 2010. WYSINWYX: What you see is not what you eXecute. *ACM Trans. Program. Lang. Syst.* 32, 6, Article 23 (aug 2010), 84 pages. <https://doi.org/10.1145/1749608.1749612>
- [4] Andrew R. Bernat and Barton P. Miller. 2011. Anywhere, any-time binary instrumentation. In *Proceedings of the 10th ACM SIGPLAN-SIGSOFT Workshop on Program Analysis for Software Tools (Szeged, Hungary) (PASTE '11)*. Association for Computing Machinery, New York, NY, USA, 9–16. <https://doi.org/10.1145/2024569.2024572>
- [5] Sam Blackshear, Nikos Gorogiannis, Peter W. O'Hearn, and Ilya Sergey. 2018. RacerD: compositional static race detection. *Proc. ACM Program. Lang.* 2, OOPSLA, Article 144 (oct 2018), 28 pages. <https://doi.org/10.1145/3276514>
- [6] Michael D. Bond, Katherine E. Coons, and Kathryn S. McKinley. 2010. PACER: proportional detection of data races. In *Proceedings of the 31st ACM SIGPLAN Conference on Programming Language Design and Implementation (Toronto, Ontario, Canada) (PLDI '10)*. Association for Computing Machinery, New York, NY, USA, 255–268. <https://doi.org/10.1145/1806596.1806626>
- [7] Daming D. Chen, Wen Shih Lim, Mohammad Bakhshalipour, Phillip B. Gibbons, James C. Hoe, and Bryan Parno. 2021. HerQues: securing programs via hardware-enforced message queues. In *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (Virtual, USA) (ASPLOS '21)*. Association for Computing Machinery, New York, NY, USA, 773–788. <https://doi.org/10.1145/3445814.3446736>
- [8] Sanchuan Chen, Zhiqiang Lin, and Yinqian Zhang. 2021. SelectiveTaint: Efficient Data Flow Tracking With Static Binary Rewriting. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 1665–1682. <https://www.usenix.org/conference/usenixsecurity21/presentation/chen-sanchuan>
- [9] Intel Corporation. [n. d.]. *libipt: an Intel(R) Processor Trace decoder library*. <https://github.com/intel/libipt> Accessed: 2024.
- [10] Cormac Flanagan and Stephen N. Freund. 2009. FastTrack: efficient and precise dynamic race detection. In *Proceedings of the 30th ACM SIGPLAN Conference on Programming Language Design and Implementation (Dublin, Ireland) (PLDI '09)*. Association for Computing Machinery, New York, NY, USA, 121–133. <https://doi.org/10.1145/1542476.1542490>
- [11] Anup Holey, Vineeth Mekkat, and Antonia Zhai. 2013. HAccRG: Hardware-Accelerated Data Race Detection in GPUs. In *Proceedings of the 2013 42nd International Conference on Parallel Processing (ICPP '13)*. IEEE Computer Society, USA, 60–69. <https://doi.org/10.1109/ICPP.2013.15>
- [12] Jeff Huang, Charles Zhang, and Julian Dolby. 2013. CLAP: recording local executions to reproduce concurrency failures. In *Proceedings of the 34th ACM SIGPLAN Conference on Programming Language Design and Implementation (Seattle, Washington, USA) (PLDI '13)*. Association for Computing Machinery, New York, NY, USA, 141–152. <https://doi.org/10.1145/2491956.2462167>
- [13] Baris Kasikci, Weidong Cui, Xinyang Ge, and Ben Niu. 2017. Lazy Diagnosis of In-Production Concurrency Bugs. In *Proceedings of the 26th Symposium on Operating Systems Principles (Shanghai, China) (SOSP '17)*. Association for Computing Machinery, New York, NY, USA, 582–598. <https://doi.org/10.1145/3132747.3132767>

- [14] Baris Kasikci, Cristian Zamfir, and George Candea. 2013. RaceMob: crowdsourced data race detection. In *Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles* (Farmington, Pennsylvania) (SOSP '13). Association for Computing Machinery, New York, NY, USA, 406–422. <https://doi.org/10.1145/2517349.2522736>
- [15] Dileep Kini, Umang Mathur, and Mahesh Viswanathan. 2017. Dynamic race prediction in linear time. In *Proceedings of the 38th ACM SIGPLAN Conference on Programming Language Design and Implementation* (Barcelona, Spain) (PLDI 2017). Association for Computing Machinery, New York, NY, USA, 157–170. <https://doi.org/10.1145/3062341.3062374>
- [16] Leslie Lamport. 1978. Time, clocks, and the ordering of events in a distributed system. *Commun. ACM* 21, 7 (July 1978), 558–565. <https://doi.org/10.1145/359545.359563>
- [17] Jian Lin, Liehui Jiang, Yisen Wang, and Weiyu Dong. 2019. A Value Set Analysis Refinement Approach Based on Conditional Merging and Lazy Constraint Solving. *IEEE Access* 7 (2019), 114593–114606. <https://doi.org/10.1109/ACCESS.2019.2936139>
- [18] Bozhen Liu, Peiming Liu, Yanze Li, Chia-Che Tsai, Dilma Da Silva, and Jeff Huang. 2021. When threads meet events: efficient and precise static race detection with origins. In *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation* (Virtual, Canada) (PLDI 2021). Association for Computing Machinery, New York, NY, USA, 725–739. <https://doi.org/10.1145/3453483.3454073>
- [19] Shan Lu, Soyeon Park, Eunsoo Seo, and Yuanyuan Zhou. 2008. Learning from mistakes: a comprehensive study on real world concurrency bug characteristics. *SIGOPS Oper. Syst. Rev.* 42, 2 (mar 2008), 329–339. <https://doi.org/10.1145/1353535.1346323>
- [20] Shan Lu, Soyeon Park, Eunsoo Seo, and Yuanyuan Zhou. 2008. Learning from mistakes: a comprehensive study on real world concurrency bug characteristics. In *Proceedings of the 13th International Conference on Architectural Support for Programming Languages and Operating Systems* (Seattle, WA, USA) (ASPLOS XIII). Association for Computing Machinery, New York, NY, USA, 329–339. <https://doi.org/10.1145/1346281.1346323>
- [21] Madanlal Musuvathi, Shaz Qadeer, Thomas Ball, Gerard Basler, Piramanayagam Arumuga Nainar, and Iulian Neamtii. 2008. Finding and reproducing Heisenbugs in concurrent programs. In *Proceedings of the 8th USENIX Conference on Operating Systems Design and Implementation* (San Diego, California) (OSDI'08). USENIX Association, USA, 267–280.
- [22] Robert O'Callahan and Jong-Deok Choi. 2003. Hybrid dynamic data race detection. In *Proceedings of the Ninth ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming* (San Diego, California, USA) (PPoPP '03). Association for Computing Machinery, New York, NY, USA, 167–178. <https://doi.org/10.1145/781498.781528>
- [23] Stefan Savage, Michael Burrows, Greg Nelson, Patrick Sobalvarro, and Thomas Anderson. 1997. Eraser: a dynamic data race detector for multithreaded programs. *ACM Trans. Comput. Syst.* 15, 4 (nov 1997), 391–411. <https://doi.org/10.1145/265924.265927>
- [24] Konstantin Serebryany and Timur Iskhodzhanov. 2009. ThreadSanitizer: data race detection in practice. In *Proceedings of the Workshop on Binary Instrumentation and Applications* (New York, New York, USA) (WBLA '09). Association for Computing Machinery, New York, NY, USA, 62–71. <https://doi.org/10.1145/1791194.1791203>
- [25] Tianwei Sheng, Neil Vachharajani, Stephane Eranian, Robert Hundt, Wenguang Chen, and Weimin Zheng. 2011. RACEZ: a lightweight and non-invasive race detection tool for production applications. In *2011 33rd International Conference on Software Engineering (ICSE)*. 401–410. <https://doi.org/10.1145/1985793.1985848>
- [26] Yan Shoshitaishvili, Ruoyu Wang, Christopher Salls, Nick Stephens, Mario Polino, Andrew Dutcher, John Grosen, Siji Feng, Christophe Hauser, Christopher Kruegel, and Giovanni Vigna. 2016. SOK: (State of) The Art of War: Offensive Techniques in Binary Analysis. In *2016 IEEE Symposium on Security and Privacy (SP)*. 138–157. <https://doi.org/10.1109/SP.2016.17>
- [27] Yannis Smaragdakis, Jacob Evans, Caitlin Sadowski, Jaeheon Yi, and Cormac Flanagan. 2012. Sound predictive race detection in polynomial time. In *Proceedings of the 39th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages* (Philadelphia, PA, USA) (POPL '12). Association for Computing Machinery, New York, NY, USA, 387–400. <https://doi.org/10.1145/2103656.2103702>
- [28] Jan Wen Voun, Ranjit Jhala, and Sorin Lerner. 2007. RELAY: static race detection on millions of lines of code. In *Proceedings of the 6th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on The Foundations of Software Engineering* (Dubrovnik, Croatia) (ESEC-FSE '07). Association for Computing Machinery, New York, NY, USA, 205–214. <https://doi.org/10.1145/1287624.1287654>
- [29] Jie Yu and Satish Narayanasamy. 2009. A case for an interleaving constrained shared-memory multi-processor. In *Proceedings of the 36th Annual International Symposium on Computer Architecture* (Austin, TX, USA) (ISCA '09). Association for Computing Machinery, New York, NY, USA, 325–336. <https://doi.org/10.1145/1555754.1555796>
- [30] Tong Zhang, Changhee Jung, and Dongyoon Lee. 2017. ProRace: Practical Data Race Detection for Production Use. In *Proceedings of the Twenty-Second International Conference on Architectural Support for Programming Languages and Operating Systems* (Xi'an, China) (ASPLOS '17). Association for Computing Machinery, New York, NY, USA, 149–162. <https://doi.org/10.1145/3037697.3037708>

Conference acronym 'XX, Xudong Sun, Zhuo Chen, Jingyang Shi, Yiyu Zhang, Peng Di, Xuandong Li, and Zhiqiang Zuo

- [31] Tong Zhang, Dongyoon Lee, and Changhee Jung. 2016. TxRace: Efficient Data Race Detection Using Commodity Hardware Transactional Memory. In *Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems* (Atlanta, Georgia, USA) (*ASPLOS '16*). Association for Computing Machinery, New York, NY, USA, 159–173. <https://doi.org/10.1145/2872362.2872384>
- [32] Zhiqiang Zuo, Kai Ji, Yifei Wang, Wei Tao, Linzhang Wang, Xuandong Li, and Guoqing Harry Xu. 2021. JPortal: precise and efficient control-flow tracing for JVM programs with Intel processor trace. In *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation* (Virtual, Canada) (*PLDI 2021*). Association for Computing Machinery, New York, NY, USA, 1080–1094. <https://doi.org/10.1145/3453483.3454096>