Optimal Prediction of Synchronization-Preserving Races

UMANG MATHUR, University of Illinois at Urbana-Champaign, USA
ANDREAS PAVLOGIANNIS, Aarhus University, Denmark
MAHESH VISWANATHAN, University of Illinois at Urbana-Champaign, USA

Concurrent programs are notoriously hard to write correctly, as scheduling nondeterminism introduces subtle errors that are both hard to detect and to reproduce. The most common concurrency errors are (data) races, which occur when memory-conflicting actions are executed concurrently. Consequently, considerable effort has been made towards developing efficient techniques for race detection. The most common approach is dynamic race prediction: given an observed, race-free trace $\sigma$ of a concurrent program, the task is to decide whether events of $\sigma$ can be correctly reordered to a trace $\sigma^*$ that witnesses a race hidden in $\sigma$.

In this work we introduce the notion of sync(hronization)-preserving races. A sync-preserving race occurs in $\sigma$ when there is a witness $\sigma^*$ in which synchronization operations (e.g., acquisition and release of locks) appear in the same order as in $\sigma$. This is a broad definition that strictly subsumes the famous notion of happens-before races. Our main results are as follows. First, we develop a sound and complete algorithm for predicting sync-preserving races. For moderate values of parameters like the number of threads, the algorithm runs in $\tilde{O}(N)$ time and space, where $N$ is the length of the trace $\sigma$. Second, we show that the problem has a $\Omega(N/\log^2 N)$ space lower bound, and thus our algorithm is essentially time and space optimal. Third, we show that predicting races with even just a single reversal of two sync operations is NP-complete and even W[1]-hard when parameterized by the number of threads. Thus, sync-preservation characterizes exactly the tractability boundary of race prediction, and our algorithm is nearly optimal for the tractable side. Our experiments show that our algorithm is fast in practice, while sync-preservation characterizes races often missed by state-of-the-art methods.

CCS Concepts: • Software and its engineering → Software verification and validation; • Theory of computation → Theory and algorithms for application domains; Program analysis.

Additional Key Words and Phrases: concurrency, dynamic analysis, race detection, complexity

ACM Reference Format:

1 INTRODUCTION

The verification of concurrent programs is one of the main challenges in formal methods. Concurrency adds a dimension of non-determinism to program behavior which stems from inter-process communication. Accounting for such non-determinism during program development is a challenging mental task, making concurrent programming significantly error-prone. At the same time, bugs...
due to concurrency are very hard to reproduce manually, and automated techniques for doing so are crucial in enhancing the productivity of software developers.

Data races are the most common form of concurrency errors. A data race (sometimes just called a race) occurs when a thread of a multi-threaded program accesses a shared memory location while another thread is modifying it without proper synchronization. The presence of a data race is often symptomatic of a serious bug in the program [Lu et al. 2008]; races have caused data corruption and compilation errors [Boehm 2011; Kasikci et al. 2013; Narayanasamy et al. 2007], and significant system errors [Boehm 2012; Zhivich and Cunningham 2009] in the past. Therefore, considerable research has focused on detecting and preventing races in multi-threaded programs.

One of the most popular approaches to race prediction is via dynamic analysis [Bond et al. 2010; Flanagan and Freund 2009; Pozniansky and Schuster 2003]. Unlike static analysis, dynamic race prediction is performed at runtime. Such techniques determine if an observed execution provides evidence for the existence of a possibly alternate program execution that can concurrently perform conflicting data accesses\(^1\). The underlying principle is that a race is present but "hidden" in a large number of different program executions; hence techniques that uncover such hidden races can accelerate the process of debugging concurrent programs significantly. The popularity of dynamic race prediction techniques further stems (i) from their scalability to large production software, and (ii) from their ability to produce only sound error reports.

The most popular dynamic race prediction techniques are based on Lamport’s happens-before partial order [Lamport 1978]. These techniques scan the input trace, determine happens-before orderings on-the-fly, and report a race on a pair of conflicting data accesses if they are unordered by happens-before. This approach is sound, in that the presence of unordered conflicting data accesses ensures the existence of an execution with a race. While happens-before based analysis fails to predict races in various cases [Smaragdakis et al. 2012], its wide deployment is based on the fact that the algorithm is fast, single pass, and runs in linear time. The principle that forms the basis of its efficiency is the following. When reasoning about alternate executions, happens-before analysis does not consider any execution in which the order of synchronization primitives is reversed from that in the observed execution. We call such alternate executions sync(hronization)-preserving executions. Other, more powerful race prediction techniques [Genç et al. 2019; Huang et al. 2014; Huang and Rajagopalan 2016; Pavlogiannis 2019; Roemer et al. 2018; Smaragdakis et al. 2012] sacrifice this principle and consider alternate executions that are not sync-preserving. Naturally, this typically results in performance degradation, as the problem is in general NP-hard [Mathur et al. 2020a], and considerable efforts are made towards improving the scalability of such techniques [Roemer and Bond 2019; Roemer et al. 2020].

Although happens-before only detects races whose exposure preserves the ordering of synchronization primitives, it can still miss simple races that adhere to this pattern. For example, consider the trace \(\sigma_1\) shown in Figure 1a. Let us name the events of this trace based on the order in which they appear in the trace; thus, \(e_i\) denotes the \(i\)th event of the trace. Here, the partial order happens-before orders the first \(w(x)\) (event \(e_1\)) and the last \(w(x)\) (event \(e_6\)), and therefore, does not detect any race in this execution. However events \(e_1\) and \(e_6\) are in race. This can be exposed by the alternate execution shown in Figure 1b, which is obtained by dropping the critical section of lock \(\ell\) performed by thread \(t_1\). Notice that the order of synchronization events (namely, \(\text{acq}(\ell)\) and \(\text{rel}(\ell)\) events) that appear in the trace of Figure 1b, are in the same order as in the trace of Figure 1a, and hence this is a sync-preserving execution. Thus, the notion of sync-preservation captures races beyond standard happens-before races.

\(^1\)Conflicting data accesses come from different threads, access a common memory location, and at least one is a write.
The trace $\sigma_2$ and is exposed by a witness that omits the critical section on lock $\ell$. On the other hand, sync-preservation is not bound to such limitations: only between successive events $e_1$ and $e_6$, both conflicting on variable $x$. Notice, however, that the intermediate event $e_3$ also accesses $x$, but is not in race with either $e_1$ or $e_6$. Partial-order methods for race prediction are limited to capturing races only between successive conflicting accesses\(^2\). Hence, distant races that are interjected with intermediate conflicting but non-racy events, are missed by such methods. On the other hand, sync-preservation is not bound to such limitations: $(e_1, e_6)$ is characterized as a race under this criterion, regardless of the intermediate, non-racy $e_3$, and is exposed by a witness that omits the critical section on lock $\ell$ in the thread $t_1$.

Our Contributions. Motivated by the above observations, we make the following contributions.

1. We introduce the novel notion of sync(hronization)-preserving data races. This is a sound notion of predictable races, and it strictly subsumes the standard notion of happens-before races. Moreover, it characterizes races between events that can be arbitrarily far apart in the input trace, as opposed to happens-before and other partial-order methods that only characterize races between successive conflicting accesses. Our notion is applicable to all concurrency settings, and interestingly, it is also complete for systems with synchronization-deterministic concurrency \([\text{Aguado et al. 2018; Bocchino et al. 2009; Cui et al. 2015; Zhao et al. 2019}]\).

2. We develop an efficient, single-pass, nearly linear time algorithm SyncP that, given a trace $\sigma$, detects whether $\sigma$ contains a sync-preserving race. In fact, our algorithm soundly reports all events $e_2$ which are in a sync-preserving race with an event $e_1$ that appears earlier in $\sigma$. Given $N$ events in $\sigma$, our algorithm spends $\tilde{O}(N)$ time, where $\tilde{O}$ hides factors poly-logarithmic in $N$, when other parameters of the input (e.g., number of threads) are $\tilde{O}(1)$.

3. Although our algorithm performs a single pass of the trace, in the worst case, it might use space that is nearly linear in the length of the trace, i.e., $\tilde{O}(N)$ space. Hence follows a natural question: is there an efficient algorithm for sync-preserving race prediction that uses considerably less space? We answer this question in negative, by showing that any single-pass algorithm for detecting even a single sync-preserving race must use nearly linear space. Hence, our algorithm SyncP has nearly optimal performance in both time and space.

4. We next study the complexity of race prediction with respect to the number of synchronization reversals that might occur when constructing a witness that exposes the race. In the case

\(^2\)When the earlier access is a read instead of a write, this statement is true per thread.
of synchronization via locks, this number corresponds to the number of critical sections whose order is reversed in the witness trace. We prove that the problem of predicting races which can be witnessed by a single reversal (of two critical sections) is NP-complete and even W[1]-hard when parameterized by the number of threads. Thus, sync-preservation characterizes exactly the tractability boundary of race prediction, and our algorithm is nearly optimal for the tractable side. Moreover, our result shows that any level of synchronization suffices to make the problem of race prediction as hard as in the general case.

(5) Finally, we have implemented our race prediction algorithm SyncP and evaluated its performance on standard benchmarks. Our results show that sync-preservation characterizes many races that are missed by state-of-the-art methods, and SyncP detects them efficiently.

2 PRELIMINARIES

In this section we establish notation useful throughout the paper. The exposition follows other related works in the literature.

2.1 Background on Dynamic Data Race Prediction

Traces and Events. Our objective is to develop a dynamic analysis technique which works over execution traces, or simply traces of concurrent programs. We work with the sequential consistency memory model. In this setting, traces are sequences of events. We will use \( \sigma, \sigma’, \ldots, \sigma_1, \sigma_2, \ldots \) to denote traces. Every event of \( \sigma \) can be represented as a tuple \( e = (i, t, op) \), where \( i \) is a unique identifier of \( e \) in \( \sigma \), \( t \) is the thread that performs \( e \) and \( op \) is the operation performed in the event \( e \). We often omit the unique identifier of such a tuple and simply write \( e = (t, op) \). We use \( \text{thr}(e) \) and \( op(e) \) to denote the thread performing \( e \) and the operation performed by \( e \). An operation can be one of read from or write to a shared memory location or variable \( x \), denoted \( r(x) \) and \( w(x) \), and acquisition or release of a lock \( \ell \), denoted \( \text{acq}(\ell) \) or \( \text{rel}(\ell) \). Forks and joins can be naturally handled, but we avoid introducing them here for notational convenience. We denote by \( \text{Events}_\sigma \) the set of events in a trace \( \sigma \). We use \( \text{Thr}_{\sigma}, \text{Vars}_\sigma \) and \( \text{Locks}_\sigma \) to denote respectively the threads, variables and locks that appear in \( \sigma \). Likewise, we use \( \text{Acquires}_\sigma(\ell) \) and \( \text{Releases}_\sigma(\ell) \) to denote the set of acquire and release events of \( \sigma \) on lock \( \ell \in \text{Locks}_\sigma \).

We require that traces obey lock semantics. In particular, every lock \( \ell \) is released by a thread \( t \) only if there is an earlier matching acquire event by the same thread \( t \), and that each such lock is held by at most one thread at a time. Formally, let \( \sigma|_{\ell} \) denote the projection of \( \sigma \) to the set of events \( \text{Acquires}_\sigma(\ell) \cup \text{Releases}_\sigma(\ell) \). We require that for every lock \( \ell \), the sequence \( \sigma|_{\ell} \) is a prefix of some sequence that belongs to the language of the regular expression \( \{ \sum_{t \in \text{Thr}_{\sigma}} (t, \text{acq}(\ell)) \cdot (t, \text{rel}(\ell)) \}^* \).

For an acquire event \( e \), we use \( \text{match}_{\sigma}(e) \) to denote the matching release event of \( e \) if one exists (and \( \bot \) otherwise). Similarly, for a release event \( e \), \( \text{match}_{\sigma}(e) \) is the matching acquire of \( e \) on the same lock. For an acquire event \( e \), the critical section protected by \( e \), denoted \( \text{CS}_{\sigma}(e) \), is the set of events \( e’ \) such that \( \text{thr}(e') = \text{thr}(e) \) and \( e' \) occurs after \( e \) and before the matching release \( \text{match}_{\sigma}(e) \) (if it exists) in \( \sigma \). For a release event \( e \), we have \( \text{CS}_{\sigma}(e) = \text{CS}_{\sigma}(\text{match}_{\sigma}(e)) \).

Orders on Traces. A partial order \( \leq_{\sigma} \) defined over a trace \( \sigma \) is a reflexive, anti-symmetric and transitive binary relation on \( \text{Events}_\sigma \), the symbol \( P \) is an optional identifier for the partial order. We write \( e_1 \leq_{\sigma} e_2 \) to denote \((e_1, e_2) \in \leq_{\sigma} \), where \( e_1, e_2 \in \text{Events}_\sigma \). For a partial order \( \leq_{\sigma} P \), we use \( <_{\sigma} P \) to denote the strict order \( \leq_{\sigma} \setminus \\{(e, e) \mid e \in \text{Events}_\sigma \} \). We write \( e_1 <_{\sigma} P e_2 \) to denote that \((e_1, e_2) \notin \leq_{\sigma} \). Events \( e_1, e_2 \in \text{Events}_\sigma \) are said to be unordered by \( \leq_{\sigma} P \), denoted \( e_1 \parallel_{\sigma} e_2 \) if \( e_1 <_{\sigma} P e_2 \) and \( e_2 <_{\sigma} P e_1 \); otherwise, we write \( e_1 \parallel_{\sigma} e_2 \) denoting that \( e_1 \) and \( e_2 \) are ordered by \( \leq_{\sigma} P \) in one or the other way. When \( \sigma \) is clear from context, we will use \( \leq_{P}, <_{P}, \parallel_{P} \) and \( \parallel_{P} \) instead of respectively \( \leq_{\sigma}, <_{\sigma}, \parallel_{\sigma}, \parallel_{\sigma} \).
Events \( \sigma \) is, let scheduling. The notion of data race captured in alternate executions of the same program that arise due to slightly different thread place. Such a notion is captured by a \( \sigma \) that ensures that data races can be detected agnostic of the program that generated \( \sigma \) reorderings of an observed trace, thereby detecting data races beyond those in just the execution of data race detection, this means that a dynamic analysis that looks for executions with enabled \( x \) \( \sigma \) in \( \sigma \) that

\[
\sigma_1 \leq \sigma_2 \quad \text{iff the event } e_1 \text{ occurs before } e_2 \text{ in } \sigma. 
\]

The thread-order \( \leq_{\text{TO}} \) defined by \( \sigma \) is the partial order on Events\( _{\sigma} \) that orders events in the same thread: for two events \( e_1, e_2 \in \text{Events}_{\sigma} \), \( e_1 \leq_{\text{TO}} e_2 \) if \( e_1 \leq_{\sigma} e_2 \) and \( \text{thr}(e_1) = \text{thr}(e_2) \).

Conflicting Events and Data Races. Let \( \sigma \) be a trace. Two events \( e_1, e_2 \in \text{Events}_{\sigma} \) are said to be conflicting, denoted \( e_1 \equiv e_2 \), if \( \text{thr}(e_1) \neq \text{thr}(e_2) \), and there is a common variable \( x \in \text{Vars}_{\sigma} \) such that \( \text{op}(e_1), \text{op}(e_2) \in \{ t(x), w(x) \} \) and at least one of \( \text{op}(e_1) \) and \( \text{op}(e_2) \) is \( w(x) \). Let \( \rho \) be a trace with \( \text{Events}_{\rho} \subseteq \text{Events}_{\sigma} \). An event \( e \in \text{Events}_{\rho} \) is said to be \( \sigma \)-enabled in \( \rho \) if \( e \notin \text{Events}_{\rho} \) and for all events \( e' \in \text{Events}_{\sigma} \) such that \( e' \prec_{\text{TO}} e \), we have \( e' \in \text{Events}_{\rho} \). A pair of conflicting events \( (e_1, e_2) \) in \( \sigma \) is said to be a data race of \( \sigma \) if \( \sigma \) has a prefix \( \sigma' \) such that both \( e_1 \) and \( e_2 \) are \( \sigma \)-enabled in \( \sigma' \). The trace \( \sigma \) is said to have a data race if there is a pair of conflicting events \( (e_1, e_2) \) in \( \sigma \) that constitutes a data race of \( \sigma \).

Example 1. Consider the trace \( \sigma_3 \) in Figure 2a. The set of events of \( \sigma_3 \) is \( \text{Events}_{\sigma_3} = \{ e_1, e_2, \ldots, e_6 \} \), \( \text{Thr}_{\sigma_3} = \{ t_1, t_2 \} \), \( \text{Vars}_{\sigma_3} = \{ x \} \) and \( \text{Locks}_{\sigma_3} = \{ \} \). For the event \( e_1 = (t_1, \text{acq}(t)) \), we have \( \text{thr}(e_1) = t_1 \) and \( \text{op}(e_1) = \text{acq}(t) \). The trace order of this trace is \( \leq_{\text{tr}} = \{ (e_i, e_j) \mid i \leq j \} \) and the thread-order is \( \leq_{\text{TO}} = \{ (e_1, e_2), (e_1, e_5), (e_2, e_3), (e_3, e_4), (e_3, e_6), (e_4, e_6) \} \). Events \( e_5 \) and \( e_6 \) conflict because they access the same variable \( x \) and are performed by different threads. For the prefix trace \( \sigma'_3 = e_1 \cdot e_2 \cdot e_3 \cdot e_4 \), both \( e_5 \) and \( e_6 \) are \( \sigma_3 \)-enabled in \( \sigma'_3 \). Thus, \( (e_5, e_6) \) constitutes a data race of \( \sigma_3 \).

Correct Reorderings. Execution traces of concurrent programs are sensitive to thread scheduling, and looking for a trace with a specific pattern is like searching for a needle in a haystack. In terms of data race detection, this means that a dynamic analysis that looks for executions with enabled conflicting events (data races) is likely to miss many data races that might have otherwise been captured in alternate executions of the same program that arise due to slightly different thread scheduling. The notion of data race prediction attempts to alleviate this problem by capturing a more robust notion of data races. The idea here is to infer data races that might occur in alternate reorderings of an observed trace, thereby detecting data races beyond those in just the execution that was observed. The set of allowable reorderings of an observed trace \( \sigma \) is defined in a manner that ensures that data races can be detected agnostic of the program that generated \( \sigma \) in the first place. Such a notion is captured by a correct reordering which we define next.

For a trace \( \sigma \) and a read event \( e \), we use \( \text{lw}_\sigma(e) \) to denote the write event observed by \( e \). That is, \( e' = \text{lw}_\sigma(e) \) is the last (according to the trace order \( \leq_{\text{tr}} \)) write event \( e' \) of \( \sigma \) such that \( e \) and \( e' \) access the same variable and \( e' \leq_{\sigma} e \); if no such \( e' \) exists, then we write \( \text{lw}_\sigma(e) = \perp \).

![Fig. 2. Traces, data races and predictable data races](image-url)
Given the above notation, a trace $\rho$ is said to be a correct reordering of trace $\sigma$ if

(a) $\text{Events}_\rho \subseteq \text{Events}_\sigma$
(b) $\text{Events}_\rho$ is downward closed with respect to $\leq^{\sigma}_{\text{TO}}$ and further $\leq^\rho_{\text{TO}} \subseteq \leq^{\sigma}_{\text{TO}}$
(c) for every read event $e \in \text{Events}_\rho$, $\text{lw}_\rho(e) = \text{lw}_\sigma(e)$.

The above definition ensures that if $\rho$ is a correct reordering of $\sigma$, then every program that generates the execution trace $\sigma$ also generates $\rho$. This is because $\rho$ preserves both intra-thread ordering, as well as the values read by every read occurring in $\rho$, thereby preserving any control flow that might have been taken by $\sigma$. This style of formalizing alternative executions based on semantics of concurrent objects was popularized by [Herlihy and Wing 1990] and by prior race detection works [Said et al. 2011; Şerbănuţă et al. 2012]. Our definition of correct reordering has been derived from [Smaragdakis et al. 2012], which has subsequently also been used in the literature [Genç et al. 2019; Kini et al. 2017; Mathur et al. 2018, 2020a; Pavlogiannis 2019; Roemer et al. 2018].

**Data Race Prediction.** Armed with the notion of correct reorderings, we can now define a more robust notion of data races. A pair of conflicting events $(e_1, e_2)$ in $\sigma$ is said to be a *predictable* data race of $\sigma$ if there is a correct reordering $\rho$ of $\sigma$ such that $e_1, e_2$ are $\sigma$-enabled in $\rho$. We remark that a pair of conflicting events $(e_1, e_2)$ in trace $\sigma$ may not be a data race of $\sigma$, but nevertheless may still be a *predictable* data race of $\sigma$.

**Example 2.** Consider the trace $\sigma_4$ in Figure 2b. Observe that there is no prefix of $\sigma_4$ in which both $e_1$ and $e_6$ are enabled. However, $(e_1, e_6)$ is a predictable race of $\sigma_4$ that is witnessed by the singleton correct reordering $\sigma_4^{\text{CR}} = e_5$ in which both $e_1$ and $e_6$ are enabled; $\sigma_4^{\text{CR}}$ is both downward closed with respect to, and respects $\leq^{\sigma_4}_{\text{TO}}$. Further, it has no read events and thus vacuously every read observes the same last write as in $\sigma_4$. The other pair of conflicting events in $\sigma_4$, namely $(e_3, e_6)$, however, is not a predictable race. These events are protected by a common lock, and there is no correct reordering in which $e_3$ and $e_6$ are simultaneously enabled — any attempt at doing so will lead to overlapping critical sections on $\ell$, thereby violating lock semantics.

**Example 3.** Now, consider $\sigma_5$ in Figure 2c. Here, the conflicting pair $(e_3, e_5)$ cannot be a predictable race as in the case of $\sigma_4$— the lock $\ell$ protects both $e_3$ and $e_6$. Now consider the other conflicting pair $(e_1, e_5)$. Let $\rho$ be a correct reordering of $\sigma_5$ in which $e_8$ is enabled. We must have $e_6 \in \text{Events}_\rho$ ($\rho$ must be $\leq^{\sigma_5}_{\text{TO}}$-downward closed) and further $e_3 \in \text{Events}_\rho$ (as $e_3 = \text{lw}_{\sigma_5}(e_6) = \text{lw}_\rho(e_6)$). Clearly, $e_1$ cannot be enabled in any such trace $\rho$, and thus, the trace $\sigma_5$ has no predictable data race.

The central theme of race prediction is to solve the problem below.

**Problem 1 (Data Race Prediction).** Given a trace $\sigma$, determine if $\sigma$ has a predictable data race.

**A Note on Soundness.** We say that an algorithm for data race prediction is *sound* if whenever the algorithm reports a YES answer, then the given trace has a predictable data race. Likewise, an algorithm is complete if the algorithm reports YES whenever the input trace has a data race. Our convention for this nomenclature ensures that no false positives are reported by a *sound* algorithm [Sergey 2019] and is consistent with prior work on data race prediction [Genç et al. 2019; Kini et al. 2017; Pavlogiannis 2019; Roemer et al. 2018; Smaragdakis et al. 2012]. Soundness is often a desirable property for dynamic race predictors for widespread adoption [Gorogiannis et al. 2019].

**2.2 Synchronization-Preserving Data Races**

In general, the problem of data race prediction is intractable [Mathur et al. 2020a], and a sound and complete algorithm for data race prediction is unlikely to scale beyond programs of even moderate size. A recent trend in predictive analysis for race detection instead, aims to develop
techniques that are sound but incomplete, with successively better prediction power (ability to report more data races) than previous techniques [Genc et al. 2019; Kini et al. 2017; Pavlogiannis 2019; Roemer et al. 2018; Smaragdakis et al. 2012]. Most of these techniques are either based on partial orders [Kini et al. 2017; Pozniansky and Schuster 2003; Smaragdakis et al. 2012] or use graph-based algorithms [Pavlogiannis 2019; Roemer et al. 2018]. In this paper, we characterize a class of predictable data races, called \textit{sync(hronization)-preserving} races, which we define shortly. We will later (Section 4) present an algorithm that reports a race iff the input trace has a sync-preserving race. Since sync-preserving races are predictable races, our algorithm will be sound for race prediction.

\textbf{Sync-Preserving Correct Reordering.} A correct reordering of a trace is called \textit{sync(hronization)-preserving} if it does not reverse the order of synchronization constructs; in our formalism, traces use locks as synchronization primitives to enforce mutual exclusion. Formally, a correct reordering $\rho$ of a given trace $\sigma$ is \textit{sync-preserving} with respect to $\sigma$ if for every lock $\ell$ and for any two acquire events $e_1, e_2 \in \text{Acquires}_\ell(\ell)$, we have $e_1 \leq_{\rho} e_2$ iff $e_1 \leq_\sigma e_2$. In other words, the order of two critical sections on the same lock is the same in $\sigma$ and $\rho$. Let us illustrate this notion on an example.

\textbf{Example 4.} Consider trace $\sigma_6$ in Figure 3a. This trace has 3 critical sections on lock $\ell$. Now consider the correct reordering $\sigma_6^{\text{CR}}$ (Figure 3b) of $\sigma_6$. Here, the critical section in thread $t_1$ is not present. But, nevertheless, the order amongst the remaining critical sections on $\ell$ (in threads $t_2$ and $t_3$) is the same as in $\sigma_6$, making $\sigma_6^{\text{CR}}$ a sync-preserving correct reordering of $\sigma_6$. This example also demonstrates that the order of read and write events may be different in a trace and its sync-preserving correct reordering (as in Figure 3).

A pair of conflicting events $(e_1, e_2)$ of a trace $\sigma$ is said to be a \textit{sync(hronization)-preserving race} of $\sigma$ if there is a sync-preserving correct reordering $\rho$ of $\sigma$ in which $e_1$ and $e_2$ are $\sigma$-enabled.

\textbf{Example 5.} Let us again consider traces from Figure 3. Events $e_1$ and $e_8$ in $\sigma_6$ (Figure 3a) correspond respectively to events $e_6$ and $e_7$ in $\sigma_6^{\text{CR}}$ (Figure 3b). These two events are $\sigma_6$-enabled in the prefix $\rho = e_1.e_2.e_3.e_4.e_5$ of $\sigma_6^{\text{CR}}$. As a result, $(e_1, e_8)$ is a sync-preserving race of $\sigma_6$. Likewise, $(e_1, e_4)$ is also a sync-preserving race of $\sigma_6$ witnessed by the singleton sync-preserving correct reordering $\rho' = (t_2, \text{acq}(\ell))$, in which both $e_1$ and $e_4$ are enabled.

In this paper we present a linear time algorithm for the following decision problem, giving a \textit{sound} algorithm for Problem 1.

\textbf{Problem 2 (Sync-Preserving Race Prediction).} Given trace $\sigma$, determine if there is a pair of conflicting events $(e_1, e_2)$ in $\sigma$ such that $(e_1, e_2)$ is a sync-preserving data race of $\sigma$. 

Comparison with Other Approaches. Here we briefly compare sync-preserving races with other approaches in the literature for sound dynamic race prediction. Races reported using the famous happens-before (HB) partial order [Pozniansky and Schuster 2003], and its extension to schedulable-happens-before (SHB) [Mathur et al. 2018] are strictly subsumed by this notion. That is, these techniques only compute sync-preserving races, but can also miss simple cases of sync-preservation, as already illustrated in the examples of Figure 1. The causally precedes (CP) partial order [Smaragdakis et al. 2012], and its extension to the weak causally precedes (WCP) partial order [Kini et al. 2017] are capable of predicting races that reverse critical sections. However, they are closed under composition with HB, and as such can miss even simple sync-preserving races, even on two-threaded traces. The doesn’t commute (DC) partial order [Roemer et al. 2018] is an unsound weakening to WCP, that further undergoes a vindication phase to filter out unsound reports. Nevertheless, DC is somewhat similar to WCP and also misses sync-preserving races. The recently introduced partial order strong-dependently-precedes (SDP) [Genç et al. 2019], while claimed to be sound in that paper, is, in fact, unsound. In our technical report [Mathur et al. 2020b], we show a counter-example to the soundness theorem of SDP, which we confirmed with the authors [Genç et al. 2020]. The partial order WDP [Genç et al. 2019] is a (unsound) weakening of DC, and can nevertheless miss sync-preserving races in the vindication phase employed for ruling out false positives. We further refer to [Mathur et al. 2020b] for a few examples that illustrate the above comparison.

3 SUMMARY OF MAIN RESULTS

Here we give an outline of the main results of this paper. In later sections we present the details, i.e., algorithms, proofs and examples. Due to limited space, some technical proofs are relegated to our companion technical report [Mathur et al. 2020b]. Our first result is an algorithm for dynamic prediction of sync-preserving races. We show the following theorem.

Theorem 3.1. Sync-preserving race prediction is solvable in \( O(N \cdot T^2 + A \cdot V \cdot T^3) \) time and \( O(N + T^3 \cdot V \cdot L) \) space, for a trace \( \sigma \) with length \( N \), \( T \) threads, \( A \) acquires, and \( V \) variables.

In many settings the number of events \( N \) and number of acquires \( A \) are the dominating parameters, whereas the other parameters are much smaller, i.e., \( T, V = \tilde{O}(1) \), where \( \tilde{O} \) hides poly-logarithmic factors. Hence, the complexity of our algorithm becomes \( \tilde{O}(N) \) for both time and space. Our next result shows that a linear space complexity is essentially unavoidable when predicting sync-preserving races with one-pass streaming algorithms.

Theorem 3.2. Any one-pass algorithm for sync-preserving race prediction on traces with \( \geq 2 \) threads, \( N \) events and \( \Omega(\log N) \) locks uses \( \Omega(N/\log^2 N) \) space.

Clearly, any algorithm must spend linear time, while Theorem 3.2 shows that the algorithm must also use (nearly) linear space. As our algorithm uses \( \tilde{O}(N) \) time and space, it is optimal for both resources, modulo poly-logarithmic improvements. Our next theorem shows a combined time-space lower bound for the problem, which highlights that reducing the space usage must lead to an increased running time, given that the algorithm is executed on the Turing Machine model.

Theorem 3.3. Consider the problem of sync-preserving race prediction on traces with \( \geq 2 \) threads, \( N \) events and \( \Omega(\log N) \) locks. Consider any Turing Machine algorithm for the problem with time and space complexity \( T(N) \) and \( S(N) \), respectively. Then we have \( T(N) \cdot S(N) = \Omega(N^2/\log^2 N) \).

Finally, we study the complexity of general race prediction as a function of the number of reversals of synchronization operations. Given our positive result in Theorem 3.1, can we relax
our restriction of sync-preservation while retaining a tractable definition of predictable races? Our next theorem answers this question in negative.

**Theorem 3.4.** Dynamic race prediction on traces with a single lock and two critical sections is W[1]-hard parameterized by the number of threads.

Note that W[1]-hardness implies NP-hardness. The theorem has two important implications.

1. Any witness of predictable races in the setting of Theorem 3.4 is either a sync-preserving reordering, or reverses the order of a single pair of acquire events. Thus, Theorem 3.1 and Theorem 3.4 establish a **tight dichotomy** on the tractability of the problem, based on the number of synchronization reversals: the problem is as hard as in the general case for just 1 reversal, while it is efficiently solvable for no reversals.

2. The general problem of dynamic race prediction was shown to be W[1]-hard in [Mathur et al. 2020a]. However, that proof requires traces with Ω(N) critical sections, and hence it applies to traces that essentially comprise of synchronization events entirely. In contrast, the class of traces in Theorem 3.4 has the smallest level of synchronization possible, i.e., just a single lock and two critical sections on that lock. Hence, Theorem 3.4 shows that any amount of lock-based synchronization suffices to make the problem as hard as in the general case.

Together, Theorem 3.1, Theorem 3.2 and Theorem 3.4 characterize exactly the tractability horizon of race prediction, and show that our algorithm is *time and space optimal* for the tractable side.

### 4 DETECTING SYNCHRONIZATION-PRESERVING RACES

In this section, we discuss our algorithm SyncP for detecting sync-preserving data races. The complete algorithm is presented in Section 4.4. The algorithm may appear complex at first glance and to make the exposition simple, we first present a high-level overview of the algorithm in Section 4.1. In the overview, we highlight important observations and algorithmic insights for solving smaller subproblems of the main problem of sync-preserving race prediction. Section 4.2 and Section 4.3 present details of the algorithms for the smaller subproblems, and pave the way for the final algorithm in Section 4.4. In Section 4.5, we present a matching space lowerbound result for detecting sync-preserving races, thereby showing the optimality of our algorithm.

#### 4.1 Insights and Overview of the Algorithm

Our algorithm, SyncP, relies on several important observations that are crucial for detecting sync-preserving races in linear time. In order to present these observations, it is helpful to define intermediate subproblems.

**Problem 3** (Sync-Preserving Race Prediction Given Pair). Given a trace σ and a pair of conflicting events (e₁, e₂) of σ, determine if (e₁, e₂) is a sync-preserving data race of σ.

**Problem 4** (Sync-Preserving Race Prediction Given Event and Thread). Given a trace σ, an event e in σ and a thread t ≠ thr(e), check if there is an event e′ ≤ₜₑ e such that thr(e′) = t and (e′, e) constitutes a sync-preserving race of σ.

Observe that a trace with N events can have O(N²) conflicting pairs of events. Thus, an algorithm for Problem 3 that runs in time O(T(N)) can be used to obtain an algorithm for Problem 4 (resp. Problem 2) that runs in time O(N · T(N)) (resp. O(N² · T(N))) by checking if every other event of the given thread t that conflicts with e is also in race with e (resp. every conflicting pair of events is a race). We will, however, present algorithms for all three problems that run in O(N) time.

**Efficiently Solving Problem** 3. Our first important observation towards a full fledged solution to Problem 3 is that when checking for the existence of a sync-preserving reordering (of a trace σ)
that witnesses a race on a given pair \((e_1, e_2)\), it, suffices to only search for those reorderings \(\rho\) of \(\sigma\) which impose the same order as in \(\sigma\), on all of its events, and not just on the critical sections. We formalize this in Lemma 4.1.

**Lemma 4.1.** If \((e_1, e_2)\) is a sync-preserving race of \(\sigma\), then there is a correct reordering \(\rho\) of \(\sigma\) such that both \(e_1, e_2\) are \(\sigma\)-enabled in \(\rho\) and \(\leq_{\text{tr}} \subseteq \leq_{\text{tr}}^\rho\).

The implication of Lemma 4.1 is the following. When we are searching for a correct reordering \(\rho\) of \(\sigma\) that witnesses a race \((e_1, e_2)\), and if we already have access to the set of events \(S \subseteq \text{Events}_{\sigma}\) of a candidate reordering \(\rho\), a simple check suffices — does \(S\) form a correct reordering of \(\sigma\) when linearized according to \(\leq_{\text{tr}}^\rho\)? In other words, we do not need to enumerate and check all the (exponentially many) permutations of events in \(S\). Thus, Problem 3 — ‘search for a sync-preserving correct reordering \(\rho\)’ — reduces to a simpler problem — ‘search for an appropriate set of events’. Of course, not all sets \(S \subseteq \text{Events}_{\sigma}\) of events can be linearized (according to \(\leq_{\text{tr}}^\rho\)) to obtain a correct reordering of \(\sigma\). At the very least, \(S\) should satisfy some sanity conditions which we outline next.

**Definition 1** (Thread-Order and Last-Write Closure). Let \(\sigma\) be a trace. A set \(S \subseteq \text{Events}_{\sigma}\) is said to be \((\leq_{\text{TO}}^\sigma, \text{lw}_{\sigma})\)-closed if (a) \(S\) is downward-closed with respect to \(\leq_{\text{TO}}\), and (b) for any read event \(r \in \text{Events}_{\sigma}\), if \(r \in S\) and if \(\text{lw}_{\sigma}(r)\) exists, then \(\text{lw}_{\sigma}(r) \in S\).

The \((\leq_{\text{TO}}^\sigma, \text{lw}_{\sigma})\)-closure of a set \(S \subseteq \text{Events}_{\sigma}\), denoted \(\text{TLClosure}_{\sigma}(S)\), is the smallest set \(S' \subseteq \text{Events}_{\sigma}\) such that \(S \subseteq S'\) and \(S'\) is \((\leq_{\text{TO}}^\sigma, \text{lw}_{\sigma})\)-closed.

We remark that any correct reordering \(\rho\) of \(\sigma\) that contains events in the set \(S\) must also contain all the events in \(\text{TLClosure}_{\sigma}(S)\).

**Definition 2** (Sync-Preserving Closure). Let \(\sigma\) be a trace. A set \(S \subseteq \text{Events}_{\sigma}\) is said to be sync-preserving closed if:

(a) \(S\) is \((\leq_{\text{TO}}^\sigma, \text{lw}_{\sigma})\)-closed, and
(b) for any two acquire events \(a_1, a_2 \in \text{Acquires}_{\sigma}(\ell)\) with \(a_1 \leq_{\text{tr}} a_2\), if both \(a_1, a_2 \in S\), then \(\text{match}_{\sigma}(a_1) \in S\).

The sync-preserving closure of a set \(S \subseteq \text{Events}_{\sigma}\), denoted \(\text{SPClosure}_{\sigma}(S)\), is the smallest set \(S' \subseteq \text{Events}_{\sigma}\) such that \(S \subseteq S'\) and \(S'\) is sync-preserving closed.

Intuitively, the set \(S' = \text{SPClosure}_{\sigma}(S)\) captures the additional set of events that must be present in any sync-preserving correct reordering \(\rho\) of \(\sigma\) given that \(\rho\) contains all events in \(S\). First, any correct reordering of \(\sigma\) containing \(S\) will contain \(\text{TLClosure}_{\sigma}(S)\) and thus \(\text{TLClosure}_{\sigma}(S) \subseteq S'\) (Condition a). Second, if a correct reordering \(\rho\) is sync-preserving and contains two acquires \(a_1 \leq_{\text{tr}} a_2\) on the same lock \(\ell\), then we must also have \(a_1 \leq_{\text{tr}} a_2\). Then, in order to ensure well-formedness of \(\rho\), \(\text{CS}_{\sigma}(a_1)\) must also finish entirely before \(a_2\) in \(\rho\), and thus \(\rho\) must contain \(\text{match}_{\sigma}(a_1)\) (Condition b).

For two events \(e_1, e_2 \in \text{Events}_{\sigma}\), we define

\[
\text{SPIdeal}_{\sigma}(e_1, e_2) = \text{SPClosure}_{\sigma}(\{\text{prev}_{\sigma}(e_1)\} \cup \{\text{prev}_{\sigma}(e_2)\}).
\]

Here, we use \(\text{prev}_{\sigma}(e)\) to denote the last event \(e'\) in \(\sigma\) such that \(e' \leq_{\text{TO}} e\); if no such event exists, then \(\text{prev}_{\sigma}(e) = \bot\) (and further we let \(\{\bot\} = \emptyset\)). In essence, \(\text{SPIdeal}_{\sigma}(e_1, e_2)\) contains the necessary set of events that must be present in any sync-preserving correct reordering that witnesses the race \((e_1, e_2)\). We next show that, in fact, it is also a sufficient set of events, given that it is disjoint from \((e_1, e_2)\).

**Lemma 4.2.** \((e_1, e_2)\) is a sync-preserving race of \(\sigma\) iff \((e_1, e_2) \cap \text{SPIdeal}_{\sigma}(e_1, e_2) = \emptyset\).

Lemma 4.2 gives us a straightforward algorithm for Problem 3 — compute \(I = \text{SPIdeal}_{\sigma}(e_1, e_2)\) and check if neither \(e_1, e_2 \not\in I\). In Section 4.2 we show how to compute this in linear time.
Efficiently Solving Problem 4. As noted before, a linear time algorithm for Problem 3 guarantees a quadratic time algorithm for Problem 4. In order to design a more efficient linear time algorithm, we will exploit monotonicity of SPIdeal(· ·), which we formalize next.

**Lemma 4.3.** Let σ be a trace and let \( e_1, e_2, e_1', e_2' \in \text{Events}_\sigma \) such that \( e_1 \leq^\sigma_{\text{TO}} e_1' \) and \( e_2 \leq^\sigma_{\text{TO}} e_2' \). Then, SPIdeal(\( e_1, e_2 \)) \( \subseteq \) SPIdeal(\( e_1', e_2' \)).

Our linear time algorithm for Problem 4 exploits Lemma 4.3 as follows. Suppose we are checking if a given event \( e \) in the given trace \( \sigma \) is in sync-preserving race with some earlier conflicting event of thread \( t \). To accomplish this, we can scan \( \sigma \) and enumerate the list \( L \) of events that belong to \( t \) and conflict with \( e \). When checking for a race with the first such event \( e_1' \), we compute \( I_{\text{first}} = \text{SPIdeal}(e_1', e) \). If a race is found, we are done. Otherwise, we analyze the next event \( e'_{\text{next}} \) in \( L \) and compute \( I_{\text{next}} = \text{SPIdeal}(e'_{\text{next}}, e) \). Here Lemma 4.3 ensures that \( I_{\text{first}} \subseteq I_{\text{next}} \). Our algorithm exploits this observation by computing the latter set \( I_{\text{next}} \) incrementally, spending time that is proportional only to the number of extra events (i.e., events in \( I_{\text{next}} \setminus I_{\text{first}} \)). This principle is applied repeatedly to subsequent events of \( L \), giving us an overall linear time algorithm (Section 4.3).

**Efficiently Solving Problem 2.** A final ingredient in our incremental linear time algorithm for Problem 2 is the following observation which builds on Lemma 4.3.

**Lemma 4.4.** Let \( \sigma \) be a trace and let \( e_1, e_2, e_2' \in \text{Events}_\sigma \) such that \( e_1 \leq^\sigma_\sigma e_2 \leq^\sigma_{\text{TO}} e_2' \), \( e_1 = e_2 \) and \( e_1 = e_2' \). If \((e_1, e_2)\) is not a sync-preserving race, then \((e_1, e_2')\) is also not a sync-preserving race of \( \sigma \).

Intuitively, this observation suggests the following. Suppose that, when looking for a sync-preserving race, the algorithm determines that \( e_2 \) is not in race with any earlier conflicting event. Then, for an event \( e_2' \) (that appears later in the thread of \( e_2 \)), we only need to investigate if \( e_2' \) is in race with conflicting events \( e_1' \) that appear after \( e_2 \). This reduces to checking \((e_1', e_2'\)) instead of \((e_1', e_2)\), where \( e_1' \leq_{\text{TR}} e_2 \).

**High-Level Overview of SyncP.** Equipped with Lemma 4.3 and Lemma 4.4, we now describe our incremental algorithm for Problem 2 that works in linear time. For ease of exposition, let us focus on the question — is there a write-write race on some fixed variable \( x \in \text{Vars} \) when accessed in two fixed threads \( t_1, t_2 \in \text{Vars} \). The algorithm scans the trace in a streaming forward pass and analyzes every event \( e = \langle t_2, w(x) \rangle \), checking if there is an earlier conflicting event \( e' = \langle t_1, w(x) \rangle \leq_{\text{TR}} e \) so that \((e', e)\) is a sync-preserving race. In doing so, it computes \( I = \text{SPIdeal}(e', e) \) in linear time and checks if \( e' \notin I \). If not, \((e', e)\) is not a race and the algorithm checks if there is a different event \( e'_{\text{next}} \leq_{\text{TR}} e \) so that \((e'_{\text{next}}, e)\) is a race. This continues until there are no earlier events remaining that conflict with \( e \). Each time, the ideal computation is performed incrementally, by using the previously computed ideals. After this, the algorithm moves to the next write event \( e_{\text{next}} = \langle t_2, w(x) \rangle \) in \( t_2 \) and checks if it is in race with some earlier event \( e'' \) of thread \( t_1 \), where this time, \( e'' \) appears in the trace after the previously discarded event \( e \) of \( t_2 \). Again, the closed set \( \text{SPIdeal}(e'', e_{\text{next}}) \) is computed incrementally. We show that all this incremental computation can be performed efficiently and present an outline for our algorithm for Problem 2 in Section 4.4.

Next, we present high-level descriptions of the intermediate steps that we outlined above, and discuss important algorithmic insights and data-structures that help achieve efficiency.

### 4.2 Checking if a Given Pair of Conflicting Events is a Sync-Preserving Race

Algorithm 1 outlines our solution to Problem 3 (check if a given pair of events \((e_1, e_2)\) is a sync-preserving race of \( \sigma \)). This algorithm computes the closure \( \text{SPIdeal}(e_1, e_2) \) in an iterative fashion and checks if it contains neither \( e_1 \) nor \( e_2 \) (see Lemma 4.2). We remark that when \( e_1 \leq_{\text{TR}} e_2 \), Definition 2 ensures that \( e_2 \notin \text{SPIdeal}(e_1, e_2) \). Consequently, the check \( e_1 \notin \text{SPIdeal}(e_1, e_2) \)
### Algorithm 1: Checking if a Given Conflicting Pair Constitutes a Sync-Preserving Race

**Input:** Trace $\sigma$, Conflicting events $e_1$ and $e_2$ with $e_1 \preceq^\sigma e_2$

1. function $\text{ComputeSPIdeal}(\sigma, e_1, e_2, I_0)$
2. $I \leftarrow I_0 \cup \text{TLClosure}_\sigma(\{\text{prev}_\sigma(e_1)\}) \cup \text{TLClosure}_\sigma(\{\text{prev}_\sigma(e_2)\})$
3. repeat
4. if $\exists \ell \in \text{Locks}_\sigma, \exists a_1, a_2 \in \text{Acquires}_\sigma(\ell)$ such that $a_1 \preceq^\sigma a_2$ and $a_1, a_2 \in I$ then
5. $I \leftarrow I \cup \text{TLClosure}_\sigma(\{\text{match}_\sigma(a_1)\})$
6. until $I$ does not change
7. return $I$
8. $I \leftarrow \text{ComputeSPIdeal}(\sigma, e_1, e_2, \emptyset)$
9. if $e_1 \notin I$ then
10. declare 'race'

(Line 9 in Algorithm 1) is equivalent to the condition ‘$\{e_1, e_2\} \cap \text{SPIdeal}_\sigma(e_1, e_2) = \emptyset$’ (due to Lemma 4.2). The function $\text{ComputeSPIdeal}$ performs a fixpoint computation, starting from the set $I = \bigcup_{i \in \{1,2\}} \text{TLClosure}_\sigma(\{\text{prev}_\sigma(e_i)\})$ (when $I_0 = \emptyset$). The correctness of the algorithm follows from the correctness of the function $\text{ComputeSPIdeal}$, which we formalize below.

**Lemma 4.5.** Let $\sigma$ be a trace, $e_1, e_2 \in \text{Events}_\sigma$ and $I_0 \subseteq \text{Events}_\sigma$ be a $(\leq^\sigma_{\text{TO}}, \text{lw}_\sigma)$-closed set. Let $I$ be the set returned by $\text{ComputeSPIdeal}(\sigma, e_1, e_2, I_0)$ in Algorithm 1. Then, $I = \text{SPClosure}_\sigma(I_0 \cup \{\text{prev}_\sigma(e_1)\} \cup \{\text{prev}_\sigma(e_2)\})$.

We now discuss the data-structures used to ensure linear time and space for Algorithm 1.

**Vector Timestamps.** Vector timestamps [Fidge 1991; Mattern 1988] are routinely used in distributed computing and also in prior work on race prediction [Flanagan and Freund 2009; Kini et al. 2017; Pozniansky and Schuster 2003; Roemer et al. 2018]. We use vector timestamps to represent sets of events that are $(\leq^\sigma_{\text{TO}}, \text{lw}_\sigma)$-closed; a formal definition is deferred to Section 4.4. In Algorithm 1, the sets $\text{TLClosure}_\sigma(\{\text{prev}_\sigma(e_i)\})$ are $(\leq^\sigma_{\text{TO}}, \text{lw})$-closed (Line 2). Further, the initial value $I_0 = \emptyset$ ensures that all subsequent values of $I$ in $\text{ComputeSPIdeal}$ are $(\leq^\sigma_{\text{TO}}, \text{lw})$-closed. All these sets can be represented as vector timestamps. The advantage of using vector timestamps is two-fold. First, these timestamps provide a succinct representation of sets — instead of representing a set explicitly as a collection of events, a vector timestamp only uses $\mathcal{T}$ integers (where $\mathcal{T} = |\text{Thr}_\sigma|$). Second, the vector timestamps for $(\leq^\sigma_{\text{TO}}, \text{lw}_\sigma)$-closed sets can be computed in a streaming fashion, incrementally, using vector timestamps of smaller subsets.

**Projecting a Trace to Threads and Locks.** Let us consider the check in Line 4. Here, we look for two acquire events $a_1 \preceq^\sigma a_2$ in the current ideal $I$ that acquire the same lock $\ell$. How do we efficiently discover two such acquire events? A straightforward but naive way is to enumerate all pairs of events in $I$ and check if they are acquire events of the above kind. But this can take $O(N^3)$ time, where $N = |\text{Events}_\sigma|$ because the number of such pairs can be $O(N^2)$ in the worst case and the number of times the ideal can change is $O(N)$. Instead, we rely on the following observation:

**Proposition 4.6.** Let $I \subseteq \text{Events}_\sigma$ be downward closed with respect to $\leq^\sigma_{\text{TO}}$. For every $t \in \text{Thr}_\sigma$ and every $\ell \in \text{Locks}_\sigma$, there is at most one acquire event $a$ with $\text{thr}(a) = t$, $\text{op}(a) = \text{acq}(\ell)$ such that $a \in I$ and $\text{match}_\sigma(a) \notin I$. When such an event $a$ exists, then $\text{match}_\sigma(a') \in I$ for every other acquire $a' <^\sigma_{\text{TO}} a$ of the form $a' = \langle t, \text{acq}(\ell) \rangle$. 

The above observation can be exploited as follows. Let \( I \) be the current ideal and let \( e_{t,ℓ}' \) be the last acquire on lock \( ℓ \) performed by thread \( t \) such that \( e_{t,ℓ}' \in I \). Let \( \operatorname{Acq}_t^I = \{ e_{t,ℓ}' \in T \text{r}_t \} \) be the set of last acquire events in \( I \) of every thread. Let \( e_t' \) be the last event (according to trace order \( \leq^ω_t \)) in \( \operatorname{Acq}_t^I \). Then, for every other acquire \( e' \in \operatorname{Acq}_t^I \setminus \{ e_t' \} \), the matching release \( \operatorname{match}_t(e') \) must be added in \( I \). Hence, if we can efficiently determine the events \( e_{t,ℓ}' \) each time, then we can also efficiently determine \( e_t' \) and thus efficiently perform the closure each time. So, how do we determine \( e_{t,ℓ}' \) efficiently each time? We achieve this by maintaining a FIFO sequence \( \operatorname{CSHist} \), for every thread \( t \in \text{Thr}_σ \) and lock \( ℓ \in \text{Locks}_σ \). For every critical section on lock \( ℓ \) in thread \( t \), there is a corresponding entry in the list \( \operatorname{CSHist}_{t,ℓ} \), and the order of these entries is the same as the order in which these critical sections were performed in the trace \( σ \). Every entry (corresponding to critical section with acquire \( a \)) is a pair \( (\text{TLClosure}_σ(a), \text{TLClosure}_σ(\text{match}_σ(a))) \), represented as a pair of vector timestamps \( (C_a, C_{\text{match}_σ(a)}) \).

Let us now see how we perform the check in Line 4 using these data structures \( \{ \operatorname{CSHist}_{t,ℓ} \}_{t,ℓ} \). For the current ideal \( I \), we essentially need to determine the last acquire \( e_{t,ℓ}' \) of each thread \( t \) and lock \( ℓ \) that belongs to \( I \). This can be done by traversing the list \( \operatorname{CSHist}_{t,ℓ} \) starting from the earliest entry, until we encounter the entry corresponding to the last acquire \( e_{t,ℓ}' \) that belongs to \( I \) (this corresponds to a simple timestamp comparison). All entries in \( \operatorname{CSHist}_{t,ℓ} \) prior to the identified event \( e_{t,ℓ}' \) can then be discarded from \( \operatorname{CSHist}_{t,ℓ} \), because the ideal now contains their information and only grows monotonically through the course of the fixpoint computation and the discarded entry will not be of use from now on. Thus, every entry in the lists \( \{ \operatorname{CSHist}_{t,ℓ} \}_{t,ℓ} \) is traversed only once and the overall fixpoint computation runs in linear time when the number of \( T \) is constant.

4.3 Checking for a Sync-Preserving Race on a Given Event with a Given Thread

Algorithm 2: Checking if There Is a Sync-Preserving Race on a Given Event with a Given Thread

\[
\begin{align*}
\text{Input:} & \quad \text{Trace } σ, \text{ Event } e = \langle t, a(x) \rangle, \text{ Thread } u \\
1 & \text{for each } b \in \{ r, w \} \text{ such that } b = a^3 \\
2 & \quad \text{let } L_b \text{ be the list of events } e' \text{ of the form } \langle u, b(x) \rangle \text{ such that } e' \leq^ω_t e \\
3 & \text{for each } b \in \{ r, w \} \text{ such that } b = a \\
4 & \quad I_b \leftarrow \emptyset \\
5 & \text{for each } e' \in L_b \\
6 & \quad I_b \leftarrow \text{ComputeSPIdeal}(σ, e', I_b) \\
7 & \text{if } e' \notin I_b \text{ then} \\
8 & \quad \text{declare 'race' and exit}
\end{align*}
\]

Let us now consider Algorithm 2. This algorithm takes as input a trace \( σ \), an event \( e = \langle t, a(x) \rangle \) and a threads \( u \neq t \), and checks if there is an event \( e' \) of thread \( u \) such that \( e' \leq^ω_t e \) and \( (e', e) \) is a sync-preserving race. Algorithm 2 works as follows. For the sake of simplicity, assume that \( e \) is a read event, i.e., \( a = r \). The algorithm accesses the list \( L_w \) of write events \( e' = \langle u, w(x) \rangle \) that appear prior to \( e \) (Line 2); these lists can be constructed in linear time, in a single pass traversal of the trace. The algorithm simply traverses \( L_w \) (according to trace order \( \leq^ω_t \)) and checks for races with

\( ^3b \neq a \) whenever not both \( b \) and \( a \) are read (r) operations
each event in \( L_w \), by computing the fixpoint closure sets as in Algorithm 1. Instead of computing the ideal from scratch, the algorithm exploits the monotonicity property outlined earlier in Lemma 4.3 by reusing the ideal \( I_{aq} \) computed so far. As with Algorithm 1, this algorithm also uses vector timestamps and maintains lists \( \{ \text{CSHist}_{t,f} \}_{t \in \text{Thr}, f \in \text{Locks}} \) for computing successive ideals efficiently. Overall again, each entry in \( \{ \text{CSHist}_{t,f} \}_{t \in \text{Thr}, f \in \text{Locks}} \) is visited a constant number of times and thus Algorithm 2 runs in linear time and uses linear space.

4.4 Algorithm SyncP for Sync-Preserving Race Prediction

The pseudo-code for SyncP is presented in Algorithm 3. This is a one pass streaming algorithm that processes events as they appear in the trace, modifying its state and detecting races on the fly. The algorithm maintains several data structures including vector clocks and FIFO queues in its state. We first describe these data structures, then discuss how the algorithm initializes and modifies them as it processes the trace, and finally discuss the time and space usage for this algorithm; many of these details have already been spelt out in Sections 4.1-4.3.

Let us first briefly explain the notion of vector timestamps and vector clocks [Fidge 1991; Mattern 1988]. A vector timestamp is a mapping \( V : \text{Thr}_a \rightarrow \mathbb{N} \) from the threads of a trace to natural numbers, and can be represented as a vector of length \( |\text{Thr}_a| \). The join of two vector timestamps \( V_1 \) and \( V_2 \), denoted \( V_1 \sqcup V_2 \) is the vector timestamp \( \lambda t, \max(V_1(t), V_2(t)) \). Vector timestamps can be compared in a pointwise fashion: \( V_1 \sqsubseteq V_2 \) iff \( V_1(t) \leq V_2(t) \). The minimum timestamp is \( \perp = \lambda t, 0 \). For a scalar \( c \in \mathbb{N} \), we use \( V[t \mapsto c] \) to denote the timestamp \( \lambda u, c \), if \( u = t \) then \( c \) else \( V(u) \).

Vector clocks are variables that take values from the space of vector timestamps. We use normal font for timestamps (\( t, r, w \)) and boldfaced font for vector clocks (\( C, C', I \ldots \)). We use \( \lambda t, \sigma \) to denote the timestamp \( \lambda u, \sigma \) and tuple \( \lambda t, \sigma \). A dedicated clock \( L \) is essentially a succinct representation of the set \( \text{TLClosure}(e) \). The algorithm ensures that an event appears only once, and additionally ensures that the entries respect the order of appearance of the corresponding events in the trace. The first kind of FIFO queues are used in the fixpoint computation. For this, the algorithm maintains queues \( \text{CSHist}_{t,f}^{(t_1,t_2,a_1,a_2,x)} \), one for every thread \( t \), lock \( f \) and tuple \((t_1, t_2, a_1, a_2, x)\). Each entry of \( \text{CSHist}_{t,f}^{(t_1,t_2,a_1,a_2,x)} \) is a triplet \((g_e, C_e, C'_e)\) and corresponds to an acquire event \( e \) in the form \((t, \text{acq}(f))\) — here, \( g_e \) is the index of \( e \) in the trace, \( C_e \) is the timestamp of \( e \) and \( C'_e \) is the timestamp of the matching release \( \text{match}(e) \). The second kind of queues are of the form \( \text{AccessHist}_{t,a,x}^{(u)} \), one for each \( t, u \in \text{Thr}, a \in \{ r, w \} \) and \( x \in \text{Vars} \). Entry in such a FIFO queue corresponds to access events of the form \((t, a(x))\). Each entry is of the form \((C_{\text{prev}(e)}, C_e)\) where \( C_{\text{prev}(e)} \) is the timestamp of \( \text{prev}(e) \) (and \( \perp \) if \( \text{prev}(e) \) does not exist) and \( C_e \) is the timestamp of \( e \). All FIFO queues are empty (\( \emptyset \)) in the beginning. We write \( L \cdot \text{first()} \) and \( L \cdot \text{last()} \) to denote the first and last elements of the FIFO queue \( L \). Further, \( L \cdot \text{isEmpty()} \), \( L \cdot \text{addLast()} \) and \( L \cdot \text{removeFirst()} \) respectively represent functions.
that check for emptiness of $L$, add an element at the end of $L$ and remove the earliest (first) element of $L$.

```plaintext

Algorithm 3: Detailed Streaming Algorithm for Checking Sync-Preserving Races

```
Let us now describe the working of SyncP. The algorithm works in a streaming fashion and processes each event $e$ as soon as it appears, by calling the appropriate handler depending upon the operation performed in $e$ (read, write, acquire or release). The argument for each handler is the thread performing the event and the object (variable or lock) accessed in the event. In each handler, the algorithm updates vector clocks to compute timestamps, and maintains the invariants of the FIFO queues. In addition, inside read and write handlers, the algorithm also checks for races using a fixpoint computation (function $\text{ComputeSPIdeal}$). We explain some of these briefly.

**Computing Vector Timestamps.** The algorithm computes the timestamp of an event $e$, denoted $C_e$ when processing the event $e$. The timestamps computed in the algorithm are such that

$$\forall \tau \in \text{Thr}, C_e(\tau) = |\{f \in \text{Events} \mid \text{thr}(f) = \tau, f \in \text{TLClosure}(e)\}|$$

Observe that, with this invariant, we have $C_e \subseteq C'_e$ iff $e \in \text{TLClosure}(e')$. The algorithm uses vector clocks $\{C_e\}_{\tau \in \text{Thr}}$ and $\{\bar{L}_x\}_{x \in \text{Vars}}$ and ensures that after processing an event $e$, (1) $C_e$ stores the timestamp $C_{e_i}$, where $e_i$ is the last event by thread $\tau$ that occurs before $e$, and (2) $\bar{L}_x$ stores the timestamp $C_{e_x}$, where $e_x$ is the last event with $\text{op}(e_x) = w(x)$ that occurs before $e$. The algorithm correctly maintains these values by appropriate vector clock operations on Lines 45, 53, 60 and 64.

Let us now describe the invariant for the vector clock $\bar{I}^t_{t_1,t_2,a_1,a_2,x}$ (given tuple $(t_1, t_2, a_1, a_2, x)$). Let $e_{t_1,a_1,x}$ be the last event with $\text{thr}(e_{t_1,a_1,x}) = t_1$ and $\text{op}(e_{t_1,a_1,x}) = a_1$ (for $i \in \{1, 2\}$). Let $I \subseteq \text{Events}_x$ be defined as follows. If $e_{t_1,a_1,x}$ does not exist, then $I = \text{SPClosure}_\sigma(\{\text{prev}_\sigma(e_2)\})$. Otherwise, let $e$ be be the first event in $\sigma$ (according to trace order $\leq^\sigma_{tr}$) with $\text{thr}(e) = t_1$ and $\text{op}(e) = a_1$ (for $x$ such that $(e, e_2)$ is a sync-preserving race of $\sigma$. If no such event exists, then let $e = e_{t_1,a_1,x}$. Then, $I = \text{SPIdeal}_\sigma(e_1, e_2)$. Then, the timestamp stored in $\bar{I}^t_{t_1,t_2,a_1,a_2,x}$ is $\bigcup_{u \in \text{Thr}} C_{e_u}$, where $e_u$ is the last event of thread $u$ which is in $I$.

**Checking Races.** When processing an access event $e = \langle t, a(x) \rangle$, the algorithm checks for a race as follows. For every other thread $u$ and for every other conflicting type $b \neq a$, the algorithm calls $\text{checkRace}$ with the list $\text{Lst} = \text{AccessHist}^t_{u,b,x}$ and the timestamp representation of the current ideal as argument. This function, similar to Algorithm 2, scans $\text{Lst}$ and reports races by repeatedly performing fixpoint computations and checking membership in some set (using the timestamp comparison in Line 38). The closure computation is performed using the optimizations discussed in Section 4.2 (use of FIFO queues $\text{CSHist}_{t,t}$).

**Space Optimizations.** Observe that, for a given thread $t$ and lock $\ell$, the algorithm, as presented, maintains $4T^2 \cdot V$ FIFO queues $\{\text{CSHist}^t_{t_1,t_2,a_1,a_2,x}\}_{t_1,t_2 \in \text{Thr}, a_1,a_2 \in \{x,v\}, x \in \text{Vars}}$. The total number of entries across these queues will then be $O(\mathcal{A} \cdot 4T^2 \cdot V)$. To this end, we observe that all the above data structures essentially have the same content, and are suffixes of a common queue corresponding to the critical sections on $\ell$ in thread $t$. Indeed, we exploit this redundancy and instead maintain a common underlying data-structure that stores all entries corresponding to acquire and releases on $\ell$ in $t$, and maintain a pointer for each $(t_1, t_2, a_1, a_2, x)$ such that $e_u$ keeps track of the starting index of the FIFO queue $\text{CSHist}^t_{t_1,t_2,a_1,a_2,x}$. With this optimization, we only store $O(\mathcal{A})$ entries along with additional $4T^3 \cdot V \cdot L$ pointers, one for every tuple $(t_1, t_2, a_1, a_2, x)$ and every shared queue indexed by $(t, \ell)$. The same observations also apply to the FIFO queues $\{\text{AccessHist}^t_{u,a,x}\}_{u \in \text{Thr}}$.

**Lemma 4.7** (Correctness). For every access event $e$ in the input trace $\sigma$, Algorithm 3 declares a race when processing $e$ if and only if there is an event $e'$ such that $e' \leq^\sigma_{tr} e$ and $(e', e)$ is a sync-preserving race of $\sigma$.

The proof of Lemma 4.7 follows directly from the correctness of the semantics of clocks and other observations outlined in Section 4.1.
The time complexity of the algorithm can be determined as follows. The algorithm visits each entry in the FIFO queues $\text{AccessHist}_{t,x}^{(n)}$ once, performing constant number of vector clock operations, each running in $O(T)$ time. The total length of all these queues is $O(T \cdot N)$ (more precisely, the number of access events in the trace). Similarly, the algorithm visits each entry in the FIFO queues $\text{CSHist}_{t,f}^{(t_1,t_2,a_1,a_2,x)}$ once, performing constantly many vector clock operations. The total number of entries in these queues is $O(T^2 \cdot V \cdot \Lambda)$. This gives us the following complexity for SyncP.

**Lemma 4.8 (Complexity).** Let $\sigma$ be a trace with $T$ threads, $L$ locks, $V$ variables and $N$ events, of which $\Lambda$ are acquire events. Then, Algorithm 3 runs in time $O(N \cdot T + \Lambda \cdot V \cdot T^3)$ and uses space $O(N + T^3 \cdot V \cdot L)$ on input $\sigma$.

The proof of Theorem 3.1 follows from Lemma 4.7 and Lemma 4.8.

### 4.5 Linear Space Lower Bound

In this section we prove the lower-bounds of Theorem 3.2 and Theorem 3.3, i.e., that any streaming algorithm for sync-preserving race prediction must essentially use linear space, while the time-space product of any algorithm for the problem must be quadratic in the length of the input trace.

**The language $L_n$.** Given a natural number $n$, we define the equality language $L_n = \{u \# v : u, v \in \{0,1\}^n \text{ and } u = v\}$, i.e., it is the language of two $n$-bit strings that are separated by $\#$ and are equal.

**Lemma 4.9.** Any streaming algorithm that recognizes $L_n$ uses $\Omega(n)$ space.

**Reduction from $L_n$ recognition to sync-preserving race prediction.** Consider the language $L_n$ for some $n$. We describe a transducer $\mathcal{A}_n$ such that, on input a string $s = u \# v$, the output $\mathcal{A}_n(s)$ is a trace $\sigma$ with 2 threads, $O(n \cdot \log n)$ events, $2 \cdot \log n + 1$ locks and a single variable such that the following hold.

1. If $s \notin L_n$, then $\sigma$ has no predictable race.
2. If $s \in L_n$, then $\sigma$ has a single predictable race, which is a sync-preserving race.

Moreover, $\mathcal{A}_n$ uses $O(\log n)$ working space. The transducer $\mathcal{A}_n$ uses a single variable $x$, two sets of locks $A = \{a_1, \ldots, a_{\log n}\}$ and $B = \{b_1, \ldots, b_{\log n}\}$, plus one additional lock $c$. The trace $\sigma$ consists of two local traces $\pi_1, \pi_2$ of threads $t_1$ and $t_2$ which encode the bits of $u$ and $v$, respectively.
The local trace $\pi_1$ is constructed as follows. For every $i \in [n]$, $\pi_1$ contains an event $e_i^1$, which is a write event $w(x)$ if $u[i] = 1$, and a read event $r(x)$ otherwise. The events $e_i^1$ are surrounded by locks from $A$ and $B$ arbitrarily, as long as the following holds. For any $i < j$, we have

$$\text{locksHeld}_\sigma(e_i^1) \cap A \not\subseteq \text{locksHeld}_\sigma(e_j^1) \cap A$$

and

$$\text{locksHeld}_\sigma(e_i^1) \cap B \not\subseteq \text{locksHeld}_\sigma(e_j^1) \cap B.$$ 

Here, the locks held at an event $e$ has the obvious meaning: $\text{locksHeld}_\sigma(e) = \{ \ell \in \text{Locks}_\sigma | \exists a \in \text{Acquires}_\sigma(\ell) \text{ such that } e \in CS_\sigma(a) \}$.

The above property can be easily met, for example, by making $A_n$ perform a breadth-first traversal of the subset-lattice of $A$ (resp., $B$) starting from the top (resp., bottom). Given the current $i$, the transducer surrounds $e_i^1$ with the locks of the current element in the corresponding lattice. Finally, every write event $e_i^1$ is surrounded by the lock $c$.

(2) The local trace $\pi_2$ is similar to $\pi_1$, i.e., we have an event $e_i^2$ for each $i \in [n]$, which is a write event $w(x)$ if $u[i] = 1$, otherwise it is a read event $r(x)$. The locks that protect $e_i^2$ satisfy

$$\text{locksHeld}_\sigma(e_i^2) \cap (A \cup B) = A \cup B \setminus \text{locksHeld}_\sigma(e_i^1).$$

Finally, similarly to $\pi_1$, every write event $e_i^2$ is surrounded by the lock $c$.

See Figure 4 for an illustration. Observe that $A_n$ uses $O(\log n)$ bits of memory, for storing a bit-set of locks for each set $A$ and $B$ that must surround the current event $e_i^1$ and $e_i^2$.

The key idea of the construction is the following. Any two events $e_i^1, e_j^2$ are surrounded by a common lock from the set $A \cup B$ iff $i \neq j$. Hence, $(e_i^1, e_j^2)$ may be a predictable race of $\sigma$ only if $i = j$. In turn, if $u[i] = v[j]$, then either both events are read events, or both are write events. In the former case the events are not conflicting, while in the latter case the two events are surrounded by lock $c$. In both cases no race occurs between $e_i^1$ and $e_j^2$. On the other hand, if $u[i] \neq v[j]$, then one event is a read and the other is a write event. Hence, the two events conflicting, and one of them is not surrounded by lock $c$, thereby constituting a predictable race. The following lemma makes the above insight formal and establishes the correctness of our reduction.

**Lemma 4.10.** The following assertions hold.

1. If $s \in L_n$, then $\sigma$ has no predictable race.
2. If $s \not\in L_n$, then $\sigma$ has a single predictable race, which is a sync-preserving race.

We now prove Theorem 3.2, and refer to [Mathur et al. 2020b] for the proof of Theorem 3.3.

**Proof of Theorem 3.2.** Consider any algorithm $A_1$ for sync-preserving race prediction, executed in the family of traces $\sigma$ constructed in our above reduction. Let $m = n/\log n$, and assume towards contradiction that $A_1$ uses $o(m)$ space. Then we can pair $A_1$ with the transducer $A_m$, and obtain a new algorithm $A_2$ for recognizing $L_m$. Since $A_m$ uses $O(\log m)$ space, the space complexity of $A_2$ is $o(m)$. However, this contradicts Lemma 4.9. The desired result follows. □

## 5 BEYOND SYNCHRONIZATION-PRESERVING RACES

In this section we explore the problem of dynamic race prediction beyond sync-preservation. We show Theorem 3.4, i.e., that even when just two critical sections are present in the input trace, predicting races with witnesses that might reverse the order of the critical sections becomes intractable. Our reduction is from the realizability problem of Rf-posets, which we present next.

**Rf-posets.** An Rf-poset is a triplet $P = (X, P, RF)$, where $X$ is a set of read and write events, $P$ defines a partial order $\leq_P$ over $X$, and $RF$: $\text{Rds}(X) \to \text{Wts}(X)$ is a reads-from function that maps...
every read event of X to a write event of X. Given two distinct events \( e_1, e_2 \in X \), we write \( e_1 \parallel P e_2 \) to denote \( e_1 \not\leq_P e_2 \) and \( e_2 \not\leq_P e_1 \). Given a set \( Y \subseteq X \), we denote by \( P|Y \) the projection of \( P \) on \( Y \), i.e., we have \( \leq_{P|Y} \subseteq Y \times Y \), and for all \( e_1, e_2 \in Y \), we have \( e_1 \leq_{P|Y} e_2 \) iff \( e_1 \leq_P e_2 \). Given a partial order \( Q \) over \( X \), we say that \( Q \) refines \( P \) denoted \( Q \subseteq P \) if for every two events \( e_1, e_2 \in X \), if \( e_1 \leq_P e_2 \) then \( e_1 \leq_Q e_2 \). We consider that each event of \( X \) belongs to a unique thread, and there is thread order \( \leq_{TO} \) that defines a total order on the events of \( X \) that belong to the same thread, and \( P \) agrees with \( \leq_{TO} \). The number of threads of \( P \) is the number of threads of the events of \( X \).

The Realizability problem of rf-posets. Given an rf-poset \( P = (X, P, RF) \), the realizability problem is to decide whether \( P \) can be linearized to a total order \( \sigma \) such that \( lw_\sigma = RF \). It has long been known that the problem is NP-complete [Gibbons and Korach 1997], while it was recently shown that it is even W[1]-hard [Mathur et al. 2020a].

Our proof of the lower bound of Theorem 3.4 is by a two-step reduction. First we define a variant of the realizability problem for rf-posets, namely reverse rf-realizability, and show that it is W[1]-hard when parameterized by the number of threads. Afterwards, we reduce reverse rf-realizability to the decision problem of dynamic race prediction, which concludes the hardness of the latter.

Rf-Triplets. Given an RF-poset \( P = (X, P, RF) \), an rf-triplet of \( P \) is a tuple \( \lambda = (w, r, w') \) such that (i) \( r \) is a read event, (ii) \( RF(r) = w \), and (iii) \( w \neq w' \). We refer to \( w, r \) and \( w' \) as the write, read, and interfering write event of \( \lambda \), respectively. We denote by \( \text{Triplets}(P) \) the set of rf-triplets of \( P \).

We next define a variant of rf-poset realizability, and show that, like the original problem, it is W[1]-hard parameterized by the number of threads.

Reverse RF-Poset Realizability. The input is a tuple \( (P, \lambda, \sigma) \), where \( P = (X, P, RF) \) is an rf-poset, \( \lambda = (\overline{w}, \overline{r}, \overline{w}') \) is a distinguished triplet of \( P \), and \( \sigma \) is a witness to the realizability of \( P \) such that \( \overline{w}' <_\sigma \overline{w} \). The task is to determine whether \( P \) has a linearization \( \sigma' \) with \( \overline{w} <_\sigma \overline{w}' \). In words, \( P \) is already realizable by a witness that orders \( \overline{w}' \) before \( \overline{w} \), and the task is to decide whether \( P \) also has a witness in which this order is reversed.

Hardness of Reverse RF-Poset Realizability. We show that the problem is W[1]-hard when parameterized by the number of threads of the rf-poset. Our reduction is from rf-realizability. We first present the construction and then argue about its correctness.

Construction. Consider an rf-poset \( P = (X, P, RF) \) with \( k \) threads, and we construct an instance of reverse rf-poset realizability \( (P' = (X', P', RF'), \lambda, \sigma) \) with \( k' = O(k^2) \) threads. We refer to Figure 5 for an illustration. For simplicity of presentation, we assume wlog that the following hold.

1. \( X \) contains only the events of the triplets of \( P \).
2. For every read event \( r \), we have \( \text{thr}(r) = \text{thr}(w) \), i.e., every read observes a local write event.

Let \( \{X_i\}_{1 \leq i \leq k} \) be a partitioning of \( X \) such that each \( X_i = P|X_i \) is a total order containing all events of thread \( i \) (i.e., it is the thread order for thread \( i \)). We first construct the rf-poset \( P' = (X', P', RF') \). The threads of \( P' \) are defined implicitly by the sets of events for each thread. In particular, \( X' \) is partitioned in the sets \( X'_i \) (which are the events of \( P_i \)), as well as two sets \( X'^i_j \) and \( Y'^i_j \) for each \( i, j \in [k] \) with \( i \neq j \), where each such \( X'^i_j \) and \( Y'^i_j \) contains events of a unique thread of \( P' \). Finally, we have two threads containing the events of the distinguished triplet \( \lambda \). Hence, \( P' \) has \( k' = k + k \cdot (k - 1) + 2 = k^2 + 2 \) threads.

We first define the set of triplets \( \text{Triplets}(P') \), which defines the event set \( X' \) and the observation function \( RF' \). We have \( X \subseteq X' \) and \( \text{Triplets}(P) \subseteq \text{Triplets}(P') \). In addition, we create a distinguished triplet \( \lambda = (\overline{w}, \overline{r}, \overline{w}') \), and all its events are in \( X' \). Finally, for every \( i, j \in [k] \) with \( i \neq j \), we have \( X'^i_j, Y'^i_j \subseteq X' \), where the sets \( X'^i_j \) and \( Y'^i_j \) are constructed as follows. We call a pair of events...
(e₁, e₂) ∈ Xᵢ × Xⱼ with e₁ <ᵡ e₂ dominant if for any pair (e'₁, e'₂) ∈ Xᵢ × Xⱼ such that e₁ ≤ᵡ e'₁, and e'₂ ≤ᵡ e₂, and e'₁ <ᵡ e'₂, we have e'ᵢ = eᵢ for each i ∈ [2]. In words, a dominant pair identifies an ordering in P that cannot be inferred transitively by other orderings. For every dominant pair (e₁, e₂) ∈ Xᵢ × Xⱼ we create a triplet (wₑ₁,ₑ₂, wₑ₁,ₑ₂, wₑ₁,ₑ₂), and let wₑ₁,ₑ₂, rₑ₁,ₑ₂ ∈ Xᵢ and wₑ₁,ₑ₂ ∈ Yᵢ.

We now define the partial order P'. For every triplet (w, r, w') of P', we have w <ᵡ r. For every i, j ∈ [k] with i ≠ j, for every two events e₁, e'₁ ∈ Xᵢ such that e₁ <ᵡ e'₁, for every two events e₂, e'₂ ∈ Xⱼ such that e₂ <ᵡ e'₂, if rₑ₁,ₑ₂ and wₑ₁,ₑ₂ are events of X' (i.e., (e₁, e₂) and (e'₁, e'₂) are dominant pairs), we have (i) rₑ₁,ₑ₂ <ᵡ wₑ₁,ₑ₂ and (ii) wₑ₁,ₑ₂ <ᵡ wₑ₁,ₑ₂. Finally, for every triplet of the form (wₑ₁,ₑ₂, rₑ₁,ₑ₂, wₑ₁,ₑ₂), we have

\[ e₁ <ᵡ rₑ₁,ₑ₂ \quad \text{and} \quad wₑ₁,ₑ₂ <ᵡ e₂ \quad \text{and} \quad wₑ₁,ₑ₂ <ᵡ rₑ₁,ₑ₂. \]

The following lemma establishes that P' is indeed a partial order.

**Lemma 5.1.** P' is a partial order.

We now turn our attention to the solution σ of P', which is constructed in two steps. First, we construct a partial order Q ⊆ P' over X' which orders in every triplet the interfering write before the write of the triplet. That is, for every triplet (w, r, w') of P', we have w' <ᵡ w. Then, we obtain σ by linearizing Q arbitrarily. The following lemma states that σ witnesses the realizability of P'.

**Lemma 5.2.** The trace σ realizes P'.

Observe that the size of P' is polynomial in the size of P. The following lemma states the correctness of the reduction. Here we sketch the argument while the detailed proof is in our companion technical report [Mathur et al. 2020b].

**Lemma 5.3.** Reverse rf-poset realizability is W[1]-hard parameterized by the number of threads.

Correctness. We now present the key insight behind the correctness of the reduction. Consider any dominant pair of events (e₁, e₂) in the initial rf-poset, i.e., we have e₁ <ᵡ e₂. Observe that the two events are unordered in P'. Now consider any trace σ' that solves the reverse rf-poset realizability problem for P'. By definition, σ' must reverse the order of the two writes of the conflicting triplet, i.e., we must have \( \overline{w} <_{\sigma'} \overline{w} \).

1. Since \( \overline{w} <_{\sigma'} \overline{w} \), we have \( \overline{r} <_{\sigma'} \overline{w} \), so the last write of \( \overline{r} \) is not violated in \( \sigma' \).
2. Since \( wₑ₁,ₑ₂ <ᵡ \overline{r} \), by the previous item we also have transitively \( wₑ₁,ₑ₂ <ᵡ \overline{w} \), and since \( \overline{w} <ᵡ \overline{wₑ₁,ₑ₂} \), we have , transitively \( wₑ₁,ₑ₂ <_{\sigma'} \overline{wₑ₁,ₑ₂} \).

![Fig. 5. Reduction of rf-poset realizability (5a) to reverse rf-poset realizability (5b).](image-url)
We construct a trace \( \sigma' \) for the problem of dynamic race prediction. We first describe the construction.

Hence, the witness \( \sigma' \) also holds for \( \mathcal{P} \). The inverse direction is similar.

**Hardness of 1-Reversal Dynamic Race Prediction.** We are now ready to prove our second step of the reduction, i.e., to establish an FPT reduction from reverse rf-poset realizability to the decision problem of dynamic race prediction. We first describe the construction.

Consider an instance \((\mathcal{P} = (X, P, RF), \lambda, \sigma)\) of reverse rf-poset realizability, and we construct a trace \( \sigma' \) such a specific event pair of \( \sigma' \) is a predictable race iff \( \mathcal{P} \) is realizable by a witness that reverses \( \lambda \). We assume wlog that \( X \) contains only events that appear in triplets of \( \mathcal{P} \). We construct \( \sigma' \) by inserting various events in \( \sigma \), as follows. Figure 6 provides an illustration.

1. For every dominant pair \((e_1, e_2)\) of \( \mathcal{P} \), we introduce a new variable \( x_{e_1, e_2} \), and a write event \( w(x_{e_1, e_2}) \) and a read event \( r(x_{e_1, e_2}) \). We make \( \text{thr}(w(x_{e_1, e_2})) = \text{thr}(e_1) \) and \( \text{thr}(r(x_{e_1, e_2})) = \text{thr}(e_2) \). Finally, we thread-order \( w(x_{e_1, e_2}) \) after \( e_1 \) and \( r(x_{e_1, e_2}) \) before \( e_2 \). Notice that any correct reordering of \( \sigma' \) must order \( w(x_{e_1, e_2}) \leq_{\sigma'} r(x_{e_1, e_2}) \), and thus, transitively, also order \( e_1 \leq_{\sigma'} e_2 \).

2. For every thread \( t_i \) that does not read \( \bar{w} \) or \( \overline{w} \) we introduce a new variable \( x^i \), a write event \( w(x^i) \), and a read event \( r(x^i) \). We make \( \text{thr}(w(x^i)) = t_i \) and \( \text{thr}(r(x^i)) = \overline{\text{thr}(\overline{w})} \). We thread-order each \( w(x^i) \) as the last event of \( t_i \), and thread-order all \( r(x^i) \) as final events of \( \text{thr}(\overline{w}) \) so far.

3. We introduce a new variable \( y \), and a write event \( w(y) \) and a read event \( r(y) \). We make \( \text{thr}(w(y)) = \text{thr}(\overline{w}) \) and \( \text{thr}(r(y)) = \overline{\text{thr}(\overline{w})} \). Finally, we thread-order \( w(y) \) and \( r(y) \) at the end of their respective threads. In particular, \( r(y) \) is thread-ordered after the events \( r(x^i) \)

![Diagram](attachment:image.png)

Fig. 6. Example of our reduction of an instance of reverse rf-poset realizability (Figure 6a) to an instance of dynamic data-race prediction (Figure 6b) on the event pair \((e_1, e_1^8)\).
introduced in the previous item. Notice that because of this ordering and the previous item, any correct reordering σ' of σ must contain all events of X'.

(4) We introduce a lock ℓ and two pairs of lock-acquire and lock-release events (acqℓ, relℓ), for each i ∈ [2]. We make thr(acqℓ) = thr(relℓ) = ti, where tj = thr(⃗ω) if i = 1 and tj = thr(⃗ω) otherwise. Finally, we surround with the critical section of acq1(ℓ), rel1(ℓ) all events of the corresponding thread, and surround with the critical section of acq2(ℓ), rel2(ℓ) the event ⃗ω. Notice that any correct reordering σ' that witnesses a race on (w(y), x(y)) is missing rel1(ℓ), and thus must order rel2(ℓ) ≤rel1(ℓ) acq1(ℓ). In turn, this leads to a transitive ordering ⃗ω ≤rel1(ℓ) ⃗ω', and since the last write of ⃗ω must be lwσ'(⃗ω) = ⃗ω, we must have ⃗ω ≤σ' ⃗ω'.

We now outline the correctness of the reduction (proof in [Mathur et al. 2020b]). Consider any correct reordering σ* that witnesses a predictable race (w(y), x(y)) on σ'. Item 2 and Item 3 above guarantee that X ⊆ Eventsσ*, while Item 1 guarantees that σ* linearizes P, and Item 4 guarantees that σ* reverses λ, i.e., ⃗ω ≤σ' ⃗ω'. Finally, note that σ' has size that is polynomial in n, while the number of threads of σ' equals the number of threads of P. This concludes the proof of Theorem 3.4.

6 EXPERIMENTS

In this section we report on an implementation and experimental evaluation of the techniques presented in this work. Our objective is two-fold. The first goal is to quantify the practical relevance of sync-preservation, i.e., whether in practice the definition captures races that are missed by the standard notion of happens-before and WCP [Kini et al. 2017] races. The second goal is to evaluate the performance of our algorithm SyncP for detecting sync-preserving races.

6.1 Experimental Setup

We have implemented SyncP (Algorithm 3) for predicting all sync-preserving races in our tool RAPID [Mathur 2020], written in Java, and evaluated it on a standard set of benchmarks.

Benchmarks. Our benchmark set consists of standard benchmarks found in the recent literature [Huang et al. 2014; Kini et al. 2017; Mathur et al. 2018; Pavlogiannis 2019; Roemer et al. 2018; Yu et al. 2018]. It consists of 30 concurrent programs taken from standard benchmark suites: (i) the IBM Contest benchmark suite [Farchi et al. 2003], (ii) the Java Grande forum benchmark suite [Smith et al. 2001], (iii) the DaCapo benchmark suite [Blackburn et al. 2006], (iv) the Software Infrastructure Repository [Do et al. 2005], and (v) some standalone benchmarks. For each benchmark, we generated a single trace using RV-Predict [Rosu 2018] and evaluated all methods on the same trace.

Compared Methods. We compare our algorithm with state-of-the-art sound race detectors, namely, SHB [Mathur et al. 2018], WCP [Kini et al. 2017] and M2 [Pavlogiannis 2019]. Recall that SHB and WCP are linear time algorithms that perform a single pass of the input trace σ. SHB computes happens-before races and is sound even beyond the first race. On the other hand, WCP is sound only for the first race report. In order to allow WCP to soundly report more than one race, whenever a race is reported on an event pair (e1, e2) (i.e., we have e1 ⊥WCP e2), we force an order e1 ≤WCP e2 before proceeding with the next event of σ. This is a standard practice that has been followed in other works, e.g.,[Pavlogiannis 2019; Roemer et al. 2018]. Finally, M2 is a more heavyweight algorithm that makes sound reports for all races by design, though its running time is a larger polynomial (of order n^4) [Pavlogiannis 2019].

Optimizations. In general, the benchmark traces can be huge and often scale to sizes of order as large as 10^6. A closer inspection shows that many events, even though they perform accesses to shared memory, are non-racy and even totally ordered by fork-join mechanisms and data flows in the trace. We have implemented a lightweight, linear time, single-pass optimization of the input
trace $\sigma$ that filters out such events. The optimization simply identifies memory locations $x$ whose conflicting accesses are totally ordered in $\sigma$ by thread and data-flow orderings, and ignores all such accesses in $\sigma$. For a fair comparison, we employ the optimization in all compared methods. This is similar to FastTrack-like optimizations [Flanagan and Freund 2009], which identify and ignore thread-local events. We note that, as each of the compared methods attempts to report as many races as possible, epoch-like optimizations were not applied.

**Reported Results.** Our experiments were conducted on a 2.6GHz 64-bit Linux machine with Java 1.8 as the JVM and 30GB heap space. Each of the compared methods is evaluated on the same input trace $\sigma$. For every such input, the respective method reports the following race warnings.

1. **Racy events.** We report the number of events $e_2$ such that there is an event $e_1$ with $e_1 <_{tr} e_2$ for which a race $(e_1, e_2)$ is detected. We remark that this is the standard way of reporting race warnings [Flanagan and Freund 2009; Genç et al. 2019; Kini et al. 2017; Mathur et al. 2018; Roemer et al. 2018], as it allows for one-pass, linear time algorithms that avoid the overhead of testing for races between all possible $\Theta(n^2)$ pairs of events.

2. **Racy source-code lines.** We report the number of distinct source-code lines which correspond to events $e_2$ that are found as racy in Item 1. This is a meaningful measure, as the same source-code line might be reported by many different events $e_2$.

3. **Racy memory locations.** We report the number of different memory locations that are accessed by all the events $e_2$ that are found as racy in Item 1.

4. **Running time.** We measure the time of the algorithm required to process each benchmark, while imposing a 1-hour timeout (TO).

### 6.2 Experimental Results

We now turn our attention to the experimental results. Table 1 shows the races and running times reported by each method on each benchmark.

**Coverage of Sync-Preserving Races.** We find that every race reported by SHB or WCP is a sync-preserving race, also reported by SyncP. On the other hand, bold-face entries highlight benchmarks which have sync-preserving races that are not happens-before races. We see that such races are found in 11 out of 30 benchmarks. Interestingly, in the 5 most challenging out of these 11 benchmarks, the same pattern occurs if we focus on source-code lines (i.e., the entries in the parentheses). Hence, for these benchmarks, sync-preservation is necessary to capture many racy source-code lines, which happens-before would completely miss. We also remark that the more heavyweight analysis M2 misses several of these races due to frequent timeouts. In total, we have 18 unique source-code lines that are racy but only detected by SyncP. On the other hand, there are only 2 source-code lines that are caught by M2 but not by SyncP.

**Running Times.** Our experimental times indicate that SyncP is quite efficient in practice. Among all algorithms, SyncP is the second fastest, being about 1.4 times slower than the fastest, lightweight SHB, while at the same time, being able to detect considerably more races the SHB (i.e., 1342 more racy events, and 21 more racy source-code lines). On the other hand, SyncP detects even more races than M2, due to timeouts, and even has almost equal detection capability with M2 on the cases that M2 does not time out. Due to the slow performance of M2 (i.e., over 8.5 hours and with several timeouts), we exclude it from the more refined analysis that follows.

**Racy Memory Locations.** We next proceed to evaluate the capability of SyncP in detecting racy memory locations. As all races detected by SHB or WCP are sync-preserving, the same follows for the racy memory locations, i.e., they are all detected as racy by SyncP. On the other hand, Table 2 shows a few cases in which SyncP has discovered racy variables that are missed by SHB and WCP.
Table 1. Dynamic race reports in our benchmarks. \( N \) and \( T \) denote the number of events and number of threads in the respective trace. For races, an entry ‘r (s)’ denotes the number of events \( e_2 \) found to be in race with an earlier event \( e_1 \), as well as the number \( s \) of unique source-code lines corresponding to such events \( e_2 \).

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>( N )</th>
<th>( T )</th>
<th>SHB Races</th>
<th>SHB Time</th>
<th>WCP Races</th>
<th>WCP Time</th>
<th>M2 Races</th>
<th>M2 Time</th>
<th>SyncP Races</th>
<th>SyncP Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>array</td>
<td>51</td>
<td>4</td>
<td>0 (0)</td>
<td>0.02s</td>
<td>0 (0)</td>
<td>0.03s</td>
<td>0 (0)</td>
<td>0.09s</td>
<td>0 (0)</td>
<td>0.04s</td>
</tr>
<tr>
<td>critical</td>
<td>59</td>
<td>5</td>
<td>3 (3)</td>
<td>0.19s</td>
<td>1 (1)</td>
<td>0.03s</td>
<td>3 (3)</td>
<td>0.11s</td>
<td>3 (3)</td>
<td>0.07s</td>
</tr>
<tr>
<td>account</td>
<td>134</td>
<td>5</td>
<td>3 (1)</td>
<td>0.06s</td>
<td>3 (1)</td>
<td>0.03s</td>
<td>3 (1)</td>
<td>0.23s</td>
<td>3 (1)</td>
<td>0.09s</td>
</tr>
<tr>
<td>airtickets</td>
<td>140</td>
<td>5</td>
<td>8 (3)</td>
<td>0.02s</td>
<td>5 (2)</td>
<td>0.03s</td>
<td>8 (3)</td>
<td>0.13s</td>
<td>8 (3)</td>
<td>0.05s</td>
</tr>
<tr>
<td>pingpong</td>
<td>151</td>
<td>7</td>
<td>8 (3)</td>
<td>1.09s</td>
<td>8 (3)</td>
<td>0.04s</td>
<td>8 (3)</td>
<td>0.17s</td>
<td>8 (3)</td>
<td>0.06s</td>
</tr>
<tr>
<td>twostage</td>
<td>193</td>
<td>13</td>
<td>4 (1)</td>
<td>0.02s</td>
<td>4 (1)</td>
<td>0.09s</td>
<td>4 (1)</td>
<td>0.20s</td>
<td>4 (1)</td>
<td>0.10s</td>
</tr>
<tr>
<td>wronglock</td>
<td>246</td>
<td>23</td>
<td>12 (2)</td>
<td>0.02s</td>
<td>3 (2)</td>
<td>0.09s</td>
<td>25 (2)</td>
<td>0.43s</td>
<td>25 (2)</td>
<td>0.15s</td>
</tr>
<tr>
<td>bbuffer</td>
<td>332</td>
<td>3</td>
<td>3 (1)</td>
<td>0.01s</td>
<td>1 (1)</td>
<td>0.05s</td>
<td>3 (1)</td>
<td>0.11s</td>
<td>3 (1)</td>
<td>0.06s</td>
</tr>
<tr>
<td>prodcors</td>
<td>658</td>
<td>9</td>
<td>1 (1)</td>
<td>0.03s</td>
<td>1 (1)</td>
<td>0.09s</td>
<td>1 (1)</td>
<td>0.20s</td>
<td>1 (1)</td>
<td>0.10s</td>
</tr>
<tr>
<td>clean</td>
<td>1.0K</td>
<td>10</td>
<td>59 (4)</td>
<td>0.04s</td>
<td>33 (4)</td>
<td>0.14s</td>
<td>110 (4)</td>
<td>0.85s</td>
<td>60 (4)</td>
<td>0.17s</td>
</tr>
<tr>
<td>mergesort</td>
<td>3.0K</td>
<td>6</td>
<td>1 (1)</td>
<td>11m10s</td>
<td>1 (1)</td>
<td>0.12s</td>
<td>5 (2)</td>
<td>0.96s</td>
<td>3 (1)</td>
<td>0.13s</td>
</tr>
<tr>
<td>bubblesort</td>
<td>4.0K</td>
<td>13</td>
<td>269 (5)</td>
<td>0.03s</td>
<td>100 (5)</td>
<td>0.27s</td>
<td>374 (5)</td>
<td>8.05s</td>
<td>269 (5)</td>
<td>0.50s</td>
</tr>
<tr>
<td>lang</td>
<td>6.0K</td>
<td>8</td>
<td>400 (1)</td>
<td>0.10s</td>
<td>400 (1)</td>
<td>0.23s</td>
<td>400 (1)</td>
<td>1.31s</td>
<td>400 (1)</td>
<td>0.31s</td>
</tr>
<tr>
<td>readwrites</td>
<td>11K</td>
<td>6</td>
<td>92 (4)</td>
<td>0.12s</td>
<td>92 (4)</td>
<td>0.41s</td>
<td>228 (4)</td>
<td>12.74s</td>
<td>199 (4)</td>
<td>0.77s</td>
</tr>
<tr>
<td>raytracer</td>
<td>15K</td>
<td>4</td>
<td>8 (4)</td>
<td>0.02s</td>
<td>8 (4)</td>
<td>0.30s</td>
<td>8 (4)</td>
<td>0.40s</td>
<td>8 (4)</td>
<td>0.30s</td>
</tr>
<tr>
<td>bufwriter</td>
<td>22K</td>
<td>7</td>
<td>8 (4)</td>
<td>0.10s</td>
<td>8 (4)</td>
<td>0.70s</td>
<td>8 (4)</td>
<td>2.65s</td>
<td>8 (4)</td>
<td>0.84s</td>
</tr>
<tr>
<td>ftpserver</td>
<td>49K</td>
<td>12</td>
<td>69 (21)</td>
<td>6.91s</td>
<td>69 (21)</td>
<td>1.34s</td>
<td>85 (21)</td>
<td>4.11s</td>
<td>85 (21)</td>
<td>4.69s</td>
</tr>
<tr>
<td>moldyn</td>
<td>200K</td>
<td>4</td>
<td>103 (3)</td>
<td>0.05s</td>
<td>103 (3)</td>
<td>1.83s</td>
<td>103 (3)</td>
<td>1.25s</td>
<td>103 (3)</td>
<td>1.86s</td>
</tr>
<tr>
<td>linkedlist</td>
<td>1.0M</td>
<td>13</td>
<td>5.0K (4)</td>
<td>7.25s</td>
<td>5.0K (3)</td>
<td>27.07s</td>
<td>TO</td>
<td>TO</td>
<td>7.0K (4)</td>
<td>5.19s</td>
</tr>
<tr>
<td>derby</td>
<td>1.0M</td>
<td>5</td>
<td>29 (10)</td>
<td>0.01s</td>
<td>28 (10)</td>
<td>16.48s</td>
<td>30 (11)</td>
<td>22.49s</td>
<td>29 (10)</td>
<td>24.07s</td>
</tr>
<tr>
<td>jigsaw</td>
<td>3.0M</td>
<td>12</td>
<td>4 (4)</td>
<td>0.41s</td>
<td>4 (4)</td>
<td>19.53s</td>
<td>6 (6)</td>
<td>11.69s</td>
<td>6 (6)</td>
<td>17.30s</td>
</tr>
<tr>
<td>sunflow</td>
<td>11M</td>
<td>17</td>
<td>84 (6)</td>
<td>39.66s</td>
<td>58 (6)</td>
<td>47.14s</td>
<td>130 (7)</td>
<td>50.24s</td>
<td>119 (7)</td>
<td>55.30s</td>
</tr>
<tr>
<td>cryptorsa</td>
<td>58M</td>
<td>9</td>
<td>11 (5)</td>
<td>3m4s</td>
<td>11 (5)</td>
<td>6m35s</td>
<td>TO</td>
<td>TO</td>
<td>35 (7)</td>
<td>9m42s</td>
</tr>
<tr>
<td>xalan</td>
<td>122M</td>
<td>7</td>
<td>31 (10)</td>
<td>0.15s</td>
<td>21 (7)</td>
<td>15m30s</td>
<td>TO</td>
<td>TO</td>
<td>37 (12)</td>
<td>10m44s</td>
</tr>
<tr>
<td>lufact</td>
<td>134M</td>
<td>5</td>
<td>21K (3)</td>
<td>7m26s</td>
<td>21K (3)</td>
<td>14m57s</td>
<td>TO</td>
<td>TO</td>
<td>21K (3)</td>
<td>10m38s</td>
</tr>
<tr>
<td>batak</td>
<td>157M</td>
<td>7</td>
<td>10 (2)</td>
<td>9m49s</td>
<td>10 (2)</td>
<td>22m56s</td>
<td>TO</td>
<td>TO</td>
<td>10 (2)</td>
<td>11m59s</td>
</tr>
<tr>
<td>lusearch</td>
<td>217M</td>
<td>8</td>
<td>232 (44)</td>
<td>12.63s</td>
<td>119 (27)</td>
<td>13m40s</td>
<td>232 (44)</td>
<td>27m9s</td>
<td>232 (44)</td>
<td>14m5s</td>
</tr>
<tr>
<td>tsp</td>
<td>307M</td>
<td>10</td>
<td>143 (6)</td>
<td>15m2s</td>
<td>140 (6)</td>
<td>29m10s</td>
<td>TO</td>
<td>TO</td>
<td>143 (6)</td>
<td>20m19s</td>
</tr>
<tr>
<td>luindex</td>
<td>397M</td>
<td>3</td>
<td>1 (1)</td>
<td>24m40s</td>
<td>2 (2)</td>
<td>31m6s</td>
<td>TO</td>
<td>TO</td>
<td>15 (15)</td>
<td>31m46s</td>
</tr>
<tr>
<td>sor</td>
<td>606M</td>
<td>5</td>
<td>0 (0)</td>
<td>38m38s</td>
<td>0 (0)</td>
<td>TO</td>
<td>TO</td>
<td>TO</td>
<td>0 (0)</td>
<td>44m36s</td>
</tr>
</tbody>
</table>

| Totals      | 2.0B   | -      | 29520 (157) | 1h51m    | 29133 (134) | ≥ 3h15m  | 1846 (131) | ≥ 8h30m  | 30862 (178) | 2h40m      |

Hence, sync-preservation is more adequate to capture not only racy program locations, but also racy memory locations. We note that, in principle, many different racy memory locations could correspond to the same static race (e.g., if memory is allocated dynamically). Note, however, that the additional reports of SyncP in Table 2 occur on benchmarks where it also makes more race reports in Table 1. Together, the two experimental tables give confidence that the new reported races are on entirely different variables.

**Race Distances.** We examine the capability of SyncP to detect races that are far apart in the input trace. Table 3 shows maximum race distance of races \((e_1, e_2)\) in various benchmarks, including the ones that contains sync-preserving races that are missed by happens-before. In each case, the distance is counted as the number of events in the input trace between \(e_1\) and \(e_2\), for every event \(e_2\) reported as racy. We see a sharp contrast between SHB/WCP and SyncP, with the latter being...
Table 2. Numbers of different memory locations that are detected as racy.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>SHB</th>
<th>WCP</th>
<th>SyncP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ftpserver</td>
<td>49</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>jigsaw</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>xalan</td>
<td>7</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>cryptorsa</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>lindice</td>
<td>1</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>sunflow</td>
<td>14</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>linkedlist</td>
<td>927</td>
<td>927</td>
<td>932</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1006</td>
<td>1002</td>
<td>1027</td>
</tr>
</tbody>
</table>

Table 3. Maximum race distances.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>SHB</th>
<th>WCP</th>
<th>SyncP</th>
</tr>
</thead>
<tbody>
<tr>
<td>tsp</td>
<td>11K</td>
<td>11K</td>
<td>224M</td>
</tr>
<tr>
<td>batik</td>
<td>1.7M</td>
<td>1.7M</td>
<td>4.8M</td>
</tr>
<tr>
<td>cryptorsa</td>
<td>7.9M</td>
<td>7.9M</td>
<td>8.3M</td>
</tr>
<tr>
<td>jigsaw</td>
<td>428</td>
<td>428</td>
<td>121K</td>
</tr>
<tr>
<td>sunflow</td>
<td>10M</td>
<td>10M</td>
<td>10M</td>
</tr>
<tr>
<td>xalan</td>
<td>4K</td>
<td>4K</td>
<td>13K</td>
</tr>
<tr>
<td>ftpserver</td>
<td>11K</td>
<td>11K</td>
<td>11K</td>
</tr>
<tr>
<td>lindice</td>
<td>165K</td>
<td>165K</td>
<td>165K</td>
</tr>
<tr>
<td>mergesort</td>
<td>57</td>
<td>57</td>
<td>1.4K</td>
</tr>
<tr>
<td>clean</td>
<td>355</td>
<td>47</td>
<td>1.2K</td>
</tr>
<tr>
<td>readswrites</td>
<td>13</td>
<td>13</td>
<td>696</td>
</tr>
<tr>
<td>wronglock</td>
<td>50</td>
<td>6</td>
<td>113</td>
</tr>
</tbody>
</table>

able to detect races that are far more distant in the input. This is in direct alignment with our theoretical observations already illustrated earlier in Section 1 (see Figure 1c). Indeed, as partial orders, SHB/WCP can only detect races between conflicting accesses that are successive in the input trace. On the other hand, sync-preserving races may be interleaved with arbitrarily many conflicting, non-racy accesses, and our complete algorithm SyncP is guaranteed to detect them. Overall, all our experimental observations suggest that sync-preservation is an elegant notion: it finely characterizes almost all races that are efficiently detectable, while it captures several races that are beyond the standard happens-before relation.

Table 4. Statistics of the core of the benchmark traces after the lightweight optimization is applied. \( \overline{N}, \overline{T}, \overline{A}, \overline{V} \) denotes respectively the number of events, threads, lock-acquire events, and variables in the core trace.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>( N )</th>
<th>( T )</th>
<th>( A )</th>
<th>( V )</th>
<th>Benchmark</th>
<th>( N )</th>
<th>( T )</th>
<th>( A )</th>
<th>( V )</th>
<th>Benchmark</th>
<th>( N )</th>
<th>( T )</th>
<th>( A )</th>
<th>( V )</th>
</tr>
</thead>
<tbody>
<tr>
<td>array</td>
<td>14</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>mergesort</td>
<td>170</td>
<td>6</td>
<td>49</td>
<td>1</td>
<td>jigsaw</td>
<td>30K</td>
<td>12</td>
<td>1.0K</td>
<td>51</td>
</tr>
<tr>
<td>critical</td>
<td>14</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>bubblesort</td>
<td>1.0K</td>
<td>13</td>
<td>119</td>
<td>25</td>
<td>sunflow</td>
<td>30K</td>
<td>17</td>
<td>585</td>
<td>20</td>
</tr>
<tr>
<td>account</td>
<td>18</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>lang</td>
<td>1.0K</td>
<td>8</td>
<td>0</td>
<td>100</td>
<td>cryptorsa</td>
<td>1.0M</td>
<td>9</td>
<td>156K</td>
<td>18</td>
</tr>
<tr>
<td>airtickets</td>
<td>27</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>readswrites</td>
<td>9.0K</td>
<td>6</td>
<td>1.0K</td>
<td>6</td>
<td>xalan</td>
<td>671K</td>
<td>7</td>
<td>183K</td>
<td>72</td>
</tr>
<tr>
<td>pingpong</td>
<td>38</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>raytracer</td>
<td>529</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>lufact</td>
<td>891K</td>
<td>5</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>twostage</td>
<td>86</td>
<td>13</td>
<td>20</td>
<td>2</td>
<td>bufwriter</td>
<td>10K</td>
<td>7</td>
<td>1.0K</td>
<td>6</td>
<td>batik</td>
<td>132</td>
<td>7</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>wronglock</td>
<td>125</td>
<td>23</td>
<td>20</td>
<td>1</td>
<td>ftpserver</td>
<td>17K</td>
<td>12</td>
<td>4.0K</td>
<td>135</td>
<td>lusearch</td>
<td>751K</td>
<td>8</td>
<td>53</td>
<td>77</td>
</tr>
<tr>
<td>bbuffer</td>
<td>13</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>moldyn</td>
<td>21K</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>tsp</td>
<td>15M</td>
<td>10</td>
<td>91</td>
<td>189</td>
</tr>
<tr>
<td>proccons</td>
<td>248</td>
<td>9</td>
<td>34</td>
<td>3</td>
<td>linkedlist</td>
<td>910K</td>
<td>13</td>
<td>1.0K</td>
<td>932</td>
<td>lindice</td>
<td>15K</td>
<td>3</td>
<td>6.0K</td>
<td>9</td>
</tr>
<tr>
<td>clean</td>
<td>871</td>
<td>10</td>
<td>239</td>
<td>2</td>
<td>derby</td>
<td>75K</td>
<td>5</td>
<td>21K</td>
<td>190</td>
<td>sor</td>
<td>1.0M</td>
<td>5</td>
<td>633K</td>
<td>4</td>
</tr>
</tbody>
</table>

Complexity of SyncP and Running Time. Recall the complexity of SyncP established in Theorem 3.1. We have argued that the complexity is \( \tilde{O}(N) \), i.e., \( \tilde{T}, \tilde{V} = \tilde{O}(1) \), meaning that the number of threads and variables are much smaller than \( N \). Here we justify this assumption experimentally, by presenting these numbers for the benchmark traces in Table 4. For each trace \( \sigma \), we report the parameters of the core trace \( \overline{\sigma} \) resulting from our lightweight optimization discussed earlier. We see that SyncP (and the other algorithms) is, in reality, executed on the core trace \( \overline{\sigma} \) where the number of threads \( \overline{T} \) and variables \( \overline{V} \) is indeed considerably smaller than \( \overline{N} \). Hence, our theoretical treatment of \( \overline{T}, \overline{V} = \tilde{O}(1) \) is justified.
7 RELATED WORK

The happens-before (HB) partial order, computable in linear time [Fidge 1991; Mattern 1988], is the basis of many race detectors [Bond et al. 2010; Christiaens and Bosschere 2001; Flanagan and Freund 2009; Pozniansky and Schuster 2003; Schonberg 1989]. However, it can miss many predictable races. Recent work improves upon this with a small increase of computational resources [Genç et al. 2019; Kini et al. 2017; Pavlogiannis 2019; Roemer et al. 2018, 2020; Smaragdakis et al. 2012].

Another common approach to race prediction is via lockset-based methods. At a high level, a lockset is a set of locks that guards all accesses to a given memory location. Such techniques report races when they discover a write access to a location which is not consistently protected (i.e., whose lockset is empty). They were introduced in [Dinning and Schonberg 1991] and equipped in Eraser [Savage et al. 1997]. The lockset criterion is complete but unsound, and various works attempt to reduce false positives by enhancements such as random testing [Sen 2008] and static analysis [Choi et al. 2002; von Praun and Gross 2001]. Locksets have also been combined with happens-before techniques [Elmas et al. 2007; Yu et al. 2005].

Another direction to dynamic race prediction is symbolic techniques that typically rely on SAT/SMT encodings of the condition of a correct reordering, and dispatch such encodings to the respective solver [Huang et al. 2014; Liu et al. 2016; Said et al. 2011; Wang et al. 2009]. The encodings are typically sound and complete in theory, but the solution takes exponential time. In practice, windowing techniques are used to fragment the trace into chunks and analyze each chunk independently. This introduces incompleteness, as races between events of different chunks are naturally missed. Dynamic techniques have also been used for predicting other types of errors, such as deadlocks, atomicity violations and synchronization errors [Chen et al. 2008; Farzan and Madhusudan 2009; Farzan et al. 2009; Flanagan et al. 2008; Kalhauge and Palsberg 2018; Mathur and Viswanathan 2020; Sen et al. 2005; Sorrentino et al. 2010].

8 CONCLUSION

In this work, we have introduced the new notion of synchronization-preserving races. Conceptually, this is a completion of the principle behind happens-before races, namely that such races can be witnessed without reversing the order in which synchronization operations are observed. We have shown that sync-preservation strictly subsumes happens-before, and can detect races that are far apart in the input trace. We have developed an algorithm SyncP that is sound and complete for sync-preserving races, and has nearly linear time and space complexity. In addition, we have shown that relaxing our definition even slightly, i.e., by allowing a single synchronization reversal suffices to make the problem W[1]-hard. Finally, we have performed an extensive experimental evaluation of SyncP. Our experiments show that sync-preservation is an elegant notion that characterizes almost all races that are efficiently detectable, while it captures several races that are beyond the standard happens-before. Given the demonstrated relevance of this new notion, we identify as important future work the development of more efficient race detectors for sync-preserving races, in a similar manner that happens-before race detectors have been refined over the years.

ACKNOWLEDGMENTS

We thank anonymous reviewers for their constructive feedback on an earlier draft of this manuscript. Umang Mathur is partially supported by a Google PhD Fellowship. Mahesh Viswanathan is partially supported by grants NSF SHF 1901069 and NSF CCF 2007428.
REFERENCES


