Detecting JavaScript Races that Matter

Erdal Mutlu
Koç University

Serdar Tasiran
Koç University

Benjamin Livshits
Microsoft Research

Abstract

As JavaScript has become virtually omnipresent as the language for programming large and complex web applications in the last several years, we have seen an increase in interest in finding data races in client-side JavaScript. While JavaScript execution is single-threaded, there is still enough potential for data races, created largely by the non-determinism of the scheduler. Recently, several academic efforts have explored both static and runtime analysis approaches in an effort to find data races. However, despite this, we have not seen these analysis techniques deployed in practice and we have only seen scarce evidence that developers find and fix bugs related to data races in JavaScript.

In this paper we argue for a different formulation of what it means to have a data race in a JavaScript application and distinguish between benign and harmful races, affecting persistent browser or server state. We further argue that while benign races — the subject of the majority of prior work — do exist, harmful races are exceedingly rare in practice (19 harmful vs. 621 benign). Our results shed a new light on the issues of data race prevalence and importance.

To find races, we also propose a novel lightweight runtime symbolic exploration algorithm for finding races in traces of runtime execution. Our algorithm eschews schedule exploration in favor of smaller runtime overheads and thus can be used by beta testers or in crowd-sourced testing. In our experiments on 26 sites, we demonstrate that benign races are considerably more common than harmful ones.

1. Introduction

When it comes to client-side Web programming, today JavaScript powers the majority of large and popular Web sites. JavaScript execution is single-threaded. Yet the complex needs of sites such as Facebook, Outlook, Google Maps, and the like have led to asynchrony becoming a common way to program complex Web applications. It is asynchronous processing that is responsible for interactive and responsive user interfaces (UIs) that operate without blocking the UI thread or requiring a reload, as did web applications of the late 1990s. Despite JavaScript lacking conventional threads, the presence of asynchrony creates a potential for races. In particular, the ordering of event execution in JavaScript as well as the timing of completions of asynchronous requests is non-deterministic, prone to be affected by network delays, etc., resulting in data races.

In this paper we argue that one should distinguish between races that have persistent consequences and those that are ephemeral. We argue that the majority of races in JavaScript have no persistent consequences — we dub these races benign. This is because these races only result in invisible to the user portions of the program state, or, at worst, UI glitches that are either unnoticed by the user, or disappear if the user reloads the page. Given the forgiving nature of JavaScript execution, where failures of individual event handlers force the scheduler to terminate the current event handler and move execution to the next one, these kinds of failures are not as important as previously believed. We suggest that a more useful way to think about data races on the web is by focusing on persistent state, such as client-local cookies, localstorage and persistent state, as well as server-based side effects. The latter are achieved via post calls to the server (get calls are designed for reads and are supposed to be idempotent, and thus are not frequently used for state updates). Our experiments confirm that the number of such harmful races is quite modest, yet it is these kinds of races that are more likely to be considered serious and fixed by developers. Previous research efforts tend to produce a large number of reports, despite the emphasis on suppressing false positives [10, 12, 17].

It is not our goal to provide a sound over-approximation of all possible races in a given application or site. Doing so is a very difficult task. Indeed, there are some fundamental limitations of program analysis for finding races in JavaScript. Static analysis is unlikely to be sound for large programs, given the presence of eval and complex language features, not to mention the fact that precise reasoning about happens-before relationship between event handlers statically is simply very difficult. Runtime analysis, including various forms of symbolic exploration of possible schedules fails victim to code coverage.

Instead, our goal is to provide a lightweight exploration algorithm that allows for the exploration of multiple schedules while only requiring a single run. This approach can be used for testing, including collaborative testing by a large number of beta- and crowd-sourced testers. A number of companies nowadays provide commercial crowd-sourced testing services for web sites (http://www.ute.com), which can be used as a cheap and practical way to increase coverage.

Contributions: We make the following contributions:

- We propose a novel view of benign and harmful data races in JavaScript web applications, and argue that harmful races should be the primary focus of analysis tools, due to them affecting the persistent client- or server-side state of the applications.
- We propose a lightweight exploration algorithm for finding data races in runtime traces of JavaScript programs. A key advantage for the scalability of our approach is that it does not require multiple program runs and can operate on the basis of a single execution.
- We find and investigate a total of 19 harmful and 621 benign races in 26 web sites, with only 2 observed false positives.
2. Overview

In the last several years, we have seen an upsurge of interest in data races in asynchronous programs and in JavaScript, more specifically. Researchers have tried applying static and runtime analysis to the problem of finding races. A fundamental challenge with static analysis for a language as dynamic as JavaScript is that it is really quite difficult to even enumerate all the relevant code, as much of the JavaScript code is produced with the help of eval calls and dynamic code loading. As a result, the traditional advantage of static analysis, namely, full path coverage largely does not apply to this problem. As such, the ability to make sound statements about the lack of data races is compromised [3, 5, 15, 17]. Additionally, reasoning about runtime events such as multiple XMLHttpRequests (XHRs), whose callbacks race against each other statically is fundamentally very difficult, because the happens-before relation is quite complex.

Runtime techniques in this space are also vulnerable to losing precision, which leads tool authors to develop heuristics to eliminate potential false positives [4, 8, 12]. They also suffer from the lack of coverage and the inability of making sound guarantees about the lack of races. Additionally, runtime techniques that involve combinatorial schedule exploration can run into scalability challenges, especially when the number of possible handlers to schedule is high [8, 12].

Our technique attempts to combine the advantages of static and runtime analysis. We execute the code only once, yet we explore multiple execution orders. As such, our technique scales well, while increasing the coverage of a single-pass runtime analysis. One way to see our approach is that in explores neighboring schedules for a particular runtime execution. We foresee this approach as being especially useful in the context of beta- or crowd-sourced testing: having a large number of users will naturally increase code coverage. At the same time, the users’ sessions will not be significantly slowed down. It should be noted that in the browser, slowing down the browser runtime runs the risk of modifying the behavior of timeout set with setTimeout and setInterval; additionally, the runtime may actively attempt to terminate slow-running events.

2.1 What is a Data Race?

Several possibly definitions of data races have been proposed for web applications [9, 10, 12]. They all center around the idea of writes to shared state that are performed by callbacks. Some of the races in prior work are caused by user interactions and browser-induced timing. In this work, our chief focus is on the XMLHttpRequest mechanism, which allows client-side code to request data from servers:

```javascript
<script>
var xhr = new XMLHttpRequest();
xhr.open("GET", "http://www.data.com/mydata.json");
xhr.onreadystatechange = function(e, d){
  state = 1;
};
xhr.send(null);
</script>
...a lot of text and images here...
<script>
state = 2;
</script>
```

Thirdly, and even more subtly, if the user opens the same site in multiple browser tabs, it is possible for these tabs to lead to concurrent execution. Two instances of the code below may race with each other when run in different tabs, resulting in a cookie-based race on line 5:

```javascript
<script>
var xhr = new XMLHttpRequest();
xhr.open("GET", "http://www.data.com/mydata.json");
xhr.onreadystatechange = function(e, d){
  state = 1;
};
xhr.send(null);
</script>
```

The happens-before relation for asynchronous callbacks is defined by the creation order. The XHR callback is preceded by the code that creates the XHR (xhr.send). A specific case of this is what we call nested (or chained) XHRs, when callbacks are defined one within another. Practically, this is about the only way for the developer to ensure that there is ordering of XHR callbacks, so we see this programming pattern quite a bit.

2.2 Motivating Examples

In an effort to understand the possible impact of data races on the web, we spent some time analyzing bug reports for open-source projects located on GitHub. Below we describe some of the examples of subtle server-side bugs from GitHub. In the interest of fairness we should mention that these examples of races reported as GitHub issues were not particularly common bugs for JavaScript projects, an intuition that is largely confirmed by our results in Section 4.

Example 1 [Old server state.] Issue #70 for the Wheaton-WHALE project\(^1\) describes the following situation:

1. The user reloads the page;
2. onbeforeunload listener fires, and the data is saved to server;
3. the page is loaded up again, and asks the server for the data;
4. the client-side JavaScript code loads up the old (outdated) data;
5. client state is saved to the server.

In the last step, the old, outdated data is saved to the server, essentially ignoring data updates. The culprit is the fact that steps 1 and 3 can race with each other: the data load request

\(^1\)https://github.com/WheatonWHALE/whaleweb/issues/79
may arrive before the save is processed. The implemented fix makes data updates synchronous.

**Example 2** [Racing for a user ID.] A somewhat similar situation that has to do with the issue of stale data obtained from the server is captured in issue #20 in a project called LikeLines\(^2\). LikeLines provides users with an in-browser video player with a navigable heat map of interesting regions for the videos they are watching. This case describes two racing XHRs that are issued to the backend server during initialization by the following functions: 1) `createSession` and 2) `aggregate`. The first call is to create a new session for recording user interactions. The second call is needed for drawing a heat map.

The problem arises when a user has not contacted the backend server before. In this case, both XHRs will be issued *without a cookie*, and in both cases the server will create a *new* user ID. This is clearly a problem because interaction sessions are tied to a user ID in this application. If the cookie from the call to `aggregate` “wins” (i.e., arrives last), then subsequent calls to the server will contain a user ID that does not match the interactions session.

In addition to the GitHub issues discussed above, below we list an illustrative example inspired by some of the samples from prior work [12, 17], although prior work did not focus on the issue of asynchronous XHRs.

**Example 3** [Racing with the browser.] Consider the code in Figure 1. For convenience, we mark every handler above with a number. The happens-before relation induced by this code example is as follows: 1 ← 2, 3 ← 4, 5 ← 6. As such, our exploration algorithm will consider the possibility 1 ← 2 ← 3 and 1 ← 3 ← 2. Similarly, because 2 and 5 are weakly ordered, traces in which 2 happens before or after 5 will be considered.

While one can explore these traces via a search in the schedule space, we choose to do so via data flow. We keep track of the event handlers that may be “concurrent” and mark the writes that they make to the same locations as weak writes.

---


---

Figure 1: Multiple XHR example.

```javascript
<html>
<script>
 var xhr = new Xhr();
 xhr.open('GET', false);
 xhr.onreadystatechange = function(){
   document.cookie = 'var1=1';
   xhr.send();
 //for(i=0;i<10000000;i++) console.trace(i);
 };
 ...

 //<!-- input id='mydiv' /> -->
</script>

<script>
 document.cookie = 'var1=2';
</script>

<script>
 var xhr2 = new Xhr();
 xhr2.open('GET', false);
 xhr2.onreadystatechange = function(){
   document.cookie = 'var1=3';
 };
 xhr2.send();
</script>
</html>

---

Figure 2: Sample trace illustrating cookie races.

Because 2 and 3 can race, the value of `document.cookie` will be either `var1 = 1` or `var1 = 2`. Similarly, for 2 and 4, the value of `document.cookie` will be either `var1 = 1` or `var1 = 5`. To preserve precision, our algorithm maintains existing happens-before relations such as those between 1 and 2 and 1 and 3.

---

2.3 Trace Processing

Figure 2 shows a simple trace obtained by running the code in Figure 1 that illustrates cookie-based races. We start processing the lines 1–4 where the XHR is opened and send to the remote server, we mark the XHR callback as an active callback which can be executed asynchronously any time in the future. As we process the first XHR callback on lines 6–10, we will record the write value made to cookie variable into the memory map where we store values for each variable id (i.e. `242159440` for cookie).

While recording the write value, we will look for values of the variable that are written by any callbacks that may be racing with each other, and complain about a race if there is any. As we continue processing the trace, we will record the value written on lines 12–14 by first checking the earlier values of `document.cookie`. As this sequential code segment can race with the earlier XHR callback (there is no happens-before edge), our processing will record a race on `document.cookie`, while adding a new value for the cookie into the memory map. As we continue to process the trace, a new XHR is opened and send to the server on lines 16–19 and added to the active callbacks list.

Later, the callback for the second XHR will be processed, when a write to the cookie is performed on lines 21–26. While processing the write operation, the values for `document.cookie` will be checked and a race will be recorded, as two XHR callbacks are marked as racing, resulting in different cookie values.

---

2.4 Algorithm Summary

Here we provide the underlying intuition for our approach, with a more formal treatment relegated to Section 3. The key idea behind our approach is to consider alternative scenarios within a given trace. We do not attempt to force exploration
of UI interactions, for example, however, we do explore the possibilities of different schedules that may occur because of the order of arrival of asynchronous handlers that are part of the trace.

Our approach effectively performs static analysis on a trace that is collected at runtime, as a way to consider different schedule orders. When considering multiple execution of XHR callbacks as shown in Figure 3, the key observation is that instead of separately considering each of the possible schedules, we can encode the effect of the possible race by merging the state and keeping track of multiple, merged values. This is analogous to doing meets in static dataflow or abstract interpretation-style analysis, as an alternative to a costly meet-over-all-paths (MOP) solution. Conceptually, given a merge point for variable $x$ with multiple values coming in for two racing XHRs, $x = v_1$ and $x = v_2$, we keep track of both values $\{v_1,v_2\}$; not of course that if $v_1 = v_2$, no need to keep two copies of the same values exists.

We formulate our race detection algorithm as a dataflow analysis on the values within a given execution trace. We flag a possible race if multiple values may flow to a sensitive location, indicating a presence of scheduling dependencies; these sensitive locations are persistent storage such as cookies, localStorage, sessionStorage, and, lastly, the DOM, etc. These latter locations serve as the sinks of our data flow analysis. Returning to the example in Figure 1, we can represent the race between handlers 2 and 5 as an assignment $\text{document.cookie} = \{\text{var1} = 2\text{'}, \text{var1} = 3\text{'}\}$ as a merge node after handler 5. This of course represents a direct flow of multiple values to a sensitive persistent storage location $\text{document.cookie}$. More interesting cases involve multiple steps of propagation.

2.5 Implementation

To collect execution traces, we have instrumented the most recent version of the Firefox web browser. Our changes span both the SpiderMonkey JavaScript engine to track data propagation through the memory of the browser as well as operations on cookies, localStorage, sessionStorage, and, the like, which are recorded by instrumenting the DOM. Our instrumentation spans over three main components of Firefox:

- **XPCOM (Cross Platform Component Object Model):** In order to record the triggered XHR callbacks, we instrumented the event queue in Firefox in `nsThread.cpp`. When events are taken from the queue for execution, we mark XHR `readystatechange` events as well as button clicks initiated by the user.
- **Gecko (Layout engine):** We also need to instrument DOM API implementation of Firefox for recording updates made to DOM elements of interest. We achieved this by modifying various DOM class implementations like `nsGlobalWindow.cpp`, `DOMStorage.cpp`, etc.
- **SpiderMonkey (JavaScript Interpreter):** Lastly, we instrumented the JavaScript interpreter for recording value manipulation on variables and objects and also to mark the start and end points of XHR callback execution. The former is achieved by instrumenting the JavaScript bytecode interpreter on `interpreter.cpp` and `jsapi.cpp`.

Overall, our instrumentation is quite sparse and we believe can be easily migrated to another open-source browser such as Chromium. We have added a total of about 430 lines of instrumentation code to Firefox to collect our traces. A total of 12 files were modified.

**Deployment strategies:** The process of race detection is something that can be performed both online, as the application is running, as well as offline, as an auditing step. We envision that as part of beta-testing, traces from multiple users can be analyzed. Note that as we will highlight in Section 4, even relatively simple-looking sites can create long traces with a large number of events. At the same time, the number of events relevant to asynchrony and scheduling is relatively small. Our analysis for finding potential races is implemented as a linear pass over the trace. However, if desired, this is something that can be parallelized as well, by splitting longer traces to be analyzed on different machines.

3. Formalization

Since JavaScript execution is single-threaded, the execution of event handlers and XHR callbacks is carried out non-preemptively, without interruption. We find it convenient to represent the executions of event handlers, XHR callbacks, and portions of script code executed without pre-emption ("sequential blocks") as execution blocks with unique IDs. Other entries in an execution trace will also be assigned unique IDs as detailed later.

There is no universally accepted definition of the happens-before relation for web-based JavaScript code. We define the happens-before relation (denoted $\leftarrow$) not at the level of low-level memory accesses, but at the level of higher-level language constructs, based on causality information we abstract from JavaScript operational semantics as was illustrated in Figure 1. Within each uninterrupted execution block, trace entries are ordered by the program, and, therefore, happens-before order. We define and record a happens-before order between blocks such that if $id \leftarrow id'$ then all trace entries in $id$ happen before those in $id'$. We order $id \leftarrow id'$ if $id'$ appears later in the trace and one of the following hold:

1. Both blocks are sequential blocks. Sequential blocks are ordered by $\leftarrow$ in the order they appear in the trace because of the browser-imposed ordering.
2. Both blocks are event-handling blocks. Event handling blocks are ordered by $\leftarrow$ by how they occur in the trace to reflect the order of user interactions.
3. $id'$ is an XHR callback, and its XHR `send` is within block $id$; this is because the XHR callback can only happen after the `send` operation.

According to the data in Figure 6, our analysis is fast enough to be run online, so the beta testers only need to use a differently compiled version of the browser or perhaps a browser with a flag that they turn on. The results of such exploration can be centrally collected and communicated to the site developers.

Note that these IDs are not to be confused with the statically-assigned numbers in Figure 1 as, for instance, the point in an execution where an XHR callback is registered and the point where it is executed are different.
3.1 Modeling and Analysis of Traces

The set of memory locations (simply referred to as “locations” from here on) manipulated by a JavaScript program is denoted by Locs, and the set of values they can take by Val. To denote the values of locations that have not yet been assigned a value, we use the symbol ⊥. Locs and use (Locs are of type key-value store and are treated specially. These elements and are written to using the setter for the each other. The set of locations that represent DOM elements and are there due to “concurrent” execution blocks and are therefore not removed.

3.2 Defining Traces

Formally, a trace is a finite sequence of trace entries \( (\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_n) \) of the following types:

- **XHR, event handling, and sequential blocks:** Trace entries \( \lambda = CBBegin(id) \) and \( \lambda = CBEnd(id) \) denote the beginning and end of the execution of the callback for the XHR with ID \( id \). \( \lambda = HandlerBegin(id) \) and \( \lambda = HandlerEnd(id) \) do the same for event handler blocks, and \( \lambda = SeqBegin(id) \) and \( \lambda = SeqEnd(id) \) for sequential blocks.

- **Key, location, innerHtml accesses** \( \lambda = key\text{Wr}(kv, ky, vl, varsRd, id) \) denotes the writing of the value \( vl \) for the key \( ky \) in the key-value store \( kv \) within the block with ID \( id \). The value \( vl \) has been computed immediately prior to the write trace entry as the result of an expression over the memory locations \( varsRd \). \( \lambda = key\text{Rm}(kv, ky, id) \) denotes the removal of the value \( ky \) from \( kv \) within block \( id \). \( \lambda = var\text{Write}(v, vl, id) \) denotes the writing of the value \( vl \) to the location \( v \) within block \( id \). Finally, \( \lambda = set\text{HTML}(hElt, hVal, id) \) denotes the setting of the innerHtml property of a DOM element \( hElt \) to value \( hVal \) within block \( id \).

- **POST, XHR send:** \( \lambda = post(url, id, id_{in}, vl, varsRd) \) is a POST request XHR call or Window with id \( id \) and the call occurs within an execution block with ID \( id_{in} \). The data posted \( vl \) is the computed result of an expression over the memory locations \( varsRd \). \( \lambda = xhr\text{Send}(id, id_{in}) \) denotes an XHR operation for the XHR object with ID \( id \) that takes place within an execution block with ID \( id_{in} \). This \text{send} is a GET request.

\[\text{Seq-Blk-Begin} \quad \lambda = \text{SeqBegin}(id) \quad \text{HB}' = \text{TransClose}(\text{HB} \cup \{(id_{seq}, id_{ext})\})\]

\[\text{Seq-Blk-END} \quad \lambda = \text{SeqEnd}(id) \quad id_{seq}' = id\]

\[\text{XHR-POST} \quad \lambda = \text{post}(url, id, id_{in}, vl, varsRd) \quad P' = P \cup \{(v, id)v \in varsRd\} \quad \text{HB}' = \text{HB} \cup \{(id_{in}, id)\}\]

\[\text{XHR-SEND} \quad \lambda = xhr\text{Send}(id, id_{in}) \quad \text{HB}' = \text{TransClose}(\text{HB} \cup \{(id_{seq}, id_{ext})\})\]

\[\text{KEY-WRITE} \quad \lambda = \text{keyWrite}(kv, ky, vl, varsRd, id) \quad V' = [\{(kv, ky) := \text{keyWrite}(kv, ky, vl, varsRd, id)\} \cup \{(v, id)\} \cup \{(‘v’, ‘id’)|id’ \leftrightarrow id \lor id’ \lor (v, id)\}]
\]

\[\text{WRITE} \quad \lambda = \text{varWrite}(v, vl, id) \quad V' = [\{v := \text{varWrite}(v, vl, id)\} \cup \{(k, v)|k \neq \text{keyWrite}(kv, ky, vl, varsRd, id)\}]
\]

\[\text{SET-DOM} \quad \lambda = \text{setHTML}(hElt, hVal, id) \quad V' = [\{hElt := \text{setHTML}(hElt, hVal, id)\} \cup \{(v, id)\} \cup \{(v, ‘id’)|id’ \leftrightarrow id \lor id’\}]
\]

\[\text{XHR-SEND} \quad \lambda = \text{xhrSend}(id, id_{in}) \quad \text{HB}' = \text{TransClose}(\text{HB} \cup \{(id_{in}, id)\})\]

Figure 4: Trace analysis rules.

3.3 Interpreting Traces

Given a trace \( \text{Trace} = (\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_n) \), our race detection algorithm analyzes it by processing it sequentially, one log entry at a time. The algorithm maintains analysis state represented by the tuple \( \Sigma = (V, \text{HB}, P, id_{seq}, id_{ext}) \). Here, \( V \) is the memory map. \( \text{HB} \) is the happens-before relation, which is a partial order over IDs. Whenever new elements are added to \( \text{HB} \) by a trace processing rule, the transitive closure of the relation is taken to obtain the resultant \( \text{HB} \). \( \text{HB} \) is initially the empty relation. \( P \) is a list of pairs of the form \( (v, id) \) where the location \( v \) has been read while computing the value submitted by a XHR POST request with ID \( id \). Finally, \( id_{seq} \) and \( id_{ext} \) are the IDs of the last sequential block processed or the last event handling block processed by our algorithm, respectively, or \( \bot \) if no such callback or block exists. Given a state \( \Sigma = (V, \text{HB}, P, id_{seq}, id_{ext}) \) reached at a point in the trace, the analysis state reached after processing the log entry is described by the rules given in Figure 4 and explained below.
Callbacks: CB-BEGIN and CB-End (not shown) keep track of the ID of the ongoing XHR callback block. Seq-BLK-BEGIN processes the log entry indicating the beginning of a sequential block with ID id by setting idSeq to id. Seq-BLK-END resets idSeq to ⊥. The rules ensure that the occurrence order of sequential blocks in the trace and their happens-before order coincide. EVT-HANDLER-BEGIN and EVT-HANDLER-END operate similarly to the corresponding Seq-BLK rules. The order of occurrence in the trace of the event handlers is the same as their happens-before order. Event handlers and sequential blocks are not ordered with respect to each other by the happens-before relation.

Location updates: Key-Write handles the case where key ky in the key-value map kv is updated. We declare a potentially-harmful race if |Val(v)| > 1 any location v read while the new value for the key is being computed (v ∈ varsRd) indicating the potential for non-determinism. The rule computes V′(kv, ky) by from V(kv, ky) by adding the pair (vl, id), removing all pairs (vl, ’id’) such that id’ ← id. Key-REMOVE (not shown) writes the value ⊥_vl to the key ky. WRITE and SET-DOM are similar to Key-Write above, updating V for a location v or a DOM element hElT by removing value-Id pairs overwritten as dictated by the happens-before relation and adding the new value-Id pair written by the current trace entry.

POST, send: The rule XHR-POST declares a potentially harmful race if |V(v)| > 1 for any location v ∈ varsRd, since at least one location used in the computation of vl the data posted, has the potential for non-determinism.XHR-SEND updates the happens-before relation. xhrSend(id, idn) indicates that the send for the XHR with id id has taken place in block idn, idn ← id. This rule and the fact that CB-BEGIN and CB-END do not modify the happens-before relation ensure the fact that only chained XHR calls are ordered by ← with respect to each other5.

3.4 Detecting Races
We say V(v) has non-determinism potential on a location v when the memory map contains at least two different values for v, i.e., when there are pairs (vl, id) and (vl’, id’) in V(v), such that vl ≠ vl’ and id ≠ id’. Our algorithm declares a race on a location v while evaluating the WRITE, SET-DOM, and Key-WRITE rules if the memory map V′(v) computed by the rule has non-determinism potential on v. We also declare a race when evaluating the XHR-POST and Key-WRITE rules, if a variable read when computing the value posted or written, v ∈ varsRd, V(v) has non-determinism potential for v. Of these races, only the ones associated with KEY-WRITE and XHR-POST rules are deemed to be harmful races.

Consider a prefix of the trace in a “quiet” state such that, at the end of the prefix, no XHR callback or execution block is in progress. Suppose that our algorithm has signaled non-determinism potential on a location v while processing a trace entry λ within the last block or callback with id id in this prefix. Then a different re-ordering of the execution blocks and/or XHR callbacks while leaving the happens-before relation in the prefix intact may result in a different final value for v, as explained next.

---

5 Note that it is possible to define the happens-before relationship differently, for instance, declaring xhr to happen before xhr_j if the send entry for xhr_j appears later in the trace than the end of the callback for xhr_i. While such a definition may capture the happens-before relationship observed in a particular schedule more precisely, our definition narrows in on ordering relationships enforced by the JavaScript semantics and excludes those that may have taken place differently in different executions of the same program driven by the same user interaction.
Let \((vl, id)\) and \((vl', id')\) be in \(Val(v)\) such that \(id' \neq id\) and \(vl \neq vl'\). It must be the case that while both \(id'\) and \(id\) wrote to \(v\), these two execution blocks are not ordered by the happens-before relationship. Otherwise, either \((vl, id)\) or \((vl', id')\) would have been removed from \(Val(v)\), according to the update rules of our algorithm and how they use the happens-before relation in variable updates. Therefore, it is possible to modify the execution by delaying the execution of the block with \(id'\) until \(after\) the execution of the block with \(id\). In this case, the final value of the location \(v\) or \((kv, ky)\) would be different at the final state of the newly-obtained execution. This points to a potentially different result produced purely as a result of XHR callback scheduling non-determinism.

There are two sources of false positives in our race detection approach. The first source is the assumption made in the argument in the previous paragraph while obtaining a new execution by delaying the execution of the block with ID \(id'\) past other blocks. The assumption is that control decisions made in the original execution based on data values written by the block \(id\) are not modified in a way that makes the reordered execution infeasible. In our empirical experience, such cases are rare and can be ruled out by inspection or replaying and validating the reordered execution. The second source is the check performed when a value is written to a persisted location or sent on the network. In these cases, if, while computing the written or sent value, a location with non-determinism is read, our algorithm signals a potentially harmful race. This approach is conservative, i.e., non-determinism in the value of one of the locations in an expression may not result in non-determinism in the value of the result. Even in cases where this is the cause of a false alarm, we believe that the potentially different data values flowing to a persisted output location may be of concern to programmers.

Let \((vl, id)\) and \((vl', id')\) be in \(Val(v)\) such that \(id' \neq id\) and \(vl \neq vl'\). It must be the case that while both \(id'\) and \(id\) wrote to \(v\), these two execution blocks are not ordered by the happens-before relationship. Otherwise, either \((vl, id)\) or \((vl', id')\) would have been removed from \(Val(v)\), according to the update rules of our algorithm and how they use the happens-before relation in variable updates. Therefore, it is possible to modify the execution by delaying the execution of the block with \(id'\) until \(after\) the execution of the block with \(id\). In this case, the final value of the location \(v\) or \((kv, ky)\) would be different at the final state of the newly-obtained execution. This points to a potentially different result produced purely as a result of XHR callback scheduling non-determinism.

There are two sources of false positives in our race detection approach. The first source is the assumption made in the argument in the previous paragraph while obtaining a new execution by delaying the execution of the block with ID \(id'\) past other blocks. The assumption is that control decisions made in the original execution based on data values written by the block \(id\) are not modified in a way that makes the reordered execution infeasible. In our empirical experience, such cases are rare and can be ruled out by inspection or replaying and validating the reordered execution. The second source is the check performed when a value is written to a persisted location or sent on the network. In these cases, if, while computing the written or sent value, a location with non-determinism is read, our algorithm signals a potentially harmful race. This approach is conservative, i.e., non-determinism in the value of one of the locations in an expression may not result in non-determinism in the value of the result. Even in cases where this is the cause of a false alarm, we believe that the potentially different data values flowing to a persisted output location may be of concern to programmers.

Let \((vl, id)\) and \((vl', id')\) be in \(Val(v)\) such that \(id' \neq id\) and \(vl \neq vl'\). It must be the case that while both \(id'\) and \(id\) wrote to \(v\), these two execution blocks are not ordered by the happens-before relationship. Otherwise, either \((vl, id)\) or \((vl', id')\) would have been removed from \(Val(v)\), according to the update rules of our algorithm and how they use the happens-before relation in variable updates. Therefore, it is possible to modify the execution by delaying the execution of the block with \(id'\) until \(after\) the execution of the block with \(id\). In this case, the final value of the location \(v\) or \((kv, ky)\) would be different at the final state of the newly-obtained execution. This points to a potentially different result produced purely as a result of XHR callback scheduling non-determinism.

There are two sources of false positives in our race detection approach. The first source is the assumption made in the argument in the previous paragraph while obtaining a new execution by delaying the execution of the block with ID \(id'\) past other blocks. The assumption is that control decisions made in the original execution based on data values written by the block \(id\) are not modified in a way that makes the reordered execution infeasible. In our empirical experience, such cases are rare and can be ruled out by inspection or replaying and validating the reordered execution. The second source is the check performed when a value is written to a persisted location or sent on the network. In these cases, if, while computing the written or sent value, a location with non-determinism is read, our algorithm signals a potentially harmful race. This approach is conservative, i.e., non-determinism in the value of one of the locations in an expression may not result in non-determinism in the value of the result. Even in cases where this is the cause of a false alarm, we believe that the potentially different data values flowing to a persisted output location may be of concern to programmers.

Let \((vl, id)\) and \((vl', id')\) be in \(Val(v)\) such that \(id' \neq id\) and \(vl \neq vl'\). It must be the case that while both \(id'\) and \(id\) wrote to \(v\), these two execution blocks are not ordered by the happens-before relationship. Otherwise, either \((vl, id)\) or \((vl', id')\) would have been removed from \(Val(v)\), according to the update rules of our algorithm and how they use the happens-before relation in variable updates. Therefore, it is possible to modify the execution by delaying the execution of the block with \(id'\) until \(after\) the execution of the block with \(id\). In this case, the final value of the location \(v\) or \((kv, ky)\) would be different at the final state of the newly-obtained execution. This points to a potentially different result produced purely as a result of XHR callback scheduling non-determinism.

There are two sources of false positives in our race detection approach. The first source is the assumption made in the argument in the previous paragraph while obtaining a new execution by delaying the execution of the block with ID \(id'\) past other blocks. The assumption is that control decisions made in the original execution based on data values written by the block \(id\) are not modified in a way that makes the reordered execution infeasible. In our empirical experience, such cases are rare and can be ruled out by inspection or replaying and validating the reordered execution. The second source is the check performed when a value is written to a persisted location or sent on the network. In these cases, if, while computing the written or sent value, a location with non-determinism is read, our algorithm signals a potentially harmful race. This approach is conservative, i.e., non-determinism in the value of one of the locations in an expression may not result in non-determinism in the value of the result. Even in cases where this is the cause of a false alarm, we believe that the potentially different data values flowing to a persisted output location may be of concern to programmers.
Trace statistics: The workload used to collect these counts was simply loading the page and applying basic user interactions like button-link clicks. The counts for DOM manipulation are generally higher than those for cookies, localStorage, or sessionStorage. It is natural to expect more races on DOM elements as well, compared to more uncommonly used persistent elements. Figure 6 summarizes information about the traces we used for our analysis. The compression ratios range between 9.97 and 33.83. The percentage of time it takes to analyze a trace compared to the time to record a trace ranges between 3% and 32%.

Detection Time: Figure 6 shows the trace collection and analysis time as a function of trace size. Analysis time grows approximately linearly with the size of the trace and is a fraction of the trace recording time shown in the second column. Since trace analysis can be parallelized or run on an unused core, we can envision this analysis being run in parallel with execution. Among other advantages, this would obviate the need for trace storage and transfer.

4.2 Detection Results
In this section we describe and analyze several representative races found with our approach.

Example 4 [sessionStorage in milliyet.com.tr] In the case of www.milliyet.com.tr, the race on sessionStorage was caused by a shared variable namespace that is used for generating the key name. The variable is written at two different locations, one being an XHR callback that is executed using jQuery.ajax method. The execution steps are listed below and the relevant code is shown in Figure 8:

1. As a page loads, an XHR is created using the jQuery.ajax method (included in www.milliyet.com.tr/D/j/base.js?v=20) and sent to the server (lines 3–5):
   ```javascript
   <script>
   ...;
   </script>
   2. once the response is received a user-defined callback is executed using jQuery.ajax.done (lines 19–22):
   ```javascript
   function done(...)
   ...;
   </script>
   3. namespace variable is set to empty string at the end of jQuery.ajax.done method (line 21);
   4. an external library is initialized by setting the namespace variable to '_uv_' (line 46):
   ```javascript
   namespace = "__uv_"; // write to sessionStorage
   </script>
   5. namespace variable is subsequently used for generating a sessionStorage key (lines 43–45).

   In the last step, the value of namespace is used for setting an item on the sessionStorage for recording user interactions with the web page. In this particular trace, we observe that multiple items are added to the sessionStorage using the namespace as a prefix (i.e., '_uv_autoprompt_disabled._uv_r'). Any future read operations on the sessionStorage keys will depend on this prefix namespace. As the XHR callback can race with the namespace writes which sets the value for namespace to the empty string, it may result in a failed read operation. □

Example 5 [innerHTML in radikal.com.tr] Another race-inducing execution trace is obtained from www.radikal.com.tr, where two XHR callbacks race with each other when it comes to writing to innerHtml property of a <div> block in the DOM. In this news web site, users can filter displayed news according to their preferred categories, by using category links on the page. Each click action on a category link will generate an XHR call to server to gather news for the corresponding category. The race is captured within the execution steps listed below; the corresponding code is shown in Figure 9:

1. User clicks on a category link (lines 2–3) for enabling a category on the web page triggering the AddCategory method (lines 8–11);
2. AddCategory method will than trigger GetArticle method (line 10);
3. GetArticle method (lines 12–15) makes an XHR call using jQuery.load for <div> with id r5_content_upd (line 14);
4. user clicks another category link (lines 4–5) triggering another XHR call;
5. both XHR calls sets innerHtml of the same <div> on the DOM (line 8).

---

6 Sample traces can be found at pastebin.com/MxF7ENP8, pastebin.com/Je5Y4FY8 and pastebin.com/lPZe5Ugb.
At the end of this execution, the displayed DOM will depend on the order of the responses returned from the server as the `<div id=r5_contentupd>` is directly affected by the asynchronous XHR callback.

Figure 9: Race on `innerHTML` manipulation in `radikal.com.tr`. To save space, we remove unrelated lines of code.

Example 6 [document.cookie in `gazetta.it`]. In this example of a `document.cookie` race from `www.gazetta.it`, the web site uses an application monitoring library (DynaTrace Real User Monitoring found at dynatrace.com) for recording and POSTing user actions and browsing experience to a remote server. Each POST request initiated by this library updates the key `dtCookie` of the `document.cookie` with the value from the server via XHR callback.

Figure 10: Race on `document.cookie` manipulation in `gazetta.it`. To save space, we remove unrelated lines of code.

Figure 11: Trace from `optimum.net` illustrating a false positive. A race occurs when multiple XHR calls try to update the new response value to `document.cookie`. The execution steps are listed below, with line references from the code shown in Figure 10:

1. The page load time, the XHR call is created for initializing the monitoring process (line 9);
2. The data is sent to a URL constructed from `dtCookie` value (lines 6–8);
3. `dtCookie` key is updated with the server’s response (lines 35–37);
4. A new XHR is created at `onLoad` event for posting loading time of the page (lines 11–13);
5. Callbacks of each XHR try to write to the same `document.cookie` key `dtCookie` (lines 35–37).

The value of `dtCookie` key depends on the order of the XHR callbacks, causing possible issues with future POST operations, as the value is used for URL generation.

4.3 False Positives

In looking for possible false positives, we have decided to focus our attention on both persistent races (19) and those on the DOM state (47). Out of the 66 races we investigated, we identified only two cases as false positives. These are cookie races in `fedex.com` and `optimum.net`.

Figure 11 shows the trace from `optimum.net`, with writes to the cookie key `fars` at multiple locations (lines 6, 10, and 22), one of which is within an XHR callback. In this case, `fars` is used for collecting site statistics where the key value is updated by concatenating new values (lines 11–22). Although the XHR callback may happen in between two cookie writes on lines 5 and 6, causing a different value for the key, at the end of the execution the cookie key `fars` will contain all the written values but in different orders. In this case, the program treats this value as a set and not a list, so, while we catch the nondeterminism correctly, this is not really a bug in the program. The false positive in `fedex.com` is very similar, with a race is on cookie key `s_sess`, updated within an XHR callback.

Benign memory races: We would also like to highlight an example of a memory race that does not have persistent consequences. On `radikal.com.tr`, there is a global variable "duration" that is used to measure time elapsed at different points throughout the execution on which there is a race...
between its update in sequential code and in an XHR callback. The value of this memory location does not propagate to any persisted state and cause non-determinism.

4.4 Discussion

Returning to the research questions in Section 4, we clearly see that according to Figure 7, races on persistent state (RQ 1) are quite uncommon, with only 19 for 26 sites. Races on session state (RQ 2) are also uncommon (only 3 races observed), in part because session state is used not as frequently as cookies (Figure 6). Lastly, races on transient state are considerably more frequent. While changes to HTML content may technically be races, one could argue that most of these are ephemeral. Indeed, the execution model of JavaScript-based web pages induces a great deal of non-determinism. This is because many if not most web pages are not the same across multiple reloads: the ads on the sides of the page change; content of pages often changes (consider a rapidly changing news site), sometimes the experience of the first visit to a page and subsequent ones is different because of cookies. The browser user has been trained over the last decade not to expect much consistency and, when everything else fails, to reload the page. Additionally, the JavaScript execution model is extremely permissive: errors are “swallowed” by the runtime (the current event handler is terminated) and in many cases pages can survive exceptions and keep on running.

Our observations mesh well with the anecdotal experience of the user encountering any problem on the web: one only needs to reload the page for the problem to go away. In a sense, web programming is very forgiving. This is different from thread-related races in desktop applications and also data races in mobile apps [4, 8], where researchers are able to replicate races with obvious visual consequences (mangled or upside down images on the screen, for instance).

5. Related Work

This section covers some of the recent work on finding races in JavaScript programs and asynchronous code.

Races in Asynchronous Programs: Hsiao et al. propose an approach to finding a subset of asynchronous races in Android apps, focusing on races that lead to use-after-free violations (i.e., uses of a freed pointer) [4]. While this tool also does offline analysis of a single execution trace, their focus is on computing an explicit happens-before relation, which they apply to the trace in order to find accesses that may lead to user-after-free possibilities. They employ some heuristics to minimize the possibility of false positives. We focus on data flow from multiple values to sensitive locations (like `document.cookie`), an approach that naturally eliminates the need to explicitly reason about commutativity.

Maiya et al. [8] et al. focus on a systematic exploration of possible schedules using a UI explorer and reasoning about the obtained traces using a race detector. A precise model of the Android execution life cycle is key to avoiding false positives, although a large number of these remain.

Races in JavaScript: Zheng et al. [17] propose a static technique to detect potential races in JavaScript applications. More recently, Petrov et al. [10] and Raychev et al. [12] have observed the potential for asynchrony creating out-of-order execution and developed a notion of race conditions for Web applications written in JavaScript. In principle, race conditions can arise because of accesses to data shared among components of a Web page which are not ordered by proper synchronization, or, more formally, a happens-before relation. Of course, on a Web page, the entire DOM is (a giant blob of) global state, creating the potential for races.

Petrov et al. [10] define a happens-before relation for Web pages and generalize the notion of race conditions to take into account cases where, logically, there are unordered accesses to the same resource. The authors present a dynamic method for detecting races in a given execution of a Web page, explore similar executions that could potentially be racy, and, in later work [12] identify and filter out large sets of benign races. As discussed earlier, our work is distinguished from these studies by the fact that we only pursue race conditions that lead to non-determinism in persisted state or data sent to the server. Our choice of the happens-before relationship follows from this design decision and only records high-level causality relationships.

Muthu et al. [9] advocate the notion of observable races, i.e., those that can be seen and visually distinguished by the end user. Our notion of persistent side effects is even stronger than that captured in this paper.

Program Analysis in JavaScript: A number of analyses have been propose for JavaScript in recent years; here we highlight only a handful. Additionally, several aspects of the language such as the use of `eval` [5, 14] and trying to understand `XMLHttpRequest` performance [11, 13], Rozzle [6] proposes the idea of lightweight multi-execution in the context of a JavaScript engine, similar to our work. The goal of Rozzle is to expand the impact of malware detectors by increasing code coverage and thereby observing more, potentially malicious, code. In terms of techniques, Rozzle is probably the closest runtime exploration approach to the work described in this paper. A project by Chugh et al. focuses on staged analysis of JavaScript and finding information flow violations in client-side code [1]. The Gatekeeper project [2, 3] proposes a points-to analysis together with a range of queries for security and reliability as well as support for incremental code loading. Gulfstream [3] is a successor of the Gatekeeper project whose focus is on incremental analysis and dynamic code loading. Sridharan et al. [16] presents a technique for tracking correlations between dynamically computed property names in JavaScript programs. Their technique allows them to reason precisely about properties that are copied from one object to another as is often the case in libraries such as jQuery. Madsen et al. [7] proposes the idea of a `use analysis` for the purposes of call graph construction in JavaScript applications that use large frameworks and libraries. Their use analysis is combined with a points-to analysis for the rest of the application.

6. Conclusions

This paper proposes an alternative way of looking at what constitutes a race in web applications written in JavaScript. We advocate a focus on races that are caused by asynchronous callbacks and their order of arrival, primarily investigating races produced by the `XMLHttpRequest` (XHR) mechanism. Unlike prior work which concluded that there is ample potential for races in JavaScript, our findings suggest that given the forgiving nature of JavaScript applications, damaging, persistent races are considerably more rare. Nevertheless, we propose a lightweight algorithm that explores different schedules in the “neighborhood” of a particular runtime trace. Our approach avoids the imprecision of static analysis and the combinatorial explosion and scalability issues of runtime schedule exploration. We find and investigate a total of 19 harmful races and 621 benign races in 26 web sites.
References


