Towards a Verified Range Analysis for JavaScript JITs

Fraser Brown  
Stanford, USA

John Renner  
UC San Diego, USA

Andres Nötzli  
Stanford, USA

Sorin Lerner  
UC San Diego, USA

Hovav Shacham  
UT Austin, USA

Deian Stefan  
UC San Diego, USA

Abstract

We present VeRA, a system for verifying the range analysis pass in browser just-in-time (JIT) compilers. Browser developers write range analysis routines in a subset of C++, and verification developers write infrastructure to verify custom analysis properties. Then, VeRA automatically verifies the range analysis routines, which browser developers can integrate directly into the JIT. We use VeRA to translate and verify Firefox range analysis routines, and it detects a new, confirmed bug that has existed in the browser for six years.

CCS Concepts: • Security and privacy → Browser security; • Software and its engineering → Just-in-time compilers; Software verification and validation; Domain specific languages.

Keywords: JIT verification; range analysis; JavaScript

ACM Reference Format:

1 Introduction

On May 30, 2019, employees of the cryptocurrency startup Coinbase were targeted by a phishing campaign that lured them to visit a Web page hosting attack code [88]. This code exploited previously unknown bugs in Firefox to take over the victim’s machine; the first bug arose from incorrect type deduction in the just-in-time (JIT) compiler component of Firefox’s JavaScript engine [15].

Earlier this year, Google’s Threat Analysis Group identified websites, apparently aimed at people “born in a certain geographic region” and “part of a certain ethnic group,” that would install a malicious spyware implant on any iPhone used to visit them. Two bugs exploited in this campaign, according to analysis by Google’s Project Zero [41, 68], were in the JIT component of Safari’s JavaScript engine [5, 34].

The JavaScript JITs shipped in modern browsers are mature, sophisticated systems developed by compilers experts. Yet bugs in JIT compilers have emerged in recent months as the single largest threat to Web platform security, and the most dangerous attack surface of Web-connected devices.

Unlike other compilers, browser JITs are exposed to adversarial program input. Remote attackers can craft JavaScript that will trigger JIT compiler bugs and break out of the sandbox on victim users’ machines. These attacks are possible in spite of the fact that JavaScript is a memory-safe language, because its safety and isolation guarantees only apply if they are correctly maintained by JIT compilation.

This is easier said than done. Consider JavaScript arrays, which are maps from index positions to values. These maps can be “sparse,” in the sense that indices can be missing in the middle. Any out-of-bounds accesses must be checked to return the special value undefined. Furthermore, values stored in an array can be of any type, so array elements must be tagged or boxed. In a naïve implementation, numerical kernel performance would be unacceptably slow.

JavaScript JITs speculatively optimize accesses to dense arrays and to arrays whose elements are all the same type, with bailouts to a slow path should the array shape change. And they perform range analysis on values that could be used as array indices to facilitate bounds-check elimination. When range analysis confirms that the array indices are guaranteed to be within a given range, the JIT compiler generates code that allocates the array sequentially in memory and allows sequential access without any bounds checks.

These optimizations are crucial—but also risky. Failing to bail out of the speculative fast path when an array’s shape changes leads to type confusion; incorrectly eliminating a bounds check allows out-of-bounds memory accesses. Both bug types can be exploited for arbitrary code execution [63]. The implications of JavaScript JIT bugs for security were recognized as early as 2011 [103]; attackers have turned to JIT bugs as other avenues for browser exploitation become more rare and more difficult to exploit (see, e.g., [121]). Since late
2017, industrial security researchers have uncovered a dozen or more bugs in the JIT compilers shipping with Chrome (e.g., [19, 22, 23]), Firefox (e.g., [3, 15, 18, 20]), Safari (e.g., [5, 16, 17, 21, 34]), and Edge (e.g., [12–14]). They have documented JIT compiler internals and developed generic techniques for exploiting JIT compiler bugs for code execution in blog posts [43, 59, 60, 101, 105, 110, 117, 124], in industrial security conference presentations [58, 64, 67, 77, 104, 115], and in industrial journals such as Phrack [69]. And, as we noted above, JIT bugs are being weaponized against real users.

In short, the status quo is untenable. The JIT’s adversarial setting means even obscure bugs that no normal program would ever hit become exploitable attack vectors. To secure the Web, it is necessary to build and deploy JIT compilers free of security bugs.

In this paper, we explore, for the first time, the possibility of using compiler verification to secure browser JavaScript JIT compilers. Verification proves that the compiler is correct using formal methods that consider all possible corner cases.

There is, of course, much prior work on compiler verification (§8). But JavaScript JITs are a new and challenging domain for verification. JavaScript is an unusual language with complicated semantics; the ARMv8.3-A architecture revision even adds an instruction, F3C VTZS, to support floating-point conversion with JavaScript semantics [1]. And because browser JITs are supposed to improve the perceived runtime performance of Web programs, compilation time is a cost. A JIT that is verified but slow will not be acceptable.

As a first step in this direction, we build a system, VeRA, for expressing and verifying JavaScript JIT range analyses. VeRA supports a subset of C++ for expressing range analysis routines, and a Satisfiability Modulo Theories [36] (SMT)-based verification tool that proves the correctness of these analyses with respect to JavaScript semantics. Browser developers can write their range analysis code using VeRA C++, prove it correct, and incorporate the verified code directly into their browser.

Compared to prior work in compiler verification, VeRA distinguishes itself by handling the details of a realistic range analysis; we use VeRA to express and verify the range analysis used in practice by the Firefox browser. This requires handling many challenges: complicated corner cases of JavaScript semantics; complex dataflow facts whose semantics are often disjunctive predicates; and complex propagation rules for those dataflow facts. Our verification uncovered a new Firefox range analysis bug and confirmed an old bug from a previous version of the browser. We also find that our verified routines work correctly in Firefox—they pass all 140 thousand Firefox JIT tests—and perform comparably to the original routines in both micro and macro benchmarks.

2 Overview

This section gives an overview of range analysis in JIT compilers and the ramifications of range analysis bugs. Then, it walks through using VeRA to verify a piece of the Firefox JIT’s range analysis logic.

2.1 Range Analysis in JITs

Range analysis is a dataflow analysis that compilers use to deduce the range—typically upper and lower bounds—of values at different points in program execution. These range deductions, or value constraints, are then used by different optimization passes to generate efficient code. For example, Firefox’s dead code elimination (DCE) pass eliminates blocks guarded by contradictory constraints [6]:

```
if (x > 0)
  if (x < 0) /* ... dead code... */
```

It also eliminates redundant checks—in the redundant check elimination (RCE) pass—when it can prove that values are within a certain range [8]:

```
var uintArray = new Uint8Array(...); // buffer

function foo(value) {
  if (value >> 0) { // always true; redundant check
    return value;
  } else { /* ... dead code ... */
}

for(let i = 0; i < uintArray.length; i++) {
  foo(uintArray[i]); // call foo with unsigned value
}
```

Here, the comparison in foo is always true—uintArray can only contain unsigned integers—and thus can be eliminated (along with the else branch).

More significantly, Firefox relies on range analysis to move and eliminate internal JavaScript array bounds checks [7]. Since JavaScript is a memory safe language, the compiler inserts bounds checks around every array access. For instance, in the example above, the array indexing operation uintArray[idx] internally performs a bounds check, which returns undefined if the access is out of bounds, i.e., when idx < 0 or when idx >= uintArray.length. In practice, this incurs overhead—real applications make heavy use of arrays—and JITs aggressively try to eliminate bounds checks. In the example above, for instance, Firefox can prove that the array accesses are in-bounds and eliminate all the internal bounds checks.

2.2 From Range Analysis Bugs to Browser Exploits

Bugs in the range analysis code can, at best, manifest as application correctness errors—and at worst as memory safety vulnerabilities. For example, an incorrect range deduction can cause the JIT to eliminate a bounds check, which can in turn allow an attacker to read and write beyond JavaScript array bounds and hijack the control flow of the browser renderer process (e.g., using JIT-ROP techniques [84, 109]).
class Range : public TempObject {
  int32_t lower_;  
  int32_t upper_;  
  bool hasInt32LowerBound_;  
  bool hasInt32UpperBound_;  
  // possibly not a whole number  
  FractionalPartFlag canHaveFractionalPart_;  
  // possibly negative zero  
  NegativeZeroFlag canBeNegativeZero_;  
  // the maximum exponent needed to represent the number  
  // @ -1023, 1024 indicates inf, 65536 indicates NaN or inf  
  uint16_t exp_;  
}

Figure 1. Parts of Firefox’s range object

range rsh(range& lhs, int32_t c) {
  int32_t shift = c & 0x1f;
  return newInt32Range(lhs.lower >> shift,  
    lhs.upper >> shift);  
}

Figure 2. Firefox’s range analysis logic for the right shift operator.

range rsh(range& lhs, int32_t c) {
  int32_t shift = c & 0x1f;
  return newInt32Range(lhs.lower >> shift,  
    lhs.upper >> shift);  
}

Figure 3. Simplified VeRA implementation of the right shift operator.

Until recently, TurboFan—the JIT compiler component of Chrome’s V8 JavaScript engine—used to deduce that indexOf, when applied to a string, would return an integer in the range \([-1, \text{String}::kMaxLength-1]\), where \text{String}::kMaxLength is the longest allowed V8 JavaScript string (\(2^{28} - 16\) characters). Unfortunately, the actual V8 implementation of indexOf can return \text{String}::kMaxLength—one more than the range analysis deduced. As the following example (from the original bug report [22]) shows, this bug allowed attacker-supplied JavaScript to create a variable i that TurboFan deduced to be 0, but actually held an arbitrary value (in this case, 100,000):

```javascript
var i = 'A'.repeat(2**28 - 16).indexOf('A'), 2**28);
```

Since TurboFan thought i was zero, it would eliminate all bounds checks on i—so attackers could use i as an index into any array in order to access that array out-of-bounds.

2.3 Why Range Analysis is Hard to Get Right

The indexOf example is not the only case of an exploitable browser vulnerability introduced by a buggy JavaScript range analysis. Similar bugs have been a problem in practice for all major browsers. This is because JavaScript range analysis is hard to get right.

First, it requires reasoning about double-precision floating point numbers. To do so correctly and efficiently, browsers can’t just implement ranges as lower-bound and upper-bound pairs; instead, their range objects are necessarily complicated structures. Figure 1 shows Firefox’s range object, which keeps track of, among other things, integer bounds, special values, and whether the range includes non-integers. Not only is the data structure itself complex, but there are also subtle invariants that it must maintain.

Second, tracking special floating-point values like NaN, Infinity, -Infinity, and -0.0 is error-prone. For example, until recently, Turbofan’s range analysis incorrectly deduced that Math.expm1, the JavaScript builtin used to compute \(e^x{-}1\), must either return a number value or NaN—it didn’t account for floating point negative zero. The browser’s implementation of Math.expm1, applied to \(-0.0\), correctly returned \(-0.0\) [23, 119]—a mismatch that again allowed an attacker to hijack the browser renderer’s control flow.

A constellation of other factors make writing range analysis routines even harder. For example, JITs internally distinguish 32-bit integer values from double-precision floats. This is crucial for performance.1 But this also means that the range analysis must correctly determine whether an output can be within the range of possible 32-bit numbers—for the Firefox range object, whether the fields hasInt32LowerBound and hasInt32UpperBound should be set. As another example, since JIT speed directly affects users’ experience, the range analysis must be fast—it must run just in time—and usefully precise—it must produce information useful enough to assist other optimization passes (e.g., DCE, RCE, and bounds-check elimination (BCE)). To this end, JIT developers implement range analysis in low-level languages like C++ and eschew verbose, readable code for fast, terse code. They also explore the trade-offs between speed and precision: Firefox, for example, tracks integer ranges precisely—by tracking lower bounds and upper bounds—but approximates wider floating-point ranges by only tracking exponents (in addition to tracking special values).

2.4 Using VeRA to Express Range Analysis

VeRA is a subset of C++ for writing verified range analysis routines. If browser developers write their analysis logic in VeRA C++, they can compile it into automatic correctness proofs for custom properties. We ported twenty-two Firefox routines to VeRA C++; Figure 2 shows Firefox’s implementation of range analysis for the right shift operator, while Figure 3 shows the VeRA version. To calculate the range of possible output values for >>, it masks constant c with 31, per JavaScript semantics [57]. Then, it shifts the left-hand

1This makes it possible for the JIT to efficiently compile array lookups into a direct memory accesses (as opposed to lookups in a hash map) [111]
operate’s lower and upper bound by the masked constant. If rsh is given an 1hs range of [10,100] and a c constant of 2, it will return a range of [2, 25].

2.5 Using VeRA to Verify Range Analysis

VeRA also provides an internal DSL for expressing the semantic meaning of each computed fact. Given a range analysis dataflow fact \( R \), its semantic meaning is a predicate \( \llbracket R \rrbracket \) over values. For expository purposes, the following is a simplified semantic meaning for ranges over 32-bit integers:

\[
\llbracket R \rrbracket(v) \equiv R.\text{lower} \leq v \leq R.\text{upper}
\]

The actual semantic meaning we use in practice (described in Section 5) is far more complicated since it includes floating point numbers, special values, and implementation-specific Firefox invariants.

Once the semantic meaning of range facts is defined, the verification condition can be described as follows. We show the case for a binary operator \( \text{op} \), but a similar approach can be used for other kinds of operators. We use \( \text{op}_{ra} \) to denote the range analysis flow function for \( \text{op} \) (where "ra" stands for "range analysis"). This flow function, implemented in a subset of C++, takes two range objects and returns a resulting range object. Finally, we use \( \text{op}_{js} \) to denote the JavaScript semantics of \( \text{op} \).

The verification condition for \( \text{op} \) is defined as follows (where ranges \( R_1 \) and \( R_2 \), and JavaScript values \( v_1 \) and \( v_2 \), are universally quantified):

\[
\llbracket R_1 \rrbracket(v_1) \land \llbracket R_2 \rrbracket(v_2) \Rightarrow \llbracket \text{op}_{ra}(R_1, R_2) \rrbracket(\text{op}_{js}(v_1, v_2))
\]

This condition states that the flow function for \( \text{op} \) "preserves" the semantic meaning of range facts: if the semantic meaning of the incoming range facts holds on certain incoming values to the operator, then the semantic meaning of the output range fact will hold on the value produced by the JavaScript semantics of the operator on those values.

Let’s apply this definition to the rsh flow function from Figure 3. In this case \( \text{op} \) is \( \gg ; \) \( \text{op}_{ra} \) is rsh; and \( \text{op}_{js} \) is \( \gg \text{js} \). While \( \gg \) is a binary operator, the version of the operator described in Figure 3 is the one where the second parameter is a constant, and so we only need to consider range inputs for the first parameter. As result, we have the following verification condition (where range \( R \) and JavaScript values \( v, n \) are universally quantified):

\[
\llbracket R \rrbracket(v) \Rightarrow \llbracket \text{rsh}(R, n) \rrbracket(v \gg \text{js}, n)
\]

Once a verification developer specifies the predicate \( \llbracket R \rrbracket \), VeRA automatically proves it (using an SMT solver) for each range analysis function. The browser developer need only write the flow functions using a familiar subset of C++; Figure 3 is an example of such a flow function. The above is intended to give a sense for how verification works in VeRA, but the details are described in Section 5.

3 VeRA C++

In this section, we describe the programming language that browser developers use to implement range analysis flow functions in VeRA. This language is called VeRA C++, and is a subset of C++. As shown in Figure 4, various standard C++ features are not included in VeRA C++, the most notable of which may be loops (including recursion). Looping would make the verification much more complicated, because it would require determining loop invariants. In practice, though, omitting looping constructs does not affect VeRA C++’s expressiveness, since flow functions for realistic range analysis don’t rely on loops; for example, none of Firefox’s analysis functions use loops. Note that even though there is no looping allowed inside a flow function, iteration does occur in the dataflow analysis algorithm that finds a solution to the flow functions: indeed, that algorithm applies flow functions repeatedly until reaching a fixed point.

To allow programmers to describe the data structures for their range analysis, VeRA supports C++ classes. For example, the VeRA C++ code in Figure 3 uses a range class, which is a modified version of the Range class from the Firefox codebase (Figure 1). Finally, since VeRA C++ is a subset of C++, all VeRA C++ range analyses can be directly incorporated into the Firefox codebase once they are automatically verified.

4 Translating VeRA C++ to SMT

Before we describe verification conditions (Section 5), we show how to translate VeRA C++ programs to SMT. We describe the challenges with C++ to SMT translation, and then describe how we address these challenges; Section 7 explains why we choose this particular approach.
The IR consists of assignment statements in SSA form. The right-hand side of each assignment is an expression tree. Each node in the tree is labeled by an IR operator, and has child nodes that represent the operator’s parameters. Figure 5 outlines the operators in the IR.

**Compiling C++ to IR.** We compile VeRA C++ into control-flow-free IR using a series of transformations. The transformations must eliminate if statements, return statements, function calls, and method calls, and must alter variable assignments to satisfy SSA (which the IR expects).

The transformation to SSA is standard. To handle control flow, VeRA re-writes if statements into straight-line code using predicated execution. For example,

```c++
if (c) { x = 1; } else { x = 2; }
```

becomes:

```c++
x_1 = c ? 1 : x_0; x_2 = !c ? 2 : x_1;
```

VeRA handles function calls using inlining, which works well because VeRA C++ disallows recursion. Thus, each call to a function gets its own set of assignments; calling “foo(1); foo(2);” will generate two separate, versioned copies of foo. The rewrites for method calls, return values, nested conditionals, and returns from within conditionals are all similar to the above translations.

**Compiling IR to SMT.** Compiling the IR to SMT requires some additional information to be stored in the IR. In particular, each IR node will include the following three pieces of information (in addition to an operator and children):

1. The SMT term generated for this IR node
2. The C++ type of this node (e.g., int32 or double), which is used to generate the correct version of SMT operations
3. An “undef” bit, which indicates whether the node is the result of an operation with undefined behavior. For example, in the statement `x = 4 << -1`, x’s undef bit would be set since a left-shift by a negative value is undefined.

This information, computed during the translation from IR to SMT, is critical for generating SMT terms that faithfully replicate the C++ semantics of the original VeRA C++ code.

We compile IR to SMT as follows. First, we translate assignment nodes into equality, which is seamless because the IR is already in SSA form. Second, we translate the expression tree that appears on the right-hand side of each assignment using a bottom-up traversal that simultaneously computes the SMT term, the C++ type, and the “undef” bit (the three pieces of information described above).

As an example, consider a comparison operator. If either argument is unsigned, the compiler generates an unsigned SMT comparison; otherwise, it generates a signed one.

As another example, consider an IR left-shift a << b. If a is unsigned, the generated SMT term is the logical left-shift of a by b; the result type is unsigned; and the undef bit is equal to `undef(a) ∨ undef(b) ∨ isNegative(b)`.

If a is signed, the generated SMT term is again the logical left-shift of a by b, and the result type is unsigned. However,
the “undef” bit is more complicated, because C++ 14 dictates that a \( \ll b \) has undefined behavior if the shift operation discards any bits of a that were set. To detect discarded bits in the 32-bit case, we perform the shift in 64-bits \((\text{int64}_a \ll (\text{int64}_b))\), and then check if any of the high 32 bits of the result are set. If so, some bit was shifted off the end of a, and the operation has undefined behavior. As a result, the “undef” bit is equal to: \( \text{undef}(a) \lor \text{undef}(b) \lor \text{isNegative}(b) \lor (a <<_{64} b)[63 : 32] \neq 0_{32} \).

5 Veriﬁcation

In this section, we describe how VeRA proves range analyses correct; dataﬂow analysis “correctness” means that the facts concluded by the analysis are correct with respect to the semantics of the program. For range analysis: if the analysis concludes that a certain variable is in a range, the analysis is correct if the variable is actually within that range according to the program’s semantics. Note that we only consider correctness of range analysis facts, not their precision.

People typically prove dataflow analyses correct through abstract interpretation [53]. In abstract interpretation, we start by setting up a connection between computed dataﬂow facts and the semantics of the program. In our setting, we do so by deﬁning, for each range fact \( R \) that is computed about a variable, a semantic meaning—a predicate \( \llbracket R \rrbracket \) over runtime values. This predicate establishes the connection between the computed dataﬂow facts and the semantics of the program: we say that a range fact \( R \) on a variable \( x \) is correct at a program point if, for all possible values \( v \) that \( x \) can take at runtime, we have \( \llbracket R \rrbracket(v) \).

The goal of our veriﬁcation is to establish that the implemented range analysis is correct, meaning that when the dataﬂow analysis algorithm terminates, all facts it has computed are correct. In abstract interpretation, this is achieved by proving local preservation conditions on all ﬂow functions. These conditions can be shown to imply that the entire range dataﬂow analysis is correct [53]. It is these local preservation conditions that we ask an SMT solver to prove.

Recall that we use \( \text{op}_{\text{ra}} \) to denote the range analysis ﬂow function for an operator \( \text{op} \) (which for simplicity we assume to be a binary operator). Recall also that we use \( \text{op}_{\text{js}} \) to denote the JavaScript semantics of \( \text{op} \). As mentioned in Section 2, we deﬁne the local preservation condition for an operator \( \text{op} \) as:

\[
\llbracket R_1 \rrbracket(v_1) \land \llbracket R_2 \rrbracket(v_2) \Rightarrow \llbracket \text{op}_{\text{ra}}(R_1, R_2) \rrbracket(\text{op}_{\text{js}}(v_1, v_2))
\]

This condition states that the ﬂow function for \( \text{op} \) “preserves” the semantic meaning of range facts: if the incoming range facts are correct, then the propagated range fact is correct.

In this section, we ﬁrst deﬁne the semantic meaning \( \llbracket R \rrbracket \) of a given range fact \( R \). Then, we discuss how VeRA translates \( \text{op}_{\text{ra}} \) to SMT (since Section 4 explains translation of \( \text{op}_{\text{js}} \)), and talk about how proofs work in practice.

\section{5.1 Semantic Meaning of Predicate Facts}

We split the semantic meaning \( \llbracket R \rrbracket \) of a range fact \( R \) into two parts: (1) a predicate that connects \( R \) to values and (2) a well-formedness invariant expressed solely on \( R \). In particular:

\[
\llbracket R \rrbracket(v) \triangleq \text{inRange}(R, v) \land \text{wellFormed}(R)
\]

The \text{inRange} predicate connects the range fact to values from JavaScript semantics. The \text{wellFormed} predicate is more complex—it isn’t about a connection to semantic values, but instead about implementation-speciﬁc Firefox invariants that are necessary for proving the ﬂow functions correct. We explain each piece of \( \llbracket R \rrbracket \) in turn below.

The \text{inRange} predicate. Figure 6 shows the \text{inRange} predicate, which we derived from Firefox code and comments. An important point here is that this predicate is unusually complex—far more complex than any of the semantic meanings used in prior automated veriﬁcation efforts for program analyses [82]. A typical meaning for a range analysis fact in prior work is:

\[
\text{inRange}(R, v) \triangleq R.\text{lower} \leq v \leq R.\text{upper}
\]

This predicate is simpler than our \text{inRange} because \text{inRange} must handle the complexities of realistic range analysis for JavaScript (§2), i.e., tracking ﬂoating point numbers, since all JavaScript values are double-precision floats.
Given a range fact $R$ and a JavaScript value $v$, the predicate $\text{inRange}(R, v)$ is true iff the following conditions hold. First, if the exponent field of $R$ is less than a special Firefox value, $\text{e\_INF}$, $v$ must not be infinite. Similarly, if $R$’s exponent is less than the special value $\text{e\_INF\_OR\_NAN}$, $v$ must not be NaN. If $R$’s canBeNegativeZero flag is not set, $v$ should not be -0.0; if $R$’s canHaveFractionalPart flag is not set, $v$ should be a whole number. Finally, if $R$ has a lower bound (hasInt32LowerBound flag), $v$ should be greater than or equal to that lower bound (lower). The hasInt32LowerBound flag indicates whether a range contains numbers that are smaller than thirty-two-bit integers; it allows Firefox to internally use integers to represent JavaScript numbers when possible.

The wellFormed predicate. Figure 7 shows the wellFormed predicate, which is unusual because it does not relate the range fact to semantic values. Instead, it simply imposes constraints on the range fact itself; it is an invariant that Firefox flow functions depend on to be correct. We derived this invariant from a set of long comments and a handful of invariant-checking functions in the Firefox codebase.

All range facts should have canBeNegativeZero set only when zero is contained within their range\(^2\), where contains is defined as follows for range $R$ and value $v$:

$$\text{contains}(R, v) \equiv v \geq R.\text{lower} \land v \leq R.\text{upper}$$

Furthermore, exponents should either be in the range 0 to 1023, or should have the special value $\text{e\_INF}$ (1024) for infinity, or the special value $\text{e\_INF\_OR\_NAN}$ (65535) for NaN or infinity. In addition, the exponent should also be consistent with the lower and upper bound, if they exist.

Now, we walk through an example showing how Firefox routines break when invariants are violated, focusing on the invariant relating hasInt32LowerBound with $R.\text{lower}$ and hasInt32UpperBound with $R.\text{upper}$. The Firefox implementation of range analysis requires that if a range’s hasInt32LowerBound is unset, then its lower field equals JS_INT_MIN; similarly, if its hasInt32UpperBound is unset, its upper field equals JS_INT_MAX. Consider the following Firefox range analysis code for max, which calculates the new lower field and the new hasInt32LowerBound flag in the returned range fact:

```plaintext
int32_t newLower = max(lhs->lower, rhs->lower)
bool newHasLower = lhs->hasInt32LowerBound ||
                  rhs->hasInt32LowerBound
```

Now consider two input ranges to the max function, $R_1$ and $R_2$, and assume the output range is $R_3$. If the Firefox invariant doesn’t hold, we can have $\neg r_1.\text{hasInt32LowerBound}$ and $R_1.\text{lower} = 1000$, and $R_2.\text{hasInt32LowerBound}$ and $R_2.\text{lower} = -1000$. This means that the above code will set $R_3.\text{hasInt32LowerBound}$ to true, and $R_3.\text{lower}$ to 1000 (since the result lower bound is the maximum of the input lower bounds).

Now consider two JavaScript values $v_1$ in $R_1$ and $v_2$ in $R_2$ (meaning that $\text{inRange}(R_1, v_1)$ and $\text{inRange}(R_2, v_2)$ are both true). Let $v_1$ be $-\infty$ (since $\neg r_1.\text{hasInt32LowerBound}$ and $v_2$ be -1000. Then, JavaScript semantics says that $\text{max}_v(v_1, v_2)$ is -1000, but -1000 is well below $R_1$’s lower bound of 1000.

However, once we add the wellFormed invariant, this problem is fixed: when hasInt32LowerBound is not set, lower must be JS_INT_MIN, so $R_1.\text{lower}$ cannot be 1000. Now that $R_1$’s lower bound is JS_INT_MIN instead of 1000, it correctly captures $\text{max}_v(v_1, v_2)$ by including -1000. This example shows that max’s range analysis relies on the invariant that lower is JS_INT_MIN when hasInt32LowerBound is not set.

Other Verification Conditions. We prove two additional properties of the Firefox range analysis: we prove the correctness of the functions for combining range facts—union and intersection—which the JIT uses within its larger range analysis loop. We will use union and intersection to refer to Firefox’s union and intersection functions for range facts. Conceptually, given two range facts $R_1$ and $R_2$, we want to show that $\text{union}(R_1, R_2)$ is an overapproximation\(^3\) of the mathematical union operation $\cup$ on the semantic values contained in $R_1$ and the semantic values contained in $R_2$. We achieve this as follows. Given two well-formed ranges $R_1$ and $R_2$, we let $R_3 = \text{union}(R_1, R_2)$. Then we want to show that for all JavaScript values $v$, we have:

$$\text{inRange}(R_1, v) \lor \text{inRange}(R_2, v) \implies \text{inRange}(R_3, v)$$

Similarly, for intersection we let $R_3 = \text{intersect}(R_1, R_2)$, and prove that:

$$\text{inRange}(R_1, v) \land \text{inRange}(R_2, v) \implies \text{inRange}(R_3, v)$$

5.2 Using VeRA to Express Predicates

VeRA exposes an internal verification domain-specific language (DSL) embedded in Haskell. Verification developers use the language to express verification infrastructure, including verification conditions and the semantic meaning of range facts. The DSL exposes a number of JavaScript operators—i.e., implementations of op\(_D\)—against which to verify range analysis functions. If verification developers wish to verify new operations, it is straightforward to expose new JavaScript routines in the VeRA internal DSL. To express predicates and verification conditions, the DSL also exposes SMT directives. They allow verification developers to make assumptions, call the SMT solver, and push and pop new incremental solver contexts.

In practice, proving the conjunction of wellFormed and inRange for each operator is suboptimal: if any component of either predicate does not finish, the entire verification

\(^2\)This is only an invariant on optimized ranges; unoptimized may have set flags even if their ranges don’t contain zero.

\(^3\)And it is necessarily an overapproximation given how Firefox’s range analysis object is implemented: consider a union of two ranges, one that can include fractions and one that can’t.
proof does not finish. As a result, we prove individual conditions within each predicate separately; Section 6 describes each condition that we prove and the time the proof takes to finish (or not). Mechanically, to prove a given predicate part, we query an SMT solver with its negation—e.g., we check whether the operation can produce any values outside of the computed range. If the formula is unsatisfiable, there are no such values and the range analysis routine is safe. If the formula is satisfiable, the model provided by the solver is a concrete counterexample that captures that input ranges and values for which the range analysis routine is unsafe.

6 Implementation and Evaluation

We implement VeRA—both the compiler for VeRA C++ and the internal verification language—in 1384 lines of Haskell. Since the implementation details of the compiler are standard, we only describe the details relevant to answering our evaluation questions. All our source code and data sets are available at https://vera.programming.systems.

We evaluate VeRA by asking six questions:

Q1 Can VeRA prove Firefox range analysis routines correct?
Q2 Can VeRA proofs catch real correctness bugs?
Q3 Are the VeRA proofs correct?
Q4 Do the verified routines work correctly in Firefox?
Q5 How do the verified routines perform in Firefox?
Q6 How hard is it to integrate verified routines into Firefox?

To answer these questions, we port 21 top-level Firefox range analysis flow functions to VeRA C++, try to prove their correctness, and then re-integrate them into the browser. We are able to: prove 137 separate facts about these routines; identify a new Firefox analysis bug; and correctly detect an old analysis bug. A version of Firefox that uses our verified routines still performs comparably to standard Firefox, and it still passes all (147,322) Firefox JavaScript tests.

6.1 Proofs

We choose to verify 19 Firefox flow functions because they are the complete set of Firefox Range-type flow functions for JavaScript operators; we discuss this further in Section 7 (e.g., as a result, we don’t check division). In addition, we verify the union and intersect predicates, which are not JavaScript operators but instead Firefox-internal functions that combine two different Range objects; this brings the total number of routines we attempt to verify to 21. Verification of both union and intersect times out, but VeRA finds a bug in our port of an older, broken version of intersect [9]. Finally, we port 25 helper functions called by each verified function—i.e., every function except the optimize function (§7). Figure 8 summarizes our results.

All operators followed by asterisks in the table (e.g., rsh) are only valid for 32-bit ranges. Thus, for each bitwise operator, we only prove (1) the simple predicate from Section 2 and (2) the absence of undefined behavior in the result range’s upper and lower bounds. We don’t worry about wellFormed for these operations, either; none of them alter the floating-point-specific fields of the range.

For floating-point operators (e.g., add), we separately prove each condition in the inRange predicate and the wellFormed predicate with two exceptions: since we do not call optimize on our result ranges, we only verify the wellFormed conditions that apply to non-optimized ranges (i.e., our output ranges are correct but may not have the tightest possible bounds). Figure 8 goes into more detail about which columns correspond to which conditions in Figure 6 and Figure 7.

Multiple proofs fail: ceil and broken intersect are both real bugs, while ursh and ursh’ require extra invariants on the input ranges. We don’t amend them because their failure is actually an interesting manifestation of a comment above both functions [26]:

```
// ursh's left operand is uint32, not int32, but for range analysis we currently approximate it as int32. We assume // here that the range has already been adjusted...
```

In other words, over all possible inputs, ursh is not correct.

Our proof code, i.e., the code that automatically verifies each part of both the inRange and wellFormed predicates, amounts to 804 lines of Haskell. We run all proofs on a virtual machine (QEMU-KVM, Linux 5.3.11) running Arch Linux with 16 GB of memory and 4 vCPUs. The processor is an AMD Ryzen Threadripper 2950X 16-Core with a base clock rate of 3.5GHz. We use the Z3 SMT solver (4.8.6), which we call with custom Haskell bindings that extend the haskell-z3 library [30], using a 20-minute timeout.

Can we prove Firefox range analysis routines correct?

We successfully prove or refute 137 conditions out of a possible 159, for a success rate of ≈86%; the shortest proofs complete in under a second, while the longest takes ≈ten minutes. The results suggest that R5.double, R6.double, and W4 are particularly challenging to verify. R5.double and R6.double are more challenging than their integer counterparts because they involve reasoning about floating-point values, which is generally more expensive. W4 is challenging because it involves proving a relationship between two properties of the range, both of which may be modified by the range analysis. Finally, R1 and W3 of Math.ceil may timeout because they involve bounding the size of an exponent, since Math.ceil involves extracting the exponent from the absolute value of the range bounds.

There is hope, however. Though VeRA does not verify conditions for certain routines (e.g., correctness of intersect),
Towards a Verified Range Analysis for JavaScript JITs

PLDI ’20, June 15–20, 2020, London, UK

Table 8. The time it takes, in seconds, for VeRA to verify predicates from Section 5. R1–7 correspond to the lines in the inRange predicate, while W1–4 correspond to line groups one through four in wellFormed. Broadly, R1 makes sure the routine handles infinities correctly, R2 NaNs, R3 -0.0, R4 fractions, R5s lower bounds (over both 32-bit integer and double values), and R6s upper bounds. W1 ensures missing upper and lower bounds imply a minimum or maximum value for lower and upper; W2 ensures lower and upper are always valid JavaScript 32-bit numbers; W3 ensures the range’s exponent is valid, and W4 ensures the exponent is consistent with the upper and lower bound. Finally, Undef shows the time it takes to verify that the range computations for upper and lower are free of undefined behavior. ∞ indicates timeout, while X indicates verification failure.

<table>
<thead>
<tr>
<th>Operation</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5,i32</th>
<th>R5.double</th>
<th>R6.i32</th>
<th>R6.double</th>
<th>R7</th>
<th>W1</th>
<th>W2</th>
<th>W3</th>
<th>W4</th>
<th>Undef</th>
</tr>
</thead>
<tbody>
<tr>
<td>add</td>
<td>15</td>
<td>2</td>
<td>5</td>
<td>386</td>
<td>2</td>
<td>∞</td>
<td>2</td>
<td>∞</td>
<td>21</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>80</td>
<td>1</td>
</tr>
<tr>
<td>sub</td>
<td>13</td>
<td>2</td>
<td>11</td>
<td>445</td>
<td>8</td>
<td>∞</td>
<td>5</td>
<td>∞</td>
<td>14</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>78</td>
<td>1</td>
</tr>
<tr>
<td>and*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>or*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>xor*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>not*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>mul</td>
<td>92</td>
<td>65</td>
<td>22</td>
<td>362</td>
<td>∞</td>
<td>∞</td>
<td>∞</td>
<td>∞</td>
<td>∞</td>
<td>94</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>lsh*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>rsh*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>ursh*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>lsh’*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>rsh’*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>ursh’*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>abs</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>∞</td>
<td>1</td>
<td>224</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>2</td>
<td>20</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>224</td>
<td>2</td>
<td>∞</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>max</td>
<td>3</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>15</td>
<td>∞</td>
<td>2</td>
<td>∞</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>12</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>floor</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>146</td>
<td>1</td>
<td>9</td>
<td>54</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>∞</td>
<td></td>
</tr>
<tr>
<td>ceil</td>
<td>∞</td>
<td>1</td>
<td>X</td>
<td>8</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>9</td>
<td>266</td>
<td>1</td>
<td>1</td>
<td>∞</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>sign</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8. The time it takes, in seconds, for VeRA to verify predicates from Section 5. R1–7 correspond to the lines in the inRange predicate, while W1–4 correspond to line groups one through four in wellFormed. Broadly, R1 makes sure the routine handles infinities correctly, R2 NaNs, R3 -0.0, R4 fractions, R5s lower bounds (over both 32-bit integer and double values), and R6s upper bounds. W1 ensures missing upper and lower bounds imply a minimum or maximum value for lower and upper; W2 ensures lower and upper are always valid JavaScript 32-bit numbers; W3 ensures the range’s exponent is valid, and W4 ensures the exponent is consistent with the upper and lower bound. Finally, Undef shows the time it takes to verify that the range computations for upper and lower are free of undefined behavior. ∞ indicates timeout, while X indicates verification failure.

it is able to catch multiple bugs in broken versions of unverified routines. For example, as we discuss later in this section, we port an older, broken version of intersect to VeRA C++. VeRA is able to detect the bug in this version in 173 seconds, even though the proof for the current version of intersect never finishes. VeRA also catches a number of errors that we introduced while copying code over from Firefox; as we improved VeRA, it went from being a custom DSL to a subset of C++, so porting was not always as easy as copy-pasting browser code. For example, we switched an upper and lower bound in sub, which caused it to fail the floating-point lower bounds check—even though this check never verifies in the fixed version of sub. VeRA caught one other porting error in sub (use of the field canBeNegativeZero instead of the function canBeZero), and at least two more in mul (switched lsh and rhs, and use of canBeFiniteNonNegative in place of canBeFiniteNegative).

A new Firefox bug. VeRA found a bug in Firefox’s range analysis for the Math.ceil operator [4], which rounds its input up to the nearest integer (e.g., the ceiling of 2.5 is three). The bug, which follows, has existed since the routine’s introduction six years ago [2]:

```c
Range Range::ceil(TempAllocator& alloc, const Range* op) {
  Range* copy = new (alloc) Range(*op);
  if (copy->hasInt32Bounds())
    copy->max_exponent_ =
      copy->max_exponentImpliedByInt32Bounds();
  else if (copy->max_exponent_ < MaxFiniteExponent)
    copy->max_exponent_++;

    copy->canHaveFractionalPart_ = ExcludesFractionalParts;
    copy->assertInvariants();
    return copy;
}
```

The routine looks straightforward. Given an input range, it adjusts the range’s exponent upwards by one—to account for upward rounding—and unset the canHaveFractionalPart flag—since the result is always a whole number.

The problem lies in what ceil doesn’t do: it never adjusts the input range’s canBeNegativeZero flag. JavaScript semantics, though, dictate that Math.ceil(x) = -0 when x is between zero and negative one. Therefore, given a range with lower and upper bound [-1, 0] and an unset canBeNegativeZero flag, ceil will not correctly set the flag.

VeRA identifies the error after about three seconds, and provides the following (shortened) counterexample:

```c
result_range.canBeNegativeZero : 0
start_range.canBeNegativeZero : 0
start_range.lower : -128
start_range.upper : 0
start : -1.166149293399505e-301
```
It finds a start value, start, within the `start_range` of $[-128, 0]$ with the `canBeNegativeZero` flag unset. It notes that ceiling of start is result, -0, but that the result range does not have the `canBeNegativeZero` flag set.

After confirming the bug with Mozilla engineers, we tried to patch the bug and use VeRA to verify the patch. VeRA rejected our first attempt—it was wrong—but approved the next one [4]:

```c
| copy |>canBeNegativeZero_ = ((copy |> lower_ > 0) ||
| copy |>upper_ <= -1) 

| copy |>canBeNegativeZero_  
| IncluesNegativeZero;
```

Now, the function sets the `canBeNegativeZero` flag to be true when the resulting range can be includes values between negative one and zero.

**An old Firefox bug.** To test VeRA further, we port a buggy version of Firefox’s `intersect` operator, which takes the intersection of two ranges [10]. The buggy operator deduces that the intersection of two ranges $r_1 \cap r_2 = \emptyset$ if the upper bound of $r_1$ doesn’t overlap with the lower bound of $r_2$. This behavior is correct—unless both ranges contain NaN, in which case the result range should include NaN, too. VeRA identifies the error after 173 seconds, and provides a counter example showing an element in the two input ranges (NaN) that was not included in the output range.

**Are VeRA proofs correct?** Verifiers can be as broken as the code they are intended to verify. To guard against this possibility, we use Haskell’s QuickCheck [52] to automatically random-test the semantics of (1) the operators in the VeRA IR and (2) the JavaScript operators that VeRA uses for verification. For each JavaScript operator, we generate JavaScript code that performs the operations, evaluates it with Node.js, and compares the result against that produced by our SMT model. We use Node.js (version 10.1.0) because it uses the Chrome V8 JavaScript engine and is thus likely to have different bugs from Firefox (and thus VeRA). Our C++ operator tests are similar; we use Clang version 9.0.0. This checking proved useful—e.g., QuickCheck found a bug in the implementation of the C++ floating-point abs operator. For further assurance, we also cross-checked our JavaScript semantics against KJS [96].

6.2 Verified Routines in the Browser

We integrate our verified range analysis into Firefox 72.0a1 (commit 10c869240d). Our versions of the Firefox range analysis functions amount to 621 lines of C++ code. In this section, we describe the effort it took to retrofit Firefox and measure the performance of our modified browser when compared with `vanilla`, unmodified Firefox.

**How hard is it to integrate VeRA code into Firefox?** We retrofit Firefox in two steps. First, we extend Firefox’s Range class with a `verifiedRange` field—pointing our verified object—and modify the class setters and getters to forward all accesses to the corresponding `verifiedRange` fields. For example, we rewrite:

```c
| Range* Range::abs(TempAllocator& alloc, const Range* op) |
| int32_t l = op | lower_; |
| int32_t u = op | upper_; |
```

```c
| Range* Range::abs(TempAllocator& alloc, const Range* op) |
| auto vrRange = verified::abs(op | verifiedRange); |
| return Range::fromVerifiedRange(alloc, vrRange); |
```

Our porting effort was surprisingly low: two engineers—neither of whom is a Firefox core developer—integrated our VeRA C++ routines into Firefox over the course of two days. We think this effort can be reduced even further: both steps are mechanical and can be automated to only require human intervention when tests fail.

**Do the verified routines work correctly?** To test our ports, we run all of Firefox’s JavaScript suites (using their mach build tool): the 7,364 JIT tests, 394 JSAPI tests (which use the Range interface directly), and 139,564 general JavaScript tests. Though VeRA passes all of Firefox’s tests now, it failed some tests along the way. After altering Firefox to call VeRA routines whenever applicable, all but three of the JIT tests passed; one timed out and two failed. This was because one of our ported functions did not initialize every field of the range object; the function was only called by a range analysis routine for a bitwise operator, so to verify it, we did not need to set the floating-point-specific range fields. We also failed two of the JSAPI tests, both due to typos in our porting of `intersect` and `sign` functions. The VeRA verification actually caught the sign bug, but due to a miscommunication within the team, the fix didn’t make it into Firefox immediately. The bug in `intersect` caused it to over-approximate the possibility of negative zeroes—but our verification specifically allows over-approximation by design (§3). Once we fixed these bugs, all JIT and JSAPI tests passed, and when we ran the other 139,564 Firefox JavaScript tests, all of those passed, too.
Both popular and unpopular websites. In particular, we use the list of 11 sites curated by the Chrome team in their recent Chrome sandboxing work [102]. For each site, we use Mozilla’s Talos benchmarking tool to measure the time it takes to render a page (i.e., the time to first paint [28]), taking the median of 50 runs (after a 5 run warm-up). Figure 9 presents our measurements. The median and average latencies of browsing these sites with our browser are within 5% of vanilla Firefox. The biggest slowdown is on reddit.com (18%), while the biggest speedup is on economist.com (-9%); like [102], we attribute these bigger differences to the inherently noise introduced by media content and the network. Overall, these results are encouraging: VeRA does not impose overheads that are prohibitive to its adoption.

7 Discussion, Limitations and Future Work

Why not existing proof tools? We started out building a DSL for range analysis verification, and ended up building a compiler from both C++ and an internal verification language to SMT. There are many existing tools that can translate programming languages into SMT [49, 100], and they primarily operate on an existing compiler IR (e.g., LLVM IR) instead of defining their own IR. Verifying JavaScript JIT optimization passes at the LLVM IR level is something to strive for in the future, but using a small language is an easier start; to our knowledge, no LLVM-IR-level tool supports JavaScript semantics, some don’t support C++ reliably [24], and the ones that do can get lost in complex class object hierarchies—in IR, series of pointer offset calculations—and as a consequence struggle to verify anything at all. Furthermore, using these tools requires integrating them with (very complicated) browser build systems. We, on the other hand, don’t provide any guarantees about the final machine code.

What we don’t verify. We only verify range analysis routines for Firefox. Chrome’s range analysis pass is tied to its type inference pass—different ranges correspond to different types (e.g., 32-bit integers have a bound). We consider extending VeRA to other browsers future work.

Within Firefox, the range analysis functions that we verify all return the basic Firefox Range object type. These functions contain local range analysis logic, and are called during the range analysis computation for different MIR nodes, Firefox’s middle-level intermediate representation of JavaScript programs. For example, the computeRange method for the MSpectreMaskIndex node (representing Math.ceil) simply wraps Range.ceil and thus reaps the benefits of our verification effort. Many MIR nodes, however, do not correspond to a JavaScript-level construct (e.g., MSpectreMaskIndex is used to represent a masked array index) and we thus do not verify them. We also do not verify the range analysis algorithm itself nor Firefox’s use of ranges in code generation or other optimization passes (e.g., DCE or BCE). These are natural
8 Related Work

VeRA lies at the intersection of work on compiler verification, JIT compilation, and browser security.

Verifying optimizations using SMT. Various prior systems have used DSLs combined with SMT solvers to express and verify compiler optimization correctness: for example, Cobalt [81], Rhodium [82], PEC [80] and Alive [86]. These systems mostly focus on proving correctness of transformations, e.g., scalar optimizations in Cobalt and Rhodium, control flow rewrites in PEC, and peephole optimizations in Alive. While there has been much less work on using DSLs and SMT solvers to prove the correctness of analyses, with the notable exception of the Rhodium system [82]. While conceptually the techniques in VeRA are similar to Rhodium’s techniques for analysis correctness, the analyses in Rhodium are relatively simple, with the semantic predicate of facts often containing a single term. In contrast, VeRA demonstrates how to use DSLs and SMT solvers to verify the correctness of analyses in a realistic setting: our work handles all the corner cases of Firefox’s range analysis, which in turn includes a semantic predicate for the range facts with 16 cases.

Another difference between our work and prior work is the type of constructs we support (in VeRA C++). Alive supports pointers and arrays (with static, known sizes), while VeRA does not support either construct—but neither system handles loops, since neither peephole optimizations nor range analysis computations typically require them.

Compiler and analysis verification in a proof assistant. Another approach to general compiler verification is foundational verification. In foundational verification, the programmer writes the compiler in a Proof Assistant, and then uses the proof assistant to interactively prove that the compiler is correct. Examples of foundational verified compilers include CompCert [83], CompCertTSO [116], Compositional CompCert [112], and CakeML [79]. Examples of semantic IR frameworks include the Vellvm system [122], which provides a formal semantics for LLVM IR that others can use to verify IR optimizations/transformations.

In addition to entire compilers or IR frameworks, there has also been work on specific analyses and optimizations that are foundational verified. For example, Versaco [76] is a foundationally verified static analyzer for CompCert; developers can use it to write their own analyses that prove properties of analyzed programs. For specific optimizations, Zhao et al. use Vellvm to verify a version of LLVM’s mem2reg transformation, which changes memory references to register references [25, 123]. Mullen et al. use the Coq proof assistant to verify peephole optimizations for the CompCert verified C compiler [83, 90]. Finally, Tatlock et al. extend CompCert with a DSL for expressing optimizations [114]—combining the DSL approach with foundational verification.

Because proofs in foundational verification are performed in full detail, foundational verification provides the strongest possible correctness guarantees. However, these proofs often require a significant amount of expert human guidance, making them very difficult to complete. In contrast, VeRA allows browser developers to express their analysis in a subset of C++, and then provides automated verification to the developer without any additional effort or verification knowledge.

Verification of JIT compilation. There has also been work specifically on verifying correctness of JIT compilers, including work on verifying a non-optimizing JIT compiler [91], and work on defining correctness criteria for trace-based JIT compilation [72]. There is also ongoing work on verifying a JIT in Coq [38], which includes support for some optimizations, though it is not clear which ones; similarly, there is ongoing work on verifying a JIT using symbolic execution [107]. None of this work focuses on the specific challenge that we are addressing, namely verifying the complex analyses that drive optimizations in browser JITs.

Translation validation. Another approach to compiler correctness is translation validation [92, 98, 106]: each time the compiler runs, a validator tries to prove that the transformed code behaves the same as the original code. While translation validation can find compiler bugs, it does not guarantee the absence of bugs, as VeRA attempts to. For translation validation to guarantee the absence of bugs, it would have to do validation on production runs, which incurs compilation overhead—not ideal for a JIT.

Recent work by Taneja et. al. goes further by proposing an algorithm for sound and maximally precise dataflow facts like integer ranges (similar to this work) and known bits, among others [113]. For a given code fragment, they (1) use their algorithm (implemented with an SMT solver) to compute dataflow facts about that fragment and (2) compare those facts to the ones LLVM has computed. This technique has identified several precision errors in LLVM’s analyses, and adapting it to help developers design tighter ranges is interesting future work.

Verification and floating point numbers. Other systems also handle the challenge of verifying floating-point code. Recently, Boldo et al. formalize IEEE-754 semantics in Coq in order to extend CompCert to support programs that use floating-point numbers [47]. More similar to VeRA, Icing [39], which builds on the verified ML compiler CakeML [79], is a language for writing “fast-math” optimizations. Multiple projects also extend the Alive system to support peephole
optimizations related to floating point numbers [89, 94]. Recently, Becker et al. use the Daisy tool [54] to verify optimizations to floating-point computations—in programs, not in compilers—that the Herbie [95] tool generates [40].

**Testing compilers.** Beyond verification, there are other approaches to compiler correctness. One is automatic testing, or “fuzzing,” which finds bugs but does not guarantee the absence of errors. For example, the Csmith tool automatically generates useful—well-formed, undefined-behavior-free—inputs for testing C compilers [120]. Dewey et al. present a system for fuzzing the Rust compiler’s type checker [56]. Finally, there are fuzzers specific to JavaScript interpreters: Fuzzili [65, 66], and CodeAlchemist [74]. Fuzzili has been remarkably effective at finding bugs in JavaScript engines; its bug showcase lists two dozen security issues [11].

**Verified JavaScript semantics.** There has been a significant effort to formalize (parts of) the JavaScript language. Most of these efforts start with a simple core language and extend it with unwieldy JavaScript features (e.g., eval, property descriptors, and with) [61, 70, 87, 99]. The JSCert JavaScript subset was even mechanized in Coq, from where they extracted a verified correct interpreter [45, 46, 62]. KJS [96] provides a complete JavaScript semantics in the K framework. Though we cross-check our semantics against KJS and JSCert’s (where possible), this work is complimentary: they focus on verifying JavaScript semantics, we focus on the JITs that should preserve these semantics.

**Verification in the browser.** A verified browser kernel, Quark [75], demonstrates that verification is possible in the browser setting. Though we’re a long way from a verified Firefox, browsers are interested in verified software. For example, both Firefox and Chrome incorporated verified cryptographic primitives into their TLS stacks [27, 125].

**Safer browser JITs.** Another approach to JIT safety is to limit the damage of bugs, not prevent them. NaClJIT sandboxes both the JIT compiler and the code it produces [31]. RockJIT applies a control-flow integrity policy to the JIT compiler and the code it produces [93]. NoJITSu prevents code-reuse and data-only attacks [97]. Browser vendors have also modified their JITs to reduce the effectiveness of JIT spraying [85], a technique that allows attackers to introduce bytes of their choice into pages marked executable in browser memory [44].

**Large, real-world verification.** Finally, beyond what we have discussed so far, there are many other large, real-world verification efforts. The Astrée static analyzer has verified safety properties of Airbus software [55]; the miTLS project provides verified reference implementations of TLS 1.0, 1.1, and 1.2 [42]; Barbosa et al. [35] give an overview of verification efforts for crypto code, including discussion of an ongoing effort to formally verify the TLS 1.3 protocol. seL4 [78] is a verified operating systems kernel; and various recent efforts have proved properties of systems software like file systems [50, 51, 108] and in-kernel interpreters [118].

### 9 Conclusion

This paper presents VeRA, a system for verifying the range analysis pass in browser JITs. VeRA allows browser developers to write range analysis routines directly in the browser (in a subset of C++) and provides a DSL that verification developers can use to encode verification properties (e.g., range analysis invariants). VeRA automatically verifies these properties using SMT.

We use VeRA to encode a semantics for Firefox’s range analysis, and then port 22 Firefox analysis routines to VeRA C++ in order to verify them. The ported version of the browser performs on-par with the original. Moreover, VeRA detects a bug that has existed in the browser for six years—and verifies the Firefox patch we wrote to fix the bug.

### Acknowledgments

We thank the anonymous reviewers for their comments, and especially our shepherd John Regehr for helping us think through both the content and presentation of this paper. Thanks to Dawson Engler for technical discussions about this work and other applications of these techniques. Thanks to Riad S. Wahby for uploading our huge artifact (and more!), and to Ranjit Jhala for early discussions. Finally, thanks to Chris Fallin, Nick Fitzgerald, Dan Gohman, Jan de Mooij, Nicolas B. Pierron, and Luke Wagner at Firefox; we can’t imagine a more thoughtful and responsive group of people. This work was supported in part by gifts from Cisco, Fastly, Google, Mozilla, and Qualcomm; by the NSF under Grant Numbers CCF-1918573 and CNS-1514435; and by the CONIX Research Center, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.

### References
