**Solver-Aided Constant-Time Hardware Verification**

**ABSTRACT**

We present *Xenon*, a solver-aided, interactive method for formally verifying that *Verilog* hardware executes in constant-time. *Xenon* scales to realistic hardware designs by drastically reducing the effort needed to localize the root cause of verification failures via a new notion of constant-time counterexamples, which *Xenon* uses to synthesize a minimal set of secrecy assumptions in an interactive verification loop. To reduce verification time *Xenon* exploits modularity in *Verilog* code via module summaries, thereby avoiding duplicate work across multiple module instantiations. We show how *Xenon*’s assumption synthesis and summaries enable us to verify different kinds of circuits, including a highly modular AES-256 implementation where modularity cuts verification from six hours to under three seconds, and the ScarVs side-channel hardened RISC-V micro-controller whose size exceeds previously verified designs by an order of magnitude. In a small study, we also find that *Xenon* helps non-expert users complete verification tasks correctly and faster than previous state-of-art tools.

**KEYWORDS**

template; formatting; pickling

1 INTRODUCTION

Timing side-channel attacks are no longer theoretical curiosities. Over the last two decades, they have been used to break implementations of cryptographic primitives ranging from public-key encryption algorithms [26, 62, 88], to block ciphers [23, 71], digital signature schemes [70], zero-knowledge proofs [35], and pseudo-random generators [33]. This, in turn, has allowed attackers to break systems that rely on these primitives for security—for example, to steal TLS keys used to encrypt web traffic [26, 33, 88], to snoop and forge virtual private network traffic [70], and to extract information from trusted execution environments [25, 33, 35, 87].

The gold standard for preventing timing side-channel attacks is to follow a discipline of constant-time or data-oblivious programming [1, 15, 22, 29, 36, 89]. At its core, this discipline ensures that secret data is not used as an operand to variable-time instructions (e.g., floating-point operations like division [19, 20, 63, 75]) and (2) the program’s control flow and memory access patterns do not depend on secrets. But for the constant-time discipline to be effective, it is crucial that the constant-time property be preserved by the underlying hardware. For example, an instruction that is deemed constant-time needs to indeed produce its outputs after the same number of clock cycles, irrespective of operands or internal state. Similarly, given that control-flow and memory access patterns are public, i.e., free of secrets, a CPU’s timing must indeed be secret-independent.

Unfortunately, simply assuming that hardware preserves constant-time doesn’t work. Incorrect assumptions about the timing-variability of floating-point instructions, for example, allowed attackers to break the differentially private Fuzz database [56]. Attempts to address these attacks (e.g., [76]) were also foiled: they relied on yet other incorrect microarchitectural assumptions (e.g., about the timing-variability of SIMD instructions) [63]. Yet more recently, hardware crypto co-processors (e.g., Intel and STMicroelectronics’s trusted platform modules) turned out to exhibit similar secret-dependent timing variability [70].

A promising path towards eliminating such attacks is to formally verify that our hardware preserves the constant-time property of the software it is executing. Such verification efforts, however, require tool support. Unfortunately, unlike software verification of constant-time, which has had a long history [21], constant-time hardware verification is still in its infancy [28, 50, 53, 89]. As a result, existing verification approaches fail to scale to realistic hardware. This is because of two fundamental reasons. First, existing tools do not help when verification fails—and inevitably it does fail: hardware circuits only preserve constant-time execution under very specific secrecy assumptions that describe which port and wire values are public or secret. In our experience, dealing with failures takes up most of the verification time. With tools like *Iodine* [50], you must manually determine whether the circuit is leaky (i.e., variable-time), or whether it is missing additional secrecy assumptions that the tool needs to be made aware of. Second, current methods fail to exploit the modularity that is already explicit at the register transfer level. Hence, they duplicate verification effort across replicated modules which leads to a blow up in verification time.

In this paper, we present *Xenon*, a solver-aided, interactive method for formally verifying that *Verilog* hardware executes in constant-time. We develop *Xenon* via five contributions.

1. **Counterexamples.** To help users understand verification failures, we introduce the notion of constant-time counterexamples (§ 4.1). A counterexample highlights the earliest point in the circuit where timing variability is introduced; this simplifies the task of understanding whether a circuit is variable-time by narrowing the user’s attention to the root cause of the verification failure (and thereby a small fraction of the circuit). To compute counterexamples, *Xenon* leverages information extracted from the failed proof attempt. In particular, the solver communicates

   (1) which variables (i.e., registers and wires) remained constant-time during the failed proof attempt, and

   (2) the order in which the remaining variables became non-constant time.

This allows *Xenon* to break cyclic data-dependencies which cause a chicken-and-egg problem that is hard to resolve when assigning blame manually.

2. **Assumption Synthesis.** To help the user resolve the verification failure, *Xenon* uses the counterexample to synthesize a suggested fix. For example, *Xenon* may find a constant-time counterexample for a processor pipeline where the two different runs may execute two different ISA instructions (say, addition and division) which take different numbers of clock cycles. Yet, the execution of each instruction (for any inputs) may be constant-time. *Xenon* uses the counterexample to synthesize a minimal candidate set of secrecy assumptions (e.g., that any two executions have the same, publicly
visible sequence of instructions) which address the root cause of the verification failure (§ 4.2). The user then decides either to accept the candidate assumption or, if they do not match their intuition for the intended usage of the circuit, reject them, in which case Xennon computes an alternative. Internally, Xennon computes candidate assumptions via a reduction to integer linear programming [73].

3. Modular Verification. Finally, to scale verification and counterexample generation to larger and more complex hardware, and to keep counterexamples and suggested assumptions local, we introduce a notion of module summaries (§ 3). Module summaries succinctly capture the timing properties of a module’s input and output ports at a given usage site. By abstracting inessential details about the exact computations performed by the module and focusing solely on its timing behavior, Xennon produces fewer and more compact constraints. Our modular verification approach also allows the user to focus attention on one module at a time, which keeps errors and assumptions local, and helps to bootstrap the verification of large circuits (§ 7).

4. Evaluation. We implement Xennon and evaluate the impact of counterexamples, assumption synthesis, and modularity on the verification of different kinds of hardware modules (§ 6). We find that Xennon’s solver-aided interactive verification process drastically reduces verification effort (e.g., verifying the largest benchmark of [50] took us several minutes instead of multiple days) and, together with module summaries, allows us to scales verification to realistic hardware (e.g., we verify the SCARV side-channel hardened RISC-V core [4], which is order of magnitude larger than the RISC-V cores verified by previous state-of-the-art tools). From a small (ten-person) user study, in which users were tasked with verifying three circuits (an ALU, an FPU, and a full RISC core), we find that Xennon has large \( d = 1.62 \), statistically significant \( t(8) = 2.56, p = .016 \) effective positive effect on correct completion: Participants using Xennon were able to correctly complete significantly more tasks in the allotted time (40 min), and their solution sizes were (on average) smaller. On the most challenging task—a full RISC-V processor with a complex assumption set—no participant in the control-group succeeded, whereas 60% of the participants using Xennon were able to successfully complete the verification task.

6. Secrecy Assumptions. As a side product of the verification of SCARV, we obtain a set of annotations (§ 7) detailing secrecy assumptions under which SCARV is guaranteed to execute in constant-time. These secrecy assumptions, together with Xennon’s source code are open source and available on GitHub (link elided for DBR). We hope that these artifacts will facilitate further efforts to provide end-to-end constant-time guarantees across hardware and software.

2 OVERVIEW

We start by reviewing how to specify and verify the absence of timing channels in Verilog hardware designs (§ 2.1), show how existing techniques fail to scale on real-world hardware designs, as these designs are often only constant-time under additional secrecy assumptions which are tedious to derive by hand (§ 2.2), sketch how Xennon helps to find secrecy assumptions automatically (§ 2.3), and finally discuss how Xennon exploits modularity (§ 2.4).

```verilog
module S (clk, in, out);
input clk; [7:0] in;
output reg [7:0] out;
always @ (posedge clk)
case (in)
  8’h00: out <= 8’h63;
  ...
  8’hff: out <= 8’h2c;
endcase
endmodule
```

Figure 1: A simple, constant-time lookup-table in Verilog, taken from [10].

<table>
<thead>
<tr>
<th>time</th>
<th>in</th>
<th>L</th>
<th>K</th>
<th>L</th>
<th>R</th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>h00</td>
<td>hff</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>h00</td>
<td>hff</td>
<td>h63</td>
<td>h2c</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Figure 2: Two runs of Figure 1 showing values and liveness-bits for input (in) and output (out). X represents an undefined value.

2.1 Verifying Constant-Time Execution of Hardware

Lookup Circuit. Figure 1 shows the code for a Verilog module, which implements a lookup table by case-splitting over the 8-bit input value. This module executes in constant-time: even if input \( in \) contains a secret value, producing output \( out \) takes the same amount of time (one clock cycle), irrespective of the value of \( in \), and therefore an attacker cannot make any inference about the value of \( in \) by observing the timing of the computation.

Specifying Constant-Time Execution. Figure 2 makes this intuition more precise, using a recent definition of constant-time execution for hardware [50]. Instead of tracking timing indirectly through information flow [66, 82, 91] the definition uses a direct notion of timing. The figure shows two runs of module \( S \): one for input \( 8’h00 \) and one for input \( 8’hff \). We want to track how long it takes for the two inputs, issued at cycle 1 to pass through the circuit and produce their respective outputs. For this, we put a “tracer” on the inputs by assigning a liveness-bit to each register. For some register \( x \), we set its liveness-bit \( x^* \) to \( \star \), if \( x \) has been influenced by the input at initial cycle 1 (we say \( x \) is \( 1 \)-live) and \( \diamond \), otherwise. Figure 2 shows how liveness-bits are propagated through the circuit. Initially, in both executions, the input is \( 1 \)-live and the output is not. In cycle 2, both outputs \( 1 \)-live due to the case-split on the value of \( in \). Assuming that an attacker can observe the liveness-bits of all outputs, here, register out, the attacker cannot distinguish the two executions, and we can conclude that the pair of executions is indeed constant-time.

Verifying Constant-Time Execution. To show constant-time execution, not only for the two runs in Figure 2, but for the whole circuit, we have to prove that for any pair of runs, that is, for any pair of inputs, and any initial cycle, the constant-time property holds. This can be achieved by constructing a product circuit [50] whose runs correspond to pairs of runs—called left and right—of the original circuit. In this product, each original variable \( x \) has

\[ Xennon \text{ computes an alternative. Internally, Xennon computes candidate assumptions via a reduction to integer linear programming [73].} \]

\[ \text{We start by reviewing how to specify and verify the absence of timing channels in Verilog hardware designs (§ 2.1).} \]

\[ \text{show how existing techniques fail to scale on real-world hardware designs, as these designs are often only constant-time under additional secrecy assumptions which are tedious to derive by hand (§ 2.2),} \]

\[ \text{sketch how Xennon helps to find secrecy assumptions automatically (§ 2.3), and finally discuss how Xennon exploits modularity (§ 2.4).} \]

\[ \text{Our modular verification approach also allows the user to focus attention on one module at a time, which keeps errors and assumptions local, and helps to bootstrap the verification of large circuits (§ 7).} \]

\[ \text{We find that Xennon’s solver-aided interactive verification process drastically reduces verification effort (e.g., verifying the largest benchmark of [50] took us several minutes instead of multiple days) and, together with module summaries, allows us to scales verification to realistic hardware (e.g., we verify the SCARV side-channel hardened RISC-V core [4], which is order of magnitude larger than the RISC-V cores verified by previous state-of-the-art tools).} \]

\[ \text{From a small (ten-person) user study, in which users were tasked with verifying three circuits (an ALU, an FPU, and a full RISC core), we find that Xennon has large \( d = 1.62 \), statistically significant \( t(8) = 2.56, p = .016 \) positive effect on correct completion: Participants using Xennon were able to correctly complete significantly more tasks in the allotted time (40 min), and their solution sizes were (on average) smaller.} \]

\[ \text{On the most challenging task—a full RISC-V processor with a complex assumption set—no participant in the control-group succeeded, whereas 60% of the participants using Xennon were able to successfully complete the verification task.} \]

\[ \text{As a side product of the verification of SCARV, we obtain a set of annotations (§ 7) detailing secrecy assumptions under which SCARV is guaranteed to execute in constant-time. These secrecy assumptions, together with Xennon’s source code are open source and available on GitHub (link elided for DBR).} \]

\[ \text{We hope that these artifacts will facilitate further efforts to provide end-to-end constant-time guarantees across hardware and software.} \]

\[ \text{We start by reviewing how to specify and verify the absence of timing channels in Verilog hardware designs (§ 2.1).} \]

\[ \text{show how existing techniques fail to scale on real-world hardware designs, as these designs are often only constant-time under additional secrecy assumptions which are tedious to derive by hand (§ 2.2),} \]

\[ \text{sketch how Xennon helps to find secrecy assumptions automatically (§ 2.3), and finally discuss how Xennon exploits modularity (§ 2.4).} \]

\[ \text{Our modular verification approach also allows the user to focus attention on one module at a time, which keeps errors and assumptions local, and helps to bootstrap the verification of large circuits (§ 7).} \]

\[ \text{We find that Xennon’s solver-aided interactive verification process drastically reduces verification effort (e.g., verifying the largest benchmark of [50] took us several minutes instead of multiple days) and, together with module summaries, allows us to scales verification to realistic hardware (e.g., we verify the SCARV side-channel hardened RISC-V core [4], which is order of magnitude larger than the RISC-V cores verified by previous state-of-the-art tools).} \]

\[ \text{From a small (ten-person) user study, in which users were tasked with verifying three circuits (an ALU, an FPU, and a full RISC core), we find that Xennon has large \( d = 1.62 \), statistically significant \( t(8) = 2.56, p = .016 \) positive effect on correct completion: Participants using Xennon were able to correctly complete significantly more tasks in the allotted time (40 min), and their solution sizes were (on average) smaller.} \]

\[ \text{On the most challenging task—a full RISC-V processor with a complex assumption set—no participant in the control-group succeeded, whereas 60% of the participants using Xennon were able to successfully complete the verification task.} \]

\[ \text{As a side product of the verification of SCARV, we obtain a set of annotations (§ 7) detailing secrecy assumptions under which SCARV is guaranteed to execute in constant-time. These secrecy assumptions, together with Xennon’s source code are open source and available on GitHub (link elided for DBR).} \]

\[ \text{We hope that these artifacts will facilitate further efforts to provide end-to-end constant-time guarantees across hardware and software.} \]
world circuits are typically not variable pipelined MIPS processor [9]. If the reset bit rst is set (Line 18), the processor sets several registers to zero (Line 19). Otherwise, the processor checks whether the pipeline is stalled (Line 21) and either forwards the current instruction from the instruction-fetch stage to the instruction-decode stage (Line 28) and advances the program counter (Line 29), or stalls by reassigning the current values (Lines 22 to 24).

The Pipeline is not Constant-Time. When using the processor in a security-critical context, we want to make sure that it avoids leaking secrets through timing, i.e., that it is constant-time. Unfortunately, our example pipeline is not constant-time without any further restrictions on its usage. For example, the execution time for a given instruction depends on whether the pipeline is stalled before the instruction is retired. This is illustrated in Figure 4. We model an attacker that can measure how long an instruction takes to move through the pipeline, i.e., from source IF_pc to sink WB_reg. Such an attacker can distinguish the two runs in Figure 4, as the liveness-bits of WB_reg differ in cycle 3. This timing difference lets the attacker make inferences about the control flow of the program which is executed on the processor, and therefore any attempt to verify constant-time execution results in a failure.

2.3 Automatically Finding Secrecy Assumptions

We may, however, still be able to use this processor safely, if we can find a suitable set of secrecy assumptions. For example, we could assume that the pipeline is public (i.e., free of secrets, which can be formally expressed by the assumption that StallL = StallR always holds). In this case, the timing difference in Figure 4 would only leak information that the attacker is already aware of. However, assuming that Stall is public may not be the best choice. Stall is defined deep inside the pipeline which makes it hard to translate this assumption into a restriction on the kind of software we are allowed to execute on the processor. Instead, we want to pick assumptions as closely as possible to the sources, i.e., external visibility computation inputs. For example, restricting program counter IF_pc to be public directly translates into the obligation that the executed program’s control flow be independent of secrets.

To discover this assumption using existing technology, the verification engineer first has to manually identify that the timing variability is first introduced in variable ID_instr (Lines 22 and 28) due to a control dependency on Stall. They then need to inspect how Stall is set (Line 14) and painstakingly trace the definitions which may involve complex combinatorial logic (excerpt starting in Line 5) and circular data-flows to identify a promising candidate register, such that marking the register as public will render the circuit constant-time. Counterexamples like Figure 4 are often of little help as they are hard to interpret and fail to focus attention on the relevant parts of the circuit.

Solver Aided Verification: Xenon’s Interactive Loop. Xenon drastically simplifies this time-consuming process through an interactive, solver-aided workflow that helps to find an optimal set of secrecy assumptions automatically.

Step 1. First, we start with an empty set of secrecy assumptions and run Xenon on the pipeline. The verification fails, as the pipeline is not constant-time, however, Xenon displays the following prompt to guide the user towards a solution.

Figure 3: MIPS Pipeline Fragment.

two copies xL and xR that hold the values of x in the left and right runs, respectively. We can then use the product circuit to synthesize invariant properties of the circuit. For example, let’s define that a variable x is constant time (and write \( ct(x) \)), if for any pair of executions, its liveness-bit in the left execution is always the same as its liveness-bit in the right execution, i.e., \( x^L_t = x^R_t \) always holds, for all initial cycles t. Then, the following invariant on the module proves constant-time execution, under the condition, that module inputs are constant-time: \( ct(in) \Rightarrow ct(out) \).

2.2 Real-World Hardware is Not Constant-Time

Unfortunately, unlike the simple lookup table from Figure 1, real-world circuits are typically not constant-time, in an absolute sense. Instead, when carefully designed, they are constant-time under specific secrecy assumptions detailing which circuit inputs are supposed to be public (visible to the attacker) or secret (unknown to the attacker). Thus, verification requires the user to painstakingly discover secrecy assumptions through manual code inspection, which can be prohibitively difficult in real-world circuits.

A Pipelined MIPS Processor. We illustrate the importance of secrecy assumptions using the program in Figure 3 which shows a code-fragment taken from one of our benchmarks—a simple, pipelined MIPS processor [9].
> Mark ‘rst’ as PUBLIC? [Y/n]

The user either answers with Y indicating that rst should indeed be considered public, or else responds n which tells XENON to exclude the variable from future consideration (i.e., not suggest it in future). Suppose that we follow XENON’s advice, and choose Y: this marks rst public and re-starts XENON for another verification attempt.

**Step 2.** Next, XENON suggests marking M_PCSrc as public. Flag M_PCSrc indicates whether the current instruction in the memory stage contains an indirect jump. But whether an indirect jump is executed depends on register values (i.e., M_PCSrc is set depending on whether the output of the ALU is zero) and therefore, indirectly, on the data memory. Assuming that M_PCSrc is public would lead to assumptions about the memory which we wish to avoid. Hence, we tell XENON to exclude it in future verification attempts and restart verification.

**Step 3.** Restarting verification causes XENON to suggest candidate variable IF_pc, the program counter of the fetch stage. We mark IF_pc as public as this directly encodes the assumption that the program’s control flow does not depend on secrets. XENON restarts the solver which proves that the program—under the inferred secrecy assumptions—executes in constant-time. This concludes the verification process. In addition to the assumptions that rst and IF_pc are public, XENON also infers a set of usage assumptions that detail the parts of the pipeline that have to be flushed on context switches. These assumptions would otherwise have to be supplied by the user as well.

**Counterexamples.** When synthesizing assumptions, XENON internally computes a constant-time counterexample which contains the set of variables that have lost the constant-time property earliest. While the user can simply follow XENON’s suggestions without further investigating the root cause of the violation, we find that—if the user chooses to do so—the counterexample often helps to further understand why the circuit has become non constant-time. In our example, XENON returns as counterexample, variable ID_instr, for all three interactions. Indeed, inspecting the parts of the circuit where ID_instr is assigned focuses our attention on the relevant parts of the circuit, that is, the conditional assignment of ID_instr under rst (Line 19) and under Stall (Lines 22 and 28). We discuss how XENON computes counterexamples using artifacts extracted from the failed proof attempt in § 4.1, and how XENON uses them to synthesize an optimal set of secrecy assumptions via a reduction to integer linear programming in § 4.2.

### 2.4 Real-World Circuits Are Not Small

While XENON’s solver-aided, interactive verification loop significantly reduces the time the user has to spend on verification efforts, large real-world circuits often also present a challenge for the solver.

<table>
<thead>
<tr>
<th>time</th>
<th>stall</th>
<th>ID_jmp</th>
<th>IF_inst*</th>
<th>ID_inst*</th>
<th>EX_rt*</th>
<th>WB_reg*</th>
</tr>
</thead>
<tbody>
<tr>
<td>L R</td>
<td>L R</td>
<td>L R</td>
<td>L R</td>
<td>L R</td>
<td>L R</td>
<td>L R</td>
</tr>
<tr>
<td>1</td>
<td>0 1</td>
<td>1 0</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>2</td>
<td>0 0</td>
<td>0 1</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>3</td>
<td>0 0</td>
<td>0 0</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Figure 4: Two runs of Figure 3, where the right run stalls in cycle 1. The liveness bits of sink wb_reg differ in cycle 3 and therefore the circuit is not constant-time.

![Figure 5: Module dependency graph of the AES-256 benchmark.](image)

Figure 5: Module dependency graph of the AES-256 benchmark.

```
1 module S4 (clk, in, out);
2   input clk;
3   input [31:0] in;
4   output [31:0] out;
5   wire [7:0] out_0, out_1, out_2, out_3;
6   S S0 (clk, in[31:24], out_3);
7   S S1 (clk, in[23:16], out_2),
8   S S2 (clk, in[15:8], out_1),
9   S S3 (clk, in[7:0], out_0);
10  assign out = (out_3, out_2, out_1, out_0);
11 endmodule
```

Figure 6: Module from the AES benchmark.

This is because computing invariants and synthesizing assumptions naively requires a whole-program analysis. Hence, efficiency crucially depends on the size of the circuit we are analyzing.

Consider, for example, the AES-256 benchmark from [10]. Fig. 5 depicts the dependency graph of its modules, where each node $m$ represents a Verilog module, and we draw an edge between modules $m$ and $n$ if $m$ instantiates $n$. Each edge is annotated with the number of instantiations. Even though there are only ten modules, the total number of module instantiations is 789. This, in turn, causes a blowup in the size of code XENON has to verify. Even though the sum of #LOC of the modules is only 856, inlining module instances causes this number to skyrocket to 135194 rendering both assumption synthesis and verification all but intractable. (In fact, XENON does manage to verify the naive, inlined circuit, however, a single verification run takes over 6 hours to complete).

Fortunately, we can avoid this blowup by exploiting the modularity that is already apparent at the Verilog level. We illustrate this process using module S from Figure 1.

**Module Summaries.** Since the value of out only depends on in, we can characterize its timing behavior as follows: the module output out is constant-time, if module input in is constant-time.

We can formalize this in the following module summary, which XENON computes automatically: $ct(in) \Rightarrow ct(out)$. Instead of inlining the module, both assumption synthesis, and verification can now use its summary thereby eschewing the code explosion. The
3 MODULAR CONSTANT-TIME VERIFICATION

We now formalize the concepts introduced in the overview. We first say that \( t_0 \) is the initial cycle.

After that, we discuss counterexamples and assumption synthesis (§ 3.3). After that, we discuss counterexamples and assumption synthesis (§ 4).

3.1 Defining Constant-Time Execution

Configurations. Configurations represent the state of a Verilog computation. A configuration

\[
\Sigma \triangleq (P, \sigma, \theta, c, t, Src)
\]

is made up of a Verilog program \( P \) (say, the processor in Figure 3), a store \( \sigma \), a liveness map \( \theta \), current clock cycle \( c \in \mathbb{N} \), initial clock cycle \( t \in \mathbb{N} \) and, finally, a set of sources \( Src \subseteq Vars \). Store \( \sigma \in Vars \rightarrow \mathbb{Z} \) maps variables \( Vars \) (registers and wires) to their current values; \( \theta \in Vars \rightarrow \{\ast, \bullet\} \) maps variables to liveness-bits; cycle \( t \) marks the start of the computation. We want to track and finally, \( Src \) identifies the inputs of the computation we are interested in.

Transition relation. Transition relation \( \Rightarrow \in (\Sigma \times \Sigma) \) encodes a standard Verilog semantics which defines how a configuration is updated from one clock cycle to the next. We omit its definition, as it is not needed for our purposes, but formal accounts can be found in [50, 51, 91]. In addition to updating the store and current cycle, the transition relation updates the liveness map \( \theta \) by tracking which variables are currently influenced by the computation started in \( t \). At initial cycle \( t \), our transition relation starts a new computation by setting the liveness-bits of all variables in \( Src \) to \( \ast \), and those of all other variables to \( \bullet \).

Runs. We call a sequence of configurations

\[
\pi \triangleq \Sigma_0 \Sigma_1 \ldots \Sigma_{n-1}
\]
a run, if each consecutive pair of configurations is related by the transition relation, i.e., if \( \Sigma_i \Rightarrow \Sigma_{i+1} \), for \( i \in \{0, \ldots, n-2\} \). We call \( \Sigma_0 \triangleq (P, \sigma_0, \theta_0, 0, t, Src) \) initial state, and require that \( \theta_0 \) maps all variables to \( \bullet \). Finally, for a run

\[
\pi \triangleq (P, \sigma_0, \theta_0, c_0, t, Src) \ldots (P, \sigma_{n-1}, \theta_{n-1}, c_{n-1}, t, Src),
\]

we say that \( \pi \) is a run of \( P \) of length \( n \) with respect to \( t \) and \( Src \) and let \( store(\pi, t) = \sigma_t \) and \( live(\pi, t) = \theta_t \), for \( i \in \{0, \ldots, n-1\} \).

Example. Consider again Figure 2. The Figure depicts two runs \( \pi_1 \) and \( \pi_2 \) of length 3 with respect to initial cycle 1 and source \( \{in\} \) of the program in Figure 1. Columns \( in \) and out show \( store(\pi, i)(in) \) and \( store(\pi, i)(out) \), for \( \pi \in \{\pi_1, \pi_2\} \) and \( i \in \{1, 2\} \). Similarly, columns \( in^* \) and out* show \( live(\pi, i)(in) \) and \( live(\pi, i)(out) \), for \( \pi \in \{\pi_1, \pi_2\} \), and \( i \in \{1, 2\} \). The Figure omits the initial state at cycle 0, where all liveness-bits are set to \( \bullet \).

Flushed, Constant-Time, Public. For two runs \( \pi_L \) and \( \pi_R \) of length \( n \), we say that variable \( v \) is flushed, if \( store(\pi_L, 0)(v) = store(\pi_R, 0)(v) \), we call \( v \) public, if \( store(\pi_L, i)(v) = store(\pi_R, i)(v) \), for \( i \in \{0, \ldots, n-1\} \) and call \( v \) constant-time, if \( live(\pi_L, i)(v) = live(\pi_R, i)(v) \), for \( i \in \{0, \ldots, n-1\} \).

Secrecy Assumptions. A set of secrecy assumptions

\[ A \triangleq \text{(Flush, Pub)} \]

consists of a set of variables \( \text{Flush} \subseteq Vars \) that are assumed to be flushed in the initial state, and a set of variables \( \text{Pub} \subseteq Vars \), that are assumed equal throughout. A pair of runs \( \pi_L \) and \( \pi_R \) of length \( n \), satisfy a set of assumptions \( A \), if, for each \( v \in \text{Flush} \), \( v \) is flushed, and for each \( v \in \text{Pub} \), \( v \) is public. We describe how XENON synthesizes secrecy assumptions in § 4.

Constant-Time Execution. We now define constant-time execution with respect to a set of sinks \( Snk \subseteq Vars \), sources \( Src \), and assumptions \( A \). We say that a program \( P \) is constant-time, if for any initial cycle \( t \) and any pair of runs \( \pi_L \) and \( \pi_R \) of \( P \) with respect to \( t \) and \( Snc \) of length \( n \) that satisfy \( A \), and any sink \( v \in \text{Snk} \), \( v \) is constant-time.

Example. Consider again Figure 2. If we assume that variables in cycle 0 have the same value as in cycle 1, then output is flushed while \( in \) is not. Neither \( in \) nor \( out \) are public, but both are constant-time. As output is constant-time in all runs, the program in Figure 1 is constant-time with respect to the empty set of assumptions and sink \( \{out\} \). In Figure 4, none of the variables are public or constant-time, however, the program in Figure 3 can be shown to be constant-time with

\[ Pub = \{IF_{pc}, rst\}. \]

3.2 Verifying Constant-Time Execution via Horn Constraints

To verify constant-time execution, we mirror the formal definition in a set of Horn clauses [24]—an intermediate language for verification. We start with the naive, monolithic encoding and discuss how to make it modular in § 3.3. At high level, the constraints

(1) issue a new live instruction at a non-deterministically chosen initial cycle \( t \), and

(2) ensure constant-time execution by verifying that the liveness-bits for each sink are always the same, in any two runs.

The clauses—shown in Figure 7—encode verification conditions over an inductive invariant

\[ inv(as_L, as_R, c, t) \]

of the product circuit, where \( as \) ranges over all variables in the circuit and their respective liveness bits, and \( c \) and \( t \) are the current and initial cycles, respectively.

Initial States and Transition Relation. Formula \( init(as_L, as_R) \) describes the product circuit’s initial states and requires all liveness-bits to be set to \( \bullet \). To ensure that the proof holds for any initial cycle, \( init \) does not constrain \( t \). Formula \( next(as_L, as_R, as'_L, as'_R, c, t) \) encodes the transition relation of the product circuit, where un-primed variables represent state before, and primed variables represent state after the transition. Like \( \Rightarrow \), \( next \) sets liveness-bits of all sources
init(\(v_{SL}, v_{SR}\)) \land flush \land pub \Rightarrow inv(v_{SL}, v_{SR}, 0, t) \quad \text{(init)}

\[
\begin{align*}
\left( inv(v_{SL}, v_{SR}, c, t) \land pub \right. \\
\left. \land next(v_{SL}, v_{SR}, v_{SL}', v_{SR}', t) \right) \Rightarrow inv(v_{SL}', v_{SR}', c + 1, t) \quad \text{(cons)}
\end{align*}
\]

\[
\text{inv}(v_{SL}, v_{SR}, c, t) \land pub \Rightarrow s^\bullet_L = s^\bullet_R \quad \text{for } o \in \text{SNK} \quad \text{(ct)}
\]

Figure 7: Horn clause encoding of the verification conditions for constant-time execution.

to ★ at clock cycle \(t\). Importantly, constructing next requires inlining all modules and therefore can lead to large constraints that are beyond the abilities of the solver.

**Assumptions.** For a set of assumptions \(\mathcal{A} \triangleq (\text{FLUSH, PUB})\), we construct formulas \(\text{flush}\) and \(\text{pub}\), both of which require the variables in their respective sets to be equal in the two runs. We let

\[
\text{flush} \triangleq (\land_{x \in \text{FLUSH}} x_L = x_R) \quad \text{and} \quad \text{pub} \triangleq (\land_{x \in \text{PUB}} x_L = x_R).
\]

**Horn Constraints & Solutions.** We then require that the invariant holds initially (init), assuming all variables in \(\text{FLUSH}\) and \(\text{PUB}\) are equal in both runs; that the invariant is preserved under the transition relation of the product circuit, assuming that public variables are equal in both runs (cons), and finally, that the liveness-bits of any sink are the same in both runs (ct). These constraints can then be passed to any of a vast array of existing Horn constraint solvers [8, 34, 37, 44, 52, 55, 57] yielding a formula which, when substituted for \(\text{inv}\), makes all implications valid and thus proves constant-time execution.

**Proof Artifacts.** To compute constant-time counterexamples and synthesize secrecy assumptions upon a failed proof attempt, as described in the next section, Xenon requires the solver to generate the following artifacts:

1. the set of variables which remained constant-time and public, during the current failed proof attempt, and
2. the order in which the remaining variables lost the respective properties.

These artifacts can, for example, be extracted from a concrete counterexample trace like Figure 4. However, even if the solver is unable to produce concrete counterexample traces, the necessary information can often be recovered from the internal solver state.

### 3.3 Finding Modular Invariants

Naively, constructing next requires all the code to be in a single module. However, this can yield gigantic circuits whose Horn clauses are too large to analyze efficiently. To avoid instantiating the entire module at each usage site, Xenon constructs module summaries that concisely describe the timing relevant properties of the module’s input and output ports.

**Per-Module Invariants and Summaries.** Instead of a single whole program invariant \(\text{inv}\), the modular analysis constructs a per-module invariant \(\text{inv}_m\), and an additional summary \(\text{sum}_m\), for each module \(m\). The summary only ranges over module inputs and outputs, and respective liveness-bits (\(\text{io}\)) and needs to include all input/output behavior captured by the invariant, i.e., we add a clause:

\[
\text{inv}_m(v_{SL}, v_{SR}, t) \Rightarrow \text{sum}_m(i_{SL}, i_{SR}, t) \quad \text{.}
\]

The analysis produces the same constraints as before, but now on a per-module basis, that is, we require module invariants to hold on initial states (init), and be preserved under the transition relation (cons), but, instead of using the overall transition relation next we use a per-module transition relation \(\text{next}_m\). It may now happen that \(\text{next}_m\) makes use of a module \(n\), but instead of inlining the transition relation of \(n\) as before, we substitute it by its module summary \(\text{sum}_m\), thereby avoiding the blowup in constraint size. Finally, we restrict sources and sinks to occur at the top-level module, and add a clause requiring that any sink has the same liveness-bits in both runs (ct). The summaries are also used to modularize our assumption synthesis algorithm § 4.3, which is crucial for our modular verification approach, as we will discuss in § 7.

**Solving Modularity Constraints.** To solve the modular Horn constraints, the solver first computes an invariant for each module, and then uses quantifier elimination [58] to project the module’s behavior onto its inputs and outputs, which yields the summary. Since a module’s summary may show up in another module’s transition relation and thereby influence its invariant, this yields an interdependent constraint system, which we solve via a fix-point iteration loop [24, 52].

### 4 COUNTEREXAMPLES & ASSUMPTION SYNTHESIS

We now explain how Xenon uses the proof artifacts to help the user understand and explicate secrecy assumptions when verification fails. We first describe how Xenon analyzes the artifacts from the failed proof attempt in order to compute a counterexample consisting of the set of variables that—according to the information communicated by the prover—lost the constant-time property first (§ 4.1). Next, we discuss how Xenon uses the counterexample to synthesize a set of secrecy assumptions that eliminate the root cause of the verification failure (§ 4.2). This is done by computing a blame-set that contains the variables that likely caused the loss of constant-time for the variables in the counterexample via a control dependency. This blame set is then used to encode an optimization problem whose solution determines a minimal set of assumptions required to remove the timing violation. Finally, we briefly discuss how Xenon uses module summaries to speed up counterexample generation and secrecy assumption synthesis (§ 4.3).

#### 4.1 Computing Counterexamples

**Dependency Graph.** To compute the counterexample from a failed proof attempt, Xenon first creates a dependency graph

\[
G \triangleq (V, D \cup C)
\]

which encodes data- and control-dependencies between program variables. \(G\) consists of

- **variables** \(V \subseteq \text{VARS}\),
- **data-dependencies** \(D \subseteq (\text{VARS} \times \text{VARS})\), where \((v, w) \in D\) if \(v\)’s value is used to compute \(w\) directly through an assignment, and
- **control-dependencies** \(C \subseteq (\text{VARS} \times \text{VARS})\) where \((v, w) \in C\), if \(v\)’s value is used indirectly, i.e., \(w\)’s value is computed under a branch whose condition depends on \(v\).
We define the counterexample $Cex$ of a graph $G$ with map $varTime$ as the set of nodes in the reduced graph $G’$ (wrt. $varTime$), that have no predecessors, i.e.,

$$Cex = \{ v | \pre(v, G’) = 0 \} .$$

**Example: Simplified Pipeline.** The code in Figure 8 shows a simplified version of the pipelined processor from Figure 3. Like in Figure 3, the pipeline either stalls (Lines 10 and 11) if flag Stall is set (Line 9), or else forwards values to the next stage (Lines 13 and 14). To avoid a write-after-write data-hazard, the Stall flag is set, if the instructions in the execute and decode stage have the same target registers (Line 6). The target register is calculated from the current instruction (Line 1), and the instruction is, in turn, fetched from memory using the current program counter (Line 3). Note the cyclic dependency between ID_instr and Stall1 that turns comprehending the root cause into a “chicken-and-egg” problem.

**Dependency Graph.** To check if the pipeline fragment executes in constant-time, we mark IF_pc as source, and ID_instr as sink and run Xenon. Since the pipeline is variable-time, the verification fails. To compute a minimal counterexample, Xenon creates the dependency graph shown in Figure 9a. Each node is annotated with information extracted from the failed proof attempt: the node is marked with (✓) if the variable remained constant-time throughout the proof attempt and (✗) otherwise. Solid edges represent data- and dashed edges represent control-dependencies.

**Reduced Dependency Graph.** Figure 9b shows the dependency graph after removing all constant-time nodes and edges that violate the causal ordering. Xenon extracts an artifact from the failed proof attempt: the node is marked with (✓) if the variable remained constant-time throughout the proof attempt and (✗) otherwise. Solid edges represent data- and dashed edges represent control-dependencies.

**Remark.** In case the proof artifact only partially resolves the cyclic dependencies, that is, $varTime$ only defines a partial order over non-constant-time variables, the reduced graph may still contain cycles, and therefore there may be no nodes without predecessor. We can however still apply our technique by computing the graph’s strongly connected components and including all nodes in the respective component in the counterexample.

4.2 Assumption Synthesis

The previous step leaves us with a set of nodes $Cex$, which lost the constant-time property first. Since these nodes must have lost the constant-time property through a control dependency on a secret value, we can compute a set of variables $\text{Blame}$ that are directly responsible: the immediate predecessors of $Cex$ in the dependency graph with respect to a control dependency. Formally, for dependency graph $G = (V, D \cup C)$, we let

$$\text{Blame} = \{ w | v \in Cex \land (w, v) \in C \} .$$

```verilog
assign ID_rt = ID_instr[20:16];
romj2 IMEM(IF_pc, IF_instr);
always @(*)
  stall = (ID_rt == EX_rt);
always @(posedge clk)
begin
  if (Stall == 1) begin
    ID_instr <= ID_instr;
    EX_rt <= EX_rt;
  end else begin
    ID_instr <= IF_instr;
    EX_rt <= ID_rt;
  end
end
```
To synthesize secrecy assumptions that remove the constant-time violation, we could directly assume that all nodes in Blame are public. But this is often a poor choice: variables in Blame can be defined deep inside the circuit, whereas we would like to phrase our assumptions in terms of externally visible input sources.

**Finding Secrecy Assumptions via ILP.** Instead, we compute a minimal set of assumptions close to the input sources via a reduction to Integer Linear Programming (ILP). To this end, we use a second proof artifact, a map secret that—similar to varTime—describes the temporal order in which the verifier determines variables have become secret, i.e., ceased being public. Let

\[ G' = (V', D' \cup C') \]

be the reduced dependency graph with respect to secret, and let No \( \subseteq V' \) be a set of variables that the user chose to exclude from consideration. Xenon produces constraints on a new set of variables: two constraint variables

\[ m_v \in \{0, 1\} \text{ and } p_v \in \{0, 1\}, \]

for each program variable \( v \), such that \( m_v = 1 \), if program variable \( v \) is marked public by an assumption, and \( p_v = 1 \), if \( v \) can be shown to be public, that is, it is either marked public, or all its predecessors are public. Then, Xenon produces the following set of constraints.

\[ m_v \geq p_v, \quad \text{if } v \in V', \text{ pre}(v, G') = \emptyset \]  
\[ m_v + \left( \frac{\sum_{w \in \text{pre}(v, G')} p_w}{\#\text{pre}(v, G')} \right) \geq p_v, \quad \text{if } v \in V', \text{ pre}(v, G') \neq \emptyset \]  
\[ p_v = 1, \quad \text{if } v \in (\text{Blame} \setminus \text{No}) \]  
\[ m_v = 0, \quad \text{if } v \in \text{No} \]  

Constraints (1) and (2) ensure that a variable is public, if either it is marked public, or all its predecessors in \( G' \) are public. Constraint (3) ensures that all blamed variables that have not been excluded can be shown to be public, and finally, constraint (4) ensures that all excluded constraints are not marked. Let \( d(v, w) \) be a distance metric, i.e., a function that maps pairs of nodes to the natural numbers. Then we want to solve the constraints using the following objective function that we wish to minimize, where for \( v \in V' \), we define as weight the minimal distance from one of the source nodes:

\[ \sum_{v \in V'} w_v m_v, \]  

and we let \( w_v = (\min_{\text{inScope}} d(v, o)) \). A solution to the constraints defines a set of assumptions \( A = (\text{Flush}, \text{Pub}) \), with

\[ \text{Flush} \triangleq \{ v \in V' \mid m_v = 0 \land p_v = 0 \} \]

and

\[ \text{Pub} \triangleq \{ v \in V' \mid m_v = 1 \}. \]

The constraints can be solved efficiently by an off-the-shelf ILP solver.

**Example: Simplified Pipeline.** Consider again the simplified pipeline in Figure 8. As we identified ID_instr as counterexample in the previous step, we need to ensure that its blame set consisting of all indirect influences is public. ID_instr only depends on Stall, and therefore we add constraint \( p_{\text{Stall}} = 1 \). Since all variables are secret (i.e., we didn’t make any public-assumptions yet), the reduced graph is equal to the original graph. For variables IF_instr and ID_instr, we get:

\[ m_{\text{IF_instr}} + p_{\text{IF_pc}} \geq p_{\text{IF_instr}} \]

and

\[ m_{\text{ID_instr}} + p_{\text{IF_instr}} + p_{\text{Stall}} \geq p_{\text{ID_instr}}. \]

We obtain the following objective function:

\[ m_{\text{IF_pc}} + 2m_{\text{IF_instr}} + 3m_{\text{ID_instr}} + \ldots. \]

Sending the constraints to an ILP solver produces a solution, where \( m_{\text{IF_pc}} = 1 \), and \( m_v = 0 \), for all variables \( v \neq \text{IF PC} \), and \( p_v = 1 \), for all \( v \). This corresponds to the following assumption set \( A = (\text{Flush}, \{\text{IF PC}\}) \), where Flush includes all variables except IF_PC. This is exactly our desired minimal solution where we only mark IF_PC as public. Note that our method does not necessarily result in all variables becoming public. We give an example in Appendix A.

### 4.3 Modular Assumption Synthesis

To avoid a blowup in constraint size and to keep counterexamples and synthesized assumptions local, we want to avoid inlining instantiated modules. We, therefore, extract a dependency graph from the module summary. Whenever the summary requires an
input in to be public for an output out to be constant-time, we draw a control dependency between in and out. Whenever the summary requires an input in to be constant-time for an output out to be constant-time, we draw a data dependency. Finally, we insert the computed summary graph into the top-level dependency graph, and connect the instantiation parameters to the graph’s inputs and outputs.

**Example.** We modify Figure 8 to factor out the updates to ID_instr into a separate module. Xenon computes the following summary invariant, from which we create the graph in Figure 9c

\[
ct(\text{IF} \_ \text{instr}) \land pub(\text{Stall}) \Rightarrow pub(\text{ID} \_ \text{instr}).
\]

Since connecting the instantiated variables to the summary graph is equivalent to the original graph (Figure 9a), our analysis returns the same result.

## 5 IMPLEMENTATION

Xenon is split into front-end and back-end. Our front-end translates Verilog to an intermediate representation (IR) and associates secrecy assumptions with input and output wires. Our back-end translates this annotated IR into verification conditions (Horn clauses); when verification fails, we generate counterexamples and secrecy assumptions and present them to the user for feedback. We implement the back-end in roughly 9KLOC Haskell, using the Liquid-fixpoint (0.8.0.2) [8] and Z3 (4.8.1) [37] libraries for verification, and the GLPK (4.65) [7] library for synthesizing assumptions by solving the ILP problem of § 4. Our tool and evaluation data sets, including the secrecy assumptions discovered for SCARV (§ 7) are open source and available on GitHub. ³

## 6 EVALUATION

We evaluate Xenon by asking the following questions:

- Q1: Are constant-time counterexamples effective at localizing the cause of verification failures?
- Q2: Are the secrecy assumptions suggested by Xenon useful?
- Q3: What is the combined effect of counterexamples and secrecy assumption generation on the verification effort?
- Q4: Do module summaries improve scalability?
- Q5: Does Xenon reduce verification time by helping users find secrecy assumptions?
- Q6: How does using Xenon affect assumption quality?

To answer questions Q1 and Q2, we use Xenon to recover the assumptions for the benchmark suite from [50]. These benchmarks include a MIPS and RISC-V core, ALU and FPU modules, and RSA and SHA-256 crypto modules. To answer Q3 and Q4, we evaluate Xenon on two challenging new benchmarks, the SCARV “side-channel hardened RISC-V” processor [4] whose size exceeds the largest benchmark from [50] by a factor of 10, and a highly modular AES-256 implementation [10]. Finally, we conduct a user study to answer Q5 and Q6, in which participants were asked to find assumptions for three benchmarks from [50]: two benchmarks with relatively simple assumptions (ALU and FPU) and RISC-V core with a more complex assumption set. ⁴

**Summary.** Xenon’s counterexample synthesis dramatically reduces the number of potential error locations users have to manually inspect (6% of its original size) and most of Xenon’s assumption suggestions are accepted by the user (on average 81.67%). Module summaries are key to reducing verification times for certain hardware designs (e.g., for AES-256 crypto core summaries reduced the verification time from six hours to three seconds). We find the counterexamples and secrecy assumptions suggested by Xenon to be crucial to reducing the human-in-the-loop time from days to (at worst) hours. Our user-study findings indicate that—using Xenon—participants were able to correctly complete significantly more tasks in the given time frame, showing a very large (d = 1.62), statistically significant (t(8) = 2.56, p = .016) positive effect on correct completion. Participants in the test group produced fewer (d = 1.03) incorrect solutions (t(5.5) = 1.63, p = .07), and solution sizes were smaller on average.

**Experimental Setup.** We run all experiments on a 1.9GHz Intel Core i7-8650U machine with 16 GB of RAM, running Ubuntu 20.04 with Linux kernel 5.4.

**Methodology.** For every benchmark used to answer Q1–Q4, we start with an empty set of secrecy assumptions and run Xenon repeatedly to recover the missing assumptions needed to verify the benchmark. We collect the following information after every invocation of the tool: the total number of variables that are variable-time and secret; the size of the counterexamples measured by the number of variables they contain; the number of assumptions Xenon suggests, and how many of these assumptions we reject; finally, we record the number of times we invoke Xenon to complete each verification task. With all the assumptions in place, we measure the time it takes for the tool to verify each benchmark; we report the median of thirty runs for all but the non-modular (inlined) AES benchmark, for which—due to its size—we report the median of three runs.

**User-Study Design.** For our user study, we recruited ten participants who had some familiarity with software constant-time execution, but had never used Xenon or Iodine. The participants were randomly split into two equally sized groups: Test who were given Xenon and Control who were given Iodine. Participants using Iodine were given access to Iodine’s counterexample outputs. After reading the instructions, both groups were given 40 minutes to complete the three tasks, i.e., find assumptions for three Iodine benchmarks. For each task, we recorded the time taken to complete the task in minutes (Time), the number of annotations in the solution set (Size), and whether the solution was correct (Crt). We rejected solutions for ALU and FPU if they contained assumptions about the operands, and for RISC-V, if they contained assumptions about memory or the register-file.

**Q1: Error Localization.** To understand whether our counterexample generation is effective at localizing the cause of verification failures, we compare the number of variables in the counterexample to the total number of non-constant-time variables. The CEX Ratio

We include the larger RISC-V core to evaluate the hypothesis that Xenon benefits users even if many of Xenon’s suggestions are eventually rejected. We chose this benchmark because Xenon achieves the lowest accept-ratio over all benchmarks.

³We omit the link for the double-blind review process.
column Table 1 reports the average ratio per iteration. We observe that fewer than 6% of non-constant-time variables are included in the counterexample. Since the total number of non-constant-time variables is typically on the order of hundreds (e.g., the median (and geometric mean) number of non-constant-time variables across all benchmarks and iteration is 97 (94)), this dramatically reduces the number of variables the developer has to inspect in order to understand the violation. For the benchmarks that were variable time, the counterexamples also precisely pinpointed where in the circuit the constant-time property was violated. For example, in the FPU2 benchmark XENON included the state register in its third iteration counterexample. This register indicates when the FPU’s output is ready. Inspecting the register’s blame-set (similar to the process described in § 2.3) revealed that its value is set depending on whether one of the operands to the division operation is NaN and thus the FPU clearly leaks information about its operands.

**Q2: Identifying Secrecy Assumptions.** To assess the quality of secrecy assumptions suggested by XENON, we record the number of suggestions that the user accepts (useful suggestions) and the ratio of suggestions to the total number of secret variables the user would otherwise have to inspect manually. We find that most (on average 81.67%) of XENON’s suggestions are useful, reported in the Accept Ratio column of Table 1. Moreover, we observe that the number of variables included in the counterexamples is relatively small (Sugg Ratio column); on average, we only had to inspect 2.77% of the secret variables.

**Q3: Verification Effort.** Finally, as a rough measure of the overall verification effort, we count the number of user interactions, i.e., the number of times we invoked XENON after modifying our set of secrecy assumptions. Verifying the largest benchmark from [50], the YARVI RISC-V core [6] took five invocations over several minutes. The final assumptions we arrived at were the same as the assumptions manually identified by the authors of IODINE in [50]; they, however, took multiple days to identify these assumptions and verify this core [61]. Verifying the SCARV core took thirty-four iterations and roughly three hours; this core is considerably larger (roughly 10x) than the YARVI RISC-V core and, we think, beyond what would possible with tools like IODINE, which rely on manual annotations and error localization. Indeed, we found the error localization and assumption inference to be especially useful in narrowing our focus and understanding to small parts of the core and avoid the need to understand complex implementation details irrelevant to the analysis.

**Q4: Scalability.** To evaluate how module summaries affect the scalability, we compare the time it takes to verify (or show variable-time) a program with and without module summaries. Columns Inlined and Modular of Table 1 give the run times of XENON with inlining (no summaries) and module summaries, respectively. On the IODINE benchmarks (the first seven benchmarks), we observe that module summaries don’t meaningfully speed up verification. Indeed, on average, module summaries only reduce the size of the query sent to our solver by roughly 5% on these benchmarks. On the more complex AES-256 and SCARV benchmarks, however, the benefits of module summaries become apparent. For AES-256, using module summaries reduces the query size by 99.7%, from 391.3 MB to 1.2 MB, which, in turn, reduces the verification time by three orders of magnitude—from six hours to three seconds. Module summaries allow XENON to exploit the core’s modular design, i.e., AES-256’s multiple and nested instantiations of the same modules (see Figure 5). For SCARV, summaries reduce the query size by 41% and speed up the verification time by 40%. Though this reduction is not as dramatic as the AES-256 case, the speedup did improve XENON’s interactivity.

**Q5: Reducing Verification Time.** To determine whether XENON helps users find annotations more quickly, we recorded the number of tasks that participants were able to correctly complete within the 40 minutes timeframe. Column #Crt of Figure 13 summarizes our results. Participants in the test group completed 2.6 tasks on average, while participants in the control group were only able to solve 1.4. Figure 10 shows the percentage of participants that were able to find a correct set of assumptions, split by task. A little over 50% of control group participants were able to complete the first
We now describe our experience verifying SCARV and discuss the set of secrecy assumptions Xenon synthesized.

**SCARV: Overview.** SCARV is a 5-stage single-issue in-order CPU, implementing the RISC-V 32-bit integer base architecture. SCARV is side-channel hardened and explicitly designed to run cryptographic code. It supports an external hardware random number generator and implements fine-grained per-stage flushing of its processor pipeline via an instruction set extension.

**Finding Assumptions Modularly.** To verify SCARV, we follow Xenon’s modular philosophy: we start with modules that occur at leaf-level in the instantiation-tree, that is, modules that have no sub-modules of their own, and iteratively work our way up such that in each stage, we already determined the assumptions for all sub-modules. At each step, we prove that the current module is unconditionally constant-time, where we set all module inputs as sources and outputs as sinks. This keeps errors and assumptions local: At every stage of the verification process, we only have to think about the current module. But this approach has a downside. We might end up with a set of assumptions that is unnecessarily restrictive. Our modular verification process ensures that all input/output paths of all submodules are constant-time. But, to ensure constant-time execution of the entire circuit, constant-time execution of only a subset of modules and their respective input/output paths might be required. Fortunately, we can use the assumptions found via our modular verification process to bootstrap a search for a minimal assumption set. As Xenon’s module summaries can express that a module is constant-time only under certain conditions, and only for a subset of input/paths, we can safely erase assumptions, as long as Xenon can still prove the circuit to be constant time. Repeating this process yields a minimal assumption set, which we now discuss.

**Sources and Sinks.** Xenon represents assumptions as yaml files that are iteratively populated during verification. Figure 14 shows assumptions for the top-level module of SCARV. Annotations src and snk define sources and sinks, respectively. We choose all module inputs as sources, and all module outputs as sinks. This captures all relevant externally observable timing behaviors, including:

- The timing of signals interacting with both instruction and data memory, including requests (Lines 5 and 8), acknowledgments (Line 6), and strobe signals (Line 7),
- the timing of flush signals to external resources, such as caches (Line 9), and
- the timing of requests to the external random number generator, such as request ready bit (Line 10) and accept response bit (Line 11).

**Secrecy Assumptions: External Devices.** Annotation pub shows the secrecy assumptions synthesized by Xenon. At top-level, these assumptions concern external signals, hardware and interrupts. They require, for example, the external reset signal (Line 14), control inputs from external devices like memory (Lines 16 to 19), memory-mapped devices (Line 25), and the external random generator (Lines 21 to 23) to be public. The assumptions on memory are
Figure 13: Results of the user study. The participants were split into two equally sized groups: Test (Fig. 13a) using XENON and Control (Fig. 13b) using IODINE. The participants were asked to find assumptions for three of the IODINE benchmarks: ALU, FPU, and RISC-V. Both groups were given 40 minutes to complete the three tasks. For each task, we record the time taken to complete the task in minutes (Time), the number of annotations in the solution set (Size), and whether the solution was correct (Crt). We reject solutions for ALU and FPU if they contain assumptions about operands, and for RISC-V, if they contain assumptions about memory or the register file. We report the average ($\mu$) and standard deviation ($\sigma$) of all completed runs, i.e., including those that yielded wrong solutions. $\mu'$ and $\sigma'$ show average and standard deviation for correct runs only. Finally, we report the overall number of correctly completed tasks $\#Crt$. On average, participants in the test group were able to complete 2.6 tasks, while participants in the control group were only able to solve 1.4 within the 40 minutes trial. This indicates that using XENON has a very large ($d = 1.62$), statistically significant ($t(18) = 2.56, p = .016$) positive effect on correct completion.

8 LIMITATIONS AND FUTURE WORK

We discuss some of XENON’s limitations.

Assumptions about Data. XENON currently only discovers secrecy assumptions, i.e., whether a given value is public or private. It may be beneficial to also discover assumptions about data (e.g., that a certain flag is always set). In future work, we would like to explore how to combine XENON’s assumption synthesis method with techniques for inferring data preconditions [38, 74].

Minimality of Assumptions. XENON inherits the limitations of the underlying Horn solver (§ 3). In particular, an assumption set could be sufficient to ensure constant-time execution of the circuit, but the solver may be unable to prove it. The minimality of our assumption set (§ 7) is therefore relative to the solver, and could potentially be improved with more precise solving methods—at the cost of reduced interactivity and scalability. As future work, one could use fast over-approximating solvers in the assumption discovery phase of our verification method (§ 7), and then slower, more precise solvers to minimize the assumption set after bootstrapping.

Mapping Back to Software. XENON discovers an assumption set that ensures constant-time execution of the verified design. But, it leaves open the question of how to map assumptions back into proof obligations on software. The assumption set XENON discovered for SCARV (§ 7) suggests that this might require a whole system effort that goes beyond current practices of constant-time programming. We hope that open-sourcing assumptions for SCARV will help future research efforts in this direction.

Guarantees on Synthesized Circuits. Finally, we prove constant-time at the Verilog level. This is convenient for error-localization, but it doesn’t ensure that guarantees carry over to the generated circuits. Proofs are guaranteed to carry over if the synthesizer produces behavior within the Verilog standard [13], as, for example, formalized in [50, 51]. In particular, we make no further assumptions

<table>
<thead>
<tr>
<th>Task 1 (ALU)</th>
<th>Task 2 (FPU)</th>
<th>Task 3 (RISC-V)</th>
<th>#Crt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Size</td>
<td>Crt</td>
<td>Time</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>✓</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>✓</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>✓</td>
<td>18</td>
</tr>
<tr>
<td>21</td>
<td>3</td>
<td>✓</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>✓</td>
<td>4</td>
</tr>
</tbody>
</table>

| µ | 7.80 | 3.00 | µ̇ | 12.00 | 3.00 |
| σ | 7.92 | 0.00 | σ̇ | 4.24 | 0.00 |

| µ' | 7.80 | 3.00 | µ̇' | 10.33 | 3.67 |
| σ' | 7.92 | 0.00 | σ̇' | 2.31 | 0.57 |

(a) Test Group: Using XENON

(b) Control Group: Using IODINE

Secrecy Assumptions: Internal Processor State. While, in our experience, top-level assumptions about IO behavior are relatively easy to find, proving constant-time execution also requires harder to find assumptions about processor internals. These assumptions encode constraints on the kind of programs that can safely be executed on the processor. Figure 15 shows the assumptions for SCARV’s pipeline module. XENON discovers the classic constant-time assumptions stating that control-flow (Lines 4 and 5) and memory-trace (Line 7) are secret independent. Similarly, memory stalls (Line 9), instruction validity (Line 11), and the computed $\mu$ (Line 13) must not depend on secrets.

But XENON also discovers systems-level assumptions that are not commonly associated with constant-time programming. For example, access errors for control and status (CSR) registers (Line 15) must not depend on secrets, and the timing of when to return from machine-mode (Line 17) and traps (Lines 21 and 22), must be public. Finally, we may not set the configuration register of SCARV’s leakage fence instruction (Line 19) depending on secrets. This is necessary as different configurations flush different parts of the pipeline and might incur different delays.

Constant-Time Subset of Instructions. The above assumptions are satisfied, when the instructions run on SCAR-V are limited to the arithmetic, and bitwise-logic subset of RISC-V; instructions must be valid, i.e., properly encoded; even division is constant-time.
9 RELATED WORK

Verifying Leakage Freedom. There are various techniques, such as ct-verif [16, 22], and CT-WASM [85], that verify constant-time execution of software, and quantify leakage through timing and cache side-channels [12, 17, 41, 65, 79, 90]. However, their analyses do not directly apply to our setting: They consider straight-line, sequential code, unlike the highly parallel nature of hardware. There are many techniques for verifying information flow properties of hardware. Kwon et al. [64] prove information flow safety of hardware for policies that allow explicit declassification and are expressed over streams of input data. SeeVerilog [91] and Caisson [66] use information flow types to ensure that generated circuits are secure. GLIFT [82, 83] tracks the flow of information at the gate level to eliminate timing channels. Other techniques such as HyperFlow [45], GhostRider [67] and Zhang et al. [90] take the hardware and software co-design approach to obtain end-to-end guarantees. dudect [77], detects end-to-end timing variability across the stack via a black-box technique based on statistical measurements. IoDine [50], like Xenon, focuses on clock-precise constant-time execution, not information flow. Unlike Xenon, none of these methods provides help in elucidating secrecy assumptions, in case the verification fails—a feature we found essential in scaling our analysis to larger benchmarks. We see the techniques presented in this paper as complementary and would like to explore their potential for scaling existing verification methods for hardware and software.

Fault Localization. There are several approaches to help developers localize the root causes of software bugs [86]. Logic-based fault localization techniques [31, 42, 59, 60] are the closest line of work to ours. For example, BugAssist [59] uses a MAXSAT solver to compute the maximal set of statements that may cause the failure given a failed error trace of a C program. Xenon is similar in that we phrase localization as an optimization problem, allowing the use of ILP to locate the possible cause of a non-constant-time variable. However, Xenon focuses on constant-time, which is a relational property, and hardware which has a substantially different execution model.

Synthesizing Assumptions. Our approach to synthesizing secrecy assumptions is related to work on precondition synthesis for memory safety. Data-driven precondition inference techniques such as [47, 48, 72, 80, 81], unlike Xenon, require positive and negative examples to infer preconditions. Xenon’s synthesis technique is an instance of abductive inference, which has been previously used to leverage analysis reports by allowing the user to interactively determine the preconditions under which a program is safe or unsafe [38] or to identify the most general assumptions or context under which a given module can be verified safe [14, 27, 39, 40, 49]. Livshits et al. [68] infer information-flow specifications for web-applications using probabilistic inference. Unlike these efforts, our abduction strategy is tailored to the relational constant-time property. Furthermore, Xenon uses information from the verifier to ensure that the user interaction loop only invokes the ILP solver (not the slower Horn-clause verifier), yielding a rapid cycle that pinpoints the assumptions under which a circuit is constant-time. In future work, we would like to see, if ideas introduced in Xenon can be applied to localization, explanation and verification of other classes of correctness or security properties.

Modular Verification of Software and Hardware. Xenon exploits modularity to verify large circuits by composing summaries of the behaviors of smaller sub-components of those circuits. This is a well-known idea in verification; for example, [78] shows how to perform dataflow analysis of large programs by computing procedure summaries, and Houdini [46] shows how to verify programs by automatically synthesizing pre- and post-conditions summarizing the behaviors of individual procedures. On the hardware side, model checkers like Mocha [18] and SMV [69] use rely-guarantee reasoning to perform modular verification. Kami [30] and [84] develops a compositional hardware verification methodology using the Coq proof assistant. However, the above require the user to provide module interface abstractions. There are some approaches that synthesize such abstractions in a counterexample guided fashion [54, 92]. All focus on functional verification of properties of a single run, and do not support abstractions needed to reason about timing-channels which require relational hyper-properties [32].
A EXAMPLE: NOT ALL VARIABLES BECOME PUBLIC

Example 3. One might think that XENON requires all variables occurring in branch conditions to be annotated as public, however, this is not the case. Appendix A shows an example of such a program. Running XENON produces the dependency graph shown in Figure 16. XENON computes root-cause candidates by eliminating constant-time nodes and edges violating the precedence order. The result is shown in Figure 17. Removing all nodes that cannot reach source out leaves only nodes r3 and out, and since r3 has no predecessors, we identify it as the earliest node that became non-constant time, and therefore the root cause of the problem. Solving the ILP constraints yields stall as candidate assumption, and marking stall as public and restarting XENON verifies constant time execution without the need to mark cond as public. This is possible because XENON is able to prove that tmp1 and tmp2 have the same liveness-bits, irrespective of the value of cond, i.e., that tmp1′=tmp2′ holds irrespective of cond.

```verilog
module test (clk, in, cond, bubble, out);
    input wire clk, in, cond, bubble;
    output reg out;
    reg tmp1, tmp2, r2, r3;
    always @(posedge clk) begin
        tmp1 <= in | r3;
        tmp2 <= in & r3;
        if (cond)
            r2 <= tmp1;
        else
            r2 <= tmp2;
        if (stall)
            r3 <= r3;
        else
            r3 <= r2;
        out <= r3;
    end 
endmodule
```

Figure 16: Example 3: Variable dependency graph.

Figure 17: Example 3: Variable dependency graph after eliminating non-ct nodes and edges that violate the precedence relation.