

oo7: Low-overhead Defense against Spectre Attacks via Program Analysis

Guanhua Wang, National University of Singapore (wangghge@gmail.com)
Sudipta Chattopadhyay, Singapore University of Technology and Design
Ivan Gotovchits, Carnegie Mellon University
Tulika Mitra & Abhik Roychoudhury, National University of Singapore

Abstract—The Spectre vulnerability in modern processors has been widely reported. The key insight in this vulnerability is that speculative execution in processors can be misused to access the secrets. Subsequently, even though the speculatively executed instructions are squashed, the secret may linger in micro-architectural states such as cache, and can potentially be accessed by an attacker via side channels. In this paper, we propose *oo7*, a static analysis approach that can mitigate Spectre attacks by detecting potentially vulnerable code snippets in program binaries and protecting them against the attack by patching them. Our key contribution is to balance the concerns of effectiveness, analysis time and run-time overheads. We employ control flow extraction, taint analysis, and address analysis to detect tainted conditional branches and speculative memory accesses. *oo7* can detect all fifteen purpose-built Spectre-vulnerable code patterns [1], whereas Microsoft compiler with Spectre mitigation option can only detect two of them. We also report the results of a large-scale study on applying *oo7* to over 500 program binaries (average binary size 261 KB) from different real-world projects. We protect programs against Spectre attack by selectively inserting fences only at vulnerable conditional branches to prevent speculative execution. Our approach is experimentally observed to incur around 5.9% performance overheads on SPECint benchmarks.



1 INTRODUCTION

The Spectre [2] vulnerabilities in processors were revealed in early 2018. The attacks that exploit these vulnerabilities can potentially affect almost all modern processors irrespective of the vendor (Intel, AMD, ARM) and the computer system (desktop, laptop, mobile) as long as the processor performs speculative execution. Speculative execution [3] is an indispensable micro-architectural optimizations for performance enhancement, ubiquitous in almost all modern processors except for the simplest micro-controllers. It is an aggressive optimization where the instructions are executed speculatively, but the temporary results created by the speculatively executed instructions are maintained in internal micro-architectural states that cannot be accessed by software. The results are committed to the programmer-visible architectural states (registers and memory) only when the speculation is found to be correct; otherwise, the internal micro-architectural states are flushed. The most common example is that of the conditional branches being predicted in hardware and the instructions along the predicted branch path are executed speculatively. Once the conditional branch direction is resolved, the instructions along the speculative path are squashed in case of wrong prediction.

Spectre attacks exploit speculation to deliberately target the execution of certain “transient” instructions. These transient instructions are speculatively executed, and are tricked to bring in secret data into the cache. These transient instructions are subsequently squashed but the secret remains, for example, in the cache. The attacker then carefully accesses the secret content (that is supposed to be hidden to the outside world) through different micro-architectural covert channels, for example, cache side-channel [4]. The

website [5] of Spectre states that “As [Spectre] is not easy to fix, it will haunt us for a long time.”

We focus on identifying program binaries that are vulnerable to Spectre attack and patch those binaries as a mitigation technique with minimal performance overhead. We present a comprehensive and scalable solution, called *oo7*, based on static program analysis. Our solution employs control flow extraction, taint analysis and address analysis at the binary level. Moreover, *our analysis needs to model the transient instructions along the speculative path that has never been required in traditional program analysis dealing with only programmer visible execution. We have successfully introduced accurate modeling of speculative execution in oo7.*

Once vulnerable code snippets are detected by *oo7*, we introduce fence instructions at selected program points to prevent speculative execution and thereby protect the code from Spectre attack. We have validated the functional correctness of our protection mechanism with all fifteen litmus test codes from [1] on Intel Xeon platform. We note that the current Spectre mitigation approach introduced by Microsoft C/C++ compiler [6], detects and protects only 2 out of 15 litmus tests for Spectre vulnerabilities [1], whereas *oo7* can detect all fifteen purpose-built Spectre-vulnerable code patterns. We can launch successful Spectre attack to access arbitrary locations in the victim code prior to the insertion of fence insertions by *oo7*; but our attempts at Spectre attacks fail after *oo7*-directed automated identification and patching of the victim code. We experimentally measure the performance overheads from our selective fence insertion and find that the overheads are 5.9% on average on SPECint benchmarks, thereby indicating the practicality of our approach. We also report the results of a large-scale experimental study on applying *oo7* to over 500 program

binaries (average binary size 261 KB) from different real-world projects.

We demonstrate that *oo7* can be tuned to defend against multiple different variants of Spectre attack (see TABLE 1) that exploit vulnerabilities in the victim code through speculative execution. We also note the limitations of our analysis-based approach in defending against certain variants of Spectre attacks. The variants that cannot be addressed by *oo7* have potential system-level solutions introduced by different vendors with reasonably low overhead [7], [8]. The Spectre variants handled by *oo7* with low performance overhead are either not amenable to system-level defense mechanisms, incur high performance overhead or escape detection with existing approaches. Thus *oo7* approach via binary analysis is complementary to all other efforts in mitigating the impact of security vulnerabilities due to speculative execution.

Contributions

The contributions of this paper can be summarized as follows. First, we present a program analysis based approach called *oo7* for mitigating Spectre attacks. Our solution is based on binary analysis and does not involve changes to the underlying operating system and hardware. It uses taint analysis, address analysis and speculation modeling to check potentially vulnerable program binaries, and inserts a small number of fences to mitigate the risks of Spectre attack. Our approach is accurate in identifying all the litmus tests for Spectre vulnerabilities [1], has low performance overhead (average 5.9% overhead for SPECint benchmark suite), and is scalable as evidenced by our analysis of over 500 large program binaries.

The main contribution of this work is in proposing and demonstrating an efficient static analysis based approach to accurately locate potential Spectre vulnerability in the code and then use well-established repair strategy (fences) to fix these selected vulnerable code fragments. We have successfully introduced accurate modeling of speculative execution in taint analysis to achieve this. The existing solutions cannot identify the vulnerable code fragments and hence repair all conditional branches to prevent speculative execution altogether resulting in significant performance overhead.

We show that our program analysis based approach can detect and mitigate certain variants of Spectre vulnerabilities in the application code, but not all (see TABLE 1). Thus our work provides an understanding of the class of attacks for which an analysis based mitigation may be suitable, and for which a system level solution is suitable.

So far, no Spectre attack has been found in the wild. We hope that the search for zero day Spectre attack in the wild can be substantially accelerated via community participation using our tool.

2 SPECTRE VARIANTS

A number of Spectre vulnerabilities that all take advantage of speculative execution in modern processors have been disclosed recently. A summary of these variants appear

in TABLE 1. We classify the different vulnerabilities into three categories:

(a) *Vulnerability in victim code*: Many Spectre attacks rely on vulnerable code snippets inside the victim process and trigger speculative execution of the code snippet to read secret data by supplying carefully selected inputs to the victim process. We detect these vulnerabilities in *oo7* by identifying the potentially susceptible code fragments via binary analysis and then introducing fences at selected program points to prevent speculative execution and thereby harden the victim software against any such attacks.

(b) *BTB or RSB poisoning*: In these Spectre variants, the attacker poisons the Branch Target Buffer (BTB) or Return Stack Buffer (RSB) in the micro-architecture. The victim process, while using the poisoned BTB or the RSB for speculative execution, is then misled to branch or return to a gadget that leaks the sensitive data. Any indirect branch or return instruction in the victim code is vulnerable to this attack and hence we do not attempt to mitigate these attacks in *oo7*. There exist potential solutions such as Retpoline [9] or RSB refilling [10] for these vulnerabilities.

(c) *Transient out-of order execution*: These attacks can be directly launched by a malicious code (malware) without the requirement of any specific vulnerable code fragment or pattern in the victim process. Since the objective of *oo7* is to detect and repair vulnerable code fragments in general software, the detection of malware is orthogonal to the objective of *oo7*. Thus, we do not consider the detection of Spectre-style malicious code fragments in this work.

Unlike the first class of attacks where the defense mechanism is to harden the victim software, here *oo7* performs malware detection, i.e., it looks for malicious code patterns within a binary.

2.1 Vulnerability in Victim Code

Spectre Variant 1: The following victim code fragment exhibits Spectre vulnerability Variant 1.

```
void victim_function_v01(size_t x) {  
    if (x < array1_size) { //TB: Tainted Branch  
        y = array1[x]; //RS: Read Secret y  
        temp &= array2[y * 256]; //LS: Leak Secret y  
    }  
}
```

In this example, the parameter x is under the attacker control in the sense that x can be influenced by external input. Hence we consider the conditional branch as a **Tainted Branch (TB)**. The attacker first trains the branch predictor to expect that the branch will be true (i.e., the array bound check will pass). The attacker then invokes the code with an input x value outside the bound of `array1`. The branch predictor expects the branch condition to be true and the CPU speculatively reads y using malicious value x outside the array bound. We call this action **Read Secret (RS)** because y can be a potential secret that is not legitimately accessible through malicious input without speculation. This is followed by the CPU speculatively accessing `array2` using an address that is dependent on the secret y leading to cache state change. We call this action **Leak Secret (LS)** because the change in the cache state lingers even after the CPU realizes that the branch prediction was wrong and squashes the speculatively executed instructions. The attacker can now launch cache side-channel attack [4] to detect this

TABLE 1: The existing speculative execution based attacks and the ability of oo7 for handling them

| Classification | Exploit name | Public vulnerability name | oo7 capability |
|------------------------------|----------------------|-----------------------------------|---|
| Vulnerability in victim code | Spectre variant 1 | Bounds Check Bypass (BCB) | Detect and patch vulnerable victim code |
| | Spectre variant 1.1 | Bounds Check Bypass Store (BCBS) | Detect and patch vulnerable victim code |
| | Spectre variant 1.2 | Read-only protection bypass (RPB) | Detect and patch vulnerable victim code |
| | Spectre-NG variant 4 | Speculative Store Bypass (SSB) | Potentially possible but not handled yet by oo7 |
| BTB or RSB poisoning | Spectre variant 2 | Branch Target Injection (BTI) | - |
| | Spectre RSB | Return Mispredict | - |

change in cache state and discover the secret y . Specifically, for Prime+Probe side-channel attack, the attacker ensures that `array2` was not cached before the memory access `LS` by evicting the cache line through priming the cache set. Then the attacker triggers `LS` action to leak the secret to the cache side channel. Finally, the attacker performs the probe phase to get the timing of the memory accesses for `array2` and discovers the value of y . The multiplier `256` in `array2[y*256]` guarantees that different values of y lead to different cache line access, and normally, this value is greater than or equal to the cache line size.

Spectre Variant 1.1: The idea behind the Spectre Variant 1.1, also known as Bounds Check Bypass Store (BCBS), is to bypass bound check and execute a store instruction speculatively [11]. In the following example, x can potentially be under attacker control, hence, the conditional $x < \text{array1_size}$ is a *Tainted Branch (TB)*. However, unlike the *Read Secret (RS)* in Spectre Variant 1, this variant uses a **Speculative Write (SW)** to modify arbitrary memory location. For instance, the example modifies an arbitrary memory location pointed to by `array1[x]` when the conditional branch is mispredicted for a value $x \geq \text{array1_size}$. Although this speculative store is squashed upon resolving the branch outcome, it can leak secret values from the program. For instance, `array1[x]` may overwrite the return address and transfer control to a gadget that leaks arbitrary secret value via a side-channel (similar to `LS` in Spectre Variant 1).

```
void victim_function_v1.1(size_t x, y) {
    if (x < array1_size) { //TB: Tainted Branch
        array1[x] = y; //SW: speculative Write
    }
}
```

Spectre Variant v1.2: This vulnerability bypasses the protection enforced by read-only memory, e.g., code pointers [11]. Consider the `victim_function_v1.1` where the valuation of the conditional captures whether x points outside the read-only memory. If x is under attacker control, then the write to a read-only memory can be speculatively executed and modify crucial data structures such as code pointers in the cache. As a result, like Spectre Variant 1.1, the program control may transfer to arbitrary location to execute attacker chosen code. Like Spectre Variant 1.1, this variant also requires the presence of *TB* and *SW*.

Spectre-NG Variant 4: Spectre Variant 4, also called Speculative Store Bypass (SSB), is based on the fact that the processor may execute a load instruction speculatively even when a prior store instruction in program order is pending because the address for the store is not yet known. Thus a speculative load may read a stale value that should have been modified by a prior store instruction if they access the same memory address; in that case, the speculative load should be squashed after the store address is known.

`oo7` can detect and patch victim binary code with potential Spectre variant 1, 1.1. and 1.2 vulnerabilities. `oo7` can potentially handle Spectre variant 4 by identifying the vulnerable code pattern but requires precise address analysis (that the load and the store are accessing the same memory address) that is not supported yet in our framework.

2.2 BTB or RSB poisoning ¹

Spectre Variant 2: Most architectures support indirect branches in the form of “`jmp [r1]`”. For such jump instructions, the program control is diverted to a location stored in the register `r1`. For improving program performance, the processor leverage Branch Target Buffer (BTB) to store the frequently used target locations of branch instructions, including indirect branches. An attacker can poison the Branch Target Buffer (BTB) to include its preferred target locations. When the victim executes an indirect branch instruction, it consults this poisoned BTB and the speculative execution can potentially be misled to a target location chosen by the attacker. Any indirect branch is vulnerable to this attack. The indirect branches can be easily identified by static analysis and mitigated by Retpoline [9] approach.

SpectreRSB: SpectreRSB vulnerability [10] is similar to the Spectre Variant 2. Instead of poisoning the BTB with attacker chosen location, the SpectreRSB vulnerability manipulates the return stack buffer (RSB), which is used by the processor to predict the return address. As a result of a successful exploit, a function may return to an attacker controlled location due to the mis-prediction of the return address inflicted by an attacker. Subsequently, the program may execute arbitrary code in the attacker-controlled location until the return address is finally resolved. All return instructions are potentially vulnerable to such exploit. RSB refilling is a potential approach to mitigate SpectreRSB [10].

Our proposed `oo7` approach focuses on identifying vulnerable code fragments (e.g. attacker controlled branch instructions) that can be fixed by the insertion of memory fence instructions. This is to prevent speculative execution with the objective to prevent the attack. Since Spectre Variant 2 and Spectre RSB need different solutions than patching fences, our `oo7` approach does not target fixing these subset of Spectre vulnerabilities.

3 RELATED WORK

3.1 Mitigation in the future processors

Intel has reportedly developed hardware fixes [12] in the form of improved process and privilege-level separation for

¹ Even though these are considered as Spectre variants, they are very different from the vulnerability in victim code.

only Spectre Variant 2. Three capabilities: Indirect Branch Restricted Speculation, Single Thread Indirect Branch Predictors, Indirect Branch Predictor Barrier will be supported in future products to mitigate the branch target injection attack. Vladimir et al. [13] proposed DAWG, which is a generic mechanism to isolate the cache side-channel by partitioning the cache ways to limit the data leakage across different secure domains. InvisiSpec [14] is another new architecture design to defend against Spectre-like attacks. InvisiSpec uses a Speculative Buffer (SB) to temporarily hold the data during speculative execution instead of directly loading the data to the cache. The data in SB will be finally visible to the cache hierarchy when the speculative load is safe to be committed. InvisiSpec slows down the execution by 21%. Obviously, both DAWG and InvisiSpec cannot be used in the legacy systems.

3.2 Mitigation in Legacy systems

Several approaches have been applied to mitigate Spectre attacks in legacy system. Microsoft Visual C/C++ compiler [15] provides a compiling option `Qspectre` to enable the mitigation of Spectre Variant 1 by inserting `lfence` serializing instruction in the potential vulnerable code locations. However, the mitigation technique can detect only 2 out of the 15 litmus tests proposed by Paul Kocher [1]. Speculative Load hardening [16] mitigates Spectre Variant 1 by inserting hardening instruction sequences that zeros out the pointers that have data dependency with the branch conditions. As it inserts hardening instructions at all conditional branches, the technique involves 36.4% performance overhead. Oleksenko et al. [17] propose the introduction of artificial data dependencies to protect from Spectre attacks. This solution is coarse-grained and will effectively disable speculation between any conditional branch and subsequent load instructions. Moreover, the authors explicitly acknowledge the absence of precise taint tracking and leaves it to the developer to examine whether the potential vulnerable locations reported by the tool can be controlled by the attacker. Microsoft has developed Windows patches [18] through CPU microcode update for Spectre Variant 2 (but not for Variant 1). Moreover, this update has been reported to cause performance overhead (specially on older platforms) and system instability. Google Chrome has developed "Site isolation" mechanism that sandboxes the memory pages associated with each website to a separate process [19] at the cost of 10–13% memory overhead. In contrast, `oo7` does not require either operating system or processor changes. Retpoline [20] has been proposed for the gcc and LLVM compiler to mitigate Spectre version 2. Retpoline replaces vulnerable indirect branches with non-vulnerable instruction sequence that forces the CPU to jump to the real destination instead of the predicted target suggested by the BTB. Recently [21] has proposed the use of symbolic execution and SMT solver for detecting Spectre Variant 1. This is a higher overhead and less scalable approach than ours. Moreover, no mitigation is proposed, as we do by inserting fence instructions.

Compared with all existing approaches for mitigation in the legacy system, `oo7` introduces the lowest performance overhead (around 5.9% on SPECint) as `oo7` only hardens a small number of branches in repairing Spectre-like vulnerabilities. Moreover, `oo7` is a flexible approach that can

be tuned to defend against different variants of Spectre (as shown in this work) as well as detect malicious code fragments for Spectre-NG variant 3a and Spectre-NG LazyFP.

4 BRIEF OVERVIEW OF OUR APPROACH

Our approach to identify vulnerable code fragments for Spectre variants 1, 1.1, 1.2 or malicious code fragments for Spectre-NG variant 3a and Spectre-NG LazyFP proceeds via static taint analysis of program binaries. *All input sources including files are initially marked as tainted.* Taint propagation across instructions proceeds by usual computation of forward data and control dependencies. Thus, for data dependency based taint propagation, if any of the operands of an instruction is tainted, the result of the instruction is tainted. For control dependency based taint propagation (also called as implicit flows in taint analysis literature), the decision taken by a branch, and hence the instructions conditionally executed owing to the decision are tainted; the identifiers (memory/registers) written by such tainted instructions are also treated as tainted. Details of the formal treatment of taint propagation policies appear in Section 5.

One of the novel aspects of our analysis is in considering speculative execution paths while capturing taint propagation. Conceptually this is handled by considering both possibilities in a branch b , and checking which instructions fall inside the speculative execution window of a branch meaning they can be speculatively executed prior to a branch's outcome being known. There is no need to explicitly maintain the speculatively executed paths as a separate set of bounded length paths, as long as we consider both directions of a branch in our analysis. An instruction i can be speculatively executed pending the outcome of a branch b , only if the distance between i and b is less than the speculative execution window set by the processor.

To check Spectre attack scenarios such as Spectre variant 1, we then need to look for a tainted branch instruction (TB), a load-like instruction which reads secret (RS) with the memory address read by RS being tainted, and RS being potentially speculatively executed owing to TB being unresolved. Note that once the occurrence of TB and RS are established, the secret data has already been speculatively accessed and can be potentially ex-filtrated via various side channels. To detect an instance of potential Spectre variant 1 vulnerability that is consummated via a cache side-channel attack, we need to also locate an instruction LS to leak the secret, where the memory address accessed by LS is dependent on the output of RS . Once again, the detailed treatment of the condition for checking Spectre variant is deferred to the next section.

For forward taint propagation along all possible paths, we use the Binary analysis Platform (BAP) tool [22]. As BAP is based on conservative analysis, it can report false positives. BAP leverages a set of techniques to construct the control flow graph for a binary. As is well known in binary analysis literature, accurate construction of control flow graph is a notoriously difficult problem owing to indirect branches. Tools like BAP use forced execution [23] and other techniques to construct a control flow graph. Forced execution (i.e., execution of both branch directions) is leveraged to construct all possible control flow edges of

a branch. As the control flow paths are constructed, taint is propagated along the paths as per the taint sources and taint policies set in our approach (please see next section for details).

On the challenges of binary analysis: During the translation of source code to the binary representation, optimizing compilers usually forfeit such properties as type preservation and control flow integrity, thus not only opening a possibility to a wide range of security vulnerabilities, such as buffer-overflow and control-flow hijacking but also complicating the task of reverse engineering and making sound static analysis of binary code merely impossible. The problem of control flow reconstruction is especially hard due to indirect branches where the branch targets are difficult to determine at compile-time. Therefore, any analysis of the binary code is doomed to be an approximation and a lot of care [24] should be taken to preserve the correctness of any analysis, and especially if that analysis involves a fixed point computation on the control flow graph, such as data flow analysis. In *oo7* we leverage Primus, a Microexecution framework that tries to preserve the program behavior of the binary, while extracting control flow graph from the binary.

On false negatives: Our taint propagation policies and rules (see Section 5) avoid under-tainting, thus, in their own merit, they do not introduce any false negatives. Nonetheless, as our approach is based on the control flow graph constructed by the underlying tools, we cannot guarantee zero false negatives, if the control flow graph constructed by BAP does not capture all possible flows. In other words, the completeness of our taint analysis is modulo the completeness of the control flow graph extraction in BAP. Also, in BAP, loops are unrolled to track the program dependencies across loop iterations and might be a source of false negatives in BAP analysis if the unrolling depth is low. However, with correctly provided loop bounds, this problem can be alleviated.

On taint analysis and symbolic execution: In identifying vulnerable code susceptible to many Spectre attack variants, the key is to find attacker controlled branches, and memory locations that can be speculatively read/written pending outcome of attacker controlled branches. For this reason, we have employed static taint analysis and considered read/write instructions that may be speculatively executed within the speculative execution window. It is possible to take a different, albeit higher-overhead approach. For example, instead of capturing the taint propagation along paths, one may summarize the execution behavior in a more fine grained fashion using symbolic execution, where the memory locations accessed by a read/write is captured as symbolic expressions over (tainted) input, instead of simply maintaining that the location accessed by a read/write is tainted. This leads to additional overhead of constraint accumulation and solving. In this way, higher overhead detection approaches can be constructed such as the recent work of [21] that handles Spectre variant 1 and potentially variant 1.1. This work [21] only reports results for different variants of the Spectre litmus tests with few lines of code. As is shown by our experiments, even a low overhead approach like ours consumes high analysis times on a few real-life programs from SPECint and OSS-

Fuzz. Hence we feel higher overhead approaches are not scalable enough for possible real-life usage. Furthermore, for any Spectre attack detection approach, it is not enough to detect leaks, the analysis needs to suggest a small number of fences to plug the leak. This is done in our approach. On the other hand, symbolic approaches like [21] simply detect leaks without suggesting concrete fence instructions to harden Spectre vulnerable code.

5 SPECTRE VULNERABILITY DETECTION

To describe our analysis, we use the notations in TABLE 2. We say that an instruction *inst* is *tainted*, i.e., $\tau(inst)$ is true, if and only if the instruction operates on some tainted operands. We first discuss the checker for Spectre Variant 1.

5.1 Detecting Spectre Variant 1

An important concept that we need for our analysis is the *Speculative Execution Window*, abbreviated *SEW*. We posit that information about *SEW* needs to be exposed by processor designers for the sake of detecting Spectre attacks. By default, it seems that *SEW* can be set to the size of the re-order buffer in an out-of-order processor. However, if the size of the re-order buffer is n , it is not sufficient to have a lookahead of n instructions from a tainted conditional branch *TB*, in our search for memory access *RS*, in order to detect Spectre attacks. For processor execution, each instruction is decoded to a sequence of micro-ops. Each micro-op will occupy one slot of the re-order buffer during execution. However, micro-ops can be fused [25] both within an instruction as well as across instruction. When micro-ops are fused across instructions (also called macro-fusion), the micro-ops of at most two instructions can be fused into a single micro-op. For this reason, if the size of the re-order buffer is n , we conservatively set the Speculative Execution Window to $2n$ in our analysis, so as to avoid any false negatives in our analysis.

$$\Phi_{spectre} \equiv \boxed{\begin{aligned} & br(TB) \wedge load(RS) \wedge mem(LS) \wedge \\ & \tau(TB) \wedge \tau(addr(RS)) \wedge \tau(addr(LS)) \wedge \\ & Dep(RS, addr(LS)) \wedge (\Delta(TB, RS) \leq SEW) \wedge (\Delta(TB, LS) \leq SEW) \end{aligned}} \quad (1)$$

We now elaborate the checking condition for detecting Spectre. *oo7* locates *TB*, *RS* and *LS* by checking $\Phi_{spectre}$. Intuitively, the first two lines of $\Phi_{spectre}$ capture the presence of tainted branch instructions *TB* and tainted memory-access instructions *RS* and *LS*. The last line shows that *RS* and *LS* are located within the speculation window of *TB*, and they are data-dependent. $\Phi_{spectre}$ reflects Spectre variant 1. Later we show that $\Phi_{spectre}$ can easily be modified to detect other Spectre variants, for example, variant 1.1.

5.2 Taint Analysis

We use taint analysis [22] to determine whether conditional branch instructions (e.g., *TB*) and the memory-access instructions (e.g., *RS* and *LS*) can be controlled via untrusted inputs. In the following, we outline the taint propagation policies and rules used to detect Spectre vulnerabilities. To

TABLE 2: Symbols used in describing *oo7*

| Symbol | Interpretation |
|------------------------|--|
| $br(inst)$ | $inst$ is a branch instruction |
| $time(inst)$ | $inst$ is an timing instruction, e.g., <code>rdtsc</code> |
| $mem(inst)$ | $inst$ is a memory access instruction |
| $load(inst)$ | $inst$ is a load instruction |
| $addr(inst)$ | the data memory address accessed by a memory-related instruction $inst$ |
| $reg(inst)$ | set of registers accessed by instruction $inst$ |
| $set(addr)$ | cache set accessed by memory address $addr$ |
| $\tau(inst)$ | instruction $inst$ is tainted |
| $\tau(x)$ | instruction operand x is tainted where x could be a register, memory location or the value located in a memory location |
| $\Delta(inst1, inst2)$ | minimum no. of instructions executed to reach $inst2$ from $inst1$. If $inst2$ is unreachable from $inst1$, then $\Delta(inst1, inst2) = \infty$. |
| $Dep(inst, x)$ | x is data-dependent on instruction $inst$ |
| $CDep(inst)$ | set of instructions control-dependent on $inst$ |
| $val(x)$ | value located at memory address x |
| SEW | Speculative Execution Window = $2n$, where n is the size of re-order buffer in the processor |

illustrate our taint propagation policy, we use the following instructions.

- $z = x \text{ op } y$: Binary operation on register x and register y . The operation `op` can be either arithmetic operation (e.g., addition or subtraction) or a logical operation (e.g., a logical comparison).
- $y = \text{op } x$: Unary operation on register x . The operation `op` can either be arithmetic (e.g., unary minus) or a logical one (e.g., logical negation).
- $y = \text{load}(x)$: Loads value from memory address x to register y .
- $y = \text{store}(x)$: Stores value from register x to memory address y .
- `branch(L, x)` : Branch to label L if the logical formula x is *true*.

Taint Propagation Policies:

Initially, all variables that read value from un-trusted sources (e.g., files, network) are tainted. The taints from these variables are then propagated via a well-defined set of rules shown in the following; for each rule, the premises appear on the top of the horizontal bar and the conclusions appear below the horizontal bar. Our taint propagation tracks both data dependencies and control dependencies (also known as implicit flows in taint analysis). Typically such implicit flows come in the form of the tainted data enabling or disabling a branch condition b , and the outcome of b affecting the computation of a variable that would not be tainted otherwise purely by tracking of data dependencies.

Taint Propagation Rules

$$[\text{Binary operation}] \frac{z = x \text{ op } y \quad \tau(x) \vee \tau(y)}{\tau(z) \wedge \tau(inst)}$$

$$[\text{Unary operation}] \frac{y = \text{op } x \quad \tau(x)}{\tau(y) \wedge \tau(inst)}$$

$$[\text{Memory load}] \frac{y = \text{load}(x) \quad \tau(val(x))}{\tau(y) \wedge \tau(inst)}$$

$$[\text{Memory load}] \frac{y = \text{load}(x) \quad \tau(x)}{\tau(y) \wedge \tau(inst)}$$

$$[\text{Memory store}] \frac{y = \text{store}(x) \quad \tau(x)}{\tau(val(y)) \wedge \tau(inst)}$$

$$[\text{branch}(L, x)] \frac{br(inst) \quad \tau(x)}{\tau(inst)}$$

$$[\text{branch}(L, x)] \frac{\tau(x) \quad Tinst = CDep(inst)}{\forall t \in Tinst. \tau(t) \wedge \forall w \in Write(Tinst). \tau(w)}$$

Taint Propagation Rules:

Based on the discussions in the preceding paragraphs, the taint propagation rules are shown. In the taint propagation rules, we assume that $inst$ captures the current instruction for which the taint propagation is being computed. $Write(Tinst)$ captures the set of operands written by a set of instructions $Tinst$. In the following, we discuss a few salient features of our taint propagation rules:

Computation: For instructions involving computations (e.g., addition, subtraction and multiplication), we consider them tainted if and only if they operate on tainted operands. Moreover, the result computed by such tainted instructions are also marked tainted. This is because the outcome of these instructions can be controlled by an attacker if they operate on tainted operands. As *oo7* operates at the binary code level, most computations can be captured via either binary or unary operations, as shown in our taint propagation rules.

Memory Load: Accounting taint propagation through memory load instruction is crucial for the effectiveness of *oo7*. To this end, we need to consider two different scenarios: (i) taint propagation through the value being loaded, and (ii) taint propagation through the accessed memory address for the load. In both cases, the loaded register can be controlled by an attacker, as either the accessed address or the value located therein can be manipulated to load an attacker controlled value. To model these taint propagation rules, we need to accurately track the tainted status of the accessed memory regions and this, in turn, involves a conservative analysis of possible aliases in the program.

Memory Store: We note that a tainted value can be stored in an arbitrary memory address y . For such an operation, we conclude that the value located in address y can be tainted, as captured by the predicate $\tau(val(y))$. Such a store operation, however, does not conclude anything about $\tau(y)$, as the store operation “ $y = \text{store}(x)$ ” cannot control the address y via the tainted operand x .

Conditional Branch: Conditional branches are involved in accounting for both the explicit and implicit (i.e., taints

through control dependencies) propagation of taints. For example, the outcome of a conditional branch can be controlled by an attacker (i.e., tainted) if the variable x in $\text{branch}(L, x)$ is tainted. Moreover, we discover the set of instructions $Tinst$ that are control dependent on a branch instruction. If the branch is tainted, then the value of any variable written by instructions in $Tinst$ is indirectly (i.e., implicitly) controlled by the attacker. Consequently, we mark all such values tainted, as captured by our taint propagation rules.

For the sake of brevity, the aforementioned taint propagation rules are described based on a simplified syntax of low-level binary code. The taint propagation rules avoid any under-tainting. However, these taint propagation rules may lead to over-tainting as described in the following.

Sources of Over-tainting: Over-tainting in *oo7* may lead to some instruction $inst$ or data element (i.e., a value or an address) x to be tainted (i.e., $\tau(inst)$ or $\tau(x)$ holds, respectively) even in the absence of any feasible execution where $\tau(inst)$ or $\tau(x)$ is true. Such a phenomenon may occur due to the following scenarios:

- 1) **Conservative extraction of CFG:** It is often challenging to extract an accurate control flow graph from the binary code. This is primarily due to the difficulty in precisely identifying the branch targets, such as the targets of indirect branches and calls. This, in turn, leads to conservative approximation of control flow edges in the extracted CFG. As our taint propagation rules walk the CFG and leverage control dependency graph (CDG) to compute implicit taint propagation, the method might lead to over-tainting due to the additional edges in the CFG and CDG.
- 2) **Conservative alias analysis:** As our taint propagation rules involve taint tracking through both the memory addresses and the values in these memory addresses, *oo7* may over-taint due to a conservative alias analysis. For example, two memory addresses x and y might be considered aliases even if x and y do not point to the same memory address in any feasible execution. Nonetheless, our taint propagation rules will conservatively assume that all values in memory locations pointed by y are tainted, given that the values in memory locations pointed by x can be attacker controlled.
- 3) **Approximate memory model:** An accurate taint tracking requires the tainted status of each possible memory address that can be accessed by a program. This might often be expensive and the static analysis may resort to conservative approximation for scalability. For instance, even if a particular data element in an aggregate (e.g., a structure) is tainted, to speed up the analysis, *oo7* might conservatively assume values in all the memory addresses occupied by the aggregate data structure to be tainted.
- 4) **Non-analyzable function calls:** The return values from non-analyzable function calls are conservatively considered to be tainted. These function calls might be libraries or other third-party software that cannot be analyzed.

Properties: In summary, *oo7* guarantees the following crucial property via its taint propagation rules:

Property 1. For a given program P , consider an arbitrary register or memory location x_P that can be controlled by an attacker. Moreover, let the outcome of instruction $inst_P$ may also be controlled by an attacker. The taint propagation rules of *oo7* guarantee that $\tau(x_P)$ and $\tau(inst_P)$ hold.

We consider all inputs arriving from external sources (e.g., network, files, command lines) are tainted initially. These are called taint sources. Our taint propagation rules then avoid any under-tainting, by considering the forward transitive closure of all control and data dependencies from the taint sources.

5.3 Detecting other Spectre variants

In the preceding section, we discussed the detection of Spectre variant 1 [2]. Note that Spectre variant 1 can leak the secret data in other ways instead of performing the LS action. Such variants can be detected via simple manipulation of $\Phi_{spectre}$:

$$\Phi_{spectre}^{weak} \equiv br(TB) \wedge load(RS) \wedge \tau(TB) \wedge \tau(addr(RS)) \wedge (\Delta(TB, RS) \leq SEW) \quad (2)$$

Our *oo7* approach can be fine tuned to detect other Spectre variants. For instance, consider Spectre Variant 1.1 (cf. TABLE 1). Such a variant can easily be detected by the following condition:

$$\Phi_{spectre}^{v1.1} \equiv br(TB) \wedge store(SW) \wedge \tau(TB) \wedge \tau(addr(SW)) \wedge (\Delta(TB, SW) \leq SEW) \quad (3)$$

where $store(SW)$ captures the presence of a speculative write instruction, as needed to exploit Spectre Variant 1.1.

Spectre Variant 1.2 (read-only protection bypass) needs exactly the same condition $\Phi_{spectre}^{v1.1}$ to be satisfied, except that the speculative write (SW) happens to be in read-only memory. For the rest of the paper, we do not distinguish between Spectre Variant 1.1 and 1.2, as *oo7* uses the same condition to detect both the variants.

To detect Spectre Variant 4, we need to check whether a load instruction (RS) follows a store (WR) to the same address, yet RS can speculatively load a value not yet written by WR . Checking for this condition requires accurate address analysis, more accurate than what we can currently support. We are currently working in this direction.

5.4 Code repair

Our repair strategy is based on systematically inserting memory fences after each tainted branch (i.e., TB) in vulnerable code fragments for Spectre variants 1, 1.1 and 1.2. The original article describing Spectre attacks [2] suggests insertion of memory fences following each conditional branch. However, using our analysis, we can obtain the exact sequence $\langle TB, RS, LS \rangle$ (for Variant 1) or the sequence $\langle TB, SW \rangle$ (for Variant 1.1 and Variant 1.2) vulnerable to Spectre attacks. As a result, we can accurately locate the program point where the memory fence should be inserted. In particular, we insert memory fences following TB instruction and immediately before the execution of RS and SW , respectively, for Spectre Variant 1 and Variant 1.1, 1.2. This prevents execution from loading the secret value

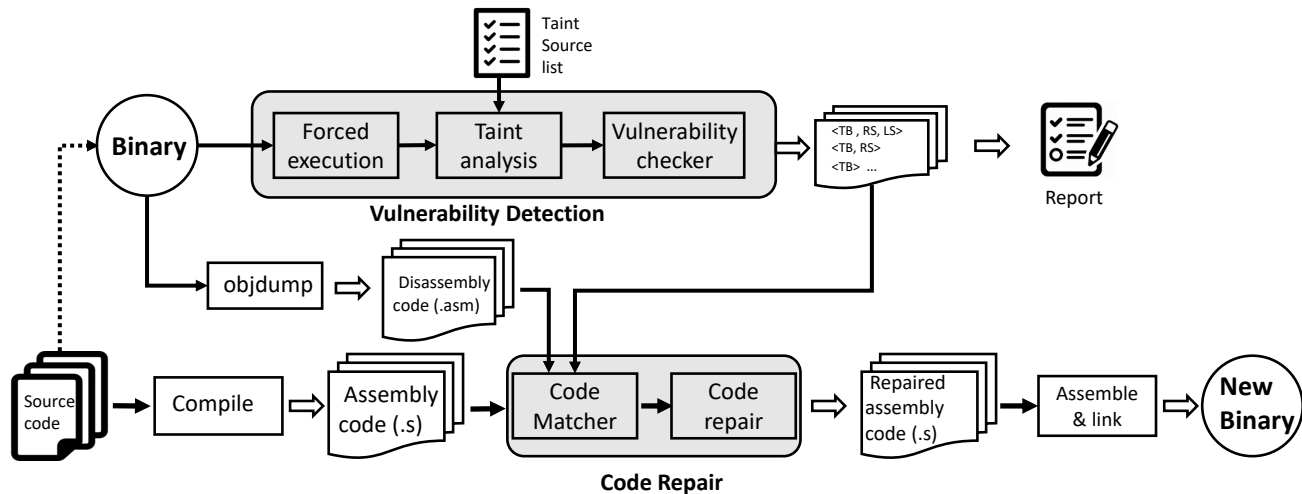


Fig. 1: Overview of *oo7* framework. The components in grey represent the core modules of *oo7*: vulnerability detection module and code repair module.

into the cache (for Variant 1) and writing to an attacker-controlled location (for Variant 1.1, 1.2) speculatively.

Nevertheless, inserting memory fences may affect the overall program performance. *oo7* inserts memory fences only for the branches identified as *TB* (for variants 1, 1.1 and 1.2). This has less overhead than inserting fences after each conditional branch, or after each tainted conditional branch. We show empirically that such a strategy has acceptable performance overheads of average 5.9% for SPECint benchmarks. When selecting the fence instruction, we observed in experiments that `cpuid` [26], `lfence` and `mfence` are able to prevent Spectre attacks (while `sfence` cannot). However, if `cpuid` is used for repairing Spectre vulnerability in the assembly code, then it modifies the general-purpose registers. Specifically, the return value of `cpuid` is stored in registers such as `EAX`, `EBX`, `ECX` or `EDX`. Thus, additional instructions are required to store and restore the impacted register(s) before and after invoking `cpuid`, which introduces additional performance overhead. On the other hand, `lfence` is officially confirmed and recommended by Intel [8] to mitigate Spectre attack. Therefore, in this work, we use only the `lfence` instruction to repair vulnerable code.

6 IMPLEMENTATION

Fig. 1 provides an overview of *oo7* tool. *oo7* contains two main modules: a vulnerability detection module for detecting the Spectre vulnerabilities, and a code repair module to fix the Spectre vulnerabilities.

Vulnerability detection module. The vulnerability detection module of *oo7* is supported by three major technologies: forced execution, taint analysis and vulnerability checker.

Forced execution [23], [27], as its name suggests, forces the program to execute along all possible paths by predicting the branch outcomes to both true and false. The capability of forced execution is in exploring the different execution paths and simulating the execution to expose the behavior of a given program. Forced execution satisfies

the semantics of speculative execution, because speculative execution may lead the processor to execute the instructions on both outcomes of a branch when the branch prediction is wrong. Forced execution engine explores all possible paths by maintaining a pool of execution paths that may be explored in future by switching more predicates. A predicate is represented as a tuple (I_{src}, I_{dst}) where I_{src} and I_{dst} denote the source instruction and forced execution target (the branch outcomes), respectively. New predicates are added when a branch instruction is evaluated. As forced execution is a well-known technique, for details of forced execution technique, we refer the readers to relevant previous works [23], [27], [28], [29].

Taint analysis performed by the taint propagation engine tracks the data and instruction that can be controlled by the attacker. The taint analysis works along with the forced execution engine. When forced execution engine evaluates a `call` instruction, it checks if the destination of the `call` is in the taint source list. The taint source list is a set of APIs which can import the data to the program from the un-trusted channels such as network, user input, file reader. In the implementation, we consider all interfaces (e.g. `fgetc()`, `recv()`) in the commonly used libraries (e.g. `glibc`) which import data from outside of program as the taint sources. If the destination of a `call` instruction executed by the forced execution engine is in the taint source list, the taint propagation engine marks the imported data as a tainted object. After the tainted object is imported from the taint source, the taint engine propagates this tainted object during the execution by applying the taint propagation rules explained in section 5.2.

Vulnerability checker detects whether the current state satisfies the conditions of arbitrary vulnerabilities presented in section 5.3. The vulnerability checker works after the taint engine taints an instruction. When the taint engine taints a new conditional branch instruction, the vulnerability checker records this tainted branch and sets up a Speculative Execution Windows (SEW) for it. SEW is decremented by one at the end of the evaluation of each instruction along the execution path. The vulnerability checker reviews whether

a memory instruction satisfies the condition for *RS* or *LS* before the *SEW* is decremented to zero. The vulnerability checker records down *TB*, *RS* and *LS* as a potentially vulnerable code fragment for the final report.

We adopt BAP [22] as our primary analysis platform. BAP provides a toolkit for implementing automated binary analysis and it supports multiple architectures such as x86, x86-64, ARM, PowerPC, and MIPS. BAP lifts binary code into RISC-like intermediate representation (IR) named BAP Instruction Language (BIL). Program analysis is performed using the BIL representation and it is architecture independent. BAP contains a microexecution framework named Primus performs forced execution and taint analysis. BAP provides several interfaces to export crucial information to other analysis modules during the analysis. We implement the vulnerability checker bases on these interfaces.

Vulnerability repair module. Once a vulnerable code fragment is detected in the binary, we locate the corresponding assembly code for repair. To this end, we first mark the address(es) of Spectre vulnerable code, as obtained during the detection stage of *oo7*. Concurrently, we obtain the disassembled code from the binary and the assembly code from the source (via “-s” option in gcc compiler). As most optimizations are employed during the compilation stage rather than assembler stage, there is no substantial difference between the assembly code and the respective disassembled code. This allows us to easily map the disassembled code back to the assembly code and locate the instructions vulnerable to Spectre attack.

Finally, our repair module directly modifies the assembly code by inserting memory fence instructions in the appropriate location (e.g., inserting `lfence` before *RS* or *SW* for mitigating Spectre Variants).

7 EVALUATION SETUP

In this section, we present the details of our evaluation setup. We first describe the programs used in our experiment. Then we introduce the platform used for the evaluation.

7.1 Subject Programs

We conduct evaluation on three sets of subject programs.

- We first apply *oo7* on 15 code examples purpose-built to demonstrate different variations of Spectre vulnerabilities from Paul Kocher’s blog post [1]. We call these *Litmus Tests*.
- Next, we conduct evaluation on SPECint benchmarks, which have been well-studied by the computer architecture community. These are detailed in TABLE 4. We concentrate on complete analysis of the SPECint (integer) benchmark suite because it includes more control-intensive code compared to SPECfp (floating point) and Spectre exploits vulnerability through conditional branches. SPECint benchmark suite contains 18.31% branches in the instruction mix compared to only 5.75% for SPECfp [30].
- Last but not the least, we conduct evaluation with a large number of software projects from Google OSS-Fuzz repository [31] and GitHub. The program binaries in these project include the main application and

miscellaneous support tools. TABLE 6 summarizes the characteristics of these projects consisting of a total of 507 program binaries with size ranging from 8.5KB to 21.8MB (average size 261.4KB).

7.2 Evaluation Platform

We conduct experimental evaluation on Intel Xeon Gold 6126 [32] running at 2.6GHz with 192GB memory. The underlying micro-architecture is Skylake with 224-entry re-order buffer (ROB) [33]. Due to the potential expansion of instructions into micro-operations and subsequent fusion in x86 micro-architectures (Section 5), we conservatively set the speculative window to twice the effective ROB size, i.e., $SEW = 448$. Intel Xeon Gold 6126 is equipped with 12 cores and 19.25MB non-inclusive shared last-level cache (LLC) with 64 byte line size. The LLC cache miss penalty is about 200 cycles. Non-inclusive LLC is more secure than the inclusive cache and can thwart certain LLC based side-channel attacks (e.g., `Flush+Flush`, `Prime+Probe`). However, it is still vulnerable to `Flush+Reload` attack. Thus the Spectre and Meltdown attacks can potentially be carried out in this platform.

8 EVALUATION RESULTS

Our evaluation investigates three different aspects:

- 1) *Effectiveness*: How effective is *oo7* in detecting Spectre vulnerabilities in program binaries?
- 2) *Analysis Time*: How long is *oo7* analysis time to detect Spectre vulnerabilities?
- 3) *Performance Overhead*: How much is the performance overhead introduced by *oo7* to protect vulnerable code fragments?

8.1 Evaluation on Litmus Tests

oo7 can correctly identify all code snippets purpose-built with different variations of Spectre vulnerabilities [1] as potential victim code fragments. 14 code examples are identified with taint propagation only along data dependencies. The remaining code example is detected with taint propagation along program (both control and data) dependencies.

The latest Microsoft Visual C++ compiler [6] has integrated `/Qspectre` switch for mitigating a limited set of potentially vulnerable code patterns related to Spectre vulnerabilities. Specifically, after compiling an application with `/Qspectre` enabled, Visual C++ compiler attempts to insert an `lfence` instruction upon detecting Spectre code patterns. Paul Kocher [1] has evaluated the Microsoft compiler using 15 litmus tests. The blog post [1] mentions that only two of the micro-benchmarks are identified and protected by Visual C++ compiler. In contrast, *oo7* can correctly detect all 15 code examples as potential victims.

The example (v13 [1]) that requires taint propagation along both control and data dependencies is given below.

```
__inline int is_x_safe(size_t x) {
    if (x < array1_size) return 1; else return 0;
}
void victim_function_v13(size_t x) {
    if (is_x_safe(x)) temp &= array2[array1[x]*512];
}
```

TABLE 3: bzip2 randomly inserted with σ vulnerable functions. VF_source: number of functions inserted in the source code; VF_oo7: number of vulnerable functions identified by oo7. Analysis time: Minutes spent by oo7 to complete analysis.

| σ | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% |
|-------------------------|-------|-------|-------|------|-------|------|-------|------|
| VF_source | 12 | 26 | 45 | 70 | 105 | 158 | 245 | 420 |
| VF_oo7 | 12 | 26 | 45 | 70 | 105 | 158 | 245 | 420 |
| Analysis time (Minutes) | 25.18 | 26.15 | 28.43 | 30.5 | 32.03 | 32.6 | 33.53 | 43.6 |

The branch in the victim function `victim_function_v13` is tainted as the return value of `is_x_safe(x)` is controlled via untrusted input `x`. However, the return value of `is_x_safe(x)` is control-dependent and not data-dependent on `x`. Thus `oo7` can detect this code pattern as potential vulnerability only if both data- and control-dependent taint propagation are applied.

8.2 Validation of Patching

We design an attacker process to steal secrets via cache side-channel from the victim process (litmus test example [1]) once the secret data is brought into the cache through Spectre attack. We successfully extracted data from arbitrary memory locations in the victim process on our platform for 10 out of 15 litmus tests. We then allow `oo7` to automatically insert `lfence` instructions at appropriate program locations to prevent speculation in vulnerable code fragments. We verify that the attacker process can no longer extract data from the victim processes running with the `oo7` fix for all 10 litmus tests even though we had successfully extracted secret data from all of them before patching.

For further evaluating the effectiveness of `oo7`, we design an experiment to check whether `oo7` can detect all potentially vulnerable code. We select the program `bzip2` from SPECint CPU benchmark suite [34], and insert several vulnerable functions to the source code of `bzip2`. The vulnerable functions are randomly chosen from the Spectre v1 variants suggested by Kocher [1]. Assume that `bzip2` has P functions, we use σ to represent the ratio of inserted vulnerable functions, where $\sigma = N/(N + P)$ and N is the number of inserted vulnerable functions. We use different values for σ to evaluate the effectiveness of `oo7`. The invocation to each vulnerable function is inserted in random locations of the `bzip2` source code. Besides, each vulnerable function contains a taint source to guarantee the vulnerable code can be controlled by the attacker.

TABLE 3 presents the number of functions inserted in the source code of `bzip2` (VF_source), and the number of vulnerable functions identified by `oo7` (VF_oo7) in the modified program with σ ranging from 10% to 80%. Program `bzip2` contains $P = 105$ functions, so we insert $\lceil \sigma \times 105 / (1 - \sigma) \rceil$ vulnerable functions. For example, we insert 12 randomly picked vulnerable functions in the `bzip2` code when $\sigma = 10\%$. As observed in TABLE 3, our analysis can identify all the vulnerable code fragments over the varying range of σ (i.e. [10%, 80%]). Moreover, as shown in TABLE 3, the analysis time increases from 25.18 minutes with $\sigma = 10\%$ to 43.6 minutes with $\sigma = 80\%$.

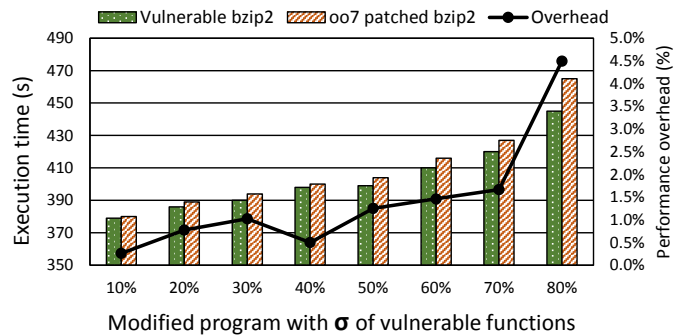


Fig. 2: Execution time and performance overhead comparison between vulnerable `bzip2` with ratio σ vulnerable functions inserted and the patched program repaired by `oo7`.

Fig. 2 show the execution time of the modified vulnerable `bzip2` and the patched version by `oo7` along with the performance overhead introduced by patching. The absolute execution time in fig. 2 are shown by bars, and the performance overhead is shown by the line with markers. As shown in fig. 2, the execution time is higher when σ is increased, for example, the test with $\sigma = 10\%$ can finish in 379 seconds, but the test with $\sigma = 80\%$ finished in 445 seconds. We use fence instructions to repair the program. We note that fence instructions introduce extra runtime overhead, as they prevent speculative execution. We can see from fig. 2 that the repaired program with $\sigma = 10\%$ takes only one second more than the vulnerable program, but the repaired program with $\sigma = 80\%$ takes 20 seconds more than the vulnerable program. Moreover, the repaired program has negligible performance overhead. Specifically, the minimum overhead is 0.26% (when $\sigma = 10\%$) and the maximum overhead is about 4.49% (when $\sigma = 80\%$).

8.3 Evaluation on Specint Benchmarks

We use SPECint CPU benchmark suite [34] to quantify the performance overhead of `oo7` protection mechanism as well as for evaluating the efficacy of our detection and repair.

SPECint benchmark suite contains twelve programs in C and C++. TABLE 4 outlines the salient features of these program: the binary size, analysis and repair time, the number of conditional branches, the number of tainted branches TB , the number of $\langle TB, RS \rangle$ pairs as well as the number of $\langle TB, SW \rangle$ pairs. We run `oo7` on the programs in SPECint benchmark suite for at most 150 hours; only `gcc` and `xalancbmk` do not complete in 150 hours. This is because `gcc` and `xalancbmk` have higher number of conditional branches.

In this set of experiments, we treat any APIs from the standard library that can import data to the program as potential taint sources. TABLE 5 lists the taint sources identified by `oo7` in SPECint benchmark suite. Most of the programs in SPECint read the input from one or more files; so the file reading functions from standard C library are taint sources in these programs. For example, in `gcc` benchmark, `fread()`, `read()`, `getcwd()`, `getenv()` and `_IO_getc()` from `glibc` are used in the code and are marked as taint sources.

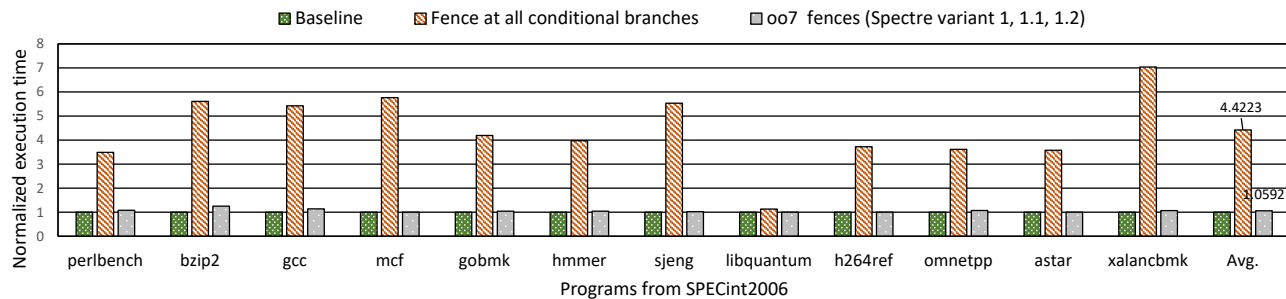


Fig. 3: Runtime overhead due to protection against Spectre Variant 1, 1.1, 1.2 compared to original code in SPECint. *All* denotes the overhead from inserting fence at all conditional branches. The overheads from fences introduced by *oo7* to protect against Spectre variants 1, 1.1, 1.2 are plotted.

TABLE 4: Results for the detection of Spectre vulnerable code fragments in SPECint. $\langle TB \rangle$ denotes the tainted conditional branches detected by *oo7*. $\langle TB, RS \rangle$ satisfies Spectre Variant 1 condition, while $\langle TB, SW \rangle$ satisfies Spectre Variant 1.1, 1.2 conditions as detected by *oo7*.

| Program | Binary Size | LoC | Analysis time(h—m) | Repair time (s) | #Conditional branches | # $\langle TB \rangle$ | # $\langle TB, RS \rangle$ | # $\langle TB, SW \rangle$ |
|------------|-------------|---------|--------------------|-----------------|-----------------------|------------------------|----------------------------|----------------------------|
| perlbench | 1.2MB | 97,314 | 125h | 5 | 21,972 | 60 | 18 | 5 |
| bzip2 | 69 KB | 5,115 | 27.4m | 1 | 942 | 102 | 81 | 5 |
| gcc | 3.6MB | 365,844 | > 150h | 11 | 59,614 | 103 | 8 | 2 |
| mcf | 23 KB | 1,370 | 3.5m | 4 | 202 | 42 | 0 | 0 |
| gobmk | 3.9 MB | 154,170 | 19.2h | 1 | 11,549 | 57 | 13 | 0 |
| hmmer | 319KB | 19,267 | 3.67h | 1 | 4,468 | 49 | 15 | 0 |
| sjeng | 153 KB | 10,147 | 2.6h | 1 | 2,146 | 18 | 3 | 0 |
| libquantum | 51KB | 2,212 | 1.35h | 1 | 444 | 30 | 0 | 0 |
| h264ref | 577 KB | 32,623 | 25.3h | 1 | 6,743 | 16 | 0 | 0 |
| omnetpp | 768KB | 22,603 | 21h | 2 | 4,812 | 90 | 26 | 3 |
| astar | 52 KB | 3,003 | 21.2m | 1 | 541 | 19 | 0 | 0 |
| xalancbmk | 5.8MB | 186,997 | > 150h | 15 | 62,209 | 72 | 47 | 2 |

TABLE 5: Detected taint sources for programs from SPECint

| Program | Taint source list |
|------------|--|
| perlbench | getpid(), getgid(), getuid(), geteuid(), getegid(), read(), fgetc(), getcwd(), getenv(), gettimeofday(), fread() |
| bzip2 | read() |
| gcc | fread(), read(), getcwd(), getenv(), _IO_getc() |
| mcf | fgets() |
| gobmk | fgets(), getenv() _IO_getc() |
| hmmer | fread(), __fread_chk(), fgets(), getenv() |
| sjeng | fgets(), _IO_getc() |
| libquantum | fgetc(), fread(), getenv() |
| h264ref | fread(), read(), __isoc99_fscanf(), |
| omnetpp | fgets(), getenv(), _IO_getc() |
| astar | read(), fscanf() |
| xalancbmk | getcwd(), fread(), getenv() |

We note that eight out of twelve programs from Specint benchmark exhibit the vulnerability pattern of Spectre variant 1 as evidenced by the presence of $\langle TB, RS \rangle$ pattern in *perlbench*, *bzip2*, *gcc*, *gobmk*, *hmmer*, *sjeng*, *omnetpp*, *xalancbmk*. By looking for the $\langle TB, RS \rangle$ pattern, we conservatively assume the strictest security requirement of reading secret data. The subsequent mechanism to leak the secret data can vary with the most common mechanism being the cache side-channel with $\langle TB, RS, LS \rangle$ code pattern. Five of the benchmarks (*perlbench*, *bzip2*, *gcd*, *omnetpp*,

xalancbmk) are vulnerable to Spectre variant 1.1, 1.2 as evidenced by the presence of $\langle TB, SW \rangle$ pattern. The analysis time varies from 3.5 minutes (*mcf*) to 150 hours (*gcc*). The analysis time not only depends on the binary size but also the complexity of the program logic, more specifically, the number of branches, for example *gcc* and *xalancbmk* contain 59,614 and 62,209 branches in there binary. Since *gcc* and *xalancbmk* contain more branches than other programs, there are more execution paths to explore. Therefore our analysis times out after 150 hours. Our repair works on the assembly code and can complete in 15 seconds for these benchmarks.

We evaluate the runtime overhead due to fence insertion by executing each modified program ten times and report the average values. Fig. 3 shows the normalized execution time. The average performance overhead is 430% when fences are inserted naively at both paths of all conditional branches. This is the safest strategy in the absence of an accurate program analyzer such as *oo7*. In contrast, *oo7* only inserts fences at detected conditional branches covering Spectre Variants 1, 1.1, 1.2 and hence incurs only 5.9% overhead on an average.

8.4 Evaluation on interactive program

lighttpd is a lightweight web server which allows programs to interact with it by sending HTTP requests. We run *oo7* on *lighttpd* to evaluate its effectiveness on interactive

```

unsigned char*
buffer_append_base64_decode(..., char*in, ...) {
    ...
    for (i = 0; i < in_length; i++) {
        unsigned char c = (unsigned char) in[i];
        int ch;
        ...
        if (c >= 128) return NULL;          /* TB */
        ch = base64_reverse_table[c];      /* RS */
        if (-3 == ch) {
            if (group < 2) return NULL;
            break;
        } else if (-2 == ch) {
            continue;
        } else if (ch < 0) {
            return NULL;
        }
        ...
    }
}
    
```

Fig. 4: Potential Spectre vulnerability in base.c within project *lighttpd*. *in* is tainted from a taint source *recv()*. *TB*, *RS* are highlighted.

programs. We identified nine taint sources in *lighttpd*, detected 129 tainted branches and 40 Read Secret (*RS*). In the identified taint sources, *recv()* is a critical interface that receives data from arbitrary clients through the network, which poses potential threats for data leakage in the server caused by deliberately constructed data from the attacker. Moreover, we evaluate the performance of repaired *lighttpd* by sending 10000 HTTP requests to the *lighttpd* server; the experiment shows our repaired code only slows down *lighttpd* by 5.6%.

Fig. 4 shows a typical Spectre variant 1 vulnerability detected in function `buffer_append_base64_decode()` within file `lighttpl.4/src/base64.c` of *lighttpd* project. `buffer_append_base64_decode()` is called to decode the `base64` string after `recv()` receives data from the network. The code extracts a sequence of characters to variable `c` from the input buffer `in` and checks whether `c` is less than 128 (*TB*). If `c < 128`, then the value of `c` is used to index the table `base64_reverse_table`. The attacker can first train the branch “`if (c >= 128)`” by using values of `c` to be less than 128. Then the attacker can pass a value of `c` to be 128. This results in accessing bytes outside of the array `base64_reverse_table[]` being read to variable `ch`. This happens when the branch “`if (c >= 128)`” is mispredicted for `c = 128` and during the speculative execution of the array access `base64_reverse_table[]`. The attacker can infer whether the leaked data is equal to -3 by probing the cache line impacted by the following `if` statement. Moreover, the type of `c` is unsigned char with a value range 0 to 255. Therefore this vulnerable code can at most leak 128 bytes outside of array `base64_reverse_table[]`.

8.5 Evaluation on various software projects

We observe that the detection of Spectre Variant 1 (as opposed to variants 1.1, 1.2) takes the longest time from our experiments with SPECint benchmarks. So we evaluate the scalability of Spectre variant 1 detection on real-world programs. In the evaluation, we select 507 binaries from OSS-fuzz repository of Google and other open-source

```

static int command_lcp(..., char **argv, ...)
{
    keylen2 = atoi(argv[3]);
    if ((keylen2 < 0) || /* TB */
        (keylen2 > (int) BITSOF(strlen(argv[2])))) {
        ...
        return -1;
    }
    ...
    lcp = fr_trie_path_lcp(..., keylen2, ...);
}

static int fr_trie_path_lcp(..., int keylen2, ...)
{
    ...
    if (end_bit > keylen2) end_bit = keylen2;
    ...
    e2 = end_bit;
    ...
    xor <<= s2;
    xor |= lcp_end_bit[e2 - s2]; /* RS */
    lcp = xor2lcp[xor]; /* LS */
    ...
}
    
```

Fig. 5: Potential Spectre vulnerability in *trie.c* within project *freeradius*. *argv* is tainted from a taint source *gets()*. The triplet $\langle TB, RS, LS \rangle$ is highlighted.

projects; most of the selected binaries receive input data from the outside world through network interface. We apply *oo7* on the program binaries from selected projects as shown in TABLE 6. The analysis are performed for at most 24 hours to detect potential Spectre variant 1 code snippets to demonstrate the scalability of our analysis.

The column “# of vulnerable binaries” in TABLE 6 shows the number of program binaries in each project with potential vulnerabilities. We identify a program as vulnerable if it has at least one $\langle TB, RS \rangle$ pattern in the code that can potentially be exploited by the attacker to read secret data, i.e., we conservatively assume the strictest security requirement. As mentioned earlier, the subsequent mechanism to leak the secret data can vary with the most common mechanism being the cache side-channel with $\langle TB, RS, LS \rangle$ code pattern. For each project, we report the number of vulnerable programs, the average number of $\langle TB \rangle$ and the average number of $\langle TB, RS \rangle$ patterns under two different taint propagation strategies: data dependencies and program (data & control) dependencies. For example, in project *samba*, out of total 230 programs, *oo7* detects 62 and 91 binaries as potential victims under data- and program-dependence taint propagations, respectively. Program-dependence based taint propagation identifies additional vulnerable code fragments compared to data-dependence only. Altogether, 233 (or 290) out of 507 programs are labeled as potential victims by *oo7* under data- (or program-) dependence taint propagation. TABLE 6 also show the analysis time in hours for detecting $\langle TB, RS \rangle$ patterns using data dependencies, and program dependencies.

Potential Spectre Vulnerability in Large-scale Code:

We show one example of a Spectre vulnerability unearthed by *oo7* in Fig. 5. This code snippet is identified by *oo7* in a program (`src/lib/util/trie.c`) within the project *freeradius*. Note that *oo7* identifies the vulnerability at binary level. For the sake of exposition and brevity, we show only the portions corresponding to the Spectre vulnerability pattern at source code level. As comments, we highlight the code fragments detected as *TB*, *RS* and *LS*. The argument *argv* is a tainted array read from an

TABLE 6: Software projects used from Github and OSS-Fuzz and the detected Spectre v1 Vulnerabilities.

| Project | Project Description | # of binaries | Avg. binary size | Data-dependency | | | | Program-dependency | | | |
|--------------|--|---------------|------------------|--------------------------|---|---|-----------------------|--------------------------|---|---|-----------------------|
| | | | | # of vulnerable binaries | Avg. # of $\langle TB \rangle$ per binary | Avg. # of $\langle TB, RS \rangle$ per binary | Avg. analysis time(h) | # of vulnerable binaries | Avg. # of $\langle TB \rangle$ per binary | Avg. # of $\langle TB, RS \rangle$ per binary | Avg. analysis time(h) |
| samba | SMB/CIFS networking protocol | 230 | 124.0KB | 62 | 16 | 9 | 0.76 | 91 | 32 | 46 | 1.2 |
| coreutils | GNU OS file, shell and text manipulation utilities | 114 | 125.3KB | 78 | 14 | 4 | 0.22 | 84 | 83 | 57 | 1.3 |
| cups | Common UNIX Printing System | 52 | 134.3KB | 30 | 40 | 31 | 1.07 | 46 | 141 | 122 | 3.09 |
| freeradius | Popular open-source RADIUS server | 47 | 49.9KB | 18 | 13 | 24 | 0.21 | 25 | 82 | 25 | 0.45 |
| openldap | Lightweight Directory Access Protocol | 31 | 1.3MB | 28 | 291 | 98 | 8.05 | 25 | 580 | 484 | 20.59 |
| openssh | Network utilities based on SSH protocol | 11 | 791.9KB | 4 | 21 | 4 | 21.13 | 6 | 72 | 15 | > 24 |
| xrdp | Remote desktop protocol (rdp) server | 10 | 107.3KB | 5 | 23 | 2 | 0.59 | 5 | 48 | 22 | 3.04 |
| ppp | PPP daemon and associated utilities | 4 | 322.0KB | 2 | 56 | 41 | 5.11 | 2 | 77 | 60 | 6.16 |
| dropbear | Small SSH server and client | 4 | 1.2MB | 2 | 148 | 20 | > 24 | 2 | 172 | 44 | > 24 |
| netdata | Distributed real-time performance monitoring | 2 | 1.9MB | 2 | 109 | 45 | 12.05 | 2 | 198 | 175 | 12.41 |
| wget | Content retrieval from web servers | 1 | 937.2KB | 1 | 134 | 25 | 16.1 | 1 | 542 | 430 | > 24 |
| darknet | Convolutional Neural Networks | 1 | 663.9KB | 1 | 76 | 42 | 3.17 | 1 | 183 | 195 | 6.8 |
| Total | - | 507 | - | 233 | - | - | - | 290 | - | - | - |

external file through the taint source `gets()`. The conditional check in the function `command_lcp` is therefore a tainted branch (TB). Taint is propagated to the function `fr_trie_path_lcp` via the parameter `keylen2`. Consequently, the array load `lcp_end_bit` may use the value of `e2` (potentially controlled by the attacker through `argv[3]`) during speculative execution. This speculative execution may take place due to the misprediction of the conditional branch in `command_lcp` that reflects a bound check. Finally, the array access `xor2lcp` may reveal information out of the boundary of array `lcp_end_bit[]` via cache side channel. Though the pattern $\langle TB, RS, LS \rangle$ is found in the wild by `oo7`, the vulnerable code fragment is executed only once at run-time, making it impossible for the adversary to poison the branch and launch an attack. The distance between TB and RS is 145 Intel x86 instructions. The example illustrates that Spectre vulnerability in real-world may span over multiple functions requiring inter-procedural program analysis.

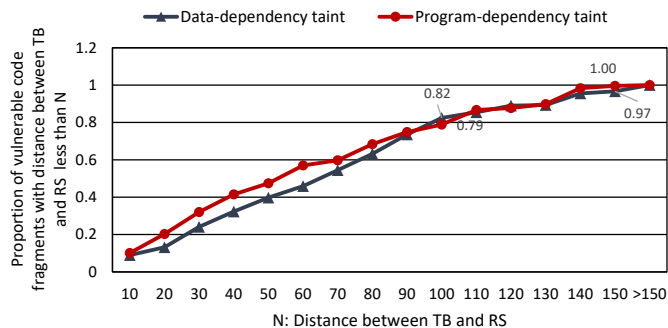


Fig. 6: Cumulative distribution of distance (#instructions) between TB and RS for binaries in TABLE 6 under data- and program dependence based taint propagation.

Sensitivity to Speculative Execution Window size:

We set the speculative execution window size $SEW = 448$ as twice the effective ROB size in our platform. This is a conservative assumption to take care of micro-operations generated from the instructions and processed within the micro-architecture. We investigate the sensitivity of our analysis on SEW value. Fig. 6 shows the distance in instructions between TB and RS ($\Delta(TB, RS)$) for vulnerable code fragments across all 507 binaries. The results show that 82% and 79% of the tainted memory accesses (RS) occur within 100 instructions from the tainted branch (TB) for data- and program-dependence based taint propagation, respectively.

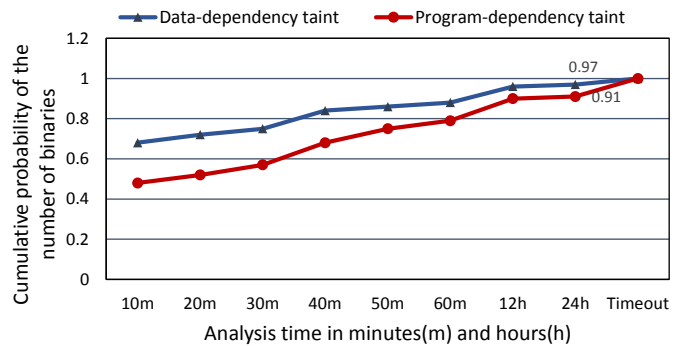


Fig. 7: Cumulative distribution of analysis time for binaries in TABLE 6 under data- and program dependence based taint propagation.

Analysis and Repair Time of `oo7`: The analysis time depends on the size and complexity of the binary. Fig. 7 shows the distribution of analysis time across all the binaries. Under data-dependence based taint propagation, the analysis time is less than 20 minutes for 72% of the binaries. Program-dependence based taint propagation increases the analysis time; still the analysis completes within 20 minutes for 52% of the binaries. Only 3% and 9% of the analysis across 507 binaries did not complete in 24 hours. The repair time is minimal; for all of the 507 binaries, it is within 30 seconds.

Quantitative Analysis of Vulnerabilities: Our analysis shows that on an average only 7.3% (variance 0.3%) of conditional branches are tainted across 290 programs with at least one tainted branch. Moreover, 217 out of 507 binaries do not have any tainted branch at all.

Next we check the percentage of conditional branches in these program binaries that are tainted (TB) and are followed by tainted memory access (RS) within speculative execution window. If we want to ensure strict security requirements, then `lfence` instruction should be inserted after all these tainted branches. On an average, our analysis shows only 3.72% (variance 0.3%) of conditional branches satisfy this criteria leading to very low overhead in fixing Spectre vulnerability.

Finally, we check the percentage of conditional branches in these program binaries that are tainted (TB) and are followed by tainted memory access (RS) within speculative execution window and a subsequent tainted memory access (LS) to leak the data to the cache. This is denoted

as $TB+RS+LS$. If we assume cache side-channel attack as the only mechanism to leak the secret brought into the cache, then *oo7* only needs to add `lfence` instruction after these branches. On an average, only 2.32% of conditional branches (variance 0.1%) satisfy this criteria. This strongly indicates that the performance overhead from inserting fences suggested by our technique will be low. However we could not collect the exact performance overhead for all of these 507 binaries because it will involve running each binary against many inputs and averaging the performance overhead across inputs. Furthermore, for some of the binaries such as `coreutils`, a large set of inputs (each input being a command) is possible. For this reason, we measured performance overheads on SPECint benchmarks instead, which is a standard benchmark suite with inputs specified for performance analysis.

Higher vulnerable code fragment percentage in program-dependency enabled analysis. : As shown in TABLE 6, *oo7* detects more vulnerable code fragments with program-dependency analysis (that considers both data dependency and control dependency) compared to only data-dependency analysis. For example, in `samba` project, *oo7* detects on an average 16 tainted branches per binary when considering only data-dependency. In contrast, with program-dependency analysis, the number of tainted branches detected increases to 32 branches on average per binary. The reason is that if a conditional branch is tainted, all the instructions that have control-dependency with the conditional branch are also marked as tainted. Thus, analysis with program-dependency enabled inevitably causes over-tainting.

9 LIMITATIONS OF OUR APPROACH

oo7 relies on BAP, which, in turn incorporates a taint analysis engine. The taint analysis statically interprets the code by unrolling the loops up to a certain depth. In order to ensure that our approach does not introduce false negatives we need to pay attention to the following three issues.

- First of all, optimistic loop unrolling may introduce false negatives (missing vulnerabilities) in *oo7*. However, with correct or worst-case loop bounds being supplied to BAP, such a limitation can be mitigated.
- Secondly, taint sources are provided to the taint analysis engine, and if taint sources are under-specified then the taint analysis may not identify all the branches that can be controlled by the attacker. We thus conservatively assume all user inputs via console, file, and network as taint sources.
- Finally, the completeness of the control-flow extraction also plays a role to decide whether our analysis will introduce false negatives. If the branch targets of register indirect jumps are not identified, the control flow graph extracted from the binary will not be complete, and as a result the taint analysis results may miss tainted branches. Thus, our approach always depends on the control flow graph being as complete as possible, in trying to ensure that we do not have false negatives in our analysis.

oo7 finds tainted memory accesses following a tainted conditional branch within a fixed speculation window. Incorrect setting of this speculation window size may lead to false positives (window size too big) or false negatives (window size too small). We conservatively set the window length to twice the size of the of the reorder buffer, as explained earlier.

oo7 works on native code for program binaries. We have not investigated Spectre detection on interpreted code.

We also assume that all memory references in the victim code have been protected with appropriate checks to prevent overflow and underflow (e.g., array bound overflow). Thus, there does not exist any overflow/underflow error in normal (i.e., non-speculative) execution traces.

The taint analysis capability gives *oo7* the flexibility to adapt to Spectre variants. We have discussed in detail (Section 2) the class of Spectre variants that we can handle and the ones that cannot be handled. In addition, new variants are constantly being found, we could face some variants in future that *oo7* cannot be adapted to handle.

The underlying technology used by *oo7* is forced execution, which is able to handle the obfuscated or self-modifying code [23]. Thus, *oo7* can potentially analyze obfuscated or self-modifying code. Moreover, code obfuscation and self-modifying code are common evasive techniques used by malware, while *oo7* focuses on identifying and repairing vulnerabilities in the general software instead of detecting malware.

oo7 is based on program analysis. As a result it can detect (and patch) vulnerable code patterns that can be exploited by malicious program inputs. It cannot detect scenarios where an external process affects micro-architectural states (e.g., flushing dynamic branch predictor) and thereby introduce vulnerabilities in another program (e.g., by exploiting the default static branch predictor as the dynamic branch predictor has been flushed). Such scenarios cannot be detected by our analysis.

10 DISCUSSION

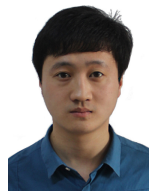
We have built *oo7* for detecting Spectre vulnerabilities in binary code and protecting against the attack with minimal overhead. Our approach is employed post-compilation on binary code to take into account all the compiler optimizations. No change to the operating system or the processor is needed as the approach proceeds by program analysis. We demonstrate that systematic analysis is useful both for detecting Spectre vulnerabilities and to repair them with minimal performance overhead. Our work also provides an understanding of the class of Spectre attacks for which an analysis based mitigation may be suitable, and for which classes of attacks a system level solution is suitable.

11 ACKNOWLEDGMENTS

This research was partially supported by a grant from the National Research Foundation, Prime Ministers Office, Singapore under its National Cybersecurity R&D Program (TSUNAMi project, No. NRF2014NCRNCR001-21) and administered by the National Cybersecurity R&D Directorate.

REFERENCES

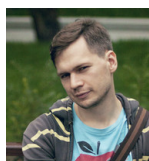
- [1] P. Kocher, "Spectre mitigations in microsoft's C/C++ compiler," <https://www.paulkocher.com/doc/MicrosoftCompilerSpectreMitigation.html>.
- [2] P. Kocher, J. Horn, A. Fogh, D. Genkin, D. Gruss, W. Haas, M. Hamburg, M. Lipp, S. Mangard, T. Prescher, M. Schwarz, and Y. Yarom, "Spectre attacks: Exploiting speculative execution," in *40th IEEE Symposium on Security and Privacy (S&P'19)*, 2019.
- [3] J. L. Hennessy and D. A. Patterson, *Computer architecture: a quantitative approach*. Elsevier, 2017.
- [4] Y. Yarom and K. Falkner, "Flush+reload: a high resolution, low noise, 13 cache side-channel attack," in *Usenix Security Symposium*, 2014.
- [5] "Meltdown and Spectre," <https://meltdownattack.com/>.
- [6] M. community, "C++ developer guidance for speculative execution side channels," <https://docs.microsoft.com/en-us/cpp/security/developer-guidance-speculative-execution>, 2018.
- [7] D. Gruss, M. Lipp, M. Schwarz, R. Fellner, C. Maurice, and S. Mangard, "Kaslr is dead: long live kaslr," in *International Symposium on Engineering Secure Software and Systems*. Springer, 2017, pp. 161–176.
- [8] Intel, "Intel analysis of speculative execution side channels," <https://newsroom.intel.com/wp-content/uploads/sites/11/2018/01/Intel-Analysis-of-Speculative-Execution-Side-Channels.pdf>.
- [9] —, "Retpoline: A branch target injection mitigation," <https://software.intel.com/security-software-guidance/api-app/sites/default/files/Retpoline-A-Branch-Target-Injection-Mitigation.pdf>.
- [10] E. M. Koruyeh, K. N. Khasawneh, C. Song, and N. Abu-Ghazaleh, "Spectre returns! speculation attacks using the return stack buffer," in *12th USENIX Workshop on Offensive Technologies*, WOOT, 2018, pp. 13–14.
- [11] V. Kiriansky and C. Waldspurger, "Speculative buffer overflows: Attacks and defenses," *arXiv preprint arXiv:1807.03757*, 2018.
- [12] T. Warren, "Intel processors are being redesigned to protect against spectre," <https://www.theverge.com/2018/3/15/17123610/intel-new-processors-protection-spectre-vulnerability>.
- [13] V. Kiriansky, I. Lebedev, S. Amarasinghe, S. Devadas, and J. Emer, "Dawg: A defense against cache timing attacks in speculative execution processors."
- [14] M. Yan, J. Choi, D. Skarlatos, A. Morrison, C. W. Fletcher, and J. Torrellas, "Invisispec: Making speculative execution invisible in the cache hierarchy," in *Proceedings of the 51th International Symposium on Microarchitecture (MICRO'18)*, 2018.
- [15] Microsoft, "Spectre mitigations in msvc," <https://blogs.msdn.microsoft.com/vcblog/2018/01/15/spectre-mitigations-in-msvc/>.
- [16] C. Carruth, "Speculative load hardening," https://docs.google.com/document/d/1wwcfv3UV9ZnZVcGiGuoITT_61e_Ko3TmoCS3uXLcJR0/edit#heading=h.phdehs44eom6.
- [17] O. Oleksenko, B. Trach, T. Reiher, M. Silberstein, and C. Fetzer, "You shall not bypass: Employing data dependencies to prevent bounds check bypass," arxiv, Tech. Rep. arXiv:1805.08506, <https://arxiv.org/abs/1805.08506>, 2018.
- [18] Microsoft, "Protect your windows devices against spectre and meltdown," <https://support.microsoft.com/en-us/help/4073757/protect-your-windows-devices-against-spectre-meltdown>.
- [19] C. Reis, "Mitigating spectre with site isolation in chrome," <https://security.googleblog.com/2018/07/mitigating-spectre-with-site-isolation.html>.
- [20] P. Turner, "Retpoline: a software construct for preventing branch-target-injection," 2018.
- [21] M. Guarneri, B. Kopf, J. Morales, J. Reineke, and A. Sanchez, "Spectector: principled detection of speculative information flows," arxiv, Tech. Rep. arXiv:1812.08639, <https://arxiv.org/abs/1812.08639>, 2018.
- [22] D. Brumley, I. Jager, T. Avgerinos, and E. J. Schwartz, "BAP: A binary analysis platform," in *International Conference on Computer Aided Verification*. Springer, 2011, pp. 463–469.
- [23] F. Peng, Z. Deng, X. Zhang, D. Xu, Z. Lin, and Z. Su, "X-force: Force-executing binary programs for security applications." in *USENIX Security Symposium*, 2014, pp. 829–844.
- [24] M. Maurer, "Holmes: Binary Analysis Integration Through Datalog," 10 2018. [Online]. Available: https://kithub.cmu.edu/articles/Holmes_Binary_Analysis_Integration_Through_Datalog/7571519
- [25] wikichip, "Macro-operation fusion," https://en.wikichip.org/wiki/macro-operation_fusion, 2018.
- [26] "Cpuid," https://c9x.me/x86/html/file_module_x86_id_45.html.
- [27] Y. Shoshitaishvili, R. Wang, C. Salls, N. Stephens, M. Polino, A. Dutcher, J. Grosen, S. Feng, C. Hauser, C. Kruegel, and G. Vigna, "SoK: (State of) The Art of War: Offensive Techniques in Binary Analysis," in *IEEE Symposium on Security and Privacy*, 2016.
- [28] K. Kim, I. L. Kim, C. H. Kim, Y. Kwon, Y. Zheng, X. Zhang, and D. Xu, "J-force: Forced execution on javascript," in *Proceedings of the 26th International Conference on World Wide Web*, ser. WWW '17. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2017, pp. 897–906. [Online]. Available: <https://doi.org/10.1145/3038912.3052674>
- [29] L. Xu, F. Sun, and Z. Su, "Constructing precise control flow graphs from binaries," *University of California, Davis, Tech. Rep*, 2009.
- [30] A. Kejarawal, A. V. Veidenbaum, A. Nicolau, X. Tian, M. Girkar, H. Saito, and U. Banerjee, "Comparative architectural characterization of spec cpu2000 and cpu2006 benchmarks on the intel® core™ 2 duo processor," in *Embedded Computer Systems: Architectures, Modeling, and Simulation, 2008. SAMOS 2008. International Conference on*. IEEE, 2008, pp. 132–141.
- [31] Google, "oss-fuzz," <https://github.com/google/oss-fuzz/tree/master/projects>, 2018.
- [32] "Intel® xeon® gold 6126 processor," <https://ark.intel.com/products/120483/Intel-Xeon-Gold-6126-Processor-19-25M-Cache-2-60-GHz->
- [33] "Skylake (client) - microarchitectures - intel," [https://en.wikichip.org/wiki/intel/microarchitectures/skylake_\(client\)](https://en.wikichip.org/wiki/intel/microarchitectures/skylake_(client)).
- [34] J. L. Henning, "SPEC CPU2006 benchmark descriptions," *ACM SIGARCH Computer Architecture News*, vol. 34, no. 4, pp. 1–17, 2006.



Guanhua Wang received the M.E degree in computer science from Shandong University, Jinan, China in 2014, and he is currently pursuing the Ph.D degree with the Department of Computer Science, School of Computing, National University of Singapore. His research interests include GPU architecture, architecture security.



Sudipta Chattopadhyay received the Ph.D. degree in computer science from the National University of Singapore, Singapore, in 2013. He is an Assistant Professor with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore. His research interests include program analysis, embedded systems, and compilers.



Ivan Gotovchits received his MS in Electrical Engineering from Moscow Technological University in 2006 and worked as an engineer developing mission-critical highly automated systems. He is a chief developer with CyLab, Carnegie Mellon University.



Tulika Mitra received a BE degree in computer science from Jadavpur University, Kolkata, India, in 1995, an ME degree in computer science from the Indian Institute of Science, Bengaluru, India, in 1997, and a PhD degree from the State University of New York, Stony Brook, NY, USA, in 2000. She is currently a Professor of computer science at the School of Computing, National University of Singapore, Singapore. Her research interests include the design automation of embedded real-time systems with particular

emphasis on software timing analysis/optimizations, application-specific processors, energy-efficient computing, and heterogeneous computing.



Abhik Roychoudhury received the Ph.D. degree in computer science from the State University of New York at Stony Brook, Stony Brook, NY, USA, in 2000. He is a Professor of computer science with the National University of Singapore, Singapore. His research interests include software testing and analysis, software security, and trust-worthy software construction. Dr. Roychoudhury is also the Lead Principal Investigator of the Singapore Cyber-Security Consortium, which is a consortium of over 40 companies in

the cyber-security space engaging with academia for research and collaboration. He has served as an Editorial Board Member of the IEEE Transactions on Software Engineering from 2014 to 2018.