

A model of speculative evaluation

CRAIG DISSELKOEN, University of California San Diego

RADHA JAGADEESAN, DePaul University

ALAN JEFFREY, Mozilla Research

JAMES RIELY, DePaul University

1 INTRODUCTION

2 MODEL

2.1 Preliminaries

We assume:

- a set of *memory locations* \mathcal{X} , ranged over by x and y ,
- a set of *registers* \mathcal{R} , ranged over by r and s ,
- a set of *values* \mathcal{V} , ranged over by v and w ,
- a set of *expressions* \mathcal{E} , ranged over by M and N ,
- a set of *logical formulae* Φ , ranged over by ϕ and ψ , and
- a set of *actions* \mathcal{A} , ranged over by a and b ,

such that:

- values include at least the constants 0 and 1,
- expressions include at least registers and values,
- expressions are closed under substitution, written $M[N/r]$,
- formulae include at least true, false, and equalities of the form $(M = N)$ and $(x = N)$,
- formulae are closed under negation, conjunction, disjunction,
- formulae are closed under substitution, written $\phi[N/\ell]$, and
- actions include at least *reads* of the form $(R x v)$ and *writes* of the form $(W x v)$.

Let the set of *lvalues* be $\mathcal{L} = (\mathcal{X} \cup \mathcal{R})$, ranged over by ℓ and k .

2.2 Pomsets

Definition 2.1. A *pomset* (E, \leq, λ) with alphabet Σ is a partial order (E, \leq) together with a function $\lambda : E \rightarrow \Sigma$.

Going forward, we fix the alphabet $\Sigma = (\Phi \times \mathcal{A})$. We will write $(\phi \mid a)$ for the pair (ϕ, a) , a for (true, a) and f for (false, a) .

We visualize a pomset as a graph where the nodes are drawn from E , each node e is labelled with $\lambda(e)$, and an edge $d \rightarrow e$ corresponds to an ordering $d \leq e$. For example:

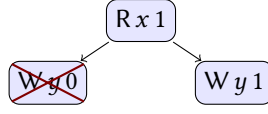
Authors' addresses: Craig Disselkoen, University of California San Diego, cdisselk@eng.ucsd.edu; Radha Jagadeesan, DePaul University, rjagadeesan@cs.depaul.edu; Alan Jeffrey, Mozilla Research, ajeffrey@mozilla.com; James Riely, DePaul University, jriely@cs.depaul.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2018 Copyright held by the owner/author(s).

XXXX-XXXX/2018/5-ART

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>



is a visualization of the pomset where:

$$0 \leq 1 \quad 0 \leq 2 \quad \lambda(0) = (\text{true}, R x 1) \quad \lambda(1) = (\text{false}, W y 0) \quad \lambda(2) = (\text{true}, W y 1)$$

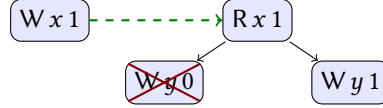
As we shall see, this is a possible execution of the program:

```
r := x; if (r) { y := 1; } else { y := 0; }
```

Definition 2.2. An *rf-pomset* is a pomset together with a $RF \subseteq E \times E$ such that for any $(d, e) \in RF$:

- $d < e$,
- $\lambda(d) = (\phi \mid W x v)$ and $\lambda(e) = (\psi \mid R x v)$, and
- there is no $d < c < e$ such that $\lambda(c) = (\chi \mid W x w)$.

We visualize rf-pomsets by drawing a dashed edge between edges in RF , for example:



As we shall see, this is a possible execution of the program:

```
x := 1; || r := x; if (r) { y := 1; } else { y := 0; }
```

Definition 2.3. An rf-pomset is *x-closed* if for $e \in E$ with $\lambda(e) = (\phi \mid a)$:

- ϕ is independent of x , and
- if a is an x read action, then there is a d with $(d, e) \in RF$.

2.3 Sets of pomsets

Let $\mathcal{P}_1 \sqcup \mathcal{P}_2$ be the set \mathcal{P}_0 where $P_0 \in \mathcal{P}_0$ whenever there are $P_1 \in \mathcal{P}_1$ and $P_2 \in \mathcal{P}_2$ such that:

- $E_0 = E_1 \cup E_2$,
- $RF_0 = RF_1 \cup RF_2$,
- if $e \leq_1 d$ or $e \leq_2 d$ then $e \leq_0 d$,
- if $\lambda_0(e) = (\phi_0 \mid a)$ then either:
 - $\lambda_1(e) = (\phi_1 \mid a)$ and $\lambda_2(e) = (\phi_2 \mid a)$ and ϕ_0 implies $\phi_1 \vee \phi_2$,
 - $\lambda_1(e) = (\phi_1 \mid a)$ and $e \notin E_2$ and ϕ_0 implies ϕ_1 , or
 - $\lambda_2(e) = (\phi_2 \mid a)$ and $e \notin E_1$ and ϕ_0 implies ϕ_2 .

Let $\mathcal{P}_1 \parallel \mathcal{P}_2$ be defined the same as $\mathcal{P}_1 \sqcup \mathcal{P}_2$ except that:

- $RF_0 \supseteq RF_1 \cup RF_2$, and for any $d, e \in E_i$, if $(d, e) \in RF_0$ then $(d, e) \in RF_i$.

Let $(\phi \mid W x v) \rightarrow \mathcal{P}$ be the set \mathcal{P}' where $P' \in \mathcal{P}'$ whenever there is $P \in \mathcal{P}$ such that:

- $E' = E \cup \{0\}$,
- $RF' = RF$,
- if $d \leq e$ then $d \leq' e$,
- $\lambda'(0) = (\phi, W x v)$, where ψ implies ϕ , and
- if $\lambda(e) = (\psi \mid a)$ then $\lambda'(e) = (\psi \mid a)$.

Let $(\phi \mid R x v) \rightarrow \mathcal{P}$ be the set \mathcal{P}' where $P' \in \mathcal{P}'$ whenever there is $P \in \mathcal{P}$ such that:

- $E' = E \cup \{0\}$,
- $RF' = RF$,

- if $d \leq e$ then $d \leq' e$,
- $\lambda'(0) = (\psi, R x v)$, where ψ implies ϕ , and
- if $\lambda(e) = (\psi \mid a)$ then $\lambda'(e) = (\psi' \mid a)$, ψ' implies $\psi[v/x]$, and $0 \leq' e$ or ψ' implies ψ .

Let $\mathcal{P}[M/\ell]$ be the set \mathcal{P}' where $P' \in \mathcal{P}'$ whenever there is $P \in \mathcal{P}$ such that:

- $E' = E$,
- $RF' = RF$,
- if $d \leq e$ then $d \leq' e$, and
- if $\lambda(e) = (\psi \mid a)$ then $\lambda'(e) = (\psi[M/\ell] \mid a)$.

Let $(\phi \mid \mathcal{P})$ be the subset of \mathcal{P} such that $P \in \mathcal{P}$ whenever:

- if $\lambda(e) = (\psi \mid a)$ then ϕ implies ψ .

Let $(vx . \mathcal{P})$ be the subset of \mathcal{P} such that $P \in \mathcal{P}$ whenever P is x -closed.

2.4 Semantics of programs

Define:

$$\begin{aligned}
 \llbracket \text{skip} \rrbracket &= \{\emptyset\} \\
 \llbracket x := M; C \rrbracket &= \bigcup_v (M = v \mid W x v) \rightarrow \llbracket C \rrbracket[M/x] \\
 \llbracket r := x; C \rrbracket &= \bigcup_v (\text{true} \mid R x v) \rightarrow \llbracket C \rrbracket[x/r] \\
 \llbracket \text{if } M \text{ then } C \text{ else } D \rrbracket &= (M \neq 0 \mid \llbracket C \rrbracket) \sqcup (M = 0 \mid \llbracket D \rrbracket) \\
 \llbracket C \parallel D \rrbracket &= \llbracket C \rrbracket \parallel \llbracket D \rrbracket \\
 \llbracket \text{var } x; C \rrbracket &= vx . \llbracket C \rrbracket
 \end{aligned}$$

3 EXAMPLES

3.1 No dependencies on writes

Consider an example with two independent writes $(x := 1; y := 2;)$. This has the semantics:

$$\bigcup_v (1 = v \mid W x v) \rightarrow \left(\bigcup_w (2 = w \mid W y v) \rightarrow (\{\emptyset\})[2/y] \right)[1/x]$$

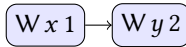
which is the same as:

$$\bigcup_v (1 = v \mid W x v) \rightarrow \bigcup_w (2 = w \mid W y v) \rightarrow \{\emptyset\}$$

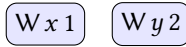
which includes the case where $v = 1$ and $w = 2$:

$$(1 = 1 \mid W x 1) \rightarrow (2 = 2 \mid W y 2) \rightarrow \{\emptyset\}$$

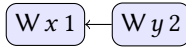
One of the executions this contains is:



but it also contains:



and:



since there is no requirement that $(W x 1) \leq (W y 2)$.

Thus, the semantics of $(x := 1; y := 2;)$ is the same as the semantics of $(y := 2; x := 1;)$.

3.2 Dependencies on reads

Whereas write prefixing introduces no new dependencies, read prefixing can. For example the program $(r := x; y := r+1;)$ has semantics:

$$\bigcup_v (\text{true} \mid R x v) \rightarrow \left(\bigcup_w (r + 1 = w \mid W y v) \rightarrow (\{\emptyset\}) [r + 1/y] \right) [x/r]$$

which is the same as:

$$\bigcup_v (\text{true} \mid R x v) \rightarrow \bigcup_w (x + 1 = w \mid W y v) \rightarrow \{\emptyset\}$$

which includes the case where $v = 1$ and $w = 2$:

$$(\text{true} \mid R x 1) \rightarrow (x + 1 = 2 \mid W y 2) \rightarrow \{\emptyset\}$$

Now, since true implies $(x + 1 = 2)[1/x]$, this contains:

$$\boxed{R x 1} \rightarrow \boxed{W y 2}$$

but since true does not imply $(x + 1 = 2)$, we have the requirement that $(R x 1) \leq (W y 2)$.

This is in contrast to the program $(r := x; y := r+2-r;)$, which has semantics:

$$\bigcup_v (\text{true} \mid R x v) \rightarrow \bigcup_w (x + 2 - x = w \mid W y v) \rightarrow \{\emptyset\}$$

Again, we consider the case where $v = 1$ and $w = 2$:

$$(\text{true} \mid R x 1) \rightarrow (x + 2 - x = 2 \mid W y 2) \rightarrow \{\emptyset\}$$

Now, since true implies $(x + 2 - x = 2)[1/x]$, this contains:

$$\boxed{R x 1} \rightarrow \boxed{W y 2}$$

but also true implies $(x + 2 - x = 2)$ (at least for arithmetic modulo n) and so this also contains:

$$\boxed{R x 1} \quad \boxed{W y 2}$$

and:

$$\boxed{R x 1} \leftarrow \boxed{W y 2}$$

Thus, the semantics of $(r := x; y := x+2-x;)$ is the same as the semantics of $(y := 2; r := x;)$.

3.3 Dependencies on conditionals

Conditionals introduce control dependencies, for example consider the program:

$r := z; \text{ if } (r) \{ x := 1; \} \text{ else } \{ y := 2; \}$

This has semantics:

$$\bigcup_v (\text{true} \mid R z v) \rightarrow (z \neq 0 \mid \llbracket x := 1 \rrbracket) \sqcup (z = 0 \mid \llbracket y := 2 \rrbracket)$$

Now $\llbracket x := 1 \rrbracket$ contains any pomset of the form:

$$\boxed{\phi \mid W x 1}$$

in particular it contains:

$$\boxed{r \neq 0 \mid W x 1}$$

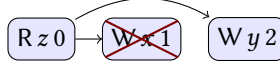
and similarly $\llbracket y := 2 \rrbracket$ contains any pomset of the form:

$$r = 0 \mid W y 2$$

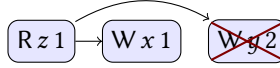
and so $\llbracket \text{if } (r) \{ x := 1; \} \text{ else } \{ y := 2; \} \rrbracket$ contains:

$$r \neq 0 \mid W x 1 \quad r = 0 \mid W y 2$$

Now, prefixing with $(R z v)$ performs substitutions, for example when v is 0 we have:



and when v is 1 we have:



3.4 Speculative evaluation

So far, the treatment of dependency has been fairly standard, we have included aborted events in the semantics of programs, but they have not had any impact. We shall now see that this is not always the case, that the execution of a conditional can depend on the branch not taken.

Consider the program:

$r := z; \text{ if } (r) \{ x := 1; \} \text{ else } \{ x := 1; \}$

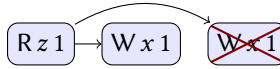
Many optimizers will consider it to be the same as:

$r := z; x := 1;$

and so there should not be a dependency between $(R z v)$ and $(W x 1)$. The semantics of this program is:

$$\bigcup_v (\text{true} \mid R z v) \rightarrow (z \neq 0 \mid \llbracket x := 1 \rrbracket) \sqcup (z = 0 \mid \llbracket x := 1 \rrbracket)$$

so, as in the previous section, we have executions such as:



but we also have the execution:



To see why, consider $\mathcal{P}_1 \sqcup \mathcal{P}_2$ where:

$$\mathcal{P}_1 = (z \neq 0 \mid \llbracket x := 1 \rrbracket) \quad \mathcal{P}_2 = (z = 0 \mid \llbracket x := 1 \rrbracket)$$

Now \mathcal{P}_1 contains the pomset P_1 where:

$$E_1 = \{0\} \quad 0 \leq_1 0 \quad \lambda_1(0) = (z \neq 0 \mid W x 1)$$

Similarly \mathcal{P}_2 contains the pomset P_2 where:

$$E_1 = \{0\} \quad 0 \leq_1 0 \quad \lambda_1(0) = (z = 0 \mid W x 1)$$

In particular, E_1 and E_2 overlap. One of the inhabitants of $\mathcal{P}_1 \sqcup \mathcal{P}_2$ is P_0 where:

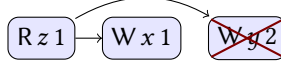
$$E_0 = \{0\} \quad 0 \leq_0 0 \quad \lambda_0(0) = (\text{true} \mid W x 1)$$

which is the execution we are interested in.

We can now see the counterintuitive behaviour of speculation, there are programs such as $r := z; \text{if } (r) \{ x := 1; \} \text{ else } \{ x := 1; \}$ with executions in which $(W x 1)$ is independent of $(R z 1)$:



while programs such as $r := z; \text{if } (r) \{ x := 1; \} \text{ else } \{ y := 2; \}$ only have executions in which $(W x 1)$ is dependent on $(R z 1)$:



so these programs have different dependency relations, depending on conditional branches that were not taken. In §3.6 we shall see that this has security implications, since relaxed memory can observe dependency. The attack is similar to Spectre, so we shall take a detour to see how Spectre can be modelled in this setting.

3.5 Spectre

We give a simplified model of Spectre attacks, ignoring the details of timing. In this model, we extend programs with the ability to tell whether a memory location has been touched (in practice this is implemented using timing attacks on the cache). For example, we can write a SPECTRE program as:

```

var a;
if (isCapability(0)) { a[SECRET] := 1; }
else if (touched a[0]) { x := 0; }
else if (touched a[1]) { x := 1; }
  
```

This is a low-security program, which is attempting to discover the value of a high-security variable SECRET. The low-security program is allowed to attempt to escalate its privileges by providing a capability which demonstrates that they are entitled to run high-security code:

```

if (isCapability(c)) { ... escalated code ... }
else { ... fallback code ... }
  
```

In this case, the `isCapability(0)` is false, so the fallback code is executed. Unfortunately, the escalated code is speculatively evaluated, which allows information to leak by testing for which memory locations have been touched.

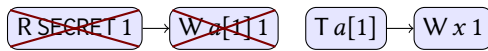
We model the touched test by introducing a new read action $(T x)$ and defining:

$$\llbracket \text{if touched } x \text{ then } C \text{ else } D \rrbracket = ((T x) \rightarrow \llbracket C \rrbracket) \sqcup \llbracket D \rrbracket$$

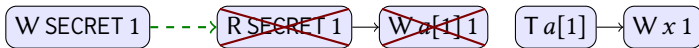
The additional requirement we need to add for x -closure is:

- if $\lambda(e) = (\phi \mid T x)$ then there is $d \not\approx e$ with $\lambda(d) = (\psi \mid R x v)$ or $\lambda(d) = (\psi \mid W x v)$.

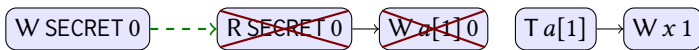
For example, one execution of SPECTRE is:



Putting this in parallel with a high-security write to SECRET gives:



but due the requirement of a-closure we do *not* have:



Thus, the attacker has managed to leak the value of a high-security location to a low-security one.

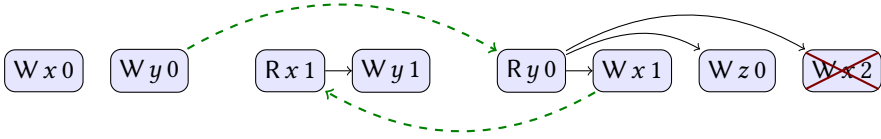
This shows how a (very abstract, untimed) model of Spectre attacks using speculative evaluation can be modelled.

3.6 Information flow attacks on relaxed memory

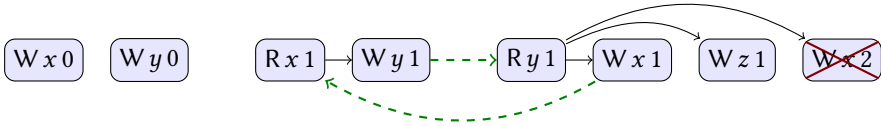
Consider an attacker program, again using security checks to try to learn a SECRET. Whereas SPECTRE uses hardware capabilities, which have to be modelled by adding extra capabilities to the language, this new attacker works by exploiting relaxed memory which can result in unexpected information flows. The attacker program is:

```
var x := 0; var y := 0;
(
  y := x;
) || (
  r := y;
  if (isCapability(r)) { x := SECRET; }
  else { x := 1; z := r; }
)
```

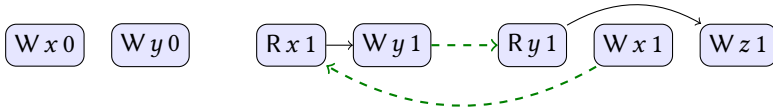
In the case where SECRET is 2, this has many executions, one of which is:



but there are no executions which exhibit (Wz1), since any attempt to do so produces a cycle:



In the case where SECRET is 1, there is an execution:



Note that in this case, there is no dependency from (Ry1) to (Wx1), which is what makes this execution possible. Thus, if the attacker sees an execution with (Wz1), they can conclude that SECRET is 1, which is an information flow attack.

This attack is not just an artefact of the model, since the same behaviour can be exhibited by compiler optimizations. Consider the program fragment:

```
r := y;
if (isCapability(r)) { x := SECRET; }
else { x := 1; z := r; }
```

Now, in the case where SECRET is a constant 1, the compiler can inline it:

```
r := y;  
if (isCapability(r)) { x := 1; }  
else { x := 1; z := r; }
```

lift the assignment to x out of the if statement:

```
r := y; x := 1;  
if (isCapability(r)) { }  
else { z := r; }
```

and then perform independent read/write reordering:

```
x := 1; r := y;  
if (isCapability(r)) { }  
else { z := r; }
```

After these optimizations, a sequentially consistent execution exhibits (W z 1).

4 EXPERIMENTS

5 CONCLUSIONS