1 APPENDIX

Here we prove the Theorem "Refined-Derivation" that is referred to in the proof in Section 3.3.3 of our OOPSLA '18 paper titled "Refinement in Object Sensitivity Points-to Analysis via Slicing".

1.1 STATEMENT

Given two alloc-site to length maps K_2 and K_1 such that $K_2(q) \geq K_1(q)$ for all q, and given a derivation R_2 based on K_2, there exists a derivation R_1 based on K_1, and there exist two functions

corrFact: from facts in R_2 to facts in R_1
corrObj: from objects in R_2 to objects in R_1

such that:

(*CorrProp1***):** If step m in R_2 processes a certain statement and uses tuples t_1' and t_2' to produce tuple t_3', then step m in R_1 processes the same statement, and uses $corrFact(t_1')$ and $corrFact(t_2')$ to produce $corrFact(t_3')$.

(*CorrProp2***):** If $corrFact(t') = t$, then for any object o' that is mentioned in t', if o is the corresponding object mentioned in t, then $corrObj(o') = o$. Also, the variable name (resp. field name) occurring in t' is the same as the variable name (resp. field name) occurring in t.

(*CorrProp3***):** For any o' in R_2, if $corrObj(o') = o$, then o' and o are allocated at the same allocation site.

(*CorrProp4***):** For any o' in R_2, $corrObj(o')$ is a suffix (proper or improper) of o'. We denote this as $corrObj(o') \leq o'$.

We say that run R_1 **corresponds** to the given run R_2.

1.2 PROOF

The proof is by induction on the number of steps in R_2. As part of the proof we constructively show the presence of R_1, *corrFact*, and *corrObj*, such that the properties mentioned above are satisfied.

The base case is the first step in R_2. This step must process an allocation site in 'main', because other rules need non-empty points-to sets to get triggered. Let this statement be $s_q : v = new$. In this case,

$$t' = t = (\epsilon, v) \longrightarrow o_q$$

We define:

$corrFact(t') = t$
$corrObj(o_q) = o_q$

It is easy to see *CorrProp1-CorrProp4* are satisfied after the first step.

The inductive hypothesis is that steps 1 to $(m-1)$ of Derivation R_2 have been processed, yielding the first $(m-1)$ steps of R_1. Also, that *corrFact* and *corrObj* have been partially defined so far, such that these partial functions in conjunction with the first $m-1$ steps of R_2 and the first $m-1$ steps of R_1 satisfy *CorrProp1-CorrProp4*.

The argument is now in cases, depending on the statement 'st' that is processed in step m.

switch (st) {

case "v = w": Say in step m of derivation R_2 the fact $(c', v) \longrightarrow o'_1$ gets generated from the fact $(c', w) \longrightarrow o'_1$.

Since $((c', w) \longrightarrow o'_1)$ was generated in the first m-1 steps of derivation R_2, by the inductive assumption of *CorrProp1* and *CorrProp2*, it follows that:

- $corrFact((c', w) \longrightarrow o'_1)$ is defined. Let this be equal to $((c, w) \longrightarrow o_1)$.
- $((c, w) \longrightarrow o_1)$ must have been generated in the first $m - 1$ steps of R_1.
- $corrObj(c') = c$, and $corrObj(o'_1) = o_1$.

We extend the *corrFact* function generated in the first $m - 1$ steps as follows:

$$corrFact((c', v) \longrightarrow o'_1) = (c, v) \longrightarrow o_1$$

(We assume that there are no repeated steps in R_2. Therefore, in Steps 1 to $m - 1$, $(c', v) \longrightarrow o'_1$ would not have been given an image under *corrFact*. This same point holds in all other cases below, whenever we extend *corrFact* or *corrObj*.)

We also define step m of R_1 as follows: it processes "v = w" using $((c, w) \longrightarrow o_1)$ to produce $((c, v) \longrightarrow o_1)$.

It is now easy to see that we have inductively re-established *CorrProp1* and *CorrProp2* for Steps 1 to m. *CorrProp3* and *CorrProp4* follow trivially because no object's image under *corrObj* was established newly in this step.

case "v = w.f": Say in step m in derivation R_2 existing facts $(c', w) \longrightarrow o'_1$ and $o'_1.f \longrightarrow o'_2$ are used to create the new fact $(c', v) \longrightarrow o'_2$.

Since $(c', w) \longrightarrow o'_1$ and $o'_1.f \longrightarrow o'_2$ were generated in the first m-1 steps of derivation R_2, by the inductive assumption of *CorrProp1* and *CorrProp2* it follows that:

- $corrFact((c', w) \longrightarrow o'_1)$ and $corrFact(o'_1.f \longrightarrow o'_2)$ are defined. Let these be equal to $(c, w) \longrightarrow o_1$ and $o_1.f \longrightarrow o_2$, respectively.
- $(c, w) \longrightarrow o_1$ and $o_1.f \longrightarrow o_2$ must have been generated in the first $m - 1$ steps of R_1.
- $corrObj(c') = c$, $corrObj(o'_1) = o_1$, and $corrObj(o'_2) = o_2$.

We extend the *corrFact* function generated in the first $m - 1$ steps as follows:

$$corrFact((c', v) \longrightarrow o'_2) = (c, v) \longrightarrow o_2$$

We also define step m of R_1 as follows: it processes "v = w.f" using $(c, w) \rightarrow o_1$ and $o_1.f \rightarrow o_2$ to produce $((c, v) \rightarrow o_2)$.

It is now easy to see that we have inductively re-established *CorrProp1* and *CorrProp2* for Steps 1 to m. *CorrProp3* and *CorrProp4* follow trivially because no object's image under *corrObj* was established newly in this step.

case "v.f = w": Say in step m in derivation R_2 existing facts $(c', w) \rightarrow o'_2$ and $(c', v) \rightarrow o'_1$ are used to create a new fact $o'_1.f \rightarrow o'_2$.

Since $(c', w) \rightarrow o'_2$ and $(c', v) \rightarrow o'_1$ were generated in the first $m - 1$ steps of derivation R_2, by the inductive assumption of *CorrProp1* and *CorrProp2* it follows that:

- $corrFact((c', w) \rightarrow o'_2)$ and $corrFact((c', v) \rightarrow o'_1)$ are defined. Let these be equal to $(c, w) \rightarrow o_2$ and $(c, v) \rightarrow o_1)$, respectively.
- $(c, w) \rightarrow o_2$ and $(c, v) \rightarrow o_1$ must have been generated in the first $m - 1$ steps of R_1.
- $corrObj(c') = c$, $corrObj(o'_1) = o_1$, and $corrObj(o'_2) = o_2$.

We extend the *corrFact* function generated in the first $m - 1$ steps as follows:

$$corrFact(o'_1.f \rightarrow o'_2) = o_1.f \rightarrow o_2$$

We also define step m of R_1 as follows: it processes "v.f = w" using $(c, w) \rightarrow o_2$ and $(c, v) \rightarrow o_1$ to produce $o_1.f \rightarrow o_2$.

It is now easy to see that we have inductively re-established *CorrProp1* and *CorrProp2* for Steps 1 to m. *CorrProp3* and *CorrProp4* follow trivially because no object's image under *corrObj* was established newly in this step.

case "s_q : v = new": Say this statement is in a method m_j. Say step m in derivation R_2 used an existing fact $(c', this_{m_j}) \rightarrow c'$ to create a new fact $(c', v) \rightarrow o'$, where $o' = mkName(c', q, K_2(q))$.

Since the fact $(c', this_{m_j}) \rightarrow c'$ was generated in the first $m - 1$ steps of derivation R_2, by the inductive assumption of *CorrProp1* and *CorrProp2*, it follows that:

- $corrFact((c', this_{m_j}) \rightarrow c')$ is defined. Let this be equal to $(c, this_{m_j}) \rightarrow c)$.
- $(c, this_{m_j}) \rightarrow c$ must have been generated in the first $m - 1$ steps of derivation R_1.
- $corrObj(c') = c$.

We extend the *corrFact* function generated in the first $m - 1$ steps as follows:

$$corrFact((c', v) \rightarrow o') = (c, v) \rightarrow o$$

where $o = mkName(c, q, K_1(q))$.

We also extend *corrObj* as follows:

$corrObj(o') = o$
We also define step m of Derivation R_1 as follows: It processes "$s_q : v = new$" using fact $(c, this_{m_j}) \rightarrow c$ to generate the fact $(c, v) \rightarrow o$.

It is now easy to that we have inductively re-established *CorrProp1* and *CorrProp2* for Steps 1 to m. Since o' and o are both allocated at site s_q, *CorrProp3* is established for Steps $1 - m$.

To establish *corrObj*, we now need to show that $o \leq o'$. By the inductive hypothesis, we know that $c \leq c'$. Therefore, $c.q \leq c'.q$. Now, o is the longest suffix of $c.q$ whose length is at most $K_1(q)$, while o' is the longest suffix of $c'.q$ whose length is at most $K_2(q)$. Because $K_2(q) \geq K_1(q)$, it follows that $o \leq o'$.

case call from "$\mathbf{a_2 = a_0.m(a_1)}$": Say in step m in Derivation R_2 this call is processed, using existing facts $(c', a_0) \rightarrow c'_1$ and $(c', a_1) \rightarrow o'_1$, to produce facts $(c'_1, this_{m_j}) \rightarrow c'_1$ and $(c'_1, p_j) \rightarrow o'_1$, where $m_j = dispatch(c'_1, m)$ and p_j is the formal parameter of m_j.

Since the facts $(c', a_0) \rightarrow c'_1$ and $(c', a_1) \rightarrow o'_1$ were generated in the first $m - 1$ steps in derivation R_2, by the inductive assumption of *CorrProp1* and *CorrProp2*, it follows that:

- $corrFact((c', a_0) \rightarrow c'_1)$ and $corrFact((c', a_1) \rightarrow o'_1)$ are defined. Let these be equal to $(c, a_0) \rightarrow c_1$ and $(c, a_1) \rightarrow o_1$, respectively.

- $(c, a_0) \rightarrow c_1$ and $(c, a_1) \rightarrow o_1$ must have been generated in the first $m - 1$ steps of R_1.

- $corrObj(c') = c$, $corrObj(c'_1) = c_1$, $corrObj(o'_1) = o_1$.

We extend the *corrFact* function generated in the first $m - 1$ steps a follows:

$$corrFact((c'_1, this_{m_j}) \rightarrow c'_1) = (c_1, this_{m_j}) \rightarrow c_1$$

$$corrFact((c'_1, p_j) \rightarrow o'_1) = (c_1, p_j) \rightarrow o_1$$

We define step m of R_1 as processing the same invoke statement, using existing facts $(c, a_0) \rightarrow c_1$ and $(c, a_1) \rightarrow o_1$, to produce facts $(c_1, this_{m_j}) \rightarrow c_1$ and $(c_1, p_j) \rightarrow o_1$. Note, this is a valid step because $dispatch(c_1, m)$ is necessarily equal to m_j. This follows from the inductive assumption *CorrProp3*, which implies that c'_1 and c_1 were both allocated at the allocation site; therefore, $dispatch(c'_1, m)$ must be equal to $dispatch(c_1, m)$.

It is easy to see that *CorrProp1* and *CorrProp2* are inductively re-established. *CorrProp3* and *CorrProp4* follow trivially because no object's image under *corrObj* was established newly in this step.

case return to "$\mathbf{a2 = a_0.m(a_1)}$": Say in step m of Derivation R_2 a return corresponding to this invoke is processed. Say this step uses existing facts $(c', a_0) \rightarrow c'_1$ and $(c'_1, ret_{m_j}) \rightarrow o'_2$ to produce the new fact $(c', a2) \rightarrow o'_2$, where $m_j = dispatch(c'_1, m)$.

Since the facts $(c', a_0) \rightarrow c'_1$ and $(c'_1, ret_{m_j}) \rightarrow o'_2$ were produced in the first $m - 1$ steps of Derivation R_2, by the inductive assumption of *CorrProp1* and *CorrProp2*, it follows that:

- $corrFact((c', a_0) \rightarrow c'_1)$ and $corrFact((c'_1, ret_{m_j}) \rightarrow o'_2)$ are defined. Let these be equal to $(c, a_0) \rightarrow c_1$ and $(c_1, ret_{m_j}) \rightarrow o_2$, respectively.

- $(c, a_0) \rightarrow c_1$ and $(c_1, ret_{m_j}) \rightarrow o_2$ must have been generated in the first $m - 1$ steps of Derivation R_1.

- $corrObj(c') = c$, $corrObj(c'_1) = c_1$, $corrObj(o'_2) = o_2$.

We extend the *corrFact* function generated in the first $m - 1$ steps as follows:

$$corrFact((c', a2) \rightarrow o'_2) = (c, a2) \rightarrow o_2$$

We define step m of R_1 as processing the return to the same invoke statement, using existing facts $(c, a_0) \rightarrow c_1$ and $(c_1, ret_{m_j}) \rightarrow o_2$, to produce the fact $(c, a2) \rightarrow o_2$. Note, this is a valid step because $dispatch(c_1, m)$ is necessarily equal to be equal m_j. This follows from the inductive assumption *CorrProp3*, which implies that c'_1 and c_1 were both allocated at the allocation site; therefore, $dispatch(c'_1, m)$ must be equal to $dispatch(c_1, m)$.

It is easy to see that *CorrProp1* and *CorrProp2* are inductively re-established. *CorrProp3* and *CorrProp4* follow trivially because no object's image under *corrObj* was established newly in this step.

} //end of switch-case