Technical Report

This document servers as the technical report of OOPSLA19 submission titled "*BDA: Practical Dependence Analysis for Binary Executables by Unbiased Whole-program Path Sampling and Per-path Abstract Interpretation*".

1 Proof of Theorem 4.1

Theorem 4.1. Using Algorithm 2, the probability $\widetilde{p}$ of any whole-program path being sampled satisfies equation 1, in which n is the total number of whole-program paths and L is the length of the longest path, which can be considered as $O(x)$ with x the number of nodes in $iCFG$.

$$(\frac{2^{63}}{2^{63}+1})^{2L} \cdot \frac{1}{n} \leq \widetilde{p} \leq (\frac{2^{63}+1}{2^{63}})^{2L} \cdot \frac{1}{n} \tag{1}$$

Proof. First, for any weight w_v, we prove that $\widetilde{w_v}$ follows $\frac{2^{63}}{2^{63}+1} \cdot w_v \leq \widetilde{w_v} \leq w_v$.

$$\begin{cases} exp = \max\left(\lfloor \log w_v \rfloor, 63\right) - 63 \\ sig = \lfloor w_v / 2^{exp} \rfloor \end{cases} \tag{2}$$

According to equation 2, if $w_v < 2^{64}$, $\widetilde{w_v} = w_v$. Otherwise, $sig \leq w_v/2^{exp} < sig + 1$, and hence $sig \times 2^{exp} \leq w_v < (sig+1) \times 2^{exp}$. As $sig \geq 2^{63}$ when $w_v \geq 2^{64}$, we have $\widetilde{w_v} \leq w_v < \frac{2^{63}+1}{2^{63}} \cdot \widetilde{w_v}$. Thus, $\frac{2^{63}}{2^{63}+1} \cdot w_v \leq \widetilde{w_v} \leq w_v$. As a result, the following holds.

$$\frac{2^{63}}{2^{63}+1} \cdot \frac{w_1}{w_1+w_0} \leq \frac{\widetilde{w_1}}{\widetilde{w_1}+\widetilde{w_0}} \leq \frac{2^{63}+1}{2^{63}} \cdot \frac{w_1}{w_1+w_0} \tag{3}$$

Let $p_1 = \frac{w_1}{w_1+w_0}$ be the accurate probability of choosing branch 1, the lighter-weight branch. $p_0 = \frac{w_0}{w_1+w_0}$ choosing the other. Thus, we can derive the following 4 from inequality 3.

$$\frac{2^{63}}{2^{63}+1} \cdot p_l \leq \frac{\widetilde{w_1}}{\widetilde{w_1}+\widetilde{w_0}} \leq \frac{2^{63}+1}{2^{63}} \cdot p_l \tag{4}$$

Next, we derive the bounds of $\widetilde{p_1}$, the probability of Algorithm **??** choosing branch 1. There are two cases.

(a) If $n < 64$, we directly have $\widetilde{p_l} = \widetilde{w_1}/(\widetilde{w_1}+\widetilde{w_0})$. According to inequality 4, we have the following.

$$\frac{2^{63}}{2^{63}+1} \cdot p_l \leq \widetilde{p_l} \leq \frac{2^{63}+1}{2^{63}} \cdot p_l \tag{5}$$

(b) If $n \geq 64$, $\widetilde{p_1} = \frac{\widetilde{w_1}.sig}{\widetilde{w_0}.sig \times 2^n}$. Note that $\frac{\widetilde{w_1}}{\widetilde{w_0}+\widetilde{w_1}} = \frac{\widetilde{w_1}.sig}{\widetilde{w_0}.sig \times 2^n + \widetilde{w_1}.sig}$. Thus, we have $\widetilde{p_1} \geq \frac{\widetilde{w_1}}{(\widetilde{w_1}+\widetilde{w_0})}$. Combining with inequality 4, we can have $\widetilde{p_1} \geq \frac{2^{63}}{2^{63}+1} \cdot p_l$. On the other hand, $\widetilde{p_1} = \frac{\widetilde{w_1}}{\widetilde{w_0}+\widetilde{w_1}} \cdot \frac{\widetilde{w_0}.sig \times 2^n + \widetilde{w_1}.sig}{\widetilde{w_0}.sig \times 2^n}$. Because $\widetilde{w_1}.sig < 2^{64} \leq 2 \cdot \widetilde{w_0}.sig$, we can have $\frac{\widetilde{w_0}.sig \times 2^n + \widetilde{w_1}.sig}{\widetilde{w_0}.sig \times 2^n} < \frac{\widetilde{w_0}.sig \times 2^n + \widetilde{w_0}.sig \times 2}{\widetilde{w_0}.sig \times 2^n} = \frac{2^{n-1}+1}{2^{n-1}}$. As $n \geq 64$ here, we can have $\widetilde{p_1} = \frac{\widetilde{w_1}}{\widetilde{w_0}+\widetilde{w_1}} \cdot \frac{\widetilde{w_0}.sig \times 2^n + \widetilde{w_1}.sig}{\widetilde{w_0}.sig \times 2^n} < \frac{\widetilde{w_1}}{\widetilde{w_0}+\widetilde{w_1}} \cdot \frac{2^{63}+1}{2^{63}}$. Combining with inequality 4, we can have $\widetilde{p_1} < (\frac{2^{63}+1}{2^{63}})^2 \cdot p_l$. Thus,

$$\frac{2^{63}}{2^{63}+1} \cdot p_1 \leq \widetilde{p_1} \leq (\frac{2^{63}+1}{2^{63}})^2 \cdot p_1 \tag{6}$$

From inequality 5 and 6, the following is true.

$$(\frac{2^{63}}{2^{63}+1})^2 \cdot p_1 \leq \widetilde{p_1} \leq (\frac{2^{63}+1}{2^{63}})^2 \cdot p_1 \tag{7}$$

Similarly, we can prove the bound for $\widetilde{p_0}$.

$\square$