A MOTIVATING EXAMPLE

A.1 Alarm Provenance and Interactive Prioritization

We now outline the computation of $\Pr(\text{Alarm}(37))$ and of $\Pr(\text{Alarm}(37) \mid \neg\,\text{Alarm}(36))$. Recall our assumption that the prior probability, $\Pr(\text{DUPath}(9,25)) = 0.9$, and that the probability of each rule application misfiring is 1%.

$$\begin{aligned}\Pr(\text{Alarm}(37)) &= \Pr(\text{Alarm}(37) \wedge \text{DUPath}(9,25)) + \\ &\quad \Pr(\text{Alarm}(37) \wedge \neg\,\text{DUPath}(9,25)) \\ &= \Pr(\text{Alarm}(37) \mid \text{DUPath}(9,25)) \times \\ &\quad \Pr(\text{DUPath}(9,25)) \\ &= 0.99^3 \times 0.9 = 0.873. \end{aligned} \tag{7}$$

The factor of 0.99^3 in the above calculation comes from the observation that there are three rule applications between the original hypothesis, DUPath(9, 25), and the final conclusion, Alarm(37).

Furthermore,

$$\begin{aligned}&\Pr(\text{Alarm}(37) \mid \neg\,\text{Alarm}(36)) \\ &= \Pr(\text{Alarm}(37) \wedge \text{DUPath}(9,30) \mid \neg\,\text{Alarm}(36)) + \\ &\quad \Pr(\text{Alarm}(37) \wedge \neg\,\text{DUPath}(9,30) \mid \neg\,\text{Alarm}(36)) \\ &= \Pr(\text{Alarm}(37) \wedge \text{DUPath}(9,30) \mid \neg\,\text{Alarm}(36)) \\ &= \Pr(\text{Alarm}(37) \mid \text{DUPath}(9,30)) \times \\ &\quad \Pr(\text{DUPath}(9,30) \mid \neg\,\text{Alarm}(36)) \\ &= 0.99^2 \times \Pr(\text{DUPath}(9,30) \mid \neg\,\text{Alarm}(36)).\end{aligned}$$

The penultimate step above follows from the conditional independence of the variables in the Bayesian network. We may apply Bayes' rule to simplify the second term:

$$\begin{aligned}&\Pr(\text{DUPath}(9,30) \mid \neg\,\text{Alarm}(36)) \\ &= \frac{\Pr(\neg\,\text{Alarm}(36) \mid \text{DUPath}(9,30)) \times \Pr(\text{DUPath}(9,30))}{\Pr(\neg\,\text{Alarm}(36))} \\ &= \frac{(0.01 + 0.99 \times 0.01) \times 0.99 \times 0.9}{1 - 0.873} = 0.140.\end{aligned}$$

Assembling these calculations, we conclude:

$$\Pr(\text{Alarm}(37) \mid \neg\,\text{Alarm}(36)) = 0.99^2 * 0.140 = 0.137. \tag{8}$$

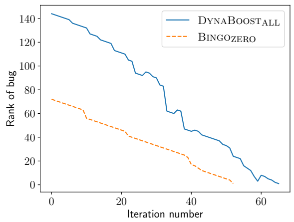
A.2 Dynamic Instrumentation as an Information Source

We transcribe the calculation of $\Pr(\text{Alarm}(38) \mid d \wedge e)$, where $d = \text{DUPath}(9,25)$, and $e = \neg\,\text{Alarm}(36)$. Recall our assumptions that the prior probability, $\Pr(d) = 0.9$, and that the probability of each rule application misfiring is 1%.

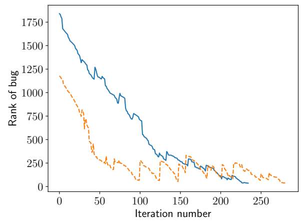
$$\begin{aligned}&\Pr(\text{Alarm}(38) \mid d \wedge e) \\ &= \Pr(\text{Alarm}(38) \wedge \text{DUPath}(9,30) \mid d \wedge e) \\ &= \Pr(\text{Alarm}(38) \mid \text{DUPath}(9,30)) \times \\ &\quad \Pr(\text{DUPath}(9,30) \mid d \wedge e) \\ &= 0.99^2 \times \frac{\Pr(d \wedge e \mid \text{DUPath}(9,30)) \times \Pr(\text{DUPath}(9,30))}{\Pr(d \wedge e)} \\ &= \frac{0.99^2 \times \Pr(\text{DUPath}(9,30))}{\Pr(d) \times \Pr(e \mid d)} \times \\ &\quad \Pr(d \mid \text{DUPath}(9,30)) \times \Pr(e \mid \text{DUPath}(9,30)) \\ &= \frac{0.99^2 \times (0.99 \times 0.9)}{0.9 \times (0.01 + 0.99 \times 0.01 + 0.99^2 \times 0.01)} \times \\ &\quad 1.0 \times (0.01 + 0.99 \times 0.01) \\ &= 0.650.\end{aligned} \tag{9}$$

The first step follows from the construction of the conditional probability distributions: because $d' = \text{DUPath}(9,30)$ occurs on every path to the alarm, $a = \text{Alarm}(38)$ can only be true when d' is also true. The second and fourth steps are justified from the conditional independencies, while the third step is a straightforward application of Bayes' rule.

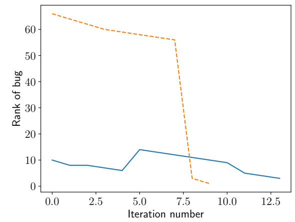
B EXPERIMENTAL EVALUATION



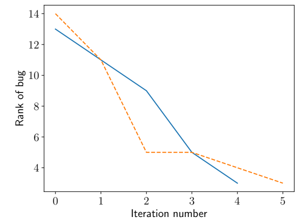
(a) grep-2.19



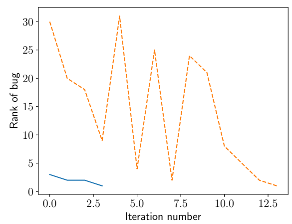
(b) gzip-1.2.4a



(c) libtasn1-4.3

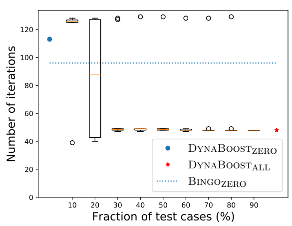


(d) latex2rtf-2.1.1

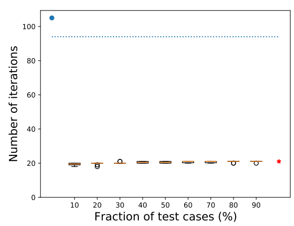


(e) optipng-0.5.3

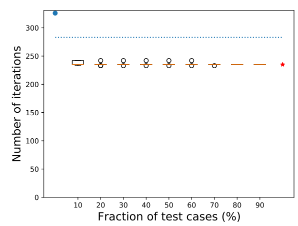
Figure 8: Plots for ranking for the true alarms within DynaBoost and Bingo.



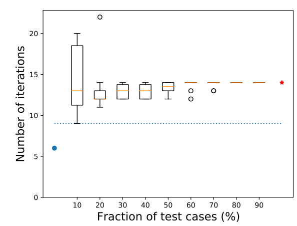
(a) bc-1.06



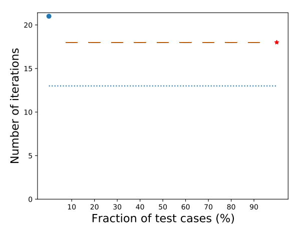
(b) cflow-1.5



(c) gzip-1.2.4a



(d) libtasn1-4.3



(e) shntool-3.0.5

Figure 9: Ranking effectiveness of DynaBoost when provided with a limited number of test cases, sampled from the full test suite.