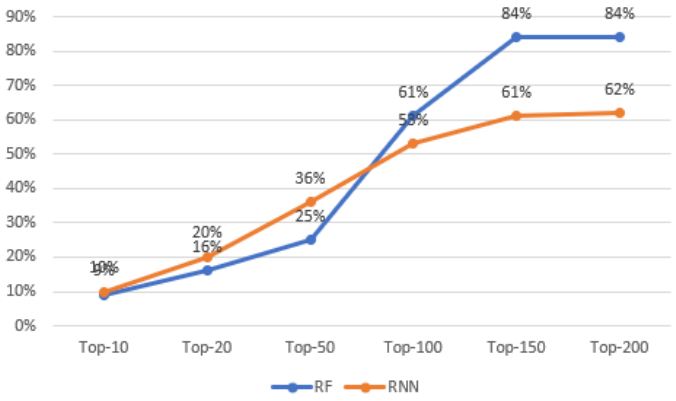
Summary of attempt one

Writer:
Fan Zhang
Tianlei Qi

Date: 25/05/2021

First prediction attempt

RNN (1085 test samples)
Top-10: 10%
Top-50: 36%
Top-100: 53%
Top-150: 61%
Top-200: 62%



RF (621 test samples)
TOP-10:9%
TOP-20:16%
TOP-50:25%
TOP-100:61%
TOP-150:84%
TOP-200: 84%

What do you think of the performance of the detection? Is it good or not good? Why?

According to the prediction results, both performances of these detection methods are higher than our expectation based on result of the paper. We think it is good due to the relatively high percentages of detected vulnerable functions in the top-200 functions of the sorted probabilities list. However, all the source data and code are provided by the paper and we strictly followed the steps mentioned in the paper, we have reasons to doubt the authenticity of the result from the paper. So, the credibility of the results still needs to be verified with other data sets.

What is your conclusion? Could you give the final conclusion about the method based on the current result?

We were strictly followed the code requirements to preprocess the train and test data and used the deep-learning methods from the paper, we need to exclude the special case that could appear during the experiment. Our first step is to train the models multiple times to exclude the special cases to validate whether there is much difference between every prediction attempt. If not, we would collect our own data set to verify these methods.