**Opening Draft**

Abstract: The rapid rise of Internet of Things (IoT) has made Industrial wireless sensor networks (IWSN) become more prevalent than ever before as it opens many possibilities with regards to control monitoring of these systems. The gained convenience however exposes additional security problem which may be fatal for these Cyber-Physical System especially for critical applications such as water treatment plant. To be filled later............ We examine an anomaly detection with a random forest classifier and propose a method for feature selection utilizing feature contribution. The method will be evaluated using a Secure Water Treatment Testbed (SWaT) dataset for evaluation.

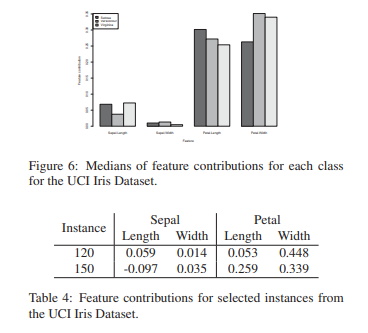
Keywords: Anomaly Detection, Industrial Wireless Sensor Networks, Random Forest Classifier, Feature Selection, Feature Contribution, Secure Water Treatment Testbed

Random forest is a popular machine learning algorithm due to its predictiveness and robustness. However, being a non-linear model, its hard to assess the affect a feature on the model, also called a black-box model. There are a couple of research in interpreting prediction made by a non-linear models. For random forests, there are 2 ways of measuring significance of a variable, variable importance and gini coefficient. Variable importance measure (vim) is the mostly used ones right now but it has a limitation. They only measure the importance of a variable as a whole using all values taken by the variable. However, sometimes there might only be one value of the feature that’s important for classification. Also, it doesn’t differentiate between positive (presence of a feature) and negative value (the lack of evidence for a property). This might be interesting because sometimes we’d like to know which features mostly that contributes to an attack vs non attack and VIM doesn’t really tell the difference as its measure the significance of a variable for the whole model. Other important mentions is that feature contribution is taken separately per instance and a median value of all feature contributions (over the training dataset) would be used to assess the whole model.

The idea is basically to utilize feature contribution as a variable selection by taking all the positive features and doing the random forest algorithm over and over until a threshold is reached where the model would have a good accuracy with the least amounts of features. Also, it would also tell us which feature that contributes the most to a certain class (in this case, attack) which might be of good insight.

In the research, I performed multiple test with the SWAT dataset. The details of the algorithm is as follows:

1. Separate training data and testing data.
2. Create a random forest model with the training data. We can use balanced random forest algorithm if the data is not balanced or we can perform a SMOTE as well (adding random data to compensate for the fewer class).
3. We perform a feature contribution on each instances and we should get a list of graphs of the median values of the feature contribution for the whole training dataset.
4. We pick a certain threshold and we cut positive/negative features that fell below the threshold.
5. Repeat until we find a number of variables that satisfy our preferences with respectable loss of accuracy (if any).



Example of feature contributions