Comparison of Machine Learning Algorithms and Their Ensembles for Botnet Detection
Songhui Ryu
Department of Computer and Information Technology, Purdue University, West Lafayette, IN 47907, USA
ryu26@purdue.edu

INTRODUCTION
Botnet is a network of compromised devices controlled by a Botmaster for malicious tasks. Botnet Detection, like Intrusion Detection System, can be done by monitoring observable botnet behaviors in network traffic. Machine learning techniques, nowadays, have been popularly used to detect abnormal traffics.

DATASET
The CTU-13 dataset is a public dataset featuring Botnet traffic mixed with normal and background traffic captured at the CTU university, Czech Republic in 2011. Among 13 different network traffic, features, #9, #10, and #11 were used.

TEST & RESULTS
Data Preparation
Among 12 features, Date and time, IP address and port number and the number of flows were excluded because Date was all the same, and there were very specific IP addresses that were set to Bots. Each test was done by splitting the dataset randomly into training and test set in ratio of 8:2. The following is the averaged values of 5 runs for each algorithm on a single machine with 64GB of memory.

Results
Accuracy scores of each algorithms and ensemble method

Although F1 score generally shows higher scores than MCC, MCC is more reliable because the true negatives are not considered in F1 score.

CONCLUSION
Findings
➢ Decision tree without any ensemble method would be the most preferable.
➢ Taking ensemble methods in a hope of enhancing the accuracy of machine learning algorithms for botnet detection is not likely to be a help to the actual detection.
➢ When a real-time detection system is considered, taking GNB or DT without any ensemble methods could be a good option.

Future works
➢ Data dependency: The accuracy results are much lower when it comes to the huge dataset. Possible reason can be overfitting or the very nature of the different bots (Neris & Rbot). Also, Boosting GNB shows near zero of MCC which means no better than random prediction. The reason for this is also to be studied.
➢ Other algorithms / dataset can be evaluated: SVM, k-NN or Random forest algorithms can be tested to see if they are appropriate for botnet detection. Also, more network attribute can be considered by making full use of NetFlow features.
➢ Scalability: Spark MLlib also provides classification algorithms. But ensemble methods are not developed yet. Finding out if it is feasible to implement the ensemble methods with Spark MLlib can be considered.