Online Malware Detection

Millions of new malware samples are submitted to VirusTotal.com every day, making the traditional detection method of hand-craft signatures obsolete. The use of supervised learning techniques to this problem becomes inevitable nowadays. However, in most state of the art works, the training set and test set are randomly generated, and albeit high accuracy is achieved (98%), this way of detection doesn't reflect the real use of such kind of system in practice. Moreover, some recent studies find that time split evaluations do report very low accuracy (67%) of such systems. We hypothesize that this is because of the evolution in behavior of malware, which traditional supervised method doesn't take into consideration.

We study the application of online learning to this problem of malware detection. In online learning, the data points are not available all at once but one at a time and reflect the real practice of malware detection. In addition, current state of art online learning algorithms are based on the static regret analysis which doesn't take into account the change in distribution of malware behavior and the result is only the best predictor in hindsight. However, dynamic regret analysis promises to derive better algorithms which can tackle non-stationary case.

In summary, we aim to achieve two objectives in this project: a novel algorithm for online learning, derived from the analysis of dynamic regret and an online malware detection system based on this algorithm.

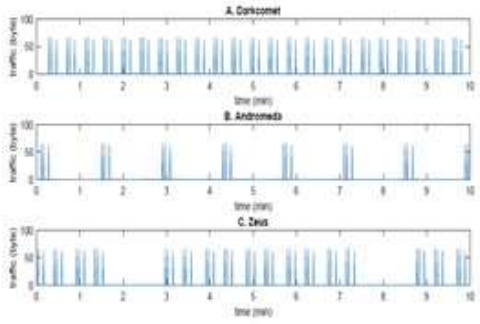
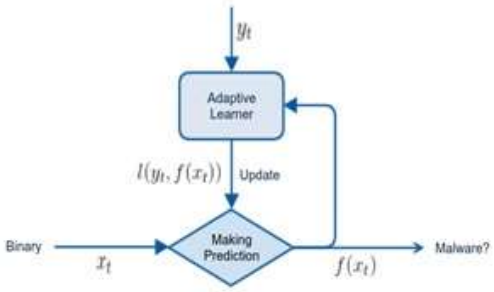


Fig. 1: Periodic network traffic generated by Darkcomet, Andromeda and Zeus, observed over a period of 10 minutes.



Principal Investigator	Assoc. Prof. Ng Wee Keong *(SCSE)*
Student	Mr. Huynh Ngoc Anh
School / Department	School of Computer Science and Engineering