PH.D
FINAL ORAL EXAMINATION

Thursday, November 3rd
10:00a.m. @ 131 Atanasoff

Liyuan Xiao

Major Professor: Carl Chang

Intrusion Detection Using Probabilistic Graphical Models

Due to the rapid growth of the internet applications, intrusion detection and prevention have become increasingly important research topics, in order to protect networking systems, such as the Web servers, database servers, cloud servers and so on, from threats. In this thesis, we attempt to build more efficient Intrusion Detection System through three different approaches, from different perspectives and based on different situations. Firstly, we propose Bayesian Model Averaging of Bayesian Network (BNMA) Classifiers for intrusion detection. In this work, we compare our BNMA classifier with Bayesian Network classifier and Naive Bayes classifier, which were shown be good models for detecting intrusion with reasonable accuracy and efficiency in the literature. From the experiment results, we see that BNMA can be more efficient and reliable than its competitors, i.e., the Bayesian network classifier and Naive Bayesian Network classifier, for all different sizes of training dataset. The advantage of BNMA is more pronounced when the training dataset size is small. Secondly, we introduce the Situational Data Model as a method for collecting dataset to train intrusion detection models. Unlike previously discussed static features as in the KDD CUP 99 data, which were collected without time stamps, Situational Data are collected in chronological sequence. Therefore, they can capture not only the dependency relationships among different features, but also relationships of values collected over time for the same features. The experiment results show that the intrusion detection model trained by Situational Dataset outperforms that trained by action-only sequences. Thirdly, we introduce the Situation Aware with Conditional Random Fields Intrusion Detection System (SA-CRF-IDS). The SA-CRF-IDS is trained by probabilistic graphical model Conditional Random Fields (CRF) over the Situational Dataset. The experiment results show that the CRF outperforms HMM with significantly better detection accuracy, and better ROC curve when we run the experiment on the non-Situational dataset. On the other hand, the two training methods have very similar performance when the Situational Dataset is adopted.

IOWA STATE UNIVERSITY
Department of Computer Science