SVM を用いた C&C セッションの分類による Botnet の検出方法

Botnet Detection Techniques by C&C Session Classification Using SVM

近藤　賢志
Satoshi Kondo

コンピュータウイルスやワームなどの悪性プログラムは，重要な社会インフラであるインターネットにおける深刻なセキュリティ脅威となっている．特に，bot と呼ばれる新たな種類の悪性プログラムは，感染した数千，数万のコンピュータを組織化し，大規模な DDoS 攻撃や大量の SPAM メール送信などによる深刻なセキュリティ被害をもたらしている．

本研究では，bot プログラムの特徴的な機能である遠隔制御セッション(C&C セッション)に着目し，C&C セッションを識別する事によって bot プログラムの検知を行う手法を提案する．

セッション識別アルゴリズムに SVM を使用し，パケットサイズと到達間隔にもとづいたパケットヒストグラムによる特徴ベクトルデータを用いた場合，学習データセット含まれない未知の C&C セッションに対しても 95%という高い識別率が得られた．

また C&C セッションの識別率だけではなく誤検知率,および処理速度においても,Naive Bayes，k-最近傍法と比較して，SVM が高い性能を示す事を確認した．これにより SVM を用いた C&C セッションの識別による bot 検知手法の有効性を示した．

A malignant program such as the computer viruses and worms is the serious security threats in the Internet that is an important social infrastructure.　Especially, the new kind of malignant program that is called bot organizes infected several thousand and tens of thousands of computers, and has caused serious security damage by a large-scale DDoS attack and a large amount of SPAM Mail Sending, etc.

In this research, it pays attention to the remote control session (C&C session) that is a peculiar function to the bot program, and it proposes the technique for detecting the bot program by identifying the C&C session.

When SVM was used for the session identification algorithm, and the feature vector data by the packet histogram that was based at the packet size and intervals was used, a high identification rate of 95% was obtained for the unknown C&C session of the study data set not included.

Moreover, it was not only an identification rate of the C&C session and it was confirmed to show the performance that SVM is high compared with Naive Bayes and k-Nearest Neighbor at the false positive detection rate and the processing speed. As a result, it was shown that the effectiveness of the bot detection technique by the identification of the C&C session that used SVM.