Malicious Code Detection and Acquisition Using Active Learning

Robert Moskovitch, Nir Nissim, Yuval Elovici
Deutsche Telekom Laboratories at Ben-Gurion University
Ben Gurion University, Be'er Sheva, 84105, Israel
{robertmo,nirni, elovici}@bgu.ac.il

Detection of known malicious code is commonly performed by anti-virus tools. These tools detect the known malicious code using signature detection methods. Each time a new malicious code is found the anti-virus vendors create a new signature and update their clients. During the period between the appearance of a new unknown malicious code and the update of the signature base of the anti-virus clients, millions of computers might be infected. In order to cope with this problem, new solutions must be found for detecting unknown malicious code at the entrance of a client's computer. Recent studies have shown that machine learning methods can be used for detecting *unknown* malicious executables based on their binary code. These methods are highly inspired by text categorization techniques, in which words are analogous to sequences in the binary code. In this approach, *malicious* and *benign* files are represented by a vector of features, extracted from the *p-header* and the binary code of the executable. These files are used to train a classifier. During the detection phase, based on the generalization capability of the classifier, an unknown (did not appear in the repository) file can be classified as malicious or benign.

Constantly training the classifier with new files (malicious or benign) is essential to maintain the detection accuracy along time. However, when a file is classified, the classifier cannot indicate whether it should be acquired as a new example or not. Additionally, in order to add the file to the training set, a labeling operation which is done by a human expert is required. The labeling is a very time consuming task since each *unknown* file (suspected as malicious) has to be analyzed by an expert. The acquisition of new files can be made using honey pots or at important network nodes. Since there are many files to inspect, malicious and especially benign, it is not feasible to label all of them by a human expert. Thus, an efficient way to identify new files which are mostly important to be labeled an acquired should be provided. We propose to use *active learning* to reduce the amount of labeled training examples while maintaining classification accuracy. Studies in several domains successfully applied active learning in order to reduce the effort of labeling examples. Unlike in random learning, in which a classifier is trained on a pool of labeled examples, the classifier actively indicates the specific examples should be labeled, which are commonly the most informative examples for the training task.

Figure 1 illustrates the acquisition circle in which files arriving from the network are classified by a classifier, which was trained based on a repository of malicious and benign files. The classifier has three outputs: malicious, benign and suspected as new which is sent for labeling by a Security Expert. Then the labeled file is added to the repository.

An active learning framework, consisting on a *Support Vector Machine* classifier, including several active learning methods (criterions) was implemented. For evaluation a dataset consisting on 1182 files represented by the top 200 features, which were extracted by 5-grams representation, was created. Preliminary encouraging results of the comparison of two active and random learning methods are shown in Figure 2. The *ErrorReduction* active learning method outperformed the *Random Selection*, used commonly by classifiers, and also the *full data* (constant line) when trained above 300 files.

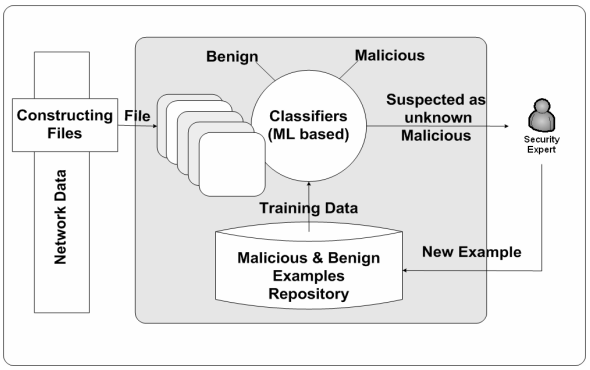

Figure 1. Acquiring and maintaining the knowledge of a classifier using active learning.

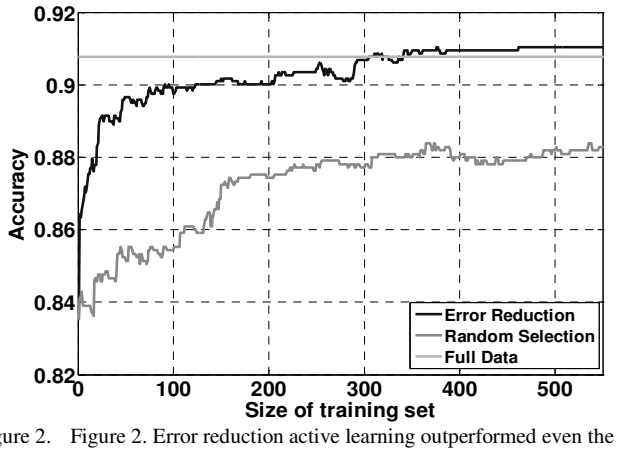

Figure 2. Figure 2. Error reduction active learning outperformed even the full data accuracy above a training set of 300 files.

We presented here the use of active learning in the acquisition of unknwon malicious code. Preliminary Results are encouraging. We are currently in the process of creating a wide test collection of more than 30,000 benign and malicious files to evaluate several active learning criterions.