ALADIN: Active Learning for Statistical Intrusion Detection

Jack W. Stokes, John C. Platt
Microsoft Research
One Microsoft Way
Redmond, WA 98052
Email: {jstokes,jplatt}@microsoft.com

Joseph Kravis
Microsoft Network Security
Redmond, WA 98052 USA
Email: jkravis@microsoft.com

Michael Shilman
Email: michael@shilman.net

To create host-based or network-based intrusion detection systems, we propose ALADIN which stands for "Active Learning of Anomalies to Detect INtrusions". ALADIN uses *active learning* combined with *rare class discovery* and *uncertainty identification* to statistically train an intrusion detection or prevention system (IDS/IPS). Active learning selects "interesting traffic" to be shown to a security expert for labeling to substantially reduce the number of labels required from an expert to reach an acceptable level of accuracy and coverage.

Our system defines "interesting traffic" in two ways, based on two goals for the system. The system is designed to *discover* new categories of traffic by showing examples of traffic for the analyst to label that do not fit any of the pre-existing models of known categories of traffic. The system is also designed to *accurately classify* known categories of traffic by requesting labels for examples which it cannot classify with high certainty. Combining these two goals overcomes many problems associated with earlier anomaly-detection based IDSs.

Recent work has investigated algorithms that combine active learning and anomaly detection [2]. Separately, an IDS has been proposed using active learning for improved classification accuracy [3]. It is possible to simply run both of these algorithms to create an intrusion detection system which both finds new intrusions and refines the rules for existing known categories. However, a security analyst would then be bombarded with labels: neither algorithm would cooperate or share labels.

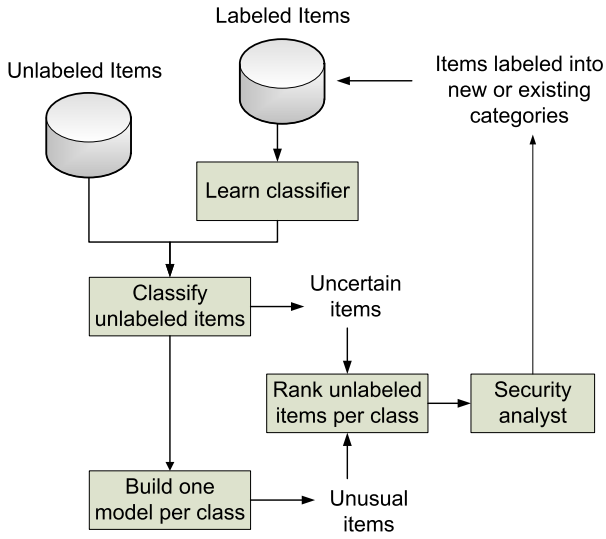

Fig. 1. ALADIN algorithm for detecting malware from intrusion detection logs.

Therefore, we propose the ALADIN algorithm in figure 1: a single intrusion detection framework for both anomaly detection and classification. A classifier is trained from labeled items which is used to predict the class for each of the unlabeled items. From the predicted labels of the unlabeled items, we choose samples for each class that lie closest to the margins of the classifier. We also build a model for each class from the labeled samples and the predicted labels for the unlabeled samples; these models are then used to identify samples which do not belong to its predicted class. A ranking function selects new samples for the security analyst to label; after labeling, the whole process is repeated. This allows analysts to quickly filter out common categories of traffic and identify rare anomalies that are new security risks. The main contribution of this work is a new algorithm that both quickly finds new classes of traffic using anomaly detection and also creates classifiers with high prediction accuracy.

To be clear, the combination of classification and anomaly detection is only used during the *training* phase. For real-time (IPS) or off-line (IDS) detection, the classifier's weights are used as a statistical misuse system. Weights

for a general purpose IDS/IPS are first generated by security analysts working for the IDS manufacturer using ALADIN. Optionally, analysts working for the organization using the IDS/IPS can further refine (i.e. train) the system on location specific traffic also using ALADIN.

In figure 2, we seek to compare how quickly ALADIN and two other supervised algorithms discover classes in the data set using the KDD-Cup 99 data set. The "SVM" algorithm is an active learning-based IDS similar to [3] which has a single classification stage using the Radial Basis Function (RBF), Support Vector Machine (SVM). The "Logistic Regression" algorithm uses logistic regression for classification instead of the SVM. For each iteration, 100 total, new samples were labeled. The results clearly show that ALADIN's use of active anomaly detection significantly outperforms standard active learning; adding the second stage of anomaly detection requires only half the number of samples to be investigated and labeled by an analyst compared to the next best alternative. Additional results show that combining a classifier and an anomaly detector improves the error rate compared to an IDS based solely on active learning.

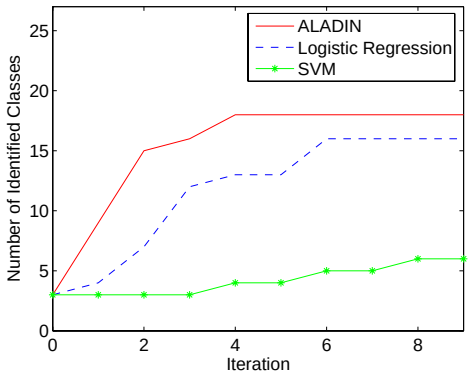

Fig. 2. Number of identified classes using the proposed ALADIN algorithm and two versions of supervised active learning algorithm.

We have also used the algorithm to analyze several daily logs of outbound network traffic, with over 13 million transfers, from Microsoft's worldwide corporate network. The algorithm discovered a previously unknown instance of malware on the corporate network in addition to a number of other forms of malware that were logged, but not yet identified.

REFERENCES

[1] A. Valdes, "Detecting novel scans through pattern anomaly detection," in *Proc. DISCEX*, 2003, pp. 140–151.

[2] D. Pelleg and A. Moore, "Active learning for anomaly and rare-category detection," in *Proc. Advances in Neural Information Processing Systems*, 2004, pp. 1073–1080.

[3] M. Almgren and E. Jonsson, "Using active learning in intrusion detection," in *Proc. IEEE Computer Security Foundations Workshop*, 2004, pp. 88–98.

[4] G. Schohn and D. Cohn, "Less is more: Active learning with support vector machines," in *Proc. Int'l. Conf. Machine Learning*, 2000, pp. 839–846.