Network intrusion detection

CS221 Artificial Intelligence Project Proposal
Amir Ziai

Task
In this project I will focus on the task of network intrusion detection. Network operators are generally aware of common attack vectors that they defend against. For most networks the vast majority of traffic is legitimate. However new attack vectors are continually designed and attempted by bad actors which bypass detection and go unnoticed due to low volume. One strategy for finding such activity is to look for anomalous behavior.

Scope
I will start with exploring unsupervised machine learning methods for detecting anomalous behavior. I will then explore the possibility of improving on this methodology by incorporating a small set of labeled data. Labeling network activity can be very time and resource intensive. In many cases security analysts need to investigate anomalous behavior for days or even weeks and this process is not very scalable. For the final part of the project I will explore active learning to guide the analyst in focusing on the most uncertain data points in an attempt to improve the sampling efficiency.

Dataset
I will use the KDD Cup 1999 dataset with 4.8M network connections in a military network environment. Each record is labeled as either “normal” or one of 22 different types of intrusion. Examples of these intrusions include smurf, IP sweep, and teardrop. Each record has 41 features including duration, protocol, and number of bytes exchanged between the source and destination.

Input and output example
The following table depicts 3 rows of data (excluding the label):

duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count
0	tcp	http	SF	181	5450	0	0	0	0	...	9
0	tcp	http	SF	239	486	0	0	0	0	...	19
0	tcp	http	SF	235	1337	0	0	0	0	...	29

The objective of the detection system is to label each row as either “normal” or “anomalous” using an unsupervised machine learning approach.

Evaluation

Since labeled data is very hard to come by in this space I have decided to treat this problem as an unsupervised learning one. Therefore the machine learning model does not receive the label information. However I will use the labels for evaluation purposes. Specifically I will use the F1 score to capture the trade-off between precision and recall. A model that is highly precise (does not produce false positives) is desirable as it won't waste the analyst's time but this usually comes at the cost of being too conservative and not catching anomalous activity that is indeed an intrusion.

Oracle and baseline

Since we have the labeled data the oracle knows all the correct answers and the F1 score is 1. To achieve a baseline I trained an isolation forests model [1] using scikit-learn.

Isolation forest is a collection of isolation trees where each tree randomly selects a feature and then randomly selects a value in the range to split on. This process is repeated to isolate each example. We then repeat the process 100 times to grow out 100 trees. Anomalous data points are on average closer to the root of the trees, i.e. it takes less cuts to isolate them.

Using this model I was able to get an F1 score of 0.9881. In contrast a random classifier that assigns random labels of "normal" and "anomalous" yields an F1 score of 0.9873. This may seem like an insignificant improvement but it's an improvement of 7x in true positives (actual anomalous activity). However there's certainly a lot of room of improvement relative to the oracle.

Prior work and proposal

Anomaly detection is an active area of research. Many algorithms exist that operate based on distribution assumptions (e.g. elliptic envelopes), partitioning (e.g. isolation forests), or density deviation (e.g. Local Outlier Factor [2]). Other methods use deep learning methods such as Recurrent Neural Networks [3] and Autoencoders. Researchers have observed that incorporating even a small set of labeled examples can significantly improve these methods. However collecting labeled data can be very expensive in this application and therefore judicious sampling is highly desirable [4]. This project focuses on combining the ideas from anomaly detection and active learning to improve network intrusion detection.

References

1- Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation-based anomaly detection." ACM Transactions on Knowledge Discovery from Data (TKDD) 6.1 (2012): 3.
2- Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000, May). LOF: identifying density-based local outliers. In ACM sigmod record (Vol. 29, No. 2, pp. 93-104). ACM.
3- Tuor, A., Baerwolf, R., Knowles, N., Hutchinson, B., Nichols, N., & Jasper, R. (2017). Recurrent Neural Network Language Models for Open Vocabulary Event-Level Cyber Anomaly Detection. arXiv preprint arXiv:1712.00557.
4- Pimentel, T., Monteiro, M., Viana, J., Veloso, A., & Ziviani, N. (2018). A Generalized Active Learning Approach for Unsupervised Anomaly Detection. arXiv preprint arXiv:1805.09411.