Applications of Generative Adversarial Networks to Intrusion Detection Systems

James Halvorsen and Assefaw Gebremedhin

GAN INTRODUCTION

- Created in 2014 by Ian Goodfellow, Generative Adversarial Networks (GANs) are a major innovation in generative machine learning.
- Previous generative models suffered from intractable problems, and GANs have enabled more serious development in this field.
- Uses two neural networks that train via an adversarial process: a generator and a discriminator.
- Generator inputs noise, outputs data samples.
- Discriminator inputs data samples, outputs a label (real or fake).
- Various improvements on GAN construction:
  - Conditional GANs (CGAN)
  - Deep Convolutional GANs (DCGAN)
  - Bidirectional GANs (BiGAN)

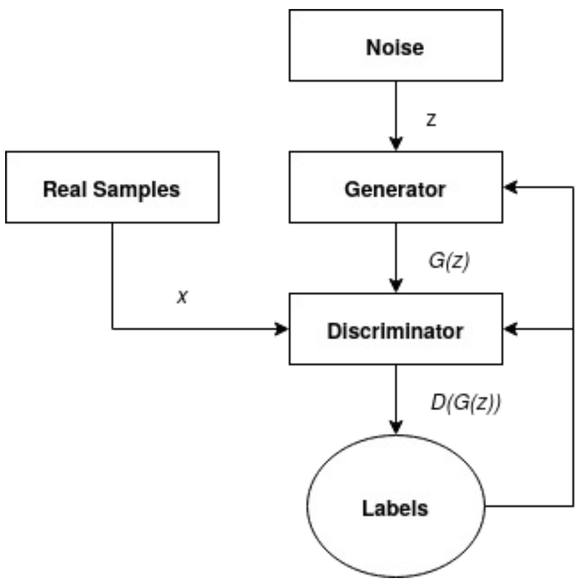

Figure 1: The training process of a Generative Adversarial Network. The generator receives noise as input and produces synthetic samples. The discriminator combines these samples with real data to produce a set of (real, fake) labels. The generator then adjusts its weights based on those labels to try to fool the discriminator on subsequent iterations.

IDS INTRODUCTION

- An Intrusion Detection System (IDS) is an application designed to detect when a network or host is experiencing a cyber attack.
- May be a part of an Intrusion Prevention System (IPS) which both detects and responds to attacks.
- Requires input from cyber monitoring sources (e.g. NetFlow, OS security logs)
- Typically developed as either:
  - Signature-Based IDS – Considers if monitored activity is similar to attacks that have previously occurred.
  - Anomaly-Based IDS – Considers if monitored activity is dissimilar to normal behavior.
- Machine Learning is seeing increasing usage for Anomaly-Based Intrusion Detection.

IDS PROBLEMS

- **Zero-Day Attacks**: These are attacks that have never been seen before and may target undisclosed vulnerabilities in either the IDS or the host system itself.
  - Signature-Based IDS have little to no protection against zero days.
  - Anomaly-Based IDS may be much more effective, depending on implementation.
- **False Positives**: This is when an IDS raises an alert for otherwise benign activity.
  - Too many false positives will make it easy to overlook actual attacks, or cause users to turn the IDS off.
  - Anomaly-Based IDS have much higher false positive rates.
- **Dataset Problem**: The quality and availability of security datasets for IDS development is very poor. Examples of dataset issues include:
  - Lack of labels
  - Unbalanced datasets (i.e. not enough attack data)
  - Dataset is too old (guess how KDD99 got its name)

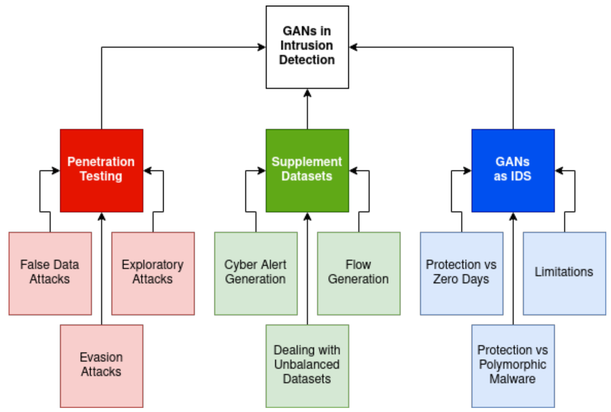

Figure 2: An overview of the applications of GANs in Intrusion Detection.

HOW DO GANS HELP IDS

Solving the Dataset Problem

- Many existing cyber datasets are highly imbalanced, with few attacks compared to mostly benign traffic.
- GANs can create synthetic attack data that is similar to existing attack data.

Hardening IDS Training

- GANs may be trained to produce attacks designed to evade a given IDS.
- The IDS may in turn be re-trained using these adversarial attacks to be more secure against unforeseen threats.

GANs as IDSs

- Can use the discriminator of a GAN to function as an IDS
- Discriminator outputs three labels (real, fake, anomaly)
- This training incorporates existing benefits of the generator (hardened training, more balanced datasets) to making a better IDS overall.
- Existing research shows GAN-based IDSs have high accuracy in detecting zero-day attacks.

SURVEY OF GANS IN IDS

Our first contribution to the subject of GANs in Intrusion Detection is a survey paper. The main topics being surveyed are depicted in Figure 2.

Why a survey?

- GANs are a relatively new technology, but there is already a lot of work made that either uses GANs alongside, or in place of, IDSs.
- A thorough review of existing works will help researchers understand current capabilities and limitations of GAN-based solutions, and areas for future work.

FUTURE WORK

The dataset problem explored through this research is of particular interest to us, and we have several planned works for this area of research. These include:

- A GAN that uses a pre-trained classifier for supervised training. We intend to use this to generate NetFlows and other data sources.
- A comparison of GANs to other types of generative models (i.e. Variational Autoencoders) for generating cyber monitoring data.
- A review of metrics for evaluating the quality of synthetic cyber monitoring data.

WASHINGTON STATE UNIVERSITY