Machine Learning for Network Security

Building Machine Learning Models to Detect and Mitigate Distributed Denial of Service Attacks

Ashrith Barthur

H_2O.ai

October 20, 2019

Contents

Overview

This presentation introduces us to three topics. These topics are:

1. The 101 on Distributed Denial of Service Attack.
2. Building a Machine Learning model to identify the attack.
3. Feature building for Identification
4. Deployment Architecture

H2O.AI Overview

Company	Founded in Silicon Valley in 2012 Funded: $147M Investors: Goldman Sachs, Paxion Ventures, Ping An, Nexus Ventures, NVIDIA, Wells Fargo.
Products	• H2O Open Source Machine Learning (18,000 organizations) • H2O Driverless AI – Automatic Machine Learning
Team	175 AI experts (Expert data scientists, Kaggle Grandmasters, Distributed Computing, Visualization)
Global	Mountain View, NYC, London, Prague, India

Industry Footprint

Driverless AI

H2O Driverless AI: Automatic Machine Learning

21 day free trial for Driverless AI

- Automatic AI and ML in a single platform
- AI to do AI
- Delivers insights and interpretability
- Provides easy to understand results and visualizations

DoS

What is a DoS attack?

A denial-of-service (DoS) attack occurs when a system floods the bandwidth or resources of a targeted system. Preventing or slowing the ability of the targeting system from further servicing any requests.

DDoS

What is a DDoS attack?

"A distributed denial-of-service (DDoS) attack occurs when multiple systems flood the bandwidth or resources of a targeted system."

Current Detection

How are we detecting DDoS up until now?

1. We largely use Rule-Based systems and constant monitoring to identify potential DDoS traffic.
2. Much of this is depending on investigation, monitoring, filtering.

Then what remains to be a problem?

1. A lot of these decisions are made after **traffic patterns are recognised, analysed, and modeled.**
2. This takes quite a bit of time causing large-scale systems and networks damage.
3. The Rules themselves are slow to identify new behaviour.

DDoS Data

1. The data is an initial sliver of DDoS traffic that was captured.
2. The attack was targeting a *http* server.

DDoS Data Transformations

1. Building these transformations requires one to decide, if this is **analytical** or **production**.
2. Analytical models are great for studying the behaviour and patterns, but could be slow when implemented.
3. Production models have simple traffic transformations but are quick.
4. By keeping the transformations simplistic, you keep pre-processing light, and allow the model to do the heavy lifting.

DEMO

How Did We Build This?

Driverless AI provides an extension.
This is a class ‘CustomTransformer‘

```
class ExampleLogTransformer(CustomTransformer):
```

How Did We Build This?

The class has:

1. Parameters that need to be provided.
2. These parameters are specific to the type of feature recipe that you are building.
3. It also has four methods which primary handle your feature engineering transformation.

Parameters - Basic

```
class ExampleLogTransformer(CustomTransformer):
_regression = True
_binary = True
_multiclass = True
```

Advantages

1. Feature engineering process standardised by:
   1.1 preset parameters
   1.2 preset methods
2. Effort minimisation leads to minimisation in time spent.
3. Build only once - Feature engineering is carried over from training/testing to production.
4. DAI automatically, runs multiple models on various sets of features to get the best model.
5. All the requirements are handled internally by DAI.

Deployment Architecture

1. Model is available as a Mojo - Java or C++, or as a Python Scorer. Depending on the infrastructure.
2. Model can score at a speed of 1.6M records per second.
3. Model is super-light, can be deployed at the edge node.
4. Model is independent, does not need help with decisions, (artifically intelligent)
5. Can independently, stop malicious traffic in a shorter amount of time, much before its intercepted, analysed, and filtered.
6. Infrastructural damages can be prevented. Service availability is assured.

References

How to build a recipe, Ashrith Barthur
https://github.com/h2oai/driverlessai-recipes/tree/master/how_to_write_a_recipe

A Streaming Statistical Algorithm for Detection of SSH Keystroke Packets in TCP Connections, DTIC
Guha, Saptarshi ; Kidwell, Paul ; Barthur, Ashrith ; Cleveland, William S ; Gerth, John ; Bullard, Carter
https://apps.dtic.mil/dtic/tr/fulltext/u2/a534101.pdf

Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization", 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018
Ilman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani

Thanks & Questions