Nvidia AI-Based Firewall

Abstract

The goal of this project is to develop a live AI-based firewall.

In this project you will get a glance of the work of system architect. You will use many components to build a complex but working network-based system.

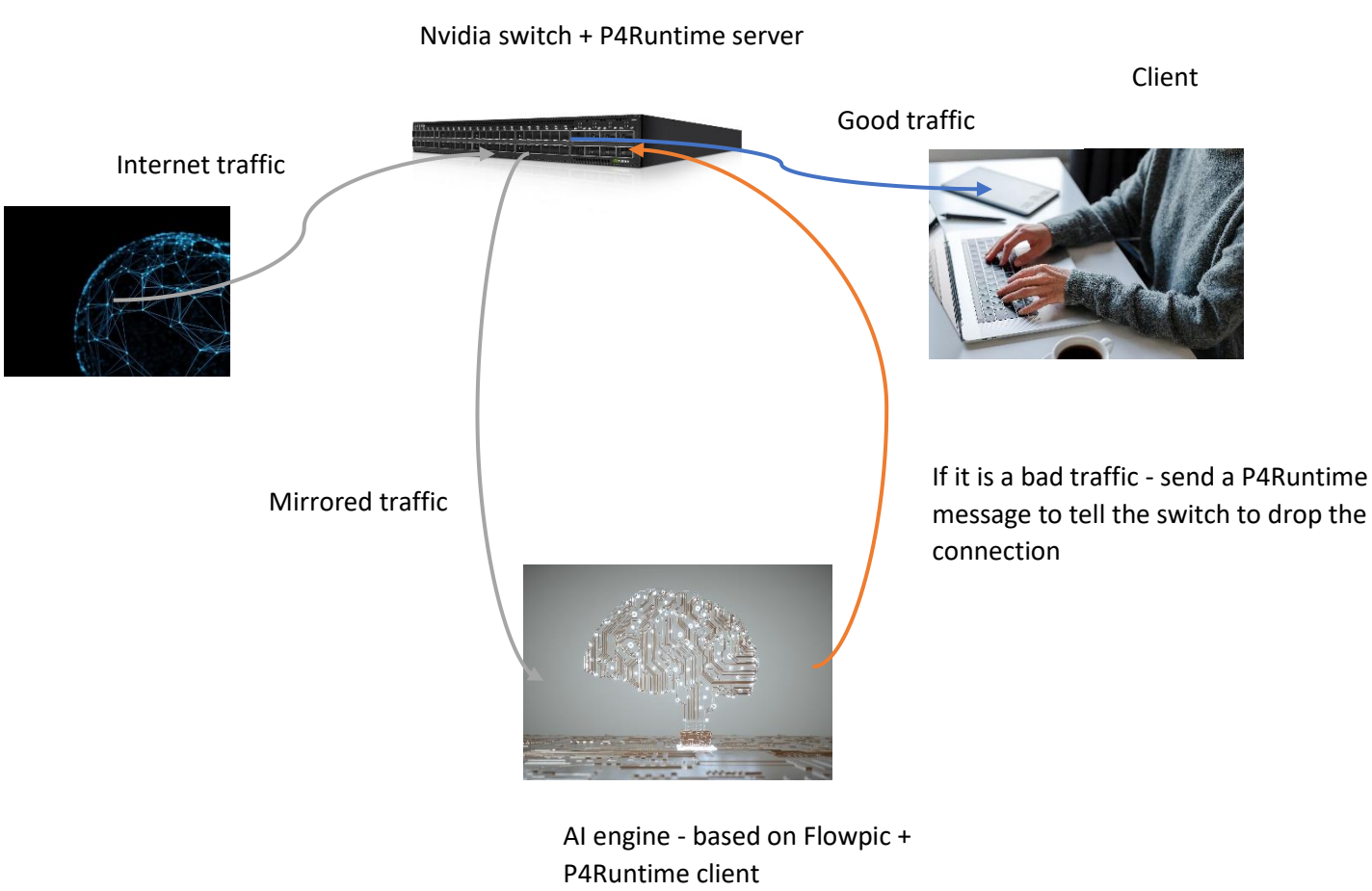
You will

- Build a live AI-based firewall
- Understand how the components are working. Among them are
    - Networking and Linux networking
    - convolutional neural network (CNN) and AI classifier
    - Wireshark
    - P4 (networking programming language)
    - HW networking (basic)
- Integrate those components into a working system
- Learn how to use your time efficiently
    - You will understand which components require a deep dive and which you can treat as black box
    - The investment in this project will be at the same level of investment as in other projects
    - You will enjoy the work of others in previous semesters
- Work closely with Nvidia employee that will help you to achieve your goals
    - Maybe a guide from Rafael will join also

The system that you will build will do the following:

1. Extract data at a high rate from the switch and deliver it to an AI engine
2. Classify the data with an AI algorithm
3. Send the classification to the switch: Tell the switch to block/forward this flow. This part will use the new P4Runtime engine
4. Drop/pass packets according to the classification. This is the goal of the firewall

The project's outcome will be probably one of the first AI-based firewalls that run in close to real-time.



AI Engine

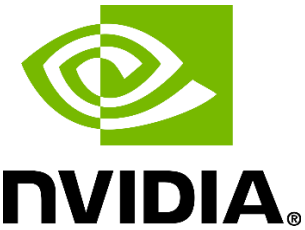
The AI is based on FlowPic, an AI algorithm that classifies flows based on the size of the packets and their arrival time. A basic CNN has developed a few semesters ago, but it is working on real traffic. You will need to improve it / build it again. When you have a good CNN, you will build a software component that extracts the data and sends it to the CNN.

If the detection tells you that this is “bad” traffic, then the software that you will build will send a request to the switch followed by the P4Runtime syntax and semantic.

Programmable Hardware and P4

Nvidia is developing a programable environment for its switches. This environment supports the P4 language. Programmable hardware is one of the most important verticals in the hardware industry.

P4



> "Programming Protocol-independent Packet Processors (P4) is a domain-specific language for network devices, specifying how data plane devices (switches, NICs, routers, filters, etc.) process packets." (p4.org)

Developing new hardware functionality is costly both in time and money. On the other side, keeping an ASIC-specific programmable language is expensive and not portable. Customers prefer that their engineers study a programming language that may be used in various ASIC vendors, instead of learning specific vendor programming tools. With P4 Nvidia can offer a portable software solution that can be developed fast. You, the students, will build a software system in which you will deploy a new engine with P4 - the P4Runtime.

P4Runtime

You will build a software program that will deploy and use Nvidia P4Runtime implementation.

> "The P4Runtime API is a control plane specification for controlling the data plane elements of a device defined or described by a P4 program." (p4.org)

The API used in this project for connecting the AI engine to the switch. When the AI will have a "bad" classification, meaning that it classifies traffic as bad, we will want to block the traffic. The software that uses the AI will send a P4Runtime request to the P4Runtime server. The P4Runtime server is located inside the switch. This request will ask to drop all the traffic related to the "bad" classification. The switch will parse this request and then will block only the related traffic.

Previous Work

You will enjoy previous work from previous semesters. Some parts of the system are built. As mentioned above, the accuracy of the classifier is in doubt as of today, so we will make it better. You will connect it to the switch with the P4Runtime engine. As written above, the goal of this project is to have a fully functional system. The students that will take this project will have a great introduction to the ML and network world, especially to the software-defined network world.

References

- Last projects
- https://p4.org/
- https://p4.org/p4-spec/p4runtime/main/P4Runtime-Spec.html
- https://talshapira.github.io/portfolio/flowpic/

Instructors: Idan Barnea (NVIDIA)

Eran Tavor (tavran@cs.technion.ac.il)