

NVIDIA P4 Runtime Smart classification

Abstract

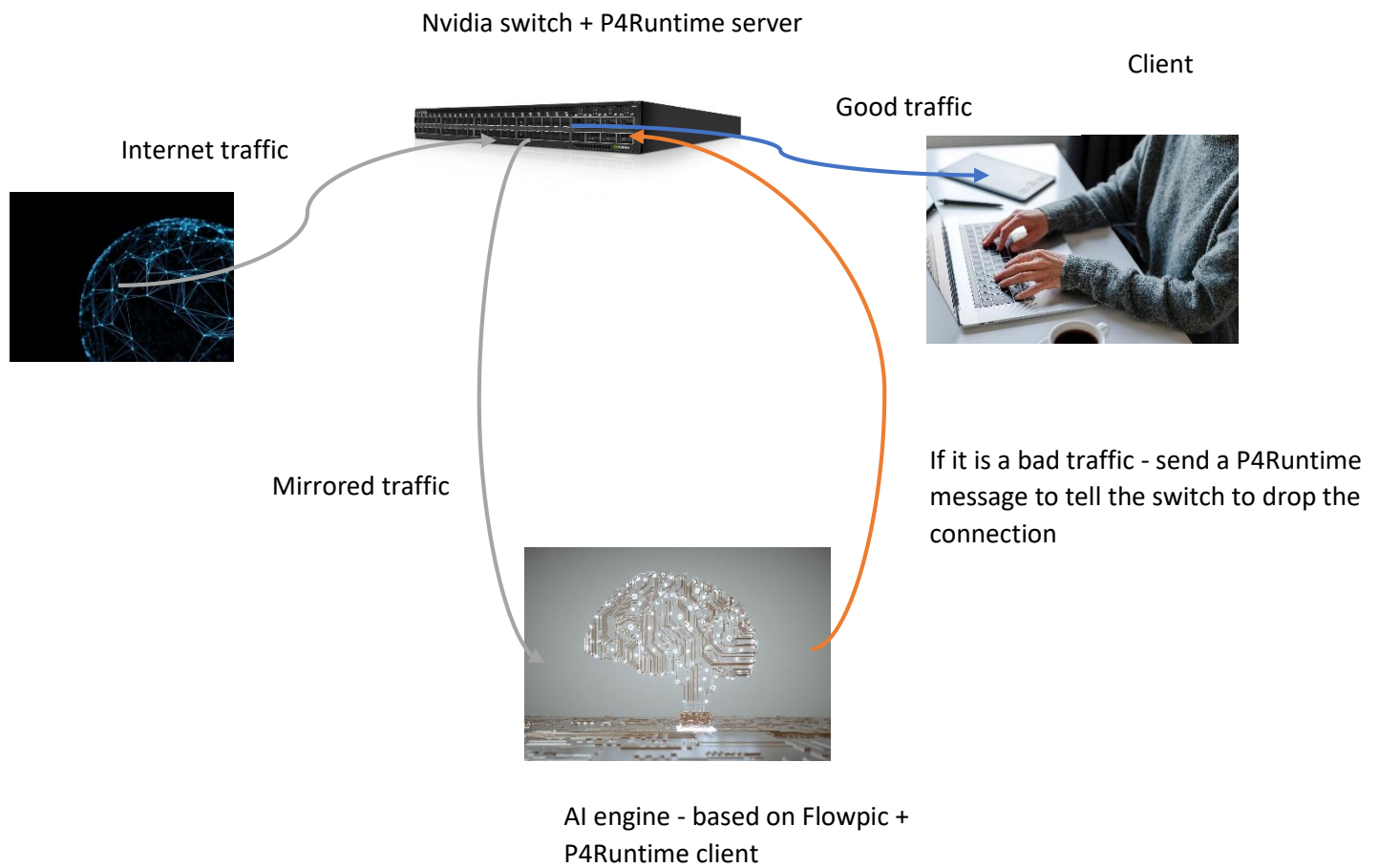
The goal of this project is to develop a live AI-based firewall.

In this project you will deploy the P4Runtime software package on a high-performance switch (25+ Gigabits per second) and use AI to create a live firewall.

The system that you will build will do the following:

1. Extract data at high rate from the switch and deliver it to an AI engine
2. Classify the data with an AI algorithm
3. Act according to the AI classification: add/delete entries to the tables in the switch. 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 real-time.



Programmable Hardware and P4

Nvidia is developing a programmable environment for its switches. This environment supports the P4 language. Programmable hardware, which is not a CPU, 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 will 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 - 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 is used in this project to connect 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, so you can focus on the implementation of the P4Runtime mechanism. 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 network world, especially to the software-defined network world.

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. You will extract the data and send it to the AI. You may use an implementation of the AI from previous semesters. 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. You will need to understand how FlowPic works in order to deliver the right data and to have the right decisions.

References

- Last semester project - Semi Real-Time Traffic Classifier on Nvidia Switch
- <https://p4.org/>
- <https://p4.org/p4-spec/p4runtime/main/P4Runtime-Spec.html>
- <https://talshapira.github.io/portfolio/flowpic/>

Guided by

NVIDIA: Matty Kadosh & Idan Barnea

Rafael: Aviel Glam

Technion: Eran Tavor

