**Questions and answers about Traffic Monitoring and Analysis**

Topics: **Traffic Monitoring and Analysis**

Name of the student: **Juan Pablo Royo Sales**

**Question 1:**

We have seen that in order to implement Traffic Sampling in Network Monitoring Analysis, we use different kinds of Data Structures that help us to implement efficient monitoring. What are the difference between implementing sampling with common Hash Tables and Bloom Filters?

**Answer to question 1:**

Bloom filters as well as regular Hash Tables uses **Hash Functions** in order to determine if some flow has already seen during the monitoring, but the main difference is Space optimization of Bloom filters against Hash Tables. Basically it uses small memory to allocate the data. This relies on the fact that Bloom filter uses bit arrays and multiple Hash functions to map the element membership to some position of this array.

It is a data structure that only answer about set membership of elements. On the other side Hash tables stores not only the hash value of the computed element but also the element itself to check set membership. This is the reason why Hash Tables are more space consuming than Bloom filters.

One of the disadvantages of using Bloom filters is that it has some rate of false positive but it is probe that if we use multiple hash function from a Universal Hash Function Family, there is a low probability of false positive. Another difference is Bloom filters does not allow deletion or elements.

**Question 2:**

Explain if the following statement is true or false and justify your answer

“In Traffic Classification techniques, it is enough to do a Port-based classification in order to accurately classify the different flow of the applications involve”

**Answer to question 2:**

It is false, because nowadays it is highly inaccurate to classify flows doing port-based inspection and it is well known that it can be easily by-passed by P2P applications for example. This implies that other techniques are required in order to do proper classification of the traffic in the SDN controller. One of those techniques that we have seen in class and it is extremely powerful is to use Statistic based classification with a combination of ML models and DPI.