**Title**: *Observing and understanding network behaviour: is your provider respecting the SLA contract*

**Advisors*:***

*Dr. Tania Cerquitelli, Politecnico di Torino, Italy*

*Dr. Genoveva Vargas-Solar, CNRS, LIG-LAFMIA, France*

**Context**:

The one of the major challenges for studying Internet performance is how to collect a large number of reliable network measurements from a sufficiently large number of locations in the network. To study and monitor the performance of Internet access links, designed Neubot, that runs on the user’s computer and periodically monitors the performance of its connection to the Internet. The results are collected on the user computer and made publicly available on a central server. A network of servers around the world allows Neubot to effectively test the performance of the clients’ broadband access network, by connecting them to the closest M-Lab server available. Since Neubot is performing a large amount of measurements each day, the problem of measuring the network is a problem of managing the available data for storage, querying and analysis purposes.

In fact, benchmarking Internet access link performance cannot be achieved by merely running a single speed test. Speed varies with time and it is affected by a number of confounding factors (i.e., home network cross-traffic, end-host configuration, wireless connection quality) that must be isolated as much as possible by proper data processing and analysis. This project will contribute to address this challenge.

**Objective**: Develop a data mining solution for modelling and discovering the behaviour of the network along time and the factors that determine.

This analysis will help to understand the use of the network by users subscribed to different providers in the world and observe how frequent the service is delivered with the quality stated in their contracts.

Observations are archived as collection of tuples representing an observation of the upload and download speed done on client machine running on a specific platform (windows, linux) by a user, at a given time in a given city using a specific IP.

Given a segmented collection of these archives, the idea is to:

* Look for the IP with the same properties given the information of an IP in each tuple.
* Aggregate observations using different time granularities, with respect to provider, city, country and cluster observations and define which are the attributes that characterize a cluster (to see the conditions in which for example a specific speed happens).