

MSc Project Proposal: Load Prediction and Forecasting in Private Cloud Systems

Background and context

The amount of energy needed to run the data centers in the world was an astonishing 196 TWh during 2020. This figure is expected to grow to between 321 TWh and 2000 TWh by 2030. In 2030, data centers are expected to account for approximately 8% of the global use of electric energy. This means that the environmental and economical impact of the electricity consumption of data centers will be extremely significant in the future.

One approach to reduce the energy consumption is to adjust the number of running computers to the current load in the system. To do so, we need to have good models for predicting the future incoming load / traffic to the system (e.g., the number of service requests of different types).

However, predicting the future incoming load to a cloud system is a non-trivial task. In the Green Clouds project, we try to exploit knowledge about the application behavior in private cloud environments to do better predictions of the future incoming traffic to the system.

Description

In this thesis work, we will develop machine learning models for predicting and forecasting the future incoming load / traffic in private cloud systems. Input to the models will typically be historical information about the incoming network traffic (i.e., service requests), etc. The output from the models should be, e.g., predictions of incoming traffic, both in the short-term and the long-term perspectives.

Examples of research questions that we would like to be addressed are:

- Which machine learning models are suitable for predicting future incoming traffic?
- Which historical data is suitable and useful in this context?
- How far into the future can we do reliable predictions of the traffic patterns?
- How can we exploit application knowledge?

The work will be performed within the Green Clouds research project, "*Green Clouds - Load prediction and optimization in private cloud systems*," and in collaboration with Ericsson AB.

Technical requirements / suitable background

We are looking for students with the following competences:

- Good knowledge about AI / Machine learning
- General knowledge about operating systems and cloud systems
- Excellent programming skills in Python, scikit learn, Pandas, PyTorch, etc.
- Good writing skills and fluency in English

Contact persons

Prof. Håkan Grahn, Hakan.Grahn@bth.se

Assoc. Prof. Anton Borg, Anton.Borg@bth.se