Infrastructure for handling high volume network traces

Supervisor: Prof. E. Biersack, erbi@eurecom.fr

The Internet Research Group at Eurecom, Sophia Antipolis, France (http://www.eurecom.fr/~btroup/) proposes a Masters Thesis (Diplomarbeit) in the area of network tomography.

Project Description:
One way to study performance issues in networking is to install a probe inside the network and capture all the packets (usually only the headers of the packets, e.g. the first 60 bytes) that pass by. As the network speeds go up, the volume of these *packet traces* keeps growing. Today, it is quite common that a trace taken at a high speed link is hundreds of Gigabytes or even Terabytes in size. After the trace has been captured, it needs to be analyzed according to certain criteria.

The aim of this MSc thesis is to investigate *how such a large amount of data can be handled efficiently*. In our group, we have previously adopted a database approach where the entire trace is put into a PostgreSQL database. However, for traces larger than a few tens of Gigabytes, the time to load the trace becomes too high.

We now propose to investigate new promising directions such as the use of
- **Dataseries**, which is an open source, generic trace format developed at HP Labs [3]. It provides streaming access to database-like tables and was developed specially for handling large volumes of data. (see http://tesla.hpl.hp.com/opensource/)
- **Pig** is a platform for analyzing large data sets that runs in a distributed fashion on a cluster. Programs are complied into Map/Reduce jobs and executed using Hadoop. (see http://wiki.apache.org/pig/)

Requirements:
- Good Knowledge of *JAVA.*
- Knowledge of Databases and Computer Networking

Reference:
M. Siekkinen, V. Goebel, and E. W. Biersack. **Object-Relational DBMS for Packet-Level Traffic Analysis: Case Study on Performance Optimization**. In *Proceedings of IEEE/IFIP Workshop on End-to-End Monitoring Techniques and Services*, April 2006.
("http://www.eurecom.fr/~btroup/BPublished/siekkinen06intrabase.pdf)