Universidade do Minho

MAP-i Thesis Proposal

Title: Internet forecasting using data mining techniques

Brief description: Internet forecasting is an important issue for any medium/large network provider that has received few attention from the Computer Networks community. The aim of this PhD is to study novel solutions for crucial TCP/IP communication tasks, such as:

- Internet TCP/IP traffic forecasting, given a network with several nodes and links;
- TCP/IP end-to-end throughputs predictions; and
- Network Coordinate estimation - Peer-to-Peer applications (which represent nowadays around 70% of traffic on networks) have much to gain from realistic Network Coordinates (similar to GPS but that allow to calculate network proximity). This task will use network delay and throughput datasets to calculate and forecast realistic network coordinates.

These are challenging tasks, in particular when few information is available. The intention is to develop advanced data mining techniques (e.g. Neural Networks or Support Vector Machines) that can provide timely and accurate forecasts. This research is expected to enhance resource management, improve anomaly detection (e.g. spread of virus), increase the quality of service and the behavior adaptation (e.g. peer selection) of computer networks.

Additional notes:

- Although data collection often requires a huge effort/time, real-world data is/will be available (e.g. UKERNA TCP/IP traffic - bytes/s of every link in the British academic network).
- Although not mandatory, we would like to encourage the candidate to visit the University College London (UCL), during a period of 3 to 6 months, to work directly under the supervision of Miguel Rio.
- This thesis is integrated in the R&D project PTDC/EIA/64541/2006 funded by FCT (http://www.fct.mctes.pt), hence there are funds for computer equipment and scientific conference traveling expenses. The project team includes also 2 PhD researchers from CCTC (Pedro Sousa, Miguel Rocha). This work can be supported with a BI grant (745 euros/month).

Pre-requisites: Solid background in programming (e.g. preferably Java) and knowledge within one or more of the following areas: Artificial Intelligence/Data Mining; Computer Networks and Operating Systems. Although not mandatory, candidatures should select the MAP-i course **Adaptive Business Intelligence**, which includes a module of forecasting.

Supervisors:

- **Paulo Cortez**, pcortez@dsi.uminho.pt, Department of Information Systems/Algoritmi R&D Centre, Univ. Minho, Guimarães, Portugal (http://www.dsi.uminho.pt/~pcortez)
- **Miguel Rio**, Department of Electrical Engineering, University College London, UK (http://www.ee.ucl.ac.uk/~mrio/)

Bibliography (previous work performed in this area):

- P. Cortez, M. Rio, M. Rocha and P. Sousa. Internet Traffic Forecasting using Neural Networks. In Proceedings of the 2006 IEEE International Joint Conference on Neural Networks, pp. 4942-4949, Vancouver, Canada, July, 2006 (http://hdl.handle.net/1822/6581).
- P. Cortez, M. Rio, P. Sousa and M. Rocha. Topology Aware Internet Traffic Forecasting using
- Neural Networks. In J. de Sá et al. (Eds.), Artificial Neural Networks - ICANN, 17th International Conference, Lecture Notes in Computer Science 4669, pp. 445-454, Porto, Portugal, September, 2007, Springer (http://hdl.handle.net/1822/7634).
- E. Mykoniati, R. Landa, S. Spirou, R. G. Clegg, L. Latif, D. Griffin and M. Rio, Scalable peer-to-peer streaming for live entertainment content, to appear in IEEE Communications.