Scheduling in data gathering networks

Description for the general public

Gathering data from remote processors is an important step of many contemporary applications. Indeed, complex computations need to be executed in distributed systems like grids, clouds or volunteer computing platforms in order to provide sufficient computing resources. In some cases the obtained results may be stored in a distributed file system, but very often they have to be collected together on a single machine. Data gathering wireless sensor networks are widely used for monitoring the environment, military purposes, surveillance, patient monitoring and in other areas. Internet services gather a lot of information about the users, their actions, habits and inclinations. Therefore, data gathering is an important and interesting subject of research.

A *star* data gathering network consists of a set of *data nodes* which obtain some datasets by computations or sensing the environment and a *base station* to which the data should be sent for further processing. A *tree* network additionally contains *intermediate nodes* responsible for transferring the data between data nodes and the base station. The intermediate nodes may be able to preprocess or compress the data. Scheduling in data gathering networks consists in organizing data transfer and processing, taking into account various parameters of the network nodes, so that the whole process is as efficient as possible. Constructing good scheduling algorithms allows to gather data in shorter time, at smaller cost and lower energy usage.

In this project a group of data gathering scheduling problems will be analyzed in star and tree networks. Our special interest lies in the following problems:

- minimizing data gathering time in a network with limited base station memory;
- minimizing data gathering time in a network with dataset release times;
- minimizing the maximum lateness in a network with dataset due dates.

The first goal is to determine the computational complexity of these problems. We conjecture that they are NP-complete, which means that fast algorithms constructing the optimum solutions for these problems probably do not exist. We also plan to identify special cases in which the problems become solvable in polynomial time. As exact algorithms solving NP-complete problems are too slow to use in practice, the next goal is to design efficient heuristics constructing good suboptimal schedules. We plan to start with implementing several simple algorithms and test them in a series of computational experiments under various combinations of network parameters. In this way we will find out what simple algorithms perform well in what situations and what schedule structures are advantageous. This knowledge will be used to construct advanced heuristics, which will be then analyzed both experimentally and theoretically. If possible, guarantees on the quality of the obtained solutions will be given. Afterwards tree data gathering networks will be studied. As the tree networks contain additional nodes in comparison to star networks, the scheduling problems become even more difficult. We will generalize the heuristics designed for star networks and combine them with different approaches suited especially for tree networks. The performance of the algorithms will be again analyzed theoretically and in computational experiments.