R

Resource Allocation in Wireless Networks

Dimitrios Katsaros
Aristotle University, Greece

Gökhan Yavas
Bilkent University, Turkey

Alexandros Nanopoulos
Aristotle University, Greece

Murat Karakaya
Bilkent University, Turkey

Özgür Ulusoy
Bilkent University, Turkey

Yannis Manolopoulos
Aristotle University, Greece

INTRODUCTION

During the past years, we have witnessed an explosive growth in our capabilities to both generate and collect data. Advances in scientific data collection, the computerization of many businesses, and the recording (logging) of clients' accesses to networked resources have generated a vast amount of data. Various data mining techniques have been proposed and widely employed to discover valid, novel and potentially useful patterns in these data.

Traditionally, the two primary goals of data mining tend to be *description* and *prediction*, although description is considered to be more important in practice. Recently though, it was realized that the prediction capabilities of the models constructed by the data mining process can be effectively used to address many problems related to the allocation of resources in networks. For instance, such models have been used to drive prefetching decisions in the World Wide Web (Nanopoulos, Katsaros, & Manolopoulos, 2003) or to schedule data broadcasts in wireless mobile networks (Saygin & Ulusoy, 2002). The intrinsic attribute of these environments is that the network records the characteristics, for example, movements, data preferences of its clients. Thus, it is possible to infer future client behaviors by mining the historical information, which has been recorded by the network.

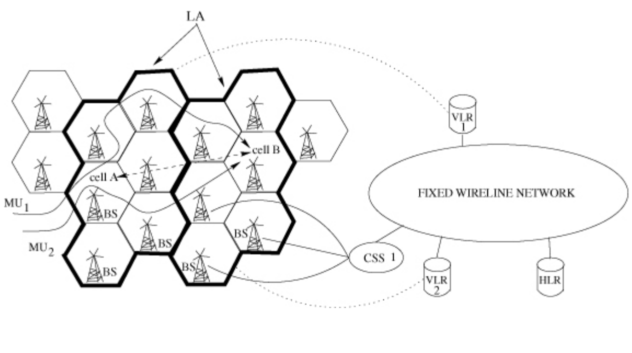
The present article will highlight the data mining techniques that have been developed to achieve efficient allocation of resources, for example bandwidth, to wireless mobile networks or the data mining methods that have been used in order to reduce the latency associated with the access of data by wireless mobile clients.

BACKGROUND

We consider a typical wireless Personal Communications Systems (PCS) (see Figure 1) with architecture similar to those used in *EIA/TIA IS*-41 and *GSM* standards. The PCS serves a geographical area, called *coverage area*, where mobile users (MU) can freely roam. The coverage area served by the PCS is partitioned into a number of non-overlapping regions, called *cells*. At the heart of the PCS lies a fixed backbone (wireline) network. A number of fixed hosts are connected to this network. Each cell is usually served by one *base station* (BS), which is connected to the fixed network and it is equipped with wireless transmission and receiving capability. We assume that each base station serves exactly one cell. MUs use radio channels to communicate with BSs and gain access to the fixed or wireless network. The BS is responsible for converting the network signaling traffic and data traffic to the radio interface for communication with the MU and also for transmitting paging messages to the MU. Finally, a *cell site switch* (CSS) will govern one or more base stations. CSS will provide access to the serving mobile network, will manage the radio resources and provide mobility management control functions (for example, location update).

The coverage area consists of a number of *location areas* (LA). Each location area consists of one or more cells. The MU can freely roam inside a location area

Figure 1. Architecture of a wireless PCS



without notifying the system about its position. Whenever it moves to a new location area it must update its position, reporting the location area it entered. This procedure is called *location update*. This is done as follows: each mobile user is assigned to one database, the *home location register* (HLR) (one for each PCS network), which maintains the *profile* information regarding the user, such as authentication, access rights, billing, position, and etcetera. Each location area is associated to one *visitor location register* (VLR), which stores the profile of the MUs currently residing in its respective location area. We assume that each VLR is associated to one location area and vice-versa. The search for mobile clients is performed by broadcasting *paging messages* to the cells, where the clients might have moved, until the client is located or the whole coverage area is searched.

The identity of the cell is continuously being broadcast by the cell's BS, thus the terminal is aware of the cell it resides. If each mobile terminal records the sequence of cells it visits and communicates them back to the network every time it reports a location area crossing, then the network can have accurate information for the *mobile user trajectories* inside the coverage region.

The concept of resource allocation in wireless mobile environments covers aspects of both network and data management issues. With respect to network management, the issue of *dynamic bandwidth allocation* is of particular importance. Instead of granting a fixed frequency spectrum to each cell, irrespective of the number and needs of the clients residing therein, the allocated spectrum varies according to the clients' demands. This necessitates prediction of both future clients' movements and future data needs. The issue of client movement prediction is also related to the order according to which the paging messages should be broadcast to the cells, so as to guarantee minimum energy consumption and, at the same time, fast client location determination.

With respect to the data management issues, a prominent problem is that of reducing the average latency experienced by the clients while retrieving data from the underline infrastructure network or the wireless network. This problem is tightly related to caching data at various places of the network. The caches can be air-located, that is, data broadcasting (Karakaya & Ulusoy, 2001), or relocated to specific base stations. Thus, we need to identify data that will be requested together by the same client or group of clients, so as to broadcast them during short time intervals. In addition, we need to deduce future client movements to cells so as to "*push-cache*" data to the base stations serving these cells (Hadjiefthymiades & Merakos, 2003; Kyriakakos et al., 2003).

DATA MINING AT THE SERVICE OF THE NETWORK

We will present the most important research areas where data mining methodologies are employed to improve the efficiency of wireless mobile networks. The first area is the mobile user location prediction and aims at deducing future client movements. Location prediction is important for both bandwidth allocation and data placement. The second area is the data broadcast schedule creation and aims at recognizing groups of data items that are likely to be requested together or during small time intervals, so as to place them "closely" in the broadcast program.

LOCATION PREDICTION

The issue of predicting future mobile client positions has received considerable attention (e.g., Aljadhai & Znati, 2001; Liang & Haas, 2003) in the wireless mobile networks research community. The focus of these efforts is the determination of the position of a mobile, given some information about its velocity and direction. Though most (if not all) of these works make unrealistic assumptions about the distribution of the velocity and direction of the mobile terminals. Only recently, data mining techniques have been employed in order to predict future trajectories of the mobiles.

Data mining techniques capitalize on the simple observation that *the movement of people consists of random movements and regular movements* and the majority of mobile users has some regular daily (hourly, weekly,…) movement patterns and follow these patterns more or less every day.

Several efforts targeting at location prediction exploited this regularity. The purpose of all these efforts