Database Support for M-Commerce

Hong Va Leong
The Hong Kong Polytechnic University, Hong Kong

INTRODUCTION

M-commerce applications have evolved out of e-commerce applications, riding on the rapid advancement in mobile communication technologies in the past decade. The diffusion of applications on the Internet into the mobile computing environment has taken an accelerating pace ever since. Virtually all e-commerce and m-commerce applications rely on the provision of information retrieval and processing capability. In this regard, database systems serve as the information source and repository for these applications, backed by efficient indexing mechanism. Bean (2003) gave a good report on supporting Web-based e-commerce with XML, which could be easily extended to m-commerce. An m-commerce framework, based on JINI/XML and a workflow engine, was also defined by Shih and Shim (2002). Customers can receive m-commerce services through the use of mobile devices, such as pocket PCs, PDAs, or even smart phones. These mobile devices together with their users are often modeled as *mobile clients*. Central to supporting m-commerce applications are three types of entities: mobile device, mobile communication, and database. In particular, we are more interested in providing efficient access mechanisms to mobile-client-enabled database servers, which are often called *mobile databases*. Mobile databases contain the core information to support the underlying m-commerce applications, while the use of mobile devices serves for the hardware platform, with mobile communication providing the necessary connection between mobile databases and mobile devices for interfacing with real users or customers.

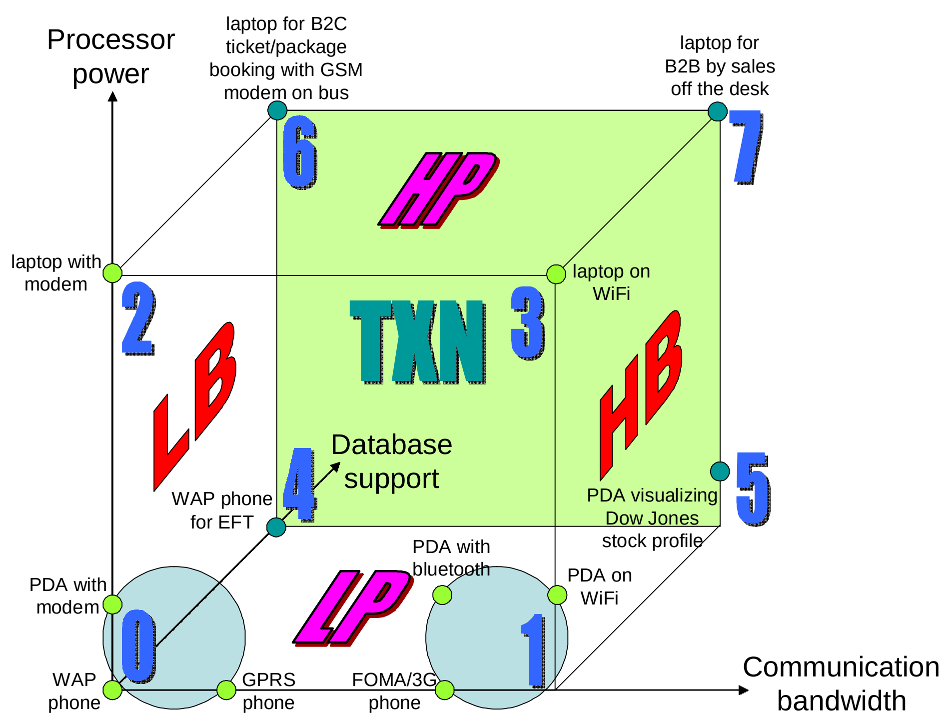
The two major types of data access requirements for a mobile database are data dissemination and dedicated data access. In a mobile environment, data dissemination is preferred, since it can serve a large client population in utilizing the high-bandwidth downlink channel to broadcast information of common interest, such as stock quotations, traffic conditions, special events, or the number of available seats at a performance. On the other hand, dedicated data access is conveyed through uplink channels with limited bandwidth. To disseminate database items effectively, the selected set of hot database items can be scheduled as a broadcast disk (Acharya, Alonso, Franklin & Zdonik, 1995) or the "air storage" (Leong & Si, 1997). The sequence of items to be broadcast is referred to as a broadcast program (Acharya et al., 1995). Proper indexes can be built to facilitate access to database items in the broadcast (Imielinski & Badrinath, 1994). Redundancy can be included to combat the unreliability of wireless communication (Tan & Ooi, 1998).

For dedicated data access, queries and updates to databases are transmitted from the client to the server. Since quite a significant portion of m-commerce services involve location-dependent queries, performance of location-dependent queries should be improved (Madria, Bhargava, Pitoura & Kumar, 2000), dictating effective location maintenance with moving object databases (Wolfson, Sistla, Xu, Zhou & Chamberlain, 1999). This enables efficient execution of common queries such as K-nearest neighbor search. As database systems become more powerful, they can help to process continuous queries (Prabhakar, Xia, Kalashnikov, Aref & Hambrusch, 2002), which are executed continuously, returning possibly location-dependent results.

BACKGROUND

The three fundamental elements in the provision of an m-commerce application, namely, mobile device, mobile communication and database support, can be considered orthogonal. First, the variety of mobile devices differs vastly in computational power, ease of programming, interoperability of operating environments, and support for auxiliary devices. Some mobile clients may be based on a high-end laptop of capacity close to that of a desktop. Other mobile clients may be based on a low-end PDA, or cellular WAP or GPRS phone. Second, mobile communication offers varying bandwidth and reliability, based on low-bandwidth and unreliable GSM connection, medium-bandwidth GPRS or Bluetooth connection, or high-bandwidth 802.11b or 3G/CDMA2000 connection with higher reliability. Third, the database may be as primitive as a file system or simple relational database like MS Access, or as complex as a high-performance Oracle database with transactional support and information retrieval ability. *Transactions* are useful in ensuring a sequence of database operations to be executed consistently without any interference experienced. This leads to a "cube"-like taxonomy as shown in Figure 1.

Figure 1. Taxonomy on m-commerce support



In Figure 1, the taxonomy for m-commerce support is displayed. Planes **LP** and **HP** represent the low computing power equipment and high computing power equipment respectively, whereas planes **LB** and **HB** reflect the availability of low- and high-communication bandwidth. With the availability of transactions in the **TXN** plane, this gives rise to eight different regions, 0 to 7.

Region 0 represents the support of standard file or simple database access from PDA, connecting through low-speed modem or WAP phones. Processing is basically performed at the server, since it is too expensive for clients to support complex mechanism. To combat the low-bandwidth problem, information distillation/extraction (Cowie & Lehnert, 1996) may be performed to reduce the amount of information transmitted. Simple client/server data access paradigm suffices. Region 1 assumes an improved wireless network, with CDMA2000 or WiFi. As a result, data access is more effective, and conventional client/server data processing techniques can be adopted in a rather straightforward manner.

Region 2 corresponds to a mobile client with higher computational power. Information transmitted can be transcoded to reduce the bandwidth consumption. Interactive and intelligent mechanisms such as multi-resolution browsing (Yau, Leong & Si, 2001) can be employed. Data items are cached to combat the low-communication bandwidth, unreliable communication, and frequent disconnection. Research work addressing this issue has drawn much attention in the past, as pioneered by the work on the Coda file system (Mummert, Ebling & Satyanarayanan, 1995). In Coda, frequently accessed files are cached by the clients, and updates made during client disconnection are reintegrated with the file system upon reconnection. Caching in an object-oriented database was studied by Chan, Leong, Si and Wong (1999). Configurations in Region 3 allow easy access to data from the server. With ample bandwidth and processing power, prefetching of data items is done to prepare for potential network disconnection (Jing, Helal & Elmagarmid, 1999). Numerous research works on mobile data access have been conducted with respect to Regions 2 and 3.

Plane **TXN** represents the transactional equivalence of the above-mentioned four regions. Regions 4 and 5 involve the use of PDAs or phones to access databases in a transactional manner. Owing to the low device capability, the only effective mechanism is to execute the transaction at the server. Clients only implement the user interface, supplying the required data and displaying the result sets. Information distillation could be needed for the low-bandwidth configurations in Region 4. The use of a proxy server helps to simplify the client/server design, since the proxy will be responsible for contacting different