University of Nebraska at Omaha
DigitalCommons@UNO

Student Work

8-1-2004

Exploring a Quality of Service (QoS) Mechanism to Enhance Multimedia Database Query Processing in Wireless Mobile Environments.

Yanpu Zhang

Follow this and additional works at: https://digitalcommons.unomaha.edu/studentwork

Recommended Citation

Zhang, Yanpu, "Exploring a Quality of Service (QoS) Mechanism to Enhance Multimedia Database Query Processing in Wireless Mobile Environments." (2004). *Student Work*. 3573.
https://digitalcommons.unomaha.edu/studentwork/3573

This Thesis is brought to you for free and open access by DigitalCommons@UNO. It has been accepted for inclusion in Student Work by an authorized administrator of DigitalCommons@UNO. For more information, please contact unodigitalcommons@unomaha.edu.

Exploring a Quality of Service (QoS) Mechanism to Enhance Multimedia Database Query Processing in Wireless Mobile Environments

A Thesis

Presented to the Department of Computer Science

And the Faculty of the Graduate College

University of Nebraska

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

University of Nebraska at Omaha

By

Yanpu Zhang

August 2004

UMI Number: EP74771

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI
Dissertation Publishing

UMI EP74771

Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code

ProQuest

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

THESIS ACCEPTANCE

Acceptance for the faculty of the Graduate College, University of Nebraska, in partial fulfillment of the requirements for the degree (Master of Science), University of Nebraska at Omaha.

Committee

Name	Signature
Dr. Zhengxin Chen	Zhengxin Chen
Dr. Jong-Hoon Youn	Jon Youn
Dr. Zhenyuan Wang	Wang Zhenyuan

Chairperson Zhengxin Chen

Date 8/4/2004

Exploring a Quality of Service (QoS) Mechanism to Enhance Multimedia Database Query Processing Under Wireless Mobile Environments

Yanpu Zhang, M.S.

University of Nebraska at Omaha, 2004

Advisor: Dr. Zhengxin Chen

Abstract:

Among the challenges of multimedia computing and mobile computing, a mechanism for data retrieval in multimedia databases under wireless mobile environments seems to be the most difficult issue. The problem is that sizes of images in a multimedia DBMS queried by mobile clients through wireless networks are different and unpredictable. Current Quality of Service (QoS) framework has no answer for it because all the QoS principles are based on users' pre-requirements. However, the issue is that in multimedia applications, it is difficult to know the size of targeted retrieval object. There should be new mechanisms of QoS to participate in query processing and provide an efficient theme around which mobile multimedia database applications can be practicably realized. In this thesis we focus on extending QoS management in wireless mobile environments to specify a range of acceptable QoS for multimedia query processing, rather than trying either to guarantee specific values or to stop the querying.

Through the investigation of current research approaches, we conclude that the statistical or empirical resource utilizations in query processing are the dominant methods to solve the problems. All proposals choose stopping query if the required QoS conditions can not meet the related statistical or empirical resources utilizations.

To address QoS in mobile multimedia DBMS issues, we explore an approach to execute query processing based on real time QoS conditions all coming from client, network, and server. We propose a QoS-based matrix to support query processing of object-relational multimedia databases in the context of wireless mobile environments. The proposed QoS-based Querying Processing Precision Matrix (QQPPM) is based on (1) real-time QoS conditions in wireless networks; (2) multimedia database's object properties; and (3) mobile client-site data processing capability.

We study related technologies as the foundations to support multimedia query processing in wireless mobile environments. Moreover, we conduct OPNET simulations, and the results indicate that our assumption is reasonable and practicable.

TABLE OF CONTENTS

LIST OF FIGURES/TABLES

Figure Number/Table	Title of Figure/Table	Page

1. INTRODUCTION

There are two trends that can be recognized in our information-driven society [UMT00]. On the one hand, users have an increasing need to gain information in an efficient manner. The amount of information accessible through the Internet is growing exponentially. Digitization of information from paper, film and tape archives and creation of new multimedia information based on image, video, voice and text all exist today. On the other hand, users have an increasing demand to do wireless mobile communications. According to [ITU02], mobile subscribers are over 1 billion worldwide, which has overtaken fixed users since 2002. Moreover, on November 10, 2003, the U.S. Federal Communications Commission issued the rules allowing consumers to disconnect their home phones and move their existing numbers to their cell phones. These rules allow more people go solely wireless mobile world. Now more and more digital multimedia data is being captured and stored. As mobile data and multimedia services continue to grow more rapidly than mobile voice services, they will overtake mobile voice services to become the dominant mobile services in the near future [CZ04]. An explosion in the combination of digital media, database and mobile Internet technologies is becoming a hot research topic [Lu99] [Jam03] [RG03] [BK03]. Virtually every computer system with which we interact in our e-Business, mobile commerce, or mobile entertainment should access database and do data transaction through mobile communication environments. In the next decades, we will become increasingly dependent on the correctness and efficiency of combined these technologies.

1.1 Combination of Multimedia Database with Wireless Networking

In today's telecommunications industry, mobile is the largest telecommunication network in many countries [ITU02]. The third generation (3G) mobile systems have been launched in the market, and the new generation high-speed cellular networks are expected to support global mobile multimedia applications and the Internet service [Jam03]. There are four characteristics in 3G platforms: broadband capacity, IP-based service, QoS support, and interoperability [CZ04]. All of these are the essential elements for the development of wireless mobile Multimedia Database Management Systems (MM-DBMSs). 3G standards are meant to become the vehicle for Internet-based multimedia and other data services to go wireless [HRM03]. As the wireless market moves towards multimedia, mobile users can fulfill their demands for all types of information in a format appropriate to each type of terminal. Digital video and still picture are today's dominant digital commodity items. Their data become major objects in the wireless mobile multimedia databases. As said in [Sau02], like the early Web, wireless mobile multimedia communications boast seemingly limitless potential market.

Meanwhile, MM-DBMSs are becoming ideal repositories for next generation multimedia applications, especially in inter-networked LAN/WAN architectures [Shi02]. By combining features from document-imaging, information-retrieval, and hierarchical storage systems, MM-DBMSs are satisfied the database requirements of a host of multimedia applications. According to Evans Data's latest biannual database developer survey [Doc03], almost half of respondents are developing database applications for mobile devices or at least plan to do so within the next year.

1.2 Impact of MM-DBMS Applications on Current Wireless Networks

However, as pointed by the work discussed in [Shi02], among the new challenges of multimedia computing and related technologies, a mechanism for content-based information retrieval in multimedia databases seems to be the most difficult issue. According to [RG03], on mobile issue, users are connected through a wireless link which bandwidth is 10 times less than Ethernet and 100 times less than ATM networks. Besides, user's locations constantly change, and mobile computers have a limited battery life. On the point of multimedia database issue, traditional DBMSs have concentrated on tables that contain a large number of tuples, each of which is relatively small. Once multimedia objects such as images, sound clips, and videos are stored in a database, individual objects of very large size have to be handled efficiently. At the same time compression techniques must be carefully integrated into the mobile MM-DBMS environment. Up to now, the fundamental technologies that are specialized for wireless mobile multimedia environments are not mature in object-oriented, object-relational as well as relational databases [Hil02] [RG03] [Wat04]. In this thesis, we focus on the studying of one of critical problems in the application of mobile multimedia databases. It is how to convenient, effective, efficient, and economical access the multimedia databases through the wireless networks. In particular, we explore an approach to integrate the new query processing technology into the MM-DBMSs in wireless network environments. This is a challenge approach because, up to now, retrieval of multimedia objects in wireless mobile environment has been considered as a difficult problem [RG03] [BK03] [MK03].

1.3 Explosion of Solutions based on Quality of Services

A key observation for above-mentioned issue is that should have extended mechanisms of Quality of Service (QoS). According to [Kaz02] [BK03], considering QoS capably participating in the query processing in mobile MM-DBMSs are a possible solution. There are two obvious problems [CMP01] [EGP02] [Kaz02] [Cgh03] should be solved. One is that a mobile communication session may be dropped if the destination cell has no enough bandwidth to support the connection. The other is that multimedia is time-dependent. Considering MM-DBMS query processing in wireless mobile environments, a satisfied query should be judged by both the response time and the quality of retrieved results. That means, to support mobile multimedia retrieval where a user need various QoS levels for different applications, the MM-DBMSs should provide specified QoS levels and efficient QoS management services correspondently.

Up to now, QoS for MM-DBMS in wireless mobile environments has been an open issue [Bar99] [UMT00] [BDW02] [Kaz02] [MT02] [RG03] [Wat04]. As regard to multimedia terminals, there are a lot of efforts needed to improve the performance. There remains little consensus on how QoS support should be provided, especially for supporting efficient query processing for the wireless mobile multimedia with the limitation of bandwidth, mobility, display space and energy. The researches in studying of the specification of QoS requirements for the multimedia databases applications in the wireless mobile environments are scarcely tangible. Even recently, [BK03] still concludes that work on service quality in the area of query processing is rare.

1.4 Contents of this Thesis

Based on the discussions given above, in this thesis, we address the above-mentioned issues, survey the fundamentals related query processing in MM-DBMS, analyze the key problems on the development of MM-DBMSs in wireless mobile environments, and further explore the enhanced multimedia database querying precision solution by extending current QoS mechanism. The rest of this paper is organized as follows. Chapter 2 studies the fundamentals of MM-DBMSs in wireless mobile environments, especially focusing on the topic of query processing. Chapter 3 discusses the research context that addresses the impacts of QoS on the application of wireless mobile MM-DBMSs. Based on previous topics, Chapter 4 proposes the framework of QoS from the theoretically and practically considerations, and discusses related issues in mobile environments. Chapter 5 gives a survey on the current researchers' approaches regarding QoS-based query processing. In Chapter 6, we present a new approach of querying precision solution which attempts to extend QoS mechanism to enhance the mobile MM-DBMSs main performance concerns like ***efficiency of query processing*** (i.e. how long it takes to answer a query), and ***effectiveness of data retrieval*** (i.e. how to solve the conflict of the user's needs and the restricted circumstances). Chapter 7 demonstrates the experimental results which come from OPNET simulations. Chapter 8 concludes this research and discusses the further proposals and related future works.

2. FUNDAMENTALS OF MOBILE MM-DBMS QUERY PROCESSING

Since a traditional DBMS does not support QoS multimedia object [Sti01], based on traditional DBMS, MM-DBMS should support QoS-sensitive multimedia data types in addition to providing all the facilities for DBMS functions. Furthermore, the crucial point is that mobile MM-DBMS should have the QoS-based capabilities to efficiently and effectively process the multimedia data in wireless mobile environments. In focus on the study of enhancing query processing mechanism efficiently and effectively for the mobile MM-DBMSs, we need to analyze all the related fundamental techniques.

The overview of the fundamentals are illustrated in the following figure 2.1, in which some techniques are available, some are to be modified, and some are to be developed. The detail explanations are described hereafter.

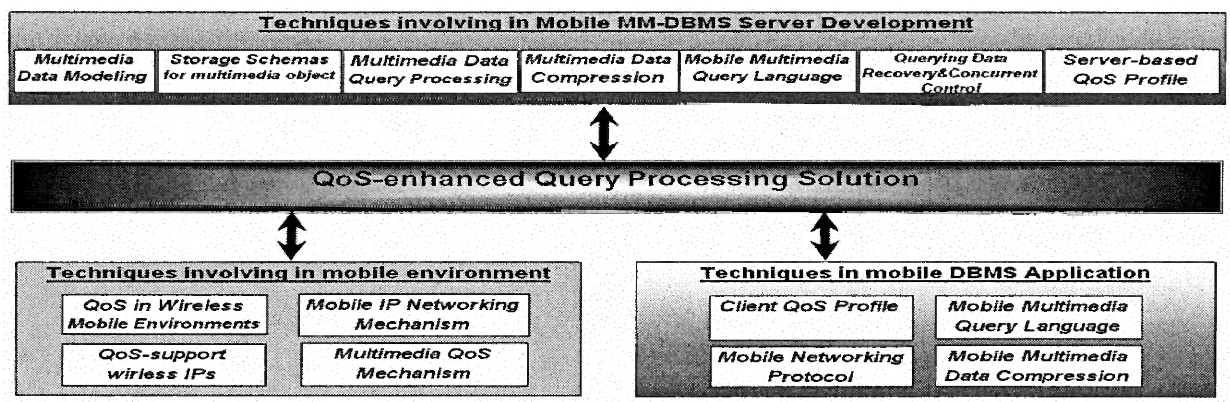

Figure 2.1: Technologies for QoS-based Query Processing in MM-DBMS

2.1 Techniques Involving Multimedia Database Server

In consideration of developing the querying and retrieval MM-DBMS in wireless mobile environments, the definition of data modeling, the special query processing and data retrieval methods, the storage schemas of multimedia objects, the data types of compression, the mobile multimedia data recovery mechanism, and Server-site multimedia object QoS profile should be necessarily involved in.

2.1.1 Data Modeling Techniques for Multimedia Information

The following multimedia concepts are widely implemented by database research community. (1) *multimedia object* describes image, audio, or video data; (2) *multimedia content* is actual image, audio, or video data; (3) *metadata* is format of multimedia content including object length, compression; (4) *method* can be performed on object such as store, deliver, extract metadata, and compress or convert image format; (5) *instance* consists of metadata, the multimedia content, and methods;.

As researches presented, the latest multimedia database technologies include the functions to process large, relatively unstructured media objects. These objects are with different formats in which large volumes of multimedia contents are loaded. Currently, multimedia applications can easily add multiple image, audio, and video columns, or mixed columns containing any of these types as objects, to existing and new relational tables (e.g. Oracle *inter*Media with Oracle 9i). Besides, the media objects can be easily imported and exported between database and their external resources.

2.1.2 Storage Schemas for Heterogeneous Multimedia Objects

The multimedia storage model includes the components of multimedia objects, which are either stored in the database under transaction control, or stored outside the database without transaction control. Multimedia contents stored outside the database can provide a mechanism for managing large, or new multimedia repositories that reside as flat files on erasable or read-only devices. The multimedia contents can be imported and exported at any time. Object metadata and methods are always stored in the database.

Taken the image as an example to explain more detail, the two-dimensional, static, digital images are stored as binary representations in most popular file formats and compression schemes. Image file formats provide a means for an application to store an image in a flat file. A digital image consists of characteristics of the image, and the image data itself (the digitized bits). The image data (pixels) can have varying depths (bits per pixel) depending on how the image was captured, and can be organized in various ways.

The basic characteristics of an image include the following:

- Size (height in scan lines and width in pixels)
- Number of bits per pixel in each of the colors sampled
- Compression type
- Content format

2.1.3 Multimedia Query Processing and Data Retrieval

Multimedia queries are diverse and fuzzy. They are diverse because the clients can specify queries in many different ways and in different types. They are fuzzy because clients know what they are looking for but cannot describe the objects exactly. Therefore, to meet the requirements of these characteristics, the current multimedia query processing mechanism provides the three procedures in practice. They are (1) search, (2) browsing, and (3) query refinement. Hereafter is the brief explanations.

- Search is the basic task of all DBMSs. In the context of MM-DBMSs, there are two types of search: search by specification and search by example. In search by specification, clients use key words and parameters to describe the main features or attributes of their query requirements. In search by example, clients specify queries in different media types or a combination of different media.
- Sometimes clients do not know exactly what they want but can recognize what they need. They need initial queries by browsing. There are three methods used to initiate browsing. The 1st method is to start with a vague query and user can navigate through the items based on the results. The 2nd method requires the information in the database be organized with some criteria so that the client can browse based on this organization. In the 3rd method, items randomly chosen from the database are presented and the client can use them as initial browsing. If the user does not find any interesting items, request other random trials.
- Since most initial multimedia queries are fuzzy. Query refinement is required. Based on client's feedback on the previous results, updated domain knowledge and client modified profile, query refinements are undergone iteratively.

There are four considerations regarding multimedia data retrieval.

- First, with the available QoS services and related media types that servers support, the queried objects are successfully presented in client site.
- Second, since the result queries may contain long audio segments, large images, or long videos, efficient extracting and presenting essential information for clients to browse and select are required.

- Third, the response time that is determined by both the network and database search should be efficiently short.
- Fourth, the query result should facilitate relevance feedback and query refinement as the result may not be final.

2.1.4 Multimedia Data Compression Technologies

Data compression is a process of converting original data into other compressed data that has a smaller size. Data compression is important in today's multimedia data communications because people hate waiting a long time for data transfers. There are many known methods for data compression. Although based on different ideas suitable for different types of data, they are all on the same principle of conducting specific algorithm to remove redundancy from original data in source file. Any nonrandom collection data has some structure that can be exploited to achieve a smaller representation of the data. Compressing data is done by changing its representation from inefficient (i.e., long) to efficient (short).

Taken digital image as an example, it is a rectangular array of dots, or picture elements, arranged in *m* rows and *n* columns. The expression $m \times n$ is called resolution of the image, and the dots are called pixels. There are several types of images.

- A bi-level (or monochromatic) image. Where pixels have one of two values, referred to as black and white. Each pixel in an image is represented by one bit, so this is the simplest type of image.
- A grayscale image. Pixels with *n* values, indicating 2^n shades of gray.

- A continuous-tone image. As a natural image, it is obtained by taking a photograph with a digital camera, or by scanning a photograph or a painting. It has many similar colors (or grayscales). When adjacent pixels differ, it is hard for eye to distinguish their colors. As a result, an image may contain areas with colors that seem to continuous as eye moves along.
- A discrete-tone image (also called a graphical image or a synthetic image). This is an artificial image. It has few colors or many colors, but it does not have the noise and blurring of a natural image.
- A cartoon-like image. It is color image that consists of uniform areas. Each area has uniform color but adjacent areas have very different colors.

Besides, JPEG (Joint Photographic Experts Group) is a sophisticated compression method for color or grayscale still images. It works best on continuous-tone images. One advantage of JPEG is that it provides many parameters, which allow user to adjust the compression ratio over a very wide range. User selects the ratio based on the tradeoff compression and quality. JPEG can reach any kind of continuous-tone image, regardless of image dimensions, color spaces, pixel aspect ratios, or other image features.

MPEG (Moving Pictures Experts Group) is a method for video compression, which involves the compression of digital images and sound, as well as synchronization of the two. There currently are several MPEG standards. MPEG-1 is intended for intermediate data rates. MPEG-2 is intended for high data rates. MPEG-4 is intended for very low data rates, which is the special standard for fixed and mobile web. MPEG-7 is a standard for describing the multimedia content that supports some degree of interpretation of the information's meaning, which can be accessed by a device or a computer code. More powerful MPEG-21 standards were released in 2001. The MPEG-21 provides a multimedia framework that allows users using a concise and powerful XML-based schema for declaring digital items that are anything from a single picture, a sound track to a complete collection of audiovisual works.

2.1.5 MM-DBMS Query Languages in Wireless Mobile Environments

MM-DBMS query languages under wireless mobile environments are critical since query languages are an integral feature of DBMS. However this issue is still only partially addressed by the database research community [CMP01]. Up to now, querying multimedia databases can be in different ways, such as querying by content (e.g. Show the movie where a cartoon character says: "How are you?"), querying by example (e.g. Show the movie which contains a popular song.), querying by time indexed (e.g. Show the movie 30 minutes after its start.), spatially querying (e.g. Show the image where UNO PKI is located.), and specific application querying (e.g. Show the video where the weather is snowing.), etc. All of these models have the impacts of user preferences on query processing, management of imprecise queries, new kinds of predicates and operators, and so on. Moreover, adding the context of mobile environment conditions, these query methods become more complex. Researchers often consider ad-hoc scenarios or completely ignore the other coordinates of the problem. Therefore, there are no proposals considering a full-fledged query language expressing all the basic MM-DBMS functionalities within a uniform framework.

There have been two MM-DBMS query languages suitable for wireless mobile environments after considerable modification. One is based on the SQL Multimedia and Application Packages (SQL/MM) [ME01] that is standardized in May 2001 by the ISO. SQL/MM introduces object types and associated methods for full-text, spatial data, and images. For images, SQL/MM provides structured object types, and methods to store, manipulate and represent image data by content. The other one that is presented in [Fla00] is based on the integration of object-relational database technology and XML techniques into wireless mobile environments. It is a framework optimizing mobile data exchange and presentation by using of multimedia specific queries, data preprocessing and reduction methods within the database system. Meanwhile it supports XML/XSL based data transformation and presentation on mobile end systems. However, all of these two query languages should be improved to process wireless network QoS parameters like local resources, available communication channels, and user preferences, etc.

2.1.6 Querying Data Recovery and Concurrency Control Mechanisms

Due to high communication latency, intermittent wireless connectivity, limited battery life, and changing client location, disconnections in the wireless mobile environments are unavoidable. Querying data recovery mechanisms should be provided, although in the context of multimedia, the data recovery activities are time consuming.

The passive solution is mainly by reducing the amount of data backup, which means to save the pointers of queried data addresses and to pass these pointers between networking layers and server. As in traditional data communications, when a packet is lost or corrupted, it should be recovered. However, this strategy is not suitable for multimedia communications. First, multimedia data can tolerate some loss. Second, recovery causes delay to the following data, resulting in more useless data at the client site. Third, implementation of the recovery strategy requires special software and large cache memory, which makes the transport protocol complicated and slow. Therefore, selective recovery techniques have been proposed. One of them requires only the lost packets are recovered rather than all packets from the last loss.

The active solution is by means of adding data caching and proxy functions in the MM-DBMSs and client-site systems. With these functions, servers and clients cache replicas of frequently accessed data. For a server (and symmetrically for a client), the proxy can cache updates intended for the client. When a connection becomes available, the proxy automatically forwards these cached updates to their ultimate destination.

Meanwhile, to ensure the ACID (Atomicity, Consistency preservation, Isolation, Durability) properties of queried multimedia objects, concurrency control techniques should be integrated in recovery mechanisms. From original status of the MM-DBMS query processing point of view, the main concern here is the queried object. When the object is being recovered due to a disconnection, it must keep the same properties as its original's. To fulfill it, traditional methods such as using timestamps to guarantee serializability of schedules can be applied in the recovery and concurrent control mechanism for the MM-DBMS's query processing.

2.1.7 QoS Profile of Server-site Multimedia Object

Up to now DBMSs do not support QoS for multimedia object. There is not commercial MM-DBMS available in the market [Sti01]. However, QoS mechanism is a critical part in designing and running a MM-DBMS. Since this thesis is focusing on addressing the QoS issues of MM-DBMS in wireless mobile environments, we should study and configure the QoS profile for multimedia object in MM-DBMS server site. Besides the QoS profile for server-based object, we will compose the Client-site QoS profile and extend QoS for multimedia applications in wireless networks in Chapter 6.

2.2 QoS Techniques Involving in Wireless Networks for Multimedia Applications

To develop wireless networks for satisfying users' multimedia applications, QoS support must be integrated throughout all underlying networks. By means of the following vital QoS fundamentals, the development of QoS-based query processing of MM-DBMS in wireless mobile environments becomes possible.

2.2.1 QoS Requirements in Wireless Mobile Multimedia Environments

A typical QoS operation model of a communications system is as follows. An application specifies its QoS requirements and submits to the system. The system determines whether it has sufficient resources to meet the requirements. If yes, it accepts the request and allocates the necessary resources to serve the application so that its requirements are satisfied. If not, it may either reject the application's request or suggest a lower QoS requirement that it can satisfy. Based on this operating model, the following elements are needed to be provided in QoS services:

- A QoS specification mechanism for applications to specify their requirements;
- Admission control to determine whether the new application should be admitted without affecting the QoS of other ongoing applications;

- A QoS negotiation process so that as many applications as possible can be served;
- Resource allocation and scheduling to meet the QoS requirement of accepted applications;
- Traffic policing to make sure that applications generate the correct amount of data within the agreed specification.

In addition to the above basic elements, other elements are needed to meet the diverse requirement of multimedia applications. First, a QoS renegotiation mechanism is required so that applications can request changes of initial QoS specifications. Second, the actual QoS provided to the ongoing sessions should be monitored so that appropriate action can be taken in case of any problem in meeting the specified QoS services. Third, media scalability and graceful quality degradation techniques should be used together with the above mechanisms to provide satisfactory services to multimedia applications.

Moreover, taking into account the major characteristics of wireless networks, such as high losses, low bandwidth, battery power constraints, and mobility, we should modify the QoS mechanism to satisfy these features. Therefore, apart from the parameters like packet delay, packet loss rate, delay jitter and minimum and maximum bandwidth, the following elements should be additionally considered [MS00]:

- A QoS extends the use of special protocols to support for making advanced resource reservation during mobility.
- Special protocols support wireless QoS parameters, such as *loss profiles* that includes trigger messages to control the data transmission and gives applications an opportunity to choose between a burst loss and a distributed loss in case of an

overload situation occurred; *probability of seamless communication* that defines the nature breaks allowed in the service, *rate reduction factor* that denote a factor by which the original resource request can be reduced in case a reservation does not go through is useful, and *power profile* that is sent along with the control messages from the mobile indicating the percentage of battery power that is left.

2.2.2 Mobile IP Networks and QoS-support Protocols

As previous mentioned, 3G mobile systems evolve by the integration of three essential domains: broadband, mobile, and the Internet. This is a technology in which user specifies the type service he needs and the cost he accepts. Thus, for the system development point of view, it would be of a great importance to ensure the availability of a reliable internetworking among various networks attached to IP backbone. The user of a mobile IP network may experience that the mobile device frequently changes its network point of attachment during movement and while an active session is ongoing.

There are two most unique characteristics of mobile IP networks. One is limited available bandwidth. Due to the wireless channels are prone to bursty and location-dependent errors. This causes further reduction in the effective available bandwidth. The other is user mobility. In wired networks, the task of admission control is to admit or reject a new traffic request depending on the probability of QoS violation. However, in wireless networks, we also need to consider the handoff traffics, that is, the continuing traffic coming from neighboring cells due to user mobility. Recently, new IPv6 is introduced to insure that data packets have less delay and enough bandwidth. The key enhancements in IPv6 is addressing, in which 128 bits are used instead of 32 bits used now, QoS and Security.

2.3 Techniques Involving in Multimedia Database Client Site

The technological developments of multimedia wireless networks and MM-DBMS allow the multimedia data query application possible. In addition, with the techniques focused on clients' applications available, the wireless multimedia database query comes to reality.

As shown earlier in figure 2.1, four components should be included in multimedia database client system: mobile multimedia query language, mobile networking protocol, and mobile multimedia compressed data decoding, and client QoS profile. Three of them have been discussed in the previous section. Same as the QoS profile in server site, the client QoS profile is considered as a vital role in the realization of running wireless mobile MM-DBMS. To study and configure the client site QoS profile is the part of core work in this thesis.

3. CHALLENGE OF QOS TO MULTIMEDIA APPLICATIONS

3.1 Importance of Quality of Service

QoS, as a terminology, is one of the most common words used by the people in the field of telecommunications in recent years. Now QoS also becomes familiar by mobile and multimedia database developers and it is being one of critical challenges for database research community [Thi98] [Fla00] [SE02] [Shi02]. Many articles discuss QoS regarding the development of the MM-DBMSs. One of them [Shi02] analyzes that multimedia resources requires a satisfied quality for presenting. In MM-DBMS, some specific file structure and program are required to support the quality of service.

3.2 Context of Multimedia Application

In today's information full-occupied world, billions of business and home clients access millions of databases. Moreover, rapid advances in wireless communication technology drive database systems to extend their services for mobile users. Besides, multimedia will be the mainly contents stored in database. Currently, we know the fact that, in the multimedia database query processing under wireless mobile environment, one of the important things is that users need querying results and efficient processing time. Thus the big issue is that, if the user finds the retrieved and stored multimedia object is neither interesting nor useful when it is displayed, the time and bandwidth used to transmit the objects have already been wasted. To save these precious time and expensive bandwidth, usually users interest in quickly viewing an object at an acceptable resolution without paying much attention to the details or at the limited device display capability. Therefore, users have to flexibly degrade or change their QoS requirements according to different applications.

The best way to illustrate the above motivation is to show an example. Figure 3.1 depicts a scenario where the client degrades his quality criteria and obtains coarse query result with his limited display capacity and within his satisfied time period.

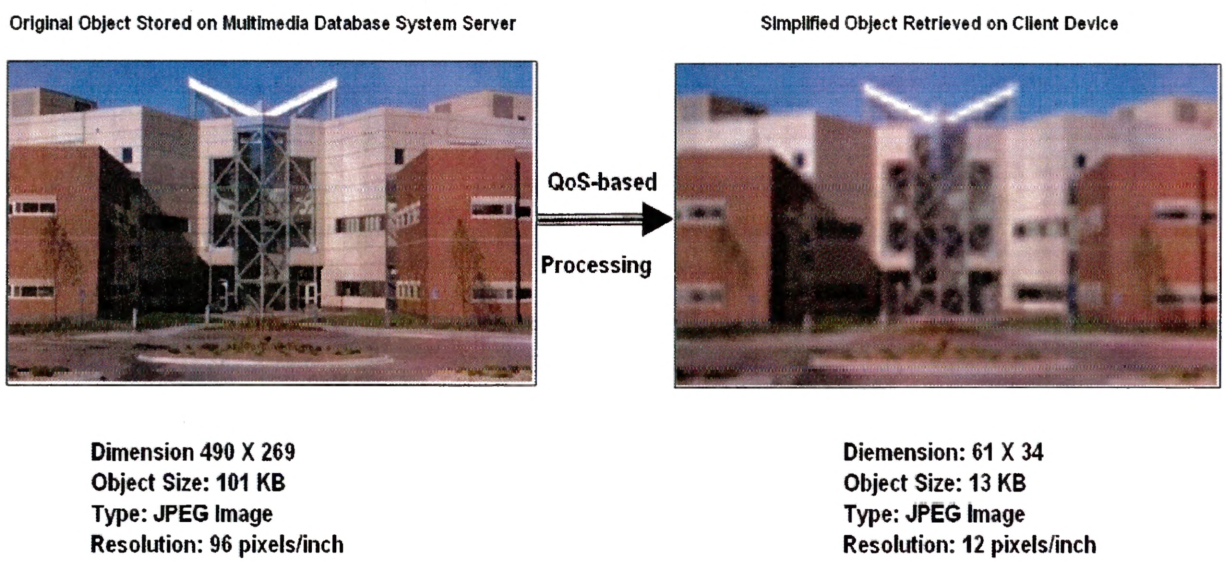

Figure 3.1: The comparison of original and processed query image enhanced by extended QoS

We illustrate the above mentioned scenario through the following figure 3.2. Suppose the client with a current available PDA device queries a University of Nebraska Peter Kiewit Institute (PKI) image in a multimedia database through wireless networks. The client uses GPRS system to connect the UMTS wireless network. The server uses wireless LAN (Wi-Fi) to connect the wireless network. There are two problems existing in client site.

(1) The display capacity is too low to present the captured image object. Since the original dimension is 490 × 269 pixels, however, the screen is 128 × 128 pixels.

(2) Suppose the client can solve the screen resolution problem, the response time is still a problem due to the out of the upper bound delay time.

The expected transmission time is over 60ms (since IPv4's one packet size is fixed at 64KByte and original image file needs two packets to transport, the available throughput for the network is 512Kbps×50%, therefore the data transport time t_t = (64Kbyte/512Kbps×50%)x2packets = 1second. Considering the query processing time t_q and network delay time t_n, the user's waiting time T ($T = t_t + t_q + t_n$) is intolerable. Therefore how to solve these problems comes to our interesting.

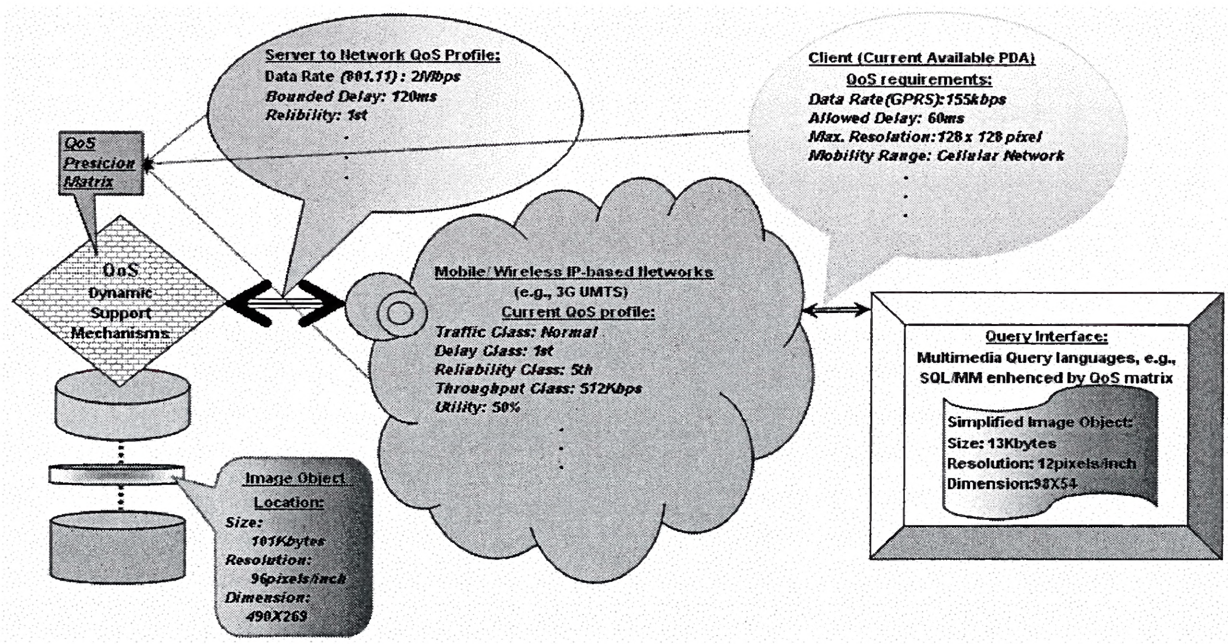

Figure 3.2: The illustration scenario of image query processing enhanced by extended QoS

Theoretically speaking, using extended QoS enhanced multimedia database's query processing under wireless mobile environments, these two issues could be solved. The basic reason is that we reduce the final quality of picture to 12% of that of original one for the compensation we obtain the simplified object size from 101Kbyte to 12Kbyte.

In this case, the data transmission time could be possibly controlled within 60ms (since t_t = 64Kbyte/512Kbyte×50% = 0.5 sec).

3.3 Tradeoff Multimedia Querying results and Qualities based on QoS

Hereby we get at the point: through fundamentally understanding of the mobile IP-QoS requirements involving in multimedia database's query processing, we explore a new approach to achieve the trade-off in querying results and qualities according to application priorities and capacities.

The primary objectives of our work are as follows.

(1) We comprehend all kinds of QoS requirements regarding multimedia database querying process in the context of wireless mobile environments;

(2) We design a QoS–based Matrix for the purpose of control the query processing efficiently and effectively;

(3) We, with the support of QoS-based Matrix, differentiate the querying results' presentation qualities according to wireless network real-time available resources and multimedia database application's limitations.

(4) We explore an approach that is general consideration which is not limited to any particular environment.

4. FRAMEWORK OF QOS AND ISSUES IN A MOBILE ENVIRONMENT

Henceforth, we observe the QoS theoretically and practically so that we can explore what the problems are in MM-DBMS in the context of wireless mobile environments. By investigating of QoS and its requirements, we establish a framework of QoS for wireless network and discuss how wireless mobile environment initiations require the QoS. Finally, we outline the issue that current QoS framework has.

4.1 Theoretically and Practically Investigating the QoS

As indicated earlier, QoS can be defined as a set of specific requirements for a particular service provided by a network to users. These requirements are usually described by quantitative figures. Therefore instead of asking for a network service, user is asked to specifically request other sorts of measures such as connection speed or delay, which can be described by a numerical value, for example, 56 Kbps or 300ms, respectively. Having a quality term such as good or bad described by a quantitative metric simplifies the process of allocation of a particular service and also prevents any possible ambiguity during the user request and service fulfillment process.

Although changing quality into measurable quantity is a good step toward clearing the meaning, there are yet other things to be done. The ambiguity in QoS definition can be pointed out in many directions by posing the following questions:

1. What types of qualities can be attributed to a particular service?
2. Which entity in the network is responsible for providing a particular service?
3. Who is the accounting entity in the network to receive a particular service?

4. Is the requested service a network service or an end-user terminal service?
5. Who judges the need for a particular service: the user or the network?
6. What is the source of the service requirements: user, network, or technology?
7. Is it possible to fulfill a service quality at a particular time and location?
8. What should be done if the service is not available at a particular time or location?

The above questions and many similar ones simply illustrate that there cannot be a definite definition for QoS. Any attempt to find a single answer would be a waste of time. In recent years, many companies have started to raise the issue that the next-generation telecommunications networks, including the Internet users, can receive a guaranteed QoS. However, the topic is too broad. No one can explain what the guaranteed QoS would be. Everyone knows that if a network can provide a full commitment on its services at all times, then it would be the ideal network. But the issue here is what type of services we should consider. To explain this issue, we should clearly separate user-based requirements of QoS from network-based requirements of QoS.

4.1.1 User-level QoS Requirements

At the user level, requirements of QoS are mainly those that can be sensed by user. This means that system-level QoS architectures are essentially transparent to user. For example, if you are using a MM-DMBS through your PDA for searching the data through wireless network to your company's remote servers, your ultimate purpose and requirement for your service is to have a reliable access and a reasonable time within the capability of your limited facilities. Meanwhile you are moving from the service area of one base station to another one. The wireless network needs to perform a complicated procedure of location updating and handoff between cells to maintain the continuity of your connection and service. For users, all these procedures that networks serve for them are transparent and ignored, unless user suffers from a problem such as a connection drop or sensible change in quality of communication happens during the handoff.

In the above particular user application example, a multimedia database application has been used and on the basis of that application, the user expectations of communications continuity and time consuming are considered as the perceived QoS metrics. Through this general user philosophy, we can generalize large scale and more complicated cases. No matter what kinds of multimedia networking facilities included, all the users have their own criteria in defining QoS requirement.

In general, user-level QoS requirements are in accordance with the perceived QoS based on data transmission and application type. To make this clear, let us consider a concrete example of multimedia transmission of a video clip with voice over the Internet. A user may consider several factors to evaluate either satisfaction or a problem with such a multimedia communication system. The possible factors are: video rate, video smoothness, picture detail, picture color accuracy, audio quality, video/audio synchronization. These factors are self-explaining and are usual in the case of video viewing on the Internet. We can judge from this example is that not all users will use the same set of factors to determine the quality of such a video transmission. One might be interested in smooth playback of a video at an acceptable rate without paying much attention to the audio quality or even the picture color quality or picture detail. Therefore, ***even for the same application we cannot put the same set of QoS requirements for all***

users. Users choose the applications and there are a limited number of QoS requirements associated with each application. However, among all these requirements there should have special criteria that make the user QoS profile to be satisfied.

4.1.2 Network-based QoS Requirements

From the view point of wireless network architecture, we can find more indicative figures to illustrate the QoS provided to users. Although many of these indications can be accessed by users, they are more or less related to the technology behind the network service and thus a user will find limited flexibility in changing the profile after subscription to the service.

We may divide these requirements mainly into four attributed types:, (1) *Bandwidth*; (2) *Timeliness*; (3) *Mobility*; and (4) *Reliability*.

4.1.2.1 Bandwidth

Bandwidth illustrates the speed or data rate available to a user application. Very loosely speaking, we can say that the more bandwidth available in the system, the higher data rate that can be provided to user application. The above statement is loose because it does not include other determining parameters that could affect actual data rate given to a particular application at a specific period. For example, when you compare 100-Base-T with 10-base-T LAN, with nominal data rates of 100 Mbps and 10 Mbps, respectively, each with N hosts attached, you may say that 100-base-T LAN is faster than 10-base-T LAN. However, this comparison does not consider loading these two particular LANs. If, for example, 100-base-T network is fully loaded and 10-Base-T network is very lightly loaded, then each of these N hosts in 10-base-T network can receive a faster connection speed than the users in the 100-base-T network. To make the definition of bandwidth more precise, we need to distinguish among three different rates, they are the *system-level data rate*, the *application-level data rate*, and the *transaction-level data rate*.

System-level data rate shows actual data rate at the physical transmission media at uplink and downlink directions. This rate relates to static network characteristics such as the type of media used for connecting user terminal to network, network technology, and network topology as well as network dynamic characteristics such as current traffic, loading, current network capacity, and also the service agreement between user and subscribed network. The system-level data rate could have a nominal value that can be a range agreed by user and network. The system-level data rate could also be further limited by interactions between protocols at the higher layers of the network stack such as transmission control protocol (TCP) and IP.

Application-level data rate could have a different value from the system-level data rate. Application protocols designed for high-bandwidth applications such as multimedia usually use different compression algorithms to reduce the amount of data exchanged between the application and transport protocols and eventually passed to the physical layer. ***Usually, the more the application data is compressed, the less the bandwidth that is required to transmit the data. However, at the same time this means to reduce the quality of received information. Therefore, there should be a trade-off between the bandwidth you may obtain to use and the quality of information you tolerantly receive at the other site of the network connection.***

The transaction-level data rate illustrates different from either of above-mentioned two rates. If you assume a transaction as one single task to be done in the transmission of certain multimedia data, the transaction rate simply shows the rate at which the user, successfully supported by the attached network environment, can perform the predefined tasks. Depending on the nature of applications run by the user, one or some of these bandwidth indicators can be used to illustrate the QoS provided by the network. A user might use different indicators to quantify the bandwidth service provided by the network.

4.1.2.2 Timeliness

The second attributed type of network-based QoS is timeliness. Timeliness can be sensed through *delay time*, *response time*, and *delay variation*. Delay can be defined as the time spent by a user from an instance the user requests some information from the network until the instance that the information is completely downloaded to the user terminal. Note that this is not the only way in defining the delay. Such a delay definition covers all type of delays that could happen within the network, including the user terminal processing time and transmission delay, link propagation delay, queuing delay at the input ports of the intermediate routers, backbone network delays, and also the processing delay at the destination host, which has the requested information.

Response time can be considered as part of the delay definition given earlier or in some occasions as a single indicator of the quality. It simply tries to illustrate how fast the network as a whole is in providing the requested information to a user.

Delay variation is the third timeliness indicator that we have specified. In the usual situation, the previous two timeliness indicators delay and response time, could be sufficient to illustrate the service timeliness quality. However, for some applications, such as real-time multimedia, it is not the delay but the delay variation that affects the quality. If network as a whole imposes a long but fixed delay at all times during the period that the application is running, it is possible to compensate the effects of delay by simple methods such as buffering and delayed replay. However, if the delay variation shows very diverse values from time to time, it would be very difficult for the application protocol to adjust to a good of compensation. Therefore, the delay variation would be necessary in some situation that include increasing usage in future networks.

4.1.2.3 Mobility

For wireless, we should add another QoS attributed type, namely, the mobility range. It is how big the geographical area in which a user can move around and still receives the service from the attached network.

Table 3.1: Mobility coverage and capacity of current different wireless networks

Wireless Network	Coverage	Data Rate
Infrared	Room	19.2 kbps – 4 Mbps
Bluetooth	10-15 m	1-2 Mbps
IEEE 802.11	100-500 m around each AP	2-11 Mbps
GSM	Cellular Network	9.6 kbps
GPRS (for GSM)	Cellular Network	155 kbps
UMTS/IMT-2000	Cellular Network	144kbps for vehicular users
		384 kbps for mobile users
		2 Mbps for static users

We see the relationship between the mobility range and other important QoS metrics as the above table 3.1. Tables 3.1 summarize several known wireless networks and list their usual coverage and the bandwidths they offer. On the top of the list in the table is the cheapest wireless connectivity technology, the infrared. Infrared ports can provide a reliable and cost-effective short connectivity between computers, handheld terminals, cellular phones, and peripheral devices. Without any infrastructure, these devices can make a computer network or a point-to-point connection. The second one is Bluetooth technology, which is a standard for wireless communications between devices in a personal area network using radio frequency for a short range. Bluetooth is better than infrared because Bluetooth does not need line of sight and it supports multipoint. The third listed system is wireless LAN, defined in IEEE 802.11 standards, which provides a high-speed computer network. In the cellular world, the 2G systems, such as Global System for Mobile communications (GSM), and the packet-switched 2.5 generation system General Packet Radio Service (GPRS), the successor of GSM, and finally the 3G systems such as Universal Mobile Telecommunications System (UMTS) offer large mobility range up to a few miles in radius around a single base station with further coverage expansion through their cellular topology and handoffs.

4.1.2.4 Reliability

The fourth attributed type we have named for the network-based QoS is reliability. Reliability in a wireless networked system could be quantified by measuring the time or the frequency. The time shows the average time for a failure to happen, the average time the system needs to recover from a failure, or the mean time between failures that happen in the network. It would be important for wireless network to avoid long failure times or frequent failures or corrupts.

4.1.2.5 Correlation between the QoS attributed types

The attributed types defined in the above are important. They can define service quality individually or together according to applications' requirements. However, it should be clear that all of the attributed types in QoS metrics for all users are neither possible nor necessary. On the basis of user requirements or application requirements at a particular time, there will only be a subset of those QoS indicators that need to be provided by the network to a user.

4.2 QoS in IP Wireless Networks

The current Internet service is based on so-called best-effort service. This service does not guarantee any thing other than delivering the IP packets. The network only promises to do its best to deliver the user information with available resources it has. However if the packet is lost or corrupted, the respective entities should try to retransmit the lost packet and recover it. The IP providing this type of service is said to be unreliable.

We can establish a QoS network with our restrictions and requirements. There are four fundamental principles in providing QoS in IP networks [BH97].

4.2.1 Packet Classification

Assume a simple example in which two multimedia database applications are running at two source hosts A and B, respectively, directly connected to a single router R_1. The host A is a video clip retrieving over IP application, which is delay and bandwidth sensitive that requires 1 Mbps bandwidth and short delays as well as bound delay variation. The host B is an image transmission, which is delay and bandwidth insensitive. A limited capacity link of 1.5 Mbps connects the routers R_1. All packets have to be routed through this connecting link to the other parts of network.

The video clip retrieving case needs typical bandwidth, delay, and delay variation requirements to conduct its sensible communications. Meanwhile, the image transmission can take longer time for delivery of packets so that no restrictive delay or bandwidth requirements are assumed here for this application. This simple example also illustrates different QoS requirements for different applications.

In an ideal situation, a network manager wants to provide 1 Mbps out of the available 1.5 Mbps to the video application and the leftover capacity of 0.5 mbps to the image application. One way to do this is to classify different packets (video packets versus image packets) at the port of router so that router can share total capacity proportionally to the two hosts. Therefore, ***the first principle in providing QoS is that we need to classify different types of packets generated from individual applications***.

4.2.2 Packet Isolation

Applications may misbehave and use more network resources than what they really need. For example, the host A uses more than 1 Mbps for its video application, it may improve its service slightly but on the other hand it will destroy the host B image application performance. So there should be some rules within the network to monitor the behavior of applications and control the network resources. Therefore, the ***second principle is that in addition to packet classification, we need to monitor and control that no application can use more resources than what it has been allocated***.

4.2.3 Efficient Resource Management

One way to provide monitoring that no application uses network resources excessively is to partition these resources. In our example, the bandwidth is the main resource and we can partition it into two parts of 1 Mbps and 0.5 Mbps allocated to host A and host B, respectively. This can be easily done by maintaining two different queues at the input port of router R_1, which can be easily implemented by the software. Note that when we say allocating resources to a user, it does not mean that this allocation is permanent. The allocation can be managed and reconfigured dynamically, for example, only for the duration of the respective application.

One problem that could be raised here is what happens if host A's video application does not use its allocated resources for some time. This means that although we have tried to discriminate the two applications to provide a good QoS, we may waste the precious network resources. Therefore, we come to the ***third principle that although we need to differentiate individual application's packets and monitor their limit on usage of the available resources in order to provide QoS in the network, we still need to make sure that our resources not be wasted at any time.*** Therefore, the resource management will be an important issue in any QoS network and we need to use the network resources as efficiently as possible. Management of the queues maintained at the routers could be thought of as part of the overall resource management in the system.

4.2.4 Traffic Load Control

Our three principles discussed so far are important and can provide a good service quality in network. But we have not included one important factor in our discussions.

That is our network resources limited. In above example, we had 1.5-Mbps capacity and thus we share it between two applications, at 1 Mbps and 0.5 Mbps. But this is not always the case. The network has eventually many users, sharing the available capacity will not be an easy task. For instance, consider our example when hosts A and B are both running video applications and both need 1 Mbps capacity to comply with their QoS requirements. Using the simple calculation we can see that it is not feasible to share the available 1.5 Mbps between the two hosts. In such a situation, we can simply forget the QoS and give each user half the available capacity. But if we want to provide QoS, then we need to do something else. That is, we need to allocate to one of the hosts the required 1Mbps and ask the second one to wait until the required bandwidth is available. The problem of who gets the capacity first and who gets it next is another issue. We may allocate the capacity on a first-in-first-served basis, or on a user subscription priority basis, or any other policy.

Thus, ***the fourth principle in providing QoS in a network is that, in order to handle the situations for allocation of the network capacity more than available resources, we need to have Call Admission Control (CAC) in the network.***

4.3 GPRS QoS support

GPRS is packet version of its predecessor GSM. It can accommodate the Internet services to the users of cellular networks. GPRS was a result of increasing demand in the Internet applications for mobile users and thus it can be considered as the first initiative in wide-area mobile Internet.

GPRS defines QoS requirements for each subscribed user in a QoS profile, which is defined and maintained at the GPRS network Home Location Register. Every subscriber to GPRS network is allocated a QoS profile with a number of QoS indicators.

- Traffic precedence class: defines the priority of service within the network, such as high, normal, or low priority;
- Delay class: defines how much delay can be tolerated for a given service, there are four classes from 1st predictive mean transfer delay class, which is less than 0.5 second for data size of 128 octets, 2 second for data size of 1024 octets, to 4th class, which is unspecified.
- Reliability class: define how much reliability (e.g. corruption, loss) can be tolerated for a given service. There are four classes, from 1st class that has 10^{-9} probability of losing or corrupt transmission to 4th class of nothing specified;
- Peak throughput class: defines the maximum rate allocated for the delivery of a given service to a user. The rates include 8, 16, 32, 64, 128, 256, 512, 1024, or 2024 Kbps;
- Mean throughput class: defines the average data rate at which the service will be available to the user, there are 19 classes from best effort to 111 Kbps.

4.4 UMTS QoS support

The advanced features of GSM and the new network architecture of GPRS make it possible that the core network of 3G system UMTS follows these two predecessors. While using similar core network architecture as the GPRS and following similar concepts for the QoS provisioning, UMTS adds specific classifications for the traffic loads. UMTS defines four traffic classes:

(1) Conversational Traffic Class

(2) Streaming Traffic Class

(3) Interactive Traffic Class

(4) Background Traffic Class

Conversational Traffic Class refers to the traffic generated from real-time applications that require a Constant Bit Rate (CBR) during communications. Typical examples are voice and videoconferencing, and network games. A bit error rate (BER) of less than 10^{-3} is necessary for these applications.

Streaming Traffic Class requires a preserved time between information entities of stream. The increasing applications of streaming multimedia, including streaming audio and video over the Internet, are the main drives for inclusion of this traffic class in UMTS. Typically, a BER of less than 10^{-5} is necessary for these applications.

Interactive Traffic Class refers to the traffic generated from applications such as web browsing and the Internet games. Request response pattern and persevered data integrity are essential characteristics of these applications. A BER of less than 10^{-8} is typically necessary for these applications for reliable data transfer, whereas the delay constraints are not as severe as in the case of conversational traffics.

Background Traffic Class is those that come from delay insensitive applications such as e-mail or FTP downloads. For this traffic class, the destination is not expecting the data within a certain period of time but the data integrity and reliable data transfer are very important. Therefore, a BER of less than 10^{-8} should be given to these traffics.

4.5 QoS Issues on Multimedia DBMS in Wireless Context

So far we have described current QoS framework in wireless networks. We have concluded that QoS could not be set as a series of parameters to be used for all networks and all applications, and QoS should be determined by required parameters of particular network facilities and specific application conditions.

However, one fact comes to our attention: if the application instances are the multimedia objects from MM-DBMS on the context of wireless mobile environment, what do the issues happen? Recall the example of PKI as shown in figure 3.1, we review that one of the characteristics of multimedia objects is their quite different storage sizes even though they have the same attributes, possibly from several kilobytes' thumbnail images to several-million-bytes' digital camera pictures. We put the special applications to evaluate the principles we have organized for providing QoS in a communication network, irrespective of the type of the network being mobile or fixed, data or multimedia. These principles are the requirements of packet classifications, packet isolation, efficient bandwidth utilization, and admission control. Can these principles be implemented in any communication network that seeks QoS guarantee? We have an example of the following to see the result. If the size of images in a MM-DBMS that be queried by a mobile client through wireless networks are big different, meanwhile the client can not judge these sizes in advance, how does the client setup the QoS profiles properly? The current QoS framework has no answer for it due to the reason that all the

QoS principles are based on the users' pre-requirements. However, the issue is that in multimedia application scheme, it is really difficult to know the targeted retrieval objects' sizes.

Same situations happen in the Internet initiatives for the QoS provisioning. The only condition that the Internet can guarantee QoS is that users should define their QoS indicators. However, it is impossible in the MM-DBMS context. If users setup QoS improperly, more likely the retrieved objects will be dropped by the Internet. Therefore as our opinion the current best-effort service can not be evolved into a guaranteed service if traditional QoS framework has not been extended.

As for the cellular networks, we know the fact that current QoS provisioning is little in common between the cellular QoS initiatives and the Internet approaches. Therefore the possible diverse approaches are taken for providing QoS in the Internet and in the cellular mobile communication system. It may cause major problems in future wireless IP networks. It is because that for the same issues, wireless IP networks takes different approaches from that of cellular networks do. It would be hard to provide a balanced efficient resolution. Moreover, this will make the end-to-end QoS service very difficult and costly. Without an end-to-end QoS solution, we cannot claim any victory in achieving QoS. In order to implement the MM-DBMS applications on the wireless network and the Internet infrastructures we have to overcome all the crucial constraints. This is primary motivation that drives us to study and explore the QoS-based query processing precision solution for MM-DBMS in wireless mobile environments.

5. REVIEW OF CURRENT QUERYING-RELATED QOS APPROACHES

Summarizing what presented so far, we have reached the crucial points that

(1) multimedia application, especially multimedia object retrieval, needs resource consumption often exceed the available resource capacities of the deployed wireless/mobile networks and portable devices;

(2) the precautions of these extra resource requirements can not be taken by server, client, even network infrastructure in advance;

(3) we need to extend QoS management in wireless mobile environments to specify acceptable QoS levels to allow for scaling of multimedia query processing, rather than trying neither to guarantee specific values nor to stop the querying.

Therefore, we should conduct related research work rather than rejecting query processing under adverse situations. Instead, with extended QoS support, multimedia database query processing in wireless mobile environments has the capability of returning quality reduced results under the current resource-limited conditions.

Recall that the QoS common scheduling disciplines are *best effort*, *priority-based scheduling*, and *reservation*. Note that only the latter allows one to construct a QoS management that can absolutely guarantee the QoS requirements for a size-expectable query. However, because of reservation entailed over-booking and inefficient resource utilization, it is rarely used for scheduling computer or network devices in wireless or mobile environments [BK03]. Therefore, we focus on the research of adaptive query processing techniques.

5.1 Initial Approaches: QoS Requirements during Query Execution

In the initial research stage, the query processing with QoS approaches focused on the idea of collecting QoS statistics at some key time points during the execution of a query. These QoS statistics are then used as empirical data to improve the resource allocation for future query requests. At that early time, query execution plans are typically generated statistically, which are based on a set of assumptions about the QoS and the costs of performance.

Derr et al [DMP93] describes a scheme in which query execution plans could be reorganized before query execution time if the QoS statistics required by query processing significantly differ from those obtained by query executing on real time. At query processing time, the statistics used by the query processor to generate the executing plan are stored with the plan in the database system. At query execution time, the actual statistics from the system catalogs are compared against the statistics stored in the plan. If they are found to differ significantly the query is re-organized before execution. There is no collection of statistics or modification of the plan during query execution.

Antoshenkov [AZ96] presents another way of dynamically determining the plan of query. In his approach, competing executions start executing using different plans with different QoS statistics. After a while, it becomes clear that one of the plans is better than the others, and the execution of the defeated plans are stopped. This model cannot be used for dynamically improving the resource allocation of a query.

However, the above query processing strategies break down in the wide-area environment because they are unable to adapt in response to unexpected delays. Due to the apparent randomness of delays when accessing remote data, it is not possible to schedule such delays in advance. Therefore, Query scrambling, which was presented by [UFA98], has been proposed to address this problem. It deals with the significant and unpredictable delays in data delivery due to the wide area networks involving a large number of remote data source, intermediate sites, and communications links, all of which are vulnerable to congestion and failures. The mechanism of query scrambling is to reschedule the query execution plans while encountering delays. Unfortunately, these approaches are passive and are very specific techniques to tackle a specific problem found in DBMS [KD98]. It is because those bad scrambling decisions can not be avoided.

Research projects described in [KD98] and [IFF99] use the special algorithms to detect the QoS problems and attempt to reschedule the execution of query. The mechanism is included dynamic resource re-allocation feature, which is designed and implement based on run-time collection of statistics. The major issue of the approach is slow down the normal execution of a query.

5.2 Current Approaches: integrating QoS into all phases of Query Processing

Although, the current research directions for query processing consider combining QoS technologies into whole query processing procedures, the inherited consideration still focus on using QoS statistics to allocate resources. Moreover, for the reason of efficiency, query processing has to be stopped at any stage as soon as the related QoS statistical or empirical data shows no sufficient resources to meet the on-site requirements.

A recent research described by [BK03] represents the current technologies for QoS-based query processing. The authors concern about the query processing on the

Internet that is not predictable and robust enough to meet the requirements of many business applications. They focus on solving the problems such as the response time of a query unexpectedly high. The goals of their QoS approaches are as follows: (1) the percentage of queries, which quality requirements could be fulfilled, should be maximized; (2) the execution of queries that cannot fulfill their QoS should be stopped as early as possible. In this way the query does not waste the user's time and money. Of course, if the missed quality constraint is a soft one, the query should not be stopped but executed in a best-effort manner.

Figure 5.1 indicates the major phases of query processing, which is used by [BK03] to support their solution. The authors divide the activities of QoS during the query processing into three query plan phases: generation, instantiation and execution. The starting point for query processing is the query itself, the QoS requirements for it, and statistics about the resources that can be used for processing the query. Based on the statistical facts that different queries often have different demands at specific resource requirements and quality constraints, query instantiation and execution use this empirical information in order to check whether the demands of a query can be satisfied. For example, if a query seems to miss its quality requirements, it could try to get a greater share on the resources at the expense of queries that can work sufficiently with a smaller share. If such an adaptation is not possible and the query will not fulfill its QoS requirements, the query has to be stopped in order to save the user's time and money.

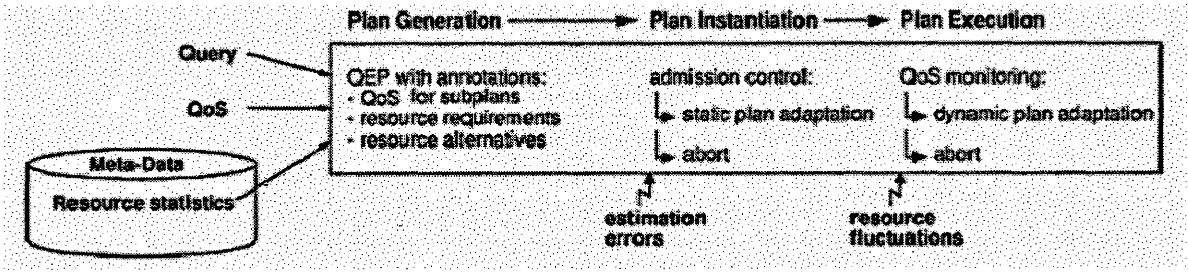

Figure 5.1: Interaction of Query Processing and QoS (Cited from [BK03])

5.3 Conclusion: Query Processing dependent on Empirical and Statistical QoS

Through the investigation of the above-mentioned approaches, we conclude that the statistical and empirical resource utilization in the query processing is the key judgment to fulfill the QoS requirements. All the proposals summarized in this chapter require the abortion of the query or generation of the query again if fails to meet the QoS. However, that conclusion conflicts with our goal. Our purpose is to explore an approach to execute the query processing regardless what conditions that QoS has. This is further discussed topic in the next chapter.

6. QOS-BASED QUERYING PROCESSING PRECISION SOLUTION

Recall the fundamentals studied in previous Chapter 2, from the perspective of enhancing query processing mechanism efficiently and effectively for MM-DBMS in wireless mobile environments, we should integrate the technologies from three aspects into a QoS-based solution. These technologies are involved in the development of MM-DBMS server system, wireless network QoS support, and MM-DBMS client-site multimedia data processing capacity. The core components of solution is to explore a QoS-based queried multimedia data precision mechanism, by which all fundamental techniques can be properly cooperated together to process multimedia query. Based on the technologies layout shown on figure 2.1, we illustrate the relationships between fundamental components and core components as the following figure 6.1.

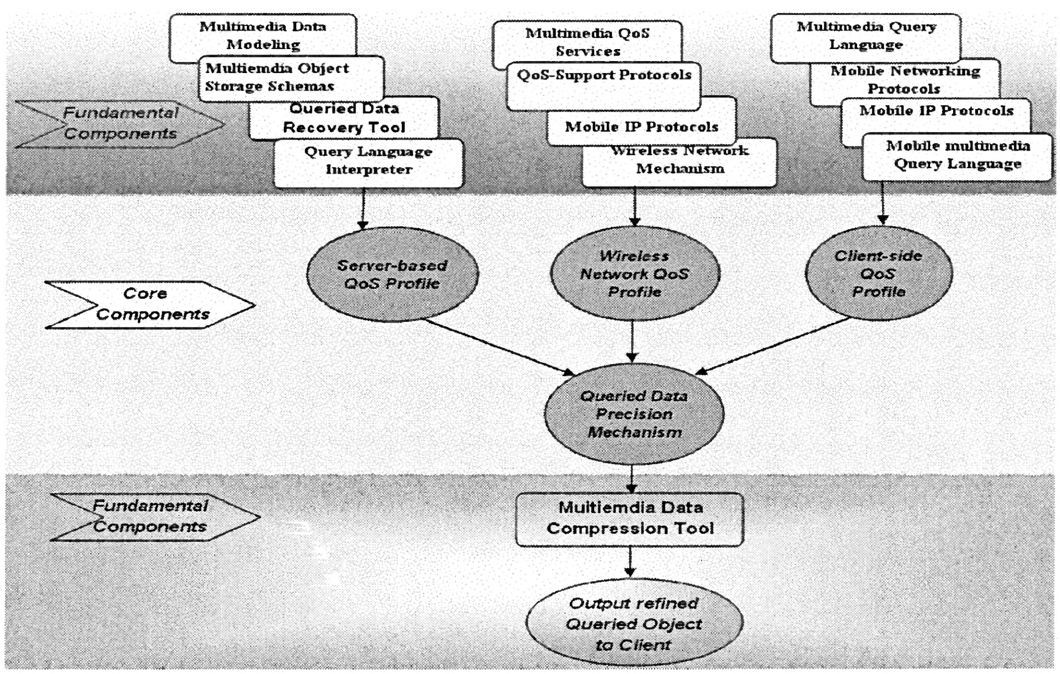

Figure 6.1: QoS-based Multimedia Query Processing in Wireless Context

In this thesis we focus on examining and establishing three categorized QoS: *server-site QoS conditions profile*, *wireless network QoS services profile*, and *client-site QoS requirements profile*. Each QoS category has four attributed types; these are *bandwidth*, *timeliness*, *mobility*, and *reliability*. Based on categorized QoS profiles, we present a QoS-based query processing precision matrix to approach the issue.

6.1 A Matrix Representation for QoS Characteristics

We address the QoS profiles having the following categorized characteristics.

- The wireless network QoS service profile is the real-time QoS for a particular application based on the conditions of on-site wireless networks;
- The server-site QoS profile is the properties of multimedia object retrieved from MM-DBMS server; and
- The client-site QoS profile is the mobile client's QoS requirements and client-site multimedia data processing capabilities.

According to QoS framework examined in previous chapters, each of the above-mentioned three categorized QoS characteristics has their own four attributed types. Although these categorized QoS profiles have their own independent functions, the attributed types that belong to each QoS categories have dependent meanings. We apply matrix notation to quantitatively represent the attributes of QoS types and the categorized QoS profiles. Through matrix analysis, we investigate the relationships among these QoS categories and related attributed types. We explain how matrix method works as follows.

As the categorized characteristics of QoS are measured objectively by quantitative values, all of the attributed types in each QoS category are represented as entries to form correspondingly type-related diagonal submatrices. Then, the attributed type-related submatrices under same category are, as blocks, composed of a categorized QoS matrix. Since we use diagonal matrix, the rows i and columns i of the matrix have the same meaning. For a matrix, the rows i and columns i mean a particular attributed type. For a submatrix, the rows i and columns i mean a particular attribute. After adjusting the attributed type-related matrices to be conformable to their counterparts in different categorized QoS matrices, we multiply all categorized QoS matrices. The product has a special function, and we name it as ***QoS-based Query Processing Precision Matrix (QQPPM).*** The objective of QQPPM is to create a smallest element as a criterion to reduce queried object quality between server's original data quality and actual quality that client can access. With the QQPPM, multimedia query result can be displayed on client site no matter what the queried object characteristics are and no matter what available network resources have.

Technically, the QQPPM solution covers the specification of client-site QoS preferences and their quantitative relationship to wireless network QoS conditions, and also the specification of server-site QoS profile and their quantitative relationship to the real time wireless network QoS conditions.

6.2 MM-DBMS Server Architecture & Related Query Data Streams

As QQPPM will be one of the core functions in MM-DBMS, we need to analyze the MM-DBMS server architecture and understand their relationship. The MM-DBMS server architecture could be presented as the left side of figure 6.2.

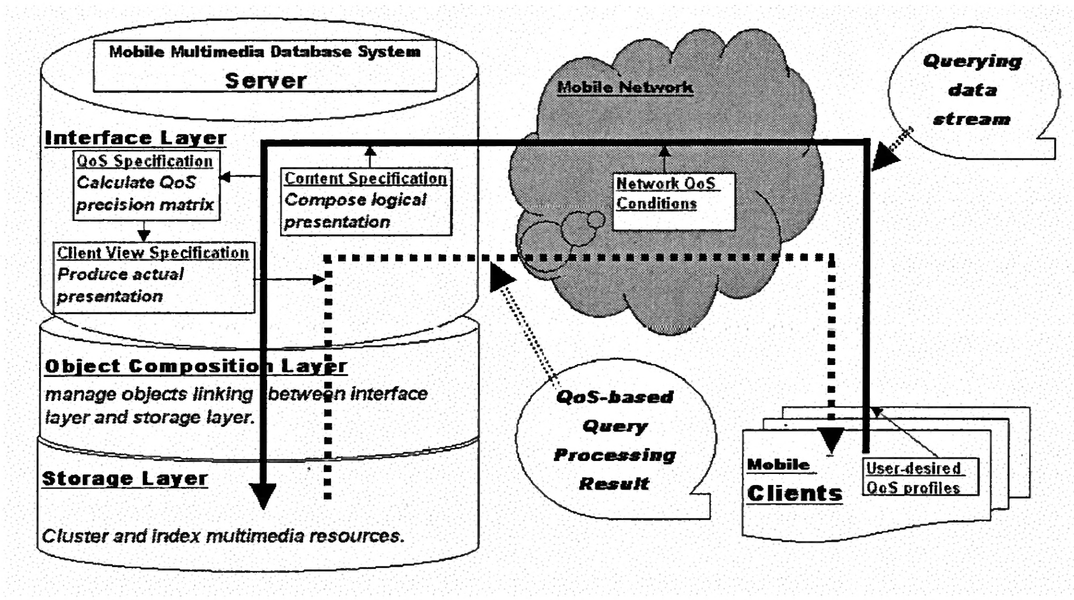

Figure 6.2: MM-DBMS Query Processing Data Streaming in Wireless Context

There are three layers in MM-DBMS server's architecture [Sti01]:

- The Interface Layer
- The Object Composition Layer
- The Storage Layer

The tasks to be dealt with the Interface Layer include object browsing, query processing, and the interaction of object composition/decomposition. Object browsing allows the client to find multimedia resource entities to be reused. Through queries, either text based or visualized, the client specifies a number of conditions to the properties of resource and retrieves a list of candidate objects. Note that multimedia resources, unlike text or numerical information, cannot be effectively located using a text based query language. Therefore, in order to allow the client to effectively find reusable multimedia objects, including pictures, sound, video, and other forms, the

Interface Layer should be included the QoS based mechanism to help the client to compose/decompose multimedia information.

The Object Composition Layer works in conjunction with the Interface Layer to manage objects. A multimedia object usually does not exist by itself. A typical multimedia presentation may contain a number of objects of various types. More importantly, the objects are embedded with some time-sensitive constraints which are ignored in traditional DBMS. To demonstrate the composed multimedia object, a number of links, such as association links, similarity links, and inheritance links, are required. These links are specified either via the database graphical user interface, or via a number of application program interface (API) functions.

The Storage Management Layer includes clustering and indexing functions. Clustering means to organize multimedia information physically on a hard disk such that, when retrieved, the system is able to access the large binary data efficiently. Usually, the performance of retrieval needs to guarantee some sort of QoS and to achieve multimedia synchronization. Indexing means that a fast locating mechanism is essential to find the physical address of a multimedia object.

6.2.1 Multimedia Queried Data Streams

Based on the relationships among query processing, QQPPM and main systematic architecture of MM-DBMS as depicted in figure 6.2, the query processing data streams can be captured by the following procedures.

1. The query processing data is combined with mobile client-site QoS profile transmitted through wireless networks to multimedia database server.

2. When the Interface Layer located in DBMS server receives the client's query request along with the client QoS profile, it transfers the client query to the Object Composition Layer. Meanwhile it saves the client QoS profile in the QQPPM that is embedded at the same layer. The query goes through the Object Composition Layer and the Storage Layer to compose the queried object. After the queried object reached the Interface Layer, the QQPPM analyzes the queried object's characteristics and figures out the server-based QoS profile for this particular queried object, while at the same time QQPPM collects on-site QoS parameters from wireless networks.
3. When three categorized profiles are available in the QQPPM, QQPPM processes the suitable customized queried object, which balances all QoS profiles. The QQPPM mechanism is sketched in the next section.

6.3 Outline of QQPPM Mechanism

When the original queried object is reached the Interface Layer, the particular QQPPM related to the object is activated. As previous expressed, the three QoS profiles, which come from client-site QoS considerations, server-site QoS requirements and network QoS services, are generated from three correspondingly categorized QoS sets. Then the QoS sets convert into three diagonal matrices, accordingly. The conduct of three matrices forms the QQPPM. Finally, the criterion query precision factor α is selected by the following steps: (1) all of the elements from QQPPM are picked up and put into a temporary set; (2) the smallest element is chosen from the temporary set as the factor α.

The particular reducing queried object's quality is refined on the basis of factor α, if necessary. With the effect of QQPPM, the QoS based queried object is conducted and the result is transmitted to the client.

The QQPPM mechanism can be illustrated as figure 6.3, where Q denotes the QQPPM:

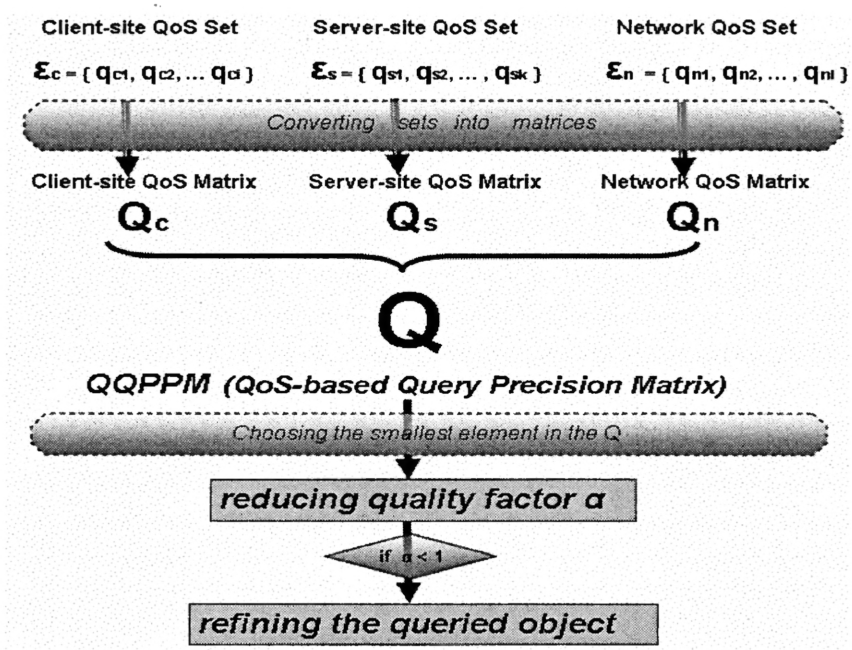

Figure 6.3: Outline of QQPPM

6.4 Construction of QQPPM

According to the definitions of QoS attributed types, as well as considering the dedicated storage schemas for multimedia objects, and the data presentation capabilities in client site, the overall QoS matrices have the following properties.

The client-site QoS matrix Q_c, the server-site QoS matrix Q_s, and the network QoS matrix Q_n, as well as their product matrix Q are diagonal matrices and are conformably partitioned by 16 blocks as shown in figure 6.4, which include four attributed type-related submatrices.

$$\begin{bmatrix} T_b & O & O & O \\ O & T_t & O & O \\ O & O & T_m & O \\ O & O & O & T_p \end{bmatrix}$$

Figure 6.4: Block Matrix used in this Thesis

The four submatrices are denoted as (1) T_b, which includes QoS attributes related to bandwidth or throughput; (2) T_t, which is related to timeliness; (3) T_m, which is related to mobility; (4) T_p, which is related reliability.

As definitions, all of the above-mentioned submatrices are diagonal matrices. The orders of all the sub-matrices with same attributed type should be uniformed, which are determined by the largest one. Similarly, the dimensions of all categorized QoS matrices equal to the sum of four sub-matrices' orders (see figure 6.5).

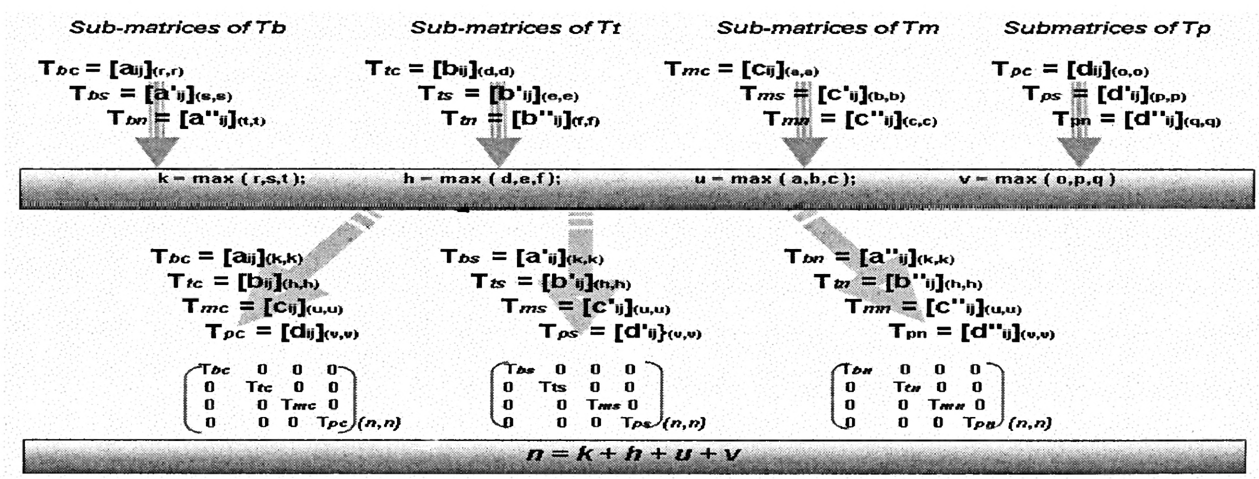

Figure 6.5: Compositions of QoS Matrices

Hereafter we build the three categorized QoS Matrices respectively.

6.4.1 Client-site QoS Considerations Matrix

Focusing on the client-site QoS requirements for multimedia data querying through wireless networks, we consider two sets of elements, one is the standard client-site QoS requirements and the other is the extended client-site QoS considerations. Figure 6.6 shows the details.

Standard Client-site QoS Requirement

Category	Parameter	Description
Perceived QoS	QcP1	Picture Detail
	QcP2	Picture Color Accuracy
	QcP3	Video Rate
	QcP4	Video Smoothness
	QcP5	Audio Quality
	QcP6	Video/Audio Synchronization
Security	QcS1	Confidentiality
	QcS2	Integrity
	QcS3	Authentication

Extended Client-site QoS Consideration

Category	Parameter	Description
Hardware Facilities	QcEH1	Power Limits
	QcEH2	Screen Display Limits
Mobility Capability	QcEM1	Data Rate
	QcEM2	Mobility Coverage
User Preference	QcEP1	Response Time Bound
	QcEP2	Data Compression
	QcEP3	Delay tolerance
Quality Criteria	QcEQ1	Percentage of Original Quality

Figure 6.6: Two Subsets of Client-site QoS Elements

On the top of figure 6.6 are the standard QoS requirements that are the perceived factors considered by a user for a particular multimedia data transmission. Normally, users define their standard QoS requirements as qualitative expressions. Therefore, to represent each application's requirements, the network systems provide the technical specifications as the correspondent quantitative expressions. On the bottom of figure 6.6 are the extended QoS requirements that include hardware facilities, user preference, mobility coverage, and critical quality acceptance. All of the extended QoS elements are quantitative expressions.

The two sets of client-site QoS profiles are converted into the following four attributed type-related sub-matrices T_{bc}, T_{tc}, T_{mc}, and T_{pc}, as shown in Eq. 6.1, Eq. 6.2, Eq. 6.3, and Eq. 6.4.

$$T_{bc} = \begin{bmatrix} q_{cp1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & q_{cp2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_{cp3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q_{cp4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_{cp5} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & q_{cp6} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & q_{cEH2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & q_{cEP2} \end{bmatrix} \qquad \text{Eq. 6.1}$$

$$T_{tc} = \begin{bmatrix} q_{cEH1} & 0 & 0 \\ 0 & q_{cEP1} & 0 \\ 0 & 0 & q_{cEP3} \end{bmatrix} \qquad \text{Eq. 6.2}$$

$$T_{mc} = \begin{bmatrix} q_{cEM1} & 0 \\ 0 & q_{cEM2} \end{bmatrix} \qquad \text{Eq. 6.3}$$

$$T_{pc} = \begin{bmatrix} q_{cs1} & 0 & 0 & 0 \\ 0 & q_{cs2} & 0 & 0 \\ 0 & 0 & q_{cs3} & 0 \\ 0 & 0 & 0 & q_{cEQ1} \end{bmatrix} \qquad \text{Eq. 6.4}$$

Finally, the client-site QoS Matrix Q_c is defined as Equation 6.5. where Qc is composed with 4 attributed type-related submatrices and other 12 zero submatrices.

$$Q_c = \begin{bmatrix} T_{bc} & 0 & 0 & 0 \\ 0 & T_{tc} & 0 & 0 \\ 0 & 0 & T_{mc} & 0 \\ 0 & 0 & 0 & T_{pc} \end{bmatrix} \qquad \text{Eq. 6.5}$$

6.4.2 Server-site QoS Requirements Matrix

The server-site QoS requirements set regarding multimedia DBMS querying in the wireless mobile environments is defined by the following elements.

- *Throughput*, which is measured in the number of bits per second and it depends on the queried object's characteristics;
- *Delay or latency*, which has a direct impact on client's satisfaction and it may cause user frustration or fail to delivery the queried object;
- *Delay variation* (*delay jitter*), which refers to the variation in the delay introduced by the components along the networking path since each packet in the network travels through different paths, and the network conditions for each packet can be different, the end-to-end delay varies and it may cause to distort the time synchronization of the original traffic; and
- *Loss or error rate*, which directly affects the perceived quality of queried object and compromises the integrity of the data.

Using the similar procedure of defining the client-site QoS sets, the server-site QoS set is obtained as shown in figure 6.7.

Server-site QoS Parameters

Category	**Parameter**	**Description**
Throughput	QsT1	Frame Size
	QsT2	Frame Rate
	QsT3	Color Depth
	QsT4	Compression
	QsT5	Constant and Variable Bit Rate
	QsT6	Burstiness
Delay	QsD1	Bound Delay
Jitter	QsJ1	Bound Variation in Delay
Loss Rate	QsL1	Tolarated data loss rate

Figure 6.7: Server-site QoS Requirements Set

From the server-site QoS set, we obtain the four attributed type-related sub-matrices, as shown in Eq. 6.6, Eq. 6.7, Eq. 6.8, and Eq. 6.9.

$$T_{bs} = \begin{bmatrix} q_{sT1} & 0 & 0 & 0 & 0 & 0 \\ 0 & q_{sT2} & 0 & 0 & 0 & 0 \\ 0 & 0 & q_{sT3} & 0 & 0 & 0 \\ 0 & 0 & 0 & q_{sT4} & 0 & 0 \\ 0 & 0 & 0 & 0 & q_{sT5} & 0 \\ 0 & 0 & 0 & 0 & 0 & q_{sT6} \end{bmatrix} \qquad \text{Eq. 6.6}$$

$$T_{ts} = \begin{bmatrix} q_{sD1} & 0 \\ 0 & q_{sJ1} \end{bmatrix} \qquad \text{Eq. 6.7}$$

$$T_{ms} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{Eq. 6.8}$$

$$T_{ps} = \begin{bmatrix} q_{sL1} & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{Eq. 6.9}$$

Since server site has no mobility consideration, we set T_{ms} as unit matrix. There is only one entry in submatrix T_{ps}, we extend it to two dimensions of matrix.

Composing sub-matrices T_{bs}, T_{ts}, T_{ms}, and T_{ps}, as well as other 12 zero submatrices, we obtain the server-based QoS matrix Q_s as following.

$$Q_s = \begin{bmatrix} T_{bs} & 0 & 0 & 0 \\ 0 & T_{ts} & 0 & 0 \\ 0 & 0 & T_{ms} & 0 \\ 0 & 0 & 0 & T_{ps} \end{bmatrix} \qquad \text{Eq. 6.10}$$

6.4.3 Wireless Network QoS Services Matrix

Wireless networks receive QoS requirements through quantitative or qualitative expressions from the clients and servers and then respond to their requests by supplying

QoS services using a number of QoS mechanisms. Generally speaking, QoS services can be determined based on following criteria: 1). What kind of service that networks can provide according to client's and server's requirements: quantitative services, qualitative services, or best effort services; 2). To which entities (individual or group class) the networks provide service: per-flow QoS services, or per-class QoS services. Wireless networks use a combination of QoS services.

To enable the above QoS services under the wireless mobile environments, the following mechanisms should be implemented in different network devices that are installed in different OSI layers:

1) Traffic handling mechanisms, which classify, handle, police, and monitor the traffic across the network. The main mechanisms are classification, channel access, packet scheduling, and traffic policing.

2) Bandwidth management mechanisms, which manage the network resources by coordinating and configuring network devices' traffic handling mechanisms. The main mechanisms are resource reservation signaling and admission control.

Based on the section of 4.1.2 of network QoS services, we consider the following attributed parameters to be included in the Network QoS services set and then convert it into network matrix Q_n.

- *Bandwidth*, which refers to system-level data rate that relates to current traffic loading, current network capacity, and also the service agreement between the MM-DBMS's client and the networks.

- *Timeliness*, which refers to the network QoS service efficiency such as delay time, response time, and delay variation.
- *Reliability*, which shows the network service deficient rate.
- *Mobility*, which indicates the network service's mobile characteristics.

Based on the above-mentioned profile, we define the correspondent network QoS set (see figure 6.8).

Network QoS Parameters

Category	Parameter	Description
Bandwidth	QnB1	System Data Rate
Timeliness	QnT1	Client-side Processing Time
	QnT2	Transmission Delay
	QnT3	Link propagation Delay
	QnT4	Queuing Delay
	QnT5	Backbone Network Delay
	QnT6	Server-side Processing Delay
Relibility	QnR1	Network Failure Frequency
Mobility	QnM1	Data Rate
	QnM2	Wireless Coverage
	QnM3	Mobile Moving Speed Bound

Figure 6.8: Network related QoS Services Set

$$T_{bn} = \begin{bmatrix} q_{nB1} & 0 \\ 0 & 1 \end{bmatrix} \quad \text{Eq. 6.11}$$

$$T_{tn} = \begin{bmatrix} q_{nT1} & 0 & 0 & 0 & 0 & 0 \\ 0 & q_{nT2} & 0 & 0 & 0 & 0 \\ 0 & 0 & q_{nT3} & 0 & 0 & 0 \\ 0 & 0 & 0 & q_{nT4} & 0 & 0 \\ 0 & 0 & 0 & 0 & q_{nT5} & 0 \\ 0 & 0 & 0 & 0 & 0 & q_{nT6} \end{bmatrix} \quad \text{Eq. 6.12}$$

$$T_{mn} = \begin{bmatrix} q_{nM1} & 0 & 0 \\ 0 & q_{nM2} & 0 \\ 0 & 0 & q_{nM3} \end{bmatrix} \quad \text{Eq. 6.13}$$

$$T_{pn} = \begin{bmatrix} q_{nR1} & 0 \\ 0 & 1 \end{bmatrix} \quad \text{Eq. 6.14}$$

Since submatrices T_{bn} and T_{pn} have only one entry, we extend them as two dimensions matrices. Based on the set, we obtain the related four attributed type-related sub-matrices T_{bn}, T_{tn}, T_{mn}, and T_{pn} (See the Eq6.11, Eq 6.12, Eq 6.13 and Eq. 14). Finally, we compose the network QoS matrix Q_n as Eq. 6.15.

$$T_n = \begin{bmatrix} T_{bn} & 0 & 0 & 0 \\ 0 & T_{tn} & 0 & 0 \\ 0 & 0 & T_{mn} & 0 \\ 0 & 0 & 0 & T_{pn} \end{bmatrix} \qquad \text{Eq. 6.15}$$

6.4.4 Mechanism of QQPPM

Based on the principles we have discussed in the previous Chapter 2, and 4, we have obtained the fundamental components to compose the QQPPM. Before we generate the QQPPM, several technical modifications should be conducted. The details are:

- Since the matrix Q is the product of categorized matrices Qc, Qn, and Qs, and all of these matrices are composed by attributed type-related sub-matrices Tb, Tt, Tm, and Tp, it is essential to adjust the order of these sub-matrices. Recall the rules expressed in the figure 6.5, after all Qc, Qs, and Qn are available , we re-arrange the order for sub-matrices Tb, Tt, Tm, and Tp, as well as for Qc, Qs, and Qn.
- As Qc consists of the elements that are the categorized client-site QoS and data processing capabilities, we keep all of their elements as original.
- For Qn, we need to change all elements' values. Qn should consist of the reciprocal of original elements in Qn, because these elements refer to the categorized network-related QoS current upper bound conditions and limitations.

- For Q_s, we do the same modification as Q_n. Q_s also consists of the reciprocal elements of the server-based queried object's categorized QoS requirements.
- After Qc, Qs, and Qn are obtained, the order of Tbc, Tbs, and Tbn are adjusted to their largest dimensions. Similarly, the order of Ttc, Tts, and Ttn, the order of Tmc, Tms, and Tmn, as well as the order of Tpc, Tps, and Tpn are adjusted by means of the method used for adjusting the order of Tbc, Tbs, and Tbn.
- The dimension of submatrix is not fixed. For each case, it is depended on the quantity of attributed entries. Accordingly, the dimension of matrix is dynamic.

After completing modifications, the form of Q should be as the Equation 6.16.

$$Q=QcXQsXQn=\begin{bmatrix} T_{bc} & 0 & 0 & 0 \\ 0 & T_{tc} & 0 & 0 \\ 0 & 0 & T_{mc} & 0 \\ 0 & 0 & 0 & T_{pc} \end{bmatrix}\begin{bmatrix} T_{bs} & 0 & 0 & 0 \\ 0 & T_{ts} & 0 & 0 \\ 0 & 0 & T_{ms} & 0 \\ 0 & 0 & 0 & T_{ps} \end{bmatrix}\begin{bmatrix} T_{bn} & 0 & 0 & 0 \\ 0 & T_{tn} & 0 & 0 \\ 0 & 0 & T_{mn} & 0 \\ 0 & 0 & 0 & T_{pn} \end{bmatrix}$$

$$= \begin{bmatrix} q_{c1} & 0 & 0 & \cdots & 0 \\ 0 & q_{c2} & 0 & \cdots & 0 \\ 0 & 0 & q_{c3} & \cdots & 0 \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & q_{cm} \end{bmatrix}\begin{bmatrix} 1/q_{s1} & 0 & 0 & \cdots & 0 \\ 0 & 1/q_{s2} & 0 & \cdots & 0 \\ 0 & 0 & 1/q_{s3} & \cdots & 0 \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & 1/q_{sm} \end{bmatrix}\begin{bmatrix} 1/q_{n1} & 0 & 0 & \cdots & 0 \\ 0 & 1/q_{n2} & 0 & \cdots & 0 \\ 0 & 0 & 1/q_{n3} & \cdots & 0 \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & 1/q_{nm} \end{bmatrix}$$

Eq. 6.16

Finally, each entry of Q is adjusted as following: if it less than one, elements keep their original values, otherwise change the value to one (see Eq. 6.17). The purpose of adjusting the elements is to guarantee that client can receive the queried object. The value of element less than one means client's particular QoS attribute can not meet the queried object's transmission. The value of element not less than one means its corresponding QoS attribute is good enough to meet the application.

$$\mathbf{Q} = \begin{bmatrix} q_{11} & 0 & 0 & . & . & & 0 \\ 0 & q_{22} & 0 & . & . & . & 0 \\ 0 & 0 & q_{33} & . & . & . & 0 \\ & & . & . & . & & \\ & & . & . & . & & \\ & & . & . & . & & \\ 0 & 0 & 0 & . & . & . & q_{mm} \end{bmatrix} \quad \text{Where } q_{ii} = \begin{cases} 1 & \text{if } q_{ii} > 1 \\ q_{ii} & \text{if } q_{ii} \leq 1 \end{cases} \quad i = 1,\ldots, m$$

Eq. 6.17

The objective of constructing the QQPPM is to select the smallest element .. which is the reduced factor as a criterion to customize the queried object's original properties. To obtain . from Q, the following equation 6.18 could be applied.

$$\alpha = \min\{q_{11}, q_{22}, \ldots, q_{mm}\}$$

Eq. 6.18

The value of α is greater than zero and not larger than one.

The overall of the construction of QQPPM could be summarized in the following pseudocode:

Build matrix of *Qc*, which is categorized client-site QoS

Build submatrices of *Qbc*,*Qtc*,*Qmc*,and *Qpc*, which are the attributed types of bandwidth, timeliness, mobility, and reliability involved in *Qc*.

Build matrix of *Qs*, which is categorized server-site QoS

Build submatrices of *Qbs*,*Qts*,*Qms*, and *Qps*, which are the attributed types of bandwidth, timeliness, mobility, and reliability involved in *Qs*.

Build matrix of *Qn*, which is categorized network QoS

Build submatrices of *Qbn*,*Qtn*,*Qmn*,and *Qpn*, which are the attributed types of bandwidth, timeliness, mobility, and reliability involved in *Qn*.

Adjust the dimensions of all submatrices

Adjust the dimensions of submatrices *Qbc*, *Qbs*, *Qbn* to the same dimensions.
Adjust the dimensions of submatrices *Qtc*, *Qts*, *Qtn* to the same dimensions.
Adjust the dimensions of submatrices *Qmc*, *Qms*,*Qmn* to the same dimensions.
Adjust the dimensions of submatrices *Qpc*, *Qps*, *Qpn* to the same dimensions.

Construct the matrices *Qc, Qs*, and *Qn* as categorized QoS profiles

Construct of matrix of *Qc* using adjusted *Qbc, Qtc, Qmc*, and *Qpc*.
Construct of matrix of *Qs* using adjusted *Qbs, Qts, Qms*, and *Qps*.
Construct of matrix of *Qn* using adjusted *Qbc, Qtc, Qmc*, and *Qpc*.

Modify the entries of *Qs* and *Qn*

Change all entries of *Qs* as reciprocal of their original elements.
Change all entries of *Qn* as reciprocal of their original elements.

Multiple *Qc, Qs*, and *Qn* to form the QQPPM

Select the smallest elements as the criteria to refine the original object if possible.

6.5 Examples of QQPPM Applications

The typical application of the Matrix Q could be explained as the following steps:

1. Client-site limitations, network delays and server-based object properties consist of corresponding attributed sub-matrices. After adjusting the orders, these sub-matrices compose the categorized matrices Q_c, Q_n and Q_s, respectively;
2. The product of Q_c, Q_n and Q_s consists of Q. The crucial elements such as the multivariate timeliness and the available display capacities are composed of Q, and the total elements accumulate the order of Q.

6.5.1 Case Study 1

Hereafter we present a real world example. Suppose a client uses current most advanced PDA (e.g. iPAQ Pocket PC model h4155) to access remote database server for querying PDA iPAQ h4155 picture (see figure 6.9) through its integrated 2.4Ghz 802.11b WLAN. For this particular case, the client-site QoS profile, server-based QoS profile, and wireless network QoS profile sets are described in the table 6.1.

Figure 6.9: Example of PDA Packet PC (Sited from Bestbuy.com)

Client-side QoS Profile		Server-based QoS Profile		Wireless network QoS Profile	
Upper Bound Display Dimension	240X320	Object Dimension	170X190		
Upper Bound Display Capacity	128KB	Object Size	96KB		
Delay Time Bound	30ms			Delay Time	3ms

Table 6.1: Case 1 QoS Profiles

Using the QQPPM construction rules, we construct the attributed type-related submatrices T_{bc}, T_{tc}, T_{mc}, and T_{pc} for Q_c; the submatrices T_{bs}, T_{ts}, T_{ms}, and T_{ps} for Q_s; the submatrices T_{bn}, T_{tn}, T_{mn}, and T_{pn} for Q_n. The following equations are shown all the submatrices

$$T_{bc}=\begin{bmatrix} 240\times 320 & 0 \\ 0 & 128 \end{bmatrix} \quad T_{bs}=\begin{bmatrix} 1/(170\times 190) & 0 \\ 0 & 1/96 \end{bmatrix} \quad T_{bn}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$T_{tc}=\begin{bmatrix} 30 & 0 \\ 0 & 1 \end{bmatrix} \quad T_{ts}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad T_{tn}=\begin{bmatrix} 1/3 & 0 \\ 0 & 1 \end{bmatrix}$$

$$T_{mc}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad T_{ms}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad T_{mn}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$T_{pc}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad T_{ps}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad T_{pn}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Then we calculate related *Qc*, *Qs*, and *Qn*. The following equations are shown all the matrices.

$$Qc=\begin{bmatrix} 240\times 320 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 128 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 30 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Qs=\begin{bmatrix} 1/(170\times 190) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/96 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Qn=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Finally we multiply *Qc*, *Qs*, and *Qn*. The conduct *Q* is obtained. The details are as follows.

$$Q=Q_c X Q_s X Q_n=\begin{bmatrix} 40.42 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.33 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Since all elements are larger than one, the client can retrieve the picture without any refinement.

6.5.2 Case Study 2

Recall the illustrated example of PKI in figure 3.1 and figure 3.2, we use the same method in Case 1 to obtain the attributed type-related submatrices T_{bc}, T_{tc}, T_{mc}, and T_{pc} for Q_c; the submatrices T_{bs}, T_{ts}, T_{ms}, and T_{ps} for Q_s; the submatrices T_{bn}, T_{tn}, T_{mn}, and T_{pn} for Q_n. The details are as follows:

$$Q_c=\begin{bmatrix} DisplayCapacity & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & Timeliness & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Q_s=\begin{bmatrix} 1/MultimediaDataSize & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Q_n=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/TotalDelay & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Then we multiply Q_c, Q_n and Q_s. The result of Q is expressed as:

$$Q=Q_c X Q_s X Q_n=\begin{bmatrix} \frac{DisplayCapacity}{MultimediaDataSize} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{BoundTimeliness}{TotalDelays} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Using Eq. 6.18, the reduced factor of α is obtained as following.

$$\alpha = \min\left\{ \frac{\dfrac{Client-Site}{UpperBoundDisplayCapacity}}{\dfrac{Server-site}{MultimediaObjectDataSize}},\ 1,\ \frac{\dfrac{Client-site}{LowBoundTimeliness}}{TotalDelays},\ 1,\ 1,\ 1,\ 1,\ 1, \right\}$$

We apply the above form to calculate the reduced factor as following:

$$Q = Q_c X Q_s X Q_n = \begin{bmatrix} \frac{128\times128pixels}{490\times269pixels} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{60ms}{120ms} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Therefore, we get α = 0.12. After the entire original object's QoS being reduced to 12%, we can solve the problems occurred in client site. The effects of QQPPM and related α have been shown in figure 3.1. When the object is reduced to 12% of the original, the query processing is successfully conducted.

6.6 QQPPM Summary

Based on the QoS conditions which come from server-site actual requirements, real time network services provided, and client-site real time considerations, the QQPPM is constructed and the smallest element of the QQPPM is calculated as the reduced factor α. The original queried multimedia object, if it cannot be presented due to the client's QoS requirements restriction and device capability limitations, should be reduced the quality and should be refined. The criterion of the reduction and refinement is exclusively based on the reduced factor α. The query processing of UNO PKI's picture, as a real world sample, demonstrates the function of reduced factor α.

7. EXPERIMENTAL RESULTS FROM OPNET SIMULATIONS

As expressed in this thesis, query processing for MM-DBMS in wireless mobile environments is still a topic as research stage. At current time, there is neither commercial MM-DBMS [EN04] nor real wireless network facility that allows us to implement the QQPPM approach. Therefore, in order to support that the conception of QQPPM is reasonable and practicable, we design the experiments for this thesis focusing on examining the response time and throughput. Due to the limited bandwidth provided by wireless networks conflicting with extraordinary resources required by multimedia queried object's transmission, the above mentioned network vital characteristics, client's querying response time and client-site wireless network throughput, should be analyzed. We conduct the simulations of transmitting different real-world multimedia data sizes with different infrastructure densities under wireless mobile environments. With proposed network QoS services, we investigate the relationships among scaleable number of users, multimedia data sizes, corresponding systems response time and network traffics. The wireless networking facilities for this thesis is the OPNET Modeler 10.5 version with Wireless LANs (IEEE802.11) and UMTS (a 3G wireless protocol). Both of WLAN and UMTS models are IP-based wireless networks. We design and construct WLAN and UMTS models with different scenarios to conduct our simulations.

7.1 Wireless LAN Model with Multimedia Database Query Application

The Wireless LAN (WLAN, IEEE802.11) model, which is illustrated on Figure 7.1, contains wireless multimedia database query clients and a wireless multimedia database server. The clients and server belong to different WLANs. These two WLANs connected to each other with two routers. The purpose of designing and constructing this model is to demonstrate the transmission results of different multimedia object sizes queried by mobile database clients with different user quantities under the WLAN environments.

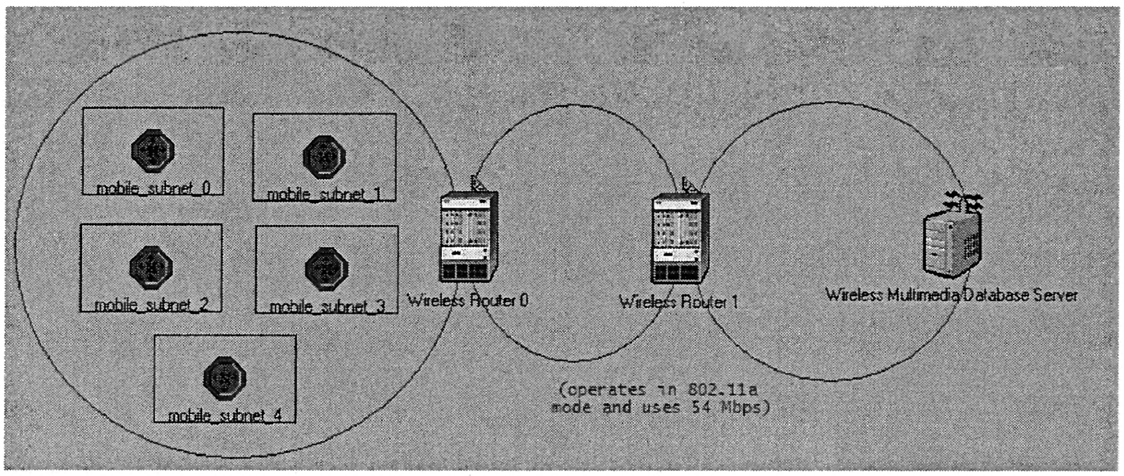

Figure 7.1: The architecture of WLAN model

7.1.1 WLAN Model Properties

The five mobile subnets each contain from 1 to 15 wireless stations. All stations comply with the WLAN protocol. The clients in the WLAN are trying to communicate with servers at the remote site via wireless routers. The clients' main applications are querying of different large size multimedia objects through wireless database server. Each workstation has only database application.

As IEEE802.11 standard definition, WLAN systems are based on best-effort. The QoS profile of the WLAN simulation model has the following attributes:

- Classification: no service differentiation is provided.
- Channel access: CSMA/CA;

- Packet scheduling: FIFO mechanism.
- There is no service guarantee in terms of bandwidth and delay.

The simulation test bed's facilities are as following:

- CPUs used in workstations, IP routers, and Servers are Sun Ultra 10 333 MHz;
- All Operating Systems used in whole model is Solaris.
- Workstations use 11 Mbps WLAN to communicate with wireless router;
- Two wireless Routers use bandwidth of 54Mbps to contact with each other;
- The server uses 11Mbps to connect with wireless router.

For the simulation of multimedia database query processing, we design 16 different scenarios, in which the client number are increased from 5,15, 30 to 75, the multimedia object sizes are transported from 40kB, 100kB, 400kB to 1000kB with Poisson distributions. All the scenarios simulate 10 minutes real world WLAN situations.

7.1.2 WLAN Simulation Results and Analysis

Under the unchangeable circumstances of QoS conditions, the same multimedia object query applications, and the same network's properties, the database query response time and the related client-site network throughput are dependent on the size of the queried object as well as the number of clients. This is the conclusion that we reach from the WLAN simulation results: increment of the size of an object is not unlimited. Meanwhile, increasing the number of clients with the same network facility is limited.

When WLAN has small clients density and clients query for small object sizes, the database query response time changes little and is acceptable. Otherwise, even small client number query for big object size, the change is significant and the response time is huge high. The simulation results are shown in table 7.1:

Clients Number / Average Response Time (sec.) / Object Size (kByte)	5	15	30	75
40	0. 255	0. 323	0. 8	1. 462
100	0. 481	0. 799	1. 964	113. 863
400	2. 168	22. 0722	134. 165	250. 611
1000	12. 096	108. 787	208. 292	319. 947

Table 7.1: WLAN Client Number and Object Size vs. Average Response Time

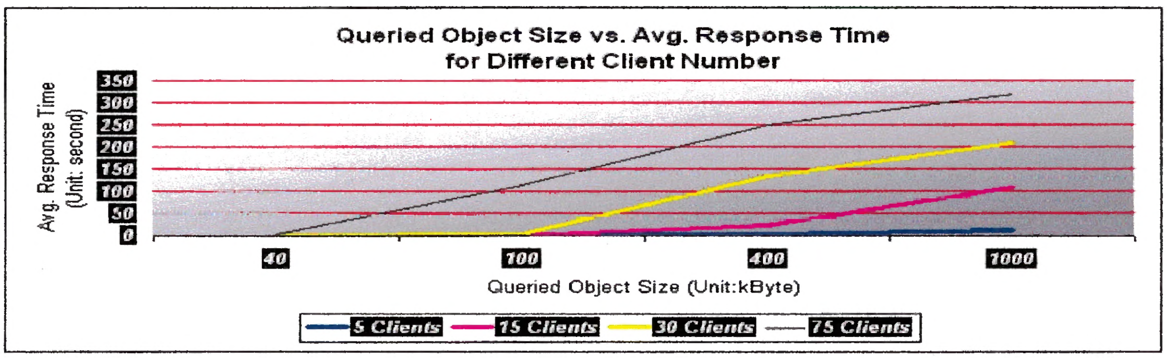

Figure 7.2: Object Size vs. Response Time in WLAN with different Client Number

Figure 7.2 illustrates the relationships among client density, queried object size and average response time through the comparison of 16 different scenarios.

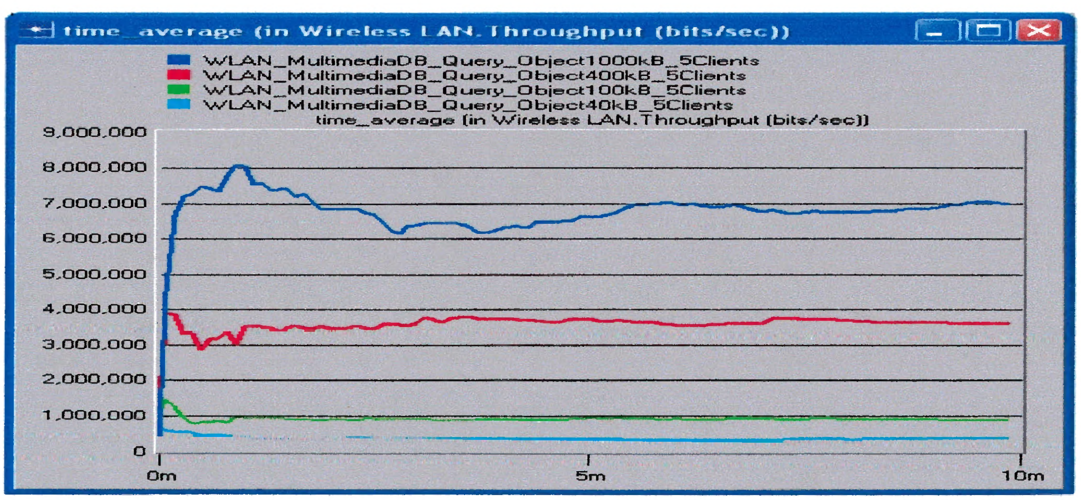

Figure 7.3: Object Sizes vs. Throughput in WLAN with same Client Density

Figure 7.3 illustrates the relationship between queried object size and client's throughput. The simulations demonstrate that even in a small client density situation, the bigger the size of queried object is, the higher the wireless network throughput the clients require to transport the data.

Through the WLAN simulations, we have had some important observations. First, the sizes of queried object can not unlimited increased, even the client density keeps very small quantity. Second, when the queried object size gradually increases and at main time client number keeps high quantity, the response time of querying for each client is over hundred second. The reason is that the current WLAN technologies can not handle multimedia heavy traffic with the existing QoS capacities. The solution is to reduce the queried object size as small as possible. Therefore, the proposed QQPPM could be a choice for conducting multimedia object query processing for MM-DBMS under WLAN environment.

7.2 UMTS Model with Multimedia Database Query Application

The UMTS model illustrated on Figure 7.4 consists of three main entities:

- 3G wireless Networks, in which five Mobile Stations Groups (constructed with mobile user equipments (UEs), in this model we have 5, 10, or 20 UEs for each group) communicating with NodeB_DCs. NodeB_DCs are controlled by the Radio Network Controller (RNC_DC). Each particular NodeB_DC controls different radio channel within its coverage area or cell;

- UMTS Core Network includes two network nodes: the serving GPRS support node (SGSN) and the gateway GPRS support node (GGSN). There are four routers connecting these nodes;
- Through IP Network, the UMTS users connect their remote database server.

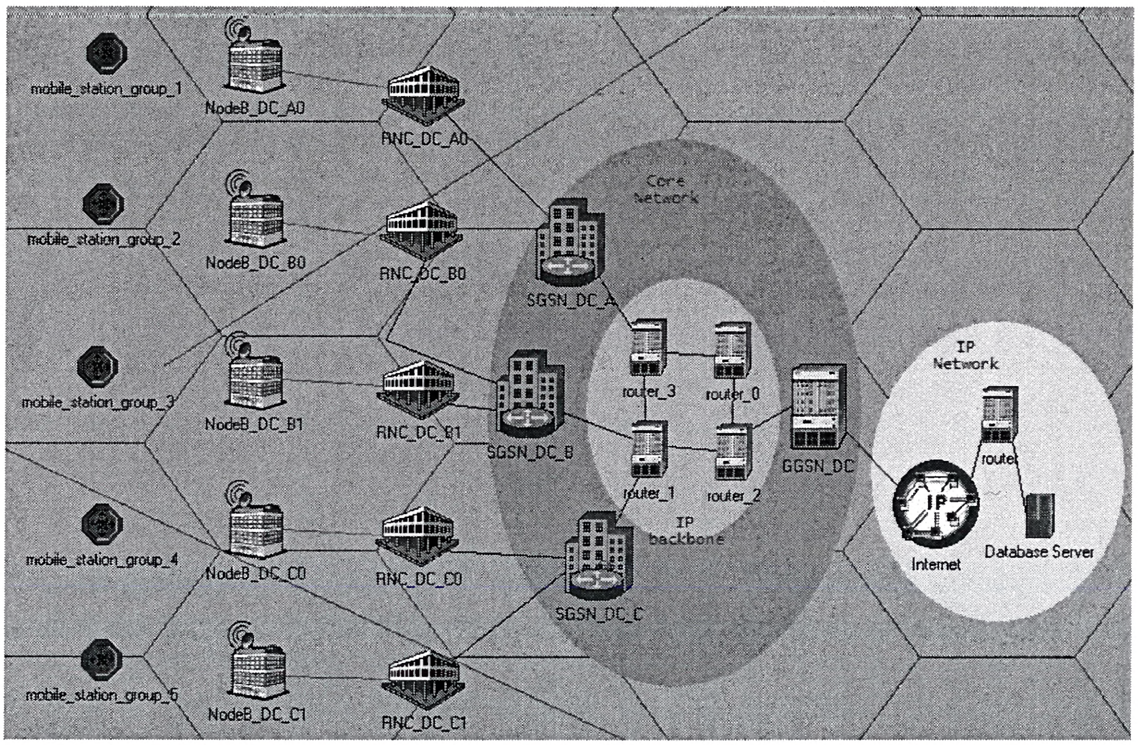

Figure 7.4: The Architecture of UMTS Model

7.2.1 UMTS Model Properties

In UMTS wireless networks, similar to WLAN, all the CPUs used in workstations, NodeB, SGSN, GGSN, routers, and Servers are Sun Ultra 10 333 MHz. A Solaris is the Operating System used in the model. There are five UMTS_NodeB groups. Each group has 5, 10, or 20 mobile workstations (called UE) which have the applications of multimedia database querying. UEs use Wideband Code Division Multiple Access (W-CDMA) access scheme to connect with NodeBs. The peak data rate is up to 2Mbps. The bandwidths among UMTs_NodeBs, Core Network, and IP network are ATM OC3, which has the data transmission capacity of 155Mbps. Among the IP network, there is a router connecting with multimedia database server. The cable is 10BaseT with data rate of 1.5Mbps.

We design the QoS for UMTS model as the Interactive class (see table 7.2)

UMTS Interactive Class Key QoS Service Attributes	
Max. bit Rate (Kbps)	2048
Delivery Order	Yes
Max.Data Size(octets)	1500
Error ratio	1.00E-05

Table 7.2: UMTS QoS Profile

Note that the applications of this model request data from remote server for multimedia data retrieval. The Interactive traffic is characterized by the request-response time delay pattern of the end user. For the simulation of multimedia database query processing, we design 15 real world scenarios to investigate the relationships among client density, queried object size, and response time.

The multimedia object sizes are transported from 40kB, 70kB, 100kB, 150kB, and to 200kB with Poisson distribution. The client densities are increased from 25, 50 to 100 client numbers. All scenarios do simulation of running 5 minutes real world database query under UMTS environment.

7.2.2 UMTS Simulation Results and Analysis

From the UMTS simulation's results, we reach a crucial conclusion that within the finite set of QoS schemes, the querying response time is significantly dependent on the queried data size and is not sensitively related to the augmenting of client number. Under current UMTS technology, the multimedia object size for query processing can not be exceeded to over 100kBytes, otherwise the processing time, especially the wireless network transaction time, is over 2 minutes that is intolerable for real world mobile clients. Meanwhile, the test samples between 50 clients and 100 clients do not affected the results. The phenomenon is shown the necessity of developing the queried data precision mechanism. It is once again demonstrated that QQPPM approach is reasonable. Hereafter is the statistics (see figure 7.5) expressed the relationship among multimedia queried object size, system query processing response time, and client density under same UMTS QoS environment.

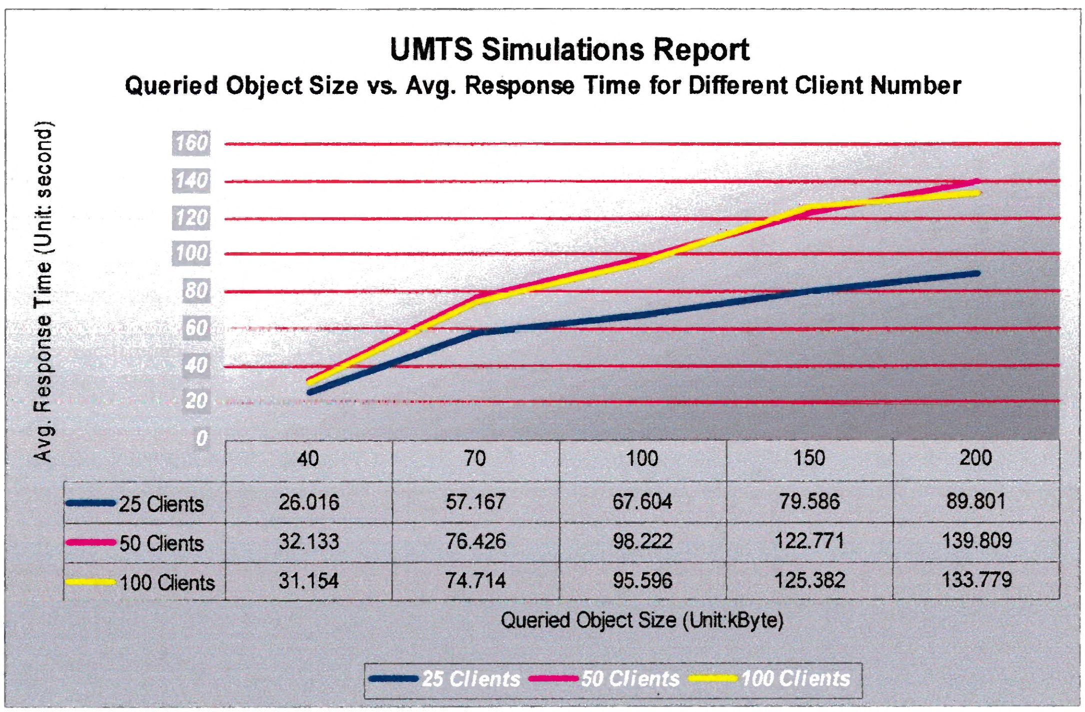

Figure 7.5: UMTS Object Size vs. Average Response Time under Different Client Density

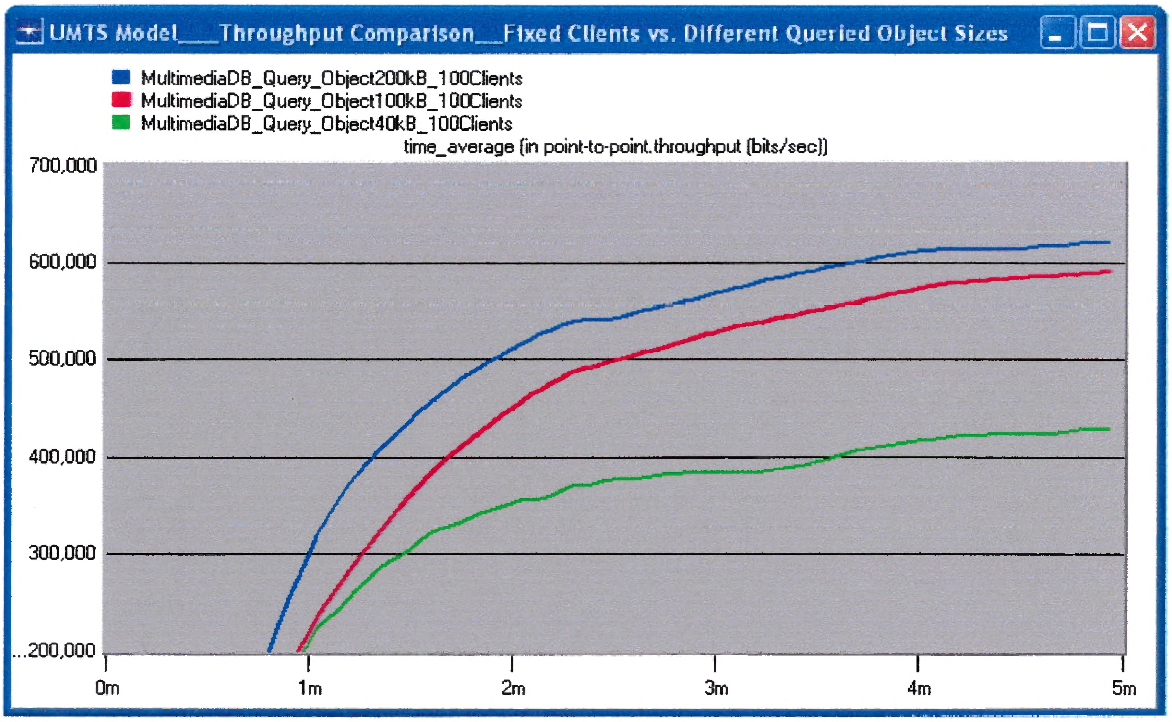

Figure 7.6: Object Sizes vs. Throughputs in UMTS with same Client Density

Figure 7.6 illustrates the relationship between queried object size and the clients' average throughput. With the same client density, the figure shows the bigger the size of queried object is, the higher the wireless network throughputs those clients require

7.3 OPNET Simulation Summary

For the simulations involved in this thesis work, we select WLAN, which is a mature and popular technology applied in today's telecommunication market, and UMTS, which is a Third Generation (3G) wireless protocol and is part of the International Telecommunications Union's IMT-2000 vision, widely accepted by over 150 nations [GGW04]. We have obtained the fundamental observations as follows:

1. The QoS support services in the network system can realize not only the traditional data transmissions but also the multimedia data applications. Moreover, the QoS can support today's multimedia wireless and mobile networks.

2. The impact of large multimedia object size on the current wireless network technologies exists. Currently, QoS only supports limited pre-defined environments according to particular application characteristics. To process the data with large size, we have to wait for the advent of more advanced technology. QoS itself can not solve this issue.

3. The network traffic in both WLAN and UMTS are sensitive with queried object sizes. WLAN and UMTS take different channel access mechanisms. While client quantity is small, WLAN transmits higher data sizes than UMTS does. However, UMTS has better scaleable client sizes than WLAN has.

4. The simulation results have verified an assumption made in this thesis: in the multimedia wireless and mobile database query processing, the time issue can be solved by reducing the queried object size.

Regarding the current OPNET modular we have on hand, we have to limit our simulations under some restrictions. There are not available functions in WLAN and UMTS models to simulate the mobility and power control. All the QoS conditions are following the default values setup by OPNET 10.5 version modular, which was released on May 13, 2004.

8. CONCLUSIONS AND FUTURE WORK

8.1 Summary

As the technologies and applications related to wireless communications and digital media are phenomenal developing and growing, these two industries are continuously listed in most worldwide academic agencies as annual top ten trends during recent years. However, to successfully combine these two technologies still challenge the research communities. In this thesis, we have studied multimedia database query processing from a QoS perspective in wireless and mobile networks, focusing on transmission and presentation of multimedia queried object which has a out of prospective large size. To explore the problems and the related reasons, we investigated the current framework of QoS and identified the QoS issues existing in wireless mobile environments. We noted that QoS can not be set for all networks and all applications, and QoS is determined as a subset of required parameters for each network and each application. However, based on our evaluation presented in this thesis, we have found that general principles of QoS are challenged in the context of mobile multimedia database query processing. The majority of the research approaches to solve this problem focus on intensifying the empirical or statistical QoS requirements and stop querying while the problem happens.

Taken into considerations of keeping the current QoS profiles and exploring the efficient approach to execute the query transmission regardless QoS conditions, we analyzed and constructed MM-DBMSs' client-site QoS considerations set, server-site

QoS requirements set, and network QoS services set. Meanwhile, we designed a QoS-based Query Processing Precision Matrix (QQPPM) to create a reduced quality strategy on the purpose of tradeoff the server's original data quality and actual quality that client can reach. We study the feasibility of the solution, and combine multidisciplinary technologies to support the methodology. The related techniques include multimedia data modeling, MM-DBMS query languages, multimedia data compression, multimedia object storage schemas, wireless and mobile QoS in MM-DBMS, and wireless Network QoS protocols. Furthermore, we constructed and performed multimedia database different object sizes querying simulations with different client densities under wireless LAN and UMTS platforms on the OPNET 10.5 modeler.

8.2 Conclusions

Just as David Oran, a Cisco Fellow and a pioneer in Voice over IP technology said that multimedia applications and quality of service is particularly welcome since it ties these technologies into the overall picture of the state of the art. We have reached the following important conclusions from this thesis:

- Multimedia applications in wireless and mobile environments are fully dependent on QoS technologies.
- Multimedia object retrieval needs resource consumption often exceeds available resource capacities of the deployed wireless/mobile networks and portable devices.
- The precautions of extra resource requirements can not be taken by server, client, even network infrastructure in advance. We need QoS management in wireless

mobile environments to specify a range of acceptable QoS levels to allow for scaling of multimedia query processing, rather than trying either to guarantee specific values or to stop the querying.

- The model of QQPPM is reasonable and practicable. The theoretical studies and experimental practices support its possibility.

8.3 Contributions

The main contributions of this thesis are as follows:

- We critically analyzed multimedia database query characteristics for the QoS point of views.
- We studied and extended standard QoS mechanism on the context of wireless mobile environments. The client consideration elements such as hardware facilities, user preference, mobility coverage, and critical quality acceptance are included in the extended QoS framework.
- Rather than using statistical resource utilizations, we combined with all client-site QoS considerations, server-site QoS requirements, and network QoS services into one simple matrix. Based on balance of all the QoS elements in the matrix, we calculated the precision criterion factor and applied it to process the queried object transmission.
- We conducted real world wireless multimedia database query processing simulations with different queried object sizes and different client densities on traditional wireless network and state-of-arts 3G wireless network, and evaluated the relationship among object size, response time, and client densities.

8.4 Future Work

QoS plays a crucial role to solve the MM-DBMS query processing problem in the wireless mobile environments. Therefore, given the basic extended QoS framework, querying processing precision solution, fundamental theories and techniques discussed in this thesis, future work will concentrate on the following further three research directions:

- Enhancing MM-DBMS query languages scheme added with QoS-based QQPPM mechanism, since there are not available query languages in commercial that have the functions to process wireless and mobile multimedia data query;
- Combining related techniques discussed in figure 2.1 into a united platform, specially the mobile multimedia data compression technology. This is the essential milestone to fulfill the functions described above.
- Improving the QQPPM algorithm in this thesis. We have just presented the simple solution to construct the QQPPM, there are possible other optimal approach to be explored.

At the end of this thesis paper, let me cite UNO Professor Stanley A.Wileman's words as the primary purpose of my thesis:

> *Quality of Service (QoS) is a term that's often misunderstood, understandably :-) It's a broad area, and research that sheds more light on understanding it should be of significant value.*

REFERENCES

[AZ96] Gennady Antoshenkov, Mohamed Ziauddin Query Processing and Optimization in Oracle Rdb. VLDB Journal 5, 4, 229-237, 1996

[Bar99] Daniel Barbara. Mobile Computing and Databases --- A Survey. *IEEE Transactions on Knowledge and Data Engineering, Vol. 11, no.1,* January/February 1999

[BH97] G Bochmann and A Hafid. Some Principles for Quality of Service Management, Distributed Systems Engineering, Vol. 4, *IOP Publishing*, 1997

[BK03] R. Braumandl and A. Kemper. Quality of Service in an Information Economy, *ACM Transactions on Internet Technology*, Vol. 3, No.4, pp. 291-333, November 2003

[BDW02] Behzad Bordbar, John Derrick, Gill Waters. Using UML to specify QoS constraints in ODP. *Computer Networks. Vol. 40, no. 2, pp. 279-304.* Oct 2002.

[Cgh03] Guohong Cao. Integrating Distributed Channel Allocation Adaptive Handoff Management for QoS-Sensitive Cellular Networks. *Wireless Networks 9,131-142,* 2003.

[CMP01] Paolo Ciaccia, Danilo Montesi, Wilma Penzo, and Alberto Trombetta. Fuzzy Query Languages for Multimedia Data. Book Chapter in Design and Management of Multimedia Information Systems: Opportunities and Challenges, M.R. Syed Ed. *Idea Group Publishing* 2001.

[CZ04] JYH-Cheng Chen, Tao Zhang. IP-Based Next-Generation Wireless Networks: Systems, Architectures and Protocols. *John Wiley* 2004

[Doc03] Database Trends for Web Design, Security, and Mobile Apps: what are database developer's priorities in designing and managing their applications and information? *Doc # 11845, ADVISOR MEDIA, Inc.* February 2003

[EGP02] Denise J. Ecklund, Vera Goebel, Thomas Plagemann, Earl F. Ecklund, Jr. Dynamic end-to-end QoS management middleware for distributed multimedia systems. *Multimedia Systems 8: 431-442*, 2002

[EN04] Ramez Elmasri and Shamkant Navathe. Fundamentals of database Systems 4th Edition. Pearson Education, Inc. 2004

[Fla00] Guntram Flach. Using Object-Relational database Systems and XML in the Context of Mobile Environments. *NEC Research Institute* 2000, retrieved on July 31, 2003 from http://citeseer.nj.nec.com/correct/457643

[GGW04] Aura Ganz, Zvi Ganz, and Kitti Wongthavarawat. Multimedia Wireless Networks Technologies, Standards, and QoS. *Prentice Hall PTR* 2004

[Hil02] Mikael Hillborg. Wireless XML Developer's Guide. *McGraw-Hill/Osborne* 2002.

[HRM03] Timo Halonen, Javier Romero, and Juan Melero. GSM, GPRS and Edge Performance: Evolution Towards 3G/UMTS. *John Wiley* 2003

[IFF99] Ives, Z., Florescu, D., Friedman, M. An Adaptive Query Execution Engine for Data Integration. In Proceedings of the ACM SIGMOD Conference on Management of Data. Philadelphia, PA, 299-310, 1999

[ITU02] International Telecommunication Union. *World Telecommunication Development Report 2002: Reinventing Telecoms.* March 2002, retrieved from http://www.itu.int/ITU-D/ict/publications/wtdr_02/material/WTDR02-Sum_E.pdf

[Jam03] Abbas Jamalipour. The Wireless Mobile Internet: Architectures, Protocols, and Services. *John Wiley & Sons Ltd.* 2003

[Kaz02] Matheos Ioannis Kazantzidis. Adaptive Multimedia in Wireless IP Networks. *UCLA PhD Dissertation* 2002

[KD98] Kabra, N. and DeWitt, D. Efficient mid-query Re-optimization for sub-optimal query execution Plans. *In Proceedings of the ACM SIGMOD Conference on Management of Data.* Seattle, WA, 106-117, 1998

[Lu99] Guojun Lu. Multimedia Database Management Systems. *Artech House, Inc.* 1999

[ME01] Jim Melton, Andrew Eisenberg. SQL Multimedia and Application Packages (SQL/MM). SIGMOD Records, 30(4), 2001, retrieved on July 29, 2003 from http://www.acm.org/sigmod/record/issues/0112/standards.pdf

[MK03] Sandor Molnar, Kin Leung, Guest Editorial. *Mobile Networks and Application* 8, 2003

[MS00] Indu Mahadevan, Krishna M. Sivalingam. Architecture and Experimental Results for Quality of Service in Mobile Networks using RSVP and CBQ. Wireless Networks 6, 221-234, 2000

[MT02] Ika Miloucheva, Sandra Tartarelli. QoS Roadmap, *Next Generation Networks Initiative Consortium,* Oct 2002.

[RG03] Raghu Ramakrishnan, Johannes Gehrke. Database Management Systems 3rd ed. *McGraw Hill Higher Education* 2003

[Sau02] Jeff Sauer. Mobile Media: the future is now. *Econtent*, June 2002

[SE02] Jurgen Stauder, Fazli Erbas. Mobile Multimedia over Wireless Network Chapter XIX, Multimedia Networking: Technology, Management and Applications/edited by Syed Mahbubur Rahman. *Idea Group Publishing* 2002

[Shi02] Timothy K. Shih. Distributed Multimedia Databases: Techniques & Applications. *Idea Group Publishing* 2002

[Sti01] Timothy K. Shih. An Introduction to Multimedia Database. *Idea Group Publishing*, 2001

[Thi98] HeikoThimm. Optimal Quality of Service under Dynamic Resource Constraints in Distributed Multimedia Database Systems. *Ph.D. Thesis, University Darmstadt*, Jun 1998.

[UFA98] Tolga Urhan, Michael J. Franklin, and Laurent Amsaleg Cost-based Query Scrambling for Initial Delays In Proceedings of ACM SIGMOD Conference on Management of Data, Seattle, WA 130-141, 1998

[UMT00] UMTS. Enabling UMTS/Third generation services and applications. *No. 11 Report from the UMTS Forum,* Oct 2000.

[Wat04] Richard Watson. Data Management: Databases and Organizations. *John Wiley* 2004