

AQuaFWiN: Adaptive QoS Framework for Multimedia in Wireless Networks and its Comparison with other QoS Frameworks

Bobby Vandalore, Raj Jain, Sonia Fahmy*
Department of Computer and Information Science
The Ohio State University, Columbus, OH, USA
E-mail: {vandalor,jain,fahmy}@cis.ohio-state.edu

Sudhir Dixit
Nokia Research Center
Burlington, MA 01803, USA
E-mail: dixit@research.nokia.com

Abstract

In a wireless environment, due to topology changes and characteristics of media (interference in radio signals) the bandwidth of a link is unpredictable and possibly very low, the error rates are variable and extremely high. Provisioning and guaranteeing quality of service (QoS) in such an environment is a very challenging problem. In this paper we propose an adaptive QoS framework to support multimedia applications in a wireless networking environment. The proposed framework is hierarchical in nature with cluster of mobile end hosts connected to a base station, base stations are connected to a supervisory node, which in turn is connected to the wired infrastructure. The changing conditions in wireless due to interference and possibly mobility, entitle that the real-time applications needing stringent QoS should be adaptable. The framework uses a generic feedback mechanism to support adaptability at all layers of the wireless network.

An overview of existing wireless architectures which support QoS is given. The architectures discussed are WAMIS (wireless adaptive multimedia information system), SWAN (seamless wireless ATM network), MMWN (multimedia support for mobile wireless networks), and QGMC (QoS guarantees in mobile computing). The proposed approach is compared with these existing methods.

1. Introduction

Cellular phone and paging services have been very successful. It is projected that by year 2000, the number of cellular telephones would be equal to 40% of the fixed-line telephones in the United States (compared to only 20% in 1995). Wireless data networks have developed relatively slow, though portable computers are the fastest growing component of the PC market. It is expected that by 2000

an amount of \$50 billion will be spent on portable computers [12]. The reasons for slow development of wireless data are complex pricing, poor performance (wide-area data transfer is only 20 Kbps), lack of killer application and form factor (today's laptop computer are too large to use on the move). Supporting multimedia in wireless environment might be the killer application needed to popularize wireless networks.

There are several challenges to be addressed for supporting QoS in a wireless network. The changing environment in wireless due to mobility and interference gives rise to varying bandwidth (possibly low) and dynamically changing error (possibly high) rates. Traditional networking assumptions such as fixed topology and fixed amount of available resources do not hold in a wireless mobile environment.

Several techniques have been proposed to support multimedia at different layers of networking. At the application layer, the real-time applications can be made adaptive to changing networking conditions. At the transport layer resource reservation can be made during connection setup to support end-to-end QoS guarantees. At the network layer techniques to provide mobility management and seamless connectivity can be used. Routing mechanism need to be QoS-aware and handle mobility. At the data link layer medium access control has to be modified so that reservations are respected and QoS guarantees can be supported. Adaptive power control techniques can be used to manage mobility and maintain active links. Error control techniques can be used to protect against varying error rates. At the physical layer several options ranging from infrared to radio is available.

An adaptive framework is proposed to provision and support QoS in a wireless network in section 2. An overview of representative class of existing architectures are given in subsequent sections. In section 8 a comparison of the proposed QoS framework and existing architectures is given. Finally, summary and conclusions are given in section 9.

*Sonia Fahmy is now with Department of Computer Sciences, Purdue University, Indiana

2. AQuaFWiN Overview

In this section we give an overview of the proposed adaptive QoS framework to support multimedia in a wireless network (AQuaFWiN). To provide QoS support in a dynamic environment, adaptability and QoS-awareness has to be supported at every layer of the network. This is the guiding principle followed in the proposed framework. In the subsequent subsections we discuss the network architecture and the various components in the proposed framework. To enable adaptability an end-to-end feedback mechanism is used. This is achieved by periodically transmitting control packets which gather feedback information that can be used at different layers.

The various features and components of the AQuaFWiN are:

- *Adaptive Applications:* Real-time applications such as voice and video need to be adaptive.
- *Network Architecture:* The network architecture should provide connectivity to wired infrastructure for scalability. A hierarchical network architecture consisting of mobile hosts, cluster of mobile hosts, base station and fixed network is used in the framework.
- *Adaptability:* To achieve adaptability every layer needs information about the current network conditions. This can be provided by end-to-end feedback. A generic feedback mechanism is proposed that can be used at different layers to achieve adaptability.
- *Flexible QoS:* In traditional networks the QoS specified is rigid. We propose flexible QoS parameters which can be used during connection admission and for supporting QoS in mobile wireless environments.
- *Separation of Wireless and Fixed network:* The fixed network should be transparent to the changes of mobile wireless part of the network. This separation is provided by using “supervisor nodes” in the proposed network architecture.

3. Components of AQuaFWiN

In this section the different components of AQuaFWiN are discussed. A simulation model of this QoS framework needs to be build to study the performance and design trade-off available at different network components.

3.1. Adaptive Applications

Applications need to be QoS-aware and adaptive. Real-time applications such as voice and video can be made adaptive. Voice and video applications are invariably encoded

(compressed) before transmission. One way of adapting to changing environment is to dynamically modify the encoding parameters.

Voice applications can be transmitted using a bandwidth which can range from 8 Kbps to 128 Kbps. Based on the network conditions appropriate encoding mechanism/parameters can be chosen for voice applications.

Video applications can be made adaptive by using the layered transmission model [7, 13]. In this model, video information is encoded into several layers of decreasing priority. In the lowest (base) layer critical timing and essential image information are transmitted. Additional layers provide increasingly better quality. Adaptation is achieved by discarding (at intermediate nodes) the lower priority layers in the event of bandwidth changes due to mobility and/or interference.

Another promising approach for adaptability is to use wavelet encoding instead of traditional discrete cosine transformation (DCT) compression method. As argued in [3] the motion compensation used in traditional compression methods such as MPEG-2 [1] suffer due to varying error rates and changing bandwidth of a wireless environment. The wavelet-based coding uses intraframe coding and less of motion compensation. A highly scalable and rate controlled wavelet-based video coder for very low bit rate can be built using the techniques proposed in [9, 14].

3.2. Network Architecture

The current Internet consists mostly of fixed nodes connected by wired network. The Internet of the future will include wireless mobile nodes connected to wired infrastructure. The Internet architecture is also becoming hierarchical in nature (for example differentiated services and multiprotocol label switching are designed to work in a hierarchical network). The motivation for AQuaFWiN are guided by the above trends. In AQuaFWiN architecture (see Figure 1) nodes are connected in a hierarchical manner. Mobile hosts are grouped into a clusters. Each cluster has a cluster head chosen by a distributed election protocol. Clusters are connected to base stations which provide connection to wired infrastructure and mobility support. A group of base stations are connected to a special node called the “supervisor node” (SN). The SN is used to logically separate the mobile and fixed part of the network. We assume that fixed part of the network can support QoS guarantees. Currently, we plan to use connection-oriented ATM network which provides support for QoS guarantees in our simulation model. A group of supervisor nodes may form a supervisory cluster to better support mobility. Finally, the supervisor nodes are connected to the wired infrastructure.

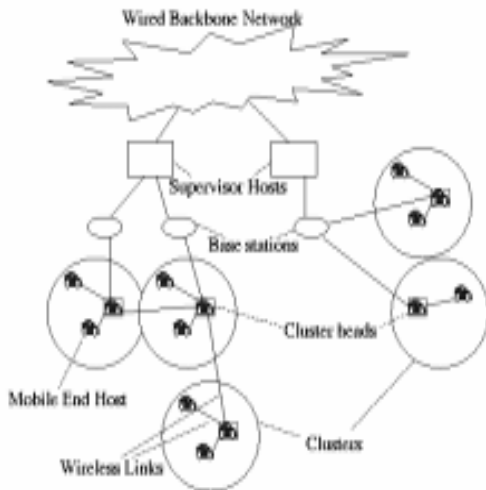


Figure 1. Network architecture of AquaFWin

3.2.1 Clustering

The mobile end hosts are partitioned into clusters. Clustering can facilitate bandwidth management needed for supporting QoS. Clusters are formed dynamically between end hosts. A distributed election mechanism (similar to one discussed in [3]) where the node with lowest ID (identifier) is used to choose the cluster head. We assume that the cluster head can communicate directly to the BSs or can reach the base station through adjacent clusters. Nodes connecting two clusters, and a cluster and a BS act as gateway nodes. In the event when the cluster head moves, it assigns the node with next lowest ID as the temporary cluster head. The temporary cluster head then starts the election process to select the new head.

3.2.2 Connection and Mobility Management

When a connection setup is requested the connection is established as a virtual circuit. The virtual circuit is split into a fixed part and a mobile part (similar to [16]). The fixed part of connection is established between the fixed network and the SN. The connection from SN to the mobile end hosts forms the mobile part. The virtual circuit for the mobile part is maintained as a soft state, which needs to be periodically refreshed (the refreshing can be achieved using the feedback mechanism). In the presence of large scale mobility there might be considerable overhead in maintaining split VC information and the mobile part may affect the performance of fixed part of the network.

Routing with QoS constraints are calculated using the link states. Each link has associated link attributes such as available bandwidth, delay experienced and error rate. Depending on the requested QoS appropriate “cost” is used in

calculation of the shortest path algorithm to find the QoS constrained route.

An end node measures periodically the power in the transmission received from the cluster head. When the power is below a threshold it initiates a handoff (which can be indicated using the feedback mechanism). The SN node receives the handoff message and communicates to adjacent SNs to find the potential new SN. The SN extends the current mobile part of the VC path through the new SN to the mobile host. During the handoff, to provide seamless connectivity packets are forwarded both to old SN and the new SN. Once the handoff is completed the mobile node continues communication through the new VC. Due to frequent movement suboptimal VCs or VCs with loops might result. These are detected by the information received in the feedback packets and VCs are updated appropriately by the new SN.

3.3. Adaptive Medium Access Control

In a cluster, communication occurs in frames. Time division multiple access (TDMA) allows channel to be shared among users without contention. Packet reservation multiple access (PRMA) in conjunction with TDMA is used for reservation [8]. The frames consists of two phases control phase which consists of mini-slots and a information phase used to transmit data. During the control phase the nodes running real-time applications reserve required number of information phase slots. Other information required for routing and slot synchronization may also be transmitted during the control phase. The cluster head broadcasts the acknowledgment (for data packets) and VC reservations. VCs which have real-time information to transmit use the reserved slots to transmit their information. The remaining slots are used by data applications using a random access scheme such as ALOHA. Efficient usage of free slots can be achieved by using a CSMA/CD (carrier sense multiple access/collision detect) scheme for the free slots. This proposed MAC layer is similar to the one used in WAMIS [3].

3.4. Flexible QoS specification

Traditional networks specify QoS in a rigid manner. QoS is specified usually by average bandwidth, bounded delay and jitter requirements. In a dynamic environment such as mobile wireless network provisioning and supporting rigid QoS is difficult. In [16] additional QoS metrics (discussed in section 7 such as probability of seamless connection and QoS degradation) are proposed for wireless environment.

A flexible QoS specification is better suited to specify the QoS requirements of adaptive applications. In our framework we allow the applications to specify a QoS in a flexible manner. For example instead of specifying an average bandwidth, the application can specify minimum and average bandwidth. The objective is to provide at least the minimum bandwidth always and provide average bandwidth in

the long term. Similarly, delay and loss requirements can also be expressed as a range. A simulation study of adaptive applications in a wireless environment needs to be done to find the exact parameters which are needed to specify the flexible QoS.

3.5. Admission Control

The admission control policy needs to use the flexible QoS when making decisions about admitting a connection. Admission control is handled by a “bandwidth manager” module at the SN. The SN collects information about various links and their attributes via the feedback packets. The bandwidth manager uses this and QoS requirements of the new connection to perform admission control. Instead of allocating all bandwidth a portion of the bandwidth (for e.g., 20%) can be used to absorb new connections introduced due to mobility.

3.6. Feedback Mechanism

In the proposed framework a generic feedback mechanism is used to provide adaptability at all layers. Periodically the application sends a feedback packet (one feedback packet for every N data packet sent) to collect information about current network conditions. The feedback packet is sent along the VC and is turned around at the destination forming a feedback control loop (similar to the ATM available bit rate (ABR) service’s feedback loop). If the application does not generate feedback the SN is assumed to have capability of generating control packets to collect feedback information. The mobile end hosts have the capability to return feedback packets to the SN or application. The feedback packet is used to collect information from all the components of the network which can be used at different levels of the architecture. We propose a generic feedback packet format (see Figure 2) and discuss how the feedback can be used at different components.

3.6.1 Application Layer

At the application layer adaptation based on the feedback packet can be achieved. For example the feedback can indicate the current available bandwidth which can be used to modify the encoding parameters (such as quantization level) of a voice or video application. In layered video encoding the feedback information can be used to decide when to drop or add a layer. Feedback information of error rate can be used to modify error protection mechanism used. Sophisticated error control mechanism such as convolutional coding can be used when higher error rates are present and simple forward error correction can be used for lower error rates.

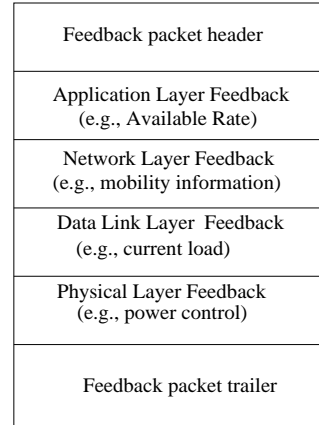


Figure 2. Generic feedback packet format.

3.6.2 Network Layer

The network layer can use feedback information which can indicate end host mobility to initiate a handoff. During handoff the SN can indicate to the gateway via the feedback packet that future packets need to be sent to both the current SN and the potential new SN. This entitles that even though the SN is part of fixed network it should be mobile-aware. Another use of feedback at network layer is to use the feedback information for calculating routes with QoS restrictions.

3.6.3 Data link Layer

The feedback can be used to indicate the current load of the network. Dynamic packet reservation mechanism where the percentage of the reserved bandwidth (similar to the technique proposed in [4]) is varied can be used based on the load indicated in the feedback.

3.6.4 Physical Layer

Adaptive power control can be used at the mobile end hosts for efficient communication and to protect active links (links which connect end host to cluster head). Mobile end hosts can measure periodically the power received from bases stations or cluster heads. If the power is below a threshold it can initiate a handoff which can be indicated in the feedback packet. The measured power can also be indicated in the feedback packet. This information can be used at the cluster head to modify its power when transmitting to that mobile host.

4. WAMIS: Architecture

The adaptive mobile multimedia networks architecture [3] for a “3M” environment (real-time Multimedia, Multihop and Mobile) was designed and implemented in

the WAMIS (wireless adaptive multimedia information system) project at UCLA. Most of the other architectures include only two of the Ms. The WAMIS project used an integrated design, evaluation and prototyping approach. The new protocols being developed were simulated before implementation. The key components of the system architecture include:

Multimedia applications: Adaptive voice/video applications are used to handle the in changing networking conditions.

Wireless adaptive network algorithms: The standard network algorithms, such as TCP/IP, run on top of these wireless algorithms. The wireless algorithms consist of the subnetwork control methods, and link and mobility control methods. The wireless algorithms interface with the wireless media via radio control.

4.1. Adaptive Multimedia Applications

Multimedia applications should use the bandwidth efficiently in wireless networks. Overhead of maintaining connectivity and routing information can be high. To adapt to changing bandwidth and error conditions the encoder for voice and video should support rate-adaptive compression. The compressed data should be transported in such manner to facilitate error protection. Multiresolution decompositions are used in which different portions are encoded and assigned a priority level (similar to layered approach). The lower priority information can be discarded in the event of rate changes or error rate changes. Usual video compression algorithms (e.g., MPEG-2) emphasize on motion compensation. Motion-compensation in wireless environment has the disadvantage of being more vulnerable to transmission errors and give rise to block effect.

WAMIS uses a wavelet-transform based adaptive video compression algorithm. In wavelet compression the image is hierarchically decomposed into a series of lower resolution ones. The hierarchical decompositions facilitate the wireless layer to deliver packet based on its priority. The coding scheme uses six-level discrete wavelet transform (DWT). The transformed coefficients are further scalar quantized and run length encoded. Error protection of the information is provided by convolutional codes. Each of the traditional encoding technique has been modified to suit the wireless environment. The scalar quantization is used so that lower power is consumed. To ensure low bit error rate (BER) punctured convolutional codes and unequal error protection of sub-bands is used.

Traditional speech coding has been mostly designed for bandwidth efficiency where available bandwidth does not vary. For example, code-excited linear prediction (CELP)-based coders achieve low bit rates, but have poor performance in presence of background noise. The speech encoding used in WAMIS provides low delay, adaptive bit allocation,

and quantization and is robust to noise [5].

4.2. Bandwidth Allocation and Topology Control

The subnet layer aims to support reliable, efficient transmission of multimedia in a mobile, multihop, self-configuring environment. The key notions used are: use of clustering for resource management, dynamic virtual circuit scheme, QoS routing and dynamic transmit power control.

In single-hop networks the nodes can learn the network topology by communicating with each other and/or with the base stations. This notion is extended to multihop networks by using the clustering concept. Though clustering concept has been used earlier for handling mobility issues in routing, WAMIS uses clustering for resource management. WAMIS uses a "soft-state" VC scheme and fast reservations, packet reservation multiple access (PRMA), to reroute and reallocate bandwidth in a mobile environment.

4.2.1 Clustering and Virtual Circuits

In WAMIS, the multihop network is partitioned into clusters, which enables controlled and accountable bandwidth sharing. Virtual circuits with associated QoS guarantees are used to provide support for multimedia. QoS routing is used to keep track of QoS metrics, such as bandwidth and delay along each path.

Clusters are formed dynamically between communicating nodes. In WAMIS nodes can communicate within cluster, and communicate with nodes up to two hops away. Optimal cluster size is formed under constraints of spatial reuse and delay minimization. Each node has a distinct ID. In each cluster a distributed election mechanism chooses a cluster head with minimum ID. Nodes that connect two clusters act as gateways. A time-slotted frame medium access control (MAC) is used in each cluster. The frames have a control phase and an information phase. The control phase uses m mini-slots (m) ($m \leq N$, N is the number of nodes). Control messages required for routing, slot synchronization, VC reservation are exchanged in this phase. Each node communicates to the cluster head. Cluster head broadcasts ACKs and slot reservation information. Each real-time application has a end-to-end VC assigned. VCs have a number of slots reserved (at call setup) in a frame depending on their bandwidth requirement. In the information phase the active VCs use their reserved slots to communicate first. The remaining free slots is accessed by datagram (best-effort) traffic using random scheme such as ALOHA. Datagram packets are ACKed but VC packets are not.

Routing uses a distributed distance vector scheme. Inter-cluster routing is via the gateway nodes. The basic routing scheme is made QoS-aware, so that various attributes such as bandwidth availability is propagated along the path.

4.3. Mobility and Multimedia support

To accommodate mobility, routes and cluster information must be dynamically updated. Distance vector supports dynamic route update inherently. The distributed cluster algorithm reconfigures clusters efficiently and quickly [11]. VCs also need to be rerouted in the presence of mobility. Static VC setup entitles a new VC setup to account for mobility. WAMIS uses VC setup based on fast reservations. Each packet of a VC is routed individually based on destination address, similar to a data packet. But the first packet in a VC stream reserves the slots for subsequent frames. If the slot is unused for certain time the reserved slot is freed based on a timeout. This is similar to the “soft-state” and flow ID proposed in RSVP (resource reservation setup protocol) [2]. The effects of route changes due to mobility is absorbed by rate-adaptive, hierarchically encoded voice/video compression algorithms. After rerouting is done the applications regain their quality.

4.4. Power Control

Power control is used to increase network capacity by mitigating wireless interference which increases spatial reuse. WAMIS uses a suite of power control algorithms for user admission and mobility management. The main protocol used is the distributed power control with active link protection (DPC/ALP). This algorithm controls powers of the links based on measurements of power at the receivers. New links increase the power by a fixed factor in each power update. The algorithm maintains the power of active links and increases the power of new inactive links. A threshold limits the increase in the power for new links. The new link is formed only if the threshold is reached. This mechanism serves as protection of existing active links. The new link voluntarily drops out if necessary. Simulation results show that this scheme works well and maintains the network structure. The DPC/ALP algorithm also provides network elasticity in maintaining connections which is suited for supporting mobility.

5. SWAN: Seamless Wireless ATM Network

SWAN is an indoor wireless network built on experimental basis to explore the combination of multimedia and wireless access. It is intended to support users in a building, using devices such as, personal digital assistants (PDAs), laptops and portable multimedia terminals. SWAN architecture uses the ATM (asynchronous transfer mode) technology in the backbone to connect base stations and wireless ATM last hops to mobile users. The aim of the project was to support integrated services (multimedia) in a wireless environment.

SWAN uses bandwidth and QoS management to support multimedia in presence of wireless access and mobility.

It aims to provide continual network connection to mobile wireless ATM hosts. The SWAN system uses *Etherware* network software which includes novel connection management for handling mobility and manage wireless resources. It provides a flexible platform for adaptive applications. The base station and mobile hosts use custom designed network adapter cards called FAWNs (flexible adapters for wireless networking [10]). FAWN has a programmable hardware for implementing different MAC control to transmit ATM cells over air. SWAN end hosts can be “smart” (laptops with FAWN card) or “dumb” (custom build dumb multimedia terminal using FAWN card).

5.1. SWAN: Communication Architecture

ATM networks is used in the backbone. Wireless hosts are present in the last ATM hop. Base stations connected to the ATM network communicate to the wireless hosts. Base stations are conventional PCs or Sun workstation with one or more FAWN adapter cards. Base stations act as mobility-aware ATM switches. Room size cells with one base station in each cell is used.

The ATM VC concept is extended to the mobile end hosts. This entitles that VCs be rerouted when the end hosts move. The main challenge is VC management in presence of mobile hosts is to minimize the disruption during handoff. The end hosts are assumed to move inside a building, hence the handoff frequency is high. Another issue is the support for ATM QoS at wireless last hop.

5.2. SWAN: Mobility Management

A signaling protocol is used to communicate between mobile hosts, base stations, wired switches and wired hosts. The signaling protocol consists of two entities, the traditional ATM signaling component and a (modified) mobility-aware component. The mobility-aware component in addition to performing connection management, also interacts with mobile hosts to negotiate resource allocation at VC setup, to reroute VCs and configure per-VC entries. The wireless link is used to schedule cells in the VCs at specified rate, perform error control using forward error correction (FEC), implement QoS-aware MAC.

Mobility is managed using fast rerouting. Each host has two threshold for measured power received from the base station. When the lower threshold is crossed the mobile hosts initiates a soft handoff. During soft handoff it periodically transmits requests for new base stations. The current base station extends its current path to the moving hosts through other base stations. Idle base stations in the vicinity of the mobile hosts start communicating with the hosts along the extended path. Optimizations to avoid loops and reduce path length are done.

5.3. SWAN: Medium Access Control

To support multimedia traffic SWAN provides multiple connections and low-latency handoffs. Each base station has multiple radio ports to communicate to mobile hosts. Channels are distributed among the base stations. Due to presence of base station in each cell, the MAC layer needs only to handle multiple channels and does not have to address the medium sharing. SWAN uses a off-the-shelf radio card which is a 625 Kbps half-duplex 2.4 GHz ISM band one. Frequency hopping is used by the radio for communication. Channels can interfere if they use same frequency. SWAN uses a 22 distinct hopping sequences for communication.

ATM cells are transmitted using air-interface packets. SWAN uses synchronous data link control (SDLC) protocol over air. In one burst several SDLC frame can be transmitted separated by SDLC sync bytes. The header of the air-interface packet carries information to indicate that the packet has an encapsulated ATM cell. Each radio port has a logical id assigned by the base station which is used in a token-passing protocol for communication between the base station and mobile hosts. The base station acts as the master, which hands out tokens to the next eligible mobile host. The token-passing approach enables the base station to manage the QoS and bandwidth within a cell.

6. MMWN Architecture

MMWN (Multimedia support for Mobile Wireless Network) system consists of adaptive link and network algorithms to support quality of service in large, multihop mobile wireless network [15]. The assumptions that nodes are fixed and availability of required bandwidth is not valid for multihop, mobile networks. Due to unpredictable movements, the capacity and bit error rate vary. The quality of service issue has to be addressed by the whole network instead of just at the last wireless hop.

MMWN consists of modular system of distributed, adaptive algorithms at link and network layer. The key features of the system are: adaptive control of link quality; hierarchical organization; routing with quality of service; and autonomous repair of multipoint virtual circuit with resource reservations.

6.1. MMWN: Hierarchical Organization

In MMWN network nodes organize themselves into hierarchical control structure. End hosts combine to form “cells” each of which contains a switch called “cellhead”. A group of cellheads forms a cluster, which operates as multihop packet-radio network. Lower level clusters, can group to form higher level clusters. Nodes interconnecting clusters act as boundary nodes. The routing information is

distributed reliably by link-states, using a hop-by-hop automatic repeat request (ARQ) protocol. Routes are calculated at switches based on link state information and QoS constraints. The route with minimum hop satisfying the QoS requirements is chosen. To avoid loops, source-routing is used for datagrams and VCs are established for stream-oriented connections. As the datagram or packets in a VC travel through the lower levels of hierarchy more detailed routing information becomes available.

6.2. MMWN: Mobility and QoS management

Servers are used to provide the various functionalities such as location management and QoS-aware route generation. A “location manager” is present in each cluster, which maintains location information of endpoints of that cluster. This also assists in locating endpoints within and outside of the cluster. A “QoS-manager” in each cluster computes the QoS attributes to be advertised in the link-state information.

Mobility is managed using a number of techniques including adaptive link control which “stretches” an existing link, local switch handoff which can bypass a particular switch, and rerouting a portion of a VC. MMWN supports multicast by using a “multicast-manager” in each cluster. This component controls the multicast groups, enabling joining and leaving of end points in an existing multicast session.

6.3. MMWN: Clustering Techniques

MMWN system has the ability to form clusters (cell-formation) autonomously. By using adaptive link control methods each endpoint has continual information on reachable switches. Quality of each link is monitored using periodic “beaconing”. QoS link attributes such as link bandwidth and error rate are tracked using link quality vector. A clustering mechanism can take place among switches (cell-heads) to form a clustering at a higher layer. Border nodes in a cluster act as virtual gateways. Virtual gateways in adjacent clusters are connected by virtual links. When hierarchical organization is used, the main issues are to bound the size of cluster, to minimize the levels in the hierarchy, and to minimize the volatility of inter-cluster connectivity. To achieve these objectives MMWN system supports cluster splitting when cluster size becomes more than a threshold, cluster merging when number of endpoints in cluster is below a threshold. Virtual gateways are created, modified or destroyed as dictated by the cluster dynamics.

6.4. MMWN: QoS Routing

The QoS manager in each cluster computes and keeps track of the service attributes of virtual gateways, attributes

of virtual link and attributes of path connecting interior points to virtual gateways. A variant of Dijkstra's shortest path first (SPF) algorithm is used to compute routes. Based on the requirements of the session, the switch chooses an appropriate attribute that will be used as the "cost" guiding the construction of the SPF tree.

All data messages use the route-directed forwarding (source-routing) to traverse to a particular destination. For stream-oriented session the route information is used for a VC. The route information is specified with fine granularity near the source and with coarser granularity towards the destination. As the packet traverses the network, intermediate switches refine the granularity of the route. Switches are capable of reserving bandwidth for individual VCs. During VC setup, each switch reserves resources along the links of adjacent switches. Reservation at VC routes at higher level cluster is just a matter of accounting. When endpoints move, the nodes use adaptive link control to prolong the existence of VC. MMWN provides mechanisms to repair VCs which include handoff and local reroute to repair lowest-level VC. End reroute mechanism is used to repair any level VC.

7. QoS Guarantees in Mobile Computing

Mobile computing is an emerging new computing environment. Wireless access up to 1-2 Mbps per user are expected to be available in the near future. The main issues in such an environment is QoS guarantee in presence of mobility and seamless connectivity. Currently used QoS parameters are inadequate to capture the QoS nature of wireless environment. New QoS parameters which are better suited for wireless environment are proposed in this architecture. A network architecture and set of transport level services to supported these QoS parameters are proposed [16].

7.1. QoS parameters

Though traditional QoS parameters can be used in mobile computing environment they do not suffice to capture the unpredictable mobile environment. Additional QoS parameters such as *graceful degradation of service*, *loss profile*, and *probability of seamless communication* can be used in mobile computing environment.

Service degradation occurs due to changing environment in mobile computing. The graceful degradation of service parameter characterizes the degree of degradation which occurs due to mobility. In a mobile computing environment though the requested resources might be available during connection setup, it might become unavailable due to mobility or interference. A loss profile can be associated with the connection which specifies a preferred in which data can be discarded when adequate bandwidth is not available. Probability of seamless communication parameter specifies

in probability terms the requirement of seamlessness of the connection. For example this parameter will be large for ftp type of application and small for real-time audio application.

7.2. Network Architecture and Internetworking

The approach used here is to view the mobile network as logically different from the fixed high-speed network at transport layer. ATM technology is used at the transport layer. The architecture has a three-level hierarchy. At the last level are the mobile end hosts (MH), which can communicate to each other and stationary hosts. These are connected to a mobile support station (MSS) which is present in each cell. A set of MSSs are connected to special nodes called "supervisor hosts" (SH) that map the transport classes to mobile network. The logical separation of mobile and fixed network is provided by the supervisor hosts. In this architecture the cells can be either a pico-cell (a room inside a building) or a cell which is a cluster of mobile nodes outside a building. The MSSs are connected via a hub or a LAN to the SH.

The addressing of nodes is solved by using the virtual internet protocol (VIP) proposed by Teraoko and Tokoro [6]. Every node has a permanent virtual address assigned to it. The transport layer uses the virtual address to locate the node. This method supports address transparency. The address resolution from VIP to actual IP address can occur at network layer of the node or at a gateway node.

7.3. Mobility Management and QoS support

Packets of variable size are used for communication between mobile users and SH. Packets have sequence numbers which can be used to discard duplicate packets. When a connection is setup it is assigned an identifier by the SH and returned to the MH. During the communication the MSSs just act as large caches. A MSS does not discard a packet unless it receives an acknowledgment from MH. When a MH moves in the same subnetwork, the SH updates the current location of MH. When MH moves between cells supervised by different SHs, then the new SH takes over all open connections of MH. It requests the old SH to forward all the pending packets and new identifiers are assigned to existing connection. Also the SH includes the VIP of MH to its list of aliases, so that future packets intended to the moved MH are sent to the new SH.

In this architecture the mobile network is separated from the fixed high-speed ATM network. A connection is split into mobile part from the SH to MH and a fixed part from SH to the fixed network. The splitting of connection protects the fixed network from the changes in the mobile environment. The QoS associated with MH to SH connection

is negotiated separately from the fixed part. For the mobile part of the connection new QoS parameters defined earlier as used in addition to the traditional QoS parameters. To handle mobility when SH moves from one domain to another one approach for the SH is to anticipate the move and inform the new SH which can renegotiate QoS parameters for the connection before MH moves. Another approach to overcome this problem is to setup a set of static connections between SHs using the high-speed network. These connections can be used to forward the pending packets when the MH moves. The second approach handles mobility better but pays the cost of maintaining connectivity between SHs.

8. Comparison of the frameworks

This section gives a comparison of the proposed QoS framework (AQuaFWiN) with existing proposals to support QoS in a wireless environment. A summary of the comparison is given in Table 1.

- **Adaptive Applications:** We believe that multimedia application should be adaptive in a wireless environment. Hence, AQuaFWiN proposes to use adaptive voice and video applications as in the WAMIS architecture. SWAN, MMWN, and QGMC have the disadvantage of not using adaptive applications.
- **Network Architecture :** The network architecture is hierarchically organized in all the frameworks. In the proposed QoS framework, MMWN, SWAN and QGMC a wired infrastructure is assumed at the backbone. In WAMIS, MMWN and AQuaFWiN clustering of end nodes with support for multihop and autonomous (ad hoc) network formation is used. SWAN and QGMC do not support multihop communication.
WAMIS is useful for quickly configuring a wireless network without support of fixed infrastructure, which is useful in emergency and military applications. AQuaFWiN and QGMC separate the mobile and fixed part of the networking which minimizes the effects of wireless part on the fixed infrastructure.
- **Network layer :** The network layer in all the frameworks support connection management with QoS constraints. VC is split into mobile and fixed parts in AQuaFWiN and QGMC. Fast VC and reservation is used in WAMIS. ATM VC concept is extended to end hosts in SWAN. A Hierarchical VC concept is used in MMWN which is the more scalable than methods used other frameworks.
- **QoS routing:** QoS routing is based on source-routing calculated using link-state information in AQuaFWiN. Source-routing and QoS manager are used for QoS routing in MMWN. VIP is used for routing in QGMC.

A signaling protocol is used in SWAN for routing. Distance vector routing protocol is used in WAMIS.

- **Mobility management:** Mobility is managed using VC (path) extension and VC optimization in AQuaFWiN, WAMIS, SWAN and MMWN. The SH maintains the location of mobile end hosts in QGMC. The VIP provides a unique address for each host in QGMC. The VIP method has the advantage of providing address transparency in QGMC. For good mobility support in QGMC overhead of maintaining connectivity between SHs has to be incurred.
- **Link Control:** Low priority drop is used in WAMIS and AQuaFWiN. Adaptive link control is used in MMWN.
- **MAC:** WAMIS and AQuaFWiN use a TDMA with PRMA for voice/video and slotted CDMA for data. SWAN uses ATM cells transmitted over air-interface using SDLC protocol and frequency hopping for channel sharing. Floor acquisition multiple access (FAMA) protocol is used for MAC in MMWN. In QGMC MSSs can communicate with different end hosts directly.
- **Feedback:** A generic feedback packet to collect information at various components is used in AQuaFWiN. Other frameworks do not provide direct feedback mechanism. They only provide indirect feedback mechanism such as mobile host initiating soft handoff.

9. Summary and Conclusions

In a wireless environment due to interference and mobility, available bandwidth and error rate are unpredictable. Provisioning and guaranteeing QoS in such a environment is a challenging problem. We have proposed a hierarchically organized network architecture to support QoS in a wireless mobile environment. All network components of the framework are QoS-aware. A generic feedback mechanism is proposed which collects information from different components. This feedback information can be used at various layer to achieve the necessary adaptability in a wireless environment. An overview of other existing architectures WAMIS, SWAN, MMWN, and QGMC that support QoS were discussed. A comparison of these frameworks with the proposed framework along various dimensions was given.

AQuaFWiN architecture has used several concepts such as hierarchical organization, adaptive power control from existing architectures. The key distinguishing feature of the proposed architecture is the generic feedback mechanism which can be used to achieve adaptability at different layers of the network.

We plan develop a simulation model for AQuaFWiN architecture to test and improve the methods proposed. Once the simulation model is available we plan to compare quantitatively our framework with existing frameworks.

Table 1. Comparison of QoS frameworks

Comparison Based on	AQuaFWiN	WAMIS	SWAN	MMWN	QGMC
Adaptive Application	Voice/ Video	Voice/ Video	-	-	-
Network Architecture	Hierarchical/ Wired Backbone	Hierarchical/ Multihop	Hierarchical/ Wired Backbone		
Network Layer	Split VC	Fast VC	ATM VC	Hierarchical VC	Split VC
QoS Routing	Source-Routing	Distance Vector	QoS signaling	SR + QoS Manager	VIP + QoS
Mobility Management	VC extension			path extension	SH using VIP
Link Control	Drop low priority		-	Adaptive links	-
MAC	TDMA + PRMA (voice/video) slotted + CSMA (data)		Frequency hopping	FAMA	MSSs to MH
Feedback	Yes (direct)	Indirect (e.g., soft handoff in WAMIS)			

References

- [1] I. 13818-2. Generic coding of moving pictures and associated audio information. Technical report, MPEG (Moving Pictures Expert Group), International Organization for Standardization, 1994.
- [2] R. 2205. Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification. In <http://ds.internic.net/rfc/rfc2209.txt>, September 1997.
- [3] A. Alwan, R. Bagrodia, N. Bambos, M. Gerla, L. Kleinrock, J. Short, and J. Villasenor. Adaptive Mobile Multimedia Networks. *IEEE Personal Communications*, 3(2):35–51, April 1996.
- [4] A. Iera, S. Marano, and A. Molinaro. A layered protocol architecture for multimedia wireless-PCS networks. *Mobile Networks and Applications*, 3(1):73–87, 1998.
- [5] A. Shen. Perceptually-Based Subband Coding of Speech Signals. Technical report, Master’s thesis, Electrical Engineering Department, UCLA, June 1994.
- [6] F. Teraoka and M. Tokoro. Host migration transparency in IP networks: The VIP approach. In *Proc. of SIGCOMM*, pages 45–65, January 1991.
- [7] H. Gharavi and M. H. Partovi. Multilevel video coding and distribution architectures for emerging broadband digital networks. *IEEE Trans. on Circuits and Systems for Video Technology*, 6(5), Oct. 1996.
- [8] J. Dunlop et al.,. Performance of a statistically multiplexed access mechanism for a TDMA radio interface. *IEEE Personal Communications*, 3(1), June 1995.
- [9] J. Tham, S. Ranganath and A. Kassim. Highly scalable wavelet-based video codec for very low bit-rate environment. *Journal of Selected Areas of Communications-Very Low Bit Rate Coding*, 1996.
- [10] J. Trotter and M. Cravatts. A Wireless Adapter Architecture for Mobile Computing. In *Proc. 2nd USENIX Symp. on Mobile and Location-Independent Comp.*, pages 25–31, April 1994.
- [11] M. Gerla and T.-C. Tsai. Multiclustet, Mobile, Multimedia Radio Network. *ACM Baltzer J. Wireless Networks*, 1(3):255–265, 1995.
- [12] M. Tatipamula and B. Khasnabish. *Multimedia communications networks : technologies and services*. Artech House, 1998.
- [13] S. McCanne, V. Jacobson, and M. Vetterli. Receiver-driven layered multicast. In *ACM SIGCOMM*, Stanford, CA, August 1996.
- [14] P. Cheng, J. Li and C.-C.J. Kuo. Rate control for an embedded wavelet video coder. *IEEE Transactions on Circuits and Systems for Video Technology*, 7(4), 1997.
- [15] R. Ramanathan and M. Steenstrup. Hierarchically-organized, multihop mobile wireless networks for quality-of-service support. *Mobile Networks and Applications*, 3(1):110–119, 1996.
- [16] S. Singh. Quality of service guarantees in mobile computing. *Computer Communications*, 19(1):359–371, 1996.