

QoS in Wireless Data Networks

[Adam Petcher](#)

Abstract

Quality of service (QoS) is an important consideration in networking, but it is also a significant challenge. Providing QoS guarantees becomes even more challenging when you add the complexities of wireless and mobile networks. This paper discusses the challenges and solutions involved in providing QoS in wireless data networks. Multiple techniques for providing service guarantees are presented, including data link layer QoS schemes, network layer schemes, integrated (multiple layer) approaches, QoS routing, dynamic class adaptation, and techniques for reservation and admission control.

Keywords

network, wireless, mobile, QoS, Quality of Service, throughput, delay, jitter, loss, media, audio, video, real-time, congestion, routing, fading, interference, MANET

Table of Contents

- [1. Introduction](#)
- [2. QoS Overview and Background](#)
 - [2.1 QoS Guarantees](#)
 - [2.2 Applications Requiring QoS](#)
 - [2.3 Challenges of Wired QoS](#)
 - [2.4 Wired QoS Schemes](#)
 - [2.5 Additional Challenges of Wireless QoS](#)
 - [2.6 User Perceived Quality](#)
- [3. QoS Schemes at the Data Link Layer and Below](#)
 - [3.1 802.11e](#)
 - [3.2 Adaptive Inter-frame spacing in 802.11](#)
 - [3.3 802.16](#)
 - [3.4 INSIGNIA](#)
 - [3.5 Cross-layer Scheduling and Adaptation](#)
 - [3.6 Scheduling for Multimedia Packets in Wireless PANs](#)
 - [3.7 Contention-based QoS in Wireless Networks](#)
 - [3.8 Service Differentiation by Window Size and ARQ Adaptation](#)
- [4. QoS Schemes at the Network Layer and Above](#)
 - [4.1 SWAN](#)
 - [4.2 Congestion Monitoring and Adaptation](#)
 - [4.3 Threshold Management via Fuzzy Logic](#)
- [5. QoS Schemes involving Data Link and Network Layers](#)
 - [5.1 Integrated Adaptive QoS](#)
- [6. Admission and Reservation Schemes](#)
 - [6.1 Admission Control in 802.11e HCF](#)
 - [6.2 Reservation Mechanisms in Wireless Networks](#)

- [7. Class Adaptation Schemes](#)
 - [7.1 Adaptive Management](#)
 - [7.2 Dynamic Class Selection](#)
 - [8. QoS Routing Schemes](#)
 - [8.1 Wireless QoS Routing](#)
 - [8.2 QoS Routing in Wireless Mesh Networks](#)
 - [9. QoS in Partitioned Networks](#)
 - [9.1 QoS Challenges and Solutions in Partitioned Networks](#)
 - [10. Wireless QoS Products](#)
 - [10.1 802.11e Wireless Access Points](#)
 - [10.2 802.11 QoS Enterprise Products](#)
 - [11. Summary](#)
-

1. Introduction

Reliable network performance has long been an important factor for many network applications. However, with an increasing amount of audio and video being sent over public, packet-switched networks, the ability to provide quality of service (QoS) guarantees may be more important in today's networks than it ever was. As such, a good deal of effort has been applied to the task of finding ways to provide reliable network performance while at the same time utilizing the total network resources in an efficient manner.

The challenges associated with providing service guarantees are numerous, but the biggest challenge for traditional networks has been congestion. However, many more challenges exist for wireless and mobile networks above those in traditional networks. For this reason, a completely different set of QoS techniques are required for wireless networks than for wired networks. These additional challenges, as well as several techniques for overcoming them, are described in this paper.

The remainder of this paper is organized as follows. The first section is an overview of QoS including what guarantees can be made, what applications require QoS guarantees, what the challenges are, and some well-known QoS approaches. The second section describes some wireless QoS schemes that work on the data link layer and below. The third section describes some wireless QoS schemes that work on the network layer and above. The fourth section describes some integrated QoS approaches that work on multiple layers in the network protocol stack. The fifth section describes some QoS schemes involving admission and reservation. The sixth section describes some QoS schemes involving dynamic class adaptation. The seventh section describes some schemes involving QoS routing. The eighth section describes the additional challenges and solutions involved with QoS in partitioned networks. Finally, some of the wireless QoS products that are emerging in the industry are discussed.

2. QoS Overview and Background

This section provides an overview on QoS. Included is a discussion of the service guarantees that can be provided, the applications that typically require QoS, the typical challenges of wired QoS, some traditional QoS schemes that meet those challenges, and some additional challenges of wireless QoS. The final part of this section describes QoS in terms of the user's perception of quality.

2.1 QoS Guarantees

Service guarantees are typically made for one or more of the following four characteristics. A guarantee of

delay assures the sender and receiver that it will take no more than a specified amount of time for a packet of data to travel from sender to receiver. A guarantee of loss assures the sender and receiver that no more than a specified fraction of packets will be lost during transmission. A guarantee of jitter assures the sender and receiver that the delay will not vary by more than a specified amount. Finally, a guarantee of throughput assures the sender and receiver that in some specified unit of time, no less than some specified amount of data can be sent from sender to receiver.

2.2 Applications Requiring QoS

There are several applications that require service guarantees in order to function properly. These applications are described in this section.

First, and probably foremost, service guarantees are required to properly transmit audio and video. In recent years, several providers have started offering audio and video services, such as telephony and video, over packet-switched networks. One motivation behind these offerings is that there is unused bandwidth in the IP network that can be utilized for a fraction of the cost of a dedicated, circuit-switched network. Another motivation is that the free-form nature of a packet-switched network makes it very versatile and allows for new forms of services, such as video on demand. A particular challenge of audio/video transmission is that, for maximum efficiency, some compression methods encode the streams at a variable bit rate (VBR). While throughput guarantees can be made at the highest possible bit rates for such streams, doing so is wasteful and a more useful QoS scheme would make throughput guarantees at the average bit rate and allow for bursty traffic with minimal additional delay and loss.

Second, real-time systems that are connected to a network require a constant stream of data which matches or exceeds their throughput rate. Consider an automated system in a factory that makes cuts in a piece of metal based on data it receives from a network. If the network cannot keep pace with the cutting machine, then there is a chance that materials can be wasted. Another example of a real-time system requiring service guarantees is a military vehicle that is connected to a network. If it takes too long for data to arrive at the vehicle, or if enough data is lost, then the situational awareness of the vehicle's operators can become degraded. This problem is amplified when non-human system components, such as guided weapons, require information from the network to support their situational awareness.

Finally, Internet service providers (ISPs) are expected to make service guarantees to their clients. In some cases, the clients have an application that requires such service and they buy a package from the ISP that meets the needs of that application. In other cases, the client has no such requirement, but still expects the ISP to meet all service guarantees because they are specified in the contract.

2.3 Challenges of Wired QoS

There are many impediments to providing service guarantees in a network. Following are some of the challenges present in both wired and wireless networks.

The primary challenge in providing QoS is network congestion. When a network is congested, the end-to-end delay increases because the packet spends more time in the queue at each hop. Loss also increases because if a queue is too full, it will start to drop incoming packets. This phenomenon also limits throughput since additional packets sent will just be dropped.

Another difficulty involves multi-path routing. When two packets are sent, there is typically no guarantee that they will take the same path to the destination. If one path has more hops than the other or is more congested, the packets will not arrive at the destination at the same time. This uncontrolled routing can cause unacceptable delay or jitter.

2.4 Wired QoS Schemes

Several schemes have been devised which provide QoS in traditional networks. While some schemes reserve network resources during a setup stage, other schemes set aside resources on a per-class basis and provide a statistical guarantee of service.

One method for providing QoS is built into the Asynchronous Transfer Mode (ATM) protocol. ATM is a packet-switched network that makes use of virtual circuits. A virtual circuit is established during a setup phase in which the path from sender to receiver is fixed and resources are allocated at each hop. Any type of service guarantee may be made by ATM since all the resources necessary for the connection are reserved for the virtual circuit. Once a virtual circuit is established, ATM is very efficient in terms of the amount of time it takes to forward a packet at a single node. However, ATM is not very efficient in terms of utilization due to the fact the resources are reserved even when no data is flowing over the link.

Another QoS scheme is called Integrated Services (usually abbreviated as IntServ). IntServ uses a call setup stage to reserve the path from sender to receiver and allocate resources at each hop. IntServ reserves resources on a “per-flow” basis where a flow contains all the network traffic associated with a single application. Unlike ATM, IntServ operates over a heterogeneous network that may have a mixture of IntServ and non-IntServ traffic flowing through each node. As a result IntServ must take measures to guarantee an upper bound on the queuing delays at each hop. IntServ also provides a “controlled-load” service that makes no hard service guarantees, but is designed for real-time multimedia applications.

The next QoS architecture is Differentiated Services (DiffServ). In DiffServ, hosts on the edge of the network mark packets with the class of service they should receive. These edge hosts also shape the data they send to ensure that they don't send too much data at once. In the core of the network, routers look at the marking on each packet and forward it according to the per-hop behavior of the packet's class. For example, a packet marked for expedited forwarding will likely spend less time queued than a best-effort packet. The primary advantage of DiffServ over IntServ is that there is much less complexity, and therefore greater efficiency, due to the fact that routers do not need to remember details for multiple flows. However, it can be difficult to implement DiffServ on a heterogeneous network, and it is possible for service guarantees to be violated across the entire network if a single edge host does not mark and shape traffic correctly.

Multi-Protocol Label Switching (MPLS) is a QoS architecture that works with IP and intends to bridge the gap between IP and ATM. MPLS has many of the same efficiencies as ATM and it can handle 1500 byte packets without significant queuing delays. By comparison, an ATM cell is only 53 bytes long and includes 5 bytes of header.

2.5 Additional Challenges of Wireless QoS

As previously stated, many obstacles must be overcome in order to provide service guarantees in traditional wired networks. All of these obstacles exist in wireless networks, and an additional set of challenges are added that only exist in wireless or mobile networks. As a result of these additional impediments, QoS schemes used in traditional networks may not be feasible in wireless networks.

The first additional challenge of QoS in wireless networks is severe loss. Loss in wired networks is typically caused by excessive congestion that causes packets to be dropped at routers in the network. A negligible amount of data is lost due to corruption during transmission on a wire. A wireless link, however, typically suffers much more loss due to data being corrupted during transmission. One cause of loss in wireless transmission is fading, in which multiple versions of the same signal are received at the destination. If these signals are out of phase with each other or Doppler-shifted, they can interfere with each other. Other types of interference may also cause problems in wireless transmissions. This interference may come from other

communication occurring on the same frequency, electrical noise, or possibly even intentional communication jamming.

Another obstacle in wireless QoS involves propagation delay. Some wireless networks span distances that are measured in kilometers. In these networks, propagation delay can be a tremendous burden to all communication, but especially to communication that requires a guarantee on total delay. This problem may exist to some extent in metropolitan area networks (MANs), and it is a significant issue in satellite communications.

Next, there are times when it is desirable to make service guarantees in a mobile ad-hoc network (MANET). Such networks may not have an infrastructure or a “coordinator” that determines which node has access to the shared channel. Instead, the nodes must enter into contention for access to the channel. This contention results in a large amount of collisions, which can result in lost data or a significant delay. In a QoS scheme for MANETs, it is generally desirable that higher-priority traffic wins the contention over lower-priority traffic, and that a minimal amount of time is wasted on contention.

Finally, it can be difficult to maintain service guarantees in a network if the nodes involved are mobile. For example, in a scheme involving resource reservation, if the sender moves, a new route to the destination must be established. The possibility exists that no route to the destination exists that can provide the previous level of service guarantees, so the connection may be dropped or degraded to a lower level of service. A similar problem can occur if a node along the route from sender to receiver moves.

2.6 User Perceived Quality

A final consideration when discussing QoS is user-perceived quality. It must be known what kind of service guarantee is required for a particular application. It would be wasteful to reserve resources beyond these needs. Also, by knowing which service characteristics are less important, the network can make better decisions regarding what to sacrifice, if necessary.

Many applications that require QoS involve a person that is consuming the information. These applications include audio and video, but users also expect a certain level of service when simply surfing the web or reading e-mail. [SAL05] explores this user-perceived quality with a focus on less time-sensitive applications such as web surfing. Sometimes there is no quantitative way to determine the service quality for a particular application, so it is necessary to have people use the application at varying levels of QoS and ask for feedback on their experience. This way, it is possible to link the “intangible” user-perceived quality with the “tangible” factors that are typically studied, such as throughput and delay. It has been found that when users are surfing the web or reading e-mail, they are more tolerant of increased loss than they are of decreased throughput. Similar studies have been performed for other applications such as audio and video.

3. QoS Schemes at the Data Link Layer and Below

QoS schemes can work on various parts of the network architecture. Much focus has been placed on the data link layer of the protocol stack and the layers below. This section contains descriptions of some such QoS architectures.

3.1 802.11e

802.11e is an extension of the popular 802.11 wireless LAN protocol. This extension enables multiple service levels and can provide service guarantees for network traffic. This extension enables audio and video over home and office 802.11 networks, and it also allows for 802.11-based wireless ISPs.

The first modification made by 802.11e applies to networks that make use of the distributed coordination function (DCF). DCF is a contention-based media access method that is present in 802.11. In DCF, the winner of the contention is the sender that transmitted first, as determined by a random function. 802.11e proposes an enhanced distributed coordination function (EDCF) in which higher priority traffic will wait for a shorter amount of time before transmitting than lower priority traffic. The amount of time that the sender spends waiting is referred to as the inter-frame spacing (IFS). As a result, the higher priority traffic will always win the contention.

The second modification made by 802.11e only applies to networks in infrastructure mode. These networks typically use a point coordination function (PCF) in which a coordinator polls the connected nodes, giving them an opportunity to transmit. 802.11e proposes a hybrid coordination function (HCF) that works as follows. If the hybrid coordinator (HC) determines that the channel is idle during the contention period, it can initiate a contention-free controlled access period (CAP). During this period, the HC transmits frames and polls stations to allow them to transmit. Many types of QoS guarantees can be made depending on the implementation of the HC and how it decides to initiate CAPs and poll stations during those CAPs.

3.2 Adaptive Inter-frame Spacing (AIFS) in 802.11

Though 802.11e is a vast improvement over 802.11 in terms of the potential for QoS, many people have identified improvements that can be made to 802.11e in order to make it even more capable. One example is the proposal for adaptive inter-frame spacing described in [KSE04].

In AIFS, the EDCF in 802.11e is extended so that the IFS is modified when the link quality changes. When link quality degrades, the IFS for high priority traffic increases and the IFS for best effort traffic decreases. Back-off periods are similarly modified. The effect of the change is that when the link quality decreases, best effort traffic is sacrificed in order to continue to meet the guarantees of the higher priority traffic.

Simulation results show that EDCF with AIFS has improved throughput and delay when compared to standard EDCF. However, the performance of best-effort traffic suffers when using AIFS.

3.3 802.16

802.16 is a standard for a metropolitan area network (MAN) that can have thousands of subscribers spread out over up to 50 km. The standard only allows for an infrastructure mode in which the base station schedules use of the channel similar to the 802.11 PCF. Since 802.16 was designed to be used by ISPs and to send audio and video, QoS is built in to every connection.

802.16 specifies four classes of data: constant bit rate, real-time variable bit rate, non-real-time variable bit rate, and best effort. Each class has its own service guarantees suited to the traffic involved.

An interesting characteristic of 802.16 is that it can grant bandwidth on a per-station or a per-flow basis. This distinction allows for situations in which a single subscriber station has multiple flows with varying levels of QoS.

3.4 INSIGNIA

One well-known and often-cited method for QoS is INSIGNIA. Though based on IP, this architecture is presented with data-link layer architectures because of the key improvement offered by INSIGNIA: in-band signaling. The architecture also makes use of soft-state resource management. The combination of these elements results in an architecture that can quickly respond to changes in network topology.

In typical reservation-based schemes, the reservation mechanism is typically transmitted in a separate control channel. INSIGNIA proposes that these control signals be sent in the same channel that carries the data. By reducing the need for a separate channel, complexity is reduced and efficiency can increase. It also takes less time to transmit these control messages in-band than out-of-band.

INSIGNIA also proposes that soft-state resource management be used in reservations. In this scheme, a reservation is only maintained for a certain amount of time and if no packet is received for some time, then the resources are released. Also, there is no need for an initial reservation before data is sent – a node will attempt to reserve resources the first time it receives a new packets for which it has no reservation. This mechanism allows for a great deal of adaptability when used in a mobile ad-hoc network.

3.5 Cross-layer Scheduling and Adaptation

Some actions can be taken in the physical layer to improve quality of service. In [LIU05], an architecture is proposed which combines QoS reservation and scheduling at the MAC layer with adaptive modulation and coding (AMC) at the physical layer.

In AMC, the method for transmission changes when the link quality changes. For example, if the link quality degrades, the physical layer may start transmitting using BPSK instead QAM-16. As a result, more time will be required to send the same amount of data, so the MAC layer must adjust its scheduling accordingly. Using this scheme, throughput performance closely matches the performance of the channel.

When the link quality is good, it will take less time to transmit the QoS-guaranteed traffic than it will take when the link quality is bad. At these times, there will be more resources available to transmit best-effort traffic. As a result, the total bandwidth is well-utilized.

3.6 Scheduling for Multimedia Packets in Wireless PANs

When discussing QoS in wireless data networks, one must not overlook the needs of personal area networks (PANs). The most common example of a PAN is Bluetooth. Though QoS for audio is built in to the Bluetooth protocol, other applications requiring service guarantees, such as video, must rely on a QoS architecture built on top of Bluetooth. [ZHA05] describes such a QoS architecture for MPEG streams.

It is possible to inform the piconet master how much channel time is required by periodically responding to queries with channel time requests. The amount of channel time required can be derived from fields that are present in MPEG streams. It is important to note that the amount of channel time required can vary within a single stream. By sending these requests periodically, the allocated channel time can vary as needed.

The proposed architecture provides an adequate QoS for MPEG streams in Bluetooth. One advantage of the architecture is that it responds well to the addition of hosts on the piconet due to the fact the bandwidth requests are sent infrequently and only the amount of bandwidth that is required at any given time is reserved.

3.7 Contention-based QoS in Wireless Networks

Several people have researched efficient ways to provide service guarantees in contention-based MAC environments since such circumstances are common in MANETs. [STI04] builds on a previous contention scheme called contention resolution signaling (CRS) in order to provide service guarantees and make more efficient use of the available spectrum. The paper also explores the abstractions used when evaluating a wireless network.

It may not be appropriate to view a wireless network in terms of the “link” abstraction since spectrum, rather than the “link” is the scarce resource with which QoS schemes are concerned. If the wireless network is viewed as a continuous vector space in which physical space and spectrum may be allocated for transmission, then all available resources can be used more efficiently.

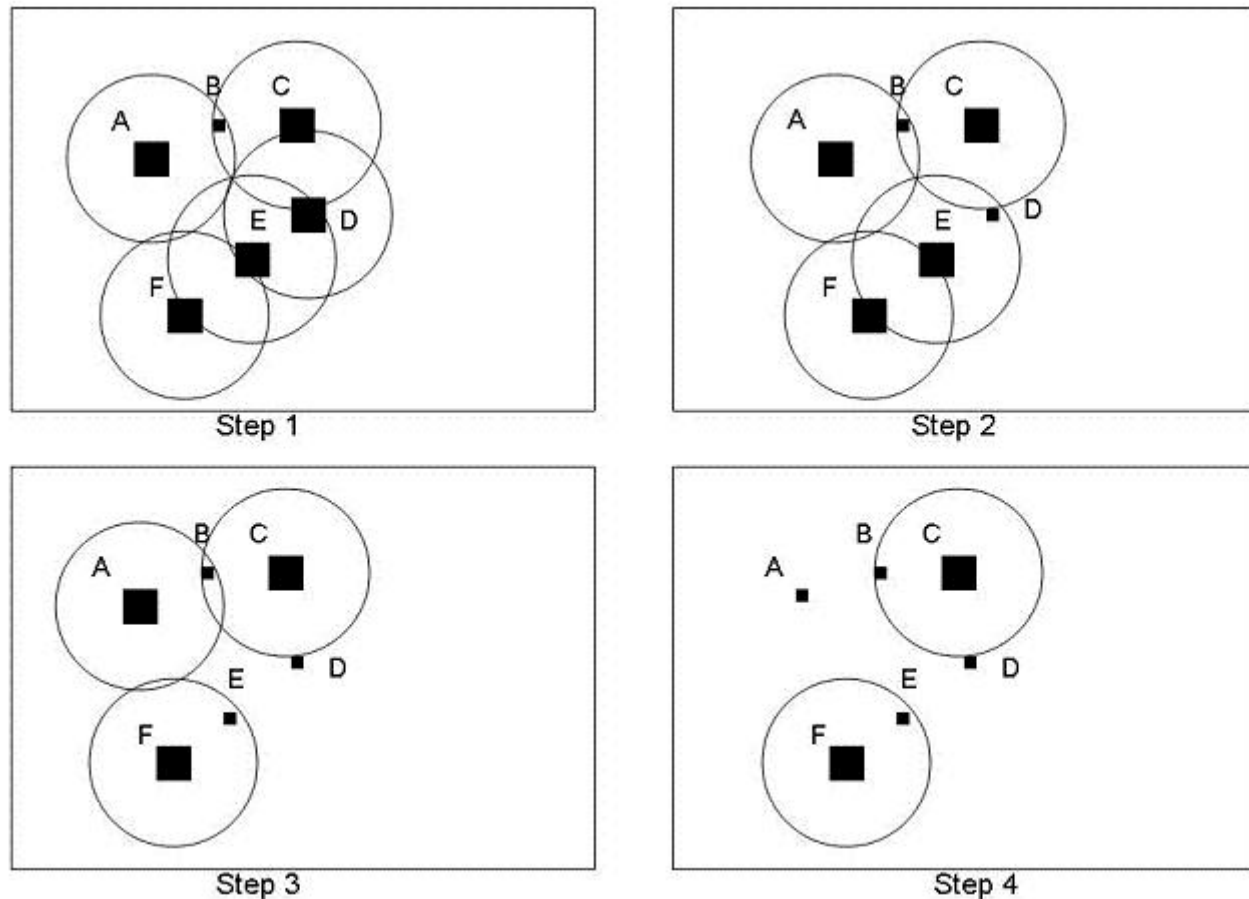


Figure 1. Contention Resolution in CRS

CRS is an architecture suited for a MANET with several nodes spread out over physical space such that all nodes cannot communicate directly with each other. A node that wants to transmit will send a signal in one of several “slots” in a CRS frame. Several rounds of contention will continue until each node that wants to transmit is not in range with another node that wants to transmit. In each round, a random backoff function determines which nodes are still in contention. An example contention scenario is depicted in figure 1. In this scenario, all hosts except for host B assert that they would like to transmit. Multiple rounds of contention follow until only hosts F and C are trying to transmit. While this system is very effective in terms of total bandwidth utilization, several improvements can be made.

First, the addition of an RTS-CTS mechanism reduces interference when there is a hidden terminal. This mechanism ensures that two senders that cannot hear each other won't transmit to the same receiver at the same time. It also allows the transmission parameters at the physical layer to be modified due to the state of the channel as perceived by both the sender and the receiver. Finally, it allows for energy conservation since only the RTS-CTS sequence must be sent at full power and the data transmission can be sent at the lowest effective power as determined during the RTS-CTS sequence.

A priority phase can be added to the beginning of the contention period. A node with high priority traffic can transmit in one of the slots in this phase. Since higher priority slots are earlier than lower priority slots, and

all prioritized slots are earlier than the typical contention slots, the highest priority traffic will win the contention. This mechanism is similar to the EDCF in 802.11e.

Next, resource reservation can be allowed by adding a QoS slot to the beginning of the priority phase. The QoS slot is used for reserving bandwidth in the network. Immediately following the QoS slot is a CBR slot which is used for holding the reservation. Finally, a VBR slot follows the CBR slot and is used to indicate the amount of throughput that is required at any given time. After a reservation is made, the sender can use the remaining slots in the frame to transmit data. A cooperative signaling slot is also added immediately before the contention slots. Once a reservation has been made, a receiver will use this slot to preempt contenders that are not in range of the sender.

Finally, better use of the spectrum can be achieved through a channelization scheme in which each node has its own channel and there is a single common “broadcast” channel used for establishing connections. A broadcast signaling slot is added to the priority phase. A node only needs to listen to the broadcast channel if there is data in the broadcast signaling slot. Otherwise the node only listens on its own channel. This scheme can be very efficient if channels are reused in such a way that no two nodes that are within range of each other share the same channel.

When all of these enhancements are combined, the result is a very effective QoS scheme for contention-based networks. Simulation results show near-perfect performance for the delivery of prioritized data.

3.8 Service Differentiation by Window Size and ARQ Adaptation

In a contention-based network, it is possible to reduce the probability that a collision will occur for any given packet simply by reducing the size of the packet. [ZIA05] suggests a way to make use of this phenomenon in order to aid in making QoS guarantees.

In the proposed scheme, real-time traffic has a smaller window size than best effort traffic. Also, the real-time data is not re-transmitted by ARQ. As a result of these changes, the real-time traffic has a much better chance of being transmitted without a collision and therefore receives a better level of service.

When combined with admission control, the proposed scheme can make guarantees on throughput, delay or loss, and on some combinations of the three.

4. QoS Schemes at the Network Layer and Above

There are many QoS architectures that work on the network layer and above. One advantage that is common to all such architectures is that they will work regardless of the transmission medium involved and can therefore be used in heterogeneous networks. This section describes some of these QoS schemes.

4.1 SWAN

One very notable network layer QoS architecture is SWAN [SWAN]. SWAN combines several features in order to provide service guarantees without requiring special QoS services in the MAC layer. SWAN forms a basis for some of the other QoS architectures described in this section.

The first feature of SWAN is an admission control system that denies access to flows requesting real-time service if the network will not be able to meet the needs of the flow. SWAN also requires nodes to shape best-effort traffic in such a way that it will not interfere with the real-time traffic. By making use of these two features, the network can make sure that it is never too congested to satisfy its service guarantees.

SWAN makes use of the congestion notification field in the IP header to inform nodes about the congestion levels in the network. The nodes can use the data in this field to make decisions about admission and shaping.

Simulation results show that SWAN is able to provide very stable delays for guaranteed traffic. However, some later work has been able to improve the overall efficiency and performance of SWAN.

4.2 Congestion Monitoring and Adaptation

One way in which SWAN can be improved is through the use of congestion monitoring and adaptation as described in [DOM04].

In the proposed system, the network is monitored and the nodes are informed when congestion levels pass a certain threshold. When a node receives this notification, it will respond by throttling down its best-effort traffic.

Simulation results show that the proposed modification improves all QoS metrics when compared to SWAN. Though the performance of best-effort is reduced compared to SWAN, it is shown that the best-effort traffic is not starved.

4.3 Threshold Management via Fuzzy Logic

In order for any QoS scheme to work, it must have some form of admission control. The methods by which a QoS architecture can decide which packets to admit are as varied as the QoS architectures themselves. In [KHO05], a fuzzy logic approach is used to determine whether to admit packets based on the level of congestion.

It is suggested that the threshold for packet admission be a function of the rate of change in buffer fullness. Fuzzy logic can be used to determine whether the buffer is getting more full and less packets should be admitted, or the buffer is getting less full and less packets should be admitted. In a sense, the network is making predictions about its future state, so several benefits are derived.

The main benefit of such an approach is that delay can be reduced and well-regulated across the network. Unfortunately, throughput generally suffers because the fuzzy logic algorithm will block admission of some packets even if the network is capable of accepting them without violating other guarantees. Another benefit of this approach is that delay and throughput are both improved when hosts are mobile.

5. QoS Schemes involving Data Link and Network Layers

The QoS architectures that have been discussed so far are mostly bound to the data link layer and below or the network layer and above. There has been some research into integrated QoS schemes that involve the data link and network layers, as well as other parts of the protocol stack. This section describes one such approach.

5.1 Integrated Adaptive QoS

An interesting approach to QoS is presented in [GOM03]. The paper presents a fully-developed QoS architecture called Havana which spans multiple layers of the network protocol stack. The architecture includes a Predictor component at the data link and physical layers, a Compensator component at the data link layer, and an Adapter component at the network layer and above. A diagram of the framework is shown in Figure 2, and the components are described in detail in the following paragraphs.

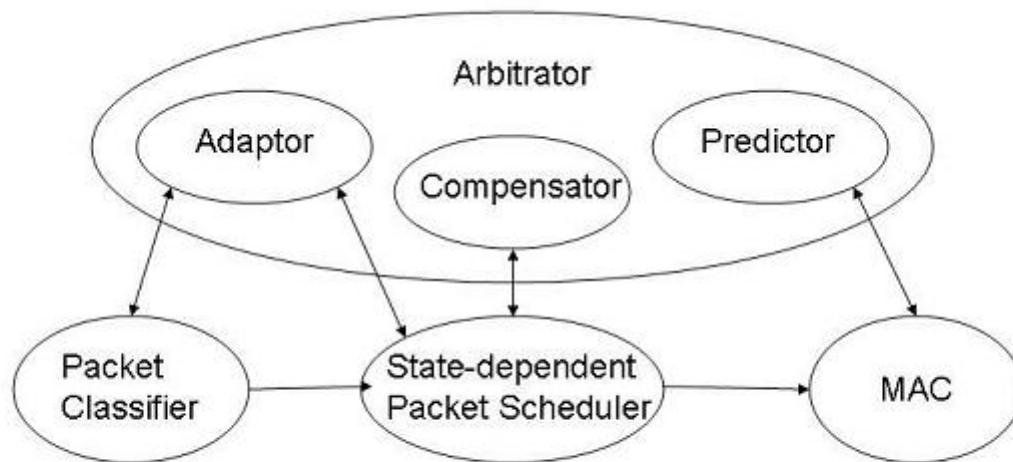


Figure 2. The Havana QoS Framework

The Predictor component works at the data link layer and physical layers. The purpose of this component is to determine whether the channel is in a “good” state. If the channel is good, then data transmitted on the channel is likely to arrive at its destination intact. It is important to note that the information provided by this component is a prediction of the future state of the channel. There are many methods for channel prediction, including monitoring for steadily increasing interference, periodic interference, and probable collisions. In this architecture, the data link layer will refrain from transmitting data when the channel is bad. This allows the network to save some bandwidth and to know ahead of time when data is not transmitted successfully.

The Compensator is a data link layer component that compensates the packets that were not transmitted because the channel state was bad. The method for compensation can be complex depending on the guarantees that were made to different classes of packets or flows. For example, traffic that has guaranteed delay jitter would be highly compensated so it is highly likely to be transmitted in the next slot. By contrast traffic that has guaranteed throughput would be compensated less and allowed to catch up over time. The general effect of compensation, however, is that data that wasn't transmitted due to a bad channel state will be compensated so it is more likely to be transmitted in the next slot.

Finally, the Adapter component works on the flow level and drops low-priority packets if the channel state is bad. This allows the higher priority traffic in the flow to be transmitted in the time when the channel state is good. The Adapter also regulates traffic at the application level if poor channel quality exists. Some applications are able to compensate when the channel is degraded so that the impact to the user is minimal. For example, an audio application might respond to a poor channel state condition by reducing the bit rate of the audio and adding more forward error correction in the transmission.

6. Admission and Reservation Schemes

The remaining sections of this paper describe some research that focuses on a particular service found in QoS-enabled networks, rather than on a portion of the network protocol stack. This section describes some research that focuses on admission and reservation schemes.

6.1 Admission Control in 802.11e HCF

[VEN04] suggests an improvement to the admission control in the 802.11e hybrid coordination function

(HCF). The problem with the HCF is that admission control is first-come-service. That means that once all the spots are taken, no more hosts will be admitted to the HCF, even if their data should get higher priority than the existing hosts.

The proposed solution involves a class-based admission policy. In this policy, portions of the bandwidth will be reserved for certain classes of traffic. For example, if 30% of the bandwidth is reserved for audio traffic, then no more audio traffic will be admitted once that 30% has been assigned to audio.

Simulation results show that improved performance over typical 802.11e HCF for voice, video, and bursty traffic. The disadvantage of this scheme is that some amount of bandwidth must be reserved, so it is difficult to make full use of all of the available bandwidth.

6.2 Reservation Mechanisms in Wireless Networks

Creating and maintaining reservations in a wireless network can be a difficult problem, especially if the hosts are mobile. A few techniques for improved reservation in mobile wireless networks are presented in [MOO04]. These techniques focus on the efficiency and effectiveness of reservations during a handoff.

The first suggested technique is to re-route the reservation in a crossover router during a handoff rather than completely re-establishing the reservation starting at the mobile host. It is typical in a handoff that most of the routers involved in the flow will still be valid after the handoff. By maintaining the reservation in these routers, only a smaller number of routers need to be involved in the new reservation. Since the reservation will be made on a smaller number of routers, there is a smaller chance that it will fail. The downside to this approach is that the routers will need to have more sophisticated software for performing a handoff.

The next technique suggested is a soft handoff. If the mobile host maintains the old reservation while making the new one, then no packets will be dropped due to the handoff. This is a well-used technique that has proven to be very effective. The downside to this technique is that up to double the resources are required in the period in which the old reservation and new reservation are both active. However, it is possible to reduce the significantly reduce the amount of duplicate reservations and wasted resources.

The final technique suggested is a dynamic resource reservation scheme that can guarantee the success of a handoff. By using the proposed method, it is possible to greatly reduce the amount of control data sent on the link to set up a connection, yet still have successful handoffs with little or no loss of data.

7. Class Adaptation Schemes

QoS-enabled networks provide different levels of service to different classes of traffic. Some QoS schemes propose that the class of traffic should be modified as a result of events occurring in the network. This section describes some such schemes.

7.1 Adaptive Management

In many QoS schemes, the service guarantees of some traffic can be compromised if significant changes to the network occur, such as the admission of new clients or a degradation of link quality. [CHA04] proposes to maintain the service level for high-priority traffic when these events occur by modifying the class of the traffic in response to the events.

In the proposed scheme, traffic is divided into a number of priority classes where some classes get a greater fraction of the channel than others. As such, the priority of the traffic is specified, but not necessarily any

other service guarantees. If a new client is added, all other clients of the same class will receive a lower level of service. So in order to receive the same level of service, some or all of the clients could raise their priority level. Whether the clients raise their priority level would depend on the application that is being used and the perceived QoS of the users.

Of course, there is a limit to the amount of increased service a client can get by raising its priority class. In order to provide an acceptable level of service to all clients, some simple admission control must be employed to block new clients if the channel is fully loaded.

7.2 Dynamic Class Selection

The most notable problem with class adaptation is that it can create a situation in which all clients are increasing their priority in a greedy fashion to the point where some or all of the clients won't receive the desired level of service. An evaluation of this problem is presented in [WAN05].

In the proposed architecture, flows will alter their class in response to past performance of that class. For example, if the desired throughput or delay was not met in the last slice of time, then the flow will increase its priority. The scheme is presented as an extension to the 802.11e EDCF, so priority is increased by reducing the IFS. Simulation shows that the scheme outperforms EDCF in terms of throughput and delays in all cases except for ad-hoc peer-to-peer communication with throughput lower bounds around 90kbps and up.

A significant part of this analysis is a game theoretic evaluation which shows that clients can alter their class in a greedy fashion, yet all clients will still receive the desired service quality given there are sufficient network resources available. Specifically, it is shown that an equilibrium state exists in all cases if a continuous set of priorities is available for use. Also, if one equilibrium is feasible, meaning all clients receive their desired level of throughput, then all equilibria are feasible. Finally, it is shown that the game is guaranteed to converge on an equilibrium.

8. QoS Routing Schemes

Any successful QoS architecture in a multi-hop network requires a routing algorithm which will select a path that can meet the desired service requirements. The following section discusses some of the issues and solutions involved in QoS routing.

8.1 Wireless QoS Routing

In ad-hoc networks, routes must be chosen in such a way that per-flow QoS requirements are met and the total bandwidth is well-utilized. In order to set up a route, it is necessary to contact every potential node to determine its level of load and whether it can provide the service level guaranteed to the flow. Unfortunately, the traffic resulting from flooding the network and contacting every node in the path can clog the network and waste bandwidth. [ZHA05] presents a QoS routing system that can satisfy per-flow guarantees without creating a large amount of control traffic on the network.

The first element of the proposed system is a hybrid routing function that stores information about each node in order to determine if a QoS-assured path is available. The network is only flooded if a path is not found using the stored information. Next, a directional search is used during flooding in order to reduce the number of nodes that are contacted. Only nodes that are geographically feasible for the desired route are contacted to determine if they can support the traffic. Finally, link delay-based scheduling of control packet forwarding helps the routing function determine the lowest-cost link from the set of available links.

Simulation results show that the proposed routing system offers the same rate of success in finding a delay-constrained path compared to purely flooding protocols, while significantly reducing channel usage for control packets compared to such protocols.

8.2 QoS Routing in Wireless Mesh Networks

When setting up a route in a wireless mesh network, special care must be taken so that nodes on the path do not interfere with each other while retransmitting the data. A solution to this problem is presented in [TAN05].

The paper proposes an interference-aware QoS routing and channel assignment algorithm that can improve overall utilization while satisfying the QoS requirements. When multiple non-overlapping channels are available, the algorithm will route traffic and assign channels such that the total harmful interference caused by retransmission with the same flow is minimized. Simulation results show that channel utilization, as measured by blocking ratio, is greatly improved over a non-interference-aware QoS routing algorithm.

9. QoS in Partitioned Networks

The final section of this paper discusses some additional challenges associated with QoS in partitioned networks. A partitioned network is a network that is divided into two or more sections. In such networks, an end-to-end connection may be made across multiple network sections, but the nodes of one section are unable to communicate with the nodes of another section. Because the node setting up the connection cannot query the individual nodes along the route to determine what kind of service they can provide, it is very difficult to provide a service guarantee for the end-to-end connection. Security concerns usually result in the creation of such a network in which some sections are encrypted differently than others.

9.1 QoS Challenges and Solutions in Partitioned Networks

Wireless military networks are typically partitioned in such a way that data must flow between encrypted and unencrypted sections of the network. Refer to figure 3 for a diagram of such a network. [ELM05] discusses QoS-related issues in these types of networks and provides some solutions that allow some level of QoS.

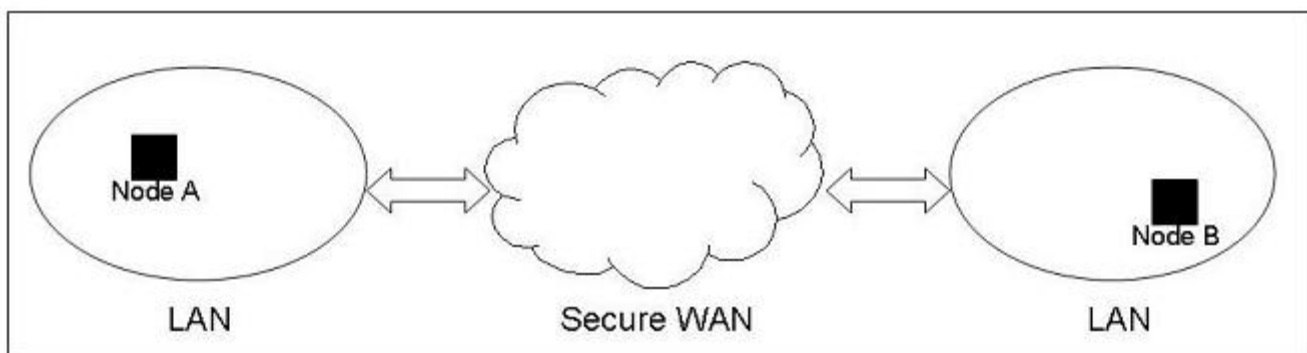


Figure 3. Partitioned Network

The first proposal for providing QoS over a partitioned network is for the nodes of one network section to measure the performance of another section as a whole. These nodes could then estimate the level of

congestion and capabilities of the other network section and make better decisions about service guarantees.

Second, it is possible for the nodes in one network section to determine if another section will suffer from blockages due to terrain, interference, or other causes. This blockage can be detected by measuring the network section in question, or by making use of other sensors to predict conditions that will cause blockage. Once a blockage is predicted, the nodes in the first section can hold packets and refuse admission until the blockage clears.

Next, it is possible to improve throughput and reduce the network utilization by combining small packets into larger packets and by compressing the data. By combining packets, a fewer number of packet headers will be sent, which will reduce the total amount of traffic sent while not reducing the amount of meaningful data sent. Compression can increase the rate of transmission for meaningful data, but only if the data transmitted can be compressed.

Finally, more reliable transmission can be achieved by treating the network sections as independent networks and setting up a TCP connection across each network. If a TCP proxy is placed on either side of a secure network, then an end-to-end connection can be created by connecting to the proxy at either side.

10. Wireless QoS Products

Due to the incredible popularity of the 802.11 standard, there are many products available or in work that offer QoS for these networks. Most of these products support 802.11e, but other QoS products are also available since 802.11e is not right for all situations.

10.1 802.11e Wireless Access Points

802.11e is expected to be an extremely popular standard for wireless QoS due to the popularity of 802.11. The immediate applications for QoS in 802.11 LANs are VOIP in the home and office, and wireless video in homes. A few wireless access points supporting 802.11e have been released, and all the major manufacturers of networking hardware will likely offer 802.11e support in their upcoming products. An example of such a product is the Netgear access point described in [WAG302]. With the combination of QoS, and higher data rates using MIMO, this new network equipment will allow an 802.11 network to accomplish some tasks that previously required an Ethernet network.

10.2 802.11 QoS Enterprise Products

Despite its prevalence and certain future popularity, 802.11e has some limitations and ways in which it can be improved. Some of these improvements were discussed earlier in this paper. According to [MERU], 802.11e will not support VOIP in an enterprise network because the number of clients is too great. So the company developed some products that will support VOIP and other systems requiring QoS in a network with up to 100 clients per access point and up to 30 VOIP calls per access point. The proprietary system also features no-loss handoffs.

11. Summary

This paper has described some of the challenges involved in quality of service for wireless networks as well as some solutions which allow service guarantees to be made. Several QoS solutions have been presented, including QoS architectures that work with the data link layer, the network layer, or both. Some solutions

have been presented that are specific to a certain aspect of QoS, such as admission control, class adaptation, and routing. Since there are many types of service guarantees that can be made, and a QoS solution must balance the network's ability to satisfy service guarantees with overall network utilization, there are many solutions available that each have their own benefits and trade-offs.

Acronyms and Abbreviations

- AIFS - Adaptive Inter-Frame Spacing
 - AMC - Adaptive Modulation and Coding
 - ARQ - Automatic Repeat reQuest
 - ATM - Asynchronous Transfer Mode
 - BPSK - Binary Phase Shift Keying
 - CAP - Contention-free Controlled Access Period
 - CBR - Constant Bit Rate
 - CRS - Contention Resolution Signaling
 - CTS - Clear To Send
 - DCF - Distributed Coordination Function
 - DiffServ - Differentiated Services
 - EDCF - Enhanced Distributed Coordination Function
 - HC - Hybrid Coordinator
 - HCF - Hybrid Coordination Function
 - IFS - Inter-Frame Spacing
 - IP - Internet Protocol
 - ISP - Internet Service Provider
 - IntServ - Integrated Services
 - LAN - Local Area Network
 - MAC - Media Access Control
 - MAN - Metropolitan Area Network
 - MANET - Mobile Ad-hoc Network
 - MIMO - Multiple In Multiple Out
 - MPEG - Moving Pictures Experts Group
 - MPLS - Multi-Protocol Label Switching
 - PAN - Personal Area Network
 - PCF - Point Coordination Function
 - QAM - Quadrature Amplitude Modulation
 - QoS - Quality of Service
 - RTS - Request To Send
 - TCP - Transmission Control Protocol
 - VBR - Variable Bit Rate
 - VOIP - Voice Over Internet Protocol
-

References

- Books
 1. C. Siva Ram Murthy, B.S. Manoj, "Ad Hoc Wireless Networks: Architectures and Protocols," Prentice Hall, 2004
 2. Sudhir Dixit, Ramjee Prasad, "Wireless IP and Building the Mobile Internet," Artech House, 2003

3. James F. Kurose, Keith W. Ross, "Computer Networking: A Top-Down Approach Featuring the Internet," Addison Wesley, 2003

- Papers

1. [CHA04]Hakima Chaouchi, Anelise Munaretto, "[Adaptive QoS Management for IEEE 802.11 Future Wireless ISPs.](#)" Wireless Networks, Volume 10, Issue 4, July 2004
2. [KHO05]Lyes Khoukhi, Soumaya Cherkaoui, "[Congestion control and clustering stability in wireless ad hoc networks: FuzzyCCG: a fuzzy logic QoS approach for congestion control in wireless ad hoc networks.](#)" Proceedings of the 1st ACM international workshop on Quality of service & security in wireless and mobile networks Q2SWinet '05, October 2005
3. [KSE04]A. Ksentini, M. Naimi, A. Nafaa, M. Gueroui, "[Routing & performance modelling: Adaptive service differentiation for QoS provisioning in IEEE 802.11 wireless ad hoc networks.](#)" Proceedings of the 1st ACM international workshop on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks, October 2004
4. [GOM03]Javier Gomez, Andrew T. Campbell, "[Havana: supporting application and channel dependent QoS in wireless packet networks.](#)" Wireless Networks, Volume 9 Issue 1, January 2003
5. [TAN05]Jian Tang, Guoliang Xue, Weiyi Zhang, "[Topology control & mobility: Interference-aware topology control and QoS routing in multi-channel wireless mesh networks.](#)" Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing MobiHoc '05, May 2005
6. [DOM04]M. C. Domingo, D. Remondo, "[QoS in wireless networks: An interaction model between ad-hoc networks and fixed IP networks for QoS support.](#)" Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems, October 2004
7. [ELM05]George F. Elmasry, C. John McCann, Robert Welsh, "[Partitioning QoS Management for Secure Tactical Wireless Ad Hoc Networks.](#)" IEEE Communications Magazine v. 43 no. 11 (November 2005) p. 116-23
8. [ZHA05]Baoxian Zhang, Hussein T. Mouftah, "[QoS Routing for Wireless Ad Hoc Networks: Problems, Algorithms, and Protocols.](#)" IEEE Communications Magazine v. 43 no. 10 (October 2005) p. 110-17
9. [ZIA05]Yang Ziao, Yi Pan, "[Differentiation, QoS Guarantee, and Optimization for Real-Time Traffic over One-Hop Ad Hoc Networks.](#)" IEEE Transactions on Parallel and Distributed Systems v. 16 no. 6 (June 2005) p. 538-49
10. [LIU05]Qingwen Liu, Shengli Zhou, Georgios B. Giannakis, "[Cross-Layer Scheduling With Prescribed QoS Guarantees in Adaptive Wireless Networks.](#)" IEEE Journal on Selected Areas in Communications v. 23 no. 5 (May 2005) p. 1056-66
11. [WAN05]Kuang-Ching Wang, Parameswaran Ramanathan, "[QoS Assurances Through Class Selection and Proportional Differentiation in Wireless Networks.](#)" IEEE Journal on Selected Areas in Communications v. 23 no. 3 (March 2005) p. 573-84
12. [STI04]John A. Stine, Gustavo de Veciana, "[A paradigm for quality-of-service in wireless ad hoc networks using synchronous signaling and node states.](#)" Selected Areas in Communications, IEEE Journal on, Volume 22, Issue 7, Sept. 2004 Page(s):1301 - 1321
13. [MOO04]Bongkyo Moon, A.H. Aghvami, "[Quality-of-service mechanisms in all-IP wireless access networks](#)" Selected Areas in Communications, IEEE Journal on Volume 22, Issue 5, June 2004 Page(s):873 - 888
14. [SAL05]J. Saliba, A. Beresford, Milosh Ivanovich, Paul Fitzpatrick, "[User-perceived quality of service in wireless data networks](#)" Personal and Ubiquitous Computing, Volume 9 Issue 6, November 2005
15. [VEN04]Balasubramanian Appiah Venkatakrishnan, S. Selvakennedy, "[Wireless LANs: An enhanced HCF for IEEE 802.11e wireless networks](#)" Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems, October 2004

16. [802.11E]IEEE 802.11 Working Group, [802.11e Standard](#)

• Web Sites

1. [SWAN]"SWAN: Service Differentiation in Stateless Wireless Ad Hoc Networks," <http://www.iks.inf.ethz.ch/education/ss04/seminar/41.pdf>, [SWAN: Service Differentiation in Stateless Wireless Ad Hoc Networks]
2. [INSIGNIA]"INSIGNIA: An IP-Based Quality of Service Framework for Mobile ad Hoc Networks," <http://www.iks.inf.ethz.ch/education/ss04/seminar/42.pdf>, [INSIGNIA: An IP-Based Quality of Service Framework for Mobile ad Hoc Networks]
3. [WAG302]"NETGEAR - ProSafe Dual Band Wireless Access Point," <http://www.netgear.com/products/details/WAG302.php>, [NETGEAR - ProSafe Dual Band Wireless Access Point]
4. [MERU]"MERU: Don't Wait for 802.11e," <http://www.wi-fiplanet.com/news/article.php/3088611>, [Meru: Don't Wait for 802.11e]