IEEE P802.11

WIRELESS LAN

DATE: September 14, 1992

TITLE: WIRELESS NETWORKING REQUIREMENTS OF MULTIMEDIA APPLICATIONS*

AUTHOR: Timothy Kwok
APPLE COMPUTER, INC.
20450 Stevens Creek Blvd., MS: 76-2H
Cupertino, CA 95014
(408) 974-8311 Fax (408) 974-0781
Email: kwok@apple.com

ABSTRACT

Multimedia applications have emerged to be among the most important future applications. The purpose of this paper is to address the key issues in designing a wireless network that can flexibly and efficiently support multimedia applications. Since multimedia applications is the most general class of applications, they also have the most diverse traffic characteristics and communications requirements. This paper presents a framework of characterizing multimedia applications and their communications requirements. Also, packet switching has been known to be the most appropriate switching and multiplexing methodology for supporting multimedia applications. Therefore, this paper explores the resource allocation schemes of packet switched wireless networks. Finally, it identifies three networking services that are necessary and sufficient for supporting multimedia applications.

Notice

* Published in IEEE 1st International Conference on Universal Personal Communications, Dallas, Sept 29, 1992.

WIRELESS NETWORKING REQUIREMENTS OF MULTIMEDIA APPLICATIONS

Timothy Kwok
Apple Computer, Inc.
20450 Stevens Creek Blvd., MS: 76-2H
Cupertino, CA 95014
(408) 974-8311
Fax: (408) 974-9793
Email: kwok@apple.com

ABSTRACT

Multimedia applications have emerged to be among the most important future applications. The purpose of this paper is to address the key issues in designing a wireless network that can flexibly and efficiently support multimedia applications. Since multimedia applications is the most general class of applications, they also have the most diverse traffic characteristics and communications requirements. This paper presents a framework of characterizing multimedia applications and their communications requirements. Also, packet switching has been known to be the most appropriate switching and multiplexing methodology for supporting multimedia applications. Therefore, this paper explores the resource allocation schemes of packet switched wireless networks. Finally, it identifies three networking services that are necessary and sufficient for supporting multimedia applications.

I. INTRODUCTION

Communications networks are designed for connectivity and sharing expensive resources. Connectivity refers to providing access to people, information and processing power [1] that are located remotely. The expensive resources that need sharing include information processing devices (computers, printers, file servers, etc.) and interconnection bandwidth. For wireless communications, sharing of the interconnection bandwidth (spectrum) is essential because spectrum is not only expensive, but also inherently limited. (Note that the needs of communications and sharing resources are not mutually exclusive. For example, access to multimedia information stored in a server by various users over the network is also a means of sharing the expensive multimedia server.)

The primary objective of designing a future wireless network is to support important future applications. Multimedia communications applications have emerged to be among the most important future applications. A multimedia application is the communication of a combination of different information types (video, images, audio, text, etc.) with real-time and nonreal-time delivery requirements [2]. Since a multimedia application is the most general form of an application, it has the most diverse characteristics and networking requirements.

Therefore, a future wireless network architecture needs to satisfy several requirements. First, the architecture must be *scalable* in supporting applications for different bandwidth requirements. Second, it must be flexible in *simultaneously* supporting applications with diverse characteristics and networking requirements. Third, it is *extensible* to both *ad hoc networking* [3] (portables communicating with portables directly, without base stations) and *infrastructure-based* (accessed through base stations that are interconnected with LANs or WANs) communications. Fourth, the architecture must be very *efficient* in supporting these diverse applications to fully utilize the available scarce spectrum. Finally, the architecture must interoperate with the emerging B-ISDN standards [4,6] and Local ATM (Asynchronous Transfer Mode) specifications [5].

One of the major challenges of designing a wireless network is that the communication environment is not only constantly changing, but also unreliable. These conditions are caused by the constant joining and leaving of portables to a wireless network, as well as the significant variation in propagation conditions on multiple time scales (because of fading and interference from a variety of sources). The design objective is to share the spectrum very efficiently in spite of these significant challenges.

The purpose of this paper is to address the key issues in designing a wireless network that can flexibly and efficiently support applications with diverse networking requirements, which include both existing data applications and future multimedia applications. First, we need to define more precisely what a multimedia application is and understand its characteristics [2]. These are discussed in section II. Then, section III explains the networking requirements [2] of an application, particularly a multimedia application. To design a future network, we need to review the existing network design methologies. Therefore, we discuss existing schemes for network resource allocation, and their suitability to support multimedia applications in section IV. Finally, in section V, we propose a minimal set of networking services that are necessary to support applications with diverse characteristics and networking requirements such as multimedia applications [6].

II. CHARACTERISTICS OF MULTIMEDIA APPLICATIONS

An application is defined as a task that requires communication of one or more information streams between two or more parties that are geographically separated. The content of each information stream can be a combination of *time-based* and *nontime-based* information, with both *real-time* and *nonreal-time* delivery requirements [2].

Time-based information is defined as those that must be presented at specific instants to convey its meaning, i.e., timing (synchronization between various elements of the information) is an integral part of the information. Time-based information includes video and audio. Nontime-based information has no time information built-in to the content. Non-time-based information includes still images, graphics and text. An application may send both time-based and non-time-based information.

Real-time delivery means that the information communicated is for immediate consumption, such as voice conversation. Nonreal-time delivery means that the information communicated is for storage at the receiver for later retrieval, such as electronic mail. Therefore, real-time delivery requires sufficient bandwidth; nonreal-time delivery needs enough buffer at the receiver station.

A multimedia application is defined as one that sends any combinations of real-time and nonreal-time transfers of time-based and nontime-based information. Traditional data communications applications, namely, those supported by existing LANs are mainly nontime-based information transfer with either real-time (e.g. distributive computing) or nonreal-time (electronic mail) delivery requirements.

From a network point-of-view, an application can be specified by its traffic characteristics and the corresponding communications requirements (the later is discussed in the next section). Each stream of information sent by an application can be characterized by a sequence of bursts of bits, each burst is referred to as a message. Therefore, the traffic characteristics (or information generation characteristics) of an application are specified by its: (1) message arrival distribution; and (2) message length distribution.

An information stream can be classified according to its message arrival distribution into periodic and bursty traffic. If message arrivals occur at regular intervals, it is called periodic traffic pattern. Periodic traffic pattern is important because time-based information typically generates periodic traffic patterns. For example, conventional 64 kbps PCM audio generates samples at 125 μs intervals, each sample has 8 bits. For uncompressed full motion NTSC video, video frames (each contains a fixed amount information) are generated at regular intervals 1/30th sec or 30 frames/s. Even for compressed video, such as those generated by the MPEG algorithm [8], a video frame is still created at regular intervals of 1/30th sec for NTSC and 1/25th sec for PAL formats.

Periodic traffic does not always implies constant bit rate. Compressed video intrinsically generates variable bit rate traffic, because each video frame is compressed to different degrees depending its image complexity as well as amount of change from previous frame. However, to adapt to existing network architecture that are primarily circuit switched, compressed video is usually preprocessed to a fixed bit rate stream. This results fluctuation of video quality with complexity (information content) of the video segments at different instants.

Bursty traffic pattern is characterized by messages of arbitrary lengths generated at random time instants and separated by gaps of silence of random duration. The period of silence is typically long compared to the duration of a packet generation, resulting in the distinctively high peak to average data rate ratios. Conventional data communications are bursty because they are typically file transfers, remote logins or more recently, traffic generated by diskless workstations [10], which are all generated randomly, unlike the tight correlation in real-time transfers of time-based information. Bursty traffic is typical for non-real-time applications (carrying either time-based or non-time-based information). The unpredictability of bursty traffic, especially the instant at which a packet is generated, is the main culprit that needs to be dealt with in designing packet switched networks. However, the statistical gain made possible by multiplexing many bursty traffic sources through the network makes packet switching more efficient than circuit switching [9]. (This was the original motivation of using packet switching, instead of circuit switching, for data communications in the 1960's.)

III. COMMUNICATIONS REQUIREMENTS OF MULTIMEDIA APPLICATIONS

Communications requirements of an application depend on whether the application has real-time or nonreal-time delivery requirements. Not surprisingly, real-time applications have more stringent communications requirements than nonreal-time applications.

Real-time application requirements are best understood from a user point-of-view, because many real-time applications involve one or more users. For a user communicating through a portable over a wireless network, the most important concern is the interactiveness and the quality of running any networking applications. This implies two networking requirements: response time and bandwidth. First, the response time for each request experienced by the user must be short (or comparable to communications with devices on the desktop). For example, if a user wants to read a file over the network, the response time for paging different parts of the document should be comparable to that of reading a file stored locally. (Note that currently, each page fault requires a burst of data transfer on the order of 10 Kbytes [10].) Second, the quality of the application should not degrade because of the user is communicating over the network. For example, if a user wants to view a movie over the network in real-time, the quality of the movie must be comparable to that from a local storage device.

To summarize, the primary communication requirement of real-time applications is sufficient networking bandwidth—on-demand and sustained. For real-time transfers of nontime-based information, instantaneous bandwidth must be available on-demand to guarantee response time of the applications. (Instantaneous bandwidth is simply the amount of information per transaction divided by the required response time.) For real-time transfers of time-based information, sufficient sustained bandwidth for the duration of the session is necessary to guarantee the quality of the applications. Table I and II show the diverse bandwidth requirements of real-time applications transferring time-based and nontime-based information, respectivlely.

Therefore, to guarantee the quality of a real-time multimedia application in a wireless network, the network architecture should require each new application to request the network for the required bandwidth before a connection be setup by the network (to ensure that there is sufficient bandwidth for this new application available). Also, the architecture can guarantee the bandwidth for the new application once it is accepted by the network. This implies two criteria for the architecture: it is connection-oriented and reservation-based.

For all other applications, namely, nonreal-time applications, there is no specific communications requirements. However, one can still characterize the network bandwidth requirement to support such applications by the amount of information divided by the maximum delay allowed by the applications.

More specifically, the communication requirements of an application (or each information stream within the application) can be classified into macrorequirements and microrequirements. Macrorequirements characterize the application's overall information transfer needs in terms of following parameters (1) bandwidth (average, peak, etc.) (2) delay (absolute delay, transmission delay), (3) error rate, (4) duration of session, (5) transaction rate, (6) information transfer per transaction, (7) total information transfer. Typically, only a subset of these parameters may be applicable in specifying an application. For example, real-time delivery of video (time-based information) can be characterized by its bandwidth, error and duration requirements only. Macrorequirements are specified during the connection setup phase.

Microrequirements characterize the needs of individual message transaction of an application, which include one or more of the following parameters: (1) Absolute delay; (2) Delay variances; (3) Error (average and burst). The idea of microrequirements of an application is very similar to the QOS (quality of service) guaranteed by the network. This should be implicitly guaranteed by the network once the connection is granted.

These communication requirements are not independent. Bandwidth is tightly related to both absolute delay and delay variance, while delay and loss rate can be tightly coupled. For example, in packet speech communication, each speech sample has a maximum delay constraint. If it is exceeded, the sample is useless for the receiver and the sample is considered lost in the network; if the network introduces more delay, the loss rate for the audio connection increases correspondingly [2].

IV. PACKET SWITCHED NETWORKS AND RESOURCE ALLOCATION SCHEMES

Packet switching has emerged to be the most appropriate switching and multiplexing methodology

[6]. The reasons are that it is not only very flexible in supporting applications with different bandwidth (delay, error, etc.) requirements (i.e., scalable), but also very efficient in simultaneously supporting many applications with diverse characteristics. (That is why B-ISDN is based on packet switching, in the form of ATM [6]). To design a wireless packet switched network to support multimedia applications with diverse communications requirements [2], it is necessary to understand the networking resources available in a packet switched network; these resources depend on whether the packet switched network is a shared medium type or point-to-point type network.

A shared medium type network has a shared communication channel directly accessible by its users. Each packet sent by any user is automatically broadcast to all other users on the network. This implies both an advantage and a disadvantage. The advantage is that the bandwidth resources consumed by a multicast or broadcast packet are the same as that for point-to-point packet. The reason is that every user automatically receives all the packets sent out by any user, and filters away those packets not addressed to itself. Therefore, it is not necessary to send individual copies to different destinations. However, the bandwidth cost of a two way connection is the sum of the two one-way connections, because each party of the two-way connection needs to consume the bandwidth from the same shared medium. Examples of shared medium networks include single hop wireless networks (such as ad hoc networking [3], or portables communicating directly with a base station) and shared bus networks (such as Ethernets [7]).

The only networking resource in the shared medium type networks is the networking bandwidth, i.e., the shared spectrum. The allocation of such resource to individual user is according to a medium access control (MAC) protocol, for which the design of a MAC protocol is essential to guarantee the performance of multimedia applications. However, although many MAC protocols have been proposed since the pioneering Aloha protocol [11] invented twenty years ago, very few of them were designed to support future multimedia applications.

A point-to-point type network is made up of a set of point-to-point links interconnected by switching nodes. A workstation accesses a point-to-point network either directly to a switching node, or through an intermediate access point that concentrates traffic from several workstations and multiplexes them to a switching node. Examples of point-to-point networks include the Internets [12] and ATM networks [9], for which the switching nodes are routers and fast packet switches [13], respectively. Their networking resources include not only transmission and switching bandwidth, but also buffers at the switching nodes and network access points. The allocation of these networking resources occurs at the switching nodes and the network access points. Each switching node determines which packet gets transmitted or delayed according to a scheduling algorithm, and which packet gets buffered or dropped according to a dropping algorithm. The most common scheduling and dropping algorithm is first-come-first-serve, which is used in traditional connectionless packet switched networks such as the Internet. However, this is not suitable for supporting multimedia applications with guaranteed performance, as explained below. Optimal scheduling and dropping algorithms for supporting multimedia applications are under active research currently.

A wireless network infrastructure probably includes both types of packet networks. Both ad hoc wireless communications and portables communicating directly with a base station would create shared medium type networks. However, very often, the portables cannot be communicated directly through the wireless medium (due to out-of-range, among other factors). Then, they may need to access an infrastructure made up of base stations interconnected by an internet or ATM networks to establish communications. This means that the portables now communicate over both shared medium and point-to-point networks.

Also, we can classify existing packet network resource allocation schemes into contention-based and reservation-based types (for both shared medium and point-to-point networks). In a contention-based network, each user requesting a data transfer will be supported by all the available network resource present at that instant; the delay experienced by the user depends on instantaneous traffic load on the network. The effective bandwidth available to each user depends on the current total traffic load generated by all the other active users. Since no guarantee of bandwidth or delay can be made to a user, contention-based schemes cannot support real-time multimedia applications (unless the network is very lightly loaded and bandwdith is abundant). A classical example of contention-based shared medium type network is the Aloha packet radio network [11]. For point-to-point networks, the connectionless service supported by IP-based [12] and Appletalk™*-based [14] internets are all based on contention-based resource allocation.

* Appletalk™ is a trademark of Apple Computer, Inc.

For a reservation-based network, before a user begins a communication application, the user must first reserve the necessary resource from the network. This application is permitted only if the network has determined that it has sufficient idle resources to serve this user. This implies that a reservation-based resource allocation scheme offers a connection-oriented service to the applications. Since the network resources are reserved ahead of time and guaranteed during the connection, the reservation-based schemes are suitable for supporting general multimedia applications (especially real-time applications). The classical example of a reservation-based network is a circuit-switched network such as the PSTN (public switched telephone network) and the cellular telephone network. However, this is not suitable for supporting multimedia applications, because circuit switching is very inflexible and inefficient to support applications with diverse communications requirements [15]. It is more appropriate, however, to extend the ATM reservation-based model in wired networks is to support multimedia application in designing future wireless networks.

V. CLASSES OF NETWORK SERVICES

To design a future wireless network architecture needed to support all possible applications, which include all traditional data applications and future multimedia applications, it is very desirable to identify the minimum classes of network services required. It follows from the previous discussion that three main classes of services are necessary and sufficient [6].

The first class is the best-effort delivery class. It is for supporting nonreal-time applications. This service class does not require an application to reserve network resource explicitly before it can begin data transfer. However, the network does not guarantee the amount of resource available to this application (or any other applications support by this service class). This is the model employed in traditional data communication application, for which data transfers do not reserve networking resources (such as bandwidth or buffer space), and each user essentially expects a "send-and-pray" quality of service without guarantee. It is called best-effort delivery because for each data transfer requested by a user, the network will attempt to use all the existing network resource available currently (which is dynamically changing and unpredictable), depending on the existing load. In other words, this is a contention-based model of network resource allocation. This quality of service is sufficient for existing data applications only when there is no congestion. When the traffic load is high, congestion begins to occur and the delay introduced to each application may become intolerable. However, it is essential for the network service to ensure fairness under congestion conditions. Another important characteristic for best effort delivery service is that it is not necessary for a user to specify the QOS or communication requirement to the network. The application does not care as long as the response time it gets is reasonable, until the network is too congested and it voluntarily withdraws from the network.

Both the second and third classes provide reservation-based services to the real-time applications to guarantee performance. They are called time-based reservation class and nontime-based reservation class, respectively. Time-based reservation class is intended to support real-time delivery of time-based information, while the nontime-based reservation is for real-time delivery of nontime-based information. They are reservation-based service because both require an application to request the necessary resource before data transfer can occur; the network needs to allocate such resource to the application and data transfer can begin only if such resource is available and the network grants such resource to the applications. Both these two classes provide connection-oriented services. The difference between them lies in the traffic characteristics and communications requirements of the applications they intend to support. For time-based real-time data transfers, the traffic characteristics are periodic and the main communication requirements are sustained bandwidth and upper bound on absolute delay and delay variance (e.g. life video communications). For nontime-based real-time applications, the traffic characteristics are bursty while the communications requirements are upper bounded on response time for specified number of transactions in a period with random request arrivals (e.g. image browsing). Finally, all three classes are expected to satisfy a certain upper bound on error probability.

VI. CONCLUSION

This paper discusses the requirements of supporting multimedia applications in wireless networks. It provides framework of characterizing multimedia applications and its communications requirements. Also, it explores the network resource allocation schemes that are suitable multimedia applications. Finally, this paper identifies three networking services that are necessary and sufficient for supporting all possible applications.

REFERENCES

[1] J. Patterson, C. Egido," Three Keys to the Broadband Future: A view of Applications," IEEE Network Magazine, March 1990, pp. 41-47.

[2] T. C. Kwok, "Communications Requirements of Multimedia Applications: A preliminary study," Proc. International Conference on Selected Topics in Wireless Communications, Vancouver, B.C., Canada, June 25-26, 1992.

[3] J. Lovette, "Data-PCS," Petition for Rulemaking to FCC for Amendment of Section 2.106, Jan. 28, 1991.

[4] A. Day," International Standardization of BISDN," IEEE LTS Magazine, vol. 2, No. 3, Aug. 1991, pp. 13-20.

[5] Apple, Bellcore, Sun, Xerox,"Network Compatible ATM for Local Network Applications" Phase 1, Version 1.0, April 3, 1992.

[6] T. Kwok," Reply comments on Wireless LANs Requirements Document (IEEE P802.11/92-20)," included in IEEE P802.11/ 92-48, March 10, 1992.

[7] R. Metcalfe and D. Boggs,"ETHERNET: Distributed packet switching for local computer networks," Communications of ACM, vol. 19, pp. 395-403, July, 1976.

[8] D. Le Gall, " MPEG: A Video Compression Standard for Multimedia Applications," Communications of the ACM, vol. 34, no. 4, April 1991, pp. 46-58.

[9] T. Kwok,"Broadband ISDN to Support Multimedia and Collaborative Applications," Proc. IEEE Compcon Spring 1992, pp. 21-24.

[10] R. Gusella," A Measurement Study of Diskless Workstation Traffic on an Ethernet," IEEE Transactions on Communications, Vol. 38, No. 9, September 1990, pp. 1557-1568.

[11] N. Abramson,"The ALOHA system—Another alternative for computer communications" in Proc. Fall Joint Computer Conf. AFIPS Press, 1970, vol. 37, pp. 281-285.

[12] D. Comer, Internetworking with TCP/IP, vol. 1, 2nd Ed., Prentice Hall.

[13] T. Kwok, "Tandem Banyan Switching Fabric: A New Fast Packet Switch Architecture And Its Performance Evaluations," Ph.D Dissertation, Stanford University, Sept., 1990.

[14] G. Sidhu, R. Andrews and A. Oppenheimer, Inside Appletalk, 2nd Ed., Apple Computer, Inc., Addison Wesley, 1990.

[15] S. Minzer, " Broadband ISDN and Asynchronous Transfer Mode (ATM)," IEEE Communications Magazine, Sept. 1989, pp. 17-24,57.

[16] M. Liou, " Overview of the p x 64 kbit/s Video Coding Standard," Communications of the ACM, vol. 34, no. 4, April 1991, pp. 59-63.

[17] R. Jurgen," The challenges of digital HDTV," IEEE Spectrum Magazine, April 1991, pp. 28-30, 71-73.

[18] W. Luplow, " Digital High-Definition Television Takes Off," IEEE Spectrum Magazine, January 1991, pp. 65-68.

[19] H. Musmann," The ISO Audio Coding Standard," IEEE Globecom 1990, pp. 511-517.

Video	Uncompressed	Compressed
videoconference	9/36 Mbps	p x 64 kbps (H.261) [16]
NTSC	~ 200 Mbps	1.5 Mbps (MPEG) [8]
HDTV	~ 1 Gbps	20 Mbps [17, 18]
Audio		
voice telephony	64 kbps	16 kbps
CD Quality Stereo	1.4 Mbps (2x706 kbps)	256kbps (MPEG) [13] (2x128kbps)

TABLE I Bandwidth requirements real-time applications delivering time-based information.

Images	Uncompressed Mbytes	Compressed (JPEG) Mbytes	Peak Bandwidth	
Response time			0.1 sec	10 mins
Photo: 1k x 1k x24 bit	3	0.06 - 0.3 (Lossy)	4.8 - 24 Mbps	8 - 40 kbps
X-ray: 2kx2kx12 bit	6	3 (Lossless)	240Mbps	0.4 Mbps

TABLE II Bandwidth requirements real-time applications delivering nontime-based information.