WIRELESS DIGITAL AUDIO DELIVERY ANALYSIS AND EVALUATION

Nicolas-Alexander Tatlas[1], Andreas Floros[2] and John Mourjopoulos[1]

[1]Audiogroup, Department of Electrical Engineering and Computer Technology, University of Patras, Patras, Greece
[2]Department of Informatics, Ionian University, Corfu, Greece

ABSTRACT

Real-time digital wireless playback of CD-quality audio in multipoint setups using Quality of Service enhancements is analyzed and evaluated in this work. A novel methodology is introduced for simulating wireless digital audio delivery as well as for theoretically deriving the playback distortions. This methodology allows the accurate wireless digital audio delivery and reproduction simulation and leads to significant results for both the wireless networking and audio playback performance, while it provides a framework for defining the optimal parameters for error-free wireless stereo audio reproduction.

1. INTRODUCTION

In recent years, the deployment of Wireless Networks in home, professional and public environments has raised the expectations and requirements for a new generation of networking applications with enhanced mobility capabilities. However, despite recent advances on the data rates supported by the Wireless Local Area Network (WLAN) technology, multimedia streaming cannot be yet efficiently implemented over WLANs, due to their inability to provide strict service guarantees to time-critical audio/video data flows.

Given the multiple possibilities for sourcing multichannel audio (DVD, PC, TV receivers etc), the need for wireless multichannel audio streaming in the home environment free of complicated connections and multiple cables is nowadays more demanding than ever. Towards this aim, the current IEEE 802.11E draft specification provides the necessary WLAN Quality of Service (QoS) support and is expected to create a new range of wireless audio products (such as mobile playback devices, wireless loudspeakers for “home theater” applications and wireless audio servers [1]) that will operate within a digital, networked, cable-free home environment. Moreover, although QoS is the basic requirement for point-to-point wireless applications (e.g. audio distribution from the source to a single audio receiver), the development of wireless audio products for multichannel playback environments, raises additional implementation issues, such as inter- and intra-stream synchronization [2] employed through various control schemes.

A number of studies have been recently published aiming to develop and evaluate network time synchronization algorithms for multiple wireless stations [3]. Furthermore, the quality assessment of multimedia applications over packet-based networks [4] as well as the development and evaluation of mechanisms for maintaining the quality of packet-audio within acceptable limits [5] has been also investigated. However, the focus in those works was mostly on lower quality audio streams suitable for teleconferencing and videoconferencing applications.

This work analyzes the real-time audio playback scenarios defined under such a protocol, resulting from the variable transmission delays and packet losses occurring during the wireless transmission of high-quality, multichannel audio traffic streams (TSs), under the 802.11E protocol rules. This presents a novel and challenging topic, since only a limited number of digital (and analogue) wireless systems have been realized up to now for use in multichannel audio applications, and these via custom transmission protocols that cannot provide fundamental interoperability and interference guarantees.

The analytic derivation and measurement of the actual relative audio channel delay and its effect on the playback quality is performed using a novel wireless audio playback simulation methodology [6], described in detail in Sections 3 and 4. The proposed methodology can be easily extended for any type of audio applications, for both uncompressed and compressed (e.g. MPEG) data, in stereo or multichannel formats.

2. OVERVIEW OF QoS OVER WLANS

The limited QoS performance of the legacy 802.11 standard has raised requirements for QoS enhancements, currently being developed by the 802.11E Task Group. The proposed Medium Access Control (MAC) layer employs two wireless channel access methods: a) the Enhanced Distributed Channel Access (EDCA), which provides distributed access using traffic differentiation and b) a centralized access method termed as Hybrid Controlled Channel Access (HCCA). In both cases, transmission rights are granted to the remote receivers within specific time intervals called Transmission Opportunities (TXOPs).

A previous work [7] has shown that the random access nature of EDCA cannot provide strict service guarantees. HCCA fulfills this requirement by centrally scheduling the access for all associated receivers. The access schedule is calculated by the Hybrid Coordinator (HC), located in the Access Point (AP) device, taking into account the admitted traffic requirements described by appropriate Traffic Specification (TSPEC) elements. The schedule is implemented through a polling mechanism, with each TXOP defined by an implicit starting time and a maximum duration. The 802.11E specification additionally defines the minimum schedule requirements, satisfied by the "simple scheduler" reference design. Recently, an adaptive scheduling scheme has been proposed termed as Scheduling based on Estimated Transmission Times - Earliest Due Date (SETT-EDD) [8], which is also considered here.

3. THEORY

The basic system under study (Fig. 1) consists of one digital audio source integrated with a wireless AP and at least two wireless audio receivers, responsible for the left and right channel reproduction.

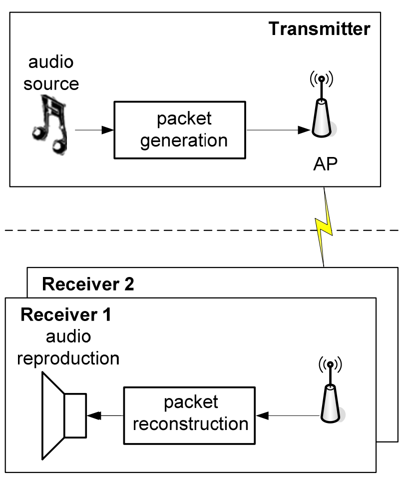

Fig. 1. Configuration of a stereo audio wireless system

3.1. Playback Analysis

The audio source generates digital audio samples at a constant rate of Nf_s (bytes/sec), where N is the number of bytes per sample and f_s (Hz) is the digital audio sampling rate. These samples are organized as packets of length L_p bytes. Assuming an integer $i=[1,2,3\ldots)$, the i-th packet generation time is:

$$t_g(i) = i\frac{L_p}{Nf_s} \tag{1}$$

Each generated packet is inserted in the AP transmission (Tx) buffer at instances $t_g(i)$. If the buffer is already full, it will be dropped. Hence, the i-th packet buffering time $t(i)$ equals to:

$$t(i) = t_g(i)\delta_{Tx}(i) \text{ where } \delta_{Tx}(i) = \begin{cases} 0, \text{if the Tx queue is full} \\ 1, \text{else} \end{cases} \tag{2}$$

If $d_{Txbuffer}(i)$ is the time delay between the packet insertion in the Tx queue and its transmission to the receiver, the i-th packet transmission time instance will be equal to:

$$t_{Tx}(i) = t(i) + \delta_{Tx}(i)d_{Txbuffer}(i) \tag{3}$$

Furthermore, provided that $\delta_{Tx}(i) = 1$, the buffering delay is:

$$d_{Txbuffer}(i) = d_{Txbuffer}(j) - (i - j)\frac{L_p}{Nf_s} + d_{Tx}(i) \tag{4}$$

with j being the index of the previous packet inserted in the Tx queue. $d_{Tx}(i)$ represents the time interval between the movement of the i-th packet to the first position of the Tx queue and its successful transmission. This delay depends on the next polling instance of the corresponding TS that results into successful packet delivery.

Upon successful reception, each packet is stored in the receiver (Rx) queue at instances:

$$t_{Rx}(i) = t_{Tx}(i)\delta_{Rx}(i) \tag{5}$$

where δ_{Rx} is defined as in eq. 2, in this case for the Rx buffer. The playback time of the i-th packet first sample equals to

$$t_p(i) = t_{Rx}(i) + \delta_{Rx}(i)d_{Rxbuffer}(i) \tag{6}$$

where $d_{Rxbuffer}(i)$ is expressed as:

$$d_{Rxbuffer}(i) = d_{Rxbuffer}(j) - [t_{Rx}(i) - t_{Rx}(j)] + \frac{L_p}{Nf_s} \tag{7}$$

and j is the index of the last packet in queue before the insertion of the i-th packet. Assuming that

$$\delta(i) = \delta_{Tx}(i)\delta_{Rx}(i) \tag{8}$$

the packet reproduction time $t_p(i)$ can be defined as:

$$t_p(i) = \delta(i)\,[t_g(i) + d_{Txbuffer}(i) + d_{Rxbuffer}(i)] \tag{9}$$

Hence, the parameter

$$d(i)=d_{Txbuffer}(i)+d_{Rxbuffer}(i) \tag{10}$$

represents the total delay (depicted in Fig. 4) induced by the wireless network, between the packet generation and its final playback time. Packet losses will cause the playout delay of proceeding packets to fluctuate, since the corresponding $t_{Txbuffer}$ and $t_{Rxbuffer}$ time instances will vary (see Fig. 4).

3.2. Reproduction Distortions

If a packet is dropped from either the Rx or Tx queue (that is $\delta(i)=0$), an audible discontinuity in playback will occur. In the case of a stereo audio setup, this would additionally cause relative channel phase shifts, unless an application level time alignment strategy is used. This situation can be met under heavy traffic congestion (e.g. excessive MAC layer retransmissions). In case the reproduction has been halted when receiving the packet, i.e. if

$$t_p(i) - t_p(j) > \frac{L_p}{Nf_s} \qquad (11)$$

then silence gaps will be heard during playback. In a stereo setup, out-of-phase distortions will also occur (one channel lagging). Obviously, both conditions occur when:

$$\begin{cases} t_p(i) - t_p(j) > \dfrac{L_p}{Nf_s} \\ j \neq i-1 \end{cases} \qquad (12)$$

In such a case, the phase difference between the receivers' reproduction depends on the number of packets lost (*i-j-1*) and the time elapsed from the playback halting until the reproduction resumes $(t_p(i) - t_p(j) - \frac{L_p}{Nf_s})$.

The distortions introduced are closely related to the throughput measured for packets delivered only within their TSPEC maximum Service Interval (max*SI*) bound, termed here as non-delayed throughput (T_{ND}) being equal to:

$$T_{ND} = \frac{1}{I_{total}} \sum_{i=1}^{I_{total}} \frac{i \cdot L_f}{d_{Tx}(i)} \text{ for } \delta(i) \neq 0 \text{ and } d_{Tx}(i) < \max SI \qquad (13)$$

where I_{total} the total number of packets transmitted.

Obviously, in case T_{ND} is less than the target, significant reproduction distortion (relative channel delay, discontinuities and silence gaps) will be introduced.

4. SYSTEM ARCHITECTURE SIMULATION

The simulation of the wireless digital audio data playback process and the evaluation of the related playback distortions was performed using a novel computer-based methodology [7] illustrated in Fig. 2. The simulation platform consists of three sub-systems: (a) a simulation pre-processing stage, which converts the audio data to a format appropriate for input to the wireless simulator (b) The wireless simulator and (c) a simulation post-processing stage deriving a new audio file containing the "reproduced" replica of the original digital audio data. This audio (wave) file can be later used for extracting distortion information for each channel, as well as for evaluating any phase differences between them.

As it is shown in Fig. 2, the audio data exchange between the above modules is realized using trace files, which contain information for the frame generation time and the corresponding packet length for a specific TS. A trace file maps the transmitted PCM data packets to specific segments of stereo wave files containing the original digital audio. This mapping includes parameters such as the desired pure audio data packet length L_p (in bytes) plus a packet header equal to 48 bytes, since the User Datagram Protocol (UDP) was employed as the transport layer protocol.

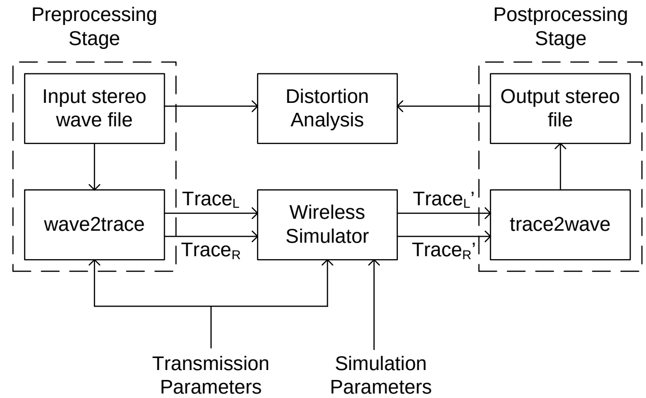

Fig. 2. Block diagram of the proposed wireless audio playback simulation methodology

The wireless simulator produces one output trace file per serviced TS, containing information such as whether the corresponding packets were transmitted or not, and the packet delay induced. These files are mapped to the stereo audio samples using an application realizing a First-In First-Out (FIFO) reception (Rx) queue, which is filled with the successfully received data packets. Upon playback, samples are read from the Rx queue at a rate equal to the audio sampling frequency f_s (Hz). The Rx queue length and the initial pre-buffering interval are defined by the user in multiples of 100ms (the nominal beacon transmission period defined by the 802.11 protocol). For all test cases considered here, the Tx and Rx queue lengths were equal to 4608 and 10000 bytes (or 5000 samples) respectively, while the initial playback latency was set to 1 beacon interval.

The wireless simulator employs the trace files to generate the appropriate traffic for a given simulation scenario. Upon a TS generation, the corresponding trace file information is also used for calculating the TSPEC parameter values, which will be used by the HC in order to calculate the access schedule. Since PCM audio has a constant bitrate, the TSPEC values depend only on the selected maxSI value (the maximum allowed time interval between two successive TXOPs granted to a specific TS), the selected audio source packet size L_p (in bytes) and the number n_L of MAC layer packet fragments of length L_f (also in bytes) that will be allowed to be transmitted within the allocated TXOP.

Table I shows the selected and calculated TSPEC parameter values. Although the pure CD-quality audio data rate for each channel equals to 705.6kbps, the transmission data rate is always higher, due to the addition of the UDP header and the possible fragmentation of the original data packets within an SI. Both the simple and SETT-EDD schedulers were considered. All transmissions were performed using the highest physical layer rate supported by the 802.11b protocol (11Mbps) under typical channel conditions as well as under heavy electromagnetic interference.

Table I

TSPEC for CD-quality PCM audio and simulation parameters

SI (msec)	10		20
L_p (bytes)	882		1764
n_L	1	3	1
L_f (bytes) (L_f+UDP header)	882 (930)	294 (342)	1764 (1812)
maxSI (msec)	10	10	20
Mean Data Rate (kbps)	744	820.8	724.8
Scheduler type	Simple, SETT-EDD		
Channel model	Typical, heavy interference		

5. RESULTS AND CONCLUSIONS

Fig. 3 shows the measured T_{ND} values (see eq.13). The sequence of tests performed has shown that under normal channel conditions, the HCCA QoS mechanism provided by the 802.11E layer can guarantee an uncorrupted, error free, high quality audio streaming and playback, for packet sizes of L_f 882 and 1764 bytes. However, the presence of high electromagnetic interference may significantly degrade the playback performance, even for high speed rates (11Mbps).

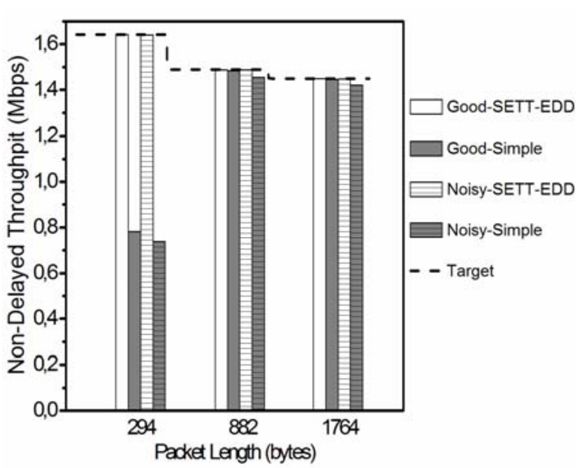

Fig. 3. Non-Delayed Throughput, 11Mbps PHY

A typical case scenario example is shown in Fig. 4, which shows the measured playout delay *d(i)* for one receiver over a 60 seconds simulation duration using the simple scheduler, under heavy interference. Note that the initial delay equals to 4410 samples, as a 100ms pre-buffering stage is employed prior to playback starting time; the delay decreases when discontinuities occur due to packet losses and locally increases when silence gaps are introduced.

The simple scheduler introduces significant data losses ($\delta=0$) and packet delays d_{Tx}, varying the playout delay. In contrast, almost acceptable playback quality is achieved for L_f=882 bytes (Fig. 4b). The audibility of the distortions introduced was confirmed through listening tests in a typical stereo reproduction environment. A set of the corresponding audio material can be obtained from [9]. In the case of the SETT-EDD scheduler, the network bandwidth was found to be optimally allocated: no buffer overflows occur during the overall simulation interval, leading to absolutely error free audio reproduction.

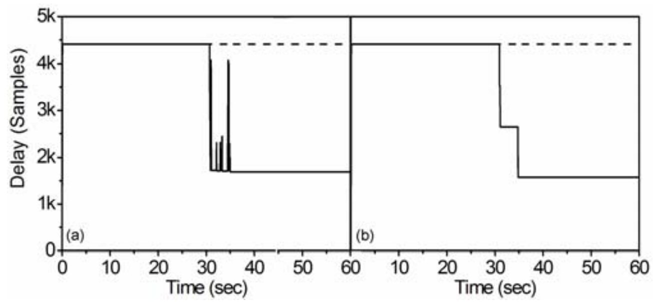

Fig. 4. Total playout delay for 11Mbps PHY under heavy interference, Simple Scheduler (a) L_f=294, (b) L_f=882. Doted Line shows error-free transmission delay.

From the above results, it is clear that the use of an adaptive access schedule can significantly improve the overall playback performance. However, the implementation complexity of such mechanisms in embedded systems significantly increase the overall design and product cost. Hence, the evolution of advanced application-level remote synchronization mechanisms combined with error concealment techniques would effectively compensate the playback distortions induced by the wireless network.

7. REFERENCES

[1] A. Floros, T. Karoubalis, "Delivering high-quality Audio over WLANs", Audio Eng. Soc. 116th Convention, Berlin, May 2004 (preprint 5996).

[2] H. Zhu, GP Zeng, I. Chlamtac, "Control scheme analysis for multimedia inter- and intra-stream synchronization", IEEE International Conference on Communications, May 2003.

[3] P. Blum and L. Thiele, "Trace-Based Evaluation of Clock Synchronization Algorithms for Wireless Loudspeakers", 2nd Workshop on Embedded Systems for Real-Time Multimedia (ESTIMedia'04), Stockholm, September 6-7, 2004.

[4] M. Claypool and J. Riedl, "End-to-End Quality in Multimedia Applications", Computer Science Technical Reports, Worcester Polytechnic Institute, WPI-CS-TR-98-18, 1998.

[5] S. B. Moon, J. Kurose, and D. Towsley. "Packet audio playout delay adjustment: Performance bounds and algorithms", ACM/Springer Multimedia Systems, 6:17--28, Jan. 1998.

[6] N.-A. Tatlas, A. Floros, J. Mourjopoulos, "An Evaluation Tool for Wireless Digital Audio Applications", Audio Eng. Soc. 118th Convention, Barcelona, May 2005 (preprint 6386).

[7] A. Floros, T. Karoubalis and S. Koutroubinas, "Bringing Quality in the 802.11 Wireless Arena", Broadband Wireless and WiMax IEC Comprehensive Report, ISBN: 978-1-931695-30-x, January 2005.

[8] A. Grillo, M. Macedo, and M. Nunes, "A Scheduling Algorithm for QoS Support in IEEE802.11E Networks", IEEE Wireless Communications, 2003, 10, (3), pp. 36 – 43.

[9] http://www.wcl.ee.upatras.gr/audiogroup/demos/