SWARM SYNCHRONIZATION FOR MULTI-RECIPIENT MULTIMEDIA STREAMING

[1)]Mika Rautiainen, [1)]Hannu Aska, [1)]Timo Ojala, [1)]Matti Hosio, [2)]Aki Mäkivirta and [2)]NikoHaatainen

[1)] MediaTeam Oulu, Department of Electrical and Information Engineering
University of Oulu, Finland
firstname.lastname@ee.oulu.fi

[2)] Genelec Oy
Iisalmi, Finland
firstname.lastname@genelec.com

ABSTRACT

IP networks allow constructing versatile device configurations for multimedia streaming. However, the stochastic nature of the packet-switched data transmission may complicate IP-based implementations of some conventional applications such as analog wired transmission of synchronized multi-channel audio. This paper introduces a multimedia streaming system based on the synchronization of multiple playback clients as a 'swarm'. The proposed 'swarm synchronization' mechanism is based on precise clock synchronization with the PTP protocol and adjusting the client-specific sampling rates according to the true playback rates of other clients. A streamlined version of the RTP protocol is employed to minimize playout delay. The proposed system is empirically evaluated in wired Ethernet LAN and in wireless IEEE 802.11g LAN. The experimental results show that in the Ethernet network the proposed streaming system is able to achieve very precise synchronization.

Index Terms — clock synchronization, Precision Time Protocol, IEEE 1588, multimedia streaming

1. INTRODUCTION

One important factor in multimedia streaming is to synchronize the playback of the elementary streams of a multimedia object with sufficient precision not to disturb the human perception. A familiar example is lip synchronization, which refers to the synchronization of speaker video with the audio of the speaker's voice. Steinmetz has studied the impact of synchronization jitter in various multimedia applications [1]. He found that lip synchronization tolerated up to 80 ms jitter between the visual and auditory signals to be imperceptible by human recipients. In other multimedia scenarios jitter for good synchronization quality ranged from 500 ms (loosely coupled audio, such as speaker and background music) to 11-20 μs (tightly coupled audio, such as stereo channels creating an auditory image) [2].

In a simple multimedia streaming application the entire multimedia object is delivered to a single recipient, e.g. a multimedia player, which constructs the playback from the elementary streams. In this study we consider a more complex application of streaming a multi-channel audio stream to multiple recipients, which are supposed to playback the individual channels back in a precisely synchronized fashion. The application has very strict performance requirements in terms of small end-to-end latency and precise synchronization of the playback between the multiple recipients. Functional requirements include flexible device configuration, scalability to larger number of recipients, straightforward deployment in different IP networks, and implementation without any special purpose hardware.

Several multimedia applications have been developed for synchronized audio streaming in IP networks such as PulseAudio [3], SqueezeCenter [4] and Axia IP-Audio Driver [5]. The last is part of a professional product suite involving dedicated hardware, while the first two are open source implementations suitable for applications with less stringent synchronization requirements. Melvin and Corcoran [10] introduced a system for synchronized playback through networked home appliances. The system used local playback adjustment using NTP synchronized clocks, which limits synchronization accuracy between devices to the order of milliseconds. Similar accuracy was obtained by Young *et al*. [11].

We present a multimedia streaming system based on 'swarm synchronization' of multiple playback clients. The proposed 'swarm synchronization' mechanism uses the PTP (Precision Time Protocol) protocol for synchronizing the clocks of the playback clients. The clients exchange information on each other's true playback rates and adjust their sampling rates according to the 'slowest' client. A streamlined version of the RTP protocol is employed to minimize playout delay. The proposed system is empirically evaluated in wired Ethernet LAN and in wireless IEEE 802.11g LAN.

2. MULTI-RECIPIENT DELIVERY WITH PRECISELY SYNCHRONIZED PLAYBACK

2.1 System architecture

Figure 1 shows the system architecture comprising of a streaming server, multiple playback clients and a network. The server sends the multi-channel stream to the swarm of clients using IP multicast. The clients join the swarm (multicast group) automatically upon receiving a multicast inquiry from the server. Upon joining the swarm the client also establishes unicast TCP control channel with the server for the purpose of dynamic swarm configuration (e.g. channel selection, client volume on/off).

The server sends the interleaved multi-channel stream to the multicast address of the swarm. This means that every client receives all media streams, but a client playbacks only the channel configured by the server. The fact that all clients receive all streams allows rapid re-configuration of the swarm without loss of synchronization. If there is no active media stream to send, the server keeps sending a 'zero signal' to maintain the synchronization between the clients.

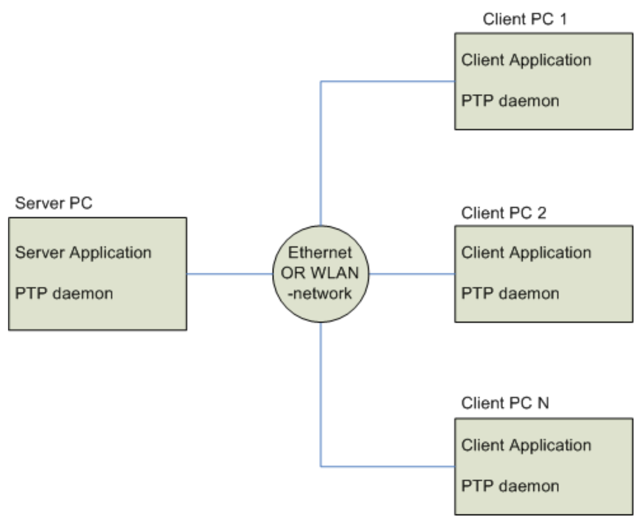

Figure 1. System architecture

The control channels are maintained by periodical alive messages. If dynamic control of the clients were not needed, e.g. with local configuration, server and client swarm could manage multi-recipient multimedia playback without any control channels, since the swarm synchronization takes entirely place between clients, independently of the server, as described in section 2.4

A client device executes a PTP process to synchronize its system clock time with the clocks of other client devices. A client software receives configuration commands from the server, media streams from the multicast address, sends synchronization messages to the swarm, adjusts its sampling rate and takes care of audio playback.

2.2 Multimedia transport

UDP was chosen as the transport protocol, as in comparison to TCP it provides finer control in terms of what data is sent and when and has lower protocol overhead.

RTP [8] is the protocol of choice for multimedia transport. The RTP specification includes sister protocol RTCP for synchronization and control purposes. However, RTCP is not designed for high precision playback synchronization of tens of microseconds between multiple recipients. To minimize the end-to-end latency we created our own streamlined RTP protocol, where QoS related features such as RTCP protocol and jitter calculation were left out of the implementation. We also employed a simple FEC (Forward Error Control) mechanism [9] as a protection against packet loss, which is very probable in wireless data transmission. FEC packets are calculated with simple XOR parities. Prior to the transmission of the next audio packet, system sends FEC codes from the previous and next packets. This gives low processing overhead but increases data bandwidth two-fold. FEC implementation ensures that the system is able to recover from the loss of two sequential packets.

2.3 Clock synchronization with PTP

PTP [6] is a protocol for accurate time synchronization in Ethernet networks. The protocol is based on slave-master architecture. The slave and the master devices periodically send messages containing send and receive timestamps. These timestamps are then used for calculating the difference between the master and slave clocks, to steer the system clocks towards a common wall clock time. The timestamps are usually received from the Network Interface Card (NIC) driver to achieve maximum accuracy and are typically used together with a specialized hardware.

PTP also uses a feedback loop with a Proportional-Integral (PI) controller for correcting both time and rate of the local clock. PTP works best in symmetrical networks achieving sub-microsecond clock accuracy that makes it better wall clock alternative than the commonly used NTP. PTP has also been implemented as an open source, software-only solution (PTPd) where special attention was put on low resource usage [7].

2.4 Swarm synchronization of playback clients

The main challenge in precisely synchronizing the playback of multiple clients is to handle the small variations in the audio consumption rates of the clients. Since the sub-millisecond synchronization precision required by our application could not be achieved with existing solutions, we have developed the swarm synchronization mechanism. The clients exchange precise information about each others' audio consumption with UDP multicast messages. The net consumption rate is determined from the ratio of audio playback buffer consumption rate and incoming audio data stream rate. The synchronization messages, containing net consumption rates, time points and sample numbers, are sent periodically to the swarm members, but client specific period start times are random to avoid bunching of messages.

Knowing the synchronization data from all swarm members, a client is able to identify what sample the other clients are consuming and at which rate. Then the client with the highest net consumption rate is chosen as the synchronization source to which all other clients synchronize their playback. Locally, each client uses the difference between the chosen and the local timepoints and sample numbers to adjust its playback speed.

The local adjustment at a client is performed as follows. The number of samples needed for the correction is added to a prior baseline value, which is thus adjusted to the direction of the error. Given the resulting adjustment value, the audio playback module changes its playback speed by adjusting its sampling rate, either by zero-padding or by dropping samples.

3. EXPERIMENTS IN WIRED AND WIRELESS NETWORKS

We evaluated the performance of the proposed system in multi-channel multi-recipient audio streaming using two different networks, a wired Gigabit Ethernet LAN and a wireless IEEE 802.11g LAN. In both cases the network had as the server and as the clients five PC computers with dual-core 2.4 GHz, 2GB of memory, integrated audio and OS Linux Fedora 7.

Synchronization error was quantified as the time difference in the playback of a pair of clients, measured with TiePie HandyScope USB oscilloscope directly from the analog audio outputs of the clients. Rising slopes of the square pulse waves were compared to obtain the synchronization time difference between client devices.

3.1 Performance in Ethernet network

The server and the clients were connected by a Gigabit Ethernet switch. The maximum throughput was measured to be 941 Mbps with 0.18 ms RTT for 1500-byte packets. The server generated a ~13 Mbps (aka 5.1 audio) multichannel bitstream.

Continuous measurement of PTPd synchronization error over a 30-minute period showed that the clock error between two PTP clients and PTP master was 2 µs or less. This is a fraction of the clock error that can be typically expected with NTP synchronization.

Figure 2 shows the histogram of the differences in playback times between four clients over a 6-minute period. Fitting a Gaussian to the histogram gives mean of 0.58 µs and standard deviation of 19.8 µs. This indicates that in an uncongested high-speed Ethernet LAN the PTPd and swarm synchronization are able to meet even the most rigorous performance requirements for tightly coupled playback of multi-channel audio.

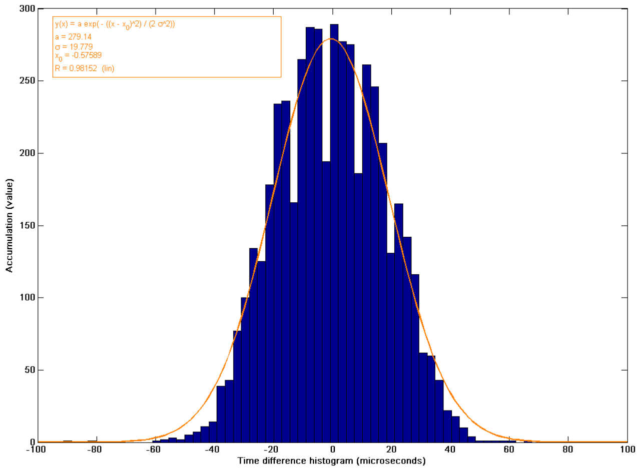

Figure 2. Histogram of the playback time differences in the Ethernet network.

3.2 Performance in IEEE 802.11g WLAN

The clients were connected to the Gigabit switch via an IEEE 802.11g access point. The maximum throughput was measured to be 29.2 Mbps with 2 ms RTT for 1500-byte packets. The server generated a ~1.4 Mbps (aka CD audio) stereo bitstream.

PTP clock synchronization accuracy was measured to be 2 ms with systematic peak error patterns, due to the PTPd clock synchronization suffering from the packet loss and retransmissions in the wireless link.

Figure 3 shows the histogram of the differences in playback times over a 6-minute period. The mean of the synchronization error is 201.9 µs and standard deviation is 60.6 µs. The synchronization suffers from a systematic offset, again reflecting the unsuitability of PTPd for wireless links.

3.3 Discussion

Our proposed system was able to synchronize client playback well below 1 ms accuracy in both wireless and wired network. In wired scenario, the playback synchronization accuracy is suitable even for the most rigorous audio playback requirements, namely tightly coupled audio delivery. In wireless scenario, however, PTPd clock synchronization suffered from the typical WLAN network characteristics rendering the accuracy not suitable for high-fidelity simultaneous playback.

Precision Time Protocol was found significantly more suitable than the Network Time Protocol [12] for applications where multimedia synchronization of very high quality is expected. NTP's typical synchronization accuracy is in a class of milliseconds, not microseconds.

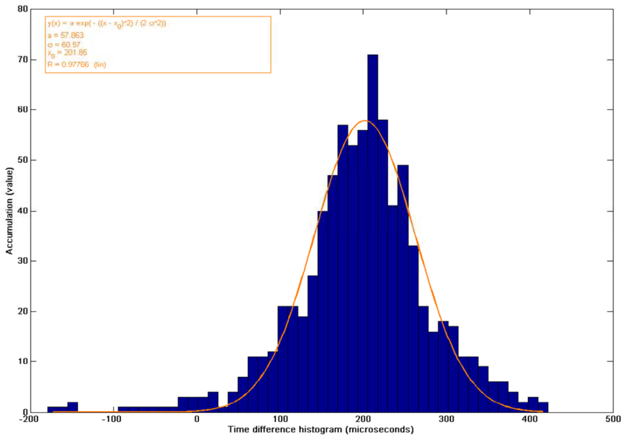

Figure 3. Histogram of the differences in playback times in the WLAN network.

3.1 RTP playout latency measurements

We also evaluated the performance of our streamlined RTP implementation. A data stream was transmitted from a sender to a receiver and PTPd timestamps were recorded before the transmission of a packet and after the reception of the packet. The difference between the timestamps corresponds to the end-to-end delay, which was measured to be 2 ms. Empirical tests showed that the minimum buffer size for successful transmission in Ethernet network was three packets. Transmission delay for three 1500-byte packets is 36 μs, thus most of the end-to-end latency is contributed by nodal delay.

4. CONCLUSIONS

We presented 'swarm synchronization' mechanism for synchronizing the playback of multi-channel audio by multiple playback clients. The proposed mechanism uses the PTP (Precision Time Protocol) protocol for synchronizing the clocks of the clients in the swarm. The clients exchange information on each other's net consumption rates and adjust their sampling rates according to the 'slowest' client. A streamlined version of the RTP protocol is employed to minimize playout delay.

The proposed system was empirically evaluated in wired Ethernet LAN and in wireless IEEE 802.11g LAN. The results showed that in the high-speed Ethernet the swarm synchronization is able to meet even the most rigorous performance requirements for tightly coupled playback of multi-channel audio. However, in the WLAN the synchronization performance was clearly worse exhibiting systematic errors, indicating that the PTP based synchronization is not suitable for 802.11g technology in applications with rigorous quality requirements. If more accurate clock synchronization existed for WLAN, our approach would naturally result in better accuracy as well.

ACKNOWLEDGEMENTS

We would like to thank Genelec Oy for financial support.

5. REFERENCES

[1] R. Steinmetz, "Human perception of jitter and media synchronization," IEEE Journal on Selected Areas in Communication, vol. 14, pp. 61–72, 1996.

[2] G. Blakowski and R. Steinmetz, "A media synchronization survey: reference model, specification, and case studies," IEEE Journal on Selected Areas in Communications, vol. 14, no. 1, pp. 5-35, 1996.

[3] PulseAudio – Trac,. URL: http://www.pulseaudio.org/, retrieved on 15.1.2009

[4] SqueezeCenter: Our powerful and free Open Source software, http://www.slimdevices.com/pi_features.html, retrieved on 15.1.2009

[5] Axia IP-Audio Driver - User's Guide. URL: http://www.axiaau dio.com/manuals/files/axia_ipaudio_driver_v2.3.pdf, retrieved on 15.1.2009

[6] IEEE Standards Committee, "Precision clock synchronization protocol for networked measurement and control systems," IEEE Std. 1588, 2004.

[7] K. Correll, N. Barendt, and M. Branicky, "Design Considerations for Software Only Implementations of the IEEE 1588 Precision Time Protocol," Conference on IEEE 1588, 2005.

[8] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," RFC 3550, 2003.

[9] J. Rosenberg and H. Schulzrinne, "An RTP Payload Format for Generic Forward Error Correction," RFC 2733, 1999.

[10] H. Melvin and P. Corcoran, "Playback synchronization techniques for networked home appliances," in Proc. of IEEE International Conference on Consumer Electronics, pp. 1-2, 2007.

[11] C.P. Young, B.R. Chang, Y.Y. Chen, and W.Z. Zhou, "The implementation of a wired/wireless multimedia playback system," in Proc. of IEEE International Conference on Innovative Computing, Information and Control, pp. 62-62, 2007.

[12] D.L. Mills, "Internet time synchronization: The network time protocol," IEEE Transactions on Communications, vol. 39(10), pp. 1482-1493, 1991.