PTP Synchronized Isosynchronous Multi-Channel Audio-Streaming over Gigabit-Ethernet based on FPGAs

Christopher Willuweit, Jan Wellmann, Stefan Goetze

Fraunhofer Institute for Digital Media Technology IDMT, Oldenburg, Email: christopher.willuweit@idmt.fraunhofer.de

Abstract

Advanced multi-channel audio-processing, e.g. localization algorithms, require highly coherent audio streams. Isosynchronous audio-streaming is usually achieved using proprietary media, hardware and protocols like ADAT or MADI. The main objective of this contribution is to stream audio data from spatially distributed modules to a host using standardized Ethernet networks isosynchronously. The synchronization is achieved by using a network protocol for industrial clock distribution, i.e. the precision time protocol (PTP). To ensure maximal clock accuracy, the system is implemented on FPGA using finite-state-machines instead of a programmable microprocessor.

Motivation

Using multiple microphone-arrays to enhance localization accuracy requires a clock skew well below the audio-sampling interval. This low clock skew is usually achieved by physically distributing a clock signal to every node. This often requires separate cables (e.g. ADAT-Sync) which is not suitable for usage e.g. in large buildings or existent installations. PTP Standard IEEE1588 speakes that - if specialized hardware is used directly above the physical layer - clock skews around 10ns can be achieved. This specialized hardware can be implemented in FPGA achieving lowest possible delay from message reception to response transmission. Although several microprocessor models for FPGA exist, the system described in this contribution is solely based on finite-state-machines to provide the best possible deterministic behavior of the system.

System structure

The network topology of the system is shown in Figure 1. The system is isolated from the (usually) very complex main network. This isolating element is a standard ethernet-switch thus no PTP-hardware is needed. PTP-synchronization only takes place in the systems subnet (PTP-domain). Thus, the audio-receiving host machine doesn't need any PTP-implementation (can be a low cost PC) and the overall network-topology has no influence on synchronization accuracy. As standard ethernet switches cannot distinguish between high priority packets (PTP-synchronization in this case) and packets with lower priority (like audio-data or other traffic), the synchronization messages undergo a non-deterministic delay within the switch (delay depends on ongoing network traffic). The PTP-standard (IEEE1588) suggests the use of specialized switches (with internal system clock), when synchronization over large networks is needed.

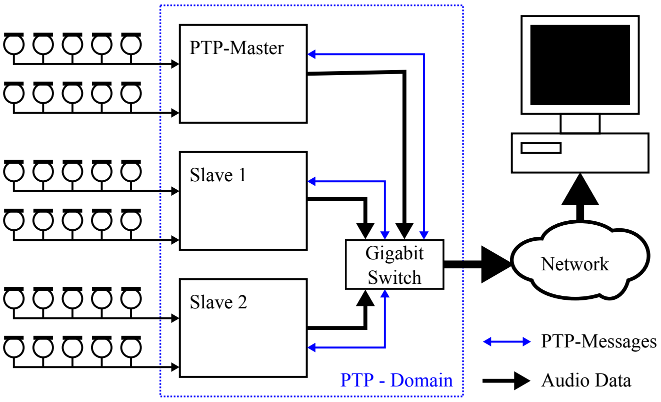

Figure 1: Network topology.

In this system, another approach is used. Due to the isolation from foreign network traffic, the delay of synchronization frames is only dependent on the ongoing audio-streaming traffic. To allow minimum and (more importantly) deterministic delay of synchronization frames, the modules will stop audio-streaming for few microseconds while the synchronization process takes place. So the internal buffers of the switch can empty and a deterministic synchronization is achieved. After the synchronization is done, audio-streaming will continue. This so-called guard-interval is shown in Figure 2.

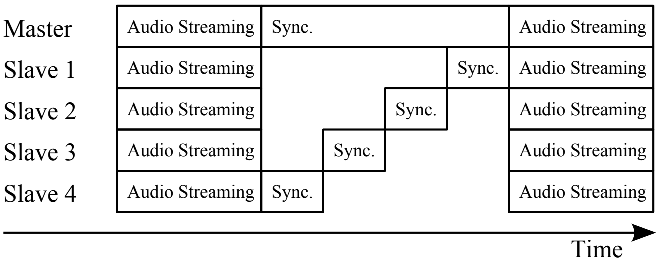

Figure 2: Guard-interval for deterministic delay of synchronization frames.

Module structure

Each module consists of two circuit-boards. The FPGA-Board is carrying the FPGA, the Ethernet-transceiver and peripheral hardware. The data-capturing board consists of amplifiers, analog-to-digital converters, analog

This work was partially funded by the EU projects Eco Shopping (project no. 609180) and Sounds for Energy Control of Buildings (S4ECoB, project no. 284628)

power supply, etc.

These two systems including the FPGA-internal structure is shown in Figure 3. The FPGA-based hardware consists of the Tx- and Rx-Stack, the system-clock, a high-speed SRAM-data-buffer and the I^2S-interface. This interface can be modified to support other digital-audio-formats like SPDIF, PCM, etc. and can transmit auxilliary data like temperature or control data.

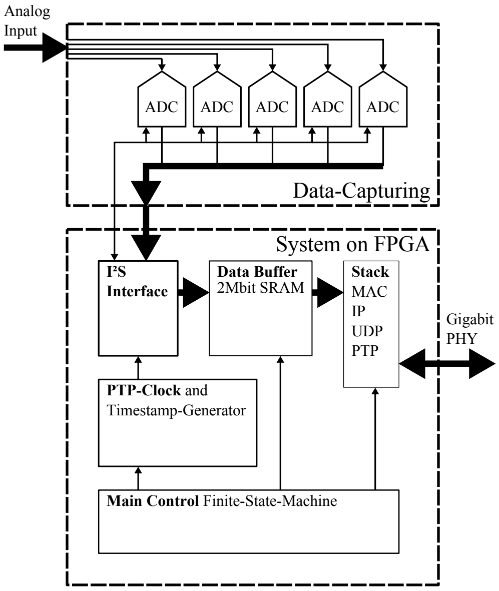

Figure 3: Module structure.

Synchronization

PTP uses three messages to determine the message delay and the clock deviation between master and slave. The synchronization is divided in two parts:

First, the delay-time for a message travelling through cables and switch is determined. Once the delay is known, the master transmits "SYNC" frames to each slave, containing the exact transmission time (measured at the master clock)

Now the slave is able to determine its clock deviation by adding the signal delay to the received master time. A sync message followed by a delay measurement is shown in Figure 4. The Sync-Messages are usually sent every 2 seconds, whereas the delay measurement is done every 30 seconds to 5 minutes. It is very important that the modules are capturing the reception and transmission times exactly. Instead of travelling through a large software protocol stack, the message can be captured directly only when using dedicated hardware.

Streaming

The analog audio signals are amplified and converted in serial digital data on the data-caputuring board. Serial audio data is directly fed into the FPGA, where it is parallelized and timestamped using the local system clock. The converted and timestamped audio-data then is stored in an internal SRAM-buffer. When enougth data is captured to fill a package, header data is attached and the package is transmitted to the ethernet transceiver. The well known user-datagram-protocol (UDP) is used for transport. Thus, the software running on the host-PC only has to open a certain UDP-Port to receive the data. Data-streams from each module have to be sorted according to their timestamps. After this is done, the audio-data is present as a single data stream consisting of synchronized audio-channels from every module. This stream is stored or processed by another software.

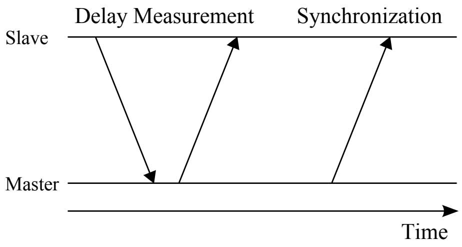

Figure 4: Synchronization process.

Expected performance

The minimum skew as well as the latency of the streaming system will be strongly dependent on the used network topology. Latency increases with every network element used between streaming system and host. For best performance the system should be isolated from other network traffic by a separate Ethernet-Switch (cf. Figure 1). The major goal of this contribution is to reach a clock skew of **below 500 ns**. The second goal is to reach a maximum latency of **below 10 μs** from the modules to a host directly connected to the isolating switch.

Outlook

In the long term the system should also be able to stream from host to module to enable isosynchronous audio output from spatially distributed modules. The used design will consume less than 20% of total FPGA-fabric on the used chip (Xilinx Spartan 6 - XC6SLX45). This enables the possibility to integrate additional fast signal processing on the modules.

Basically, the FPGA implementation is practical for all functionalities that rely on parallelization (e.g. a filter-structure for every audio-channel). For transmitting aditional data, an additional I^2C-Interface is provided in the design to support many microelectronic sensors (temperature, humidity, camera, etc.).

The usage of PTP allows the system to be integrated in already existent Audio-Video-Brigding networks, which use the same standard for synchronization. In AVB, the already mentioned PTP-Switches are used, so that the guard-interval shown in Figure 2 is not necessary anymore.

Using the proposed approach, isosynchronous audio can be transmitted on existing IP-installations with less than 500ns skew between different nodes.