IPTV Inter-Destination Synchronization:  
A Network-Based Approach

H.M. Stokking, M.O. van Deventer, O.A. Niamut, F.A. Walraven, R.N. Mekuria  
TNO  
Delft, The Netherlands  
{hans.stokking, oscar.vandeventer, omar.niamut, fabian.walraven, rufael.mekuria}@tno.nl

Abstract— This paper introduces a novel network-based approach to inter-destination media synchronization. The approach meets the need for synchronization in advanced TV concepts like social TV and offers high scalability, unlike conventional end-point based approaches. The solution for inter-destination media synchronization has been standardized in upcoming ETSI protocol standards for IPTV. This paper explains the inner workings of this standardized solution.

Inter-destination synchronization, advanced TV services, IMS

I. INTRODUCTION: NEED FOR SYNCHRONIZED PLAY-OUT AND END-POINT-BASED SOLUTIONS

In the television world, IPTV has quickly gained ground on more traditional broadcasting means. On the one hand, IPTV enables Internet Service Providers, such as DSL, Fiber or LTE providers, to deliver digital TV and video services to their customers. This stimulates competition; we now see a variety of providers offering TV services in a certain geographic region. On the other hand, IPTV offers capabilities beyond the traditional linear broadcast. The IP connection is used for streaming content to end users, but can also be used for communication between end-user and service provider or between end-users.

There are many types of interaction that have evolved around content. One term for such services is combinational services. Van Deventer et al [1] present four use cases in this area, of which the social TV use case is perhaps the most prominent one, see e.g. [2,3].

One of the problems in the social TV use case, is the synchronized play-out of content at different geographical locations. This problem is known in literature as inter-destination media synchronization (IDMS), or group synchronization [4].

Fig. 1 shows the need for IDMS. If you are watching a soccer match together with a friend at two separate locations, delay differences between content play-out can spoil the experience. Other reasons for synchronization include voting shows. If users pay to vote on e.g. a candidate in a pop star competition, and you indicate in the show that ‘the lines are closed’, then any vote before that announcement must be valid, and any vote after that announcement must be discarded. Also, television programs may refer to real world time, e.g. the 8 o’clock news starts exactly at 8 o’clock, or the new years eve clock should hit midnight at exactly midnight. Note that such exact playout times at user’s homes will require broadcasting slightly ahead of time.

As [1] shows, there can be significant delay differences between various means of broadcasting. These differences can e.g. occur between fixed and mobile, or between SD and HD signals. To validate that this problem is actually a real-world problem, we have performed measurements by recording a side-by-side view of different commercially available broadcast channels (Fig. 2).

Figure 1: Synchronization problem when “Watching Apart Together”

Figure 2: Frames 114, 141 and 197 (25 fps) comparing analogue signal with DVB-T signal

We have used a laptop with a DVB-T receiver as a reference signal, to which all others signals are compared. By writing down the precise frame numbers on which a scene change occurs for both broadcast signals, we have determined the delay differences between two signals. We performed a total of 293 measurements in 9 different situations, comparing various broadcast technologies at different geographic locations. Fig. 3 shows the results of these measurements, clearly showing delay differences of several seconds.
Figure 3: Comparison of the delays of various broadcasting technologies, using DVB-T on a laptop as a reference signal

Boronat et al [4] provide an extensive overview of existing approaches for IDMS. Although this article shows that there are numerous approaches, all approaches approximate to the same solution: A client on an end-user device receiving a multimedia stream reports on arrival time or presentation time of media packets in that stream, and a server is used to calculate delay differences between various clients. Clients receive instructions on delay, and buffer the media stream to achieve synchronization.

Problems related to end-point based solutions are:

- The approach does not scale very well, since all clients will have to report status and receive instructions to a central server. This can cause similar problems as with the presence framework, such as possible network flooding, see e.g. [5].
- The approach requires all end-user devices to support the synchronization solution, and current (legacy) devices do not support this.
- The IPTV provider cannot give guarantees on the achieved play-out times, if the end-user uses a media client not under complete control of the provider.
- Channel changing delays can increase when end-user devices are buffering media streams to achieve inter-destination synchronization.

II. PROPOSAL: A NETWORK-BASED APPROACH

This article proposes a novel approach with a different solution direction. Instead of reporting on media stream arrival time and buffering at the receiving client, such functionality is incorporated in the network. This network functionality can be implemented in the edge node, e.g. a DSLAM or CMTS, or may even be implemented higher up in the network hierarchy. The synchronization point in the network is selected such that further downstream delays are considered acceptable for the combinational service. This approach solves the found problems related to end-point based solutions to a large extend.

- It scales very well. Because many UEs can be synchronized by a single edge node, the number of synchronization messages is limited. This will also limit the needed capacity at a synchronization server, at the cost of functionality on the edge nodes. Note that the synchronization buffer for a media stream is also shared by many UEs.
- The approach does not require UEs to support the synchronization solution, so legacy devices can also be used. For IPTV providers that provide their users with a (free) set top box (STB), this saves costs for those STBs, at the cost of functionality in the network.
- Since the edge node is under complete control of the IPTV provider, it can guarantee the stream synchronization for streams sent to the UEs. When implemented at the edge of the network, little or no delay differences will occur between SC and UE. Although jitter buffer settings between UEs may vary, this will not cause significant delay differences between UEs.
- Because buffering is done in the network, channel changing delays will not increase due to inter-destination synchronization. This assumes that all broadcast channels are being buffered for a short period of time at the edge nodes. If various UEs switch to a new channel as part of a social TV experience, the new channel should also be delivered synchronously. This may mean that the new channel is delayed for certain UEs compared to other UEs.

Figure 4: Network-based media synchronization approach

Fig. 4 shows a typical situation for an IPTV television broadcast. There is a source for the media stream. The media stream goes through the (IP) network towards the User Equipment (UE). For television broadcasts, typically many users are watching the same channel, so many UEs receive the same media stream. The synchronization client (SC) consists of a synchronization buffer and control functionality which can be implemented on the edge nodes. Further, a media synchronization server (MSAS) is used to manage the synchronization between SCs. This approach has several advantages:

- The approach does not scale very well, since all clients will have to report status and receive instructions to a central server. This can cause similar problems as with the presence framework, such as possible network flooding, see e.g. [5].
- The approach requires all end-user devices to support the synchronization solution, and current (legacy) devices do not support this.
- The IPTV provider cannot give guarantees on the achieved play-out times, if the end-user uses a media client not under complete control of the provider.
- Channel changing delays can increase when end-user devices are buffering media streams to achieve inter-destination synchronization.

II. PROPOSAL: A NETWORK-BASED APPROACH

This article proposes a novel approach with a different solution direction. Instead of reporting on media stream arrival time and buffering at the receiving client, such functionality is incorporated in the network. This network functionality can be implemented in the edge node, e.g. a DSLAM or CMTS, or may even be implemented higher up in the network hierarchy. The synchronization point in the network is selected such that further downstream delays are considered acceptable for the combinational service. This approach solves the found problems related to end-point based solutions to a large extend.

It scales very well. Because many UEs can be synchronized by a single edge node, the number of synchronization messages is limited. This will also limit the needed capacity at a synchronization server, at the cost of functionality on the edge nodes. Note that the synchronization buffer for a media stream is also shared by many UEs.

The approach does not require UEs to support the synchronization solution, so legacy devices can also be used. For IPTV providers that provide their users with a (free) set top box (STB), this saves costs for those STBs, at the cost of functionality in the network.

Since the edge node is under complete control of the IPTV provider, it can guarantee the stream synchronization for streams sent to the UEs. When implemented at the edge of the network, little or no delay differences will occur between SC and UE. Although jitter buffer settings between UEs may vary, this will not cause significant delay differences between UEs.

Because buffering is done in the network, channel changing delays will not increase due to inter-destination synchronization. This assumes that all broadcast channels are being buffered for a short period of time at the edge nodes. If various UEs switch to a new channel as part of a social TV experience, the new channel should also be delivered synchronously. This may mean that the new channel is delayed for certain UEs compared to other UEs.
Disadvantages of a network-based approach also exist. The proposed solution will not work for over-the-top IPTV services, since network control is required. Control could of course be offered by a network provider, but experience has shown that network providers are not eager to open up their networks in this manner. Also, any delay differences introduced behind the synchronization point, are not yet taken into account and require further study. When users forward a media stream to e.g. a mobile device, this creates a situation for which this solution has to be adapted.

The proposed IDMS solution shares some properties of other recent application-layer service capabilities for IPTV, such as solutions for retransmission or forward error correction. For example, the DVB retransmission solution [6] can be implemented as an add-on in an existing IPTV deployment. While this add-on does require changes on the client-side functionality, it leaves the streaming solution intact. IPTV providers may decide on the introduction of an IDMS solution after roll-out of their IPTV platforms.

III. ETSI STANDARD FOR INTER-DESTINATION SYNCHRONIZATION INCLUDES NETWORK-BASED APPROACH

Traditionally, operators value standardized solutions, and standardization is important for inter-destination synchronization. It reduces development costs, and allows for multi-vendor approaches and interconnection; cross-domain solutions are pivotal for services like social TV to succeed.

Within these specifications, ETSI TISPAN has specified two new NGN functional entities, see Fig. 3:

- A Media Synchronization Application Server, MSAS. This MSAS server collects synchronization status information from synchronization clients, calculates delay settings instructions and it sends these instructions to the clients.
- Synchronization Clients, SCs. These clients report on media arrival or presentation time to the MSAS, and they delay the play-out based on instructions received from the MSAS. A requirement for an SC is that it is clock-synchronized using for example NTP or SNTP.

A new reference point for session and media information is defined between MSAS and SC, called the ‘Sync reference point’ which reuses existing reference points. The session related part of the sync reference point is exchanged over the Gm and ISC reference points (SIP) and possibly Xc reference point (RTCP). Further, the Service Control Function (SCF) plays an important role in synchronization. The SCF is an IMS application server containing IPTV functions.

This paper will first the discuss the session related parts of the ETSI standardized UE-based specification and of the network-based approach. Then, it will discuss the media related part, which is the same for both the UE-based and network-based approach.

A. Session related part of the UE-based interdestination synchronization

Inter-destination synchronization requires the UEs to be synced to set up a synchronization session between them. The synchronization session is setup between UEs containing an SC and the MSAS. The session setup piggy-backs on the existing media session setup. For broadcast, these sessions are setup using SIP and SDP [16], using the Gm and ISC reference points. A session is created between UE and SCF. After session setup the UE can use IGMP to actually retrieve the media stream. For video-on-demand (VOD) these sessions are setup using a combination of SIP and RTSP, also using SDP. SIP is used to setup a session to a specific Media Function (MF). Then, RTSP is used to select the specific content, using reference point Xc. The SDP description for VOD can be part of the SIP session setup or the RTSP session setup.

The authors of this paper have contributed the proposed IDMS solution in the third release of the ETSI TISPAN IPTV specifications [7-11]. Specifically, it focuses on the IMS-based specification with session control based on SIP [12]. ETSI also specified an NGN integrated IPTV subsystem with session control based on HTTP [13]. This integrated subsystem also contains the inter-destination synchronization functionality and works quite similar to the IMS-based specifications. For an overview of the IMS-based IPTV architecture, based on the second release of the specifications, see [14]. The third release of these specifications, which contain the synchronization solution described in this paper, are expected to be finalized and published in September 2010.

Figure 5 The IMS-based IPTV architecture

Figure 6 SIP signaling for the session set-up of inter-destination media synchronization of a BC service
The synchronization session information is contained in the SDP media description. This SDP contains the following items:

- The address of the MSAS to be used for the synchronization session. This address is allocated by the SCF and will usually be the same for all UEs in a synchronization group. See also [17].
- A SyncGroupPId, which specifies the synchronization group. The SyncGroupPId can be allocated by the SCF or can be indicated by the UE. See also [18].
- In case of VOD, the SSRC of the media stream. This SSRC can be used to correlate various RTCP messages. See also [19].

At the end of a synchronization session, the session can be ended in two ways. The media session itself can be ended, thereby also ending the synchronization session. Alternatively, a SIP re-INVITE can be sent, containing an exact duplicate of the session description but omitting the synchronization parameters.

B. Network-based interdestination synchronization

The IDMS solution that is being standardized by ETSI can be used in different settings. By specifying the functional entities and the protocol used between them, the IPTV operator can independently choose the best locations for these entities. The ETSI specifications offer guidance for this selection, and one of the options is to put the synchronization client in the transport network. This is different in a number of ways compared to synchronization on UEs as described above.

The first difference is its intended purpose. The intended purpose for synchronization in the network is for broadcast television, where many users watch the same media stream. In theory a network-based approach can also work for VOD. But, in that case, the network would have to synchronize individual streams to UEs. This will not give any advantage in scalability, and the advantage of smaller channel changing delays is not applicable to VOD.

Also, network-based synchronization requires the configuration of SCs beforehand, instead of using SIP for synchronization session setup. The network elements, i.e. edge nodes, are not involved in session setup. This means they cannot setup synchronization sessions in the same way as SCs on an UE can. Also, the media streams of a television broadcast are used in many media session for different UEs.

To enable network-based interdestination synchronization, the network elements containing the SCs need to be configured with:

- The SyncGroupPId’s, one for each television channel to be synchronized. A television channel in the network is defined by e.g. the multicast address used.
- The MSAS addresses, which should be the same for each television channel to be synchronized.

This can be compared with the configuration for retransmission solutions, where the retransmission server has to be configured in a retransmission client.

C. Media-related part of the sync reference point, used in the network-based and UE-based approach

After configuration of network elements or synchronization session setup for UEs, synchronization messages can be exchanged between SCs and their MSAS. SCs send synchronization status information to the MSAS, indicating the arrival time and/or presentation of media packets to the MSAS. The MSAS sends synchronization settings instructions to the SCs, indicating for each SC individual delay instructions. RTCP is used for this communicating of status information and delay information, as proposed by Boronat et al [20]. Since RTCP is used, this requires the use of RTP as transport protocol for the media. A new RTCP XR block type has been specified for the purpose of synchronization, as shown in Figure 7.

This new block type contains the default RTCP XR headers, followed by the new block type. It contains the following new fields:

- **Synchronization Packet Sender Type (SPST):** This field identifies the role of the packet sender for this specific XR. Because the same block type is used for sending status information and for sending settings instructions, the packet sender determines the interpretation of the values.
- **Packet Presented NTP timestamp flag (P):** This flag indicated if packet presentation times are used.
- **Payload Type (PT):** This field identifies the format of the media payload. The media payload is associated with an RTP timestamp clock rate. This clock rate provides the time base for the RTP timestamp counter.
- **Media Stream Correlation Identifier:** This field contains the value of the SyncGroupPId, indicating to which synchronization group the report belongs.
- **SSRC:** This is set to the SSRC value of the associated RTP stream.
- **Packet Received NTP timestamp**: This is the timestamp of the wall clock time the XR relates to.

- **Packet Received RTP timestamp**: This timestamp has the value of the RTP timestamp to which the XR relates.

- **Packet Presented NTP timestamp**: This timestamp reflects the actual presentation time of the associated RTP packet. This field is optional.

For synchronization status information, the use of this block type is straightforward. For synchronization settings instructions, an XR report should be interpreted as if it is a status information report of the most lagged SC. Inter-destination synchronization requires all SCs to match to SC that is most behind in media reception or presentation.

**D. Other specified synchronization aspects**

Two other aspects of inter-destination synchronization are also specified by ETSI.

1) **Modification or re-origination of streams**

   First, inter-destination synchronization may occur between SCs that are receiving different RTP streams. This may e.g. be the case when one IPTV implementation serves both HD streams and SD streams. The IPTV implementation may receive an HD stream and transcoded this for serving SD streams at lower bandwidth. Different media streams will use different RTP streams for transport. The synchronization solution described in this paper will not work in that case, because it uses the RTP timestamps for synchronization purposes.

   To solve this issue, an additional Synchronization Client (SC’) is developed. This client is used for entities that modify or re-originate streams, and thereby introduce new RTP timestamps for example when transcoding a stream. An SC’ sends synchronization correlation information to the MSAS using the sync’ reference point. This information consists of arrival time information of an incoming stream combined with timing information of the outgoing stream. For this purpose, two XR blocks are stacked. A different SPST is used for the incoming and outgoing stream.

   2) **Partial SC functionality in the network**

   Earlier in this paper, it was mentioned that delay differences introduced behind the synchronization point are not yet taken into account. The ETSI specifications do offer a solution for this. The network can contain partial SC functionality, supporting SC functionality in the UE. This requires that the SC in the network is on the network path between the SC in the UE and the MSAS. In that case, the SC in the UE can give arrival time information on RTP packets and the SC in the network can perform buffering, i.e. carry out delay settings instructions.

   This does require that the media stream from the SC in the network is a unicast stream towards the UE. In an IPTV environment, television broadcasts are usually sent out using IP multicast. But, from the edge node of the network individual streams are sent to individual UEs, even though it is still using IP multicast.

**IV. FUTURE WORK**

Future work on IDMS includes interconnection of synchronization solutions, and in determining delay differences in the last mile of the network. Because social TV services can span multiple (IP)TV Service Providers, these service providers may have to work together to jointly offer combinatorial services. Research should focus on network integrity and on interoperability between various synchronization domains. Furthermore, research on delay differences between the synchronization point and the end user devices needs to determine whether delay differences will cause problems at the service level, and if so, what the possible solutions are to resolve these problems.

**REFERENCES**


[8] draft ETSI RTS 182 027, Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IPTV Architecture; IPTV functions supported by the IMS subsystem, V3.3.4 (2009-10).

[9] draft ETSI RTS 182 028, Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IPTV Architecture; IPTV functions supported by the IMS subsystem, V3.3.4 (2009-10).

[10] Draft ETSI TS 183 063, Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IMS-based IPTV stage 3 specification, V3.3.0 (2010-02).


[18] IETF RFC 3611, RTP Control Protocol Extended Reports (RTCP XR), November 2003

[19] IETF RFC 5576, Source-Specific Media Attributes in the Session Description Protocol (SDP), June 2009