Towards Streamed Services for Co-located Collaborative Groups

Ben Falchuk
Applied Communication Sciences
Piscataway, NJ USA
bfalchuk@appcomsci.com

Tomasz Żernicki
Telcordia Poland Sp. z o.o
Poznań, Poland
tzernick@telcordia.com

Michał Koziuk
Telcordia Poland Sp. z o.o
Poznań, Poland
mkoziuk@telcordia.com

***Abstract* — From both technical and social viewpoints there is great value in services that require devices (and therefore people) to be co-located. The very act of co-location brings with it entirely new dynamics and collaboration; furthermore, devices in coalition can render services and provide experiences that a single device might not be able to. In this paper we describe the motivation, design, and uses of high experience coalition-based services and outline how such services could be architecture on both the server and client sides. Extensive use of video transcoding and region-of-interest techniques - to segment and stream only portions of video frames - makes delivering experiences like "social cinema" across several co-located devices feasible. On the client side smartphone-based interactive coalition setup and control is very viable. In this paper we explore and document our functional architecture and take a closer look at the similarities and differences between OnLive and our proposed architecture and services. The rising prominence of hi-resolution LED devices together with services such as OnLive make coalition services technically viable, desirable, and worthy of both industrial and academic investigation alike.**

Keywords – multimedia, standards, video streaming, mobility, services, co-location, collaboration

I. INTRODUCTION

Today's ubiquitous mobile networks and capable mobile devices mean that work and entertainment are almost always at hand. Furthermore, multimedia information is not only being stored in personal devices such as iPhones, tablets, and notebooks, but increasingly often in the cloud (e.g., Amazon, Apple, Spotify). While storage, virtualization, and Internet-based IT services are the underpinnings of such cloud offerings, one aspect of our industrial work focuses on the flip-side of server virtualization: user co-location. In other words, while virtualization implies that a system resource exists in any number of physical locations, users exist in only a single physical place and human-to-human interaction in close proximity remains an essential part of our lives. While it is true messaging and video-conferencing tools make co-location irrelevant for many daily tasks, humans nonetheless crave and thrive upon real interactions with each other in the real places we inhabit - the playgrounds, pubs, social gatherings, and with family. Our work focuses on the technologies that support novel and meaningful co-located service interaction.

Prevailing multimedia services – particularly streaming video, audio, and games - are largely designed to be delivered to, and rendered upon, a single device. In principle, however, collections of devices can contribute multiple distinct resources in cooperation to enrich the media experience if only mechanisms existed on the serving-side to intelligently transmit the appropriate information and on the client-side to coordinate resources. Our work establishes methods for collections of devices to pool resources for synthesizing an enriched, coordinated media experience across a collection of co-located devices. Indeed, collaborative head-to-head play is already an intrinsic part of the gaming and entertainment landscape, sometimes requiring physical co-location, sometimes not, as seen in products such as Sony PSP Mobile, NGage, NintendoDS, and the (now retired) Microsoft Zune. Massively multiplayer online role-playing games (MMORPG), on the other hand, are client-server based, do not require co-location[1], but *do* require social teamwork and strategy amongst players.

The goal of this paper is to describe the new technologies we are working on in the context of mobile collaboration. In this context the technologies enable new social uses of mobile devices that involve multimedia and co-location. Our quintessential use-case is one in which *a coalition of mobile users come together, choose a multimedia service, delegate member roles, and experience the service in an interactive inclusive way not possible in isolation*. Figure 1 depicts the notion of user-location.

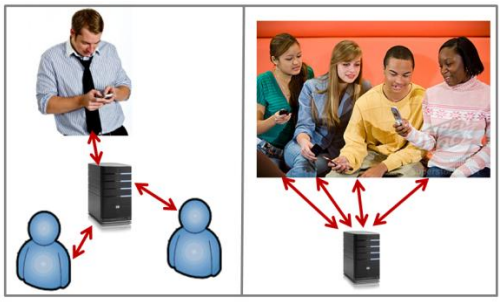

Figure 1. Distributed play (left) versus co-location (right); the latter adds social, collaborative, and commercial value.

[1] Gaming parties, in which users co-locate to play in a space with multiple computers (e.g., dorm room, lab), are also popular

COLLABORATECOM 2012, October 14-17, Pittsburgh, United States

DOI 10.4108/icst.collaboratecom.2012.250426

From a business perspective, co-location clearly enables the still-intriguing notion of location-based advertising and marketing. Can distributed players be targeted by advertisement campaigns? Of course, but co-location adds potentially important new marketing and technical dimensions:

- Increased local spending power (as a group) at local venues
- Increased influence (crowd versus one)
- Potentially increased localized bandwidth requirements
- A common desire for a particular kind of service
- Common marketable attributes (as participants are likely already friends in a strong or loose sense)

On the technical side there is much that can be gained from recent advances in video coding, compression, human-computer interaction, mobile device positioning and existing near-field communications protocols such as Bluetooth.

Key terms we use henceforth in this paper are as follows:

Coalition - a set of co-located mobile devices, put together collaboratively for the purposes of a multimedia experience (e.g., "social cinema" when referring to video and audio)

High experience service (HES) - a specialized service designed to be compatible with one or more coalition instances and delivered to the each device in the coalition over a wireless network independently. The service is based on the assumption of co-located devices. Some service content will stream (with tight inter-stream synchronization) while other content may be static.

Co-location – the act of bringing together devices (and people) such that a coalition can be formed and an HES assigned[2].

Video region of interest (VROI) – a broad notion in video processing in which regions of video frames are manipulated with special care (e.g., split-stream, object tracking, etc.). VROI plays a key role in the delivery of cinematic experiences across several video output devices (see section III).

Table 1 informally describes the notions of user experience (UX) for a gamut of traditional and co-located services. And by way of definitions we note that Nielson Norman Group, for example, describes user experience as "..encompassing all aspects of the end-user's interaction with the .. service" and that qualitative UX can be measured via usability studies of various sorts [28].

[2] Often the devices will have to be placed close to each other in space, with fairly fine placement. Such a requirement is at best awkward and at worst quite limiting but we believe that it is not a show-stopper.

There are several technical challenges that stand in the way of our vision of high experience co-located services including: synchronization of media streams across carriers, the temporary physical co-location of devices in the correct orientation, and the simplicity of initiating a complex co-located service involving several devices. Many – but not all - of these issues are at least partially addressed in standards, especially as related to digital television and multimedia streaming. Notable commercial solutions come from Apple, OnLive, Microsoft, Skype and Adobe. Open solutions are developed in various standardization bodies, such as Moving Picture Experts Group ISO/IEC JTC1/SC29/WG11 (MPEG), the Internet Streaming Media Alliance (ISMA), the MPEG Industry Forum (MPEGIF) and the 3rd Generation Partnership Project (3GPP). Moreover, network protocols used in multimedia streaming are developed by the Internet Engineering Task Force (IETF). These standards apply primarily to the receiver (decoder) side, not the encoding process, and it remains a challenge to deal with co-located services from the encoder side perspective (e.g., splitting audio and video amongst several devices)(see Section III).

TABLE I. HIGH EXPERIENCE SERVICES AND DEVICES

Services	**Description**	**Potential UX***
Streamed audio, slide-shows	Low demands on network; easily supported by most mobile devices from clamshell phones to tablets.	Low
Web, traditional streamed video, gaming	More demanding services require capable devices and networks; designed for (disparate) single-user consumption	Medium
Co-located HD video, cinema and gaming (proposed herein)	Collaborative use of heterogeneous devices can deliver new kinds of co-located experiences; older devices can participate up to their capabilities (e.g., provide a single mono audio channel); service attributes can scale to the ***combined capabilities*** of the participating devices.	High

(* UX = user experience)

II. RELATED WORK AND INDUSTRY STANDARDS

A. Related Work

The history of device co-location and segmented multimedia streaming is varied and interesting. On a fundamental level, "video walls" – typically found in public places such as busy street corners (Times Square) and sports stadiums but also in situation rooms used for joint surveillance and monitoring are a primitive form of composite service. Dedicated software such as e.g., the open-source VideoLAN software suite can provide some video wall functionality in which a video signal is striped across a series of output devices.

The Microsoft Zune media player device (circa 2006) included the novel notions of both tagging interesting music heard on the streaming radio station and of sharing songs from one device to another nearby device via Wi-Fi. The Zune never did

best the iPod's simplicity and co-located music sharing did not enter the social media lexicon.

Current mobile social applications bring gaming to mobile users while keeping the games within their circles of friends. Having friends or friends-of-friends handy to play with is seen as a key enabler to the success of this paradigm. Mobile games like CityVille and Sims currently dominate this market (over 60 million such gamers in the US alone in 2011 [1]). While CityVille-like games do not require physical co-location, other niche social applications do. Mobile dating relies on physical meet-ups. In [2], mobile users battle in a game against other people in nearby cars at stop-lights. Recently, beverage giant PepsiCo has begun mobile marketing with innovative startups, including those that employ smartphone users as "mobile workforces"[3], relying on their mobility *in the physical world*.

Collaborative multi-device Web browsing, described in [3] , is an early example of sharing Web content across participants [3]. This approach required that Web pages be annotated with XML to specify how particular HTML elements could and should be available to different devices. Device capabilities are, in turn, registered into a directory and a proxy server plays an orchestrating role. In an idealized run, a user could experience Web content split and redirected onto several devices (e.g., images onto one device, audio on another). Over the years this broad approach of matching devivce capabilities to information attributes has been gradually but mostly insufficiently exploited within emerging technology areas such as personal area networks (e.g. Bluetooth device pairing mechanism), body-area networks (e.g. through the use of Group Device Pairing protocol [4]), wireless sensor networks (e.g. using an autonomous agent-based peer to peer negotiation protocol) or Wi-Fi networks (e.g. Wi-Fi Direct, DLNA). A mobile app called Bump was noteworthy, allowing two people to bump their phones together to exchange data. In [5], mobile – but disparate - users can watch the same synchronized video stream at the same time to give that "in the living-room" feel. While watching they engage in real-time commentary about what they see. In other work the notion of "federated devices" captures (in some senses) capabilities of our system; for example, in [27] they define the term as "a set of devices which cooperatively and concurrently renders a user interface." In comparison, our work features more pragmatic analysis of today's standards. Additionally, while there are certainly some conceptual similarities to our work, Web content splitting (of the kind in [3]) focuses on Web content (pages) and does not address video. We focus on the pragmatic issues related to segmenting and transmitting audio and video to modern devices or modern networks.

Google Plus allows several users in a group to watch a YouTube video 'in synchrony' with each other (including live comments and voice chat), while Flickr photo sessions allow one user to drive a slideshow that is seen in real-time by all other members of a group. OnLive is a cloud-based service that hosts, renders and streams video games and other entertainment to Internet-connected devices. We have more to say about OnLive in subsequent sections.

Current work in video regions of interest (ROI) focuses on the automated extraction of ROIs based on the tracking of objects within the frame – furthermore, such tracking can be automatic [6] or manual [7]. Other aspects of ROI research strive to determine ROIs based on user attention using model-based algorithms over visual features of the video [8][9][10]. The models for these algorithms are often determined based on eye-tracking experiments, though recent results propose to substitute it with crowd-sourced data where multiple users mark ROIs in a given video [11]. As we will show later in this paper, ROI is important to high-experience co-located services that involve video as we anticipate that the ability to define, implement, and stream video ROI's will be essential to such services (though it should be noted that not all services need to involve video in such fashion).

B. Standards

Current audio/video standards enable many aspects of coalition services. Stepping back, recall that for the vision of collaborative multimedia coalitions to become reality we require that individual components of the multimedia service can be delivered to individual devices (in synchrony). For video, this means ROI's must be generated and streamed in appropriate resolution. For audio, it implies that one or more devices in the coalition can serve as audio output and that (optionally) audio channels can be distributed amongst these audio output devices. Furthermore, audio and video segments must retain a level of synchronicity relevant to the presentation. Standards such as MPEG DASH provide a means to this end.

Audio can indeed be divided into channels (e.g., surround sound) much like video can be partitioned into ROI's but it becomes very important to ensure stream synchronization over time between coalition devices. The first key is the appropriate audio and video codec. Audio data is compressed using a codec such as: MPEG-1 Layer 3 (MP3) [14], MPEG-2/4 Advanced Audio Coding (AAC) [15] or the recently standardized MPEG-D Unified Speech and Audio Coding (USAC) codec [16] which provides transparent sound quality for bit-rates between 16 – 24 kb/s. For the purpose of video compression, codecs such as MPEG-2 or MPEG-4 Advanced Video Coding (AVC) [17] are used. A related aspect of multimedia streaming is the usage of the appropriate bit-rate container, such as an ISO base media file format [23] as used by MPEG or 3GP defined in the ETSI 3GPP technical specification [24]. MPEG transport contains packetized information related to multimedia content, such as audio and video, and timing which is included for the purpose of stream synchronization. RTP [18], on the other hand, is one of the most prevalent solutions for media streaming on IP networks. The Real-time Transport Control Protocol (RTCP) was

[3] "PepsiCo Selects 10 Startups to Pilot Digital Marketing Projects", Oct.6. 2011, http://on.mash.to/n5zr21

introduced to support improved quality of service control and, principally speaking, achieves this via a common reference clock for inter-stream synchronization. The main disadvantage of RTP-based streaming is the reliance upon a few network ports (RTP, RTCP, RTSP) which often creates deployment challenges on firewalled or NAT-ed networks. RTP streaming also requires the server to manage separate sessions for client.Presently, HTTP-based streaming – which doesn't share RTP limitations – is seen as a popular alternative to RTP and several commercial embodiments exist such as Apple's HTTP Live Streaming [19], Microsoft's Smooth Streaming [20], and Adobe's HTTP Dynamic Streaming [21]. Recently, the MPEG group announced a new HTTP-based media streaming standard – MPEG Dynamic Adaptive Streaming (MPEG DASH) [22], an integral part of which is a so-called media presentation description (MPD) which describes available content (e.g., number of segments), codecs, and parameters. Moreover, timing information is available as a relation to other segments which facilitates inter-stream synchronization and content alternatives are expressed as URL's allowing MPEG DASH clients to decide which version of content should be fetched. This decision capability is especially useful in the context of co-located device coalitions and could be used, for example, in the choosing of video region or audio channel selection. Moreover, MPEG-DASH supports various MPEG codecs (see Section III).

III. Architecture Aspects of High Experience Services

While considering the functional properties of possible approaches we also considered the state of the art, standards, and other pragmatic approaches. Our architecture is a novel combination of new and existing techniques and we feel it is a good start towards supporting collaborative coalition services.

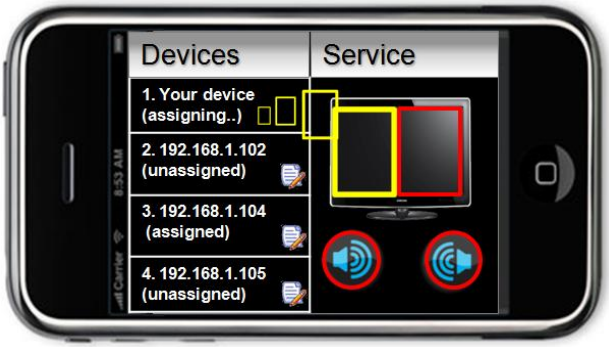

Figure 2. Coalition mobile client ("wireframe" only); in the coalition setup screen the user assigns 1 or more devices (left) to service functions (right). For example, devices may play roles as one of 2 audio channels or as video regions of interest displays.

A. End-user centric Components

A part of the coalition controller functionality on the end-user device is the mobile client whose role it is to present service discovery and service control features to participating end users. Coalition setup is a potentially complicated procedure involving the local coordination of several devices as well as with the coalition controller on the server-side. We envision that software clients for different platforms (e.g., iOS, Android) will be possible but note that this does require a messaging protocol be established between controller client and server.

The coalition setup wizard is comprised of a series of modules and screens visible to the user which step the user through the process of registering a coalition and choosing a networked service to experience. The following set of steps outlines the process of coalition setup:

1. A group of co-located mobile users/devices M,N,O, P, decides to begin a coalition for service S offered from the server
2. Each user launches a specialized mobile app (which may optionally be triggered by a message from the server) – the app registers the device IP address and attributes with a server and optionally attempts to link the devices via short range communication protocol (e.g., ad-hoc WiFi or Bluetooth)
3. A master device – say M - is selected and a graphical user interface (GUI) is displayed on it while only optionally on others that may not support high resolution graphics
4. Optionally, a visual metaphor is chosen to represent the Service functions (e.g. a living room with 2 speakers and a screen) and to help the users make assignments of device-to-function
5. M (the user) assigns service roles to (already registered) devices and the devices commit to the assignment
6. The devices communicate with the server to begin the service
7. Server optionally initiates a "test" phase in which a short segment of the service S is streamed to the devices. Users are asked if they would like to continue and each device in the coalition is passed a reference to the service media(s) that are associated to it. The service begins to stream across all devices

Figure 2 illustrates the essence of coalition setup GUI on the client application. In this application the user is already co-located with other users and a service has been selected. The task that this aspect of the GUI assists with is the assignment of users to service functions which must take place with both device capabilities and radio access network in mind. In principle each device in the coalition could be connected to a different provider radio access network. The figure shows an exemplary interface in which the registered devices in the imminent coalition are listed on the left and the interface offers a means to assign the devices to the essential parts of the chosen service such as video output and left and right audio channel output.

B. Multimedia Streaming

Figure 3 presents main components of proposed architecture for multimedia streaming of co-located devices. Server side **Coalition controller** receives the information about number of devices, hardware specification (e.g., resolution) and co-location setup. As a result it provides the encoding/transcoding parameters, such as number and size of ROIs, number of channels and bitstream specification (bitrate, number of layers in scalable coding, number of streams). If a bitstream which meets the criteria is located in Multimedia database then it is routed directly to the HTTP streaming server for transmission; otherwise, a transcoding step takes place.

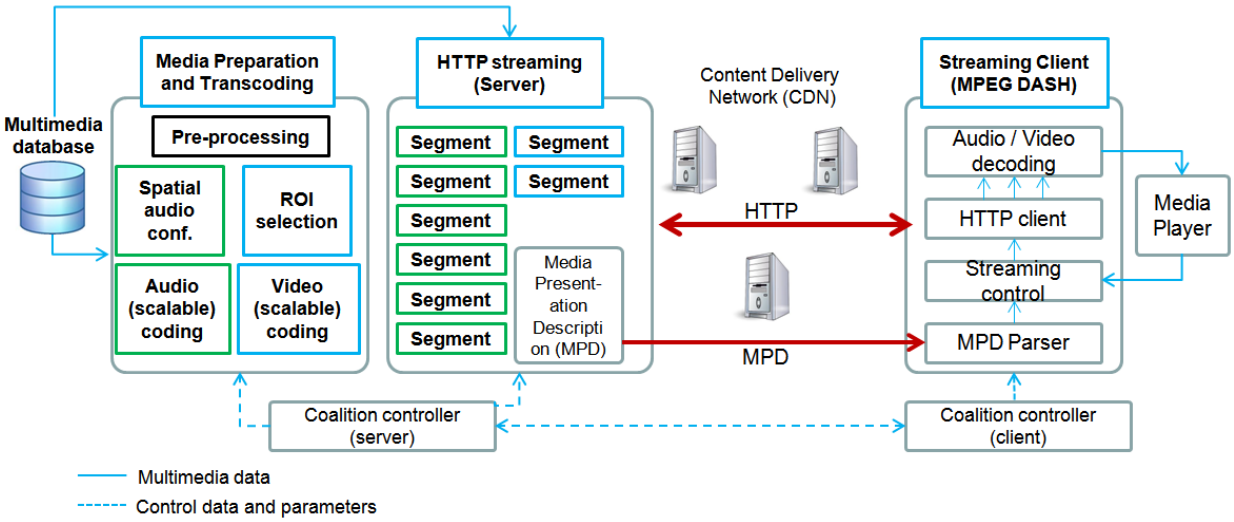

Figure 3. High-level functional architecture for coalition services

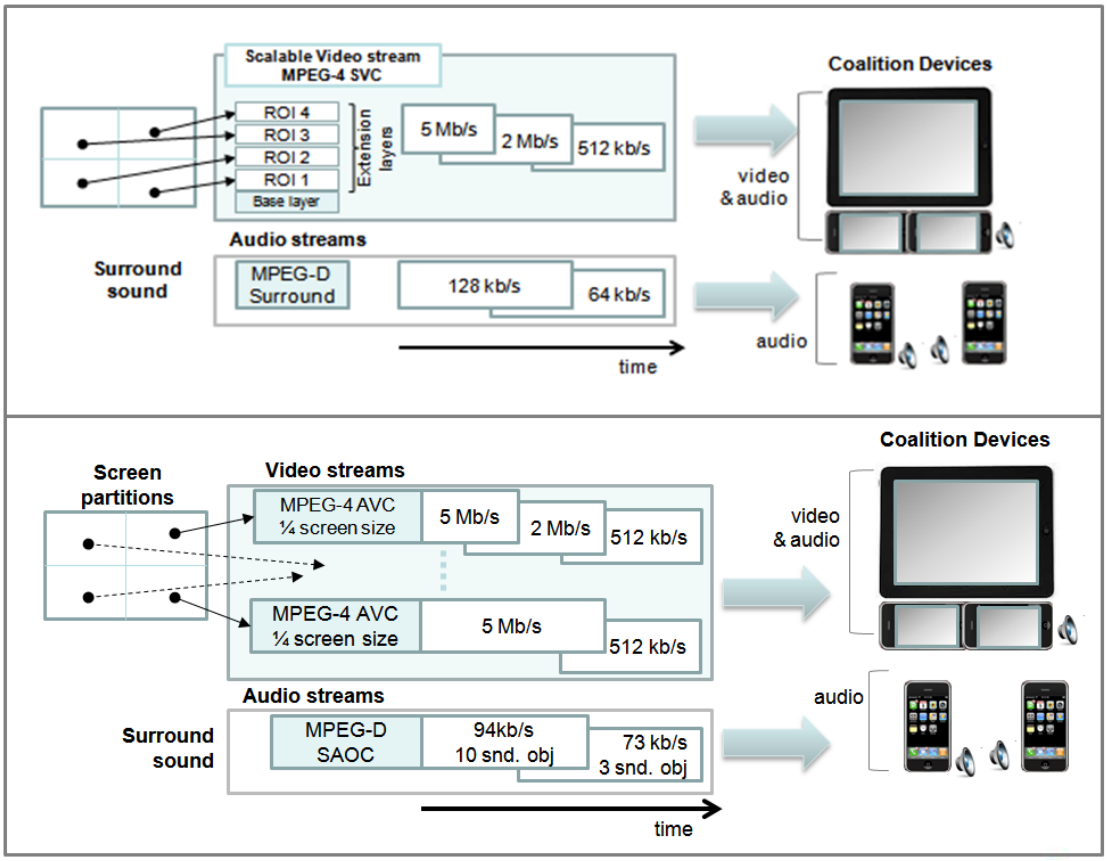

Figure 4. Technical aspects of social cinema coalition service: a) scalable video coding (MPEG-4 SVC), b) non-scalable video coding and flexible sound object coding (MPEG-D SAOC).

1) Media Preparation and Transcoding

Audio and video streams are pre-processed separately for coalitions. In the first step data is acquired from database. Next, data is passing to HTTP streaming server directly or further preprocessed and an optional last step is transcoding (e.g., converting a non-scalable video stream to scalable, multi-layer video stream). The main disadvantage of transcoding is that it can introduce artifacts and so streams

should be generated based on original source data and should cover common use cases.

The main video pre-processing operations includes ROI definition or optionally splitting the video signal into several ROI signals based on the coalition controller parameters. In the simple case video streams should be divided into corresponding number of ROIs. That is, each ROI have to be cropped in each video image and added to new video stream containing a sequence of ROIs. Next, individual streams are encoded using video codec, such as MPEG-4 H.264/AVC standard. Several ROI coding techniques are supported by H.264 but for coalition purposes we propose the use of Flexible Macroblock Ordering (FMO) (slice grouping). FMO allows for dividing image/frame into slices containing group of macroblocks whose ordering within slices is controlled by the encoder. The MPEG-4 H.264/AVC standard defines seven slice group map type. For coalition enablement, FMO Type 2 will be used which enables creating of rectangular slice patterns (group of macroblocks). Each slice is transmitted independently in separate units called packets. Therefore, it is possible to send only required ROI to each device, dropping rest of ROI's packets. Simultaneously, such operation should be supervised by coalition controller on encoder and decoder side. However, the main drawbacks of such solution are transcoding artifacts introduced during decoding and re-encoding process of streams from media repository.

Alternatively, Scalable Video Coding (amendment of H.264) with region of interest coding can be used. Main advantage of such solution is dynamically adjustable bitstream according to the network bandwidth or power constraints of resource-limited systems (eg. mobile devices). In that case, for each ROI to show a part of the "big picture" we propose forming a base image layer (layer 0) with low resolution, which will represent the whole picture and enhancement layers (layer 1,2,3,…) which represent the ROIs. Layer 0 will be transmitted to each device and following layers (ROI with increased resolution) to each devices respectively. As a result, only base and single extension layer will be transmitted to each device, decreasing overall bitrates transmitted to single device and reducing decoding overhead. In comparison to cropped ROI video compression, this would generate a 10% larger bitstream (i.e., average overhead produced by Scalable Video Coding in relation to non-scalable MPEG-4 H.264/AVC coder given the same output resolution). To enable random access to the bit-streams we must encode video stream using the same key-frame (Intra frame) locations in time.

2) Audio Pre-processing

Each co-located device will reproduce audio depending on device position (and role in the service), e.g., left/right channels, surround sound 5.1/7.1. The first approach to audio rendering assumes that each audio signal is transmitted separately to the decoder. Coalition controllers negotiate the number of channels and assign it to particular devices. During data transmission no additional controlling mechanism is needed apart from synchronization of audio streams. This requires separate compression of each audio channel at the encoder side and is power intensive. Moreover, independent encoding of channels increases the amount of bandwidth required to store and transmit the audio data. A more suitable solution could be streaming the complete audio signal to all devices. For example, total bit-stream generated by MPEG Surround for 5.1. channel setup is 64 - 96 kb/s assuming HE-AAC v2 as the core codec. Side information using for matrix process is around 10% of the total HE-AACv2 bit-rate. Therefore, audio decompression will be controlled by a coalition controller at the decoder side. That is, each co-located device will decode only the audio signal to which it is assigned. For example, if the device is assigned as left channel then the audio decoder on that device decodes the entire audio bitstream but keeps only the relevant parts of it (left channel in this case). Such a solution does not require additional preprocessing of audio data at the server side but the total cumulated bitstream will be about 10% higher than in the first scenario we described.

To support the random access point to the bitstream audio data should be synchronized with video, implying that audio codecs cannot operate on data with length longer than video group of pictures (GOP) sequences.

3) Media Presentation Description (MPD)

Data generated during the pre-processing phase are stored than at the HTTP streaming server. We propose standardized adaptive streaming based on MPEG DASH where the server contains different type of encoded multimedia data stored in segments optionally stored in a content delivery network (CDN). Additionally, MPD metadata is stored along with segments. MPD contains information about type, location and relation of multimedia streams in segments. The client application can request segments based on MPD information using HTTP GET and can also control and adjust session parameters, such as changing the media source or choosing media with different bitrate depending on user preferences or network conditions.

Synchronization between client, server and media stream is realized by MPD timestamps of MPEG DASH. However, client and server should operate in the UTC time, to ensure inter-stream synchronization. Such reference clock can be obtained by using e.g., Network Time Protocol (NTP).

IV. PRACTICAL USE CASES

This section describes viable use cases that involve the components that we have described above and that comprise high-experience co-located services.

A. Social cinema

In the "social cinema" use case a number of co-located devices display a movie stream (video and audio). To support this scenario the multimedia stream is encoded in various

configurations. Video is stored in a scalable stream (MPEG-4 SVC) where each ROI represents a part of the frame (e.g., top-left quadrant) and is treated as a separate layer in the stream (spatial scalability). There are three layers with 4, 16 or 64 ROIs (regions of interest) per picture, respectively. That is, the picture is divided equally into smaller fragments which represent ROIs. Two additional layers represent spatial (temporal) scalability, i.e., lower quality video. At the same time, the audio stream is stored using MPEG-D Surround and the server makes two alternative audio bit-streams available, one for 64 kb/s and another for 128 kb/s (higher quality). Alternatively, audio signal can be encoded using MPEG-4 SAOC (Spatial Audio Object Coding) with flexible audio object manipulation. The full quality signal (audio and video) is provided in a 5 Mb/s stream. It also becomes possible to reduce the overall bit-stream by switching to a configuration in which a fewer or greater number of ROIs are encoded.

Let us assume, however, that the coalition in question features four people with five co-located devices: one tablet and four smartphones. The service offers two ½ size ROI layers and four ¼ size ROI layers to cover the entire video frame. The coalition assignment requires the tablet and two of the smartphones to convey the full aspect ratio of the video frame while the two remaining smartphones will be used as satellite speakers in the four-channel audio streaming (i.e., surround sound). At service-request time the co-located devices synchronize clocks with the server. Next, the server begins to stream the requested content. The tablet requests the video stream which consists of two encoded (quarter-sized) ROIs whereas each smartphone requests single (quarter-sized) ROI's such that the full image extent is covered[4]. To adapt to lower bandwidth situations, the video clients switch to a lower bit-rate experience (e.g.,the next step-down may be a 2 Mb/s bit-stream) by rejecting the video layer containing the highest resolution data in favor of a lower resolution one (SNR scalability) while the audio bit-stream remains unchanged. If available bandwidth subsequently again decreases, the audio stream might be switched to the 64 kb/s stream. Figure 4 illustrates the key architectural aspects of social cinema while Figure 5 illustrates one possible orientation of devices to support this use case (e.g., 2 audio roles and 3 video roles).

B. Gaming

Social gaming is a strong use case for mobile device coalitions. With the strength of today's mobile CPUs and the rapid development on mobile GPUs, smartphones are quickly approaching the capabilities of the desktops of only a few years ago. For example, the recently released Nvidia Tegra 3 SoC features an 1+GHz quad-core CPU and a 12 core GPU with video output capabilities up to 2560×1600. A constraint that cannot be easily lifted is the physical screen size, yet mobile coalitions can provide a viable solution.

As we will see in a subsequent section, when a cloud-based approach is employed (such as the one used by OnLive) the result is "virtualized gaming" in which users can experience a multi-player game that would otherwise be incompatible with their current platform. With coalitions and high experience gaming we envision that co-located devices (eg., audio and video outputs) will be used to create a single conglomerated video gaming experience. For example, the gaming graphics may be "wiped" across multiple screens.

For joint display of 3D graphics, each device in a coalition would be responsible for rendering a piece of the player's view depending on its location with respect to other screens. In a typical (standalone) single device 3D game, the 3D content is located on the device, and the device itself controls the position of both the virtual camera and the viewport. There are 3 principle options for implementing coalition games (e.g., imagine, for example, a first-person shooter game with graphics wiped across 3 iPads):

1. Localized views – each tablet understands its layout in the

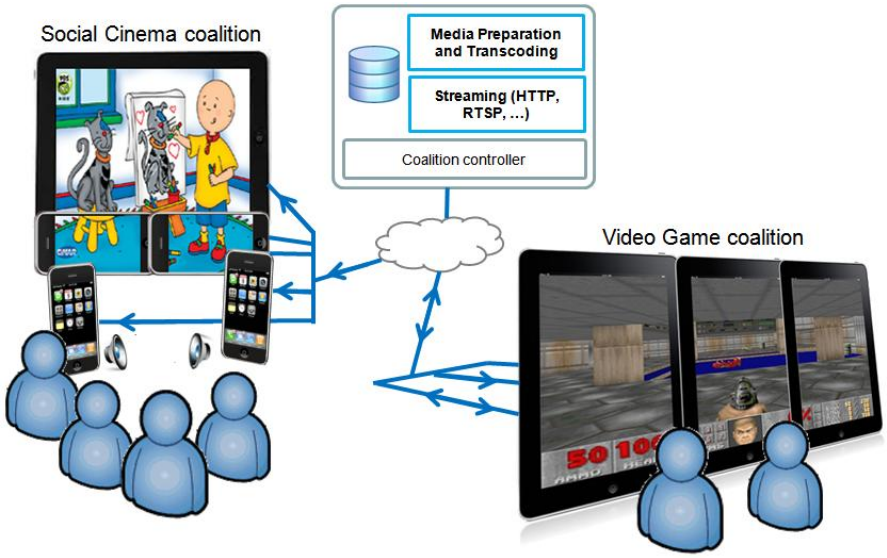

Figure 5. Use cases big picture – support for various coalition-based services will be enabled by the server and appropriate client-side interfaces and capabilities.

coalition, each runs a game instance and each adapts its camera and viewport such that the effect is that of a single wide view into the scene, wiped across the 3 screens. This approach requires relatively little bandwidth but needs a low network delay in order to keep all devices in sync. One of the devices is elected as coalition "leader", and instructs rendering engines on other 2 devices on virtual camera positioning, and collects and processes input events from all devices in the coalition.

2. Server based video streaming - Much like the OnLive service (see www.onlive.com) a dedicated server maintains game state, renders 3D content, and streams the appropriate view (Region of Interest) to each of the 3 tablets who, in turn, send commands to the server. The mobile devices act purely as input/output terminals, i.e., they receive region of the rendered video intended for their viewport. This mode requires a good broadband connection to support high resolutions and low ping times (which are also affected by the delay introduced by codec). OnLive proprietary streaming shows us that typical broadband settings are sufficient for interactive 3D game and other sorts of video streams. The challenges include the IT costs of maintaining such a gaming center.
3. Server based audio streaming – The MPEG-4 SAOC codec is a viable candidate for spatial audio streaming. Such technique enables to flexible control of encoded audio object (e.g. speech, instruments, etc.) in the sound scene. It would therefore be possible to fairly easily manipulate sound sources in the 3D space. MPEG-4 SAOC decoder decides which sound source should be decoded at particular devices. In such a scenario, for example, mobile devices are used as satellite (auxiliary) speakers within a high experience service.

Game control (within a 3-tablet shooter game) can be addressed by either controlling touches on the screens as if the 3 screens comprise a single large virtual screen or to single out one of the devices as a controller. The approach chosen depends on the type of game/application in the coalition service and would take into account whether a touch-screen, keyboard, or console is more relevant, and so on. Regardless, during social co-located gaming action the control information is much more frequent and requires a significantly shorter response time than in the video streaming use case.

As for adapting off-the-shelf games for coalition-based play, we see this as a forward-looking capability that might occur in a middleware layer residing between the OS and the application. Such adaptation – e.g., separating game-play elements (screens, audio, tactile output, etc.) into independently stream-able streams - probably won't be possible unless standards lead the way.

C. Other use cases

Past related works have shown the value of mobile technology in emerging economies to grow and inform communities [12][13]. We envision this kind of architecture being employed on two fronts: a) to provide entertaining educational content to groups of people wherein there are several operational (but far from high-end) mobile devices in the group, and b) to enhance medical communications and imagery presentation in situations where a professional may have access to imagery but only to mobile phones with small screens. By using services that can 'scale' to multiple device output displays a hard-to-read but important set of data – such as X-Rays or medical scans – can be visualized in an ad hoc fashion with the devices that are currently nearby (e.g., a set of professionals happen to be together to interpret some data). We also anticipate an advertising use case in which co-located users can receive multimedia streams in such a way as to enhance the total overall effect of the delivered message (e.g., advertising). Figure 5 illustrates the essence of the look-and-feel of coalition services in the big picture.

V. In Context: OnLive

OnLive (see www.onlive.com) is a popular and fairly successful (though going through business change at the time of writing) cloud-based service that hosts, renders and streams video games and other entertainment to mobile devices. Like other cloud gaming companies (e.g., GaiKai, CiiNOW) OnLive notably delivers cloud-based entertainment as well as virtual Windows 7 operating systems. Its delivery mechanism essentially requires only that the receiving device has sufficient downstream bandwidth (and a software client) – the service is delivered as a compressed video stream, whether it be Web browsing or video gaming. Therefore, a major advantage for OnLive users is that games and OS's can be delivered to a gamut of devices and there is no requirement, for example, that you need a Windows compatible device to experience Windows 7 – an iPad could be the access device.

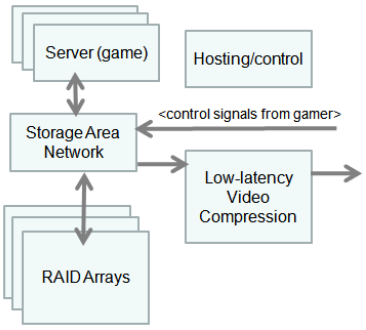

Figure 6. OnLive server-side architecture (adapted from US Patent 2009/0118017A).

OnLive employs an overlay network of streaming servers, to which users must be about within "about 1000 miles". A key quality of experience decider for gaming is latency and OnLive claims that end to end latency should top out at no more than 80ms for paid users near points of presence. But while latency is affected by many aspects (controller input, network, processing) a key notion is OnLive's proprietary video compression algorithm which resides in server hardware and client software. As virtual sessions are created, network channels are opened up from the client to OnLive's servers. Interestingly, with gaming, "two video streams are created for

each game. One (the live stream) is optimized for game play and real-world Internet conditions, while the other (the media stream) is a full HD stream that is server-side and used for spectators or for gamers to record videos of their game play."[5] Additionally, as end users make use of the proprietary game controllers (or other I/O devices) those control commands are streamed directly back upstream to the server, the game state is updated and the imagery is streamed downstream, and so on. OnLive claims to stream adaptively, cutting back on imagery when no changes are occurring on the virtual application but all services are delivered via video streams (even, for example, a virtualized Web browsing session).

TABLE II. COMPARISON WITH ONLIVE

	OnLive	**Architecture and system proposed herein**
Approach	Compress virtualized server-side screens and streaming them as video to a **single** end device, or the same thing (possibly scaled) to multiple devices.	Divide up the multimedia aspects of the service into **distinct parts** and stream individually to **different** end devices.
Openness	Proprietary; pay per use.	TBD – open standards are desirable.
Services	Only services that OnLive chooses to host will be available.	In principle supports complex multi-device services.
Service initiation	Servers and compression modules are instructed to send appropriate stream to new client (depending on size of screen); client controls are sent via a new channel towards server(s).	Since multiple devices allow multiple possibilities for display configuration we require a "device-assignment" step not found in OnLive. Resulting services more "componentized" (separate streams).
Client	Requires proprietary client software (and optionally custom game I/O devices such as game controller).	Requires a software client to coordinate with other local devices and with server-side controller.
Extensibility	Proven to be extensible; games of various sorts have been supported as have OS's and other applications.	Extensible –with a combination of different transcoding techniques.

While OnLive's compression algorithm is proprietary, some interesting technical information is gleaned from the founder's US patents [25][26] in which he describes delivering video games to wireless users via set-top boxes, a method for streaming multi-user games, and adaptively compressing video based on "tiling". Some aspects of OnLive's high-level server-side architecture are shown in Figure 6.

Table 2 describes some of the similarities and differences between OnLive and our proposed architecture.

[5] Wikipedia, "http://en.wikipedia.org/wiki/Onlive#Architecture" Feb.24, 2012

VI. RESULTS AND CONCLUSIONS

The principle benefits of our technology include:

- Enabling systematic creation of mobile coalitions through a series of interactive steps on an intuitive interface
- Allowing multimedia service content to be experienced even when no single mobile device is capable of rendering it in whole
- Creating new social collaborative services that could not otherwise exist.

The practical realization of this sort of service is not easy, despite successes of earlier related work [3]. At a high level the user experience of co-located streaming services could easily be marred by:

- Devices falling out of sync with others
- Devices with inadequate resources (CPU, RAM for buffers, etc.)
- Network congestion

These issues are of continuing interest. For example, in order to effectively setup local devices for a high experience collaborative service we presume that a) personal area network technology (e.g., Bluetooth™) will provide a communication backbone and b) the human participants will help organize, define, and assign devices to the service. As another example, today's streaming standards support dynamic adaptation (see MPEG DASH) so we presume that will, in some ways, be sufficient (indeed successful DASH-based interactive television pilots in-the-large have show practical applicability). Also, in-the-large Inter-device synchronization is commonly achieved via NTP (UTC) and, while out of scope for this paper, we presume such techniques will suffice. Finally, we can do little about network congestion except use protocols that adapt streams accordingly (again, see MPEG-DASH).

Our work thus far, therefore, comprises a pragmatic examination of current multimedia standards and possible architectures to support what we consider will comprise a new user experience: co-located collaborative services over the Internet. We are inspired by past techniques (e.g., Web page adaptation, video ROI techniques) and use them as underpinnings so as not to reinvent needlessly. That we have not found examples of service providers offering adaptive co-located collaborative services in the manner that we describe is exciting as we believe that this new niche is imminent and not at all "out of reach".

From a marketing point-of-view, we hypothesize that mobile content providers may be able to generate new revenues from offering co-located collaborative services as pay-to-play, while network providers could see new services and revenues revolving around service setup and targeted advertisement. Device manufacturers may embed the required multimedia software components into their devices or make the devices "coalition-ready". Again, in our opinion, the services we describe are not a huge leap away given the

business models and user experiences of today's cloud-gaming (e.g., OnLive, GaiKai).

There is much still to be done such as addressing service discovery, adapting to device resolutions and network conditions, and so on. In addition, without extensive prototyping we do not have a qualitative basis for claiming improved quality but we are encouraged by our results thus far and continue to work on the challenging issues of region-of-interest control, multimedia layering, and user experience. Our future work will certainly include practical streaming and synchronization tests on mobile devices and viability tests of our principle use cases, such as social cinema.

VII. WEB LINKS

CityVille – http://apps.facebook.com/cityville
MeetMoi – http://www.meetmoi.com
Sonar.me – http://www.sonar.me
VideoLAN – http://www.videolan.org
Spotify – http://www.spotify.com
Geocaching - http://www.geocaching.com/
Bluetooth - http://www.bluetooth.com
Flickr photo sessions - http://www.flickr.com/photosession
Google Plus - https://plus.google.com
Epson MegaPlex Projector - http://bit.ly/qUjiot
OnLive Desktop – http://www.onlive.com

REFERENCES

[1] P.Verna, "Behind emarketers social game numbers", http://emarketer.com/blog/ index.php/numbers-emarketers-social-gamers/

[2] L.Brunnberg, "The Road Rager: making use of traffic encounters in a mobile multiplayer game", *Proc. ACM 3rd international conference on Mobile and ubiquitous multimedia (MUM'04)*, Maryland, 2004

[3] R. Han, V.Perret, M.Naghshineh, "WebSplitter: a unified XML framework for multi-device collaborative Web browsing." *Proc. ACM Conf. on Computer supported cooperative work* (CSCW '00), pp.221-230, Philadelphia, 2000

[4] M.Li, S.Yu, W.Lou, K.Ren, "Group Device Pairing based Secure Sensor Association and Key Management for Body Area Networks," Proceedings of INFOCOM 2010, vol., no., pp.1-9, 14-19 March 2010

[5] F.Cricri, S.Mate, I.Curcio, M.Gabbouj, "Mobile and Interactive Social Television – A Virtual TV Room", *Proc. IEEE Symp. On World of Wireless, Mobile and Multimedia Networks & Workshops*, pp.1-8, Greece, 2009

[6] X.Sun, J.Foote, D.Kimber, B. S. Manjunath,"Region of Interest Extraction and Virtual Camera Control Based on Panoramic Video Capturing" IEEE Transactions on Multimedia, 2004.

[7] A. Mavlankar, D. Varodayan, and B. Girod. Region-of-interest prediction for interactively streaming regions of high resolution video. In Proc. IEEE Packet Video Workshop, pages 68--77, Nov. 2007.

[8] H. Cheng et al., "Automatic video region-of-interest determination based on user attention model",*Proc. IEEE Int. Symposium on Circuits and Systems*, 4, pp. 3219-3222, 2005.

[9] J.Zhang; L. Zhuo; L.Shen; , "Regions of Interest extraction based on visual attention model and watershed segmentation," Neural Networks and Signal Processing, 2008 International Conference on , vol., no., pp.375-378, 7-11 June 2008

[10] C.Qing-hua; X. Xiao-fang; G. Tian-jie; S. Lei; W. Xiao-fei; , "The Study of ROI Detection Based on Visual Attention Mechanism", Int'l Conf. on Wireless Communications Networking and Mobile Computing, pp.1-4, 23-25 Sept. 2010

[11] F. Ribeiro and D. Florêncio, "Region of Interest Determination Using Human Computation," in *Proc. IEEE International Workshop on Multimedia Signal Processing*, 2011

[12] B.Falchuk, R.Fisher,"Strategies for Adaptive Online Health Communities*", Proc. of 15th Annual International Meeting and Exposition of the American Telemedicine Association* (ATA 2010), San Antonio, 2010

[13] J.Veijalainen, W.Rehmat, "Mobile Communities in Developing Countries", *Proc. 11th Int'l. Conf. on Mobile Data Mgmt.,* Kansas City, 2010

[14] International Standard ISO/IEC 13818-3, Generic Coding of Moving Pictures and Associated Audio Information - Part 3: Audio, 1998.

[15] International Standard ISO/IEC 13818-7, Generic Coding of Moving Pictures and Associated Audio Information: – Part 7: Advanced Audio Coding (AAC), 2006.

[16] T. Żernicki, M. Bartkowiak, M. Domański, "Enhanced Coding of High-Frequency Tonal Components in MPEG-D USAC through Joint Application of eSBR and Sinusoidal Modeling," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, 2011.

[17] ISO/IEC 14496-10 (MPEG-4 AVC) / ITU-T Rec. H.264. "Advanced Video Coding for Generic Audiovisual Services", 2003–2007.

[18] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.

[19] R. Pantos and E.W. May, "HTTP Live Streaming", IETF Internet draft, work in progress, Mar. 2011.

[20] Microsoft, IIS Smooth Streaming Transport Protocol, Sept. 2009; http://www.iis.net/community/files/media/smoothspecs/[MS-SMTH].pdf.

[21] Adobe Systems Inc., "HTTP Dynamic Streaming on the Adobe Flash Platform.", available at http://www.adobe.com/products/httpdynamicstreaming, 2010

[22] ISO/IEC FCD 23001-6, "Part 6: Dynamic Adaptive Streaming Over HTTP (DASH)," MPEG Requirements Group, Jan. 2011.

[23] ISO/IEC 14496-12, Information Technology—Coding of Audio-Visual Objects—Part 12: ISO Base Media File Format, 2008.

[24] 3GPP TS 26.244, Transparent End-to-End Packet Switched Streaming service (PSS); 3GPP file format (3GP).

[25] S.Perlman, US Patent 7,849,491, "Apparatus and method for wireless video gaming", available via www.uspto.gov

[26] S.Perlman, US Patent 2009/0118017A, "Hosting and broadcasting virtual events using streaming interactive video", available via www.uspto.gov

[27] E.Braun, M.Muhlhauser, "Interacting with Federated Devices", Proc. Advances in Pervasive Computing, 2005

[28] T.Tullis, B.Albert, "Measuring the User Experience", Morgan Kaufman, Amsterdam, 2008.