Mobile Serverless Video Communication

Hans L. Cycon
FHTW Berlin, Germany

Thomas C. Schmidt
HAW Hamburg, Germany

Matthias Wählisch
FHTW Berlin, Germany

INTRODUCTION

Voice and video conferencing have been well established as regular communication services within the wired Internet. Facing the paradigm of ubiquitous computing and mobile communication, they are on the spot to be launched within a wireless Internet infrastructure. Following an 802.11, 802.16 or 3G standard, wireless networks provide enough bandwidth to support data intensive communication services such as videoconferencing. The vision of nomadic users at roaming devices performing synchronous communication, such as voice or videoconferencing over IP (VoIP/VCoIP), is around, but raises new challenges for the Internet infrastructure.

In conferencing scenarios addressability raises the first major issues. To globally call a device, a routable IP address must be in use. On a large scale such address space is only provided by IPv6. To identify a communication partner's current device, a supplementary global user locating scheme is needed. In wireless infrastructures, where users share limited bandwidth from a restricted frequency space, multicasting is needed to enable group conferencing compliant to transmission resources and without placing the burden of dedicated group-server infrastructure.

At the same time, synchronous real-time applications, such as VoIP and VCoIP place new demands on the quality of IP mobility services: packet loss, delay and delay variation (jitter) in a constant bit rate scenario need careful simultaneous control. A spoken syllable is about the payload of 100 ms continuous voice traffic. Each individual occurrence of packet loss above 1%, latencies over 100 ms or jitter exceeding 50 ms will clearly alienate or even distract the user. Audio and visual streams in video conferencing additionally require tight synchronization. Inter-stream latencies should remain below 30 ms for audio arriving ahead, 40 ms for audio being behind. While uni-directional distribution may compensate quality deficits by buffers, available techniques of hiding packet loss at the cost of delay and jitter or vice versa are of limited use in conferencing. Their requirements impose strong challenges on a mobile Internet scenario. Challenges are even tightened by multicast-based group communication, since in conferencing scenarios each member commonly operates as receiver and as sender. Real-time requirements consequently are a major driving force for the development of a seamless mobile Internet layer.

In concordance with communication capabilities, video coding techniques have evolved, as well. The latest standard for video coding H.264/AVC (ITU H.264, 2005), although designed as a generic standard, is predestined for applications like mobile video communications (Stockhammer, Hannuksela, & Wiegand, 2003). Besides enhanced compression efficiency, it delivers also network friendly video representation for interactive (video telephony) and non-interactive applications (broadcast, streaming, storage, and video on demand). H.264/AVC provides gains in compression efficiency of up to 50% over a wide range of bit rates and video resolutions compared to previous standards. While H.264/AVC decoding software has been successfully deployed on handhelds, high computational complexity still prevents pure software encoders in current mobile systems. There are, however, also fast hardware implementations available (see a list in Wikipedia, H.264, 2006). Next generation codes, like scalable video coding (SVC) are already in a design state (Reichel, Schwarz, & Wien, 2005; Schwarz et al., 2004). The main new feature, scalability, addresses schemes for delivery of video to diverse clients over heterogeneous networks, particularly in scenarios where the downstream conditions are not known in advance. The basic idea is that *one* encoded stream can serve networks with varying bandwidths or clients with different display resolutions or systems with different storage resources, which is an obvious advantage in heterogeneous networks prevalent in mobile applications.

BACKGROUND

Video conference communication is a person-oriented, session-based service. A caller requesting contact to one or several partners will expect to address a personal identifier, but establish the corresponding conference session with the devices currently in use by the callees. Unlike in mobile telephony, the Internet architecture is required to locate users and mask the user-device mapping, following the paradigm of location transparency, like in e-mail services. Once established, sessions need to persist while mobile devices roam. Intermediate handovers thereby should unnoticeably comply with quality of service measures for real-time communication. Operating on portable devices with limited capacities of CPU, batteries and displays, a video conferencing solution needs to balance out network efficiency and adaptability versus coding complexity. Lightweight flexible software systems as introduced by Cycon et al. (2004) are preferably employed in mobile communication.

Conference Signaling

The traditional, ISDN compatible architecture of VCoIP systems has been defined in the standard ITU H.323 (2000). Central parts of this model are derived from a client-server principle with a Gatekeeper, providing connection control and address translation, and a multipoint control unit (MCU) serving video streams in multipoint conferences. The H.323 architecture must be considered local and immobile in the sense that all participants need to agree on common MCU and pre-configured Gatekeeper servers, which, at least for the MCU, suffer from severe scaling deficiencies.

A flexible, fairly general Internet signaling solution has been presented with the session initialization protocol (SIP) (Rosenberg et al., 2002). Beside user localization, SIP covers negotiations about user capabilities, user availability, the call set-up by session description protocol (SDP) data and the handling of the calls itself. SIP introduces its own infrastructure of servers, which actively perform a peer-to-peer routing by using SIP-URLs. SIP is based on an extensible method framework and open to store persistent data. SIP liberates the rigid addressing scheme of telephone numbers used in H.323, proposing addresses of the "e-mail-like" form <user>@<SIP-server>. A basic interaction with IP multicast is defined in SIP through the *maddr* address attribute in the VIA header.

Mobility Management

As an application layer protocol SIP provides some mobility support to session-based services (Wedlund & Schulzrinne, 1999), which requires implementation at the application layer. Employing the regional SIP server as an application specific home agent, handoff notifications are traded via regular SIP messages to the home server (register) and the correspondent node (re-invite). As SIP mobility operates above the transport layer, it remains self-consistent and transparent to the Internet infrastructure, but inherits all underlying delays in addition to its own signaling efforts.

The fundamental approach to mobility management in the next generation Internet is the Mobile IPv6 (MIPv6) RFC (Johnson, Perkins, & Arkko, 2004). MIPv6 transparently operates address changes on the IP layer as a device moves from one network to the other by sustaining original IP addresses in a home address destination option and hiding the different routes to the socket layer. In this way, hosts are enabled to maintain transport and session connections when they change locations. An additional infrastructure component, the MIPv6 Home Agent, preserves global addressability, while the mobile node is away from home.

Local handovers in MIPv6 are rapidly completed. In the presence of layer 2, triggers for movement detection, the time needed for address reconfiguration and local updates remains well below 10 ms (Schmidt & Wählisch, 2003). Distributed Mobile IPv6 scenarios, though, inherit strong topology dependence from binding updates with the Home Agent and correspondent nodes. To resolve topologically originated delays, Hierarchical MIPv6 (Soliman et al., 2005) has been introduced for micro mobility scenarios and Fast Handovers (Koodli, 2005) for delay hiding by means of handover predictions. Even though an expected disappearance of predictive handover delays does not hold in practice, these accelerating schemes arrive at real-time compliant handover performance (Schmidt & Wählisch, 2005).

IP layer handovers thus can be considered capable of a mobility management for real-time voice and video communication. SIP application layer handovers have been found by Kwon, Gerla, Das, & Das (2002)) to significantly fall behind Mobile IPv4, which itself is largely outperformed by MIPv6. Handoff disruptions of the underlying layer 2 add to service degradation, admitting vendor specific, but large values in 802.11 systems (Mischra, Shin, & Arbaugh, 2003). Further quality of service issues of wireless transmission technologies and of the general IP routing layer need consideration, as well.

Video Coding

H.264/AVC, or MPEG-4 Part 10, (formally, ISO/IEC 14496-10) is a digital video codec standard. It was defined by a collective effort of the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) Video Coding Experts Group (VCEG) and the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) Moving Picture Experts Group (MPEG) known as the Joint Video Team (JVT). The final drafting work on the first version of the standard was completed in May of 2003.

H.264 is a name related to the ITU-T line of H.26x video standards, while AVC (advanced video coding) relates to the ISO/IEC MPEG side of the partnership project that completed the work on the standard. H.264/AVC has also been referred