Connecting the Worlds: Multipoint Videoconferencing Integrating H.323 and IPv4, SIP and IPv6 with Autonomous Sender Authentication

Hans L. Cycon[§], Gabriel Hege[*], Detlev Marpe[†], Mark Palkow[¶], Thomas C. Schmidt[*], and Matthias Wählisch[*‡]
[*]HAW Hamburg, [†]Fraunhofer HHI Berlin, [‡]Freie Universität Berlin, [§]FHTW Berlin, [¶]daViKo GmbH Berlin
{h.cycon,hege}@fhtw-berlin.de, marpe@hhi.fraunhofer.de, palkow@daviko.com, {t.schmidt,waehlisch}@ieee.org

***Abstract*—In this paper we present a multipoint media conference software with extended network capabilities. Its core components combine an advanced, highly efficient H.264/AVC video coding system. Session signaling follows the SIP standard and simultaneously supports IPv4 and IPv6. Built upon an underlying peer-to-peer communication scheme, it further supports advanced integration with legacy MCUs and thereby enables hybrid sessions that extend into the H.323 world. We demonstrate how our software can span fully functional conferences across these worlds by deploying a passive gateway peer. Finally, we address security issues arising in distributed conferencing systems. It is shown, how the use of cryptographically generated identifiers enables the application to authenticate data on a packet level, thereby preventing abuse and impersonation of the conference overlay network.**

***Index Terms*—P2P group communication, distributed SIP conferencing, autonomously verifiable member authentication, cryptographic identifiers**

I. Introduction

Videoconferencing systems enriched with various media applications are exposed to growing demands on functionalities and range of applications, as well as on interconnectivity capabilities within heterogeneous network protocols and various deployment scenarios. On the one hand, there are hardware videoconferencing systems with centralized MCU-based architectures using H.323 [1] signalling, connected with company or university LANs or ISDN lines. On the other hand, separated from these homogeneous networks, there are soft clients with videoconferencing applications implemented on PCs or mobile devices using SIP standards [2] on IP-based peer-to-peer (P2P) communication schemes. Beyond the established IPv4 network, which is heavily burdened by middle boxes, the next generation Internet (IPv6) is spreading at accelerating speed. Solutions which integrate soft clients into centralized systems on H.323 basis do exist, but we are neither aware of fully integrated hybrid H.323/SIP communication systems, nor a fully functional application peer-wise adapting to the different Internet protocols.

Applications for video/media conferencing systems range from peer-to-peer chat over synchronous eLearning and medical consultations up to high profile professional management meetings supported by telepresence systems. This implicates growing demands on security and information confidentiality for the communication system. Conferencing solutions are required to preserve confidentiality and resilience against denial of service attacks. In addition, fully distributed p2p applications need to protect their internal distribution network from bogus participants that may misuse the overlay infrastructure for amplified flooding attacks.

In this paper, we present a multipoint video/media conference software with extended network capabilities. In Section II, we introduce the core components combining an advanced, highly efficient H.264/AVC [3], [4] coding system with a network-adaptive data distribution layer. Section III presents the distributed hybrid conferencing scheme where signaling follows the SIP standard and which simultaneously supports IPv4 and IPv6. We show that the advanced integration with legacy MCUs enables hybrid sessions that extend into the H.323 world. In Section IV, we address security issues of the distributed conferencing system. Based on cryptographically generated identifiers, the application is efficiently enabled to authenticate data on a packet level, thereby preventing abuse and impersonation of the conference overlay network. We conclude and give an outlook in Section V.

II. The Videoconferencing Software

The basic digital audio-visual conferencing system called daViKo is realised as a serverless multipoint videoconferencing software [5]. It has been designed basically in a peer-to-peer model as an Internet conferencing tool. The system is built upon a fast H.264/AVC video codec. The codec along with the H.264/AVC design also includes some suitable mechanisms to recover quickly from video packet loss.

In addition to its video conferencing capabilities, daViKo provides an application-sharing/application-broadcasting facility for collaboration and teleteaching. It enables participants to share or broadcast not only static documents, but also any selected dynamic PC actions such as animations. All audio/video streams including the dynamic application sharing can be recorded on any site.

The system is applied in various distributed synchronous and asynchronous eLearning scenarios. Synchronous learning is realized by teaching and collaborative learning using the video conferencing system on standard PC technology over IP. The approach enables audio/video-based distance learning on a

This work is supported by the German BMBF within the project Moviecast (http://moviecast.realmv6.org).

Fig. 1. Sample Video Conference Integrating a Multi-point Control Unit (MCU) and a Mobile Presented at CeBIT '09

lowest technical level. All sessions can be real-time converted into a live streaming format, suitable for serving a broader audience. For asynchronous distributed learning, the recorded sessions will be transcoded into a variety of different formats for off-line streaming or downloads to replay on almost any mobile device [6].

The conferencing system is available for desktop computers running MS-Windows or Linux and on handhelds equipped with the Windows Mobile operating system [7]. This enables heterogeneous conferences where any mobile and desktop user can participate in established Intranet conferences.

A. P2P Adaptive Architecture

The communication subsystem of daViKo is designed along a hybrid network architecture that avoids infrastructure dependencies, but procures end-to-end accessibility in the presence of NATs and firewalls.

Focusing on ad hoc groups of limited size, users retrieve address references from an LDAP-based presence server [8] and directly connect to the callee. This directory server simultaneously can serve as a NAT traversal assisting super peer, if needed. For limited group sizes this super peer, or conference focus can easily channel the audio/video streams, attaining the role of a reflector. For larger conferences, extended versions of distributed conference focus points are provided that mutually balance the load of media stream replication and can assist weak clients, e.g., from the mobile world.

If all peers, however, are located within an Intranet which provides appropriate network support, a pure multicast streaming is used among clients and the conference communication system fully scales in the number of passive participants. Active contributors are of course limited to the number of simultaneously processable video streams.

III. Integration of IPv6/IPv4 and Legacy MCUs

A SIP [2] stack is implemented as part of the clients. This provides connectivity to customary SIP-based video conferencing systems and in particular to SIP-enabled MCUs of the H.323 [1] world. Basic SIP negotiations for client initiated conferencing [9], [10] are supported according to the standards. However, a full-featured integration of a lightweight, peer-centric software with infrastructure constraints and legacy entities in the network requires a number of additional efforts and solutions. In particular, session border controllers (SBCs) are needed to organize the transit between network and application layer protocols.

A. Managing the Internet Protocol Versions

A globally addressable and sufficiently powerful peer can act as a SIP session border controller, which is a realistic scenario in the IPv6 world. Our system, though, is designed for IPv4/IPv6 dual-stack operation, inheriting the feasible version from the network layer, and protocol knowledge from the SIP signalling stack. However, a transparent support for both protocols in conferencing requires a seamless, reliable media agreement that needs to be achieved within SDP session negotiation. Members are required to mutually explore and decide on the Internet protocol version they are entitled to use, which in turn depends on knowing the connectivity type of each party in advance. As shown in the upper part of the dialog displayed in Figure 2, we solve this by sending the initial INVITE without SDP, leaving the offer to the callee after it already has experienced network (protocol) contact. Transmitting SDP within the SIP OK message then includes multi-protocol endpoints using the "Alternative Address Type" Tag (ANAT) [11]. After completing this dialog, both parties are aware of Internet protocol versions available for communication. In this way, IP versions can be reliably determined for media streams, giving preference to IPv6.

In a pure IPv4 world, there are many scenarios, where global connectivity is likely to fail. We advise for and offer a permanently deployed 'silent' relay-peer at some unrestricted place. This 'hidden helper' can serve as an assistant negotiator and take the role of a representative SBC.

B. Incorporating Legacy MCUs

Typically, an existing conference proceeding on an MCU can be joined by calling a conference specific SIP URI. To participate in a full-featured conference, however, procedures which are not standard SIP are needed in order to gain further control over MCUs or other SIP enabled legacy devices that have been originally designed for H.323 or the telephone network. For example in the case of MCUs, it is common that the client can initiate a SIP session to the MCU, which then displays a user interface via the video channel. Using this interface, the client can access a list of existing conference rooms or create new ones. Such features are commonly controlled in an H.245 fashion by the media streams itself. Support for the latter user commands of interacting with non-SIP interfaces are available in the daViKo application in a twofold way: (1) Codes for DTMF tones can be sent via the RTP audio channel according to [12], and (2) "H.224/H.281 Far-end Camera Control" commands can also be sent via RTP

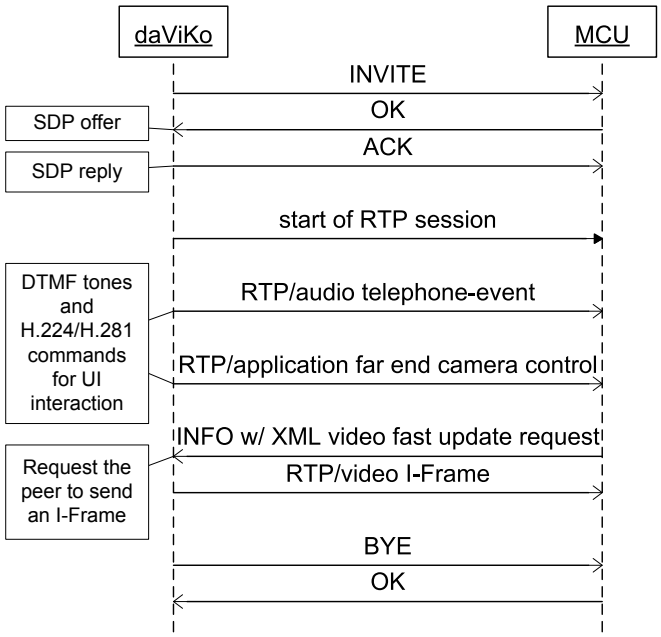

Fig. 2. SIP Call Flow for Integrating IP Versions and H.323 Legacy MCUs based on a Session Border Controller (SBC) and a Multi-point Control Unit (MCU)

[13], [14]. The former is also especially useful for interaction with any types of devices accessible through a SIP/telephone gateway.

In the centralized conference environment provided by legacy MCUs, additional functionalities are needed. Receivers experiencing excessive packet loss or having recently joined a fully distributed multiparty conference may not have received the previous frames referenced by the blocks that are currently transmitted. Thus an intra frame is needed to enable the decoding of the upcoming video stream. An out-of-order intra frame transmission can be signaled to the sender by a "Fast Update Command". The latter consists of an XML document, encoded according to the "XML Schema for Media Control" [15]. This selective update trigger is transmitted within a SIP INFO-Message [16].

IV. Member and Stream Authentication

A. Security Threats in Distributed Multi-Party Conferences

In videoconferencing systems, new parties are commonly authorized for joining the session by off-line credentials or a manual admission through established members. The daViKo system operates in ad hoc mode and relies on the latter scheme that allows for personal identification of a callee.

In an established conference, however, a number of security threats remain valid. At first, a threat of impersonation aiming at a theft of service arrives from the ability of SIP to redirect session membership. By spoofing the SIP contact URI, an adversary may issue a re-INVITE into the dialog and redirect media streams. While media encryption does prevent eavesdropping, this redirecting may disturb or even terminate the conference.

Second, group communication is inherently predestined to facilitate Distributed Denial of Service (DDoS) attacks as data will automatically be replicated to several nodes. An attacker could inject bogus packets using spoofed overlay identifiers, and conference members would unwillingly assist in amplifying the unwanted traffic. This becomes even more severe in the context of P2P networks, as overlays place enhanced stress onto the underlying infrastructure. The absence of authentication mechanisms thus leads to simple leaks in protecting conference parties, as well as the infrastructure itself.

B. Autonomous Sender Authentication

The abuse of the conference session and distribution infrastructure can be prevented, if packet forwarders and receivers are enabled to verify the legitimacy of a sender, i.e., require a source to authenticate with respect to the group.

The traditional way of organizing authentication and authorization in a group of previously unknown members relies on a trusted third party. Such a certifying authority may issue credentials that serve as valid authenticators. However, lightweight ad hoc conferencing aims at avoiding such an infrastructure entity. Its overlay content distribution is organized among independent peers that follow user instructions and autonomously agree on a distribution scheme and a conference identifier. Following this paradigm, authentication should proceed by an autonomously verifiable scheme, as well.

Currently, the only known method for self-certifying authenticity is by the use of cryptographically generated identifiers (CGIs). Having its seeds in cryptographically generated IPv6 addresses (CGAs) [17], cryptographic identifiers have been transferred to SIP URIs [18] and overlay addressing [19]. Based on public key cryptography, a sender creates its CGI from the public key and signs the message with its private key. Any receiver is thus enabled to jointly verify the message *and* the identifier of the sender on message reception without the need for an external authority.

In the context of group conferencing with SIP, we now generalize the approach of cryptographically generated identities to a combined authentication scheme of messaging and group communication. In detail, a caller contacts a conference member using INVITE with its common SIP URI in the regular from field, but with its SIP CGI in the CONTACT header field. On reception, the callee will verify the SIP CGI. Only then the call may be interactively accepted by a user dialog at the callee, which will respond according to the CONTACT header, likewise issuing its own CGI in the CONTACT field of the reply. The caller will implicitly accept callee's identity by continuing the dialog after CGI verification. Following this accept, a mutual key verification has completed and both parties are aware of each others public keys.

At this stage, the SIP messaging is protected from unauthorized frauds, and all peers can mutually exchange private

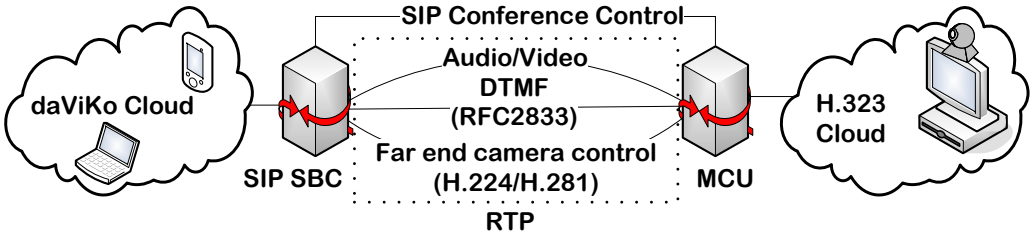

Fig. 3. Integrating H.323 and SIP based on a Session Border Controller (SBC) and a Multi-point Control Unit (MCU)

credentials for a peer-wise secure media transmission. In a group conference scenario, though, point-to-point transmission will require a transcoding, which for efficiency reasons is omitted by using a common group key for encryption. This group key can be securely distributed among peers using the established public key cryptography, but does not allow for individual sender authentication.

To authenticate media streaming sources, the scheme needs the following extension. The creator of a group or group controller that has generated its SIP CGI in the first step, will configure a cryptographic group address or conference URI G from the same public-private key pair $(\mathcal{K}_{pub}, \mathcal{K}_{sec})$.

In signing the packets using $\mathcal{K}_{sec}$ and attaching $\mathcal{K}_{pub}$, the group controller will provide cryptographically strong proof of conference ownership (beside proof of identity) to any receiving peer of the packet. After extracting $\mathcal{K}_{pub}$, an intermediate node can reconstruct source and group address, match them to the pre-established SIP URI and validate the signature. Having verified that the source is the valid owner of the group, data will be forwarded according to the P2P protocol in use. In any case of failure, the P2P forwarder drops the packet, thereby cutting distribution along multicast branches. Authentication and authorization extend to the multi-source scenario in conferencing with the help of certificates issued by the group controller. Details can be found in [20] and are omitted here for the sake of brevity. Note that for this 'AuthoCast' group authentication scheme a native implementation for IPv6 exists based on standard protocol elements [21].

V. Conclusion & Outlook

We have presented a distributed management software for high-quality videoconferencing using a highly optimized H.264/AVC codec. The system integrates IPv6 with IPv4 on signalling and media application level. A SIP/H.323 passive gateway enables hybrid sessions between SIP signalling soft clients and participants of an MCU-backed conference. Cryptographically strong, autonomously verifiable authentication and key establishment is part of our solution, as well.

In future work, we will concentrate on further optimization and generalization of the video coding software to meet emerging SVC features. SVC encoded video streams are of vital importance for seamlessly aiding mobile conference members. Particular focus will be given to optimize stream authentication in the presence of SVC, aiming at a lightweight and selective treatment of the codec layers.

References

[1] ITU-T Recommendation H.323, "Infrastructure of audio-visual services - Systems and terminal equipment for audio-visual services: Packet-based multimedia communications systems," ITU, Tech. Rep., 2000, draft Version 4.

[2] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol," IETF, RFC 3261, Jun. 2002.

[3] ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC, "Advanced Video Coding for Generic Audiovisual Services," ITU, Tech. Rep., 2005, draft Version 3.

[4] J. Ostermann, J. Bormans, P. List, D. Marpe, N. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video Coding with H.264/AVC: Tools, Performance and Complexity," *IEEE Circuits and Systems Magazine*, vol. 4, no. 1, pp. 7–28, April 2004.

[5] M. Palkow, "The daViKo homepage," 2009, http://www.daviko.com.

[6] H. L. Cycon, A. C. Wagner, F. Topfstedt, and H. Regensburg, "Distribution & Communication Tools for Video-based m-Learning," in *Wireless Communication and Information*, J. Sieck and M. Herzog, Eds. Boizenburg: Verlag Werner Hülsbusch, 2008, pp. 173–184.

[7] H. L. Cycon, T. C. Schmidt, G. Hege, M. Wählisch, and M. Palkow, "An Optimized H.264-based Video Conferencing Software for Mobile Devices," in *12th IEEE International Symposium on Consumer Electronics ISCE 2008 Proceedings*. IEEE Press, April 2008.

[8] T. C. Schmidt, M. Wählisch, H. L. Cycon, and M. Palkow, "Global serverless videoconferencing over IP," *Future Generation Computer Systems*, vol. 19, no. 2, pp. 219–227.

[9] A. Johnston and O. Levin, "Session Initiation Protocol (SIP) Call Control - Conferencing for User Agents," IETF, RFC 4579, Aug. 2006.

[10] T. C. Schmidt and M. Wählisch, "Group Conference Management with SIP," in *SIP Handbook: Services, Technologies, and Security of Session Initiation Protocol*, S. Ahson and M. Ilyas, Eds. Boca Raton, FL, USA: CRC Press, 2008, pp. 123–158.

[11] G. Camarillo and J. Rosenberg, "The Alternative Network Address Types (ANAT) Semantics for the Session Description Protocol (SDP) Grouping Framework," IETF, RFC 4091, Jun. 2005.

[12] H. Schulzrinne and T. Taylor, "RTP Payload for DTMF Digits, Telephony Tones, and Telephony Signals," IETF, RFC 4733, Dec. 2006.

[13] "A real time control protocol for simplex applications using the H.221 LSD/HSD/MLP channels," ITU, Recommendation, 2005.

[14] "A far end camera control protocol for videoconferences using H.224," ITU, Recommendation, 1994.

[15] O. Levin, R. Even, and P. Hagendorf, "XML Schema for Media Control," IETF, RFC 5168, Mar. 2008.

[16] S. Donovan, "The SIP INFO Method," IETF, RFC 2976, Oct. 2000.

[17] T. Aura, "Cryptographically Generated Addresses (CGA)," IETF, RFC 3972, Mar. 2005.

[18] J. Seedorf, "Using Cryptographically Generated SIP-URIs to Protect the Integrity of Content in P2P-SIP," in *3rd Annual VoIP Security Workshop*, Berlin, Germany, 2006.

[19] I. Baumgart, "Peer-to-Peer Name Service (P2PNS)," IETF, Internet Draft – work in progress 00, Nov. 2007.

[20] M. Wählisch, T. C. Schmidt, and G. Hege, "Overlay AuthoCast: Distributed Sender Authentication in Overlay Multicast," in *Proceedings of the 28th IEEE INFOCOM*. IEEE Press, April 2009.

[21] T. C. Schmidt, M. Wählisch, O. Christ, and G. Hege, "AuthoCast — a mobility-compliant protocol framework for multicast sender authentication," *Security and Communication Networks*, vol. 1, no. 6, pp. 495–509, December 2008, special issue Secure Multimedia Communication.