Multimed Tools Appl (2011) 53:349–370
DOI 10.1007/s11042-010-0500-8

Distributed SIP conference management with autonomously authenticated sources and its application to an H.264 videoconferencing software for mobiles

Thomas C. Schmidt · Gabriel Hege · Matthias Wählisch · Hans L. Cycon · Mark Palkow · Detlev Marpe

Published online: 16 March 2010

Abstract The design of conferencing systems for achieving efficient and flexible communication in a fully distributed, infrastructure-independent fashion is a promising direction, both in terms of research and practical development. In the particular case of video communication, the seamless adaptation to heterogeneous mobile devices poses an additional strong challenge to those seeking for interoperable and easy-to-deploy solutions. In this paper, we make several contributions towards a generic peer-to-peer (P2P) videoconferencing solution that extends into the mobile realm. We describe the essential building blocks for conference management and media distribution that are necessary for a distributed conferencing approach.

T. C. Schmidt (✉) · G. Hege · M. Wählisch
Department Informatik, HAW Hamburg, Berliner Tor 7, 20099 Hamburg, Germany
e-mail: t.schmidt@ieee.org

G. Hege
e-mail: hege@fhtw-berlin.de

M. Wählisch
Inst. für Informatik, Freie Universität Berlin, Takustr. 9, 14195 Berlin, Germany
e-mail: waehlisch@ieee.org

H. L. Cycon
HTW Berlin, FB 1, Allee der Kosmonauten 20-22, 10315 Berlin, Germany
e-mail: h.cycon@htw-berlin.de

M. Palkow
daviko GmbH, Am Borsigturm 40, 13507 Berlin, Germany
e-mail: palkow@daviko.com

D. Marpe
Fraunhofer Institute for Telecommunications—Heinrich Hertz Institute (HHI),
Einsteinufer 37, 10587 Berlin, Germany
e-mail: marpe@hhi.fraunhofer.de

Establishing a distributed SIP conference focus, participants share the conference according to their individually given capabilities and resources in terms of bandwidth and processing power rather than in a centralized and fixed way. Overall concepts and SIP-primitives for such an autonomous organization are presented. Security issues that derive from this decentralized identity management are resolved by so-called *Overlay AuthoCast*, a novel use of cryptographically generated identifiers. Furthermore, this work is dedicated to the development of a software-based H.264 video codec implementation and the specific aspects resulting from tuning such a highly resource-intensive software codec to the given target platform of a standard consumer smartphone.

Keywords SIP-based multimedia conferencing · Distributed conference management · P2P group communication · Sender authentication · Cryptographically generated identifiers · Mobile videoconferencing · Mobile video coding

1 Introduction

Multimedia conferencing is one of the fastest growing application areas in the Internet today, as deployment is accelerated by two driving forces. On the one hand, traditional telephone communication or local H.323 [14] installations are continuously replaced by pure IP-based solutions. On the other hand, consumers discover more and more the appeal of rich media applications that are ready to be performed on their Internet connected devices, no matter whether stationary or mobile. The common expectation is that of an instantaneously available, easy-to-use communication service that offers speech, chat, and with increasing importance, also visual communication, most preferably at no extra costs.

The idea of augmenting voice calls by video has been around for several decades, but only with the flexibility of the Internet a wider deployment of video communication applications beyond classical video conferencing has taken place. As compared to speech and audio, processing of video signals requires much higher resources for an end-to-end system, both in terms of processing power and network transmission capabilities. However, the continuous and rapid evolution of networks and processors have paved the way for such applications to be performed in software on standard Internet connected personal computers. While video coding for handhelds was bound to the domain of proprietary experiments only a few years ago [11], mobile phones and networked consumer portables are now on the verge of delivering sufficient performance for rich multimedia applications and communication, as well.

Videoconferencing though, which requires simultaneous decoding and encoding in real time, still poses a grand challenge to the mobile world. Limited capacity of wireless channels, on the one hand, and high demands on visual quality, on the other hand, require applications to take advantage of the latest video coding technology as, e.g., given by the H.264/AVC video coding standard [13]. 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, which, however, comes at the price of a considerably increased computational complexity. While H.264/AVC decoding software has already been successfully deployed on handheld devices, high

requirements on computational resources still prevented pure software encoders to be implemented in current mobile systems. Though fast hardware-based implementations of H.264/AVC for specific devices are already available, the use of such hardware is mostly restricted to dedicated video services connecting those specific devices through networks of selective operators only [2].

The widely adopted standard for conference call and media management throughout the Internet is given by the Session Initiation Protocol (SIP) [23]. SIP allows for scheduled or ad hoc conference initiation and media negotiations in a device- and location-transparent fashion. While SIP is inherently a peer-to-peer protocol, current multimedia conferencing solutions mostly rely on an infrastructure of central controllers. Peer-assisted group communication solutions have neither been designed, nor has a corresponding authentication and trust management been taken into account.

This paper deals with the question how to enable infrastructure-agnostic ad hoc videoconferencing in a standard-compliant and secure manner by involving both standard PCs and mobile devices. First, we discuss problems of peer-to-peer conferencing and related work in Section 2. In Section 3, we present a peer-to-peer group communication scheme, which performs well for medium-sized conferences and accounts for the heterogeneous nature of mobile and stationary participants. This includes, on the one hand, SIP compliant session signaling with respect to group communication and, on the other hand, efficient serverless media distribution, self-adjusting to the actual network infrastructure support. Further on, we focus on media distribution that is assisted by remote third parties. Such helping peers are needed to overcome blocking middle boxes, but its utilization raises severe issues in authentication and trust. Therefore, Section 4 is dedicated to Overlay AuthoCast, a group and sender authentication scheme based on cryptographic identifiers that allows for authentication on a per packet level. Finally, we introduce our pure software-based solution for real-time video communication on standard smartphones in Section 5. These mobile clients extend a feature-rich conferencing application which—by means of the previously presented scheme—was developed for an infrastructure-compliant use on standard PCs. Conclusions and an outlook are presented in the final Section 6.

2 The mobile peer-to-peer conferencing problem and related work

A mobile peer-to-peer conferencing application faces the grand challenge to remain robust with respect to both, the infrastructure and its overlay peering system. While the infrastructure is likely to provide only restricted support of point-to-point unicast routing, peers may additionally encounter limited resources, churn from joining and leaving of the session, and insufficient mutual trust. The role a user agent is able to attain in a distributed scenario needs to be adaptively determined according to constraints of its device and current network attachment, but also according to infrastructural hindrances as for instance inherited from NATs and firewalls.

In general, a globally distributed conference relies on the presence of at least one globally addressable, sufficiently powerful peer that acts as a conference focus and relay [5]. As there are many scenarios where this assumption remains unaccomplished, assistance of suitable third parties outside the conference may be required. Such need

for uninvolved bystanders raises concerns about incentives, distribution of load and abuse of resources.

In a community of nomadic users like established in Skype [32], incentives may derive from mutual aid at changing occasions which, in a joint act of cooperation, leads to an ubiquitous operability of the communication system. Users that can provide assistance may do so on the promise to receive the same services whenever they need it. However, solutions that dynamically adapt to load conditions and allow for traffic authentication under the individual control of each peer still await a design. In the following sections, we will introduce a SIP-compliant scheme for distributing such a conference focus and subsequently derive a method to allow for individual packet authentication at focus peers.

Traditional conferencing architectures have been first designed in H.323 [14], and rely on a central multipoint control unit (MCU) provided within the infrastructure. Up until now, the IETF has taken up paths to conferencing with SIP in two distinct directions. On the main trail targeting at tightly coupled conferences, it followed the centralized approach of a single, powerful conference controller. This controller or focus identifies a conference and is addressed by the conference URI. Alternatively, a standard design for loosely coupled conferences was formulated based on multicast, which does neither foresee a mutual awareness of conference members, nor initial SDP [22] negotiations. Several proposals using source-specific multicast (SSM) in tightly coupled scenarios that include SDP negotiations have been contributed from the perspective of a wider SSM deployment in the near future, cf. [25]. Multicast deployment continues to remain hesitant, though, and its extensions into the mobile world are complex and only gradually taken up by IETF working groups [27].

Sharing conference control among distributed entities is equivalent to splitting the conference focus. Little work has been accomplished in this direction, as it bears an inherent complexity: The concept of tightly coupled SIP conferences ranks around a unique conference URI which serves as a routable locator for the focus. Splitting this focus poses the requirement of defining a meaningful mapping from the conference URI to the group of focus instances. Cho et al. [6] have defined such a mapping in form of a focus hierarchy. A primary focus represents the conference URI and serves as initial contact, as well as a load dispatcher. The concept of replicating conference focuses has been recently also brought into IETF [21]. However, group conferencing in these approaches remains bound to a central entity, and thus does not comply to scalability and robustness constraints of a pure peer-to-peer paradigm. Aside from conferencing, there are strong activities in the P2PSIP working group to move SIP proxy functions to a structured peer-to-peer layer [5, 15]. A distributed hash table is envisaged to aid user location and point-to-point session management. This early work has not touched the more intricate topic of group communication at the present time.

In conferencing systems, new parties are commonly authorized for joining the session by off-line credentials or a manual admission through established members. Network access authentication in heterogeneous mobile environments has also been addressed in recent work [10]. 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, the group communication of media 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 identifiers, and conference members would unwillingly assist in amplifying the unwanted traffic. This becomes even more severe in the context of P2P overlay networks, as third parties, which are not able to decrypt or authorize content, may be needed for content replication. A comprehensive overview of corresponding security issues is given in [7].

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 call handling 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) [1], cryptographic identifiers have been transferred to SIP URIs [30], as well as overlay addressing [4] and do not conflict with current KBR implementations such as Chord or Pastry. 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.

Session initiation can be authenticated by appropriately applying SIP CGIs [30], while CGAs have been recently used in AuthoCast [26] to derive a generic framework for mobile multicast source authentication in IP. Based on cryptographic identifiers and passport packets, we will extend this scheme in Section 4, such that any overlay peer is enabled to verify the origin of data prior to forwarding and to repel its misuse. Dynamic ingress filtering and individually established gradual trust will optimize a lightweight protection of the distribution system in structured overlays.

3 Distributed conference management with SIP

In this section, we describe extended SIP operations for distributing a conference focus dynamically according to user needs. Our objective lies in simple, flexible, and cost-efficient ad hoc conferencing functions, which scale appropriately well, but avoid any infrastructure assistance. Such a solution requires group session management and media distribution at peers, which for the sake of infrastructure compliance we arrange concordantly. We rely on the group conferencing primitives in SIP, cf. [16, 17, 25]. To facilitate dialog-oriented scenarios, we purposefully restrict our solution space to conferences among mutually aware parties that in particular include SDP negotiations.

Traditional architectures for tightly coupled conferences rely on a single focus entity which is addressable by a globally routable conference URI. This URI attains the simultaneous role of an identifier of the conference and of its locator. The major task in distributing this focus lies in splitting the functions of identifier and locator. Even though all conference members are required to logically join the same conference, they physically attach to different instances of the focus located at distinct peers. We solve this identifier-locator splitting with the help of source routing options available in SIP.

With a distributed conference control, the requirement for a consistent view of conference states arises. All focus instances need to possess an identical view about conference members. We solve this by instructing focus points to mutually subscribe to the conferencing event state package [24]. Changes at one focus instance will automatically trigger updates at all other instances using the NOTIFY method.

Finally, user agent peers are exposed to severe restrictions in real-world deployments. Often they are located behind NATs and firewalls with network capacities confined to asymmetric DSL or wireless links. Realistically, spontaneous conferences may occur between restricted peers and require the assistance of third parties that are not members of the same session. Assuming the model of incentives discussed in Section 2, initiating parties will request aid from unrestricted peers of its choice. There are many, well deployed ways to obtain such peer lists [31], its details remain beyond the scope of this paper.

3.1 P2P adaptive architecture

In a simplified scenario, clients may be divided into two groups, distinguished by their ability to act as a SIP conference focus or not. A focus must be globally addressable and have access to necessary processing and network resources.

This elementary adaptation scheme can be based on individual decisions of user agents and gives rise to a hybrid architecture of super peers representing potential focus nodes, and remaining leaf nodes. To decide on its potential role of building a focus, a client at first needs to determine NATs and firewalls. Aside from address evaluation, this is done by a simple probe packet exchange. As the implementation is CPU-type aware, processing restrictions are easily evaluated, as well. However, an a priori judgement on available network bandwidth cannot be easily obtained. An evaluation of the local link capacity is frequently misleading, as wireless devices may be located behind wired transmitters of lower, asymmetric capacity such as in ADSL. Current experiments to quickly retrieve reasonable estimates of up- and downstream capacity are ongoing on the basis of variable packet size, unintrusive estimators, cf. [20]. Note that network capacity detection is of vital use for temporal adaptation of the video codecs, as well.

The initiator of a conference either forms a focus itself, or it identifies an appropriate peer among the callees or helper peer list. The first focus of a conference takes ownership of the conference URI and accepts as many leaf nodes as it is willing to serve. On the occasion of overbooking, the super peer decides to split up the focus function and delegates requests to additional super peers acting under the same conference URI. Using the identical SIP primitives as described in the following section, a super peer may decide to leave the conference and hand over leaves to a neighboring focus.

Aside from conference management, super peers provide global connectivity among each other and NAT traversal assistance to leaves, while leaf nodes experience super peers in different roles: A leaf node sees its next hop super peer as the conference focus, while the remote super peers act as proxies on the path to the leaves behind. This set-up corresponds to the well known architecture of Gnutella 0.6 and successive hybrid unstructured peer-to-peer systems, cf. [31]. Despite this architectural analogy, a routing layer for standard-compliant real time group applications should follow a more efficient design and will be presented in the following sections.

3.2 SIP representation

To explore the distributed conferencing scenario in detail, consider the arrival of a new member that exhausts the service capabilities of the current focus. Such party may arrive by a direct call to any conference member, or, as shown in Fig. 1, by a third party invite from some participant. The latter operations are compliant with conference-unaware user agents. The request arrives at the established focus (Charlie), which is fully booked and in turn refers the call to a potential second focus (Snoopy). Reference is done to the conference URI (`hypnotic-talks@...` held by Charlie), using a Route header to direct the request to Snoopy.

```
REFER sip:hypnotic-talks@my-focus.circles.com SIP/2.0
Route: sip:snoopy@dog.net
...
CSeq: 9380 REFER
Refer-To: <sip:lucy@psychic.org>
Content-Length: 0
```

Snoopy intercepts the message, accepts the call reference, and both focuses mutually subscribe to their conference event states. As instructed, the newly established conference focus INVITEs the arriving party (Lucy), adding a Record Route header to its Contact field that carries the conference URI.

```
INVITE sip:lucy@psychic.org SIP/2.0
...
CSeq: 1199 INVITE
Contact: <sip:hypnotic-talks@my-focus.circles.com>; isfocus
Record Route: <sip:snoopy@dog.net>
Content-Type: application/sdp
```

In further communication, Lucy will use this Contact address including the Record Route option to contact its focus, which will guide messages to Snoopy.

Next let us consider a caller that is willing to join a conference with already distributed focus. It may do so by contacting any established member, which will route the INVITE to the nearest available focus. Alternatively, a user agent may be aware of the conference URI and directly submit a call as shown in Fig. 2. The (primary) focus (Charlie) will accept the call either for permanent service, or it will refer the caller to another, less loaded focus. In Fig. 2, a decision is to transfer the arriving party (Lucy) to an alternate focus (Snoopy). The accepting focus then issues a re-INVITE into the call, using a Record Route option in its Contact field as described above.

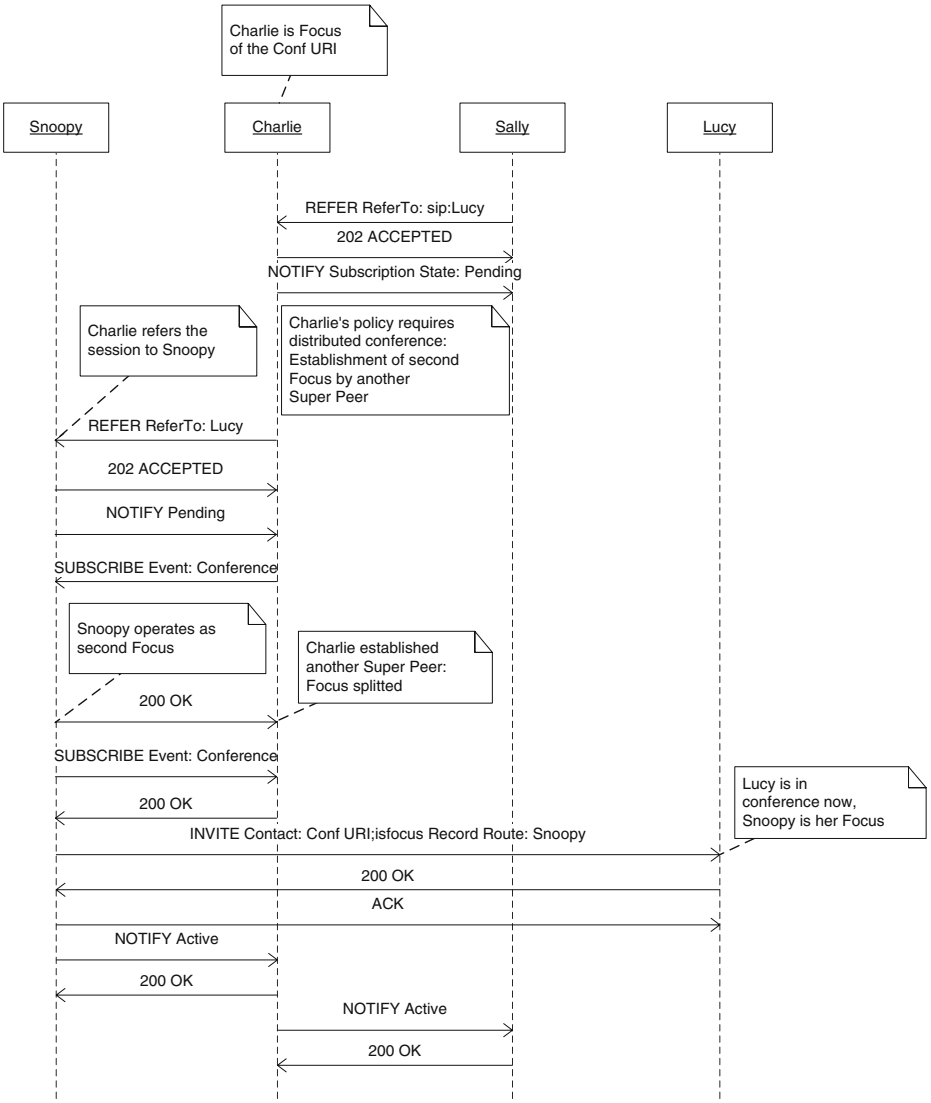


Fig. 1 Split of the primary conference focus following a third party invite

Following this line of conference management, a balanced distribution of leaves to super peers is attained only by standard-compliant SIP signaling. Focus peers can re-arrange and leave according to their needs. Note that SDP media negotiations have been part of the initial arrival steps for each party. Media distribution will naturally follow the paths of the established routing topology, where super peers act as two-sided media controllers: They can combine media streams arriving from their attached leaf nodes before peering them within the focus mesh, but withhold media arriving from neighboring super peers, which are not needed at clients in order to assure a lean transmission to lightweight or mobile leaves.

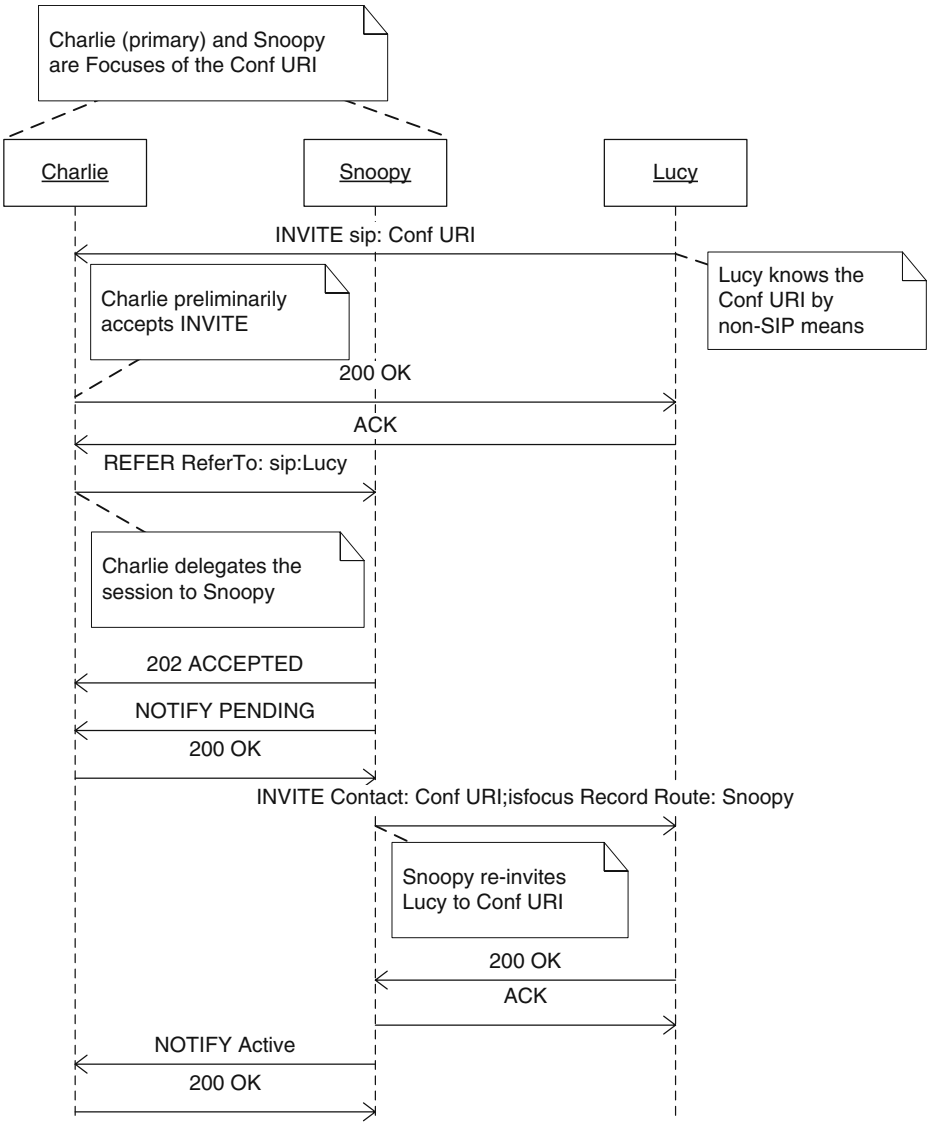
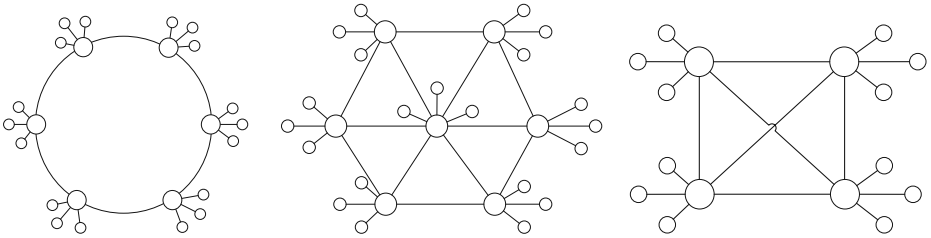


Fig. 2 Joining a conference of previously distributed focus nodes

3.3 Routing design

Routing will be performed between focus peers and is essentially open to implementation. However, its design will admit critical impact on scalability, application performance, as well as forwarding and maintenance load of the super peers. The three characteristic topologies for routing between N super peers as displayed in Fig. 3 explore the problem space: On the one extreme, routing on a ring will minimize neighbor states and forwarding load of each peer, but requires $\mathcal{O}(N)$ hops and thus induces large, varying delays. A full mesh, on the other extreme, places the burden of



(a) Ring (b) Polygon (c) Full mesh

Fig. 3 Peer-to-peer routing topologies on the overlay

$N-1$ neighbor states to be fed in replicated forwarding, but guarantees a rigid 3-hop forwarding limit and minimal delays. A polygonal mesh keeps replication load constant, but dependent on its dimension d, while its corresponding path lengths grow as $\mathcal{O}(\sqrt[d]{N})$. Forwarding on a polygonal mesh will require routing intelligence, which is neither needed on a ring nor in a full mesh topology. Alternatively, forwarding may be established based on some distributed hash table, which essentially scale logarithmically at higher processing costs. As routing paths in our conferencing scenarios are equivalent to the signaling relationships, mesh robustness respectively redundancy of the schemes is equivalent to the number of neighbor states at each peer.

Focusing the problem on moderately sized peer-to-peer conferences of simple and robust nature, a favorable routing scheme is easily identified: The full mesh topology outperforms alternative schemes in forwarding efficiency and robustness, while scaling well up to some hundred nodes. In addition, this scenario is bound to low complexity, since no routing intelligence beyond standard SIP logic of next hop proxying is required. We thus use a full mesh topology here as the favorable approach to mid-size multi-party conversations.

4 Security in distributed conferences: Overlay AuthoCast

Group conferencing with SIP that solely operates among mutually known parties can be protected by the methods reviewed in Section 2. Care should be taken to preserve user friendly names and addresses. 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 [30] 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. Subsequent communication and media streams in particular may be symmetrically encrypted, e.g., using SRTP [3].

As pointed out in Section 2, realistic scenarios require the involvement of third parties which neither have means to authorize members in person, nor share the group keys to verify traffic. In this situation, the abuse of the conference session

and virtual distribution infrastructure can only 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. To achieve this goal, we now generalize the approach of cryptographically generated identifiers to a combined authentication scheme of messaging and group communication among third parties. The concept which was first introduced in [35] is visualized in Fig. 4.

4.1 Single-source authentication

Base Scheme The creator of a group or group controller that has generated its cryptographic overlay ID from a public-private key pair $(\mathcal{K}_{pub}, \mathcal{K}_{sec})$, will use the same $\mathcal{K}_{pub}$ to configure the group address G as a cryptographic identifier. Conflicts within the overlay node ID space, e.g., occurring from identical building rules, can be avoided by adding a counter.

In signing the packets using $\mathcal{K}_{sec}$ and attaching $\mathcal{K}_{pub}$, the group controller will provide cryptographically strong proof of ownership to any receiving peer of the packet: After extracting $\mathcal{K}_{pub}$, an intermediate node can reconstruct source and group address and validate the signature. Having verified that the source is the valid owner of the group, data will be forwarded according to the routing scheme in use. In any case of failure, the forwarder drops the packet, thereby cutting distribution along subsequent branches.

Optimized Scheme Depending on the key length in use, multicast packets may be unreasonably enlarged by the public key piggybacked with data. RSA signature validation in addition is laborious and may not be applicable to every packet traversing. These security overheads can be mitigated by securing multicast forwarding relationships separate from data and offering peers an option to gradually acquire trust in upstream neighbors.

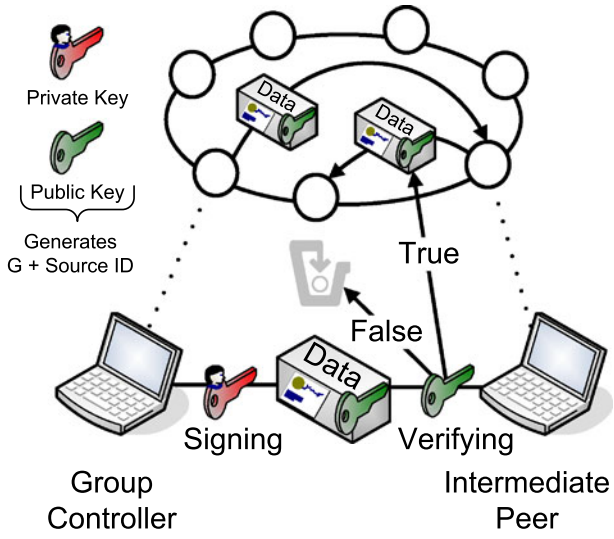


Fig. 4 Packet authentication based on cryptographic group and source identifiers in conference overlays

To establish source-specific authentication at forwarding links throughout the group distribution network, the source initially sends a *passport* packet down the routing paths, once. This signed passport contains the complete CGI extensions, including the public key. Peers store this passport, and augment group forwarding states by $\mathcal{K}_{pub}$, as well as by the verified overlay source address. In the absence of churn and group dynamics, any peer is thus enabled to match a group address to the valid source address *and* to its public key. Subsequent data packets need not carry $\mathcal{K}_{pub}$, but only the signature to allow for authentication. Whenever conference membership changes, or in the presence of mobility or churn, a peer may face new downstream neighbors. To them, it simply forwards the passport packet which will allow for a fully authenticated maintenance of augmented states at newly arriving peers. Note that cached passport packets remain autonomously verifiable and resilient to spoofing.

To further avoid the overhead of signature verification, overlay nodes may simply check for the cached source address. This however will raise the threat of global impersonation. To prevent spoofing, peers can establish ingress filters with respect to the underlay address of their upstream neighbor. In structured overlays, packet forwarding deterministically follows the routing scheme, and upstream neighbors are well defined. Each peer can reliably restrict source-specific traffic to the legitimate upstream forwarder of a group by verifying the address triple of group, source and ingress port. The need for cryptographic signature validation ceases to apply with increasing trust in the upstream forwarder.

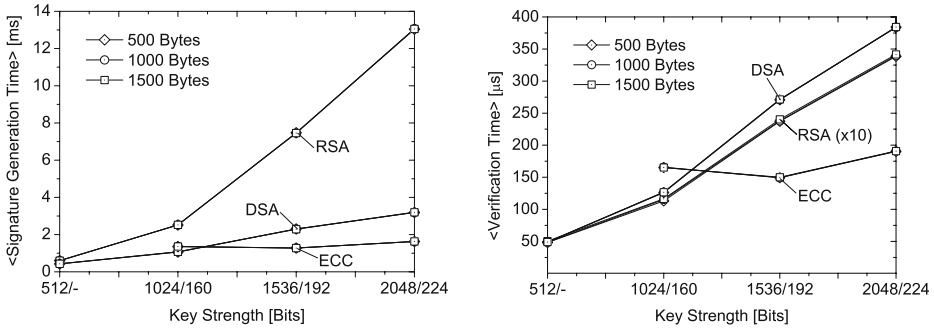
As each peer can detect unwanted traffic from invalid signatures, it can individually decide on a strategy of gradual trust establishment or continued validation. In the presence of overlay routing schemes that allow for multipath transport, a node may even employ this degree of trust for a dynamic path selection.

4.2 Multi-source authentication

The common conference scenario admits multiple senders contributing to the same group, which require admission by the group controller. This admission authority has created the cryptographically generated group address. Before an additional conference source S injects data, it requests a certificate. The group controller authenticates the sender and—according to an application policy—issues the certificate, which includes S, the peer membership of G and an optional lifetime. The certificate is signed with the private key corresponding to the creation of G. A conference source that wants to transmit data attaches this certificate and signs packets with its own private key. An overlay router verifies whether the group certificate is valid and the group address G has been generated from the group public key. Additionally, the router authenticates the source CGI according to the certificate and the peer identifier as described in the single-source case. All optimizations derived from extended state caching and ingress filtering at forwarding peers remain likewise applicable.

4.3 Protocol performance

To quantify the processing overhead of the CGI verification, we have implemented the scheme for CGI signature generation and validation on a standard Linux



(a) Signature Generation (b) Signature Verification

Fig. 5 Processing times for CGI signatures for RSA/DSA-512…2048 and ECC-160…224 with varying packet sizes

platform (2.4 GHz AMD Athlon X2 processor) using the OpenSSL library [8]. For typical ranges of key lengths and packet sizes we measured absolute processing times, and compared RSA and DSA for typical ranges of key lengths (512–2,048 bits) with elliptic curves (ECC) of corresponding strength (secp160r1/secp192r1/secp224r1) [28]. Averages were taken over 50 randomly generated keys, each of which employed to verify 10.000 packets. Results are displayed in Fig. 5. Strikingly, processing costs remain independent of data packet sizes, since the overhead of evaluating SHA-1 hashes is negligible as compared to RSA, DSA or ECC signature processing.

Verification overheads for RSA are in the order of 10–20 µs and appear fully compliant to the overall routing performance attained in overlay networks, while DSA and ECC signature validations show higher expenses by about an order of magnitude. Packet authentication, though, remains well within the bounds of real-time processing.

In the contrary, signature generation poses significantly higher computational efforts on end systems. About 1–10 ms CPU time is required for signing a packet. When applied to every packet of a video stream, this is likely to exceed capacities of weak, i.e., mobile nodes. Consequently, the optimization procedures described above gain enhanced relevance and my be used to limit packet signing to a suitable period for sustaining node-by-node gradual trust relationships.

This fully distributed and autonomously verifiable method remains valid under varying node and forwarding conditions. In particular, it can be equally applied in schemes of multipath transport, or in the presence of mobile conferencing parties.

5 The daViKo videoconferencing software for stationary and mobile participants

In this section, we give an overview of our reference platform, so-called *daViKo* [19], a commercially available software-based audio-visual conferencing system for standard PCs. daViKo has been recently extended to the first H.264/AVC-based video conferencing software solution for mobile phones [9].

The daViKo system is designed as a serverless peer-to-peer IP-based conferencing. At the heart of the system, daViKo contains a highly optimized general-purpose implementation of an H.264/AVC [18] compliant video codec called *DAVC*. DAVC,

as will be shown in this section, turns out to be competitive with other state-of-the-art software-based H.264/AVC conforming real-time implementations in terms of rate-distortion (RD) performance as well as throughput capabilities. In addition, it allows for scalable adaptation of frame rate on a bitstream level.

Audio data is compressed using a 16 kHz speech-optimized variable bit rate codec [33] with extremely short latencies of 40 ms (plus network packet delay). All streams can be transmitted by unicast as well as via multicast protocols. Within the application, audio streams are prioritised over video since user experience is usually more sensitive to losses in audio packets than those of video packets, which both may result from transmission errors or network congestions.

An application-sharing facility is included for collaboration and teleteaching. It enables participants to share or broadcast not only static documents, but also any selected dynamic PC actions like animations including mouse pointer movements. All audio/video streams including dynamic application sharing actions can be recorded on any site. The system is equally well suited to intranet and wireless video conferencing on a best effort basis, since the audio/video quality can be controlled to adapt the data stream to the available bandwidth. The daViKo conferencing system is available for personal computers running MS-Windows or Linux as well as for mobiles or handhelds with Windows Mobile operating system.

5.1 The generic DAVC codec

DAVC, the core of the videoconferencing system, is a fast, highly optimized H.264/AVC implementation. It is based on the Constrained Baseline profile and is optimized for real-time encoding (as well as real-time decoding) by means of a fast motion-estimation strategy including integer-pel diamond search as well as a fast subpel refinement strategy up to $\frac{1}{4}$-pel motion accuracy. Motion estimation includes the choice of several different macroblock (MB) partitions and multiple reference frames, as permitted by the H.264/AVC standard. For choosing between different MB partitions for motion-compensated (i.e., temporal) prediction and MB-based intra (i.e., spatial) prediction modes, a fast rate-distortion (RD) based mode decision algorithm with early termination conditions has been employed.

In comparison to the well-known open source H.264/AVC encoder implementation of x264 [34], our DAVC encoder implementation achieves a similar RD performance and a slight increase in encoding speed when using comparable encoder settings. In Fig. 6, a typical example of such a comparison between x264 and DAVC is shown. In addition to the RD performance of those two real-time encoder implementations, this plot also shows the RD behavior of two non real-time encoder implementations, as given by the H.264/AVC Joint Model (JM) reference software (with Constrained Baseline profile settings) and a MPEG-4 (Part 2) Advanced Simple Profile implementation. The latter two encoders were operated using a high-complexity RD-based mode decision strategy for demonstrating the capabilities of both video coding standards when neglecting any real-time constraints. Figure 6 also contains the number of encoded frames per second (fps) for selected RD points as a measure for maximum encoding speed (e.g., 284 fps for DAVC as compared to 210 fps for x264). Similar results were also achieved for other test sequences.

Note that the DAVC codec also includes an adaptive block-based mechanism for quick recovery from video packet loss.

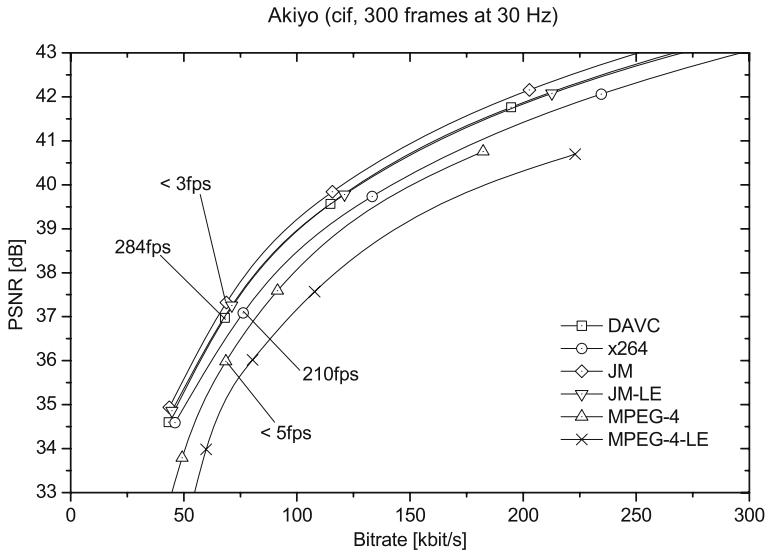


Fig. 6 RD plot for the test sequence "Akiyo" in CIF resolution comparing three different H.264/MPEG-4 AVC encoder implementations as well as a RD-optimized MPEG-4 (Part 2) Advanced Simple Profile implementation

5.2 The lightweight DAVC version for mobiles

The generic DAVC codec, as described in the previous section, has been scaled down and tuned to the specific needs of our target platform of a mobile device. Our tuning includes efficient use of the wireless MMX instruction set available at our current target system. Portability is sustained by an ANSI compliant C version, to be augmented incrementally by platform specific injections.

In order to enable real-time encoding performance, even when appropriately reducing the resolution of the input video to QCIF format, the DAVC codec has been restricted to perform motion estimation only for integer-pel displacements. As shown in Fig. 7 for the Akiyo test sequence in QCIF format, this results in a moderate loss in rate-distortion performance relative to our full DAVC solution. Thus, our mobile-based H.264/AVC video encoder produces acceptable video quality when conforming to 3GPP/UMTS bandwidths constraints.

The full daViKo application together with our lightweight DAVC version was tested on a 520 MHz Xscale processor in an Asus P735 system. On this platform, the Akiyo test sequence was encoded at a rate of 45 fps. In realistic deployment scenarios, the application can reliably and simultaneously encode and decode a QCIF video at 10–15 fps, without CPU exhaustion or frame dropping. QCIF @15 fps is the maximal raw source video rate that can be obtained from the front camera in our test equipment and we expect to arrive at realistic real-time performance at this encoding rate after further optimizations. The maximal battery life time at continuous conferencing, i.e., both encoding and decoding permanently with moderate motion complemented by 802.11 WLAN transmission and activated display was measured to slightly exceed 2 h (Fig. 8).

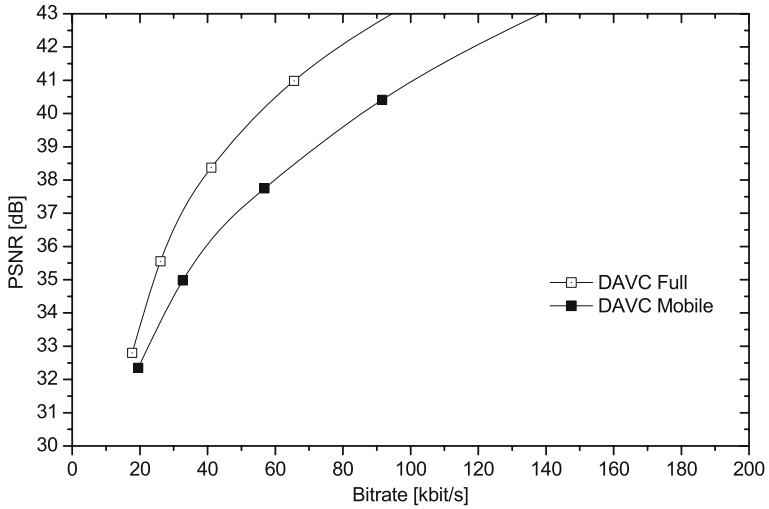


Fig. 7 RD plot for test sequence "Akiyo" in QCIF resolution at 10 fps, comparing the DAVC mobile encoder to the full DAVC encoder implementation. Real-time encoding performance on the tested mobile platform was at 45 fps

Performance values from an empirical test at vivid camera motion are shown in Table 1. Comparison is made between the full DAVC codec running on a standard PC and the mobile DAVC version for handheld and mobile devices. Reduced encoding complexity results in an increased bit rate for the same reconstruction quality relative to the full, i.e., generic DAVC, but the gross total rate for a bidirectional video exchange at 15 fps complies to 3GPP/UMTS bandwidths constraints. Note however that the video source obtained from the front camera of a mobile device is usually significantly more noisy than that from a standard USB camera connected to a PC, which increases complexity of the source and thereby the resulting bit rate.

Fig. 8 The mobile video application

Table 1 Comparison of full DAVC and lightweight DAVC for mobiles at QCIF resolution

	Frame rate (fps)	Bit rate (kbit/s)
Desktop	30	190
Desktop	15	120
Smartphone	15	135

6 Conclusions and outlook

Taking the view that user-centric end-to-end applications will soon extend into the mobile realm, we identified lightweight solutions for a distributed, standard-compliant secure conferencing and media management as major open issues. Its augmentation by video flows raises the additional challenges of a scalable, resource-adaptive media processing.

This paper has addressed several fundamental issues. First, from a generic scheme to perform an identifier-locator split for conference focuses we obtained a transparent way for distributing conference management without changing SIP standard signaling. Conferences thereby gain the abilities of enhanced adaptivity and scalability, but also an increased resilience against infrastructural deficits, node failures and mobility-related changes. Second, we introduced Overlay AuthoCast, an extension of CGI-based host authentication to multicast sources in structured P2P networks. This protocol enables overlay peers to detect unauthorized data independently and on an individual packet level. An efficient caching of authentication credentials, and protected upstream neighbor relations mitigate security overheads, and offer a path to gradual trust establishment at individual peers. Any peer that decides for traffic validation will not only protect itself from unwanted forwarding loads, but will keep subsequent overlay members free of malicious flows. In offering shared benefits, both schemes nicely follow a co-operative P2P paradigm where the incentive offered to the individual enhances the overall system quality.

Finally, the application to a peer-to-peer software for videoconferencing on mobiles was presented that admits utmost flexibility with respect to end systems, operators and network provisioning. This professional system includes a high-quality video codec efficiently adapted to mobile commodity hardware. To the best of our knowledge, and at the time of its initial release (March 2008), this was the first software-based implementation of an H.264/AVC video encoder that operates in real-time on mobile phones.

Our future work will extend in several directions. We are working on adding the scalable video coding extension (SVC) [29] of H.264/AVC to our videoconferencing system. This layered scheme will allow for scaling and selection of partial video streams. Its hierarchical packet organisation should also enable a more efficient stream authentication as well as reduced processing overhead. Distributed conferencing has great potential in linking with recent works on structured overlay multicast. New, highly efficient distribution techniques like BIDIR-SAM [36] are on our agenda, as well as an integration of the work with IETF standard activities. Finally, our conferencing application is well suited to become part of domain-specific applications, e.g., for eLearning or Web Service related systems [12].

Acknowledgements Alexander Knauf provided several experimental implementations of the SIP conference focus splitting. This is gratefully acknowledged. This work has been supported in part by the German Bundesministerium für Bildung und Forschung within the project *Moviecast* (http://moviecast.realmv6.org).

References

1. Aura T (2005) Cryptographically generated addresses (CGA). RFC 3972, IETF
2. Basso A (2006) Beyond 3G video mobile video telephony: the role of 3G-324M in mobile video services. Multimed Tools Appl 28(1):173–185
3. Baugher M, McGrew D, Naslund M, Carrara E, Norrman K (2004) The secure real-time transport protocol (SRTP). RFC 3711, IETF
4. Baumgart I (2007) Peer-to-peer name service (P2PNS). Internet draft—work in progress 00, IETF
5. Bryan D, Matthews P, Shim E, Willis D, Dawkins S (2008) Concepts and terminology for peer to peer SIP. Internet draft—work in progress 02, IETF
6. Cho YH, Jeong MS, Nah JW, Lee WH, Park JT (2005) Policy-based distributed management architecture for large-scale enterprise conferencing service using SIP. IEEE J Sel Areas Commun 23(10):1934–1949
7. Chopra D, Schulzrinne H, Marocco E, Ivov E (2009) Peer-to-peer overlays for real-time communication: security issues and solutions. IEEE Commun Surv Tutor 11(1):4–12
8. Cox M, Engelschall R, Henson S, Laurie B et al (2009) Openssl. http://www.openssl.org
9. Cycon HL, Schmidt TC, Hege G, Wählisch M, Marpe D, Palkow M (2008) Peer-to-peer videoconferencing with H.264 software codec for mobiles. In: Jain R, Kumar M (eds) WoWMoM08—The 9th IEEE international symposium on a world of wireless, mobile and multimedia networks—workshop on mobile video delivery (MoViD). IEEE, Piscataway, pp 1–6
10. Durresi A, Durresi M, Barolli L (2008) Secure authentication in heterogeneous wireless networks. Mob Inf Syst 4(2):119–130
11. Faichney J, Gonzalez R (2001) Video coding for mobile handheld conferencing. Multimed Tools Appl 13(2):165–176
12. Gehlen G, Aijaz F, Zhu Y, Walke B (2007) Mobile P2P web services using SIP. Mob Inf Syst 3(3–4):165–185
13. ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC (2005) Advanced video coding for generic audiovisual services. Tech. rep., ITU (draft version 3)
14. ITU-T Recommendation H.323 (2000) Infrastructure of audio-visual services—systems and terminal equipment for audio-visual services: packet-based multimedia communications systems. Tech. rep., ITU (draft version 4)
15. Jennings C, Lowekamp B, Rescorla E, Baset S, Schulzrinne H (2008) Resource location and discovery (RELOAD). Internet draft—work in progress 00, IETF
16. Johnston A, Levin O (2006) Session Initiation Protocol (SIP) call control—conferencing for user agents. RFC 4579, IETF
17. Mahy R, Sparks R, Rosenberg J, Petrie D, Johnston A (2008) A call control and multi-party usage framework for the Session Initiation Protocol (SIP). Internet draft—work in progress 10, IETF
18. Ostermann J, Bormans J, List P, Marpe D, Narroschke N, Pereira F, Stockhammer T, Wedi T (2004) Video coding with H.264/AVC: tools, performance and complexity. IEEE Circuits Syst Mag 4(1):7–28
19. Palkow M (2009) The daViKo homepage. http://www.daviko.com
20. Prasad R, Dovrolis C, Murray M, kc claffy (2003) Bandwidth estimation: metrics, measurement techniques, and tools. IEEE Netw 17(6):27–35
21. Romano S, Amirante A, Castaldi T, Miniero L, Buono A (2008) Requirements for distributed conferencing. Internet draft—work in progress 04, IETF
22. Rosenberg J, Schulzrinne H (2002) An offer/answer model with Session Description Protocol (SDP). RFC 3264, IETF
23. Rosenberg J, Schulzrinne H, Camarillo G, Johnston A, Peterson J, Sparks R, Handley M, Schooler E (2002) SIP: Session Initiation Protocol. RFC 3261, IETF
24. Rosenberg J, Schulzrinne H, Levin O (2006) A Session Initiation Protocol (SIP) event package for conference state. RFC 4575, IETF
25. Schmidt TC, Wählisch M (2008) Group conference management with SIP. In: Ahson S, Ilyas M (eds) SIP handbook: services, technologies, and security. CRC, Boca Raton, pp 123–158
26. Schmidt TC, Wählisch M, Christ O, Hege G (2008) AuthoCast—a mobility-compliant protocol framework for multicast sender authentication. Secur Commun Netw 1(6):495–509 (special issue on secure multimedia communications)

27. Schmidt TC, Wählisch M, Fairhurst G (2010) Multicast mobility in mobile IP version 6 (MIPv6): problem statement and brief survey, Internet RFC, No 5757
28. Schneier B (1995) Applied cryptography, 2nd edn. Wiley, Hoboken
29. Schwarz H, Marpe D, Wiegand T (2007) Overview of the scalable video coding extension of the H.264/AVC Standard. IEEE Trans Circuits Syst Video Technol 17(9):1103–1120
30. Seedorf J (2006) Using cryptographically generated SIP-URIs to protect the integrity of content in P2P-SIP. In: 3rd annual VoIP security workshop. Berlin, Germany
31. Steinmetz R, Wehrle K (eds) (2005) Peer-to-peer systems and applications. LNCS, vol 3485. Springer, Berlin
32. The Skype homepage (2009) http://www.skype.com
33. The Speex projectpage (2009) http://www.speex.org
34. VideoLan: x264—a free h264/avc encoder (2009) http://www.videolan.org/developers/x264.html
35. Wählisch M, Schmidt TC, Hege G (2009) Overlay authocast: distributed sender authentication in overlay multicast. In: Proceedings of the 28th IEEE INFOCOM. IEEE, Piscataway
36. Wählisch M, Schmidt TC, Wittenburg G (2009) BIDIR-SAM: large-scale content distribution in structured overlay networks. In: Younis M, Chou CT (eds) Proc. of the 34th IEEE conference on local computer networks (LCN). IEEE Computer Society, Los Alamitos, pp 372–375

Thomas C. Schmidt is professor of Computer Networks & Internet Technologies at Hamburg University of Applied Sciences (HAW) and leads the Internet Technologies research group (INET) there. Prior to moving to Hamburg, he headed the computer centre of FHTW Berlin for many years, and continued work as an independent project manager later. He studied mathematics and physics at Freie Universität Berlin and University of Maryland. He has continuously conducted numerous national and international projects. His current interests lie in next generation Internet (IPv6 & beyond), mobile multicast and multimedia networking, as well as XML-based hypermedia information processing. He serves as co-editor and technical expert in many occasions and is actively involved in the work of IETF.

Gabriel Hege is a member of the INET scientific team at Hamburg University of Applied Sciences. He holds a B.Sc. in chemistry from Freie Universität Berlin and is currently pursuing his graduate studies in bioinformatics. Since more than 5 years he has been active in several projects from multimedia networking and involved in the design and development of solutions with real-world deployment. His major fields of interest lie in stack design and security operations of computer networks.

Matthias Wählisch studied computer science and contemporary German literature at Freie Universität Berlin, where he completed his master thesis on structured hybrid multicast routing. He continues his research at the Computer Systems & Telematics group there, and is also with the INET research team at HAW Hamburg. He started professional activities at the networking group of the computer centre of FHTW Berlin while at high school. He is the co-founder of link-lab, a start-up company in the field of next generation networking. His major fields of interest lie in the design and analysis of networking protocols, with a special focus on mobility and group communication in underlay and overlay, where he looks back on 10 years of professional experience in project work and publication.

Hans L. Cycon received his diploma in physics in 1975, his PhD in mathematics in 1979 and his Habilitation in 1984 from the Technical University Berlin, Germany. He taught mathematics and physics at the Telekom Fachhochschule Germany and since 1995, he is a Professor at FHTW Berlin teaching mathematics and signal processing. His publications fields (including a book) are mathematical physics and signal processing i.e., image coding. Hans Cycon is leading several projects in developing wavelet-based still image and video compression codecs. He is member of the German delegation of the ITU/ISO standardisation committee for JPEG2000 still image standard.

Mark Palkow is presently the Managing Director and Chief Developer at the daViKo Gesellschaft für digitale audiovisuelle Kommunikation mbH that was founded in 2000. He received his diploma in communication engineering from the Fachhochschule Telekom Berlin in 1996. Since then he has worked on several research projects at FHTW Berlin, the Old Dominion University Norfolk and Fraunhofer HHI Berlin, all concerned with efficient video encoding, including MPEG-4 and Motion-JPEG2000, as well as multipoint conferencing and collaboration tools.

Detlev Marpe received the Dipl.-Math. degree (with highest honors) from the Technical University Berlin and the Dr.-Ing. degree from the University of Rostock, Germany. As Chief Scientist and Senior Project Manager at Fraunhofer HHI, he is currently responsible for research projects focused on the development of advanced video coding and video transmission technologies. Since 1997, he has been an active contributor to the standardization activities of ITU-T VCEG, ISO/IEC JPEG and ISO/IEC MPEG for still image and video coding. Detlev Marpe received several awards for his scientific and standard contributions including the Emmy Engineering Award in 2008, together with the Joint Video Team of ITU/ISO/IEC as a key contributor and co-editor of the H.264 | MPEG-4 AVC standard. He is also co-founder of daViKo GmbH.