Non-Repudiation in Internet Telephony

Nicolai Kuntze[1], Andreas U. Schmidt[1], and Christian Hett[2]

[1] Fraunhofer–Institute for Secure Information Technology SIT
Rheinstraße 75, 64295 Darmstadt, Germany
{andreas.u.schmidt,nicolai.kuntze}@sit.fraunhofer.de
[2] ARTEC Computer GmbH
Robert-Bosch Straße 38, 61184 Karben, Germany
christian.hett@artec-it.de

Abstract. We present a concept to achieve non-repudiation for natural language conversations over the Internet. The method rests on chained electronic signatures applied to pieces of packet-based, digital, voice communication. It establishes the integrity and authenticity of the bidirectional data stream and its temporal sequence and thus the security context of a conversation. The concept is close to the protocols for Voice over the Internet (VoIP), provides a high level of inherent security, and extends naturally to multilateral non-repudiation, e.g., for conferences. Signatures over conversations can become true declarations of will in analogy to electronically signed, digital documents. This enables binding verbal contracts, in principle between unacquainted speakers, and in particular without witnesses. A reference implementation of a secure VoIP archive is exists.

1 Introduction

The latest successful example for the ever ongoing convergence of information technologies is Internet based telephony, transporting voice over the Internet protocol (VoIP). Analysts estimate a rate of growth in a range of 20% to 45% annually, expecting that VoIP will carry more than fifty percent of business voice traffic (UK) in a few years [1]. The success of VoIP will not be limited to cable networks, convergent speech and data transmission will affect next generation mobile networks as well. The new technology raises some security issues. For eavesdropping traditional, switched analogue or digital phone calls, an attacker needs physical access to the transport medium. Digital networks are generally more amenable to attacks, as holds already for ISDN and to a yet greater extent for IP networks. Efforts to add security features to VoIP products are generally insufficient, though proposals exist for the protection of confidentiality and privacy. Secure VoIP protocols, using cryptographic protection of a call, would even be at an advantage compared to traditional telephony systems. Protocols like SRTP [2] can provide end-to-end security to phone calls, independently of the security of transport medium and communication provider.

With VoIP maturing, it becomes natural to ask for application-level security in the context of IP telephony. Our purpose is to achieve non-repudiation in this context, i.e., for speech over packet-oriented, digital channels, and in particular for VoIP conversations. This means the capability to produce tenable evidence

that a conversation with the alleged contents has taken place between two or more parties. Ancillary information, e.g., that the conversation partners have designated, personal identities, and the time at which the conversation has taken place, may be of utmost importance in this regard, either to establish a supporting plausibility, e.g., 'caller was not absent during the alleged call', or to express relevant semantic information, e.g., 'telephonic order came in before stock price rose'. For electronic documents this kind of non-repudiation is commonly achieved by applying electronic signatures based on asymmetric cryptography. In the communication between several parties, the desired result is a binding contract, and in analogy the central goal of the present contribution is a technology to establish binding verbal contracts without witnesses.

This subject has a long pre-history: As early as 1905, Edison proposed the recording of voice, which was patented 1911 [3]. With the advent of digital signature technology, Merkle [4] envisioned, referring to Diffie and Hellman that "Digital signatures promise to revolutionize business by phone". However, work on non-repudiation of digital voice communication is scarce. The work most closely related to ours is the proposal in [5], resting on the theory of contracts and multi-lateral security [6]. It comprises a trusted third party ('Tele-Witness') that is invoked by communicating parties to securely record conversations and make them available as evidence at any later point in time.

Non-repudiation of inter-personal communication is interesting because of its inherent evidentiary value, exposed by forensic evaluation of the contained biometric data, e.g., as an independent means of speaker identification [7, 8]. Methods for the latter are advanced [9], yielding to recorded voice a high probative force, e.g., in a court of law. In comparison to other media, specific features of voice contribute to non-repudiation. Voice communication is interactive [10] and enables partners to make further enquiries in case of insufficient understanding. This mitigates to some extent problems to which signed digital documents are prone, e.g., misinterpretations due to misrepresentation, lack of uniqueness of presentation, and inadvertent or malicious hiding of content [11].

We set out requirements for non-repudiation which are very particular in the case of VoIP and other multi-media communication over IP, in Section 2 and propose the method to meet them in Section 3. Section 4 analyses the security of the method by listing and assessing the auditable information secured by it. A section which describes the implementation of a secure VoIP archive could, unfortunately, not be included for space restrictions. It may be found in [27] and at `http://arxiv.org/abs/cs.CR/0701145`. Conclusions and an outlook are found in Section 5.

2 Requirements for non-repudiation of conversations

From the schematic characterisation of non-repudiation in the standards [12, 13], we focus on the secure creation of evidence for later forensic inspection. This overlaps with the basic information security targets integrity and availability of the well-known CIA triad. To account for the particularities of the channel, we here take a communication-theoretical approach to derive require-

ments for non-repudiation. The general characteristics of the class of electronic communication that we address are the same for a wide media range, comprising audio, video, and multi-media. In essence it is always a full duplex or multiplex channel operating in real time using data packets, and we subsume communication over those under the term *conversation*. Generic requirements for the non-repudiation of conversations can be profiled for specific media, and we sometimes exemplary allude to the case of speech and VoIP. They are grouped around the top level protection targets *congruence* and *cohesion*. We describe the latter and devise for each a minimal set of specific, but application- and technology-neutral requirements. The requirements are necessary preconditions to achieve the protection targets, and are ordered by ascending complexity.

T1 Congruence. Communication theory and linguistics have established that the attributions of meanings can vary between a sender and a receiver of a message [14, Chapter 6], [15] — a basic problem for non-repudiation. Apart from the ambiguity of language, this implies particular problems for electronic communication channels and media. For digital documents bearing electronic signatures, the presentation problem is addressed by invoking the 'What You See is What You Sign' (WYSIWYS [11]) principle. It is often tacitly assumed that presentation environments can be brought into agreement for sender and receiver of a signed document [16]. We term this fundamental target *'congruence'*. It has special traits in the case of telephony. Essential for non-repudiation is the receiver's understanding, which leads in analogy to the principle 'What Is Heard Is What Is Signed'. But additionally it is indispensable to assure senders (speakers) about what precisely was received (heard).

R1.1 Integrity of the data in transmission, including technical environments for sending and receiving them. For VoIP, this is to be addressed at the level of single RTP packets and their payloads *and* of an entire conversation.

R1.2 Treatment of losses in the channel must enable information of senders about actually received information. This is independent of methods for *avoidance* or *compensation* of losses, such as Packet Loss Concealment (PLC). Rather it means a secure detection of losses (enabled by fulfilled R1.1), enabling a proper handling on the application level as well as a later (forensic) inspection.

R1.3 User interaction policies and their enforcement finally use fulfilled R1.1 and 1.2 to ensure congruence in the inter-personal conversation. For electronic documents this can simply amount to prescriptions about the technical environments in which a electronically signed document must be displayed. Or it can be an involved scheme to guarantee the agreement of contents of documents undergoing complex transformations [17, 18], e.g., between data formats. For speech, it can be realised in various ways taking into account the interactive nature of the medium. This is elaborated on in Section 3.5.

T2 Cohesion regards the temporal dimension of conversations. It means in particular the protection and preservation of the sequence the information flows in all directions of the channel. Again this is at variance with signed documents, where temporal sequence of communication is immaterial. Cohesion means to

establish a complete temporal context of a conversation usually even *in absolute time*, since the temporal reference frame of a conversation can be meaningful.

R2.1 Start times of conversations must be determined and recorded. This is analogous to the signing time of documents (the assignment of which is a requirement for qualified signatures according to the EU Signature Directive).

R2.2 Temporal sequencing of conversations must be protected and related to the reference time frame established by fulfilling R2.1.

R2.3 Continual authentication of communication devices and if possible even communication partners is necessary, e.g., to prevent hijacking.

R2.4 Determined break points must allow for non-repudiation of conversations until they are terminated intentionally or inadvertently.

From the requirements analysis it is apparent that congruence and cohesion are complementary but not orthogonal categories. A specific profile for VoIP is not formulated here for brevity, but rather included in the development of the method below. It is understood that additionally the known standard requirements for electronic signatures as declarations of will and for non-repudiation of electronically signed documents, which are rooted in the theory of multilateral security [19], must be taken into account. We do not address details of user authentication, consent to recording, general privacy, confidentiality, and interaction with respect to the signing as a declaration of will proper. Nonetheless, the method proposed below enables the secure recording and archiving to preserve the probative value of a conversation, as demonstrated in [27].

3 The method

The requirements (R2.4) entail that signing a entire conversation with a single RSA signature by A is not viable, since this yields full disposal to determine (maliciously) the end time of signing of a conversation, and deprives B of any possibility to control and verify this *during conversation*. The opposite approach to secure single packets does not assure cohesion (R2.2 in conjunction with R1.1), since single RTP-packets contain only little audio data which may then easily be reordered. Apart from that, it would be computationally expensive. This is the prime motivation for the method we now present in general for the case of a bilateral conversation between A and B, using, e.g., the SIP/RTP protocol combination [20, 21]. In a basic model A secures the conversation as an unilateral declaration of will. We proceed in a bottom-up fashion from the base concept of intervals of VoIP data, over securing their integrity by a cryptographic chain, to coping with inevitable packet loss. For later reference we call the technique presented in 3.1— 3.4 below the *interval-chaining* method.

3.1 Building intervals

Intervals are the logical units on which the protection method operates. Intervals span certain amounts, which may be nil, of RTP packets for only one direction. As bi-directional communication needs formation of intervals for both directions, A and B hold buffers for packets both sent and received. Since directions are handled differently w.r.t. packet loss, as described in Section 3.3,

directionally homogeneous intervals are advantageous from a protocol design viewpoint. To resolve the full duplex audio stream into an interval sequence we determine that intervals in the directions from and to A alternate. Intervals are enumerated as I_{2k-1}, I_{2k}, $k = 1, \ldots, N$ for directions $A \to B$ and $B \to A$, respectively. Interval I_l comprises RTP packets $(p_{l,j})$, $j = 1, \ldots, K_l$, sent or received by A. For the moment we assume that there is no packet loss.

The length of an interval (in appropriate units) is a main adjustable parameter, and an important degree of freedom. Adjustable sizes of, e.g., data frames are not very common in communication technology, but recent proposals [23] show that they can be advantageous in certain situations, like the present one. We determine that interval boundaries are triggered by the elapse of a certain time, called *interval duration* and denoted by D. If T is the duration of the conversation then $N = \lceil T/D \rceil$. Basing intervals on time necessitates the formation of intervals without voice data payload when a silence period exceeds D. This design choice entails some signalling, transport, and cryptographic overhead. This is however outweighed by some favourable properties. In particular, the maximum buffer length is known from the outset, and control of the interval duration is a direct means to cope with the (known) slowness of (public key) cryptographic soft- and hardware. Adjustment of D therefore allows for an, even dynamical, trade-off between security and performance, as it controls the ratio of security data to payload data. The alternative of triggering intervals by full-run of packet buffers at both sides causes concurrency problems.

Since the communication channel is fully duplex, the sequence of intervals does not reflect the temporal sequence of audio data, rather I_{2k-1} and I_{2k} comprise approximately concurrent data sent in both directions. But this is immaterial since intervals are only logical units and security data for intervals can be stored separately from the RTP streams. This is a key feature of our method. It does not affect the VoIP communication at all but can be run in complete — logical and even physical (extra hardware) — separation from it. VoIP communication is therefore not impeded by our method.

3.2 Cryptographic chaining

The basic idea is to cryptographically secure the payload contained in each interval and include the generated security data in the subsequent interval to form a cryptographic chain. We use the shorthand $(\cdot)_X \stackrel{\text{def}}{=} \text{Priv}_X(h(\cdot))$ for entity X' digital signature by applying a private key Priv_X and a hash algorithm $h(\cdot)$. TS is a time-stamping authority. The notation $\longrightarrow$ signifies the sending of some data. To sign a conversation A performs the following operations.

$$
\begin{aligned}
\text{Sec}_I\colon\ & M_I \stackrel{\text{def}}{=} (D, \mathsf{SIP\_Data}, \mathsf{Auth\_Data}, \mathsf{nonce}, \ldots) \longrightarrow B;\\
& S_0 \stackrel{\text{def}}{=} \big((M_I)_A\big)_{TS} \longrightarrow B;\\
\text{Sec}_l\colon\ & S_l \stackrel{\text{def}}{=} (I_l, S_{l-1})_A \longrightarrow B; \quad l = 1, \ldots, 2N\\
\text{Sec}_F\colon\ & M_F \stackrel{\text{def}}{=} (\mathsf{termination\_condition}, \ldots) \longrightarrow B;\\
& S_F \stackrel{\text{def}}{=} \big((M_F, S_{2N})_A\big)_{TS} \longrightarrow B;
\end{aligned}
$$

In the initial step Sec_I, $(\cdot)_{TS}$ means a time-stamp applied by TS, e.g., according to RFC 3161 [22], and is enveloping the meta-data M_I signed by A (R2.1). This may include some authentication data Auth_Data , e.g., A's digital certificates. To provide a broad audit trail for later inspection, data from the call negotiation and connection establishment, here subsumed under SIP_Data, should be included. The final time-stamp can be used optionally to detect drift, and narrows down the conversation in time. Since this is sufficient to secure the temporal context required for cohesion, the application of time-stamps in every step, which may be costly, is not proposed. A nonce is included in M_I to prevent replay attacks. By including S_{l-1} in the signed data S_l and S_{2N-1} in S_F, and alternation of interval directions, R1.1 and R2.2 are satisfied. Signatures of A and additional authentication data in M_I support R2.3. If communication breaks inadvertently, interval chaining is verifiable up to the last interval, thus R2.4 is satisfied, with a loss of at most one interval duration of conversation at its end. A controls interval timing and the operations Sec_I, Sec_l, and Sec_F occur at times 0, $\lfloor l/2 \rfloor \cdot D$, and $N \cdot D$, respectively.

3.3 Treatment of packet loss

Digital voice communication offers a rather high reliability leading generally to a higher understandability of VoIP communication in comparison with all predecessors. However, packet loss may occur and must be treated as explained in R1.2. Denote by $\delta_l \subset \{1, \ldots, K_l\}$ the sequence of identifiers of packets actually received by A respectively B. Intervals are reduced accordingly to $I'_l \stackrel{\text{def}}{=} (p_{l,j})_{j \in \delta_l}$. The steps Sec_l are modified by a protocol to report received packages.

Sec'_{2k-1}: repeat
  repeat
    interval_termination $\longrightarrow B$;
  until $\delta_{2k-1} \longrightarrow A$;
until $S_{2k-1} \stackrel{\text{def}}{=} (I'_{2k-1}, S_{2k-2})_A \longrightarrow B$;

Sec'_{2k}: repeat
  $S_{2k} \stackrel{\text{def}}{=} (I'_{2k}, S_{2k-1})_A \longrightarrow B$;
  $\delta_{2k} \longrightarrow B$;
until success;

This accounts for losses in the VoIP (RTP) channel as well as failures in the channel for transmission of signing data. The loop conditions can be evaluated by explicit (Sec'_{2k}) or implicit (Sec'_{2k-1}) acknowledgements by receivers.

3.4 Extension to multilateral conversations

Here we present the simplest way to extend the method above to conference-like situations. Multilateral non-repudiation means mutual agreement about the contents of a conversation between all parties. For implementing it for M participants $A_0, \ldots, A_{M-1}$ a round-robin scheme [24] can be used to produce the required chain of signatures as in Lemma 1. Round-robin is a simple algorithm to distribute the required security data between the participants of the

conference. Other base algorithms of distributed systems like flooding, echo, or broadcast might be used, depending, for instance, on the particular topology of the conference network. During the round, a token is passed from participant to participant, signalling the signer role. If participant A_m carries the token, he waits for time D and buffers packets sent by himself. When A_m terminates the interval a signalling and signing protocol is processed, which, in contrast to the scheme above, only concerns data *sent* by A_m. The numbering of intervals is as follows. In the time span from 0 to D the packets $(p_{m;j})$ sent by A_m are in the interval I_m. The packets emitted by A_m during $[D, 2D]$ are in I_{M+m}, and so on. It is here not feasible to sign merely the packets received by *everyone*, because cumulative packet loss could be too high. Instead, an additional hashing indirection is included and hashes $H_k^\theta \stackrel{\text{def}}{=} (h(p_{k;j}))_{j\in\theta}$ of all packets θ received by at least one person from A_m in interval k are distributed and can be used to check the signature in spite of packet loss. Let δ_k^σ denote the list of packets sent by A_m and received by A_σ in interval k. Set $R_m \stackrel{\text{def}}{=} \{0, .., M-1\}\backslash m$ and let $r \geq 0$ be the round number. In order to account for latencies in reporting of packet loss, computing hashes, and signing, we introduce a parallel offset in the round-robin scheme. In round r participant A_m carrying the token terminates interval with number $\widehat{k}(r,m) \stackrel{\text{def}}{=} rM^2 + (M+1)m + 1$. He secures the set of intervals $\widehat{I}(r,m) \stackrel{\text{def}}{=} \big(\widehat{k}(r,m) - M \cdot \{0, \ldots, M-1\}\big) \cap \mathbb{N}$.

$\text{Sec\_mult}_{r,m}$: $\forall\sigma \in R_m$ do
  repeat
    interval_termination $\longrightarrow A_\sigma$;
  until $(\delta_k^\sigma)_{k\in\widehat{I}(r,m)} \longrightarrow A_m$;
od;
$\theta_k \stackrel{\text{def}}{=} \cup_{\sigma\in R_m}\delta_k^\sigma$ for $k \in \widehat{I}(r,m)$;
$D_{r,m} \stackrel{\text{def}}{=} \big((\delta_k^\sigma)_{\sigma\in R_m}, H_k^{\theta_k}\big)_{k\in\widehat{I}(r,m)}$;
$S_{r,m} \stackrel{\text{def}}{=} (D_{r,m}, S_{\text{pred}(r,m)})_{A_m}$;
$\forall\sigma' \in R_m$ do
  repeat
    $(S_{r,m}, D_{r,m}) \longrightarrow A_{\sigma'}$;
  until success;
od;

The preceding security value $S_{\text{pred}(r,m)}$ bears indices

$$\text{pred}(r,m) = \begin{cases} (r, m-1) & \text{if } m \geq 1; \\ (r-1, M-1) & \text{if } r \geq 1,\ m = 0; \\ I & \text{otherwise,} \end{cases}$$

where I stands for the initialisation interval which can be constructed as in the preceding sections, replacing single sending by broadcast with acknowledgements. The numbering scheme for Intervals and the evolving sequence of

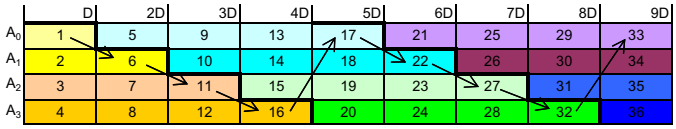

Fig. 1. Numbering of intervals in the case of 4 participants along the time axis. Arrows indicate the sequence of security values S. Thicker borders separate rounds. Equally coloured intervals are secured in a single operation $\text{Sec\_mult}_{r,m}$.

security values is shown in Figure 1 below. In effect, A_m broadcasts (with acknowledgement) a signature over hashes of all packets received by at least one other participant. This is the *common* security data with which the chain can be continued. According to Lemma 1, non-repudiation of the total, multilateral conversation for the first interval duration from time 0 to D is achieved after execution of $\text{Sec\_mult}_{2,M-1}$ at time $2M \cdot D$. With each further execution of Sec_mult a subsequent piece of conversation of length D obtains multilateral non-repudiation.

In case of call termination, $2M+1$ finalisation steps without audio data (two final rounds plus finishing by the participant carrying the token at the time of termination) are required to obtain non-repudiation of the last interval in time. Joining and leaving a signed multilateral call while the signature is created by the participants can be enabled through finalisation. If participant B requests to join the call, A_m, who posses the token, initiates a finalisation and B can join after this (inserted as $m+1$). In the case that a participant likes to leave he awaits the token and finalises including a leave message.

3.5 Operational policies

We do not lay out a complete set of rules for the operation of a system using the non-repudiation method above. Rather we list the most obvious ones and stress the most important point of monitoring and treatment of packet loss, or rather understandability.

To account for requirement R1.3, users should be signalled at any time during a conversation about the signature status of it. This necessitates to an extent specified by application-specific policies the cryptographic verification of the interval chaining, and continual evaluation of relevant information, see Section 4.1. Additionally a *secure voice signing terminal* should control every aspect of user interaction and data transmission. This is elucidated in [25].

To maintain congruence and mitigate attacks aiming at mutilating a conversation, packet loss and the ensuing level of understandability must permanently be monitored. When the packet loss is above a configurable threshold, an action should be triggered according to determined policies. The principle possibilities are: 1. ignore; 2. notify users while continuing signing; 3. abort the signing; and 4. terminate call. The first two options open the path for attacks. Termination of the call is the option for maximum security. From a practical viewpoint, the loss threshold is seldom reached without breakdown of the connection anyway due to insufficient understandability or timeouts.

Options 3 and 4 provide a 'Sollbruchstelle' (predetermined break point) for the probative value of the conversation. In contrast, most other schemes for securing the integrity of streamed data, e.g., the signing method of [26] aim at loss-tolerance, for instance allowing for the verification of the stream signature with some probability in the presence of packet loss. We suggest that for the probative value of conversations, the former is advantageous. A signed call with an intermediate gap can give rise to speculations over alternatives to fill it, which are restricted by syntax and grammar, but can lead to different semantics. Using this, a clever and manipulative attacker could delete parts of the communication to claim with certain credibility that the remnants have another meaning than intended by the communication partner(s). If the contents of a conversation after such an intentional deletion are unverifiable and thus cannot be used to prove anything, this kind of attack is effectively impeded.

4 Security considerations

We corroborate the statement that interval chaining can achieve non-repudiation for VoIP conversations, based on the information generally secured by interval chaining. An analysis based on an instance of a system architecture and possible attacks is contained in [27].

4.1 Auditable information

In this section we analyse the information that can be gained and proved to have integrity in a call secured by interval chaining. Table 1 gives a, perhaps incomplete, overview over this audit data, which may be amenable to forensic inspection, e.g., by an expert witness in court, or, on the other extreme, applicable during the ongoing conversation, or both.

4.2 Comparison with SRTP and IPsec

The well-known security methods SRTP and IPsec address the protection of confidentiality, authenticity and data integrity on the application, respectively network layer, and can be applied to VoIP and as well in parallel with interval-chaining. We want to show salient features of interval-chaining, which distinguishes it from both standards and in our view provides a higher level of non-repudiation and even practicality. On the fundamental level, both SRTP and IPsec necessarily operate on the packet level and do not by themselves provide protection of the temporal sequence and cohesion of a VoIP conversation. While it is true that pertinent information can be reconstructed from the RTP sequence numbers, in turn protected by hash values, such an approach would have some weaknesses, which taken together do not allow full non-repudiation. In particular, RTP sequence numbers can suffer from roll-overs and though their integrity is secured in transmission, they can still be rather easily be forged by the sender, since they belong to protocol stacks which are not especially secured in common systems. While packet loss can be detected or reconstructed using sequence numbers, interval chaining yields a well-defined, tunable, *and cryptographically secured* means to deal with it during an ongoing conversation,

Auditable item	Req.	Protection target	Verifies/indicates	When applicable
Initial time stamp	2.1	Cohesion	Start time	Always
Initial signature & certificate	2.3	Cohesion	Identity of signer	Always
Interval Chaining	2.2, 1.1	Cohesion	Interval integr. & order	Always
Packet loss in intervals	1.2, 2.4	Congruence	QoS, understandability	Always
Monotonic increase of RTP-sequence numbers	1.1, 2.2	Integrity & cohesion	RTP-stream plausibility	Always
Relative drift of RTP-time marks against system time	2.2	Cohesion	RTP-stream plausibility	During convers.
Relative drift of RTP-time marks against $\lfloor l/2 \rfloor \cdot D$	2.2	Cohesion	Packet & stream plausibility	Ex post
No overlaps of RTP-time marks at interval boundaries	2.2	Cohesion	Packet & stream plausibility	Always
Replay-window		Integrity	Uniqueness of recorded audio stream	Always
Final time stamp	2.2	Cohesion	Conversation duration	Ex post
Forensic analysis of recorded conversation		(Semantic) authenticity	Speaker identity, mood, lying, stress, etc.	Ex post, forensic

Table 1. Auditable information of a conversation secured with interval chaining. Columns: Secured data item audited, Non-repudiation requirement addressed, Protection target supported, Actual information indicated or verified, and when is the check applicable.

significantly limiting potential attack vectors. In essence, RTP sequence numbers are not designed to ensure a conversation's integrity and thus have lower evidentiary value in comparison to chained intervals. From the viewpoint of electronic signatures, their level of message, respectively, conversation authentication can only be achieved with an protocol-independent means to manage authentication data such as asymmetric keys, i.e., a Public Key Infrastructure. The connection and session dependent key handling of IPsec and SRTP, relying on HMACs and merely allowing for symmetric keys deprived of authentication semantics, are generally insufficient for non-repudiation. Interval chaining is an independent means to control the cryptographic workload benefiting scalability. Finally, NAT traversal is a problem for network layer integrity protection like IPsec since rewriting IP headers invalidates corresponding hash values (a solution has been proposed by TISPAN [28]). This problem does not occur with the interval chaining method, since only RTP headers, not IP headers of packets need to be (and are in the implementation below) signed.

5 Conclusions

IP-based multimedia communication is not restricted to VoIP, for instance by now, several video conferencing systems are maturing, some of which are based on sophisticated peer-to-peer communication [29]. Moreover the service quality and availability of the new communication channels is constantly increasing through developments like packet loss concealment (PLC) for audio [30] and

even video [31] streams. Our proposed method for non-repudiation is applicable in all these contexts. Its adoption would pave the way for a new paradigm for trustworthy, inter-personal communication. An efficient self-signed archive for VoIP calls and its system architecture was implemented as a prototype together with a verification and playback tool. It requires no modification to the terminal equipment, and secures the ongoing conversations 'on the fly' [27].

The next steps in our research are to i) implement the operational context for electronic signatures over speech, i.e., user interaction and signalling, ii) devise a trustworthy signature terminal for that purpose, preferably using Trusted Computing technology on mobile devices, e.g., to secure audio I/O and processing, iii) extend the method to conferences and other media than VoIP.

References

1. Kavanagh, J.: Voice over IP special report: From dial to click. `http://www.computerweekly.com/Articles/2006/02/14/214129/VoiceoverIPspecialreportFromdialtoclick.html`, visited 1.3.2006.
2. Baugher, M., et al.: The Secure Real-time Transport Protocol (SRTP). RFC 3711, IETF, March 2004. `http://www.ietf.org/rfc/rfc3711.txt`
3. Edison, T.A.: Recording-telephone. United States Patent P.No.:1,012,250, United States Patent Office, Washington, DC (1911) Patented Dec. 19, 1911.
4. Merkle, R.C.: A certified digital signature. In Brassard, G., ed.: Advances in Cryptology (CRYPTO '89). Number 435 in LNCS, Springer-Verlag (1989) 218–238 Republication of the 1979 original.
5. Strasser, M.: Möglichkeiten zur Gestaltung verbindlicher Telekooperation. Master's thesis, Universität Freiburg, Institut für Informatik und Gesellschaft (2001)
6. Kabatnik, M., Keck, D.O., M. Kreutzer, A.Z.: Multilateral security in intelligent networks. In: Proceedings of the IEEE Intelligent Network Workshop. (2000) 59–65
7. Poh, N., Bengio, S.: Noise-Robust Multi-Stream Fusion for Text-Independent Speaker Authentication. In: Proceedings of The Speaker and Language Recognition Workshop (Odyssey). (2004)
8. Rodriguez-Linares, L., Garcia-Mateo, C.: Application of fusion techniques to speaker authentication over IP networks. IEEE Proceedings-Vision Image and Signal Processing **150** (2003) 377–382
9. Hollien, H.: Forensic Voice Identification. Academic Press, London (2001)
10. Goodwin, C.: Conversational organization: Interaction between speakers and hearers. Academic Press, New York (1981)
11. Landrock, P., Pedersen, T.: WYSIWYS? What You See Is What You Sign? Information Security Technical Report, **3** (1998) 55–61
12. ISO: Information Technology: Security Frameworks for Open Systems: Non-Repudiation Framework. Technical Report ISO10181-4, ISO (1997)
13. ISO: Information Technology: Security Techniques - Non Repudiation - Part 1: General. Technical Report ISO13888-1, ISO (1997)
14. Searle, J.R.: Mind, Language and Society. Basic Books, New York (1999)
15. Austin, J.L.: How to Do Things with Words. Harvard University Press, Cambridge, Mass. (1962)
16. Schmidt, A.U.: Signiertes XML und das Präsentationsproblem. Datenschutz und Datensicherheit **24** (2000) 153–158

17. Schmidt, A.U., Loebl, Z.: Legal security for transformations of signed documents: Fundamental concepts. In Chadwick, D., Zhao, G., eds.: EuroPKI 2005. Volume 3545 of Lecture Notes in Computer Science., Springer-Verlag (2005) 255–270
18. Piechalski, J., Schmidt, A.U.: Authorised translations of electronic documents. In Venter, H.S., Eloff, J.H.P., Labuschagne, L., Eloff, M.M., eds.: Proceedings of the ISSA 2006 From Insight to Foresight Conference, Information Security South Africa (ISSA) (2006)
19. Rannenberg, K., Pfitzmann, A., Müller, G.: IT Security and Multilateral Security. In Müller, G., Rannenberg, K., eds.: Multilateral Security in Communications. Volume 3 of Technology, Infrastructure, Economy., Addison-Wesley (1999) 21–29
20. Rosenberg, J., et al.: SIP: Session Initiation Protocol. RFC 3261, IETF, June 2002. http://www.ietf.org/rfc/rfc3261.txt
21. Schulzrinne, H., et al.: RTP: A Transport Protocol for Real-Time Applications. RFC 1889, IETF, January 1996. http://www.ietf.org/rfc/rfc1889.txt
22. Adams, C., et al.: Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP). RFC 3161, IETF, August 2001. http://www.ietf.org/rfc/rfc3161.txt
23. Choi, E.C., Huh, J.D., Kim, K.S., Cho, M.H.: Frame-size adaptive MAC protocol in high-rate wireless personal area networks. ETRI Journal **28** (2006) 660–663
24. Shreedhar, M., Varghese, G.: Efficient fair queuing using deficit round-robin. IEEE/ACM Transactions on Networking **4** (1996) 375–385
25. Hett, Ch., Kuntze, N., Schmidt, A. U.: Security and non repudiation of Voice-over-IP conversations. To appear in: Proceedings of the Wireless World Research Forum (WWRF17), Heidelberg, Germany, 15-17 November 2006.
26. Perrig, A., Tygar, J.D., Song, D., Canetti, R.: Efficient authentication and signing of multicast streams over lossy channels. In: SP '00: Proceedings of the 2000 IEEE Symposium on Security and Privacy, Washington, DC, USA, IEEE Computer Society (2000) 56–75
27. Hett, C., Kuntze, N., Schmidt, A.U.: A secure archive for Voice-over-IP conversations. In et al., D.S., ed.: To appear in the Proceedings of the 3rd Annual VoIP Security Workshop (VSW06), ACM (2006) http://arxiv.org/abs/cs.CR/0606032
28. Telecoms & Internet converged Services & Protocols for Advanced Networks (TISPAN) http://www.tispan.org/, see also the Whitepaper http://www.newport-networks.com/cust-docs/91-IPSec-and-VoIP.pdf
29. Zühlke, M., König, H.: A signaling protocol for small closed dynamic multi-peer groups. In: Proceedings of High Speed Networks and Multimedia Communications, 7th IEEE International Conference (HSNMC 2004), Toulouse, France. Volume 3079 of LNCS., Springer-Verlag (2004) 973–984
30. Perkins, C., Hodson, O., Hardman, V.: A survey of packet loss recovery techniques for streaming audio. IEEE Network **12** (1998) 40–48
31. Zhu, Q.F., Kerofsky, L.: Joint source coding, transport processing, and error concealment for H.323-based packet video. In Aizawa, K., Stevenson, R.L., Zhang, Y.Q., eds.: Visual Communications and Image Processing '99. Volume 3653 of Proceedings of SPIE., SPIE (1998) 52–62
32. Kolletzki, S.: Secure internet banking with privacy enhanced mail - a protocol for reliable exchange of secured order forms. Computer Networks and ISDN Systems **28** (1996) 1891–1899
33. Grimm, R., Ochsenschläger, P.: Binding Cooperation. A Formal Model for Electronic Commerce. Computer Networks **37** (2001) 171–193