Integrität und Nicht-Abstreitbarkeit von VoIP-Kommunikation

Kai-Oliver Detken[1] · Marcel Jahnke[1] · Bernd Röllgen[2]

[1]DECOIT GmbH
{detken | jahnke}@decoit.de

[2]Global IP Telecommunications Ltd.
roellgen@globaliptel.com

Zusammenfassung

Die vollständige Digitalisierung der Nachrichtentechnik [DETK16] erlaubt neue Angriffsszenarien, die mit bisherigen Telefon-Technologien nicht möglich waren, da für das Abhören analoger Telefonate oder Videosignale stets der physikalische Zugang zu dem Transportmedium notwendig war. Bei der internetbasierten Kommunikation, zu der auch Voice-over-IP (VoIP) gehört, die von einer prinzipiell unbegrenzten Anzahl zwischenliegender Knoten aus abgefangen werden kann, ist ein Abhören für einen Angreifer potenziell erheblich einfacher. Auch das Einspeisen unerwünschter Kommunikation (SPAM, SPIT) ist in digitalen Netzen wesentlich leichter zu bewerkstelligen [BCS06]. Eine Verschlüsselung der Verbindung ist dabei nicht als ausreichend zu bewerten, da es ebenfalls relevant ist zu wissen, welche Kommunikationspartner miteinander sprechen und der Inhalt nicht nachträglich abstreitbar sein darf. Deshalb hat sich das Forschungsprojekt INTEGER dem Schutz der Integrität einer VoIP-Kommunikation, die sichere Authentifizierung der Kommunikationspartner und Nicht- Abstreitbarkeit, durch das elektronische Signieren der Kommunikation, zum Ziel gesetzt. Hierfür soll ein Hardware-Vertrauensanker (z.B. ein TPM-Chip der Trusted Computing Group) zur IP-Telefonie angebunden und genutzt werden.

1 Einleitung

Ein Ziel sollte es sein, eine sichere Kommunikation *(Verschlüsselung)* ebenfalls in VoIP-Netzen zu ermöglichen. Neuere Anforderungen legen zusätzlich Wert auf *Integrität* (Fälschungssicherheit) und *Nicht-Abstreitbarkeit* (wer ist der Gesprächspartner und was ist der Inhalt) von internetbasierter multimedialer Kommunikation [HKS06]. Die bei VoIP verwendeten offenen Übertragungsstandards ermöglichen neben der Sprachkommunikation auch weitere Anwendungen. Das Session Initiation Protocol (SIP) [RSC+02] kann beispielsweise auch zum Initiieren der Übertragung von beliebigen Multimedia-Datenströmen dienen. Gerade aber im B2B- und B2C-Bereich fehlt den aktuellen Lösungen der Nachweis über die Vertraulichkeit und die Verlässlichkeit der Kommunikation.

Die von INTEGER angestrebte, völlig neuartige Form der Nicht-Abstreitbarkeit mündlicher Kommunikation ermöglicht grundlegend effizientere Geschäftsabläufe (u.a. als Beweis für mündliche Vertragsabschlüsse) und wir zum Teil bereits heute im Finanzsektor von der Europäischen Union gefordert [EURO14]. Auch im Verhältnis zwischen Unternehmen und Out-

P. Schartner · N. Pohlmann (Hrsg.) · D•A•CH Security 2018 · syssec (2018) 392-403.

sourcing-Dienstleistern, wie zum Beispiel Call-Centern, sind die geplanten Ergebnisse einsetzbar und ermöglichen durch die dargestellte Sicherheit eine verbesserte Arbeitsteilung. Das weitreichende Ziel mündlich geschlossener Verträge zwischen zuvor unbekannten Partnern und ohne Zeugen zu beweisen, stellt zudem ein neues Paradigma in der Digitalisierung der Arbeitswelt dar und eröffnet neue Felder des verbindlichen „Collaborative Commerce".

Aber auch ohne die weitreichenden Ziele aus der Forschung muss VoIP auch dem Telekommunikationsgesetzt (TKG) genüge tun, indem nur einzelne Gespräche gesetzeskonform auf richterliche Anordnung abgehört und aufgezeichnet werden. Dies wird heute von den meisten SIP-Providern aber nicht umgesetzt. Im Gegenteil, die Gespräche werden größtenteils völlig ungeschützt und im Klartext über das Internet übertragen. Ein Abhören ist daher auf Basis standardkonformer Technik auch Einzelnen möglich, die über kein Detailwissen verfügen. Dies sollte zukünftig auf jeden Fall geändert werden und ist auf Basis vorhandener Sicherheitsstandards auch möglich, beinhaltet aber entsprechende Investitionskosten bei den SIP-Providern.

2 VoIP-Szenarien

Voice-over-IP (VoIP) bzw. IP-Telefonie bezeichnet die Telefonie über das Internet oder Computernetzwerke. Für das Führen von Telefongesprächen benötigen Benutzer ein software-basiertes Telefon oder hardware-basierte Telefonapparate mit VoIP-Fähigkeit. Dabei werden für die „Telefonie" typische Informationen, d.h. Sprache und Steuerinformationen für den Verbindungsaufbau, über ein auch für Datenübertragung nutzbares Netz übertragen. VoIP nutzt für den Telefondienst die IP-Infrastruktur, welche durch einen einheitlichen Netzaufbau Kosten reduziert und traditionelle Telefonnetze inzwischen fast vollständig ersetzt.

VoIP kann in sehr unterschiedlichen Szenarien umgesetzt und implementiert werden. Je nachdem, in welchem Szenario es zum Einsatz kommt, müssen auch andere Kriterien an die Sicherheit gestellt werden. Zusätzlich sind eine Vielzahl von Protokollen im VoIP-Umfeld im Einsatz wie SIP, RTP, RTCP, SRTP, ZRTP, H.323, MINET, IAX/IAX2 (Asterisk), MGCP, MEGACO, SCCP (Cisco). Manche Protokolle sind proprietär, und andere arbeiten nach offenen Standards. Die größte Verbreitung besitzt heute allerdings das Session Initiation Protocol (SIP).

Bei dem SIP-Protokoll handelt es sich um das Transportprotokoll für Steuerdaten und wird nahezu von allen Telekommunikationsanbietern zur Realisierung von VoIP genutzt. SIP wurde ursprünglich von der IETF Multi-Party Multimedia Session Control Working Group, bekannt als MMUSIC, entwickelt. Version 1.0 wurde als Internet-Draft im Jahr 1997 eingereicht. Eine mit wesentlichen Änderungen behaftete Version 2.0 wurde 1998 als Internet-Draft nachgereicht und im April 1999 als RFC-2543 veröffentlicht. Die im September 1999 durch die IETF gegründete SIP-Arbeitsgruppe veröffentlichte im Jahr 2000 das heute zugrundeliegende RFC-3261 [RSC+02] und ersetzte damit die ursprüngliche RFC-2543-Spezifikation.

Zwar steigen VoIP-Implementierungen immer stärker an, jedoch ist auch die Migration zur traditionellen Telefonie nach wie vor zu berücksichtigen, da kaum ein Unternehmen auf völlig neuer Infrastruktur aufsetzt. Dementsprechend müssen verbindungsorientierte (Circuit Switching) und verbindungslose Technik (Packet Switching) miteinander effektiv kombiniert werden.

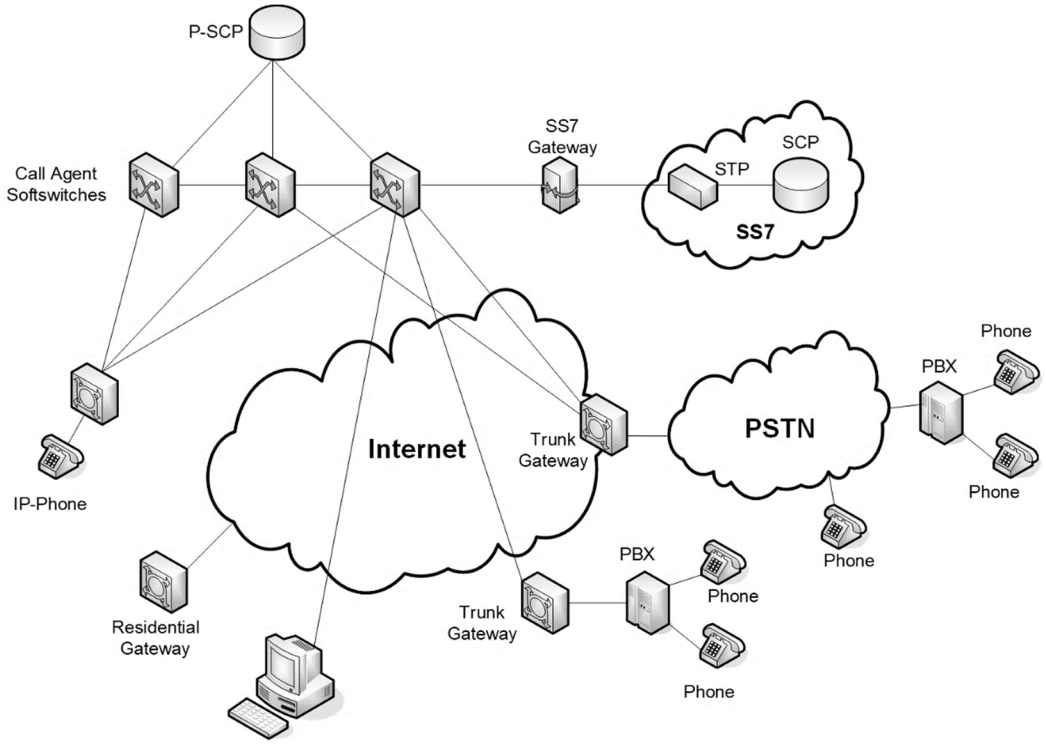

Abb. 1: Beispiel einer VoIP-Architektur

Daher kann man bei VoIP grundsätzlich folgende Szenarien unterscheiden:

- **Campus VoIP:** In einer Campus-VoIP-Umgebung wird eine Nebenstellenanlage auf IP-Basis verwendet, die auch als IP-PBX (Private Branch eXchange) bezeichnet wird. IP-Telefone und/oder Softphones sind mit dieser Nebenstellenanlage verbunden. Der Verbindungsaufbau in das öffentliche Telefonnetz wird über Gateways ermöglicht. Realisiert werden kann dieses System hardware-basiert (aufgerüstete Nebenstellenanlage mit VoIP-Interface) oder software-basiert (Serversystem mit VoIP-Software). Beide Varianten sind schwer von außen zu attackieren, da die Telefongespräche nicht über das Internet oder andere unsichere Netze geführt werden. Potenzielle Attacken müssen daher hauptsächlich aus dem Intranet kommen oder von außerhalb über die Firewall.
- **IP Centrex/Hosted IP:** Diese VoIP-Variante beinhaltet eine virtuelle IP-basierte Nebenstellenanlage, die von einem Provider zur Verfügung gestellt wird. Der Provider ist hierdurch in der Lage, eigene Sprachdienste anzubieten, ohne dass ein Unternehmen eigene Gateways oder PBX-Systeme anschaffen muss. Aus Sicht des Unternehmens muss nur eine ausreichende Internetanbindung vorhanden sein, und IP-Telefone und/oder Softphones müssen angeschafft werden. Attacken auf das VoIP-System können über das Intranet oder das Internet (aus dem Providernetz) erfolgen.
- **VoIP-Trunks:** Dieses Szenario beinhaltet direkte Punkt-zu-Punkt-Verbindungen zwischen verschiedenen Standorten. VoIP-Trunkverbindungen lösen dabei zunehmend herkömmliche verbindungsorientierte Telefonverbindungen ab. Dies liegt an der zunehmenden Konvergenz der Netze und an den sich daraus ableitenden Kostenersparnissen. Auch erhöht sich die Flexibilität, wenn keine T1- oder PRI-Interfaces mehr verwendet werden müssen. Allerdings kann es hierbei auch zu einem höheren Angriffspotenzial kommen, wenn die Übertragung über unsichere Netze realisiert wird. Speziell die Attacken aus dem Internet führen dazu, dass Unternehmensnetze verletzlicher werden.

Abbildung 1 zeigt eine typische VoIP-Architektur, die eine große Flexibilität bzgl. der Skalierbarkeit beinhaltet. In dieser Architektur sind die Gateways an der Netzwerkgrenze angeordnet. Rufverwaltung und -handhabung werden von separaten Systemen (Call Agents, Media Gateway Controller) übernommen, die für die Signalisierung der Sprachverbindungen verantwortlich sind. Wenn ein Call Agent bzw. Media Gateway Controller aus irgendwelchen Gründen nicht verfügbar sein sollte, übernimmt ein anderes System diese Funktion. Zusätzlich wurde die Intelligenz der Signalisierung von einem reinen Hardware-System (Media Gateway) auf ein allgemeines Serversystem übertragen. Durch die zunehmende Abschaltung von ISDN entfällt allerdings zukünftiger immer stärker der Gateway-Übergang in die alte Telefonwelt. [DEER07]

3 Aufbau des VoIP-Protokollstapels

Das Telefonieren mittels IP-Protokoll kann sich für den Teilnehmer genauso darstellen wie in der klassischen Telefonie. Wie bei der herkömmlichen Telefonie teilt sich das Telefongespräch hierbei in drei grundsätzliche Vorgänge auf:

a. Verbindungsaufbau

b. Gesprächsübertragung

c. Verbindungsabbau

Im Unterschied zur klassischen Telefonie werden bei VoIP aber keine „Leitungen“ durchgeschaltet, sondern die Sprache wird in kleinen IP-Paketen transportiert (siehe Abbildung 2).

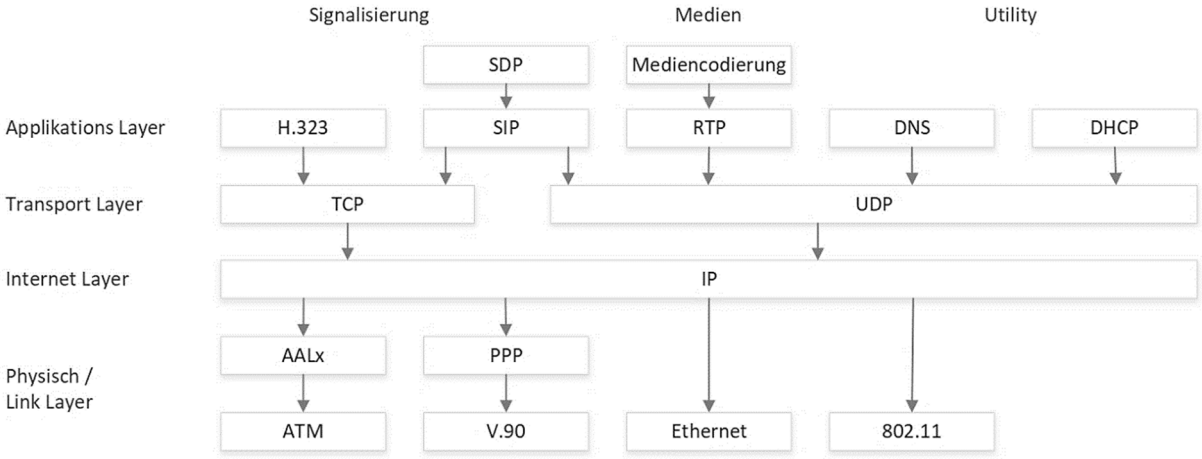

Abb. 2: Protokollstapel bei VoIP-Anwendungen

Der Auf- und Abbau von Calls (Rufsteuerung) erfolgt über ein von der Sprachkommunikation getrenntes Protokoll. Auch die Aushandlung und der Austausch von Parametern für die Sprachübertragung erfolgt über andere Protokolle, als die der Rufsteuerung. Um in einem IP-basierten Netz eine Verbindung zu einem Gesprächspartner herzustellen, muss die aktuelle IP-Adresse des gerufenen Teilnehmers innerhalb des Netzes bekannt sein, jedoch nicht notwendigerweise auf der Seite des Anrufers. Feststehende Anschlüsse wie im herkömmlichen Festnetz (Public Switched Telephone Network – PSTN) gibt es in rein IP-basierten Netzen nicht. Die Erreichbarkeit des Angerufenen wird, ähnlich wie in Mobilfunknetzen, durch eine vorangegangene Authentifizierung des Gerufenen und einer damit verbundenen Bekanntmachung seiner momentanen Adresse ermöglicht.

Aufgrund z.B. von Ortswechsel des Teilnehmers, Wechsel des Teilnehmers am gleichen PC oder die dynamische Adressvergabe beim Aufbau einer Netzwerkverbindung ist eine feste Zuordnung von Telefonnummern zu IP-Adressen nicht möglich. Die allgemein angewandte Lösung besteht darin, dass die IP-Telefonie-Teilnehmer bzw. dessen Endgeräte ihre aktuelle IP-Adresse bei einem Dienstrechner (Registrar-Server) unter einem Benutzernamen hinterlegen. Der Verbindungsrechner für die Rufsteuerung, oder manchmal sogar das Endgerät des Anrufers selbst, kann dann bei diesem Server die aktuelle IP-Adresse des gewünschten Gesprächspartners über den angewählten Benutzernamen erfragen und damit die Verbindung aufbauen.

Durch Nutzung des gleichen Netzes und der damit verbundenen Teilung mit anderen Teilnehmern wird die Sprache ungeschützt übertragen. Zwar besteht die Möglichkeit, die Übertragung zu verschlüsseln, jedoch wird dies häufig von den Anwendern nicht genutzt oder von den Herstellern bzw. Anbietern nicht angeboten. Einerseits liegt dies an fehlenden Implementierungen oder an der Unkenntnis über diese Möglichkeit, andererseits kann eine Verschlüsselung auch die Sprachqualität beeinträchtigen, weshalb sich häufig Anwender zugunsten der Sprachqualität entscheiden.

Der eigentliche Datentransport (siehe Abbildung 2) erfolgt über das Real-Time Transport Protocol (RTP) [SCFJ03], gesteuert durch das Real-Time Control Protocol (RTCP) [HUIT02]. RTP verwendet zur Übertragung in der Regel das User Datagram Protocol (UDP). UDP kommt zum Einsatz, da es ein minimales, verbindungsloses Netzwerkprotokoll ist, das nicht auf Zuverlässigkeit ausgelegt wurde wie beispielsweise das Transmission Control Protocol (TCP). Dies bedeutet, dass der Empfang der Sprachpakete nicht bestätigt wird, also keine Übertragungsgarantie besteht. Der Vorteil von UDP ist aber dessen geringere Latenzzeit gegenüber der von TCP, da nicht auf eine Bestätigung gewartet und fehlerhafte Pakete erneut gesendet werden müssen. Dadurch verzögert sich der Datenfluss insgesamt nicht. Eine vollkommen fehlerfreie Übertragung ist ohnehin nicht nötig, da die gesprochene Sprache eine hohe Redundanz aufweist und heutige Codecs in der Lage sind, Fehler bis zu einer bestimmten Toleranz zu korrigieren. Für ein flüssiges Gespräch ist eine geringe Antwortverzögerung daher wesentlich wichtiger. Die Signalisierung des Gesprächs wird hingegen über SIP vorgenommen.

Da es verschiedene Möglichkeiten gibt, Sprachdaten (oder auch Videodaten) zu codieren, müssen sich die Gesprächsteilnehmer beim Aufbau der Verbindung darauf einigen, welche Codierung sie verwenden wollen. Dazu wird das Session Description Protocol (SDP) verwendet (siehe Abbildung 2). SDP ist ein Unterprotokoll von SIP und beschreibt die zur Verfügung stehenden Codecs. SIP und RTP stellen in der Grundausstattung keine Mechanismen zur Verfügung, um die Kommunikation abzusichern. Eine Authentifizierung des Kommunikationspartners findet, ebenso wie eine Integritätsprüfung, nicht statt. Die Vertraulichkeit kann ebenfalls nicht gewährleistet werden. Dies führt zu Schwachstellen und leichter Abhörbarkeit und kann nur durch den Einsatz der Protokollerweiterungen SIPS und SRTP kompensiert werden.

4 Das INTEGER-Projekt

Ein Ziel des BMWi-Forschungsprojekts INTEGER ist die Integrität und Nicht-Abstreitbarkeit der internetbasierten multimedialen Kommunikation. Als Paradigma dient dem Projekt die IP-basierte Telefonie. Die hier verwendeten, offenen Übertragungsstandards ermöglichen neben der Sprachkommunikation weitere Anwendungen. Die von INTEGER angestrebte, völlig neu-

artige Form der Nicht-Abstreitbarkeit mündlicher Kommunikation ermöglicht grundlegend effizientere Geschäftsabläufe, die auch im Verhältnis zwischen Unternehmen und Outsourcing-Dienstleistern, wie zum Beispiel Call-Centern, einsetzbar sind.

Allgemein kann die multimediale Kommunikation zwischen zwei Parteien als eine Transaktion angesehen werden, die in verschiedenen Kontexten (z.B. bei der Entscheidungsfindung für einen Auftrag) einen hohen Schutzbedarf besitzt und deshalb vollständig geschützt werden muss. Die *Nicht-Abstreitbarkeit* einer Konversation ist hierbei eine grundlegende Voraussetzung [HKS07]. Um diese Voraussetzung zu erreichen, müssen drei generelle Aufgaben mit Hilfe einer digitalen Signatur gelöst werden:

1. **Integritätsschutz der Konversation:** Die Integrität einer digitalen Konversation unterscheidet sich im Vergleich zur Integrität von anderen digitalen Daten in Bezug auf die Bedeutung des zeitlichen Kontexts. Eine digitale Konversation wird in einem direkten zeitlichen Zusammenhang geführt. Aus diesem Grund erfordern insbesondere die Folge der übertragenen Pakete, Verlust von Paketen auf dem Transportweg und der Zeitpunkt der Kommunikation eine besondere Betrachtung.
2. **Authentifizierung der Kommunikationspartner:** Um ein Gespräch auf seine Teilnehmer zurückführen zu können ist eine Authentifizierung der Teilnehmer notwendig. Die Kombination kryptografischer Authentifizierungsmethoden mit den inhärenten biometrischen Merkmalen von Sprache, stellt hierbei einen grundlegenden Ansatz dar. Die Authentifizierung kann dabei prinzipiell ausschließlich über den Transportkanal erfolgen. Allerdings ist es vorteilhaft, wenn das gesamte Gespräch und die vorhergehende Authentifizierung durch authentisierte Geräte geführt werden, um den in 1) vorgestellten Integritätsschutz nicht zu gefährden und „End-to-End“ zu gewährleisten.
3. **Revisionssichere Speicherung der Kommunikation:** Die revisionssichere Dokumentation der Kommunikation, und damit maßgeblich das sichere und unveränderliche Speichern der Inhalte, ist in diesem Zusammenhang ein zentraler Punkt für die Nicht-Abstreitbarkeit der Konversation. Hierbei muss im Besonderen auch auf die Sicherung gegen mutwillige Manipulation geachtet werden.

Zur Erfüllung der ersten beiden Punkte will das INTEGER-Projekt nicht auf eine komplexe, aufwendige Public-Key-Infrastruktur zurückgreifen, sondern aufzeigen, dass mit Hilfe eines *Vertrauensankers* (z.B. Trusted Platform Module, TPM 2.0) die Nicht-Abstreitbarkeit einer Aufzeichnung erreicht werden kann. Zusätzlich werden effiziente und sichere Protokolle zur Signierung bilateraler, digitaler Kommunikation evaluiert, validiert und ggf. erweitert.

Zur *revisionssicheren Speicherung* werden die Kommunikationsdaten blockweise gespeichert. Der erste Block soll die Metadaten, wie z.B. SIP-URL der Teilnehmer, und den Zeitpunkt des Gesprächsbeginns enthalten. Danach folgen die Blöcke mit den Gesprächsdaten, wobei jeder Block signiert ist. Ist das Gespräch beendet, wird beim Auflegen ein spezieller Block hinzugefügt, der die Zeit und den Grund des Gesprächsabbruches enthält. Die jeweiligen Kommunikationspartner speichern die gesammelten Pakete zusammen mit der Signatur und können diese Aufzeichnung später vorlegen und somit die Existenz einer Kommunikation zweifelsfrei belegen. Anhand der inhärenten biometrischen Merkmale von Sprache sind mit dieser Aufzeichnung auch im Nachhinein die Identitäten der Personen verifizierbar.

Die VoIP-Kommunikation wird auf der Protokollebene durch das SIP- und RTP-Protokoll umgesetzt. Für eine Signatur über das Gespräch ist es hierbei wichtig, dass SIP (Aushandlung der

Kommunikationsparameter) und RTP (der Kommunikationsinhalt) erfasst werden. Im Unterschied zu einer konventionellen Signatur wird allerdings ein Dokument signiert, noch bevor es vollständig erstellt ist (ein Brief wird erst signiert, nachdem er geschrieben wurde, während bei INTEGER das Gespräch bereits signiert wird, während noch gesprochen wird). Das INTEGER-Projekt strebt eine Signatur des Gesprächs im Endgerät an, was der frühestmögliche Zeitpunkt in der Kommunikationskette ist. Es werden im INTEGER-Projekt dabei keine externen Smartcards eingesetzt, sondern die Inhalte durch entsprechende Trusted-Computing-Techniken signiert.

4.1 Technische Funktionalität

Zwei Geschäftspartner einigen sich auf Bedingungen und Konditionen eines Vertrages über eine VoIP-Telefonverbindung. Als Basis kann hier beispielsweise ein mobiles Endgerät wie ein Handy oder Smartphone zum Einsatz kommen. Zu einem bestimmten Zeitpunkt innerhalb dieser Verhandlungen wird durch die Geschäftspartner beschlossen, den Vertrag mündlich abzuschließen und so Zeit und Ressourcen zu sparen.

Beide Parteien besitzen ein zur sicheren Archivierung und Signierung der Gesprächsdaten geeignetes Endgerät, das eine PIN-Eingabe erfordert, ein biometrisches Merkmal oder Anderes, mit dem sie den Beginn der Signatur auf beiden Seiten einleiten. Beiden Seiten wird ab diesem Zeitpunkt signalisiert, z.B. über eine entsprechende Anzeige am Telefon, dass der Anruf signiert und auf beiden Seiten aufgezeichnet wird. Der Signierprozess wird am Ende der Verhandlung explizit beendet oder abgebrochen, wenn der Hörer aufgelegt wird. Der aufgezeichnete Vertrag wird zur Beweissicherung und Dokumentation durch die beteiligten Gesprächspartner archiviert.

Im Fall eines Rechtsstreits über die Inhalte oder die Existenz des Vertrags, kann diese Aufzeichnung auch vor Gericht verwendet und durch sog. „Inaugenscheinnahme“ zur Feststellung beweiserheblicher Tatsachen dienen. Eine Inaugenscheinnahme ist jede sinnliche Wahrnehmung von Beweismitteln, wozu auch akustische Abläufe gehören.

Ein technischer Gutachter kann anhand der Aufzeichnung nachweisen, dass die digitale Signatur gültig und in einem unveränderten Zustand ist. Hierdurch werden die Existenz und der entsprechende Inhalt des Gesprächs nachgewiesen. Ein weiterer Gutachter kann darüber hinaus anhand der biometrischen Merkmale von Sprache die Identität der Gesprächspartner nachweisen. Hier eingeschlossen ist der Nachweis der Echtheit der Sprache. Auch können Aussagen über die Sprachqualität und der Verständlichkeit getroffen werden. Auf dieser Basis können die Existenz und der Inhalt des Vertrags durch ein Gericht zwischen möglicherweise zuvor unbekannten Gesprächspartnern über eine VoIP-Verbindung festgestellt werden.

Nach einem expliziten Startsignal und einem erfolgreichen Handshake mit der Gegenseite, wird die Aufzeichnung und Signierung des Gespräches gestartet. Während des Handshake werden unter anderem auch die Zertifikate der Gesprächspartner ausgetauscht.

Der Kommunikationsablauf kann wie folgt exemplarisch dargestellt werden (siehe Abbildung 3). Teilnehmer *A* und *B* sammeln alle Pakete für den Kanal in einem Puffer (1). Wann immer die vordefinierte Chunk-Größe erreicht wurde, generiert *A* eine Signatur (2) über diesen und sendet die Signatur an *B* (3). Teilnehmer *B* überprüft nun die Signatur, indem er das Chunk mit dem von ihm empfangenen Paketen nachbildet und ebenfalls eine Signatur darüber bildet (4). Parallel dazu bildet Teilnehmer *B* ebenso eine Signatur über das zweite Chunk. War die Überprüfung der Signatur von Teilnehmer *A* erfolgreich, wird die bereits vorbereitete Signatur von

Teilnehmer *B* an *A* gesendet und signalisiert damit implizit, dass die empfange Signatur und die Integrität der Daten von *A* korrekt ist (6). Parallel dazu leitet Teilnehmer *B* das überprüfte Chunk an sein Archiv weiter (5).

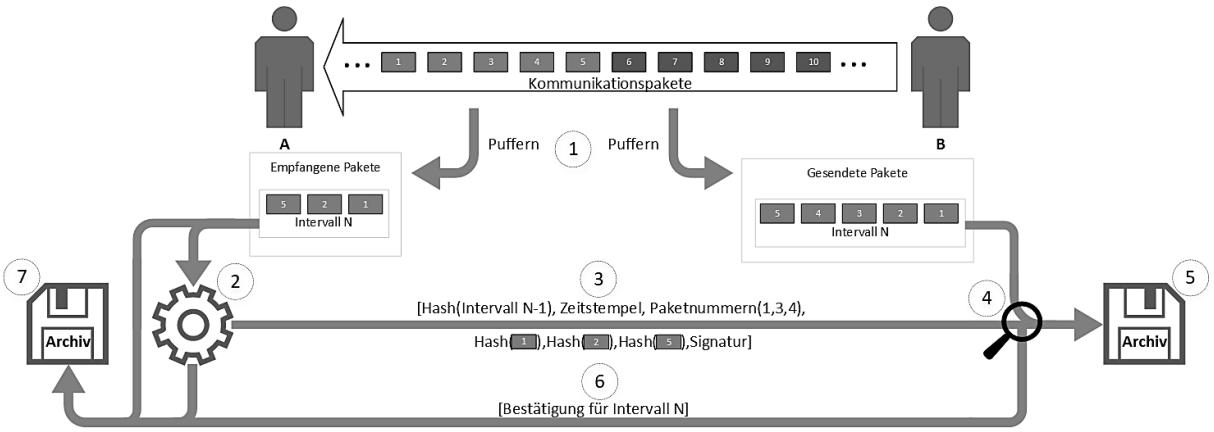

Abb. 3: Darstellung des Flussdiagramms der Kommunikation von A zu B

Teilnehmer *A* prüft nun ebenfalls die Signatur von *B*. Ist auch hier die Signatur korrekt, trägt Teilnehmer *A* sein eigenes Chunk und den von Teilnehmer *B* in die vom TPM verwaltete Hash-Kette ein und leitet beide an das Archiv weiter (7). Es werden Grundsätzlich nur Paare verarbeitet, da das zweite Chunk immer als implizite Bestätigung für die Korrektheit des ersten Chunks angesehen wird.

Nachdem der erste „Chunk-Zyklus“ durchlaufen ist, beginnt der nächste Zyklus, der dann die weiteren Chunk-Paare umfasst. Auch hier werden von beiden Teilnehmern wieder parallel die eigenen Chunks generiert und mit den entsprechenden Signaturen gebildet. Da der Teilnehmer *A* wieder als erster an der Reihe ist, schickt er seine Signatur an Teilnehmer *B*. Dieser prüft nun erneut die Signatur und sendet im Erfolgsfall seine eigene Signatur wieder an Teilnehmer *A*. Parallel dazu leitete Teilnehmer *B* nun das erste Chunk-Paar an sein Archiv weiter. Dies ist jetzt erst möglich, da er auf die implizite Bestätigung, dem Versand der zweiten Signatur von Teilnehmer *A* an ihn, warten musste. Dieser beschriebene Ablauf wiederholt sich fortlaufend, bis einer der Teilnehmer den Abschluss-Handshake einleitet und der Andere diesem zustimmt.

Um ein Gespräch zu beenden, wird auf Protokollebene ein explizites Stoppsignal erwartet. Dieses leitet den Abschluss-Handshake ein, durch den die finalen Signaturen und weitere Informationen ausgetauscht werden. Beide Partner müssen dem Beenden des Gespräches auf Protokollebene zustimmen. Dies bedeutet nicht, dass dies der Anwender selbst bestätigen muss, sondern kann von der VoIP-Anwendung automatisch erfolgen.

Die jeweiligen Kommunikationspartner speichern die gesammelten Chunks jedes Zyklus zusammen mit der Signatur und können diese Aufzeichnung später vorlegen, um die Existenz einer Kommunikation zweifelsfrei belegen zu können. Anhand der inhärenten biometrischen Merkmale von Sprache sind mit dieser Aufzeichnung auch im Nachhinein die Identitäten der Personen, unabhängig von den verwendeten Zertifikaten, verifizierbar.

4.2 TPM-Einsatz

Der Schutz von Geheimnissen kann durch die Mechanismen des *Trusted Computing* [8] erreicht werden, wie sie von der Trusted Computing Group (TCG) definiert werden. Das *Trusted Platform Module (TPM)* [TCG16] als spezieller Hardwarebaustein kann dabei Schlüssel erzeugen,

speichern und ermöglicht die Benutzung der Schlüssel auf sichere Weise. Beim Einsatz eines TPM kann vor Gebrauch sichergestellt werden, dass die verwendete Hard- und Software nicht manipuliert wurde bzw. dass eine Manipulation vorher bemerkt wird. Im INTEGER-Projekt wird man die neue TPM2.0-Spezifikation nutzen, die Anfang 2015 veröffentlicht und bereits von namenhaften Herstellern wie Infineon umgesetzt und nach CC EAL4+ zertifiziert wurden. Mit dieser wird es ermöglicht, die sichereren Algorithmen SHA256, SHA512 sowie Elliptischen Kurven Kryptographie (ECC) einzusetzen. Des Weiteren ermöglicht das TPM das sichere Speichern von Hash-Werten, die über eine Datenstruktur gebildet wurden, mit Hilfe des TPM Quote.

Damit eine Manipulation der Daten nachgewiesen bzw. entdeckt werden kann, wird eine Hash-Kette über die ausgetauschten Chunk-Paare gebildet. Diese Kette soll parallel zu dem Prozess der erfolgreichen Signierung erstellt werden (siehe Abbildung 4). Hierbei wird auf dem TPM nur asynchron zugegriffen, da die Zugriffszeiten auf das TPM keine Echtzeitverarbeitung erlauben.

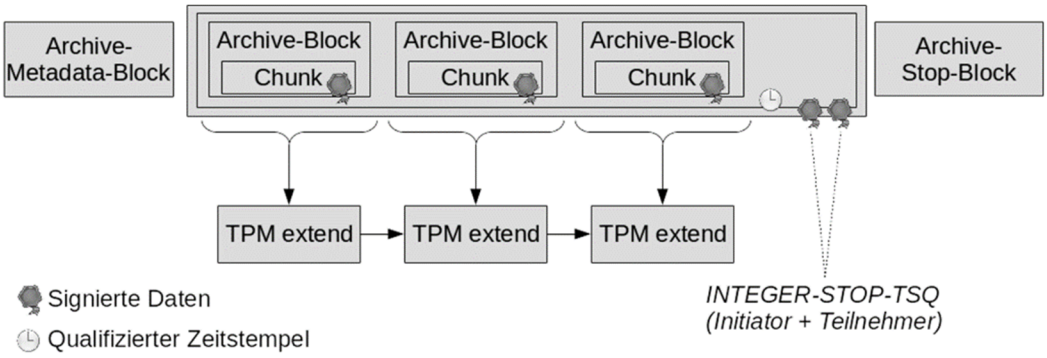

Abb. 4: Hash-Kette mit dem TPM

Durch diesen Ansatz kann theoretisch eine Vertauschung der Pakete mit der selben Blocknummer möglich sein. Die tatsächliche Reihenfolge wird erst durch den Austausch eines Stopp-Signales kontrollierbar, da diese einen qualifizierten Zeitstempel (siehe [SIGG01] und [SIGV01]) enthält. Mit diesem Zeitstempel ist eine nachträgliche Umsortierung, der Pakete ohne das dieses bemerkt werden, würde nicht mehr möglich. Alternativ kann auch ein ausreichend großer Zahlenraum für die Paketnummern verwendet werden, so dass keine Dopplung der Paketnummern auftreten kann.

Eine weitere Alternative, die davon abhängig ist, wie viel Zeit das Ansprechen des TPMs auf dem System letzten Endes benötigt, wäre es, die Signaturbildung durch ein TPM-Quote zu ersetzen. Der Quote enthält den jeweils aktuellen Stand der Hash-Kette und wird bei der Erstellung durch den jeweiligen TPM signiert. Es wird also anstelle einer Signatur über ein einzelnes Paket immer die „Signatur“ über alle bisher ausgetauschten Pakete gebildet und über die Signatur-Blöcke übertragen. In diesem Fall würde allerdings das parallele Erstellen der Signatur-Blöcke verloren gehen, da die für die Hash-Kette relevanten Pakete erst in das TPM eingetragen werden können, wenn sie vom Gesprächspartner bestätigt wurden. Weiterhin müssten initial auch die PCR-Basiswerte der Gesprächspartner ausgetauscht werden, da beide Gesprächspartner jeweils ihren eigenen TPM verwenden und entsprechend nicht davon ausgegangen werden kann, dass die zur Überprüfung benötigten PCR-Startwerte gleich sind. Entsprechend gäbe es auch nicht eine, sondern zwei Hash-Ketten, die jeweils jeder Partner bildet und die auch ausgetauscht werden müssten.

Die Möglichkeit, mit der hier vorgestellten Lösung per Telefon verbindlich Verträge abzuschließen, bedeutet sicherlich eine Unterstützung und Erleichterung für Unternehmen, da sie Ressourcen sparen, schnell und effizient mit einem System rechtsverbindliche Verträge abschließen und Angebote abgeben können. Die Einfachheit der Lösung beim Einsatz für den Anwender, der zum Signieren und Aufzeichnen eines Gesprächs nur eine entsprechende Taste drücken und seine PIN eingeben muss, führt sicherlich zu einer hohen Akzeptanz bei den Anwendern, was eine große Breitenwirkung impliziert. Die Arbeitswelt und vor allem die Prozesse eines Vertragsabschlusses oder einer Angebotsabgabe werden hierdurch deutlich vereinfacht werden, da nicht mehr alles schriftlich niedergelegt und per Post ausgetauscht werden muss.

4.3 Clearmode-Protokoll

Der *Clearmode Codec* nach RFC-4040 [KREU05] ermöglicht den transparenten Transport von RTP-Daten. Üblicherweise werden im VoIP-Umfeld ausschließlich Sprach- oder Videocodecs zum Transport ausgehandelt. Durch Clearmode können beliebige Informationen transportiert werden. Die Deutsche Telekom nutzt beispielsweise den RFC-4040 innerhalb ihrer Netze zur Steuerung hochwertiger ISDN-Anlagen über Voice-over-IP (VoIP). Daten aus dem ISDN-B-Kanal-Protokoll (z.B. X.75, V.110 oder V.120) werden mit Hilfe des RFC-4040 auf entsprechende IP-Datagramme abgebildet, die dann bei der entfernten Vermittlungsstelle auf einen herkömmlichen ISDN-Anschluss (oder einen VoIP-Anschluss mit RFC-4040-fähigem Router) übertragen werden (siehe Abbildung 5).

Abb. 5: Aufbau eines Clearmode-Codecs

Der Clearmode ist ein Grundmerkmal von VoIP-Medien-Gateways. Dieser Modus wird oft auch als „clear-channel data“, „64 kbit/s unrestricted“ oder „64k clear channel“ bezeichnet. Es wird kein Encoding oder Decoding verwendet, sondern lediglich Paketierung unterstützt. Im Clearmode wird davon ausgegangen, dass der Datenaustausch sampleorientiert und mit einem Oktett pro Sample fließt. Eine Einschränkung in der Anzahl der Samples je Paket gibt es nicht, lediglich die durch das IPv4-Protokoll herrschende Begrenzung auf 64 kByte stellt eine Limitierung dar. In der Praxis ist jedoch wegen der Maximum Transmission Unit (MTU) bei etwa anderthalb Kilobyte pro Paket die praktische Grenze für UDP-Transport erreicht.

Durch die Möglichkeit unveränderbare Daten mittels Clearmode austauschen zu können, bietet sich dieses Protokoll als Pseudo-Codec für INTEGER optimal an, um einen kombinierten Datenstrom zur Integritäts- und Beweissicherung von Telefonaten zu realisieren. Sämtlichen Steuer- und Nutzdaten können idealerweise innerhalb dieses Codecs bidirektional transportiert werden.

Der Clearmode-Codec ist bei INTEGER zunächst als reiner Mantel für unkomprimiertes Audio mit einer Abtastfrequenz von 8 kHz gehalten. Tests mit Equipment, welches dem der Telekom AG entspricht, wird CLEARMODE/8000 (Standard laut RFC-4040) unterstützen. In allen Fällen sollen Datenpakete nicht größer als 1.300 Byte ausfallen, wovon 81 Byte für unverschlüsseltes oder verschlüsseltes Audio und Steuerdaten reserviert sind.

INTEGER-fähige Endpunkte müssen sich zunächst identifizieren. Dazu ist es notwendig, behutsam mit den Ohren der Anwender umzugehen, und „Discovery-Pakete“ nur selten und gefolgt von leeren RTP-Paketen zu senden. Die bereits implementierte Peer2Peer-Verschlüsselung verfügt bereits über einen adäquaten Mechanismus. Zudem wird ein Schlüsseltausch zwischen den beiden Endpunkten durchgeführt, wodurch Daten entstehen, die nur den beiden Endpunkten bekannt sind. Auch diese Daten sollen über die Callbacks der INTEGER-Funktionalität zugänglich sein.

Nach erfolgreicher Beendigung der Discovery-Phase wird dem INTEGER-Funktionsblock zyklisch beim Empfang und vor dem Senden von Sprachpaketen über Callbacks Opus-komprimiertes Audio, sowie Steuerdaten zugänglich gemacht. Die genaue Definition der Daten erfolgt in einer späteren Projektphase. Die Daten ermöglichen in jedem Fall, die beiden Gesprächsteilnehmer anhand ihrer jeweiligen SIP URI zu identifizieren. Ein eigener 64-Bit-Paketzähler, der für die Sprachverschlüsselung zwingend benötigt wird, ist ebenfalls zugänglich. Für den Fall, dass den Paketen INTEGER-Daten hinzugefügt werden sollen, genügt es die Daten in ein zu diesem Zweck bereitgestelltes Array hineinzukopieren und die Anzahl der hinzuzufügenden Bytes zu benennen.

5 Resümee

Ein Ziel des Forschungsprojektes INTEGER ist die Integrität und Nicht-Abstreitbarkeit der internetbasierten multimedialen Kommunikation für IP-basierte Telefonie. Gerade im B2B- und B2C-Bereich fehlt den aktuellen Lösungen der Nachweis über die Vertraulichkeit und die Verlässlichkeit der Kommunikation. Der Anwendungsbereich von INTEGER zielt auf den Schutz der Integrität einer Kommunikation und die sichere Authentifizierung der Kommunikationspartner, durch das elektronische Signieren der Kommunikation. Hierfür soll ein Hardwarevertrauensanker – ein TPM-Chip 2.0 der Trusted Computing Group – zur IP-Telefonie mittels Softphones angebunden und über beliebige SIP-Provider genutzt werden. Im Rahmen des Forschungsprojektes wird ein System zur Signierung verbaler Kommunikation erarbeitet, dessen Marktbedarf bisher noch nicht gedeckt ist. Angestrebt ist eine völlig neuartige Form der Nicht-Abstreitbarkeit mündlicher Kommunikation, die grundlegend effizientere Geschäftsabläufe (u. a. als Beweis für mündliche Vertragsabschlüsse) ermöglicht.

Danksagung

Das INTEGER-Projekt (www.integer-project.de) ist ein gefördertes BMWi-Projekt mit einer Laufzeit von zwei Jahren, welches im Juli 2017 seine Arbeiten aufnahm und im Juni 2019 voraussichtlich enden wird. An dem Projekt sind die Firmen DECOIT GmbH (Projektleitung), Global IP Telecommunications Ltd. (Softphone-Hersteller), reventix GmbH (Provider) sowie die deutsche Forschungseinrichtung Hochschule Bremen beteiligt. Als assoziierte Partner nehmen der Hersteller Infineon Technologies AG und die Forschungseinrichtungen Fraunhofer SIT und Hochschule Mainz an dem Projekt teil. Daher gilt der Dank den Partnern des Projektes, die durch ihre Beiträge und Arbeiten die erfolgreiche Projektarbeit erst ermöglichen.

Literatur

[BCS06] R. Baumann, S. Cavin, S. Schmid: Voice Over IP – Security and SPIT. 2006.

[DEER07] K.-O. Detken, E. Eren: VoIP Security – Konzepte und Lösungen für sichere VoIP-Kommunikation. Hanser 2007.

[DETK16] K.-O. Detken: Alles wird IP - VoIP-Migration im Geschäftskundenbereich. NET 06 2016.

[EURO14] Europäisches Parlament und Rat: Richtlinie 2014/65/EU. Amtsblatt der Europäischen Union über Märkte für Finanzinstrumente sowie zur Änderung der Richtlinien 2002/92/EG und 2011/61/EU, Brüssel 2014.

[HKS06] C. Hett, N. Kuntze, A. U. Schmidt: Security and Non-Repudiation for Voice-Over-IP Conversations, ISSA 2006.

[HKS07] C. Hett, N. Kuntze, A. U. Schmidt: Non-Repudiation in Internet Telephony, SEC 2007.

[HUIT02] C. Huitema: Real Time Control Protocol (RTCP) attribute in Session Description Protocol (SDP). RFC-3605 (Standards Track), Network Working Group, IETF, Oktober 2003.

[KREU05] R. Kreuter: RTP Payload Format for a 64 kbit/s Transparent Call. RFC-4040 (Standards Track), Network Working Group, IETF, April 2005.

[RSC+02] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler: SIP: Session Initiation Protocol. RFC-3261 (Standards Track), IETF, Juni 2002.

[SCFJ03] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson: RTP: A Transport Protocol for Real-Time Applications. RFC-3550 (Standards Track), Network Working Group, IETF, Juli 2003.

[SIGG01] Bundesministerium der Justiz: Gesetz über Rahmenbedingungen für elektronische Signaturen (Signaturgesetz – SigG). Ausfertigungsdatum: 16.05.200.

[SIGV01] Bundesministerium der Justiz: Verordnung zur elektronischen Signatur (Signaturverordnung – SigV). Ausfertigungsdatum: 16.11.2001.

[TCG16] Trusted Computing Group: Trusted Platform Module Library Specification, Family "2.0". Level 00, Rev. 01.38 – Sept. 2016 Part 1: Architecture, 2016.