

1 Grundlagen der IP-Telefonie

Dieses Kapitel beschäftigt sich mit den Grundlagen der Übertragung von Telefongesprächen über Netzwerke mit Hilfe der Protokolle der IP-Protokollsuite. Es soll die zum Verständnis der späteren Kapitel notwendigen Grundlagen vermitteln. Wer sich sofort mit der praktischen Umsetzung beschäftigen möchte, kann direkt zum Kapitel 3 vorblättern und erst bei Bedarf die technischen Einzelheiten aus diesem Kapitel nachlesen.

*Für Eilige:
Gleich zu Kapitel 3
weiterblättern und mit der
Installation beginnen.*

1.1 Geschichte

Im Bereich der Universitäten und der Forschung hat man sich schon seit Mitte der 80er-Jahre Gedanken über die Möglichkeit gemacht, Sprache in paketbasierten Netzen zu übertragen. Erste Standards und Protokolle wurden definiert und Versuche in verschiedenen Konfigurationen wurden vorgestellt.

Da für alle diese Verfahren ein recht großzügig dimensioniertes Netz Voraussetzung war, hatten sie keine Chance, sich aus den Laboren hinaus weiter zu verbreiten. Zur gleichen Zeit durchlief die Übertragung von Sprache eine Revolution. In den Vermittlungen und Nebenstellenanlagen wurde zunehmend digital gearbeitet. Die Übertragung zwischen Vermittlungsstellen erfolgte in immer mehr Fällen in digitaler Form. Bei der ITU (International Telecommunications Union), die zu diesem Zeitpunkt noch unter dem Namen CCITT (Comité Consultatif International Télégraphique et Téléphonique) bekannt war, entstanden die Normen, die heute die Kommunikation bestimmen: ISDN und die zugehörigen Weitverkehrsprotokolle.

*Die ITU legte die Basis
für digitale
Sprachübertragung.*

Datennetze und Sprachnetze waren noch völlig getrennte Welten. Für den Betrieb über die damalige Breitbandwelt wurden von der ITU Vorschläge zum Aufbau von Videokonferenzen gemacht, die wenig später dann aus Kostengründen auch an lokale Netze mit IP-Daten-

übertragung angepasst wurden. Aus diesem Fundus von Protokollen entwickelten sich die ersten Lösungen zur Sprachübertragung im Internet.

Die ersten Produkte der 90er-Jahre konnten nicht überzeugen.

Bereits zu Anfang der 90er-Jahre wurden erste Produkte auf Grundlage dieser Standards mit kommerziellem Hintergrund zur Übertragung von Sprache über das IP-Netz oder genauer das Internet am Markt verfügbar. Obwohl zuerst euphorisch begrüßt und mit viel Vorschusslorbeeren versehen, fristeten diese Produkte wegen mangelnder Sprachqualität und umständlicher Bedienung nur ein Nischendasein. Hatte man zuerst den Untergang der herkömmlichen Telefongesellschaften prognostiziert, so musste man schnell feststellen, dass der normale Internetzugang in diesen Jahren einfach keine ausreichende Sprachqualität ermöglichte.

Gleichzeitig drängten so viele unterschiedliche Lösungen auf den Markt, die nicht miteinander kompatibel waren, dass der Anwender sich völlig verwirrt und enttäuscht wieder auf ein normales Telefon zurückgezogen hat und das Feld Sprache über Datennetze (Voice over IP – VoIP) den Forschern und Entwicklern und einem hartgesottenen Kern von Computerliebhabern überlassen hat.

VoIP wurde zuerst zwischen Nebenstellenanlagen kommerziell erfolgreich eingesetzt.

Nach dem Abklingen der ersten Euphorie über Sprache im Internet kam es zu einer Verschiebung der Schwerpunkte. Statt Sprache von einzelnen Endgeräten über das Internet zu transportieren, wurde jetzt in Unternehmen verstärkt daran gedacht, Sprache zwischen Unternehmensteilen über das vorhandene Datennetz auszutauschen. Es ging dabei mehr um die Verbindung zwischen Nebenstellenanlagen als um den Anschluss von einzelnen Telefonen. Dies stellte sich bald als technisch gut realisierbar und gleichzeitig als Möglichkeit zur Kostensenkung dar. Die Hersteller nahezu aller Nebenstellenanlagen (PBX – Private Branch Exchange) boten im Lauf der Jahre ihren Kunden die Möglichkeit, die Anlage einer Außenstelle oder einer Filiale mit der PBX der Zentrale über das Datennetz zu verbinden. Dabei wurden teilweise private Protokolle eingesetzt, teilweise verwendete man aber auch existierende Standards aus dem Bereich der ITU-Protokolle. Der Einsatz erfolgte meist auf privaten fest angemieteten Leitungen, sodass Durchsatz oder Störungen von außen kein Thema waren.

Nach und nach erweiterte sich das Angebot, es wurden neben der rein internen Kommunikation Lösungen für das Einwählen von außen und Verfahren zum Weiterleiten von Anrufen über das interne Netz bis nahe zum Zielpunkt und der Übergabe an das öffentliche Netz möglichst direkt beim Empfänger entwickelt. Alle diese Entwicklungen waren in erster Linie vom Gedanken der Einsparung von Gebühren bei der Nutzung der öffentlichen Telefonnetze geprägt.

Als die Welt der Nebenstellenanlagen immer mehr die Vorteile der Datentechnik verinnerlichte und gleichzeitig die Kostenstruktur für breitbandige Internetanschlüsse günstiger wurde, kam es bei der Entwicklung von Lösungen für VoIP zu einem neuen Aufschwung. Die IETF (Internet Engineering Taskforce), bei der historisch nahezu alle Standards für Datennetze entwickelt wurden, hat zusammen mit Gruppen der ETSI (European Telecommunications Standards Institute) eine Reihe neuer Standards für die Steuerung der Übertragung von Sprache in Datennetzen entwickelt. Diese Familie von Standards (SIP [Session Initiation Protocol] und mittlerweile ergänzend auch ENUM [tElephone NUmber Mapping]) sollte den Aufbau von Gesprächen einfacher und sicherer machen. Gab es zunächst nur wenige Endgeräte, die das Prädikat benutzerfreundlich auch nur annähernd verdienten, war es jetzt durch die neuen Protokolle möglich, auch ohne Computerwissen über das Netz zu telefonieren. Es entstanden erste Geräte, die sich wie herkömmliche Telefone bedienen lassen und dennoch eine Reihe von neuen Leistungsmerkmalen aufweisen. Gleichzeitig wurden die Protokolle so erweitert, dass auch Funktionen und Eigenschaften möglich wurden, die bisher nur mit proprietären Protokollen von Nebenstellenanlagenherstellern verwirklicht werden konnten.

*IETF und ETSI
beschäftigen sich
mit VoIP.*

Der Einsatz von VoIP-basierten Telefonen und Nebenstellenanlagen ist heute in der Geschäftswelt schon fast zum Alltag geworden. Es gibt reale Einsparungspotenziale, wobei die Differenz zum herkömmlichen Sprachnetz durch die laufende Senkung der Tarife deutlich kleiner geworden ist. Aber gerade im Geschäftsumfeld gibt es eine Reihe weiterer Vorteile von VoIP: die Einbindung in den Arbeitsalltag auf dem Computer, Funktionen wie direktes Wählen aus dem zentralen Adressbuch oder der Identifikation des Anrufers mit direktem Aufblättern der zugehörigen Kundenakte am Bildschirm.

*VoIP kann Kosten sparen
und bietet gleichzeitig
neue Möglichkeiten.*

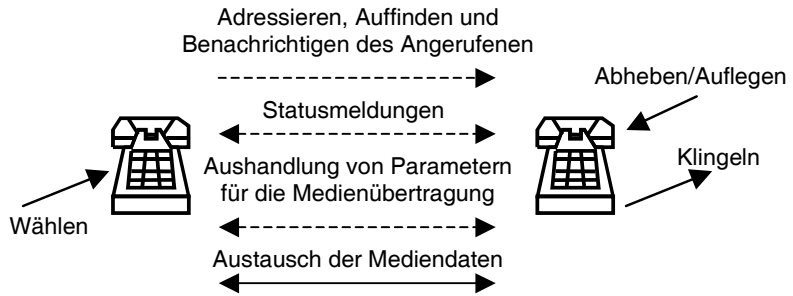
Gleichzeitig haben in den letzten Jahren extrem kostengünstige Angebote für VoIP, meist kombiniert mit Internet-Zugangstarifen ohne Zeit- und Volumenbindung, den Markt der Privatanwender auf den Ansatz »Sprache im Datennetz« aufmerksam gemacht. Die Qualität ist zwar nicht immer befriedigend, der Preis ist jedoch unschlagbar.

Was bisher noch etwas stiefmütterlich behandelt wurde, ist der Bereich der kleinen Unternehmen, der Ein-Mann-Büros und der kleinen Außenstellen von Firmen. Hier passen kommerziell verfügbare Lösungen oft nicht, gleichzeitig sind aber die Einsparungspotenziale durch VoIP gerade in diesem Bereich immer noch sehr groß. In diesem Umfeld bieten sich Lösungen mit einer kleinen Nebenstellenlage auf Basis von Linux mit Asterisk direkt an.

1.2 Grundlegende Mechanismen und Protokolle

In der IP-Telefonie werden mit den Operationen zur Signalisierung und dem eigentlichen Transport der Sprachdaten zwei grundlegende Teilbereiche unterschieden. Die nachfolgende Abbildung 1–1 macht dies deutlich und zeigt die zeitliche Abfolge und Abhängigkeiten.

Abb. 1–1
Signalisierung und
Medienpfade

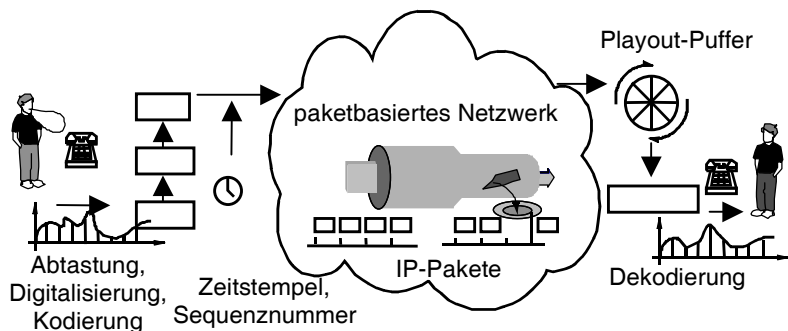


In diesem Kapitel werden zunächst die Mechanismen zum Medien-transport vorgestellt, die im einfachsten Fall durch Angabe der jeweiligen Endpunkt-Adressen und Ports eingestellt werden. Wie dies komfortabler und in der für den Nutzer gewohnten Art und Weise, z.B. durch Wahl einer Nummer, geschehen kann, zeigt die dann folgende Vorstellung einiger wichtiger Signalisierungsprotokolle.

1.3 Mechanismen und Standards zur Medienübertragung

Die nachfolgende Darstellung befasst sich mit der eigentlichen Sprachübertragung. Wie diese grundsätzlich abläuft, zeigt die folgende Abbildung 1–2.

Abb. 1–2
Pfad für Audiodaten



Es ist zu sehen, dass die Sprache zunächst durch ein Mikrofon aufgenommen, dann digitalisiert, in Pakete verpackt, über ein auch potenziell verlustbehaftetes Netz übertragen, beim Empfänger wieder zusammengesetzt, in ein Analogsignal umgewandelt und letztlich wiedergegeben wird. Damit dies effizient und auch unter dem Einfluss einer Reihe von möglichen Störfaktoren in guter Qualität und dennoch ohne unnötig Bandbreite zu verbrauchen erfolgt, sind eine Reihe von Protokollen notwendig.

Sprache wird in Paketen im IP-Netz übertragen.

1.3.1 Kodierung von Audiodaten

Um Sprache über digitale Verbindungen übertragen zu können, muss die Sprache digitalisiert werden. Dafür steht eine Reihe unterschiedlicher Verfahren (Codecs) zur Verfügung. Dazu wird die Sprache von einem Analog-Digital-Wandler in diskrete Werte umgewandelt und anschließend kodiert.

Ziel dieser Kodierung ist es einerseits, den Umfang der zu übertragenden Daten möglichst gering zu halten, andererseits aber keine zu große zusätzliche Verzögerung im Mediendatenpfad zu verursachen und nach Möglichkeit auch sicherzustellen, dass selbst bei einem Verlust oder einer Verfälschung von Sprachpaketen möglichst geringe Störungen beim Empfänger hörbar sind.

Im einfachsten Fall werden die Daten nur digitalisiert und es werden keine zusätzlichen Maßnahmen unternommen, um dem kodierten Signal Redundanz zu entziehen und seinen Umfang damit zu verringern. Das dazu gebräuchliche und praktisch von allen Endgeräten sowie den IP-Telefonie-Anlagen unterstützte Verfahren ist G.711, bei dem die Audiodaten mit einer Rate von 8000 Hz abgetastet werden und ein einzelnes Sample in einem 8-Bit-Wert kodiert wird. Daraus resultiert eine typische Datenrate von 64 kbps, wie sie bereits vom ISDN bekannt ist, wo dieses Kodierungsverfahren ebenfalls Anwendung findet.

Sprache muss digitalisiert werden, bevor sie im Netz übertragen werden kann.

Komprimierende Verfahren nutzen statistische Eigenschaften des Signals, indem sie wie im Falle der adaptiven differenziellen Pulsecode-modulation (ADPCM) versuchen, jeweils nur Differenzen zu einem vorher versendeten Wert zu übertragen. Andere Verfahren erreichen eine deutlich bessere Komprimierung, indem sie Parameter und Stimuli eines nachgebildeten menschlichen Sprachapparates berechnen und übertragen. Solche Codecs sind insbesondere bereits aus dem Mobilfunk (GSM) bekannt, bei dem Bandbreite naturgemäß eine knappe Ressource darstellt. Grundsätzlich setzen sie eine höhere Verarbeitungsleistung sowohl auf der Sender- als auch Empfängerseite voraus.

Starke Komprimierung ergibt kleinere Pakete, erfordert mehr CPU-Power und erzeugt Verzögerungen.

Sie führen zu einer generell vorhandenen zusätzlichen Verzögerung, da das Sprachsignal zunächst für eine gewisse Zeit analysiert werden muss. In einigen Fällen existieren Patente und die Verfahren sind daher in ihrer Nutzung mit Lizenzgebühren verbunden.

Die nachfolgende Tabelle fasst Codecs, genutztes Verfahren, resultierende Datenrate, Verzögerung und Aufwand zusammen und stellt damit eine Basis für die Auswahl eines geeigneten Codecs dar.

Tab. 1-1
Codecs

Codec	Verfahren	Bitrate (kbps)	Abtastgröße (Bytes)	Abtastabstand (ms)	Bandbreite im LAN (kbps)
G.711	μ-Law/A-Law	64	80	10	87,2
G.729	CS-ACELP	8	10	10	31,2
G.723.1	LPAS	6,3	24	30	21,9
G.723.1	LPAS	5,3	20	30	20,8
G.726	ADPCM	32	20	5	55,2
G.726	ADPCM	24	15	5	47,2
G.728	LD-CELP	16	10	5	31,5

Die Wahl des Codecs muss sich nach lokalen Gegebenheiten richten.

Geeignete Kalkulationshilfen für die Auswahl eines passenden Codecs unter bestimmten Nutzungsbedingungen sind sowohl online als auch als standalone nutzbare Version vorhanden [1, 2]. Innerhalb einer Asterisk-Lösung werden sowohl G.711, frei verfügbare Implementierungen höher komprimierender Codecs, aber auch lizenzpflichtige und gegen eine Gebühr als Objektmodul angebotene kommerzielle Codecs wie G.729 unterstützt. Durch eine entsprechende Systemkonfiguration, die in den Abschnitten 3.3.1 und 3.3.3.2 ausführlich beschrieben wird, ist festlegbar, in welcher Reihenfolge sie mit den beteiligten Endsystemen ausgehandelt werden. Generell ist dabei abzuwägen, dass der Vorteil einer besseren Rohdatenrate eines höher komprimierenden Codecs teilweise durch den bei der Übertragung anfallenden Overhead aus IP-, UDP- und RTP-Header wieder aufgehoben wird. Dies gilt insbesondere deshalb, weil im Falle höher komprimierender Codecs kleinere Sprachpakete versandt werden und sich das Verhältnis aus Nutzdaten und Header-Informationen damit ungünstiger gestaltet. Es ist nur innerhalb gewisser Grenzen möglich, zu versuchen, entsprechend mehr Audiodaten in einem Paket zu übertragen, da diese natürlich vorab zunächst aufgesammelt werden müssen, wobei eine entsprechende Zeit vergeht, die man auch als Paketierungsdelay bezeichnet. Um die Verzögerung im Mediendatenpfad nicht zu groß werden zu lassen (Werte über etwa 250 ms werden bereits als sehr unangenehm bis hin zu unakzeptabel empfunden), sind hier Grenzen gesetzt. Nicht

zuletzt deshalb ist in vielen Anwendungsfällen die sehr einfache und durchgängig unterstützte Kodierung nach G.711 durchaus gebräuchlich.

1.3.2 Transport der Mediendaten

Bei der Übertragung der Mediendaten kommt zunächst UDP zum Einsatz. TCP mit seinen Flusssteuerungs- und Retransmissionsmechanismen ist ungeeignet, da es sich bei der übertragenen Sprache um verzögerungssensitive Echtzeitdaten handelt. Kommt ein Mediendatenpaket nicht innerhalb einer gewissen Zeitspanne beim Empfänger an, so ist auch eine Retransmission nicht sinnvoll.

Zur Gewährleistung einer guten Sprachqualität auch bei Paketverlust und wechselnden Verzögerungen bei der Übertragung (man bezeichnet diesen Effekt auch als Jitter) kommen beim Empfänger in der Regel so genannte Playout-Buffer-Mechanismen zum Einsatz. Bei diesen werden die Daten nicht sofort beim Eintreffen wiedergegeben, sondern zunächst in einem Ringpuffer zwischengespeichert und dann mit einem wegen des interaktiven Charakters eines Telefonats nicht beliebig verlängerbaren Versatz ausgespielt. So kann Jitter ausgeglichen werden und im Fall eines Paketverlustes ist es möglich, fehlende Samples auch zu interpolieren, anstatt einfach einen Nullwert auszugeben.

Mediendaten werden vorzugsweise mit UDP übertragen.

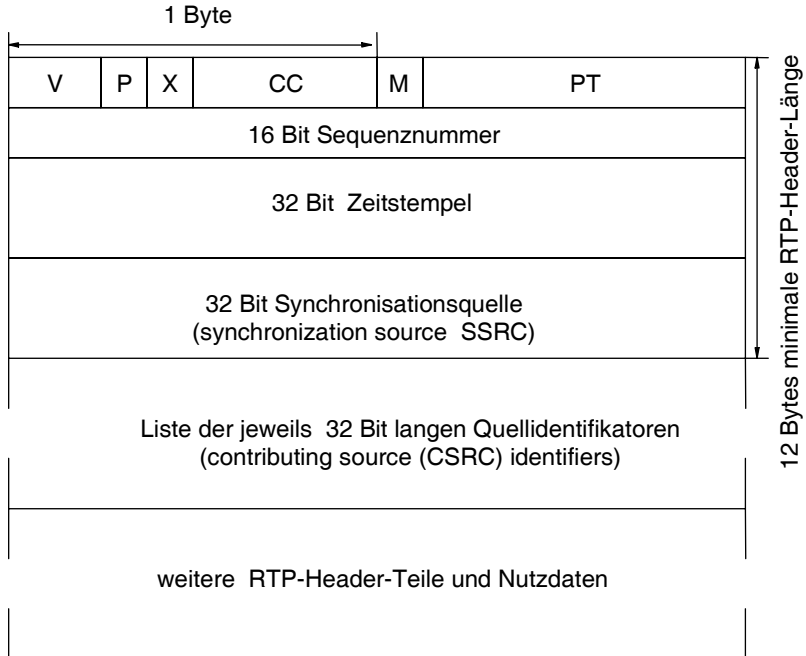
Unterstützt werden diese Mechanismen durch den Einsatz des Protokolls RTP (Real-Time Transport Protocol) [16]. Dessen Name ist etwas irreführend, es handelt sich weder um ein Transportprotokoll noch hat das Protokoll unmittelbare Mechanismen zur Unterstützung von Echtzeitanwendungen. Vielmehr folgt das Protokoll dem Ansatz des Application Level Framing (ALF), arbeitet damit auf der Anwendungsschicht (Schicht 5 des Internetmodells bzw. Schicht 7 des OSI-Modells) und kommt oberhalb des Transportprotokolls UDP zum Einsatz.

Die nachfolgende Abbildung 1–3 zeigt den Aufbau eines RTP-Paketes.

In der Abbildung bezeichnet V die Protokollversion, P kann benutzt werden, um ein Padding hinter den Nutzdaten anzuzeigen, X erlaubt es, auf so genannte Extension-Headers hinzuweisen, CC (CSRC Count) enthält die Anzahl der Contributing Sources und M ist ein Marker, der z.B. Sprachpausen anzeigen kann. Wie zu sehen ist, führt das RTP-Protokoll eine Beschreibung der übertragenen Daten in einem so genannten Payload Type PT (hier kann z.B. der genutzte Codec angegeben werden oder auch beschrieben werden, dass es sich

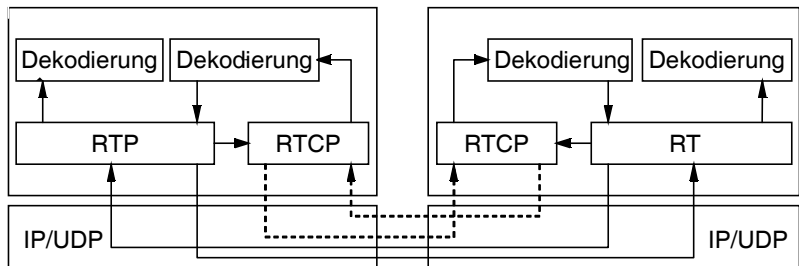
RTP dient zur Steuerung des Datenflusses.

Abb. 1-3
Aufbau eines RTP-Paketes



um ein RTP-Kontrollpaket handelt) sowie Sequenznummern und Zeitstempel ein. Mit deren Hilfe ist es möglich, Paketverlust oder eine Umordnung während der Übertragung zu erkennen und auf diese zu reagieren und ein Monitoring des Übertragungskanals hinsichtlich seiner Eigenschaften durchzuführen. Wird damit erkannt, dass wiederholt Pakete verloren gehen oder stark variierend bzw. sogar übermäßig verzögert werden, kann darauf reagiert werden. Dies ist in der nachfolgenden Abbildung 1-4 zu sehen.

Abb. 1-4
RTP-Feedback



Zur Umsetzung des gezeigten Feedbacks vom Empfänger zum Sender kommt das Real-Time Control Protocol (RTCP) zum Einsatz, und zwar jeweils in Kombination mit RTP. Spricht man von RTP, dann ist in der Regel RTP in der Kombination mit RTCP gemeint. Das RTCP-

Protokoll ermöglicht so genannte Sender- und Receiver-Reports, in denen die zusammengefassten Daten zum Status der Medienübertragung in gewissen Abständen übertragen werden.

Das RTP-Protokoll kommt für die Medienübertragung fast durchgängig bei den unterschiedlichen VoIP-Protokollen zum Einsatz und hat auch bereits mit der Übertragung von Mediendaten per Multicast im so genannten Mbone [17] eine lange Tradition. Zur Erhöhung der Übertragungseffizienz zum Beispiel auf Punkt-zu-Punkt-Links wurde es mittlerweile auch durch Standardisierungsvorschläge zur Header-Komprimierung oder auch zur Aggregation von mehreren Gesprächen (wie es beim so genannten Trunking zwischen unterschiedlichen IP-Telefonie-Anlagen sinnvoll ist) ergänzt [3, 5]. Auch gibt es die Möglichkeit, verschlüsselte Mediendaten mittels des Secure-RTP-Protokolls (SRTP) [6] zu übertragen und sogar die Schlüssel für diese gesicherte Übertragung in-band auszuhandeln oder auszutauschen [7].

RTP und RTCP sind die Grundlage für den Transport von Sprache im Internet.

Im Rahmen der Kopplung von Asterisk-Anlagen kommt das nachfolgend noch detaillierter beschriebene IAX2-Protokoll (Inter Asterisk Exchange) zum Einsatz, bei dem Signalisierung und Medientransport in nur einem Typ von Paketen (mit einem entsprechenden den Typ beschreibenden Feld) kombiniert werden. In diesem Fall wird die RTP-Medienübertragung durch IAX ersetzt. Die angewandten Mechanismen wie Sequenznummern, Monitoring der Übertragungsqualität und Möglichkeit zur Aggregation von Gesprächen sind aber äquivalent. In den Standardisierungsdokumenten [8] gibt es eine umfangreiche vergleichende Bewertung zu den genutzten Mechanismen und deren Effizienz.

IAX stellt eine hervorragende Alternative zu RTP dar.

1.4 Signalisierungsprotokolle

Das Konzept der Asterisk-Lösung, wie diese in der nachfolgenden Abbildung 1–5 gezeigt wird, kombiniert leistungsfähige Funktionen im Kern der Asterisk-PBX mit der Möglichkeit, unterschiedlichste VoIP-Protokolle, aber auch den Übergang ins klassische PSTN (Public Switched Telephone Network) über so genannte Channel-Interfaces zur Verfügung zu stellen.

Diese Channel-Interfaces sind hinsichtlich ihrer Eigenschaften und der zur Verfügung zu stellenden Funktionen standardisiert und erlauben dadurch eine sehr gute Erweiterbarkeit. Die Charakteristika wichtiger Signalisierungsprotokolle, die so angebunden werden können, werden nachfolgend vorgestellt.