

Digitale Sprache und Video im Internet

Kapitel 6.1 Digitalisierung und Paketvermittlung von Sprache

Gliederung

- Audio Streaming System
- Sprachdigitalisierung
- Sprachkodierer
- Paketvermittelte Sprache G.764
- Voice over IP
- Voice over FR

Sprachübertragung über Paketnetze

□ Verarbeitungsschritte

1. Sprache aufnehmen und digitalisieren
2. Daten komprimieren
3. Daten in Pakete unterteilen
4. Daten versenden (Vermittlung, Wegewahl, Forwarding!)
5. Daten erreichen den Empfänger
6. Datenpakete aneinanderfügen
7. Dekomprimieren
8. Daten in ein analoges Signal zurückwandeln
9. Analoges Signal akustisch ausgeben

Schritte 1.- 3., 7. - 9. Sind Codec-relevant

Schritte 3. - 7. netzrelevant

Schritte 2. - 7. verzögerungsträchtig

QoS-Aspekte bei Sprachübertragung (1)

Service Quality	Sound Quality	Conversation Quality
<ul style="list-style-type: none"> • offered services—such as calling card, 1-800/900 services, follow-me, and voice mail • availability of users in other countries or regions • network availability—down time, busy signals • reliability—such as dropped calls or wrong number • price 	<ul style="list-style-type: none"> • loudness • distortion • noise • fading • crosstalk 	<ul style="list-style-type: none"> • loudness • distortion • noise • fading • crosstalk • echo • end-to-end delay • silence suppression performance • echo canceller performance

QoS-Aspekte bei Sprachübertragung (2)

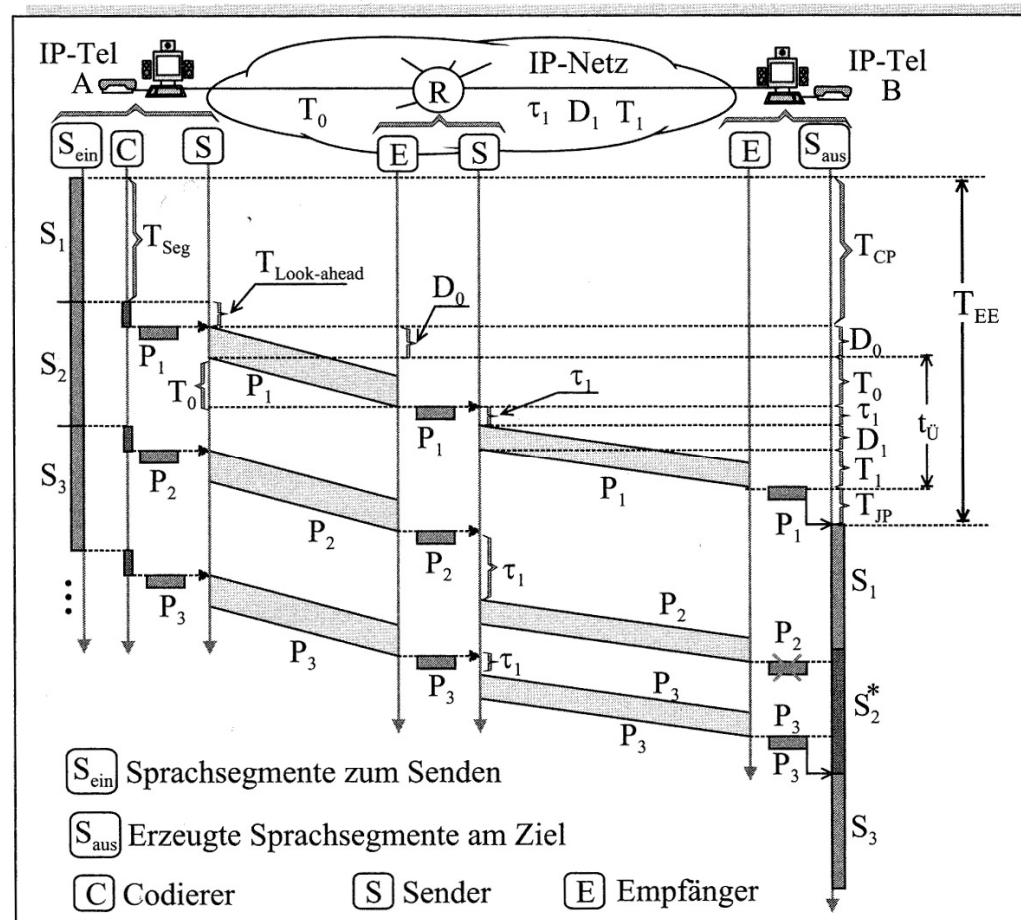
- ❑ Sprachqualität hängt ab von einer Kombination der Faktoren:
Delay, Verlustrate, Burst-Verluste, Codec, Packet Loss Concealment
- ❑ Codec und Delay
 - Je höher Codec-Komplexität (Güte der Kompression), desto höher Delay
 - Je höher Delay, desto schlechter Qualität
- ❑ Verlustrate und Sprachqualität
 - Je stärker die Kompression (niedrige Datenrate), desto höher Verlustauswirkungen
 - Höhere Verlustraten bedingen höhere Delay-Jitter
- ❑ Jitterpuffer und Verlustraten, Auslastung
 - Größere Jitter bedingen größere Jitterpuffer, erhöht Delay
 - Je höher die Auslastung, desto höher Delay und Verlustrate
 - Je mehr Frames pro Sprachpaket, desto niedriger effektive Datenrate, aber erhöhtes Delay

QoS-Aspekte bei Sprachübertragung (3)

□ Anforderungen von VoIP an IP-Netze

- Bandbreite der virtuellen Verbindung zwischen IP-Telefonen
 - fi Queue-Management in den Komponenten: Klassifizierung, Scheduling (Priority, Weighted Round Robin, Class-based)
 - fi Traffic Shaping
- End-End-Verzögerung (Delay T_{EE})
 - gemäß G.114: $\leq 150\text{ms}$ (akzeptabel für alle), ≥ 300 (nicht akzeptabel)
 - Beiträge sind: Kodierungs- u. Paketisierungszeit T_{CP} ,
 - Serialisierungsverzögerung $D = \text{Paketgröße} / \ddot{U}\text{-Rate}$, Übermittlungszeit T ,
 - Zwischenspeicherungszeit T_{JP}
- Schwankungen in der Übermittlungszeit (Jitter)
- Paketverlustrate (Packet Loss Rate)

QoS-Aspekte bei Sprachübertragung (4)



QoS-Aspekte der IP-Schicht (1)

- Einordnung von QoS-Parametern gemäß ITU-T I.350

Kriterium Funktion	Geschwindigkeit	Genauigkeit	Zuverlässigkeit
Verbindungs- aufbau			
Informations- übermittlung	Paketlaufzeit Paketlaufzeit- schwankungen	Paketfehler- häufigkeit Paketfehlein- fügungsrate	Paketverlust- häufigkeit
Verbindungs- abbau			

QoS-Aspekte der IP-Schicht (2)

□ Festlegungen der IETF

- allgemeine Eigenschaften (RFC 2330)
- A One-way Delay Metric for IPPM (RFC 2679)
- A Round-trip Delay Metric for IPPM (RFC 2681)
- IP Packet Delay Variation Metric (RFC 3393)
- IPPM Metrics for Measuring Connectivity (RFC 2678)
- One-way Loss Pattern Sample Metrics (RFC 3357)
- Network performance measurement with periodic streams (RFC 3432)

QoS-Aspekte der IP-Schicht (3)

□ Festlegungen der ITU-T

- Definition von IP-QoS (Y.1540) und quantitative Vorgaben (Y.1541)
- IP Packet Transfer Delay (IPTD)
- IP Packet Delay Variation (IPDV)
- IP Packet Error Ratio (IPER)
- IP Packet Loss Ratio (IPLR)
- Spurious IP Packet Rate (SIPP)
- IP Service Availability

QoS-Aspekte der IP-Schicht (4)

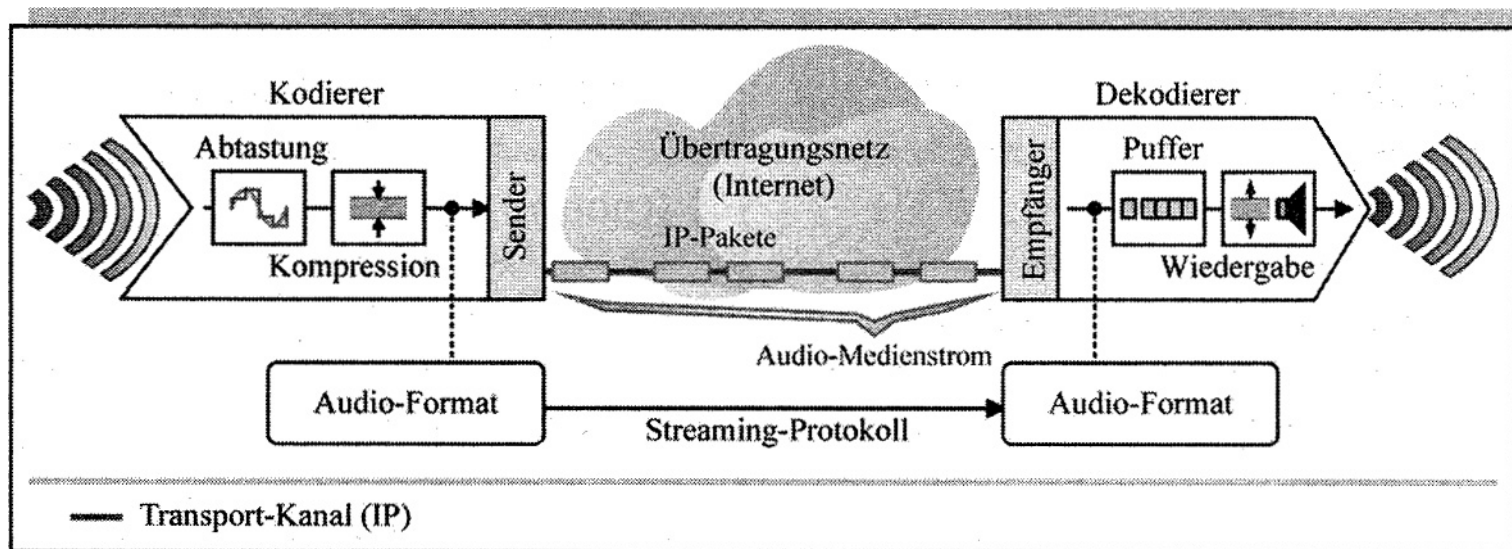
□ Dienstgüteklassen für IP gemäß ITU-T Y.1541

	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5
IPTD	100 ms	400 ms	100 ms	400 ms	1 s	U
IPDV	50 ms	50 ms	U	U	U	U
IPLR	10^{-3}	10^{-3}	10^{-3}	10^{-3}	10^{-3}	U
IPER	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	U

U: Unspecified

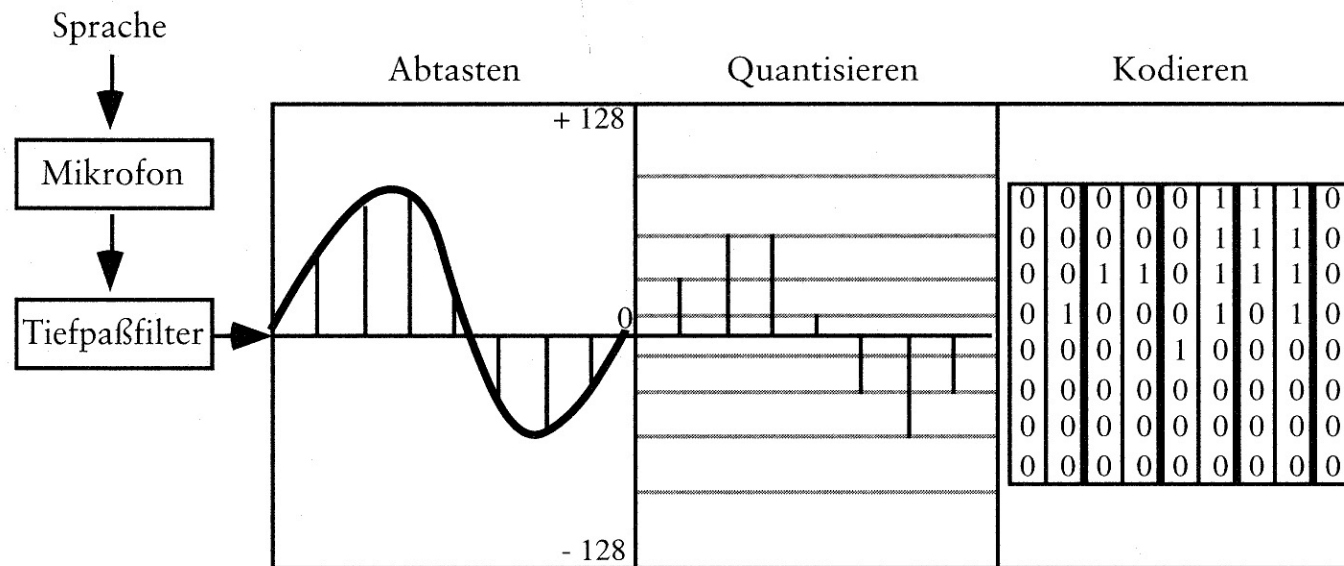
Audio - Streaming - System (1)

□ Gesamt-Überblick



Audio - Streaming - System (2)

□ Funktionsgruppen der Audio-Codierung (Vocoder)

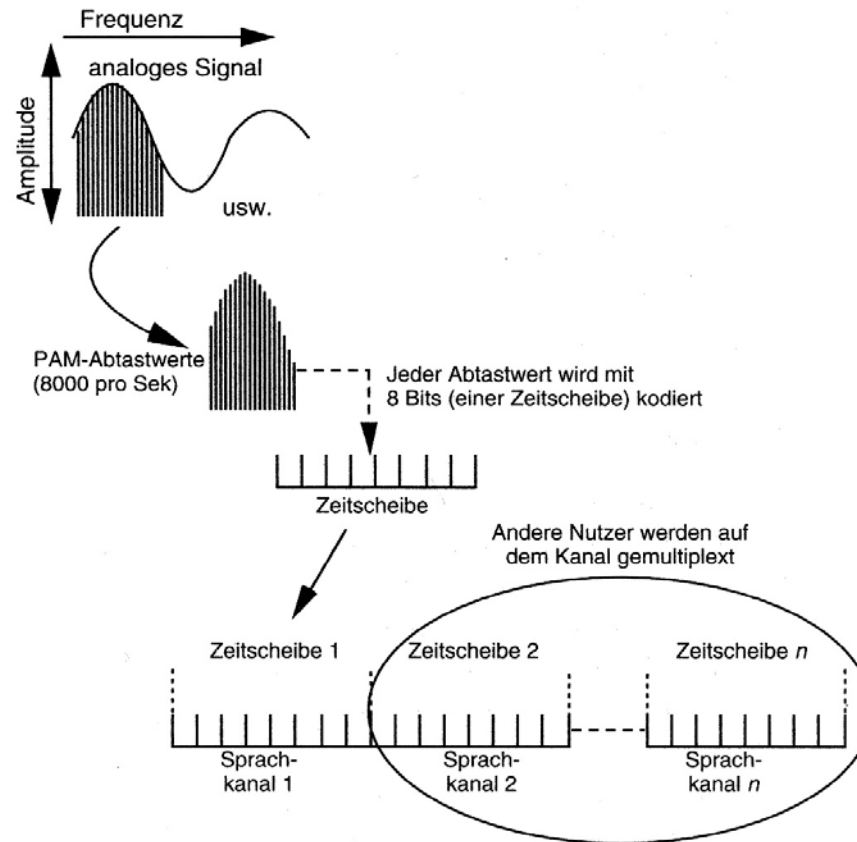


Sprachdigitalisierung (1)

Für die Codierung der Sprache bei VoIP gibt es verschiedene Verfahren

- ❑ Abtastwert-orientiert (Sample-based, Signalform-Codierung)
 - Abtastrate 8000 Hz, Kodieren der einzelnen Abtastwerte (samples) erzeugen i.d.R. eine Bitrate $\geq 16\text{Kbit/s}$
 - Beispiele PCM, ADPCM
- ❑ Segment-orientiert (frame-based)
 - Sprachsignal aufgezeichnet in Zeitsegmenten von 10 - 30 ms (Sprachsegmente)
 - Segment der Länge 10 ms entspricht 80 Abtastwerte, 20 ms ergibt 160 Werte
 - Jedes Segment wird im Sender analysiert, um die Parameter einer Empfängerschaltung zur Sprachsynthese zu bestimmen. Diese Parameter werden übertragen.
 - Beispiel: LPC, CELP mit Bitraten $< 16\text{ kbit/s}$

Sprachdigitalisierung (2)

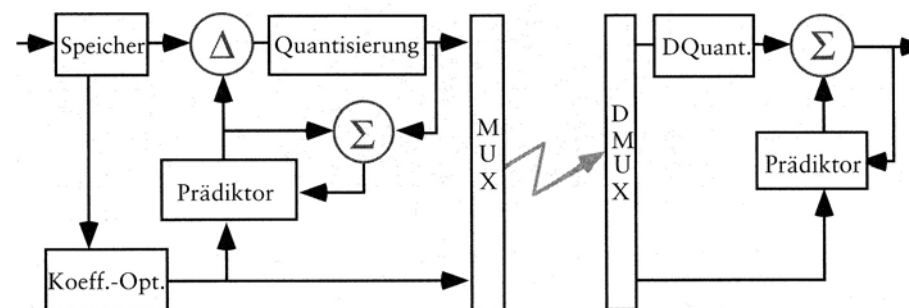


Sprachdigitalisierung (3)

- ❑ Bandbreite Telefonsprache: 300Hz – 3400Hz
Abtasttheorem (Nyquist): $f_s > 2 \times$ Grenzfrequenz
- ❑ Sampling : Diskretisierung der Zeitachse. Abtasten der stetigen Tonspannung $U(t)$ mit Frequenz f_s . $U(t)$ wird dadurch zur stückweise konstanten Funktion
 $\bar{U}(t) = U[\text{entier}(f \times t)/f]$ ersetzt
- ❑ Quantisierung: Diskretisierung der Tonspannung. Spannungsbereich (Dynamikbereich) wird in n Stufen eingeteilt, sodass n dem durch die Binärcodierung darstellbaren Zahlenbereich entspricht $\hat{U}(t) = \text{entier}[\bar{U}(t) - U_{\min}) / \Delta U$ mit ΔU Freiheit der Spannungsstufen.
- ❑ Codierung: \hat{U} wird durch Dualzahl kodiert

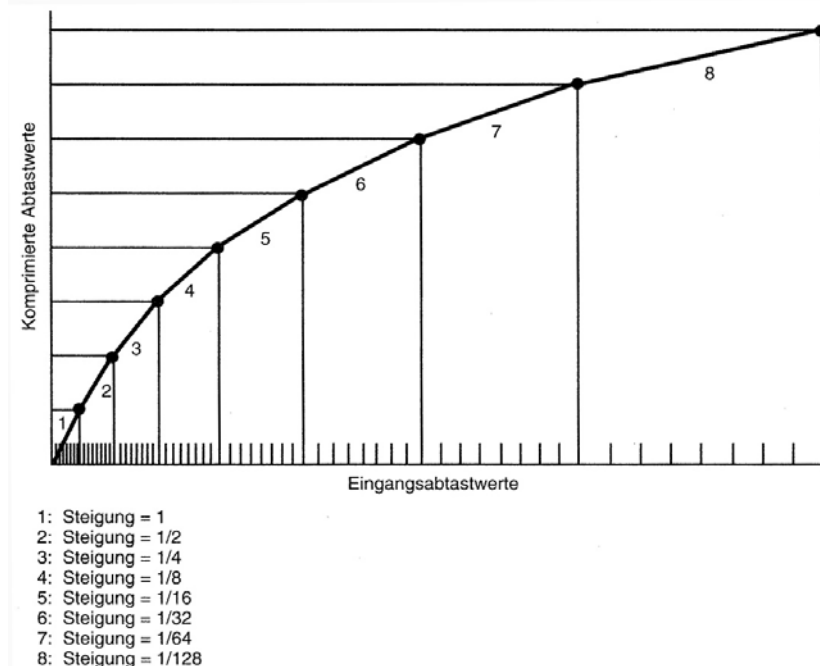
Sprachkodierer (1)

- ❑ Pulscodemodulation PCM (G.711)
 - Abtastrate $f_s=8\text{kHz}$, Abtastung alle $125\mu\text{s}$, Impulsdauer $5,2\mu\text{s}$, 256 Spannungsstufen entsprechen 8 Bit, somit benötigte Rate 64kbps
- ❑ Adaptive-Differential PCM
 - In einer 4 Bit-Zahl wird die Differenz des tatsächlichen Signals zu einem Vorhersagesignal kodiert. Der Schätzwert wird durch Extrapolation aus vorausgegangenen Signalwerten gewonnen. Dadurch Sprachübertragung mit 40, 32, 24, 16 kbps möglich



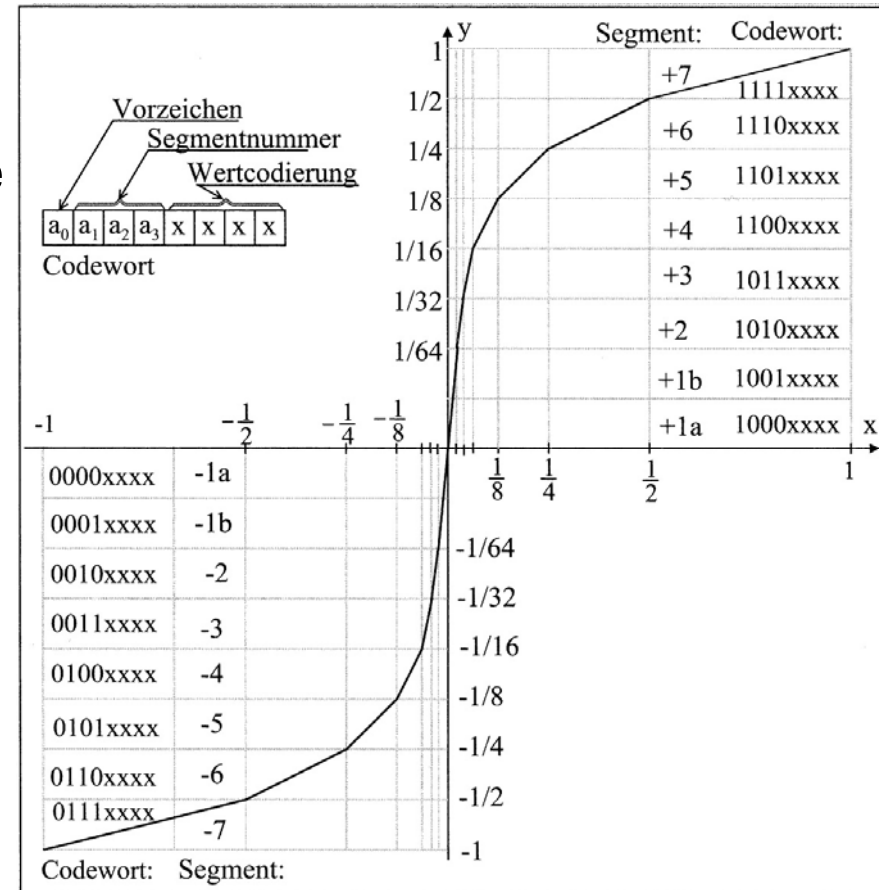
Sprachkodierer (2)

- Quantisierungsrauschen (Fehler durch Quantisierungsgranularität)
Idealerweise hätte man gerne viele Stufen. 2048 äquidistante Stufen reichen, aber benötigen $8000 \times 11 \text{ bps} = 88 \text{ kbps}$. Dynamik Telefon 30dB, somit ist es sinnvoll, variable Quantisierungsstufen vorzusehen. So z.B. bei μ -255 PCM



Sprachkodierer (3)

- ❑ Empfindlichkeit des Ohrs proportional zum $\log(\text{Lautstärke})$
- ❑ Verteilung Lautstärke Sprachsignale unregelmäßig
- ❑ Nichtlineare Quantisierung bei G.711 mit A-Kennlinie



Sprachkodierer (4)

- ❑ Breitband-Sprachkodierer G.722
ausgelegt für 7kHz Bandbreite und 64, 56, 48 kbps
- ❑ Rückwärts ADPCM, G.726/727
Schätzer nutzen die beiden letzten Ausgabewerte und die letzten 6 quantisierten Werte. Arbeitet mit 2, 3, 4 oder 5 Bit/Abtastwert.
Bedeutet Raten 16, 24, 32, 40 kbps
- ❑ Linear vorausschauende Analyse durch Synthese (LPAS). Sind musterbasierte Kodierer, die Koeffizienten und Pegel der Schätzfilter explizit übertragen, Abgleich mit LPC-Filter.
 - G.729 (8kbps), G.723.1 (5.3 und 6.3 kbps): Vorwärts adaptiv
 - G.728 (16kbps), LD-CELP: Rückwärts adaptiv

Sprachkodierer (5)

Standard	Kodierung	Bitrate Kbit/s	MOS	Kom- plexität	Verzögerung ms
G.711	PCM	64	4,3	1	0,125
G.726	ADPCM	32	4,0	10	0,125
G.728	LD-CELP	16	4,0	50	0,625
GSM	RPE_LTP	13	3,7	5	20
G.729	CSA-CELP	8	4,0	30	15
G.729A				15	
G.723.1	A-CELP	6,3	3,8	25	37,5
	MP-MLQ	5,3			
US Dod FS1015	LPC-10	2,4	synthetisch	10	22,5

Sprachqualität nach MOS-Skala (1)

- MOS (Mean Opinion Score) für Sprachqualität abhängig von
 - Verständlichkeit der Sprache
 - Akzeptanz der Lautstärke
 - Akzeptanz von Laufzeitschwankungen und Echos

MOS-Wert	Bedeutung
5 = excellent	keinerlei Anstrengung zum Verständnis der Sprache notwendig; totale Entspannung möglich
4 = good	keine Anstrengung notwendig, Aufmerksamkeit nötig
3 = fair	leichte, moderate Anstrengung nötig
2 = poor	merkbare, deutliche Anstrengung nötig
1 = bad	trotz Anstrengung keine Verständigung

Sprachqualität nach MOS-Skala (2)

□ MOS-Werte für Codierungsverfahren

Verfahren	Bitrate [kbit/s]	MOS-Wert
PCM	64	4.3 - 4.5
ADPCM	16/24/32/40	3.4/3.6/3.9/4.2
CS-ACELP	8/6.4	4.0/3.8
LD-CELP	16	4.0 - 4.1
ACELP	5.3	3.5
MP-MLQ	6.3	3.7

Paketvermittelte Sprache mit G.764 (1)

- ❑ Digitalisierte Sprache wird zur Übertragung in Pakete (Frames, Blöcke, Zellen) gepackt. Fehlerquellen sind
 - Diskretisierung/Quantisierung
 - Komprimierung
 - Paketierung
 - Paketverluste
 - Verletzung von Ratenerwartungen, Zeitbedingungen
- ❑ Synchronisationsfehler bei üblicher Sprache bereits ab 120ms wahrzunehmen, in Korrelation mit Video (Lippensynchronisation) schon ab 80ms, Jitter ab 20ms

Paketvermittelte Sprache mit G.764 (2)

□ Sprachpaketformat

HDLC Header-Typ	
Protokoll-Diskriminator	
Anzeige ausgelassener Blöcke	
Zeitstempel	
M/P-Bits	Kodierungstyp
Folge- nummer (SN)	Rauschen
nicht auslaßbare Blöcke	
optional auslaßbare Blöcke	
Übertragungsrahmen-Prüfsequenz	

wobei:
M/P = Mehr-/Poll-Bits

Paketvermittelte Sprache mit G.764 (3)

□ Sprachpaketformat-Felder

- HDLC Header Typ (Adressfeld DLCI, Rahmentyp UI, CRC)
- Protokoll: 010001000
- Anzeige auslassbarer Blöcke: zeigt an, wieviel für Auslassen zulässig sind: 00 (keiner), 01 (einer), 10 (zwei), 11 (drei). Zeigt jeweils Resterlaubnis an
- Zeitstempel: aufsummierte Verzögerung im Netz <200ms
- Kodierungstyp: zeigt spezielle A/D-Technik an
- M-Bit: für alle Pakete eines Sprachintervalls (außer letztem) = 1
- Sequenznummer: für Verlufterkennung und mit Zeitstempel zur Synchronisation
- Rauschpegel: muss wegen Digital Speech Interpolation DSI in Paketpausen wiedergegeben werden

Paketvermittelte Sprache mit G.764 (4)

□ Sprachpaketformat: Kodierungstypen

Kodierungstyp	Anzahl Blöcke
8 Bit/Abtastwert	8
1 Bit/Abtastwert	1
2 Bit/Abtastwert	2
3 Bit/Abtastwert	3
4 Bit/Abtastwert	4
5 Bit/Abtastwert	5
6 Bit/Abtastwert	6
7 Bit/Abtastwert	7
8 Bit PCM (A-Law oder μ -Law)	8
2 Bit/Abtastwert ADPCM	2
3 Bit/Abtastwert ADPCM	3
4 Bit/Abtastwert ADPCM	4
5 Bit/Abtastwert ADPCM	5
(4,2) eingebettete ADPCM	4
(5,2) eingebettete ADPCM	5
(8,6) eingebettete ADPCM	8

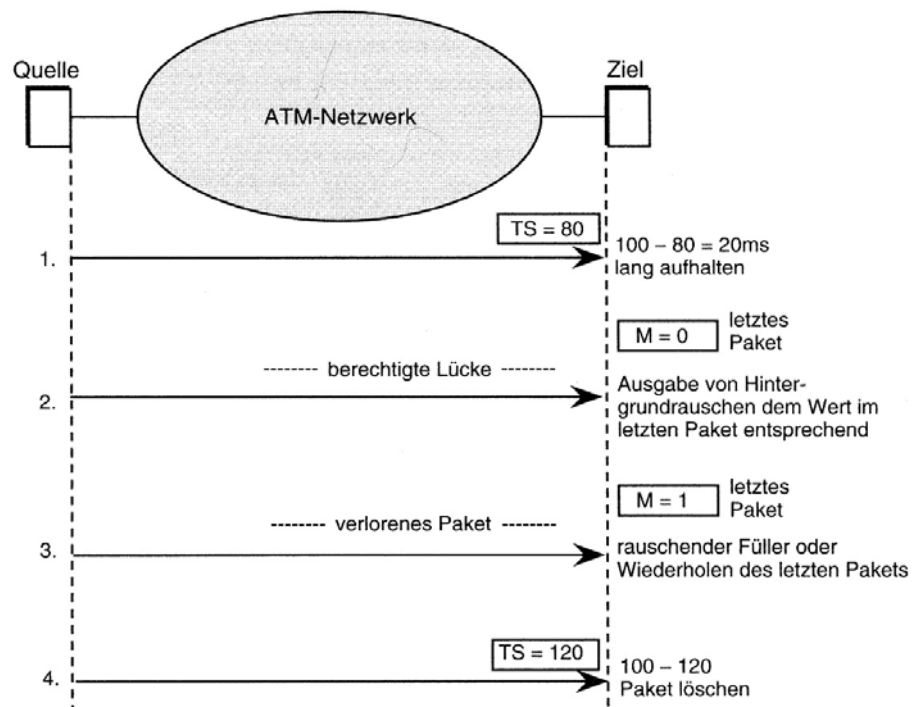
Paketvermittelte Sprache mit G.764 (5)

□ Sprachpaketformat: Rauschfeldformat

Bitnummer	Geräusch-Level
4321	(dBmC0)
0000	Leerlauf-Code
0001	16,6
0010	19,7
0011	22,6
0100	24,9
0101	26,9
0110	29,0
0111	31,0
1000	32,8
1001	34,5
1010	36,2
1011	37,9
1100	39,7
1101	41,6
1110	43,8
1111	46,6

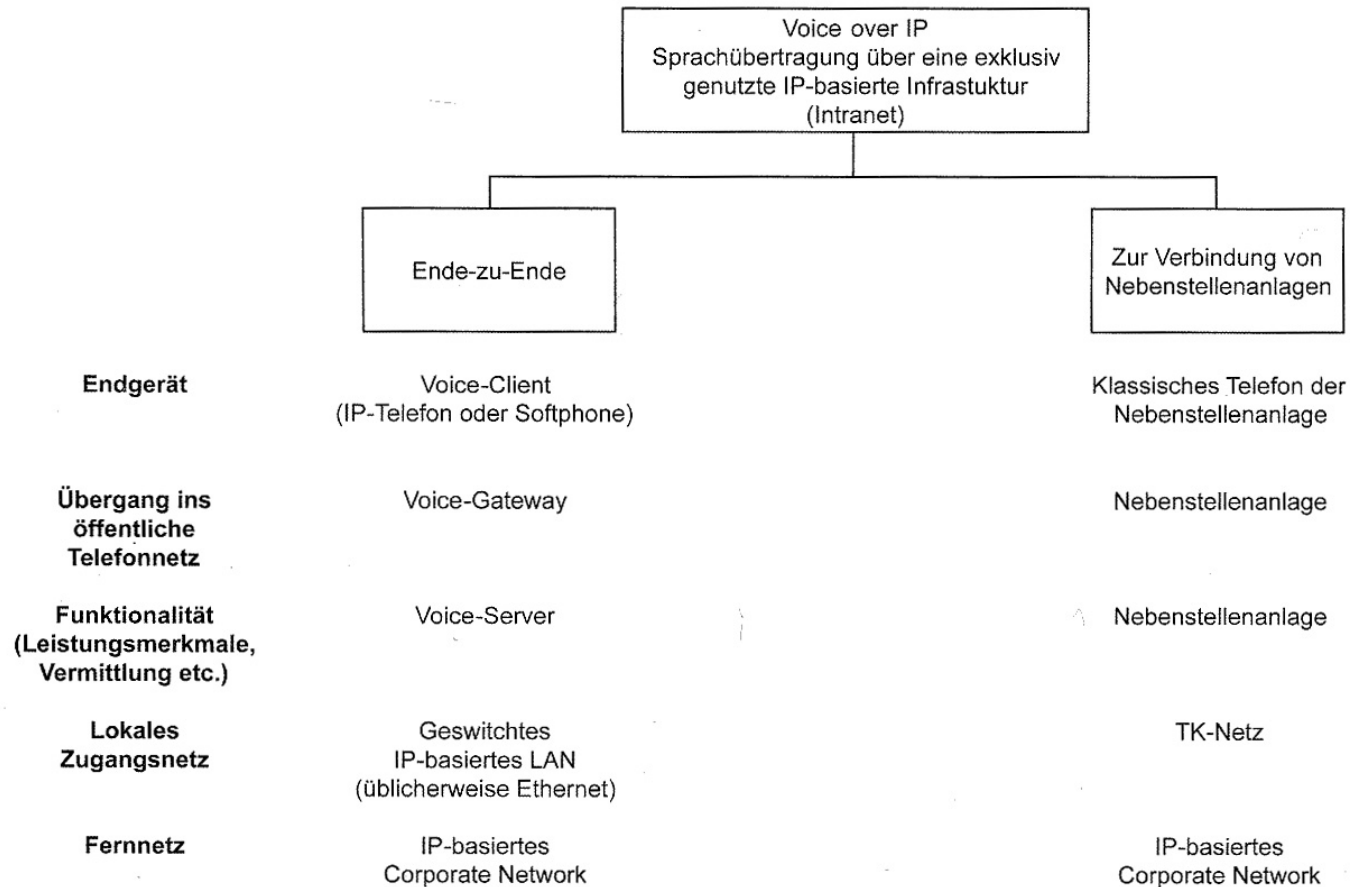
Paketvermittelte Sprache mit G.764 (6)

- Empfänger prüft Zeitstempel gegenüber z.B. Constant Delay Value
z.B. CDV=100 und ATM-Netz



Voice over IP (1)

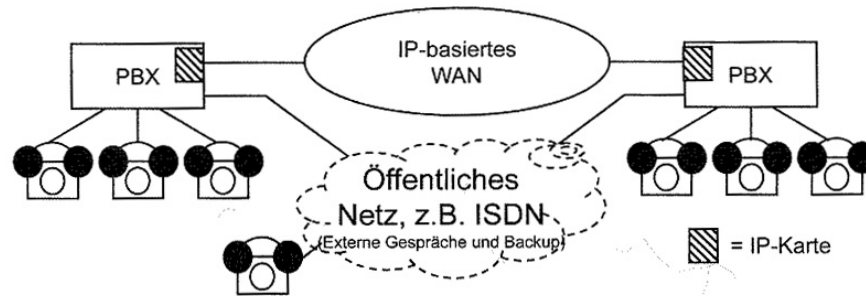
Überblick über verschiedene Arten von Voice over IP



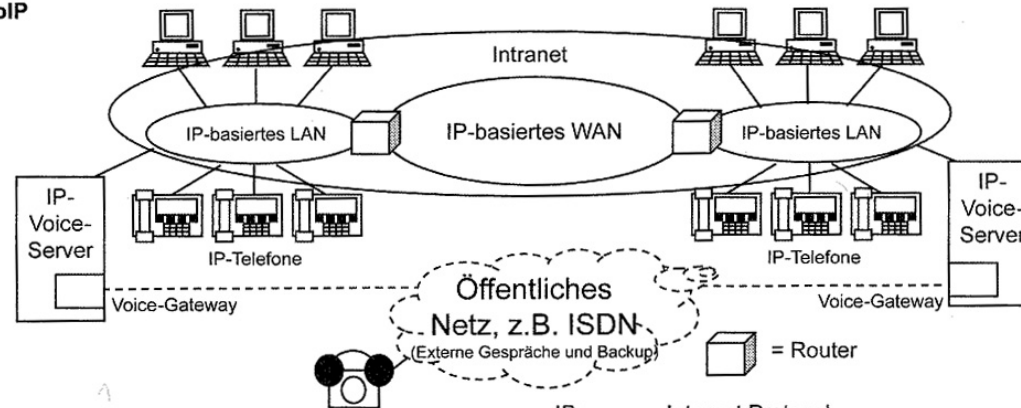
Voice over IP (2)

Technische Infrastruktur für Voice over IP

Verbindung von Nebenstellenanlagen



Ende-zu-Ende VoIP



- IP: Internet Protocol
- ISDN: Integrated Services Digital Network
- LAN: Local Area Network
- PBX: Private Branch Exchange (Nebenstellenanlage)
- VoIP: Voice-over-IP
- WAN: Wide Area Network

Voice over IP (3)

Komponenten einer VoIP-Infrastruktur

- Voice Client (IP-Phone): enthält Vocoder, erzeugt Sprachpakete
- Voice Server: i.a. PC-Programm, das eine Nebenstellenanlage simuliert, mit i.a. mehreren 100 parallelen Sessions

IP-Telephone-Gateway: verbindet VoIP-Welt mit TK-Welt

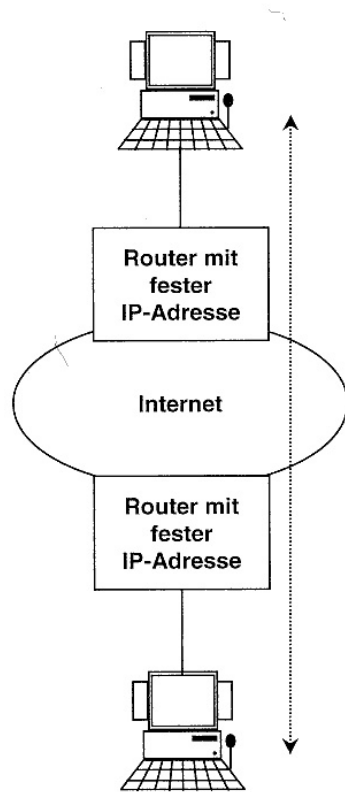
- Umsetzen Rufnummern/Adressen (ENUM-Konzept RFC 2916)
- Sicherstellen von Leistungsmerkmalen / Signalisierung QSIG
- Umrechnen Accounting
- Umkodieren von Codec-Formaten

IP-Netz: sollte QoS-fähig sein (Jitter, Delay, Bandbreite)

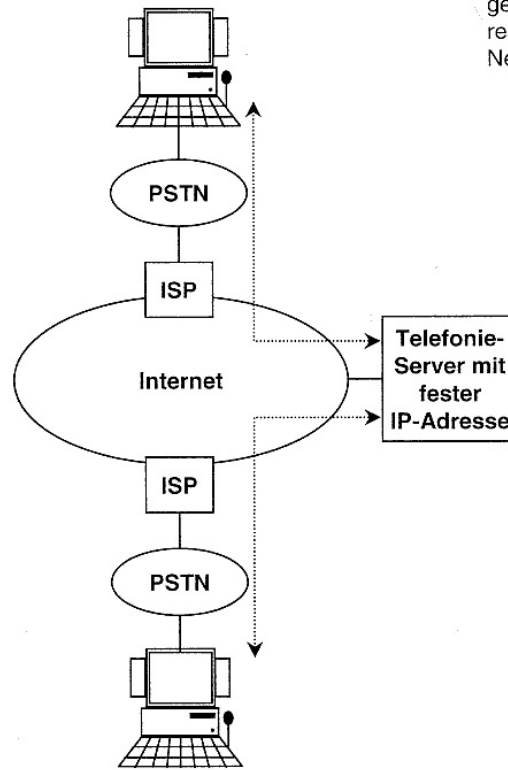
Voice over IP (4)

Varianten der Ende-zu-Ende Internet-Telefonie

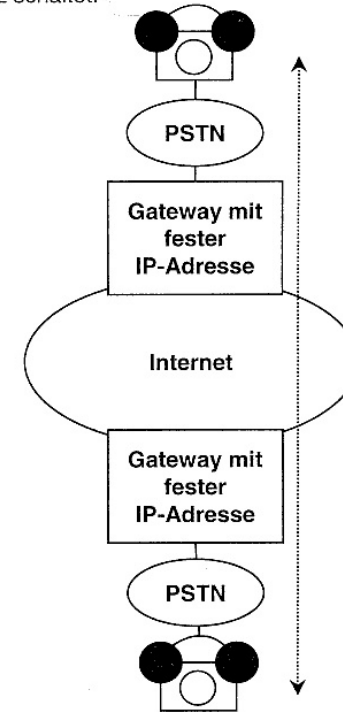
Wenn feste IP-Adressen vorhanden sind: Etablierung der Verbindung direkt durch das Internet



Wenn keine festen IP-Adressen vorhanden sind: Etablierung der Verbindung über einen Server mit permanenter IP-Adresse



Per Telefon kann über das öffentliche Netz ein Gateway angewählt werden. Es übersetzt die Telefonnummer des gewünschten Verbindungspartners in eine IP-Adresse des zum Partner nächstgelegenen Gateways, welches die restliche Verbindung durch das öffentliche Netz schaltet.



Voice over IP (5)

- Wichtiges Normungspaket: H.323 bzw. SIP
- Codecs

Codec	G.723.1	G.729	G.729 A
Bitrate [kbps]	5.3. , 6.3	8	8
Ü-Rahmengröße [ms]	30	10	10
Verarbeitungsverzögerung [ms]	30	10	10
Vorschauverzögerung [ms]	7.5	5	5
Rahmenlänge [Byte]	20 / 24	10	10

Voice over IP (6)

□ Probleme

- Adressverwaltung für angemeldete Clients, temporäre IP-Adressen
- Suchfunktionen über LDAP
- Nichterreichbarkeit, falls Client (z.B. PC) offline
- unzureichende Interoperabilität wegen heterogener Systeme
- Sicherheit (Access Control)
- Bandbreiten (→ gute Komprimierungsverfahren)
- Jitter (→ Ausgleich über Puffern beim Empfänger)
- Verschiedene Signalisierungsverfahren
- Handling der Software
- Ausfallsicherheit geringer
- Datenschutz geringer

Protokolle für VoIP (1)

- ❑ Protokolle für die Sprachübermittlung
 - Sprachkommunikation verläuft in Echtzeit, daher RTP und RTCP
 - Neues Internet-Transportprotokoll SCTP (Stream Control Transmission Protocol, RFC2960)
- ❑ Signalisierungsprotokolle

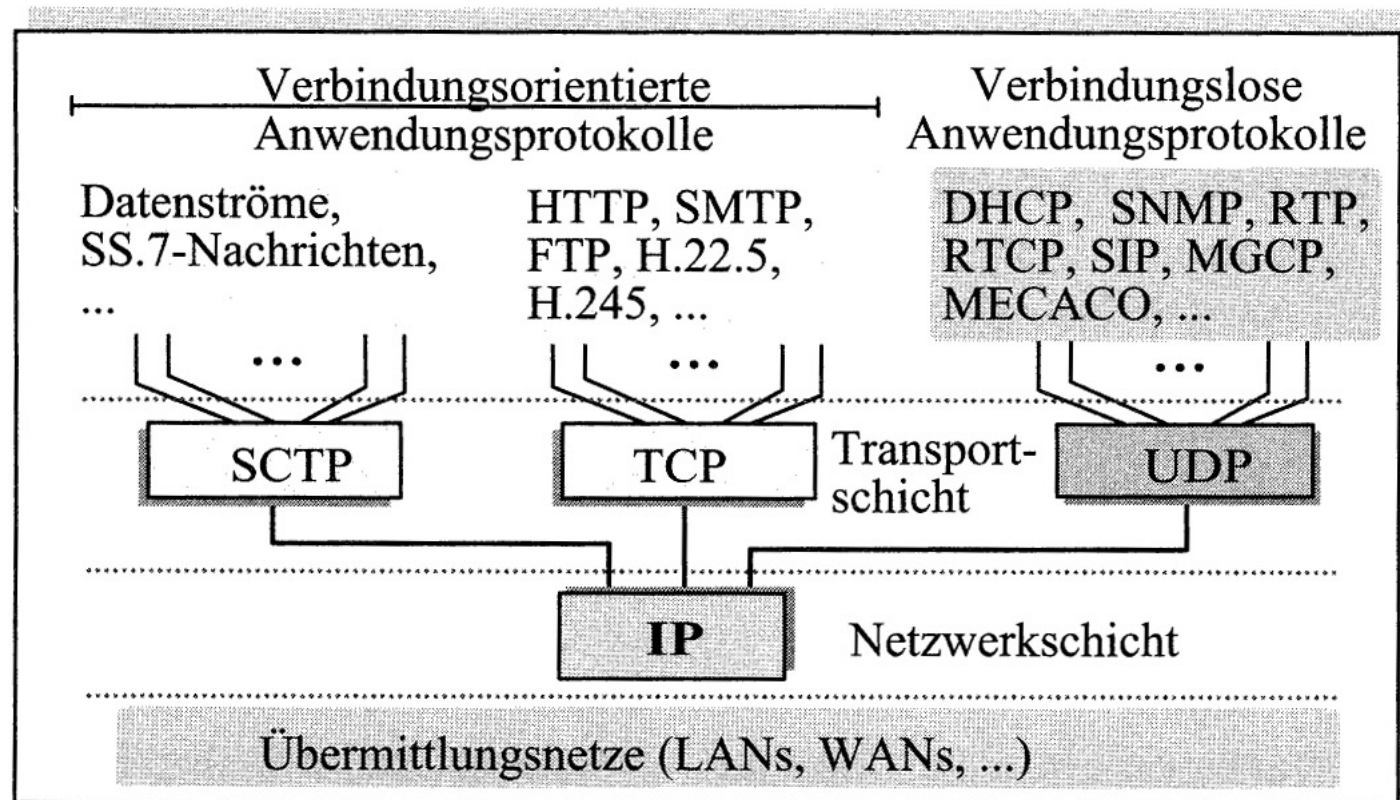
d.h. für Aufbau / Abbau von Verbindungen zwischen IP-Telefonen oder VoIP-Komponenten

 - Protokolle H.225 und H.245 aus der Standardfamilie H.323 „Packet-based multimedia communication systems“ (ITU-T)
 - Protokoll SIP (Session Initiation Protocol) der IETF
- ❑ Protokoll zur Steuerung von Media Gateways

d.h. zum Übergang zwischen VoIP und ISDN bzw. POTS oder TK-Anlagen sowie Mobilfunknetzen

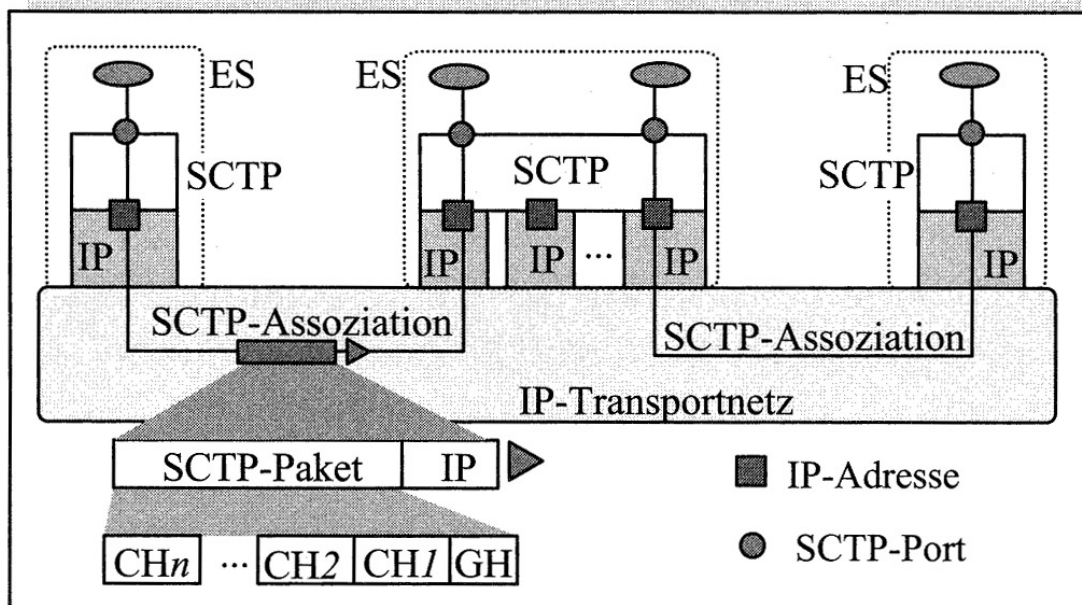
 - MGCP (Media Gateway Control Protocol), RFC2705 / 3535)
 - Megaco (Media Gateway Control, RFC3015)

Protokolle für VoIP (2)



Protokolle für VoIP (3): SCTP

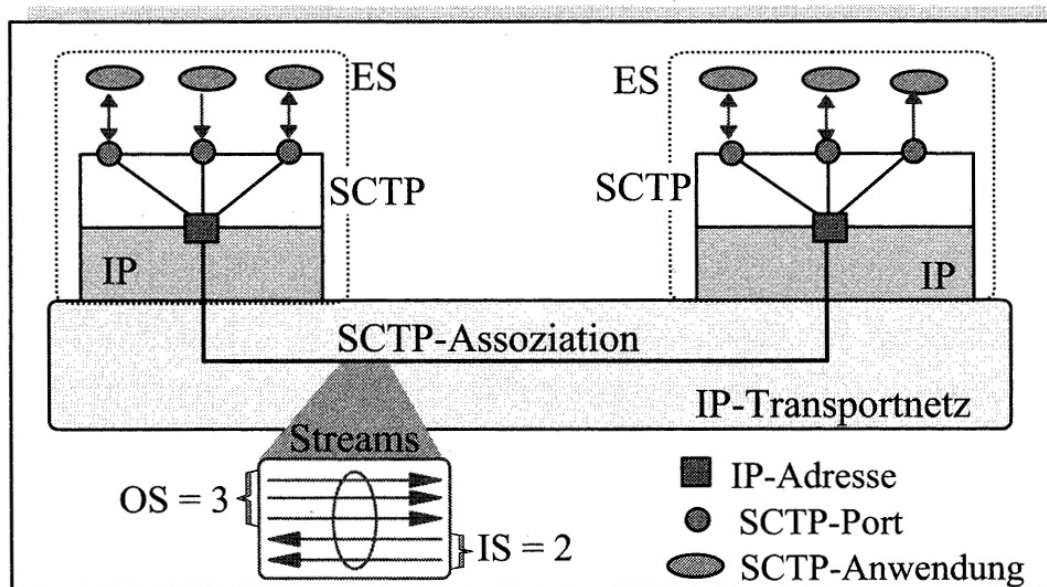
- ❑ SCTP wurde als Ergänzung von TCP entwickelt zum Transport von SS7-Nachrichten und Dateiströmen, RFC2960
- ❑ SCTP ermöglicht die gesicherte Übertragung mehrerer SCTP-Streams über SCTP-Assoziationen



ES: Endsystem
CH: Chunk
GH: Gemeinsamer Header

Protokolle für VoIP (4): SCTP

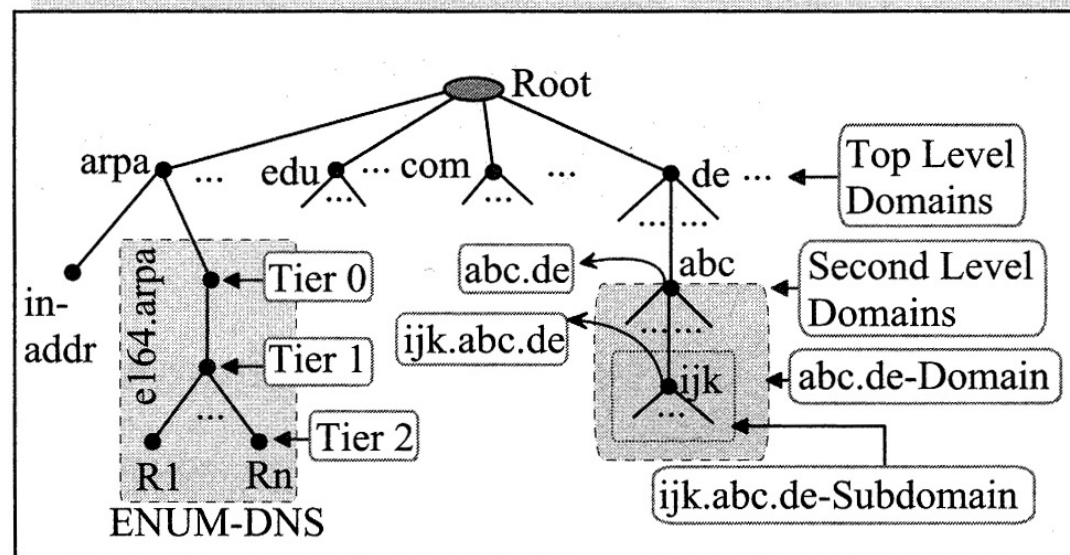
- ❑ SCTP-Streams entsprechen unidirektionalen virtuellen Verbindungen (Inbound / Outbound Streams)
- ❑ SCTP-Nachrichten werden in DATA-Chunks transportiert mit Stream Identifier



ES: Endsystem
IS: Inbound Stream
OS: Outbound Stream

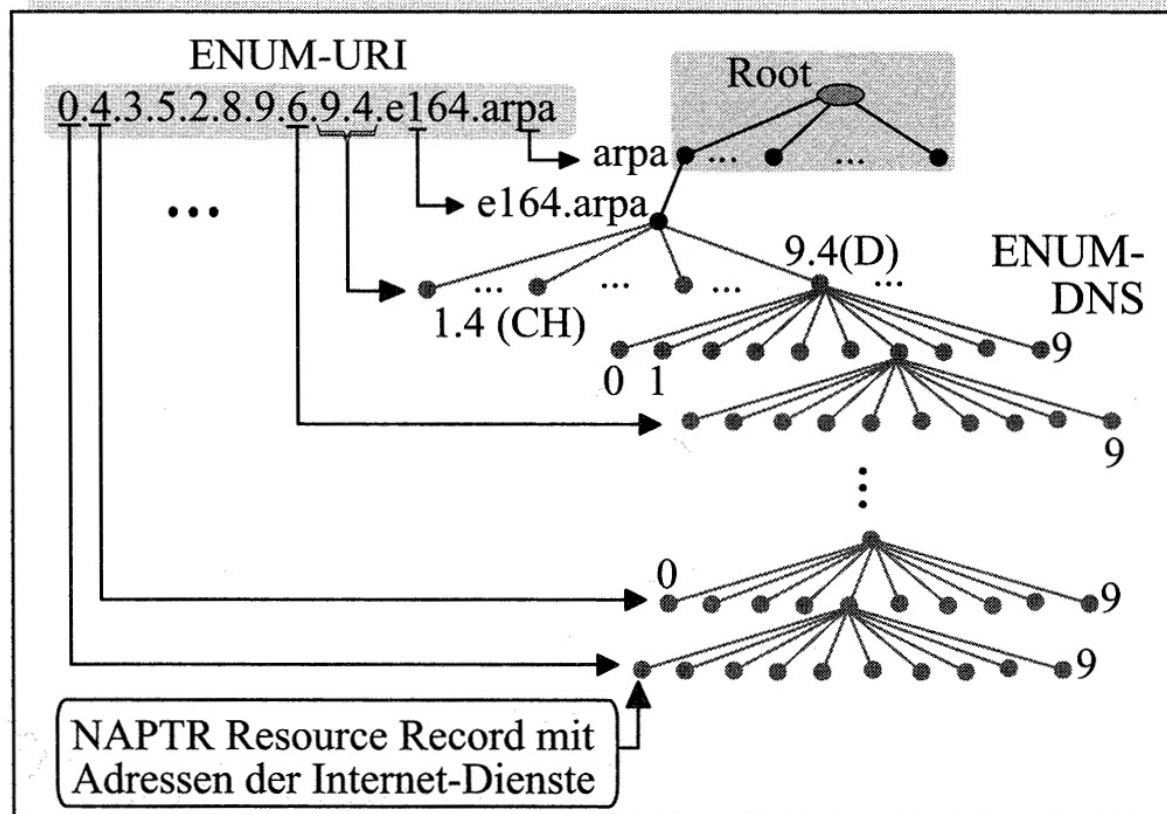
VoIP: ENUM (1)

- ❑ ENUM (RFC2916) dient der Adressumsetzung zwischen Internet-Welt und Telefon-Welt
- ❑ ENUM definiert die Umwandlung einer Telefonnummer gemäß ITU-T-Standard E.164 in einen Uniform Resource Identifier ENUM-URI
- ❑ ENUM verwendet DNS und dienstspezifische Domain e164.arpa. In den Servern unterhalb ENUM-DNS werden sog. Naming Authority Pointer Resource Records (NAPTR RR) abgespeichert, die Infos über die einer E.164-Nr zugeordneten Internet-Dienste enthält



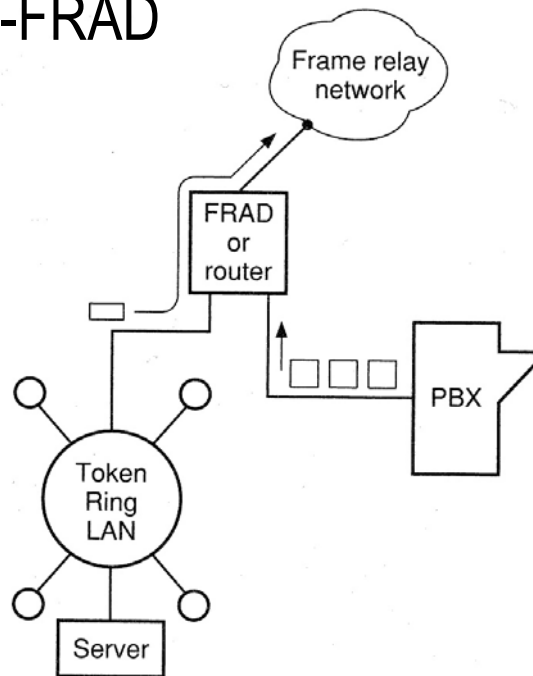
VoIP: ENUM (2)

□ Beispiel: E.164 Rufnummer +49-69-82534-0



Voice over FR (1)

- Vom FR Forum Technical Committee gibt es Dokument FRF.11 (Voice over FR Implementation Agreement), definiert für Sprachübertragung spezielles Gateway-FRAD



Voice over FR (2)

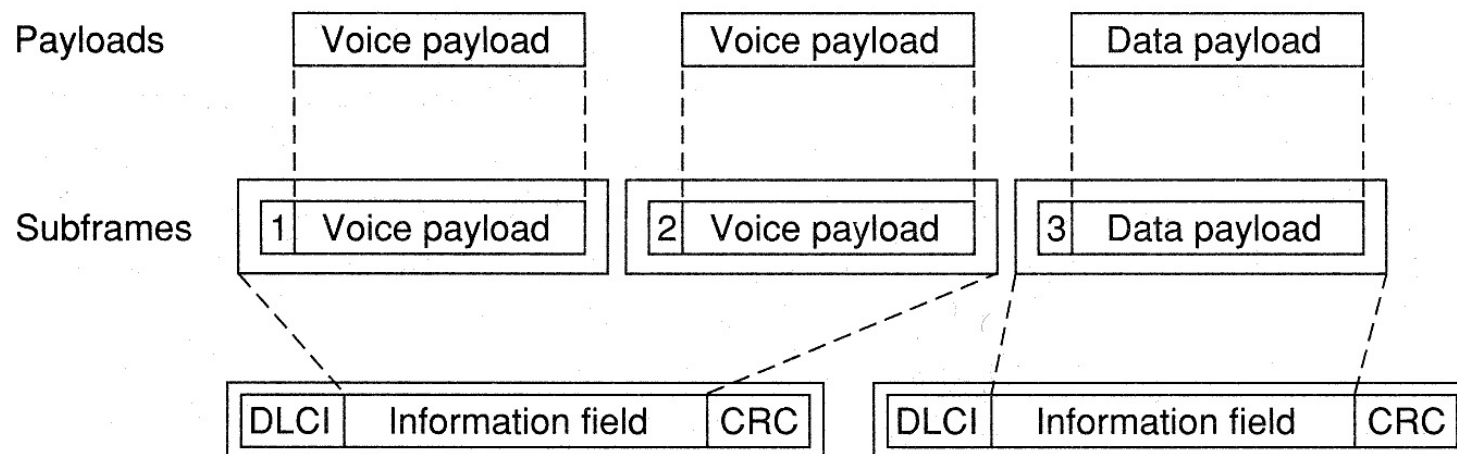
□ Technische Problemfelder

- Frame Länge: bestimmt in Abhängigkeit von Access Rate die Verzögerung
- Frame Priorisierung: Voice over Data
- Frame Verlust: Möglichkeiten generate period of silence oder generate speech for missing interval (DSI)
- Frame Delay und Jitterkompensation: Zwischenpuffern beim Empfänger unter Ausnutzen von Zeitstempeln
- Echo Cancellation
- Silence Suppression
- Voice Compression
- Telephony Signaling: Abbildung auf PBX-Signalisierung erforderlich
- Multiplexing: sowohl Logical Link Mux als auch Subchannel Mux

Voice oder FR (3)

- Ausnutzen von Subframes für Mux

Beispiel: über eine DLCI-Connection sollen zwei Audio- und ein Datenkanal gemuxt werden (bis zu 256 auf einem DLCI)



Voice over FR (4)

□ Subframe Format

Bit position	8	7	6	5	4	3	2	1
Octet 1	EI	LI	Subchannel identifier (CID)					
Octet 1a	CID		Spare	Spare	Payload type			
Octet 1b	Payload length							
Octet <i>n</i>	Payload							

Notes:

- When the EI bit = 1, the structure of octet 1a's payload type field applies as follows:

Bits	4	3	2	1	
	0	0	0	0	Primary payload transfer
	0	0	0	1	Dialed digit transfer
	0	0	1	0	Signaling bit transfer
	0	0	1	1	Fax relay transfer
	0	1	0	0	Silence information description

- When the LI bit = 1, the structure of octet 1b applies.

Voice over FR (5)

- Komposition von Subframes innerhalb Frame in Abhängigkeit von EI und LI

