VoIP

VOICE OVER IP

OVERVIEW OF VOICE OVER IP TECHNOLOGIES, NETWORK ARCHITECTURES AND PROTOCOLS

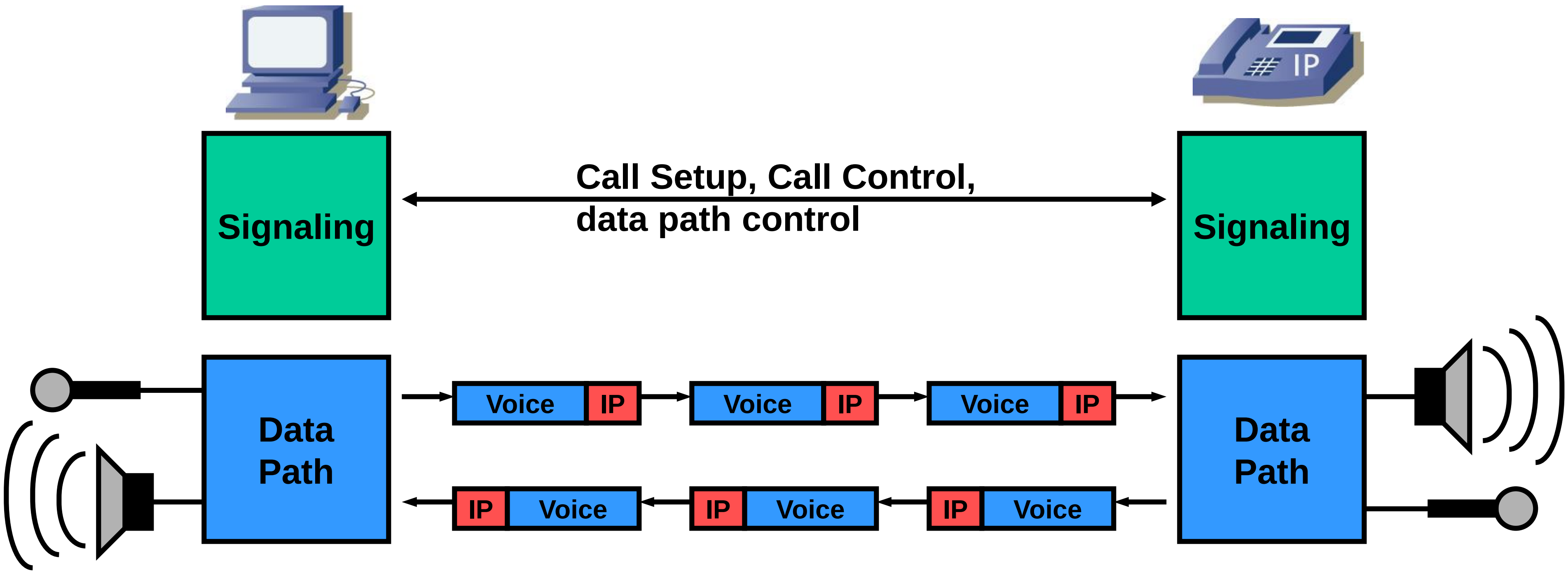
Peter R. Egli
peteregli.net

Contents

1. VoIP functions

➔ Signaling comprises all functions to set up, control and teardown a VoIP call/session. Examples of VoIP signaling protocols: H.323, SIP, MGCP, H.248, NCS, Skype. UDP and TCP are used for signaling transport.

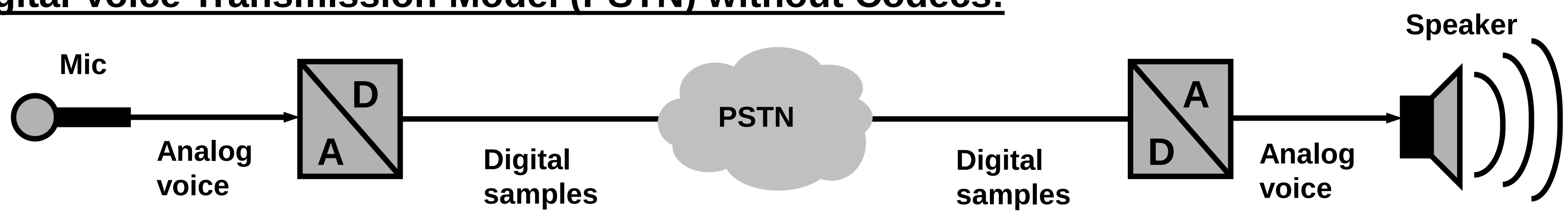
➔ The data path is responsible for encoding, packetizing and compressing the voice. UDP is always used for the data path since:
a. TCP would introduce too much delay and
b. Retransmissions are not necessary and only distort the voice in case of packet loss.



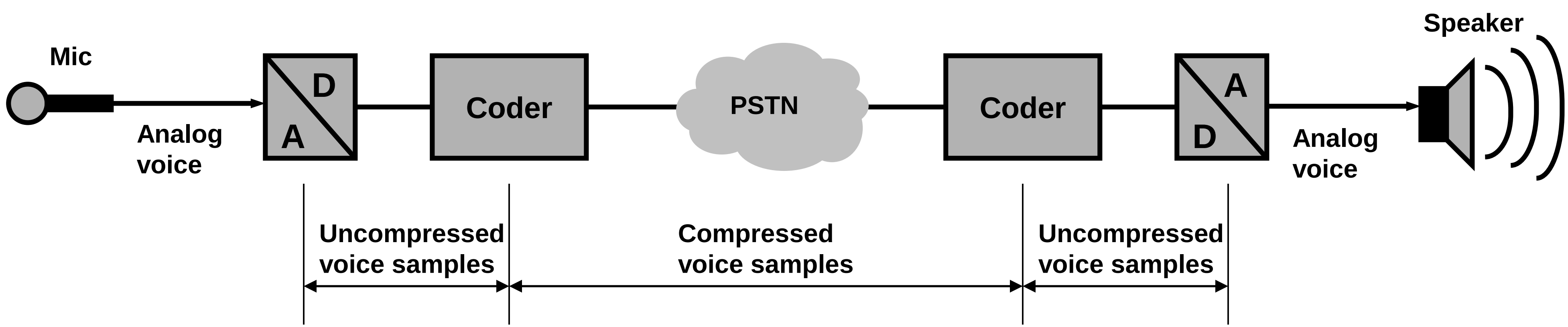
2. Voice Codecs (1/12)

➔ **Codec means Coder Decoder. Coding means encoding the (already digitized) voice samples into a different format, e.g. for compression (reduction of data rate).**

- **Digital Voice Transmission Model (PSTN) without Codecs:**



- **Digital Voice Transmission Model (PSTN) with Codecs:**

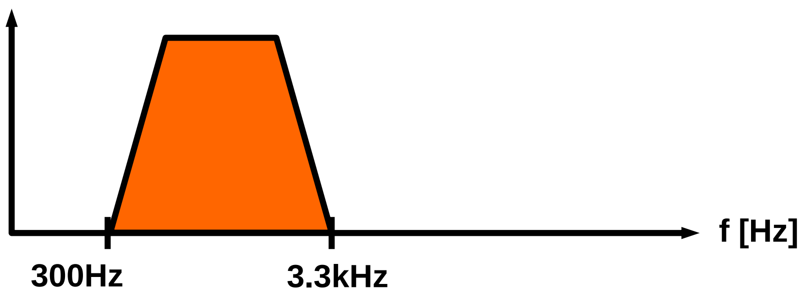


2. Voice Codecs (2/12)

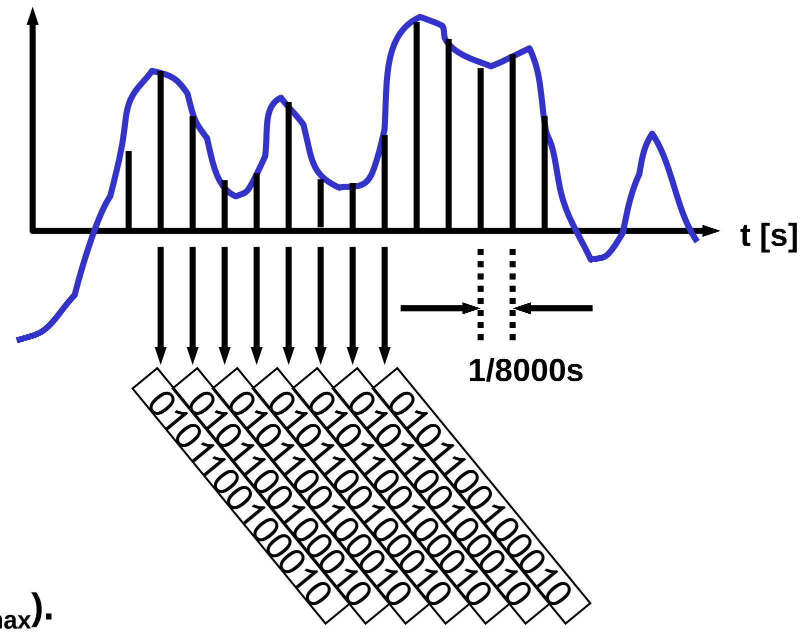
➔ Pulse Code Modulation PCM:

Sample (measure) amplitude at equal time intervals and encode the amplitude as digital value.

POTS (analog) signal in frequency domain:



POTS (analog) signal in time domain:



Sampling:

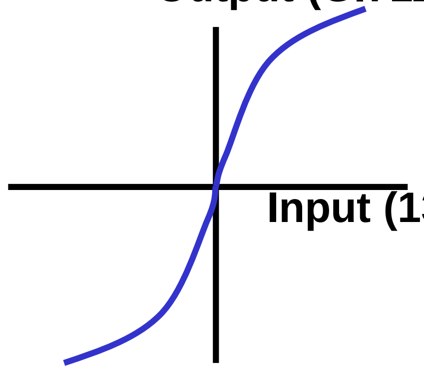
The analog signal is sampled at aequi-distant time intervals.
The sampling frequency must be at least double the highest signal frequency (Nyquist theorem: sampling frequency >= 2^*f_{max}).
This means the sampling frequency must be 2*3.3kHz ~= 8kHz.

Quantization of samples:

The samples are digitized (A/D converter) which results in a stream of 13 (A-law) or 14 (µ-law) bit samples (voice over analog lines requires >12 bits due to ~60dB dynamics = power range).

2. Voice Codecs (3/12)

➔ **G.711 Codec:**



The G.711 Codec performs compansion (COMPression and ExpANSION) for reducing the data rate and amplify weak signals in order to increase S/N ratio:
➔ Reduction of 13 (A-law) and 14 (µ-law) bits to 8 bits according to a non-linear compression curve:
1. Step: Raise power of weak signals
2. Step: Linear quantization

➔ A-law and µ-law differ in the compansion curve.

➔ G.711 is the standard codec used in PSTNs.

➔ Sampling:
8kHz sampling rate, 8bits / sample ➔ 64kbps channels.

2. Voice Codecs (4/12)

➔ **Voice Compression Codecs:**

➔ **Purpose of compression: bandwidth reduction.**

➔ **Voice / speech contains a lot of redundancy (same information contained multiple times); lossy codecs can remove this redundancy whithout reducing the voice quality too much (lossy = reconstructed signal at receiver != signal at sender before transmission).**

➔ **Compressing voice codecs use principles like (examples):**
a. Masking of tones: If 2 tones have almost the same frequency then only the louder tone is audible. Compression removes the masked tone information.

b. Only transmit difference between 2 subsequent voice samples (Differential PCM).

Toll quality: Quality good enough to charge money for the service: MOS 4-5; communication quality: MOS 3-4; synthetic quality: MOS < 3
PCM: Pulse Code Modulation
ADPCM: Adaptive Differential PCM
ACELP: Algebraic Codebook Excited Linear Prediction
LD-CELP: Low Delay CELP
MP-MLQ: MultiPulse-Maximum Likelihood Quantization
CS-ACELP: Conjugate Structure ACELP
DSP: Digital Signal Processor
MIPS: Million Instructions Per Second
VAD / DTX / CNG: Voice Activity Detection / Discontinuous Transmission / Comfort Noise Generation

2. Voice Codecs (5/12)

➔ Overview of codecs (1):

	G.711	G.721	G.722	G.723 .1	G.723.1	G.726	G.727	G.728	G.729A
Date	1972	1984	1988	1995	1995	1990	1990	1992	1995
Toll-quality	Yes	Yes	Yes	Near toll	No	Near toll	N/A	yes	Near toll
MOS	4.20	4.00	N/A	3.65	3.90	3.85	4.00	3.61	3.70
Bit rate [kbps]	64	32	48 / 56 / 64	5.3	6.3	16/24/32/40	16/24/32/64	16	8
Audio bandwidth	3.4kHz	3.4kHz	7kHz	3.4kHz	3.4kHz	3.4kHz	3.4kHz	3.4kHz	3.4kHz
VBR	No	No	No	No	No	No	No	No	No
Algorithm	a-law/u-law	ADPCM	SB-ADPCM	ACELP	MP-MLQ	ADPCM	ADPCM	LD-CELP	CS-ACELP
Algorithmic delay [ms]	0.125	20	1.5	30	30	0.125	0.125	0.625 - 2.5	10
Lookahead delay [ms]	0	N/A	N/A	7.5	7.5	0	N/A	N/A	5
Voice frame size	sample	sample	?	20bytes	24bytes	Sample	Sample	0.625ms	10bytes
Complexity [DSP MIPS/RAM/ROM]	0.1MIPS 2w RAM 50w ROM	10MIPS 256w RAM 4kw ROM	10MIPS 256w RAM 4kw ROM	18MIPS 2.1kw RAM 7kw ROM	16MIPS 2.1kw RAM 7kw ROM	12MIPS 256w RAM 12kw ROM	12MIPS 256w RAM 12kw ROM	33MIPS 3.4kw RAM 8kw ROM	22MIPS 2.5kw RAM 9.5kw ROM
Pass fax/modem	Yes	No	No	No	No	No	No	yes	No
Tandeming	14	4	3	1	2	4	4	3	2
Packet loss tolerance	n.a.	N/A	N/A	< 3%	< 3%	N/A	N/A	N/A	< 5%
PLC	Yes (annex I)	N/A	No	Yes	Yes	N/A	N/A	N/A	Yes
Bit-robustness	Yes	Yes	Yes	N/A	N/A	Yes	Yes	N/A	Yes
VAD / DTX / CNG	No	No	No	Yes	Yes	No	No	No	Yes
# of patents / # of patent holders	N/A	N/A	N/A	~ 18 / 8	N/A	N/A	N/A	N/A	~ 20 / 9
License	No	N/A	N/A	Yes	Yes	N/A	N/A	N/A	N/A
Application	ISDN	Obsolete	VoIP	VoIP	VoIP	VoIP	VoIP	PSTN	VoIP
Comments	Standard high qualitiy VoIP codec.	-	Audio encoder	Standard low bit rate VoIP codec.	Standard low bit rate VoIP codec.	-	Embedded version of G.726	-	Std. medium Q/bit rate VoIP codec.

2. Voice Codecs (6/12)

➔ Overview of codecs (2):

	GSM EFR	Speex	iLBC (RFC3951)	AMR-NB	G.719
Date	1996	2003?	2004	1999	2008
Toll-quality	Near toll	Near toll	Near toll	Yes	Yes
MOS	3.5 - 3.9	n.a.	3.4 - 4	3.79 - 4.14	N/A
Bit rate [kbps]	12.2	2 - 44kbps	15.2kbps or 13.3kbps	4.75-12.2	32 - 128
Audio bandwidth	3.4kHz	N/A	3.4kHz	3.4kHz	20Hz - 20kHz
VBR	No	Yes	N/A	Yes	Yes
Algorithm	CD-ACELP	CELP	LPC	ACELP	Adaptive time resolution etc.
Algorithmic delay [ms]	20	30ms (@ 8kHz s. rate)	25ms (15.2), 40 (13.3)	20ms	40ms (end-to-end)
Lookahead delay [ms]	N/A	10ms (@ 8kHz)	5ms (20ms frame size)	0 (@ 12.2kbps)	N/A
Voice frame size	22.5ms	N/A	30ms (13.3), 20ms (15.2)	20ms	20ms
Complexity [DSP MIPS/RAM/ROM]	15.4MIPS 4.7kw 5.9kw	Variable	22 MIPS	~7MIPS	18 floating point MIPS
Pass fax/modem	no	N/A	N/A	N/A	N/A
Tandeming	2	N/A	N/A	N/A	N/A
Packet loss tolerance	n.a.	10%	Very good	N/A	N/A
PLC	Yes	Yes	Yes	Yes	N/A
Bit-robustness	Yes	N/A	N/A	Yes	N/A
VAD / DTX / CNG	VAD / DTX / CNG	VAD / DTX	N/A	VAD / DTX / CNG	N/A
# of patents / # of patent holders	2 / 2	0 / 0 (open source)	N/A	N/A	2
License	N/A	Free	Free	N/A	2 (Polycom & Ericsson)
Application	GSM	VoIP, online game voice comm.	VoIP	GSM	VoIP, voice mail
Comments	-	Open source, well suited for VoIP; RFC	-	-	Based on Polycom Siren22, G.722.1

2. Voice Codecs (7/12)

➔ Overview of codecs (3):

Toll quality:	**Quality for which a toll can be reclaimed.**
MOS:	**Mean Opinion Score (quality measure).**
VBR:	**Variable Bit Rate.**
Algorithmic delay:	**Delay of voice codec algorithm.**
Lookahead delay:	**Delay introduced by codec by „looking“ into the following voice frame.**
Voice frame size:	**Number of bytes per voice frame (usually a codec processes a voice frame and sends the data as a packet).**
Complexity:	**Measure for the complexity of the codec (required processing resources in terms of DSP MIPS, RAM, ROM).**
Pass fax/modem:	**Ability to pass analog signals like modem and fax.**
Tandeming:	**Number of codecs in a row.**
Packet loss tol.:	**Impact of packet loss to speech quality.**
PLC:	**Packet Loss Concealment (ability to hide packet loss, e.g. by re-playing the last good packet).**
Bit-robustness:	**Ability to conceal bit errors (important on wireless links where the bit error rate is higher than on wire-based links.**
VAD / DTX / CNG:	**Voice Activity Detection, Discontinuous Transmission, Comfort Noise Generation.**

2. Voice Codecs (8/12)

➔ **Codec technology (1):**

Waveform coders:

- ➔ Send directly voice samples or sample differencies.
- ➔ Background noise is also coded and sent to receiver.
- ➔ Such coders usually provide high voice quality.
- ➔ High bit rate (>16kbps).
- ➔ Waveform coders work in the time domain.

Vocoding:

- ➔ The encoder builds a set of parameters from voice, derives the perceptual feature of the voice and sends the parameters to the receiver.
- ➔ The receiver has a synthesizer and reproduces the original voice based on the parameters received.
- ➔ The reproduced voice sounds „synthetic“ and is not good enough for telephony.
- ➔ PBX systems sometimes employ Vocoders for storing messages.
- ➔ Very low bit rates (1...4kbps).
- ➔ Vocoders work in the frequency domain.

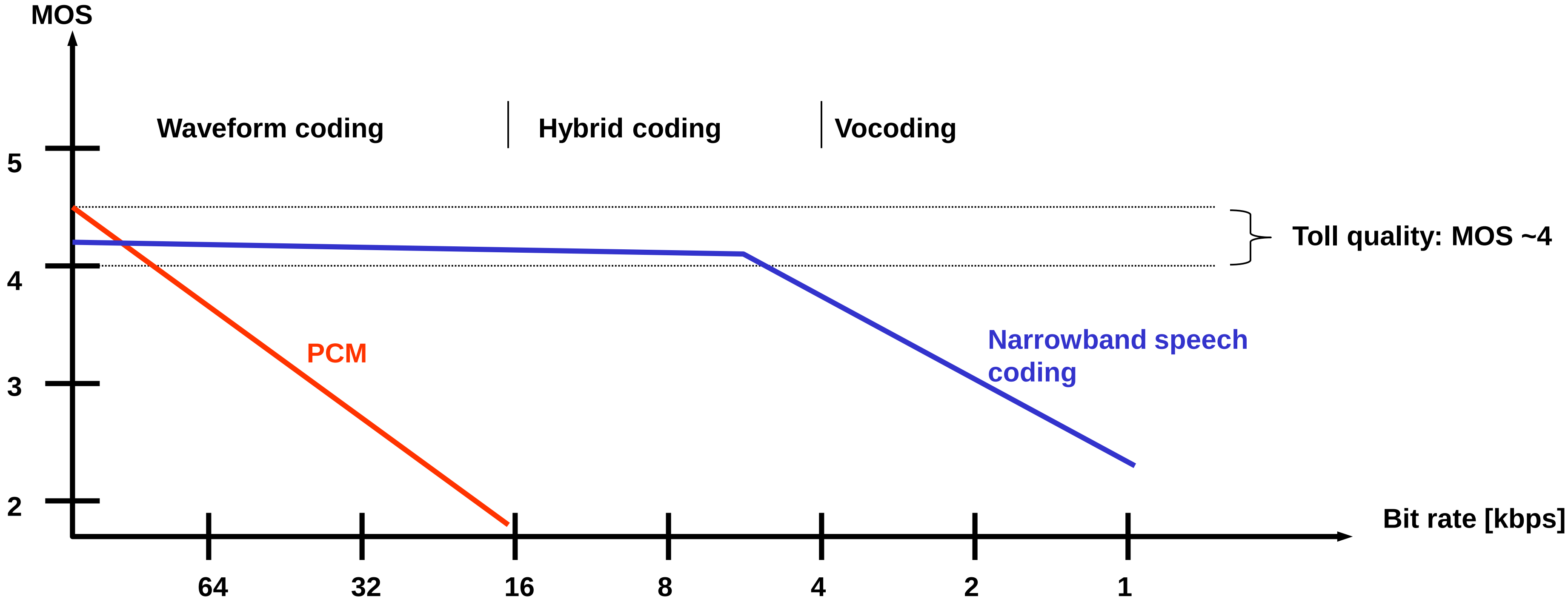
Hybrid coders:

- ➔ Mixture of waveform coders and vocoders.
- ➔ Operate from 4kbps to 16kbps.

2. Voice Codecs (9/12)

➔ Codec technology (2):

Bit rate versus quality:
PCM based codec‘s speech quality deteriorates with higher compression ratio.
Narrowband speech coders are able to maintain a high level of speech quality over a wide range of compression ratios (bit rates).



MOS: Mean Opinion Score; speech quality assessment by representative group of people.

2. Voice Codecs (10/12)

➔ Characteristics of voice coders (1):

• a. Bit rate (usually higher compression results in lower voice quality).

• b. Complexity of the coding algorithm (MIPS required to process the voice):

➔ High complexity = higher costs (more powerful and more expensive DSP).

➔ High complexity = higher power consumption.

➔ High complexity = higher delay.

Codecs require ~10...20MIPS (per voice channel).

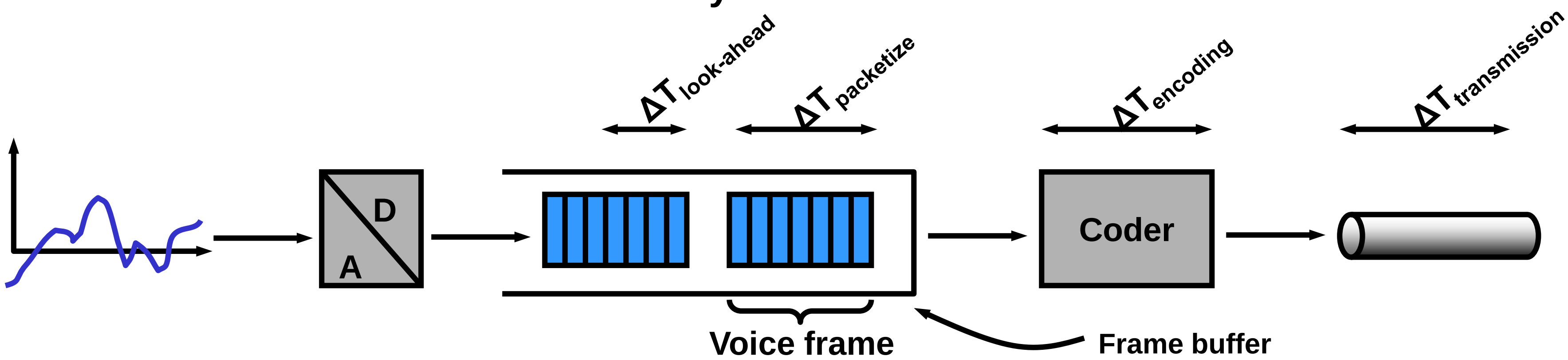
• c. Delay:

• How much delay is acceptable?

➔ ITU-T G.113 / G.114 states that max. 150ms delay in 1 direction for acceptable quality.

➔ Satellite delay: 250ms uplink, 250ms downlink = 500ms end to end delay.

➔ Factors that contribute to delay:



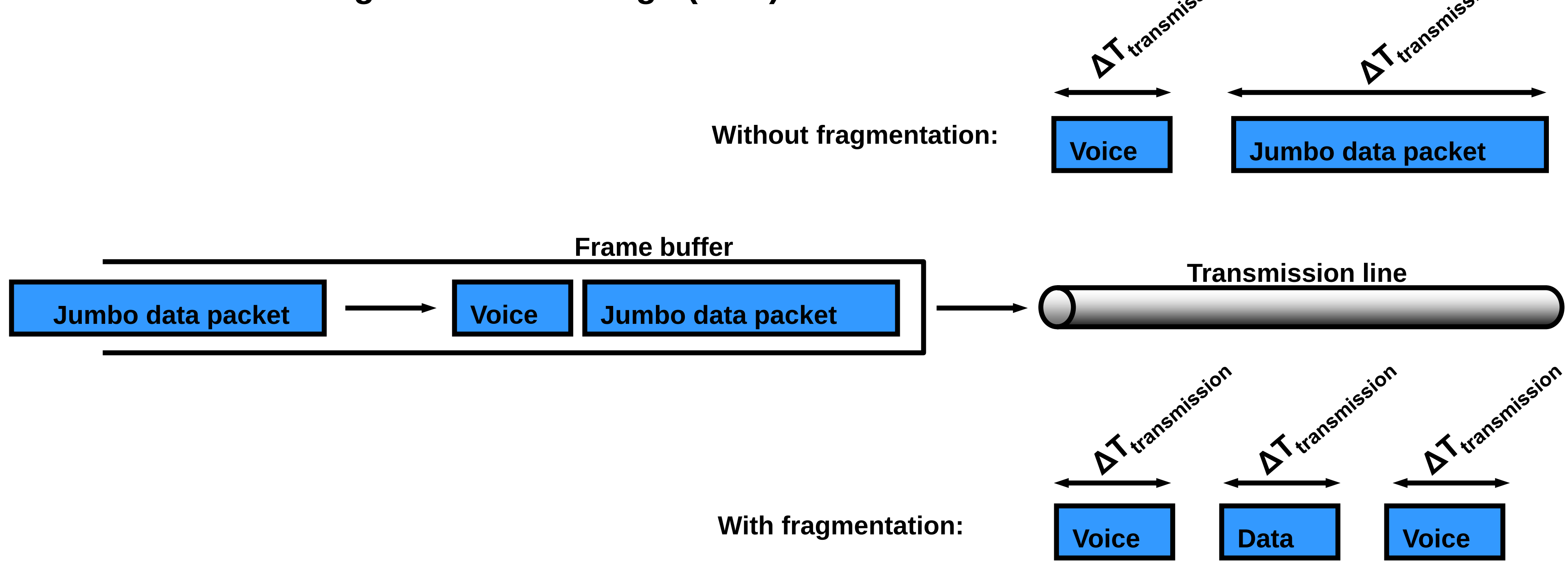
2. Voice Codecs (11/12)

➔ Characteristics of voice coders (2):

- **c. Delay (cont'd):**
  - **Delay on low speed links:**

    E.g. on a link with 256kbps a 1500 bytes data packet takes ~40ms for transmission; during the data packet transmission, the voice frame is blocked (queued) for transmission thus introducing delay.

    Solution: fragmentation of large (data) frames.



2. Voice Codecs (12/12)

➔ **Characteristics of voice coders (3):**

- **d. Quality:**

  How can quality be measured?

  a. Objective measurement (harmonics distortion, S/N):

  E.g. P.862/PESQ (Perceptual Evaluation of Speech Quality).

  Problem: objective measurements do not correlate well with subjective assessment of voice quality.

  b. Subjective measurement („how good does it sound“):

  E.g. MOS Mean Opinion Score:

  ➔ Assessments are carried out by a group of people.

  MOS scale:

  5 = excellent (HiFi).

  4 = toll-quality (G.711, PSTN standard quality).

  1 = lowest (poor) quality.

- **e. Error tolerance (susceptibility to packet loss):**

  Codecs that use entire frame for voice compression (and even look-ahead) are susceptible to packet loss („resync“ DSP).

  Solution: E.g. G.729 error concealment: replay last packet if current packet lost.

  Packet loss tolerance is expressed in [%], e.g. G.729 max. 5% packet loss.

  Packet loss in the Internet is a real problem, see e.g.

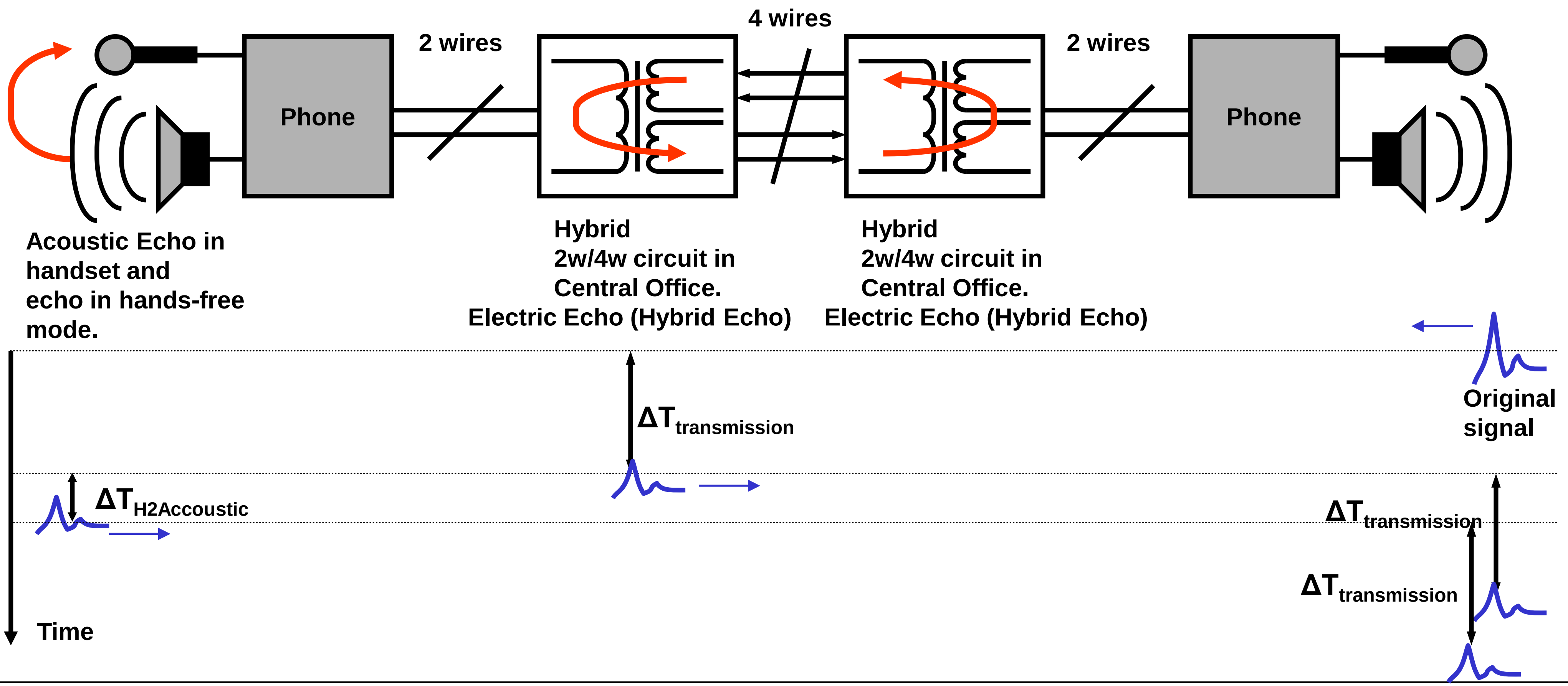
  http://www.internettrafficreport.com/

3. Echo problem with VoIP (1/5)

➔ Echo in traditional PSTN (Echo exists also in PSTN):

Echo (reflection) occurs in hybrid circuit (impedance mismatch) and handset (coupling of loudspeaker signal to microphone).

ISDN phones have separate receive and transmit paths and thus do not need a hybrid circuit; but ISDN phones, like analog phones, have acoustic echo in hands-free mode.

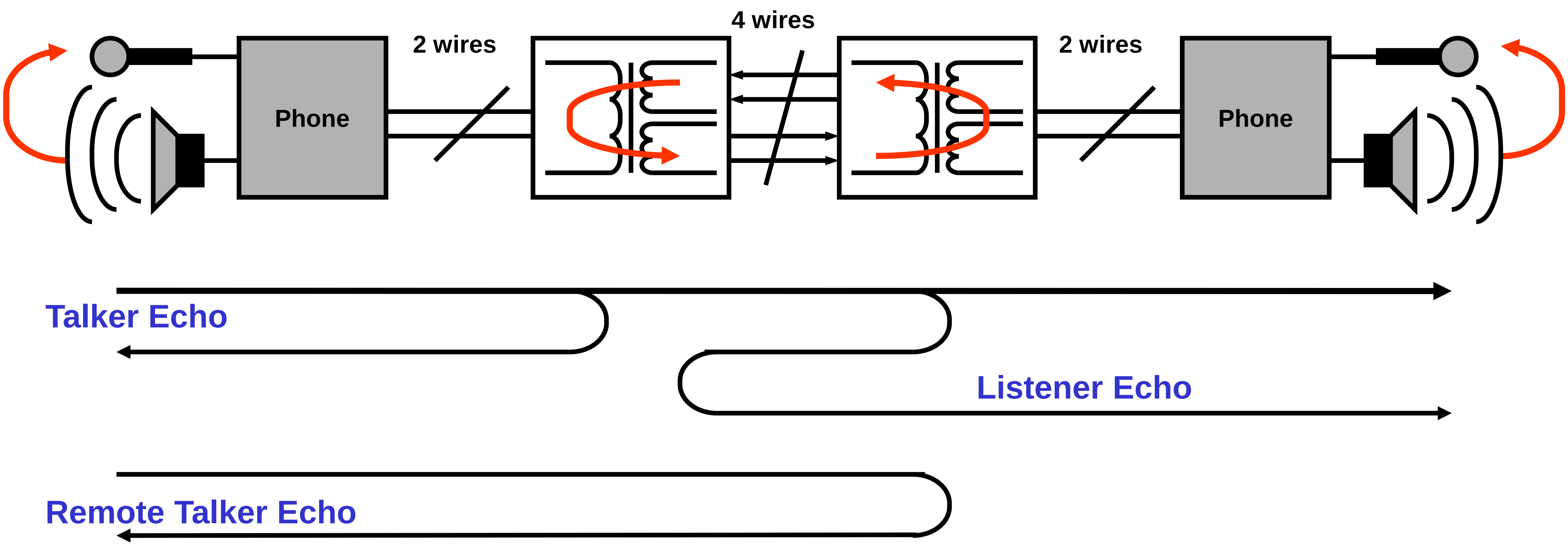


3. Echo problem with VoIP (2/5)

➔ **Echo types:**

Echoes (reflections) occur at different points in the transmission path. Echoes are again reflected at these points but are also dampened (amplitude reduced).

Talker echo: Echo that talker hears (his own voice).
Listener echo: Echo of talker signal that listener hears.
Remote talker echo: Echo of talker signal that talker hears but that is generated at the far end.

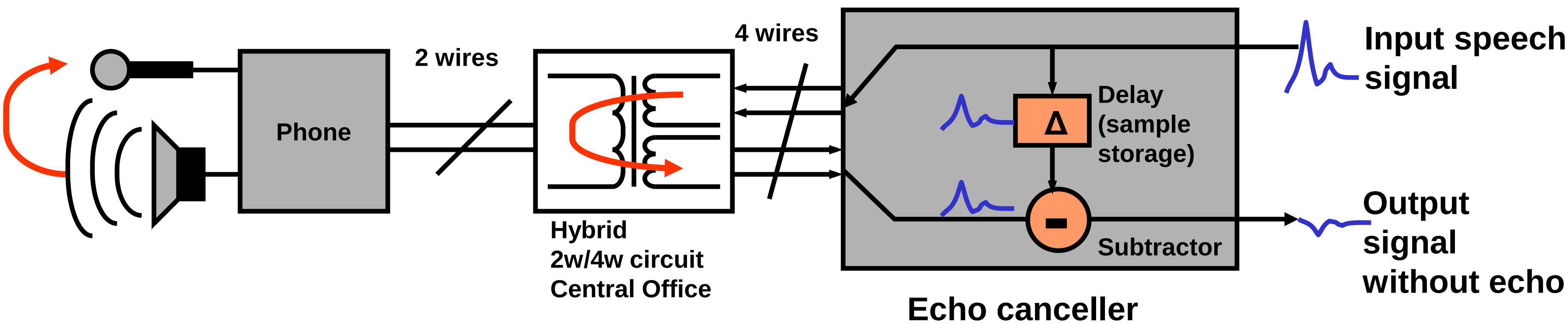


3. Echo problem with VoIP (3/5)

➔ **Solution: Echo Canceller which cancels near-end echo for far-end.**

Each side of a speech connection cancels the locally generated echo for the far side.
The echo canceller must find out the delay between signal and its reflections. The delay of the echo canceller is then configured with this measured delay.
This delay is continusuously adjusted during a speech conversation. Additionally the echo canceller must also be able to handle multiple echoes (reflections) and even echoes of echo signals.

Echo tail: time that signal needs to travel from echo canceller to point of echo and back to echo canceller; typical tail length of echo cancellers are 32ms and 64ms (maximum delay an echo canceller can handle).



3. Echo problem with VoIP (4/5)

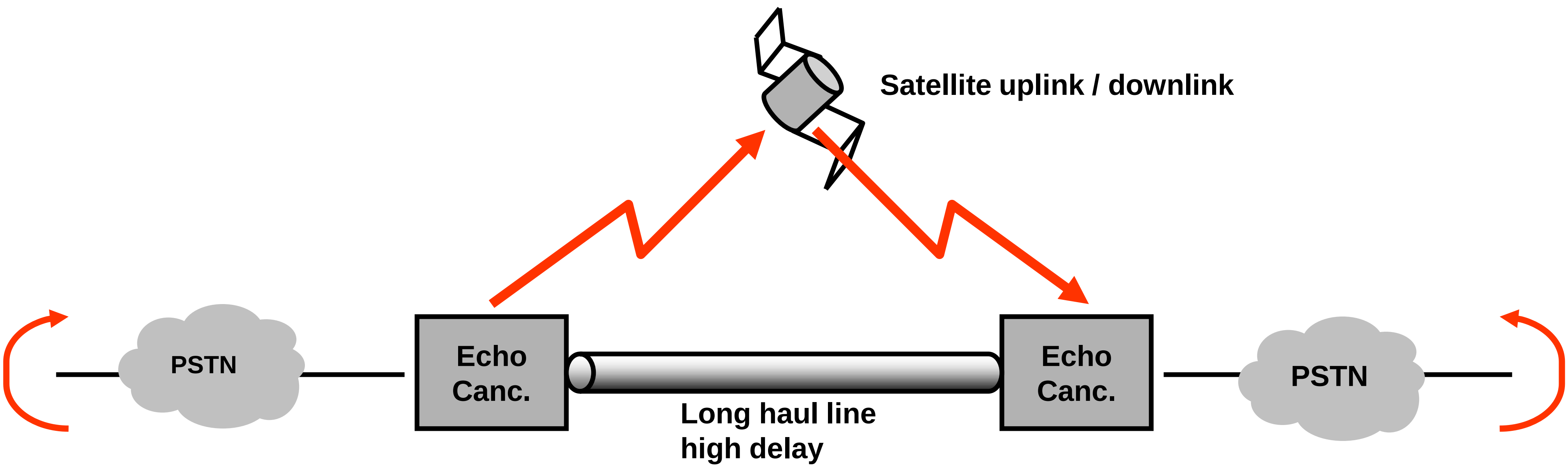
➔ 2 factors contribute to the echo problem:

a. Signal reflections (hybrid, acoustic).

b. Transmission delay.

➔ Thus the echo problem depends on the transmission delay which can not be controlled (satellite links, long haul transatlantic lines).

➔ Echo cancellation needs to be done at long delay lines' ingress points.



3. Echo problem with VoIP (5/5)

➔ When to use echo cancellers:

Rule of thumb: If network delay exceeds 30ms...50ms, echo cancellers need to be used.

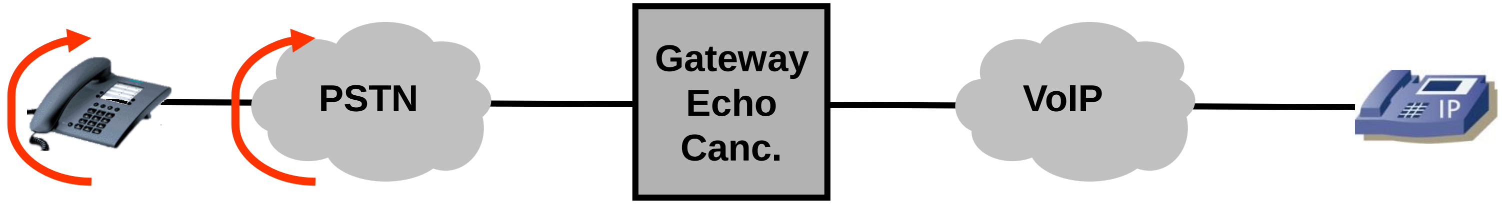
<10ms RTT:	Echo not audible.
10-30ms RTT:	„Tunnel sound“, but communication possible without echo cancellers.
>30ms RTT:	Not ok, echo cancellers must be used.

➔ PSTN (POTS, analog):
Very low transmission delay, thus (almost) no echo problem.

➔ ISDN:
Transmit and receive path are separated (because it is digital), thus no echo is present (except acoustic echo in hands-free mode).

➔ Satellite links:
Typically 250ms uplink and 250 downlink, thus echo cancellers needed.

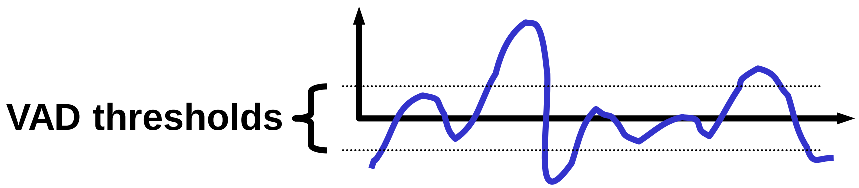
➔ VoIP:
Considerable delay (in packet network), thus need to cancel echo generated in PSTN (echo canceller removes echo produced in PSTN for VoIP client).



4. Voice Activity Detection / Comfort Noise Generation (VAD / CNG)

➔ **Voice data rate reduction through silence suppression:**

In a (reasonable) conversion at most 50% of the bandwidth is used (usually one party is silent while the other speaks). With VAD voice packets are only sent if there is speech thus saving bandwidth.



➔ **Problems of VAD:**

1. Hangover:

Codec remains active for some time (typ. 200ms) after voice level has fallen below threshold.

2. Front end clipping:

VAD needs some time to detect if signal amplitude has exceeded the threshold. The first syllable may be cut off („Meier Eier" problem).

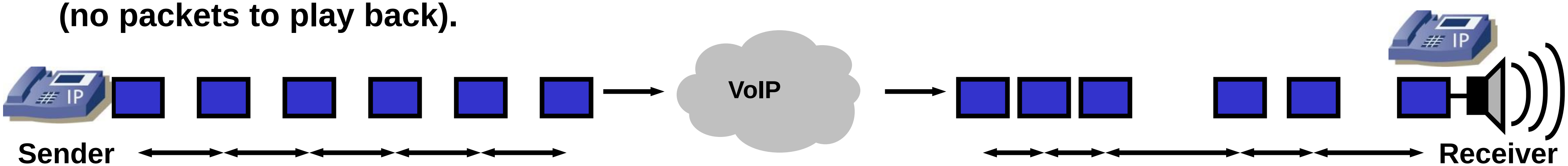
3. Silent periods:

Silent periods are very disturbing during a conversation (line appears to be „dead"). Comfort Noise Generation (CNG) produces an artificial background noise so that the line does not appear to be dead. CNG measures the background noise level and spectral distribution and transmit this information to peer which plays back the noise signal.

5. Jitter = inter-packet arrival variations

➔ Control of packet spacing at the receiver:

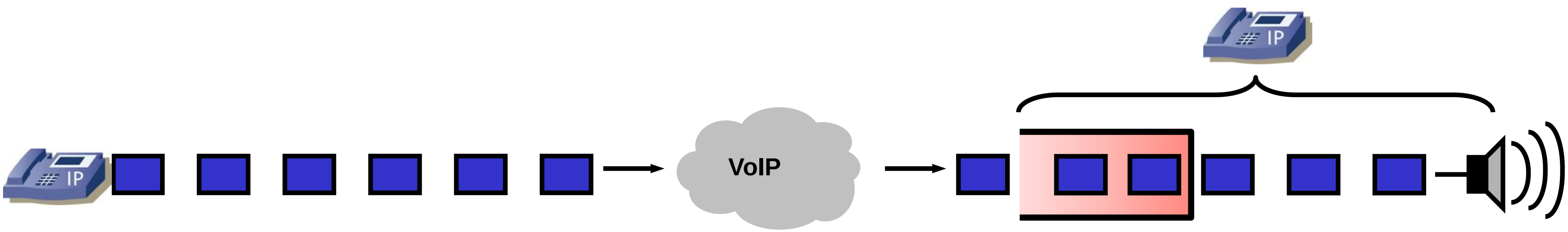
The receiver must make sure that the voice decoder or phone never has a packet underrun (no packets to play back).



Voice packets nicely spaced (traffic shaping).

Packet network introduces non-uniform delay.

Packets arrive at destination with uneven delays.



Phone has built-in dejitter buffer

➔ The dejitter buffer stores packets and replays them evenly towards the speaker thus ensuring that there are no dropouts.

But: The dejitter buffer introduces additional delay!

6. VoIP relies heavily on DSP technology

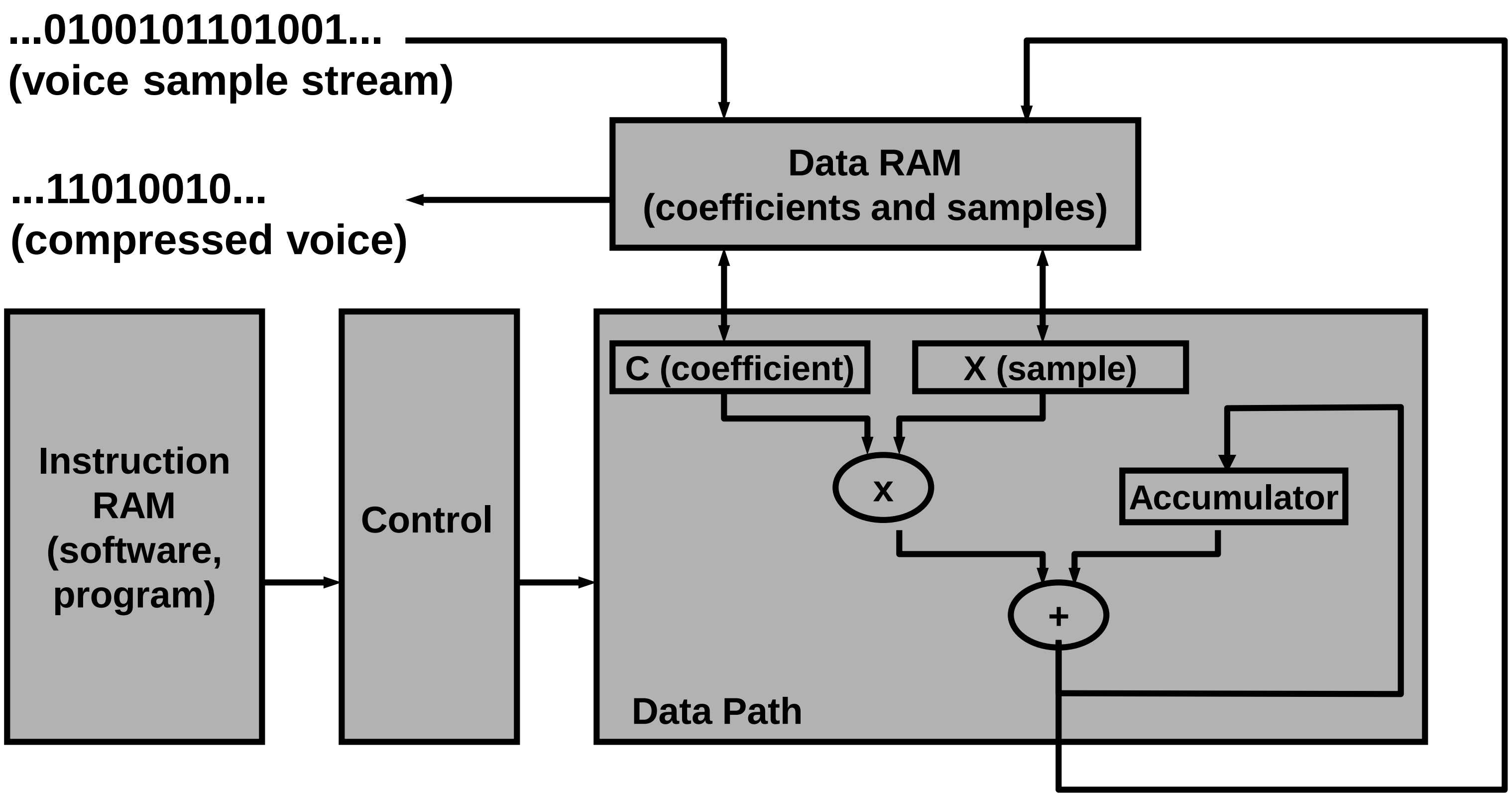
Who‘s doing all the coding, echo cancellation, voice activity detection, fax/modem detection/modulation etc.? ➔ DSP Digital Signal Processor.

The DSP digital signal processing is mainly MAC: Multiply ACcumulate operations.

E.g. Finite Impulse Response filter (FIR):

$$Y_N = X_N * C_1 + X_{N-1}*C_2 + X_{N-2}*C_3 \text{ .... } + X_0*C_N$$

➔ The DSP is optimised for these calculations (harward architecture).



7. Transport of real-time traffic: RTP / RTCP RFC1889 (1/2)

➔ Almost all VoIP protocols (H.323, SIP, MGCP, Skinny) use RTP over UDP for the transport of voice or video.

➔ RTP does not itself provide real-time characteristics. Instead it transports information that help the application achieve real-time behavior.

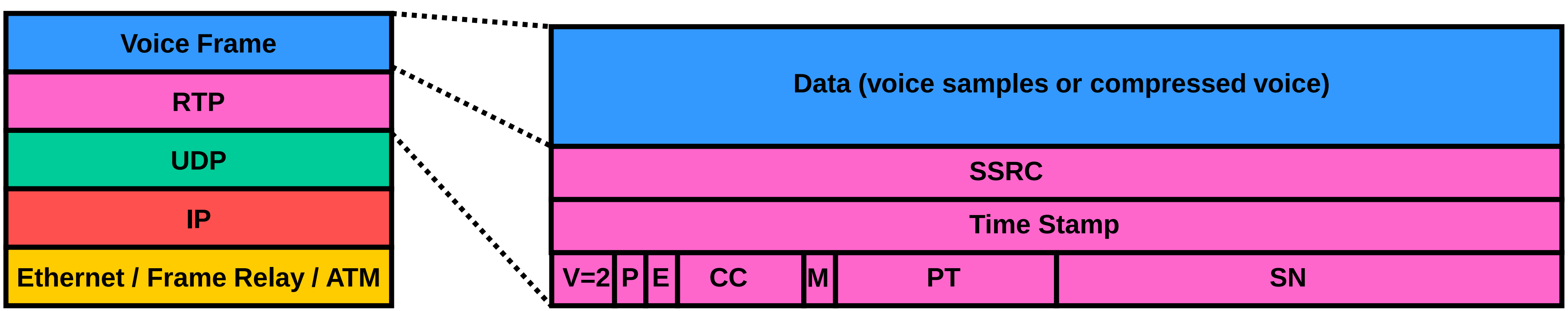
RTP: Real Time traffic Transport Protocol functions:

1. Sequencing (SN) (reordering of voice packets).
2. Time stamping (dejitter buffer control).
3. Payload type (PT) indication (which codec was used for voice in RTP packet).
4. Multiplexing (SSRC) (indication of source in case of conferencing).
5. Layer 4 framing (M) (indication of video frame).

RTCP: Real Time Transport Control Protocol functions:

1. Long term delay and packet loss statistics (5s).
2. Quality monitoring.

RTP protocol stack:



7. Transport of real-time traffic: RTP / RTCP RFC1889 (2/2)

➔ RTP uses UDP:

TCP retransmissions would garble the voice. Voice must be delivered to the loudspeaker as quickly as possible. TCP retransmissions introduce (variable) delay. Timely delivery is more important (UDP) than error-free delivery (TCP) as long as the error rate is below an acceptable level.

➔ RTP has a high overhead:

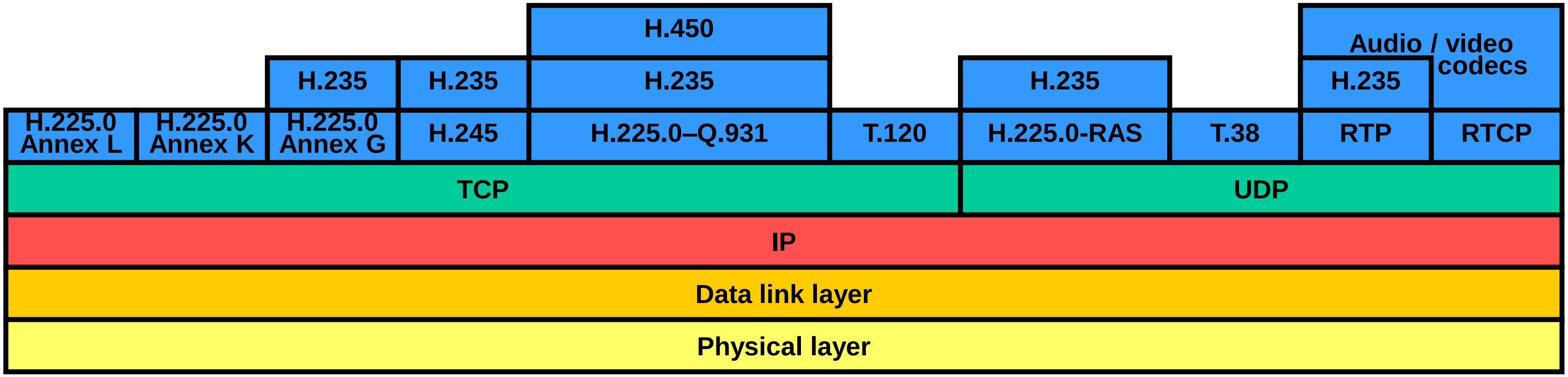
12 bytes RTP +
8 bytes UDP +
20 bytes IP =
<u>40 bytes headers</u>
E.g. G.729 with 10 bytes payload ➔ 80% overhead!
Overhead can be reduced with compressed RTP (cRTP):
→ 40 bytes are compressed to 4 bytes (with UDP checksum);
→ 40 bytes are compressed to 2 bytes (without UDP checksum).
But: cRTP is only possible on point to point links (since IP header is compressed).

➔ RTCP may be used for long term traffic monitoring (5s between RTCP reports between 2 RTP endpoints). But RTCP is usually not used to monitor voice quality.

8. H.323 (1/8)

➔ H.323 = ITU-T „all in one“ protocol suite for voice, data, fax and video over IP. H.323 is not a protocol but a protocol suite (also called „umbrella standard“).

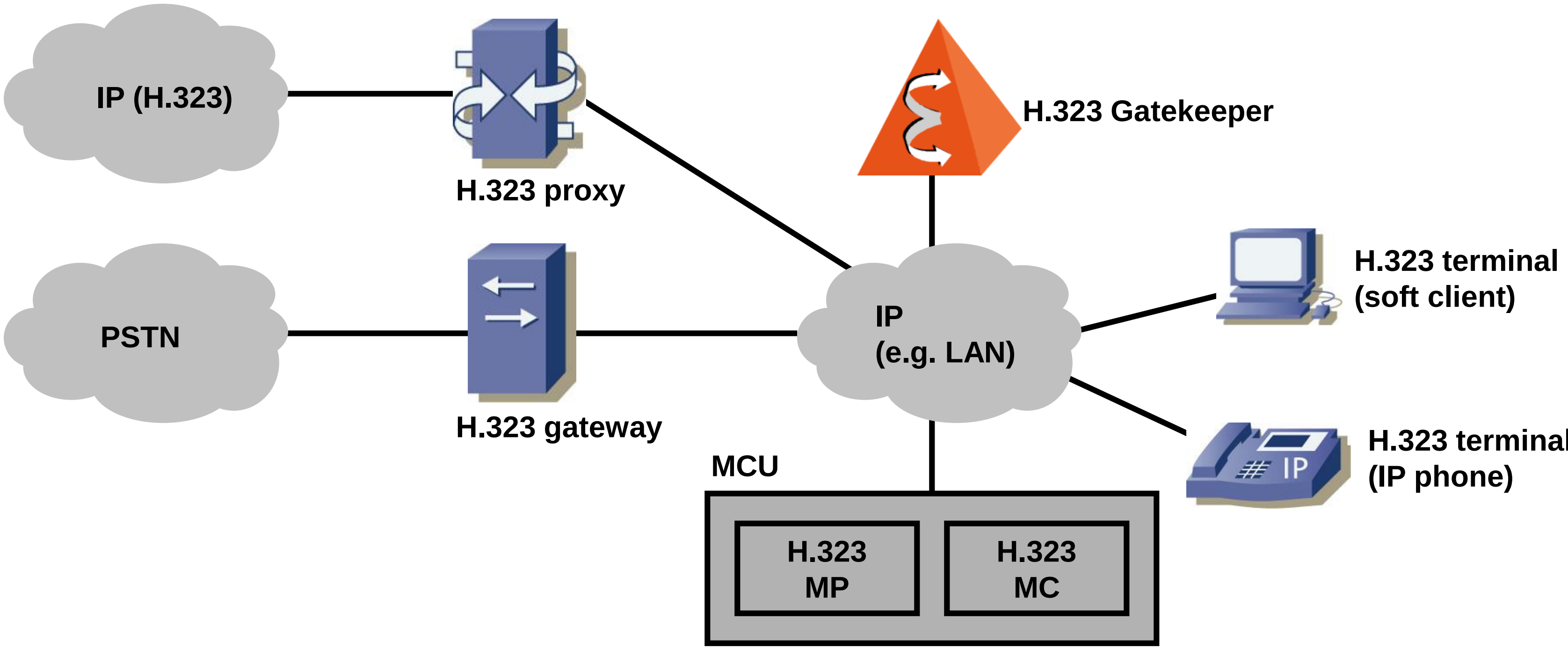
➔ H.323 Protocol components and protocol stacks:



• H.225.0-Q.931:	Call signaling protocol (similar to Q.931 in ISDN).
• H.245:	Logical channel (data/media channel) control protocol for set up and tear down media (voice) channels.
• H.225.0-RAS:	Registration, Admission, Status (registration with central gatekeeper).
• H.235:	Security (message data integrity, authentication, privacy).
• H.450:	Supplementary services for H.323 (CF, CW, 3PTY).
• T.120/T.12x:	Data sharing (data, video).
• T.38:	Fax over IP.

8. H.323 (2/8)

➔ H.323 Components (1):



Gatekeeper (=RAS Server):

All devices (clients, gateways, MCU) register with gatekeeper. For each new call the clients contact the gatekeeper for address resolution. The gatekeeper mainly has following functions:

- Access Control (who is allowed to place calls and to whom).
- Registration (of phone number and according IP address).
- Address Translation (phone number to IP).

N.B.: H.323 also allows direct signaling between clients without a gatekeeper in between („gatekeeper-less signaling“).

8. H.323 (3/8)

➔ H.323 Components (2):

MCU: Multipoint Control Unit (MC + n*MP):
Conferences between >2 parties need a multipoint unit for mixing the voice streams so that each party can hear all other conversation partners. The MCU consists of a control unit (Multipoint Controller MC) and 1 or many MPs (devices that actually mix audio streams for a conversation). The MPs are either specialized hardware devices with DSPs or powerful general purpose processors.

H.323 gateway:
The gateway interfaces the H.323 network (IP) to the PSTN (packet to circuit conversion). It consists of a signaling gateway (e.g. H.323 to ISDN signaling) and a data path gateway (e.g. RTP G.723 to G.711 transcoding).

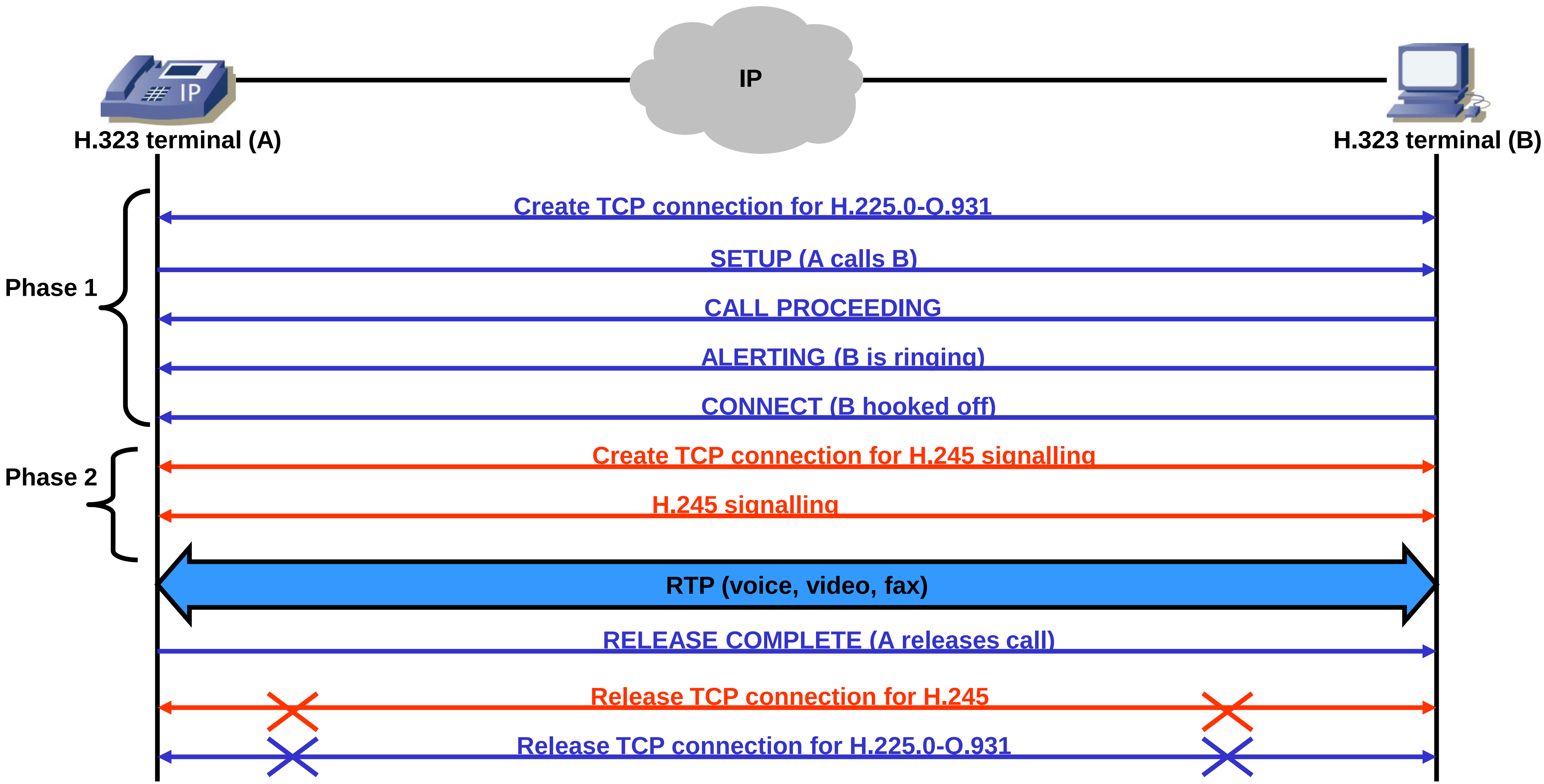
H.323 Proxy:
Proxies allow to connect an internal H.323 network (private) to an external H.323 network (public). In addition proxies afford firewall functionality (firewall for H.323 services).

H.323 Terminal:
Either softphones (soft clients) or hardphones.

8. H.323 (4/8)

➔ Signalling (terminal to terminal) (1):

Message flow for direct signaling between 2 H.323 clients.



8. H.323 (5/8)

➔ Signalling (terminal to terminal) (2):

H.323 signalling phase 1:

H.225.0-Q.931 protocol messages are used for call setup (setup, alerting, disconnect).
As its name implies this protocol is very similar to the ISDN signalling protocol (=Q.931).

H.323 signalling phase 2:

H.245 data channel signalling capability exchange (similar to PPP LCP) where each peer tells the other its capabilities. The 2 parties agree on the set of capabilities (codec to be used, VAD etc.) for the session.
If both parties disagree on media channel settings one party becomes master and resolves the conflict (Master slave determination).
The media channel characteristics may be changed during the call (optional mode request procedure), e.g. change of codec for a fax transmission (see below).

8. H.323 (6/8)

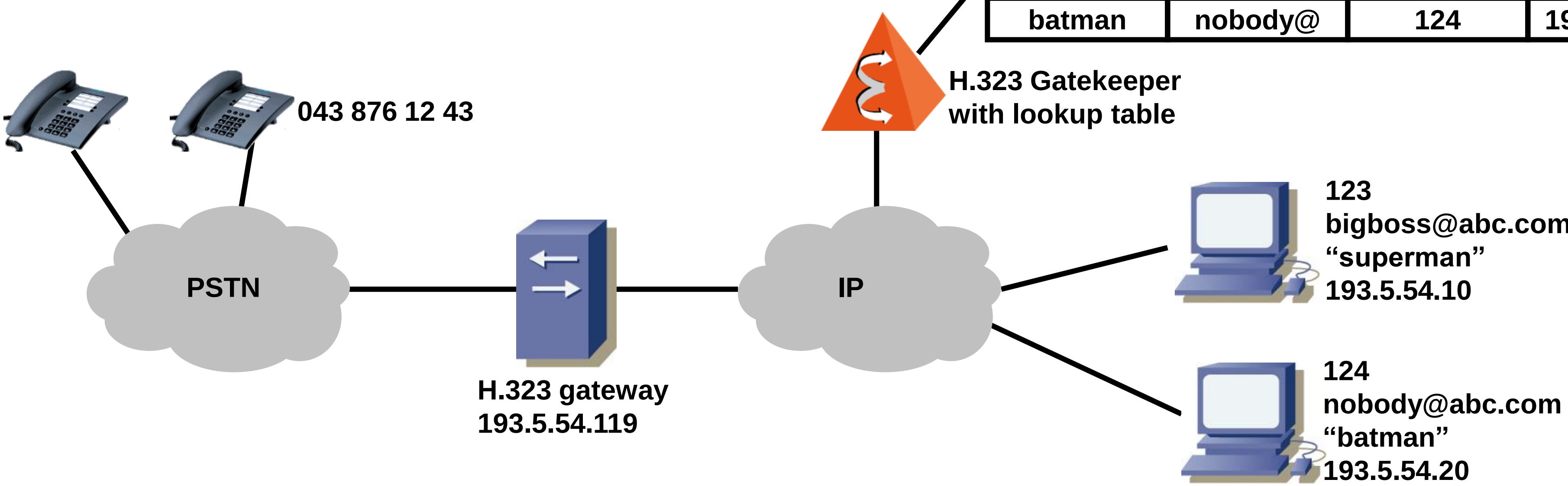
➔ **Addressing:**

H.323 supports multiple classes of addresses:

- **E.164: International PSTN phone number.**
- **E-Mail address („user@company.com“).**
- **URL (H323://user1@isp1.com).**
- **IP address (some IP phones, e.g. NetMeeting can be addressed by an IP address).**
- **String, alias name.**

At startup H.323 clients (phones, gateway, MCU) register their addresses, aliases etc. with the gatekeeper.

Alias	Mail addr.	E.164	IP
-	-	0438761243	193.5.54.119
superman	bigboss@...	123	193.5.54.10
batman	nobody@	124	193.5.54.20



8. H.323 (7/8)

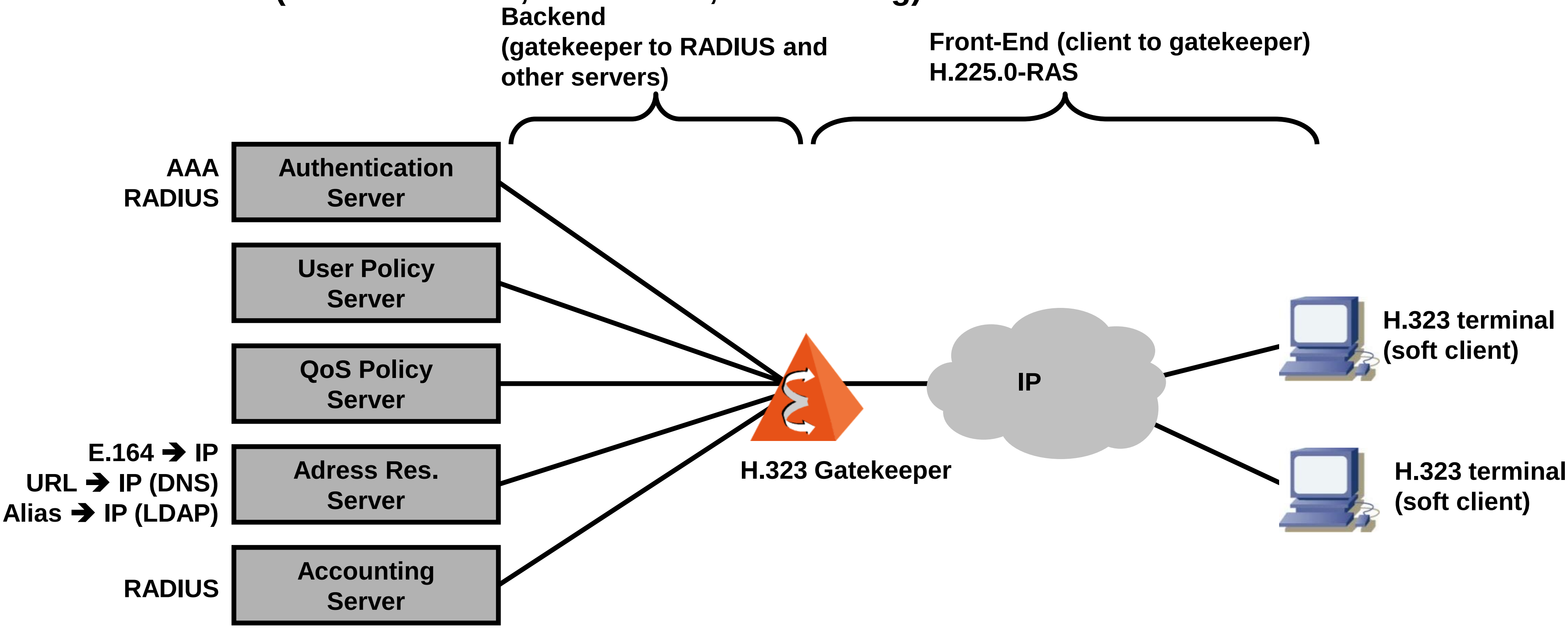
➔ H.225.0 RAS (1):

After startup H.323 clients register at the gatekeeper with the H.225.0 RAS protocol.
The gatekeeper talks to backend servers through other protocols (e.g. RADIUS).
RAS provides the following services:

a. Address resolution.

b. QoS (bandwidth allocation).

c. AAA services (Authentication, Admission, Accounting).



8. H.323 (8/8)
➔ H.225.0 RAS (2):
RAS messages:

Gatekeeper Discovery (find gatekeeper):
- Gatekeeper Request GRQ
- Gatekeeper Confirm/Reject GCF/GRJ

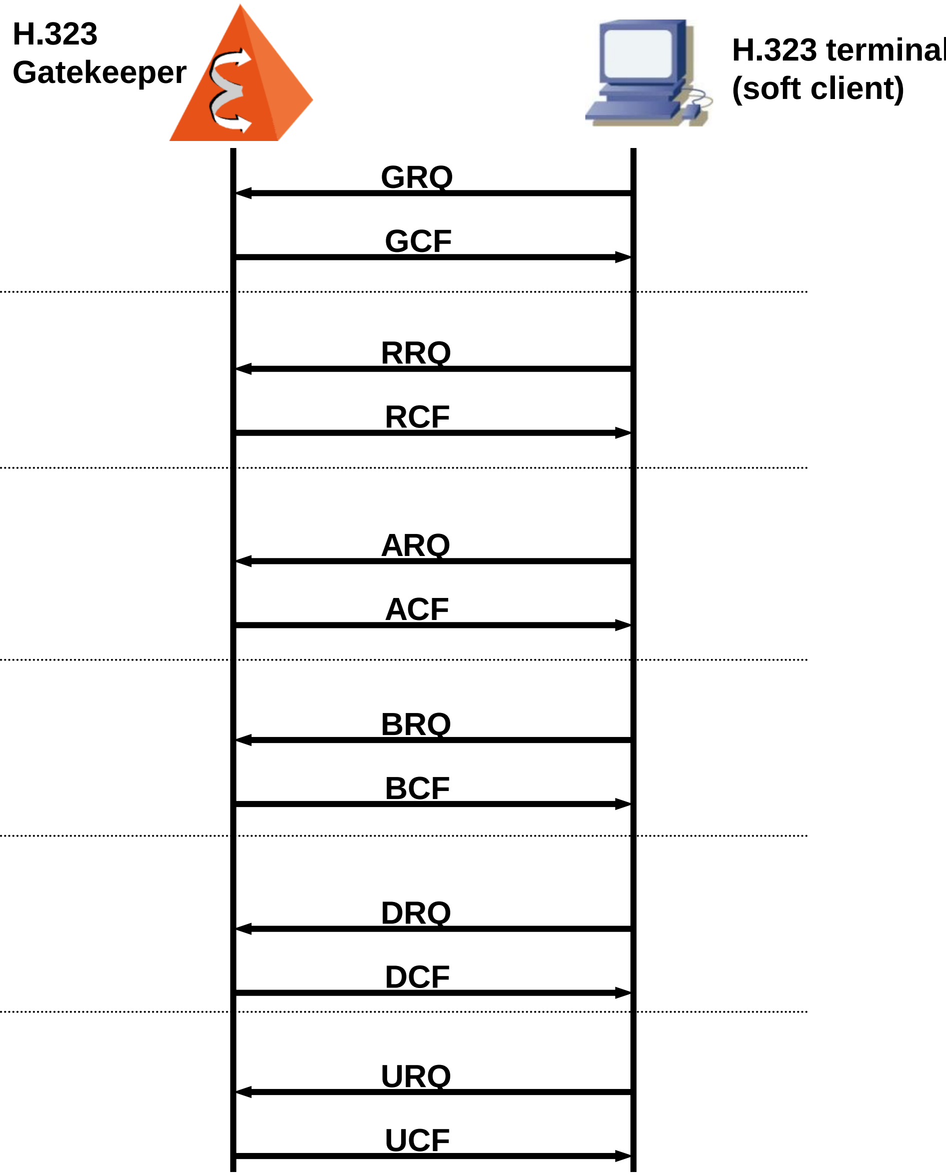
Gatekeeper Registration (register with gatkeeper):
- Registration Request RRQ
- Registration Confirm/Reject RCF/RRJ

Admission Request (for each call):
- Admission Request ARQ
- Admission Confirm/Reject ACF/ARJ

Bandwidth Request (optional, request BW for call):
- Bandwidth Request BRQ
- Bandwidth Confirm/Reject BCF/BRJ

Disengage Request (at end of call):
- Disengage Request DRQ
- Disengage Confirm DCF

Unregister Request (un-register with gatekeeper):
- Unregister Request URQ
- Unregister Confirm/Reject UCF/URJ



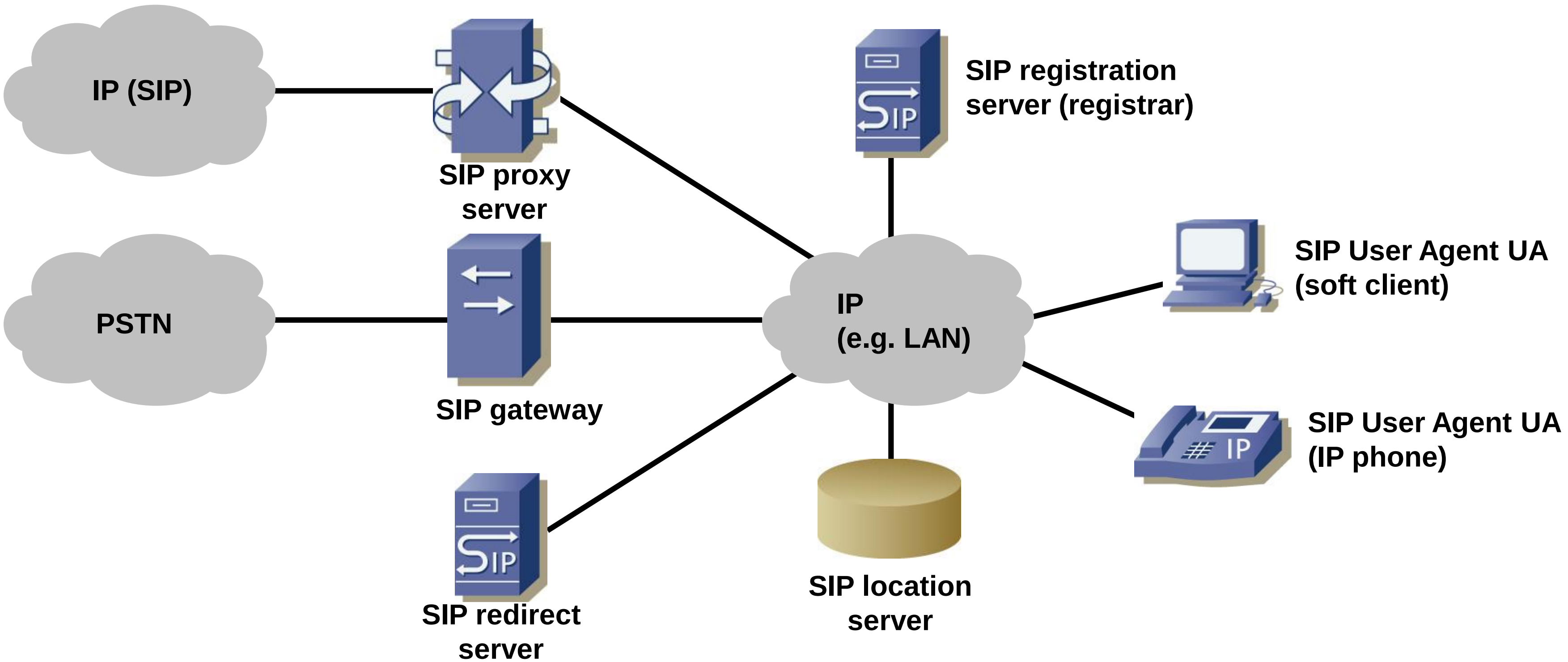
9. SIP - Session Initiation Protocol - RFC3261 (1/8)

➔ SIP Components (1):

A SIP network consists of:

- a. SIP User Agents (clients, phones)
- b. SIP servers (SIP proxy server, redirect server, registration server)

A user agent UA is a SIP client. However SIP servers (proxy server, registration server) also contain the UA functionality.



9. SIP - Session Initiation Protocol - RFC3261 (2/8)

➔ SIP Components (2):

SIP Gateway:

User Agent that connects SIP to other protocols (like ISDN).

SIP User Agent UA:

SIP-enabled endpoint. Device that can send and receive SIP INVITE and ACK messages.
A UA consists of a UA client (UAC) and UA server (UAS) akin to email client.

SIP Proxy Server:

SIP enabled device that acts both as SIP server and client. A SIP proxy server receives a SIP request (thus acting as SIP UA server), performs some application-specific action on the SIP message (e.g. changing the URLs) and forwards the SIP request to another SIP server (thus acting as SIP UA client).

SIP Redirect Server:

SIP server that accepts a SIP request, maps the incoming address to one or more new addresses and returns these new addresses to the requesting SIP client.
Unlike a SIP proxy server does initiate SIP requests on its own but only redirects a requesting client to another server.

SIP Registration Server (Registrar):

A server that accepts SIP REGISTER messages. The registrar stores the address information in a location service via a non-SIP protocol (e.g. LDAP).

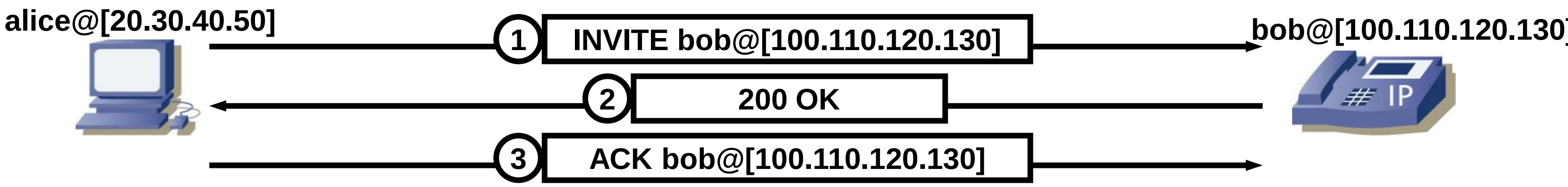
SIP Location Server:

A non-SIP device that is accessed by SIP redirect servers and SIP registrars. SIP location servers are used for address resolution (URI).

9. SIP - Session Initiation Protocol - RFC3261 (3/8)

➔ SIP session setup (1):

Call setup using SIP Direct Mode (address of callee known to caller):

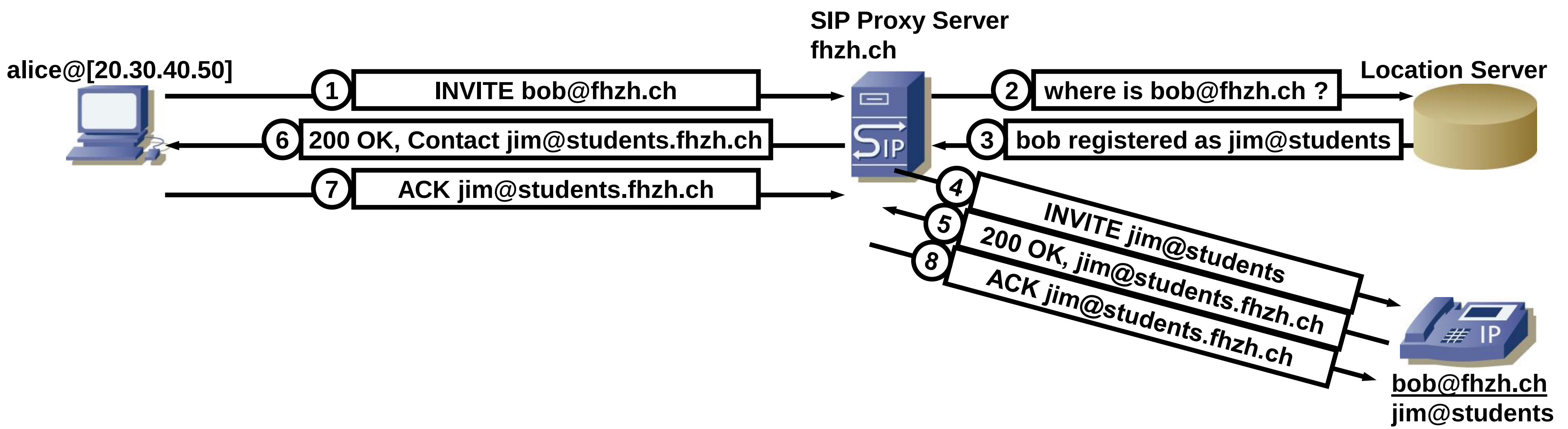


1. **Alice sends an INVITE message to Bob.**
2. **Bob accepts the call and responds with code 200 OK.**
3. **Alice acknowledges with an ACK message.**
   Both Alice and Bob open the data path (RTP) and start the conversation.

9. SIP - Session Initiation Protocol - RFC3261 (4/8)

➔ **SIP session setup (2):**

Call setup via SIP Proxy Server:



1. Alice sends an INVITE message to fhzh.com server (which acts as proxy server).

2./3. The proxy server looks up bob in its location server (through a non-SIP protocol like LDAP) and determines that bob is registered as jim@students.

4. The proxy server constructs a new URL jim@students.fhzh.ch and sends the INVITE message to Bob's PC (or SIP phone).

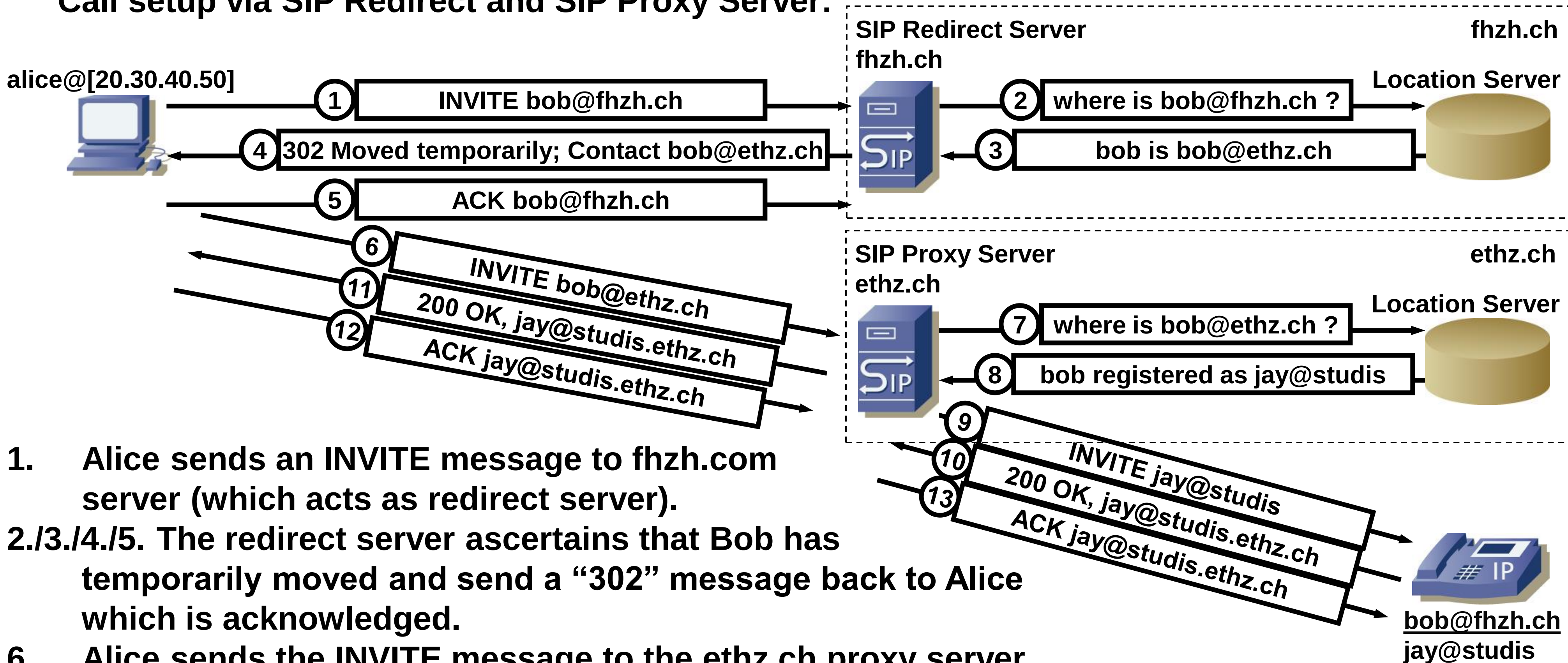
5./6. Bob accepts the call and sends back an ACK message to the proxy server which in turn sends it to Alice.

7./8. Alice acknowledges with an ACK message (sent to the proxy server and from there to bob.

9. SIP - Session Initiation Protocol - RFC3261 (5/8)

➔ SIP session setup (3):

Call setup via SIP Redirect and SIP Proxy Server:



1. Alice sends an INVITE message to fhzh.com server (which acts as redirect server).

2./3./4./5. The redirect server ascertains that Bob has temporarily moved and send a “302” message back to Alice which is acknowledged.

6. Alice sends the INVITE message to the ethz.ch proxy server.

7./8. The proxy server determines through the location server that bob is registered asjay@studis).

9. The proxy server sends an INVITE message on behalf of Alice to jay@studis.

10.-12. OK and ACK messages.

9. SIP - Session Initiation Protocol - RFC3261 (6/8)

➔ SIP message format (1):

*** SIP message use the RFC2822 format like FTP, SMTP or HTTP.**

*** SIP messages are text based (UTF-8, ASCII). This allows easy parsing with regular expressions or parsers.**

*** SIP messages are either requests (client → server) or responses (server → client).**

*** SIP message header fields have the form „field-name: field-value".**

*** Mandatory SIP header fields:**

To:	Desired local recipient (may or may not be ultimate recipient of the request). To: field in response must equal To: field in request.
From:	Logical source address (who is calling). Logical means that it must not contain an IP address or FQDN (DNS name). From: field in response must equal From: field in request.
CSeq:	CSeq (Command Sequence, sequence number) contains an 32bit integer and a method name. The CSeq number is incremented for each new request within a dialog. CSeq serves to order transactions within a dialog (a dialog is a session).
Call-ID:	Unique identifier for a session. Call-ID should have the same value in all messages of a specific session (call).
Max-Forwards:	Used for limiting the max. number of hops a request can make on its way to the destination. Like the IP-TTL this field is decremented by each SIP hop.
Via:	Indicates the path taken so far and indicates the path that should be followed in routing responses. The „branch" ID parameter in the via header field serves as a transaction ID for loop detection by proxies (branch value must start with the magic cookie "z9hG4bK„). The Via field in responses should have the same value as the Via field in requests. The Via field contains: transport protocol, the client's host name or network address and possibly the port number (if it is missing the default port 5060 is assumed).

9. SIP - Session Initiation Protocol - RFC3261 (7/8)

➔ SIP message format (2):

SIP request messages are INVITE, ACK, BYE, REGISTER, CANCEL, OPTIONS. Example:

1. Start Line (request line):

```
INVITE sip:anonymous@172.16.61.101:5060 SIP/2.0
```

2. Message header:

```
Via: SIP/2.0/UDP 172.16.61.100:5060;branch=z9hG4bKc3ebaa836
Max-Forwards: 70
To: sip:anonymous@172.16.61.101:5060
From: sip:200@172.16.61.100:5060;tag=555a194141ced29
Call-ID: 6839c0c17b344f111720b75710b9c05b@172.16.61.100
CSeq: 587023346 INVITE
Supported: timer
Content-Type: application/sdp
Content-Length: 196
Contact: sip:200@172.16.61.100:5060
Supported: replaces
User-Agent: MDD1200 MxSF/v3.2.5.21
```

2. Request body (optional, can contain session description in SDP format, body separated from header by an empty line):

```
v=0
o=MxSIP 0 9 IN IP4 172.16.61.100
s=SIP Call
c=IN IP4 172.16.61.100
t=0 0
m=audio 4864 RTP/AVP 8 101
a=rtpmap:8 PCMA/8000
a=rtpmap:101 telephone-event/8000
a=fmtp:101 0-15
a=sendrecv
```

9. SIP - Session Initiation Protocol - RFC3261 (8/8)

➔ SIP message format (3):

SIP response messages are messages with status code (3 digit code). Example:

1. Start Line (status line):

```
SIP/2.0 200 OK        Message type (GET), path, protocol & version.
```

2. Message header:

```
Call-ID: 6839c0c17b344f111720b75710b9c05b@172.16.61.100
CSeq: 587023346 INVITE
From: sip:200@172.16.61.100:5060;tag=555a194141ced29
To: sip:anonymous@172.16.61.101:5060;tag=1b4a558777daa09
Via: SIP/2.0/UDP 172.16.61.100:5060;branch=z9hG4bKc3ebaa836
Content-Length: 179
Content-Type: application/sdp
Supported: replaces
Contact: sip:anonymous@172.16.61.101:5060
User-Agent: MDD1400 MxSF/v3.2.5.21
```

2. Response body (optional, can contain session description in SDP format, body separated from header by an empty line):

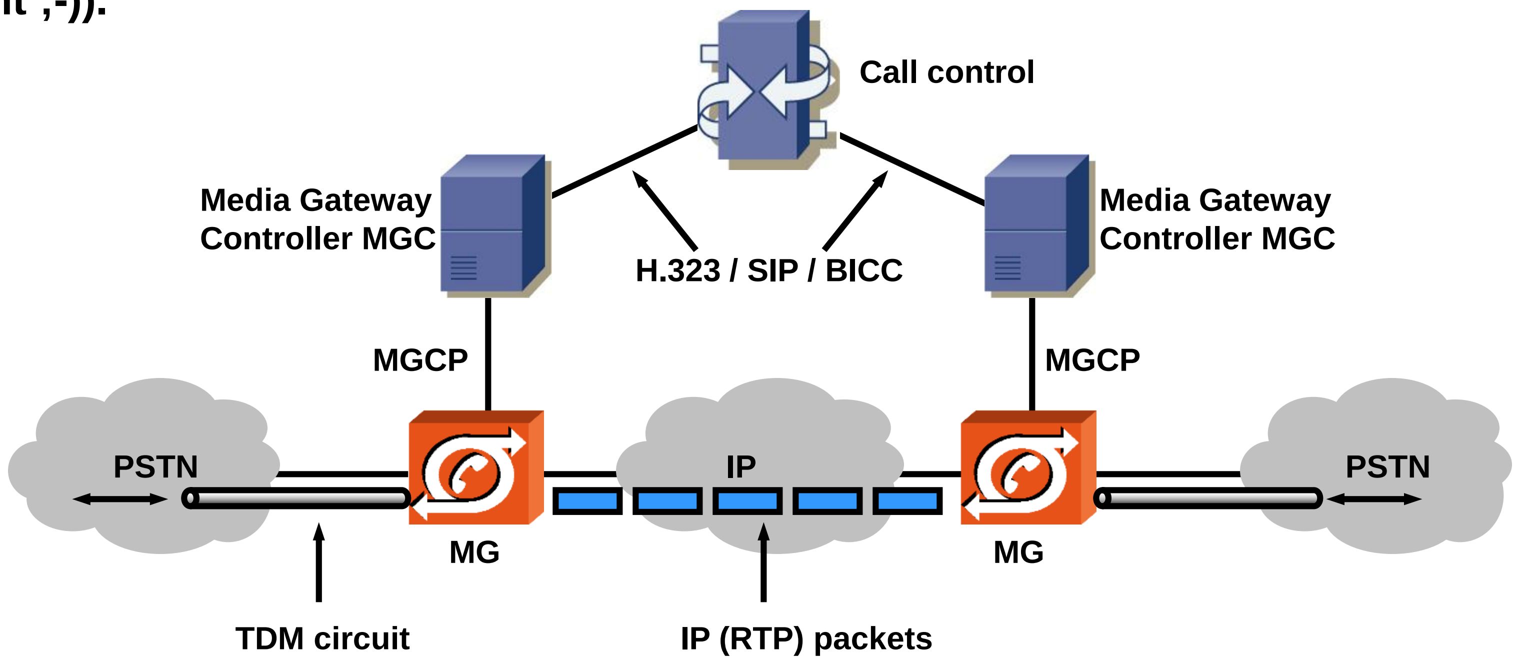
```
v=0
o=MxSIP 0 5 IN IP4 172.16.61.101
s=SIP Call
c=IN IP4 172.16.61.101
t=0 0
m=audio 4864 RTP/AVP 8 101
a=rtpmap:8 PCMA/8000
a=rtpmap:101 telephone-event/8000
a=sendrecv
```

10. MGCP - RFC2705/2805 (1/6)

➔ MGCP is a protocol to control media gateways (data path).

➔ MGCP allows the MGC to switch on/off media/voice channels on the media gateway, switch on/off specific tones (e.g. dial tone) etc.

➔ A newer protocol called MEGACO (MEdia GAteway COntrol, ITU-T H.248.1) is meant to replace MGCP in the future. MEGACO is very similar to MGCP (basically the same but different ;-)).

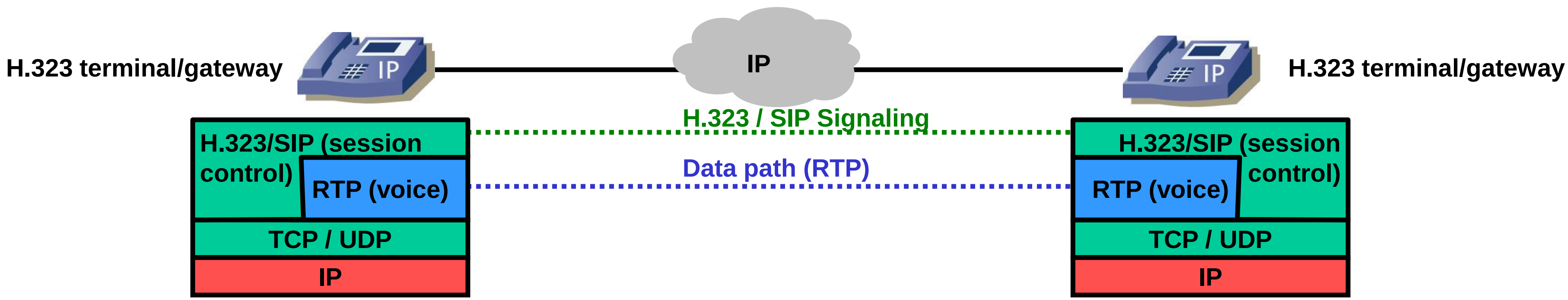


MG: Media Gateway (converts from voice in TDM circuit to RTP (IP) packets).
MGC: Media Gateway Controller (controls MG via MGCP protocol).
Call control: Central control of call. Call control box controls MGC boxes via H.323 or SIP.

10. MGCP - RFC2705/2805 (2/6)

➔ Comparison H.323/SIP model vs. MGCP model (1):

H.323/SIP call model (peer model):



➔ Each H.323/SIP endpoint is fully „aware“ of calls / sessions. The endpoints „talk“ directly to each other (with H.323 or SIP).

➔ The endpoints have wider control over local functionality (tone generation, codec selection...).

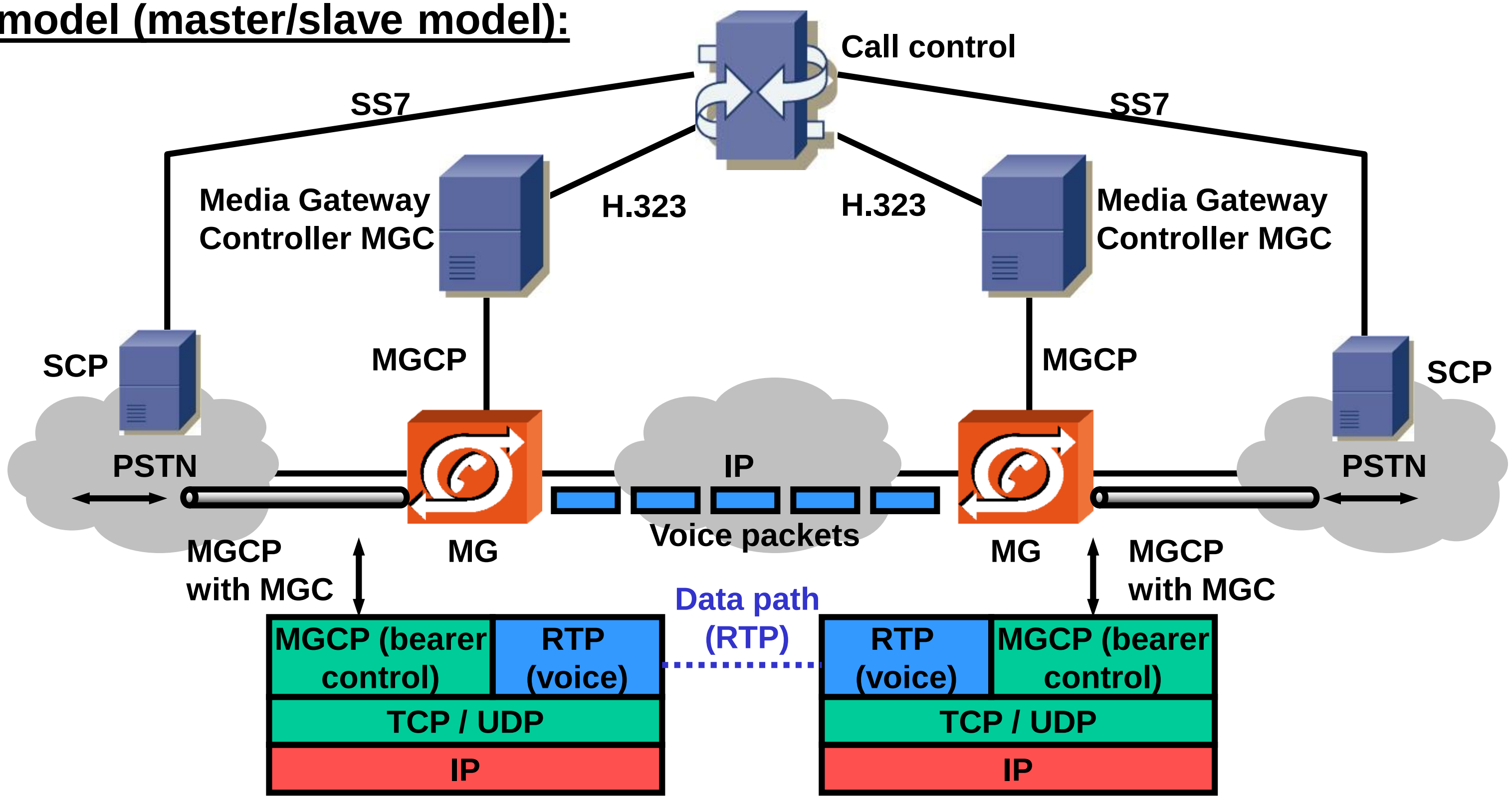
➔ In the peer-to-peer-model the session control protocol (H.323/SIP)is symmetric. Each side (left or right) can initiate a call. The protocol messages can flow in either direction.

➔ Peer model = „intelligent terminal“ phylosophy: The terminals have all the session control and bearer control (voice channel) functionality built-in. Supplementary features like Call-Forwarding are handled in the terminals.

10. MGCP - RFC2705/2805 (3/6)

➔ Comparison H.323/SIP model vs. MGCP model (2):

MGCP model (master/slave model):



➔ The MGCP model is asymmetric: A master (MGC) controls a slave (Media Gateway MG). The MGs do not know the state of a call. They just open voice channels (PSTN side) and RTP streams (IP side) and pass voice between IP and PSTN.

➔ Gateway decomposition:
Media gateway does e.g. ISDN bearer to RTP.
MGC is feature server and implements the intelligence.

10. MGCP - RFC2705/2805 (4/6)

➔ MGCP applications:

1. Virtual Trunking with MGCP and H.323 (see picture Slide 40):

➔ Virtual trunking means the replacement of traditional PSTN trunk lines with IP connections. This saves money since the IP backbone networks of providers (e.g. Swisscom) have more and more capacity (for data applications). Traditional PSTN lines however become more and more costly to maintain (expensive equipment, expensive spare parts, old technology).
Additionally with VoIP the physical and data link layer are independent of the application („IP over anything"). This means that providers can upgrade backbone links with higher capacity lines independently (which is not the case in the TDM world).

➔ MGCP is used to control the media gateways (open/close voice channels).

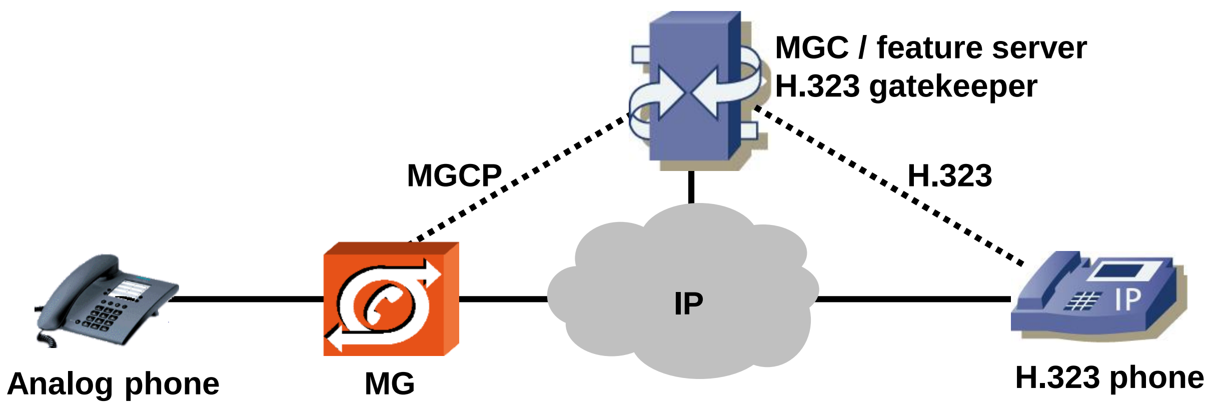
➔ The call control box controls the MGC with the H.323 protocol.

---

2. Gateway for analog phones:

➔ Old analog phones are hooked up to an IP (VoIP) network through MGCP gateways.

➔ The MGC controls the generation of ringing voltage, dial tone, busy tone etc. on the analog phone port.



10. MGCP - RFC2705/2805 (5/6)

➔ MGCP messages:

1. MGC → MG (from controller to gateway):

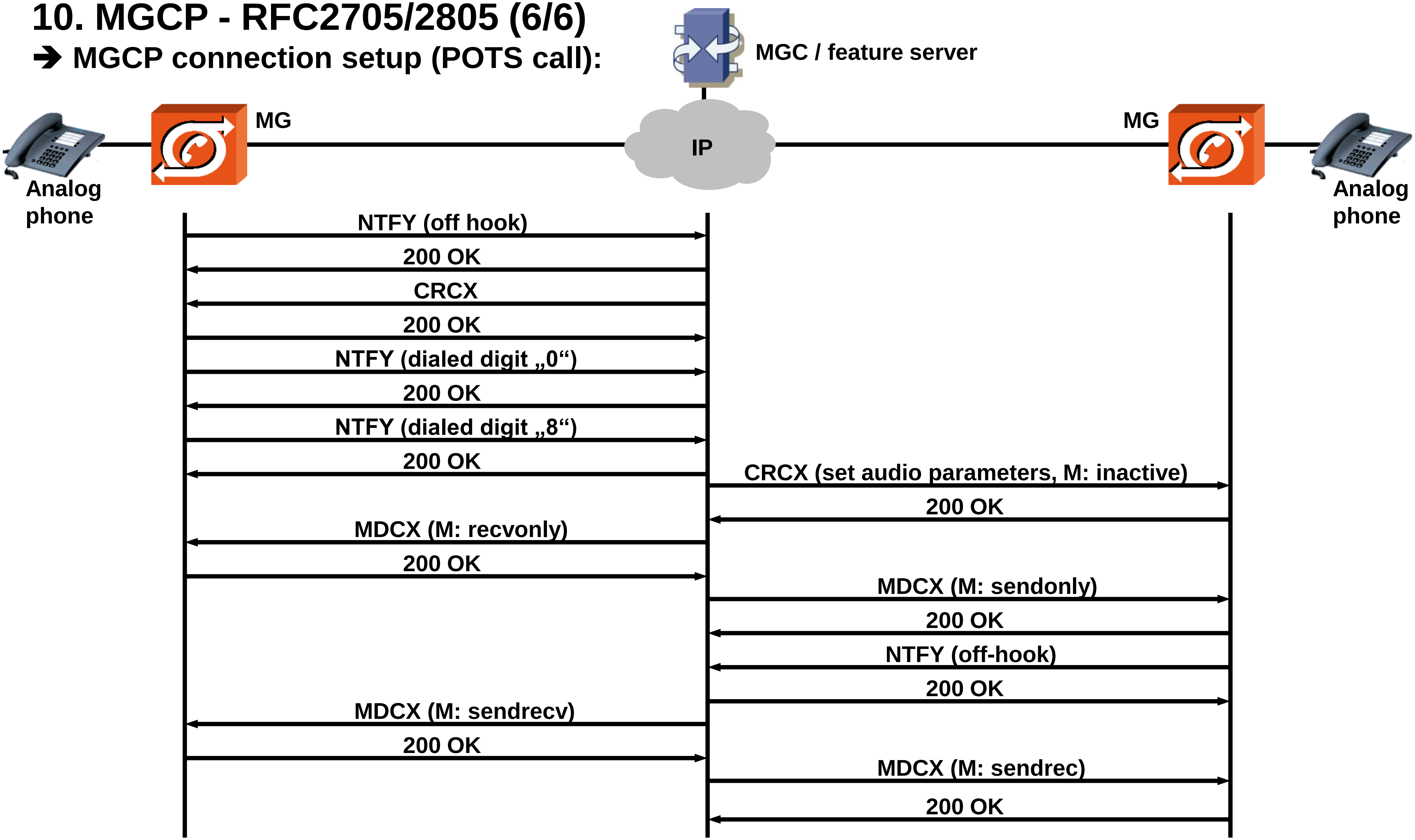
• Create Connection	CRCX	Create a new connection between two endpoints.
• Modify Connection	MDCX	Modify parameters of an existing connection (e.g. codec).
• Delete Connection	DLCX	Delete an existing connection.
• Notification Request	RQNT	Instruct the MG to watch for specific events or generate signals (e.g. off-hook event).
• Endpoint Configuration	EPCF	Instruct the MG about coding characteristics of an endpoint.
• Audit Connection	AUCX	Audit the status of a connection (check if connection still ok).
• Audit Endpoint	AUEP	Audit the status of an endpoint (properties, capabilities, events and signals).

2. MG → MGC (from gateway to controller):

• Restart in Progress	RSIP	Inform the MGC about MG restart or state change of endpoints.
• Notify	NTFY	Inform the MGC when requested events occur (if requested by MGC with RQNT message).
• Delete Connection	DLCX	Inform the MGC about the deletion of an existing connection. This message is usually only used in MGC→MG direction.

10. MGCP - RFC2705/2805 (6/6)

➔ MGCP connection setup (POTS call):

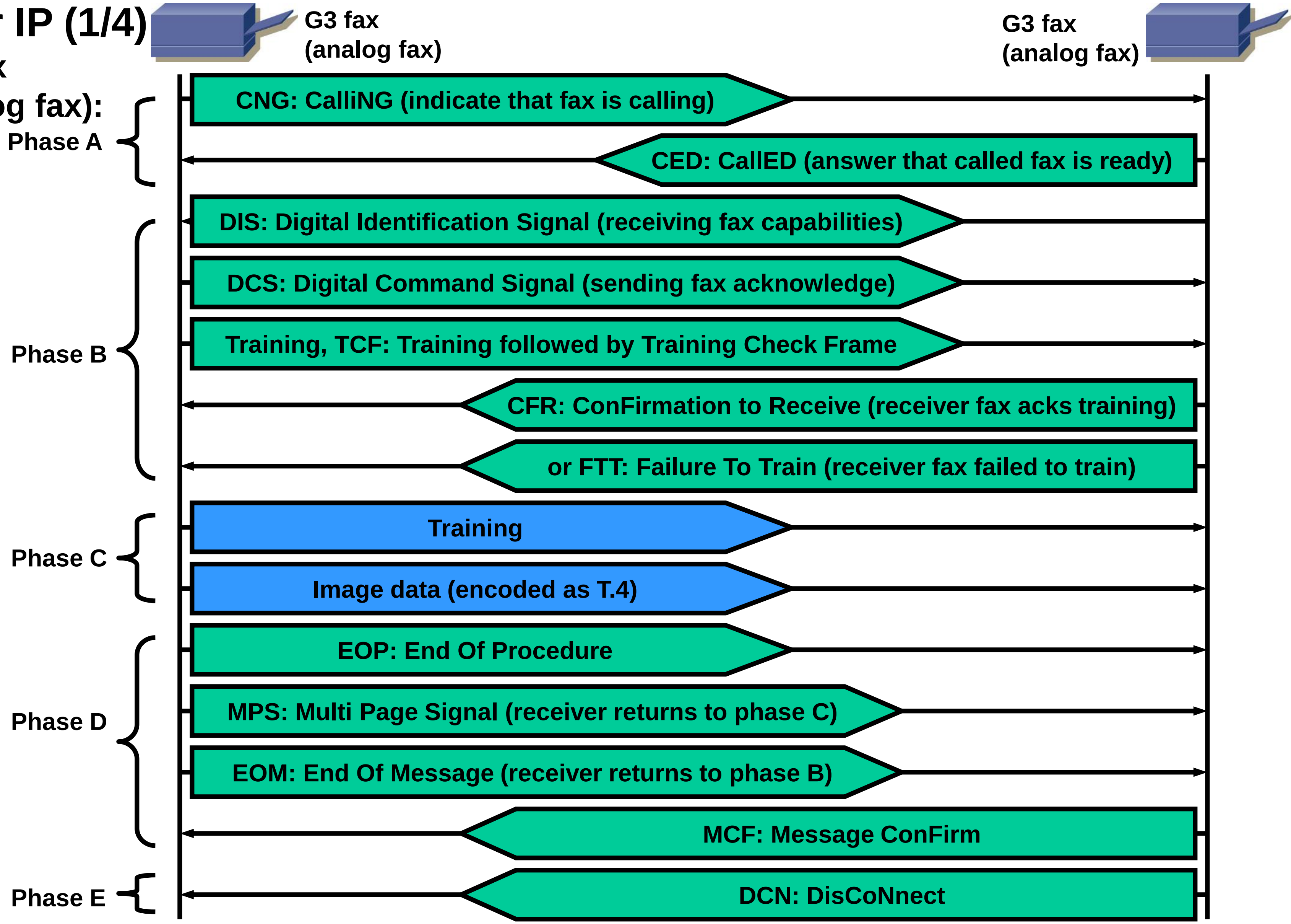


11. Fax over IP (1/4)

➔ T.30 (G3) fax protocol (analog fax):

T.30: HDLC frame

T.4: Image Data



11. Fax over IP (2/4)

➔ Problems with Fax over IP (FoIP):

1. Synchronisation:

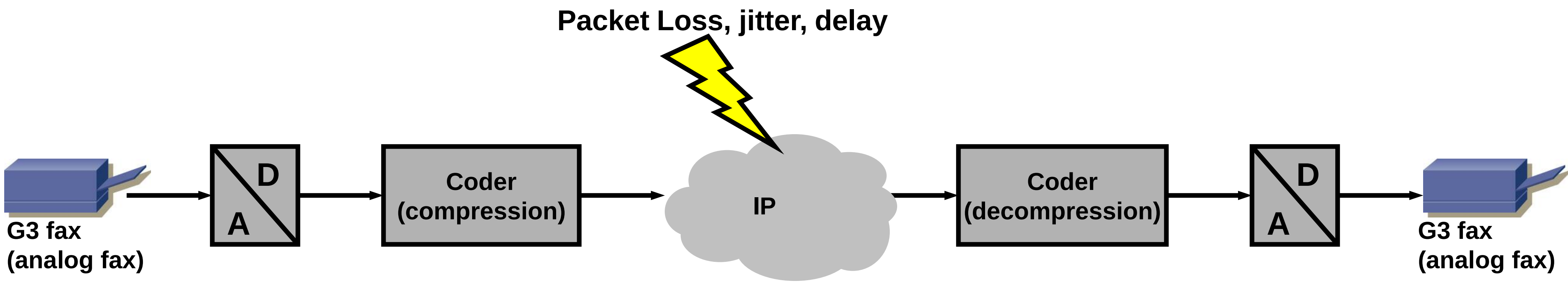
**Fax needs tight time-synchronisation between sender and receiver.
This synchronisation is given in PSTNs. But if fax is transmitted over a packet based network (e.g. IP) this synchronisation may be lost (packet loss, delay, jitter).**

2. Compression codecs:

The voice codec compresses the fax („lossful compression“) thus the receiver cannot 100%-ly reconstruct the initial fax signal (fax signal tones arrive distorted at the receiver).

3. Packet loss:

**T.4 images are composed of black and white traces. If the change between black and white is lost (due to packet loss in netwok) the image can not be reconstructed at the receiver.
If portions of fax control frames (HDLC) are lost the fax transmission will abort prematurely.**

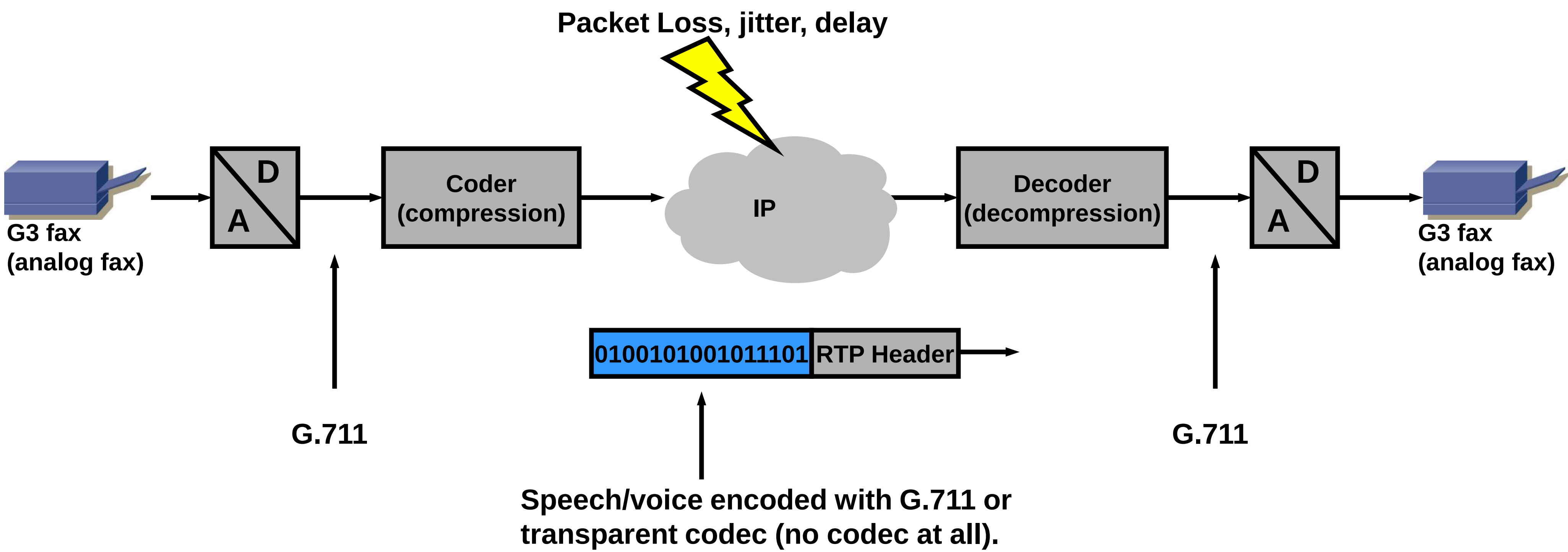


11. Fax over IP (3/4)

➔ **Solution 1:**

Use „loss-less“ codec (G.711 or transparent):

- ➔ **Works reasonably good, but some fax brands/models will exhibit problems.**
- ➔ **Even though fax is half duplex this solution will consume full-duplex bandwidth.**



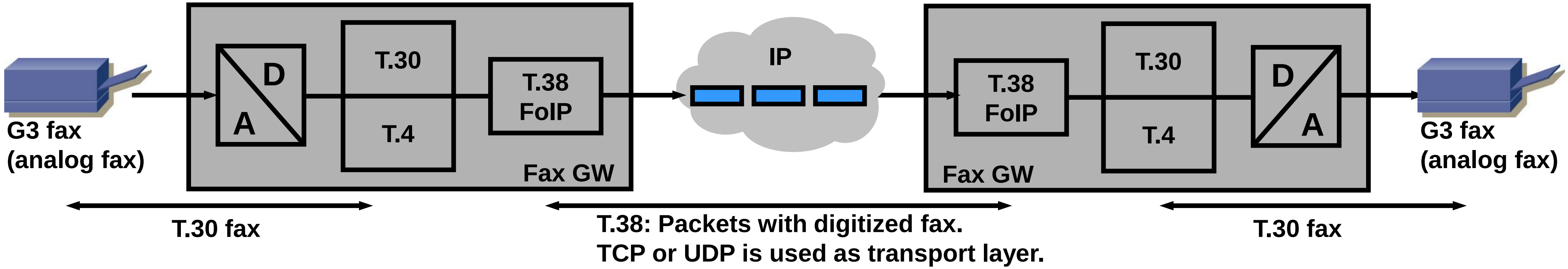
11. Fax over IP (4/4)

➔ **Solution 2:**

Use a fax over IP protocol:

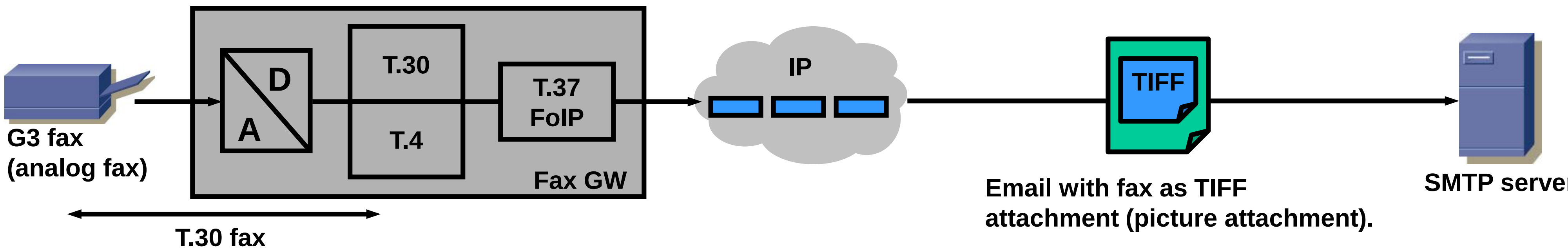
A. T.38 (real-time fax over IP):

The fax signals (tones) are converted to messages and sent to the receiver fax where these messages are converted back to tone signals (e.g. CNG tone signal is converted to *T.38:T30 IND:CNG* message).



B. T.37 (store and forward fax):

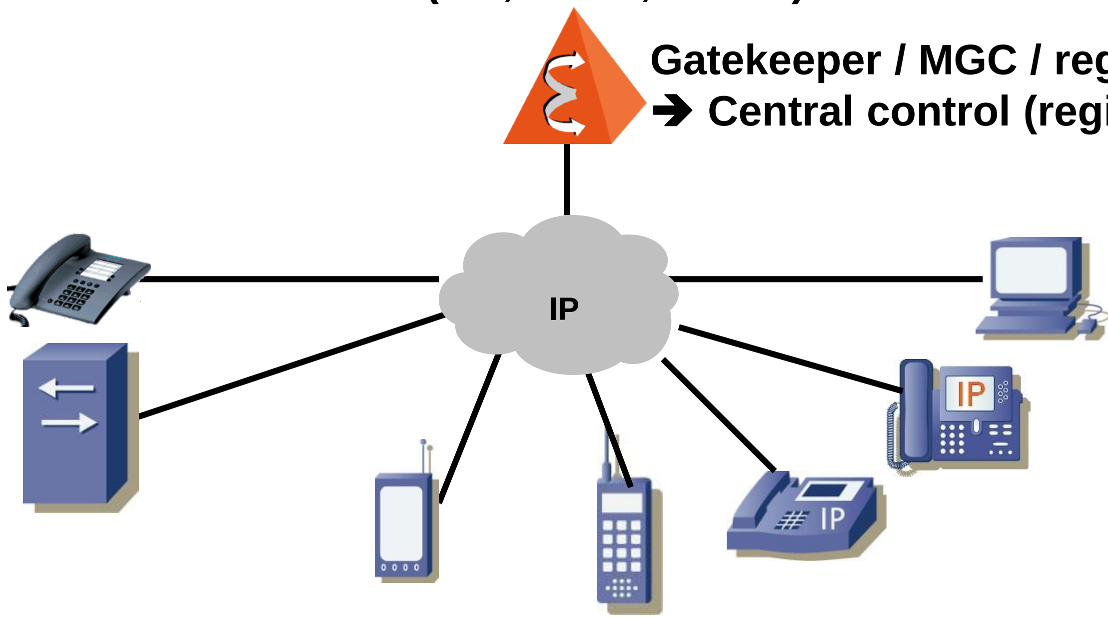
Fax image is converted to a picture (e.g. TIFF) and simply sent as email attachment. While this works fine it has the drawback that there is no acknowledgment that the fax arrived at the dest.



12. SIP / H.323 / MGCP centralized model vs. Skype peer2peer model

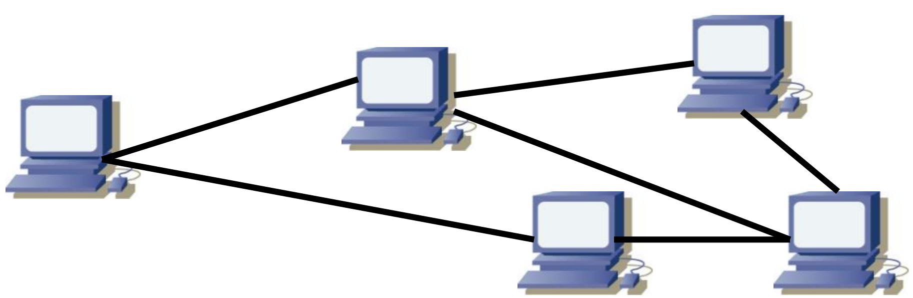
➔Centralized model (SIP, H.323, MGCP):

Gatekeeper / MGC / registrar:
➔ Central control (registration, admission, accounting...)



VoIP clients (hardphones, softclients, gateways, mobiles)

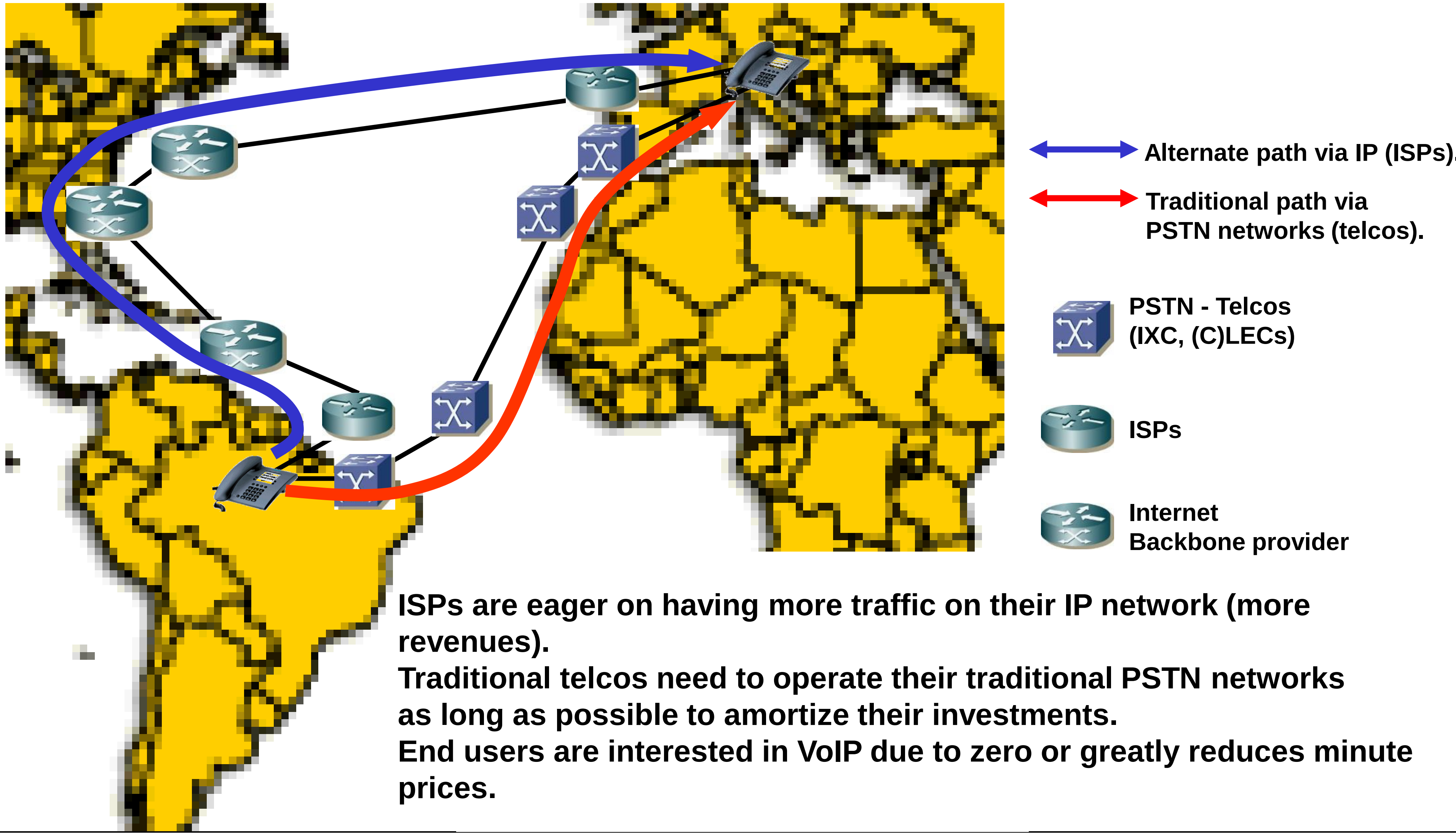
➔ Peer2peer model (Skype http://www.skype.com/):



+ No central control.
+ No single point of failure.
+ Bandwidth demand distributed.
- How do billing?

13. VoIP regulatory issues (1/2)

➔ VoIP bypass threatens the business of the traditional telcos!



ISPs are eager on having more traffic on their IP network (more revenues).
Traditional telcos need to operate their traditional PSTN networks as long as possible to amortize their investments.
End users are interested in VoIP due to zero or greatly reduces minute prices.

13. VoIP regulatory issues (2/2)

➔ VoIP bypass threatens the business of the traditional telcos!

1. License:

In most countries offering voice services requires a license and thus used to be restricted to only a handful of providers. VoIP technology allows almost everyone to offer voice service (relatively cheap equipment, ease of deployment).
Therefore VoIP is banned in many (developing) countries (originating and terminating traffic).
Some countries impose artificial restrictions on quality; e.g. Hungary demands VoIP providers to guarantee min. 250ms delay and only then grant a license to a provider.

2. Ease of deployment:

Traditional voice service (TDM) means multi-million investments. VoIP on the contrary is a technology that allows to start off with offering voice services at comparatively little costs.

3. VoIP in different markets:

- **Worldwide: VoIP already accounts for 10% of long distance calls.**
- **EU: no regulations as to VoIP (no license needed), but slow adoption since the quality is considered inferior and voice service already cheap; EU's liberal stance on VoIP may change when the national carriers start to lose significant amounts of money. VoIP in Western Europe is mostly restricted to enterprise applications.**
- **Mid East: VoIP banned in most countries except VAE.**
- **Latin America / Eastern Europe / India / China: strong VoIP development.**
- **U.S.: VoIP is considered an advanced service and as such not subject to FCC regulations.**