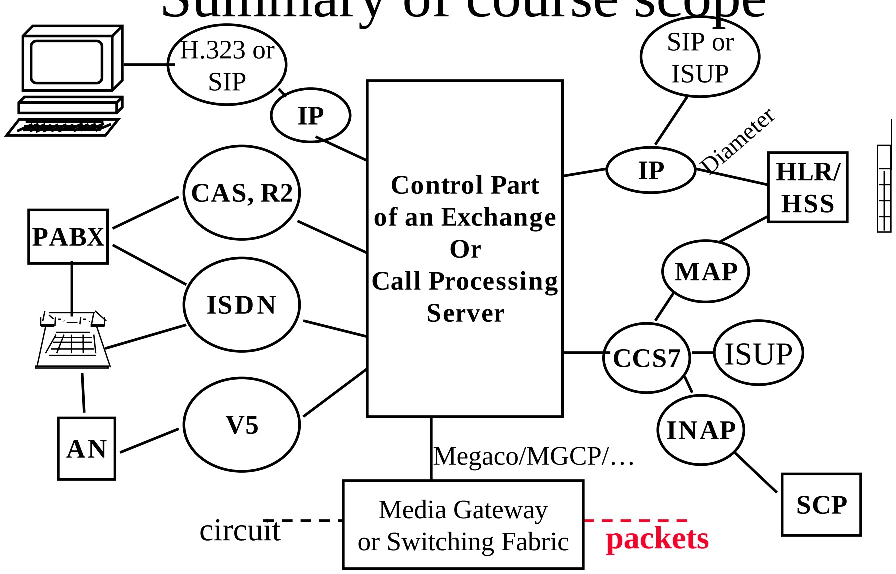
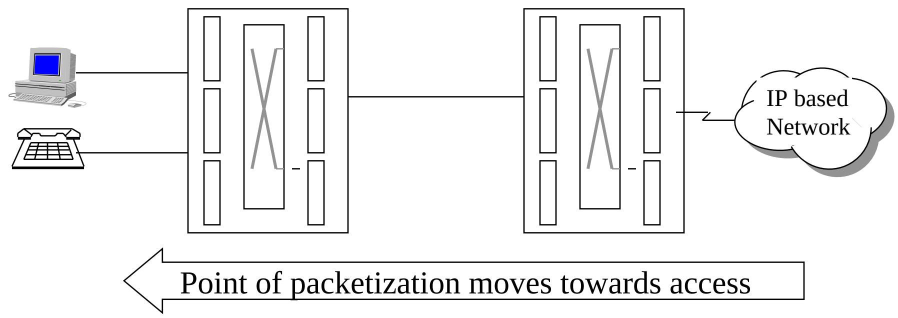
IP Telephony

Overview of IP Telephony
Media processing, RTP, RTCP
Packet header overhead
Quality of Service

Summary of course scope

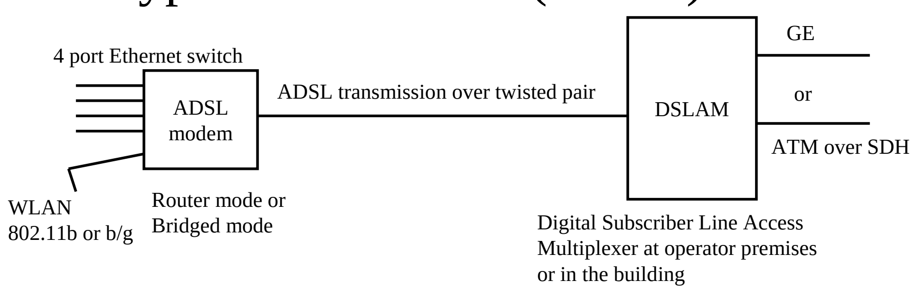


Data traffic already dominates voice in volume. Therefore Data will drive the Network Architecture.



- Broadband Networks are based on packet switching (later optical switching)
- BB network emerges from the existing Internet
- Each step of Development pays for itself.

A typical Broadband (ADSL) home



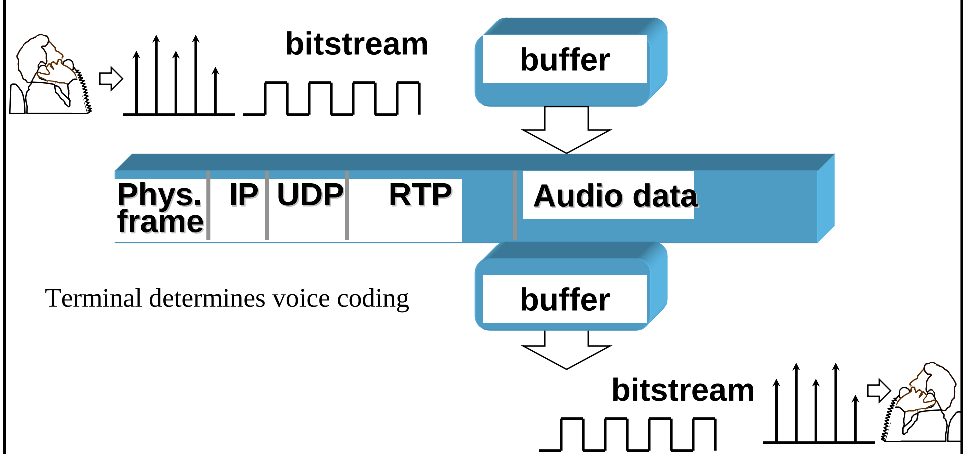
- ADSL →ADSL+ →ADSL2+ →VDSL etc
- In Japan Fiber to the Building (FTTB) has overtaken ADSL in new BB connections per month, provides > 50Mbit/s service to end-users
- Even in Europe access speeds are growing > 1Mbit/s
- Home wiring at least Cat5, twisted pairs to homes are mostly much worse than Cat5.
- 166M global BB subscribers in 1Q 2005, 50% growth per year.
  << 2500M cellular subscribers!
- Growth of BB takes place in developed and newly industrialized countries. Cellular growth in newly industrialized and developing countires.

Operator transport core is moving to packet technology

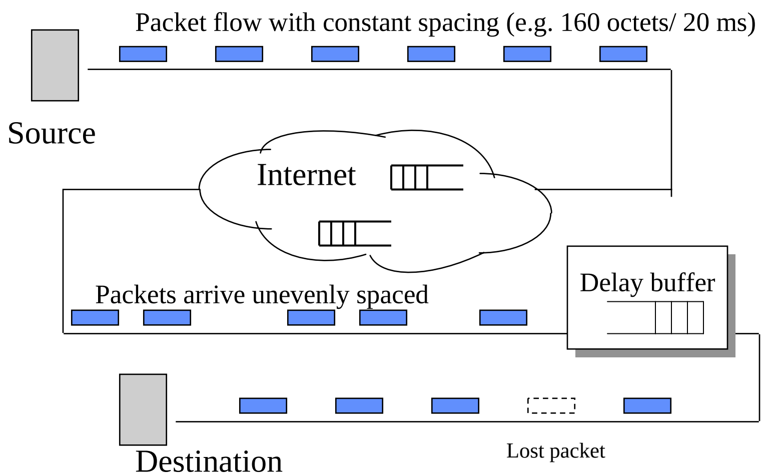
- 1GE, 10GE, Metro-Ethernet etc are replacing SDH →
  - PCM-timeslots are disappearing
  - TDM timeslots need to be carried over Ethernet: e.g. take 20 bytes from a timeslot and carry that in an Ethernet frame
- With GE transmission, Switching Fabric is packet based but still e.g. CSS7 MTP is used:
  - Must open GE frames, retrieve timeslots carrying a signaling channel (n * 64 kbit/s) and feed them to MTP level 2 etc…
- Such arrangements are ***temporary*** and stretch the life of PSTN, GSM and other CS networks.

VoIP in action

Coded samples (G.711, G.729B, G.723.1)



Delay variance is compensated at reception by buffering



Limitations of IP for VOIP

- Security – imagine that hackers could listen to all your conversations or find a record of with whom you have communicated.
- QoS – Internet is a Best Effort network, under heavy load quality is very poor and can not be quaranteed or even assured.
- Availability performance is poor, fortunately tends to recover from failures.
- IPv4 is an A-subscriber's network – NAT traversal has slowed down development.

IP creates a level playing field for everybody, anyone can be a Service Provider → Competition is hard → Operators can not assume to continue their business as before, but users reap the benefits.

IP Telephony Standardization is active on de-jure and de-facto fora

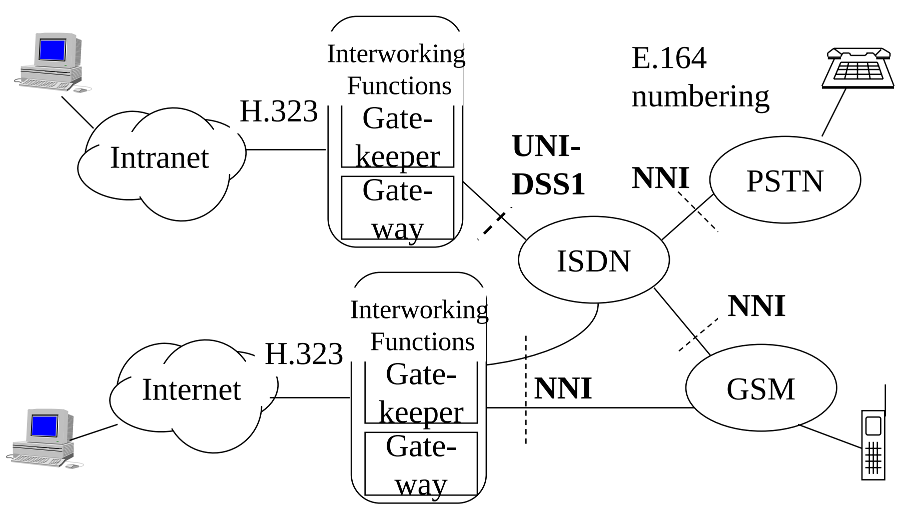
- ITU-T - H.3xx, H.2xx series
- ETSI - TISPAN (NGN – next generation networks… took over from TIPHON which was a project)

IETF working Groups

- IPTEL (IP telephony) and PINT
- MMUSIC (Multiparty Multimedia Session Control)
- SIMPLE – SIP for Instant Messaging and Presence Leveraging Extensions
- AVT – audio video transport
- ENUM – tElephone NUmber Mapping
- Megaco – media gateway control
- Rohc – robust header compression
- SIP – Session Initiation Protocol
- SIPPING – Session Initiation Protocol Investigation
- SIGTRAN (ISUP and other CCS7 over IP)

VOIP - Voice over IP by IMTC - Int'l Multimedia Teleconferencing Consortium

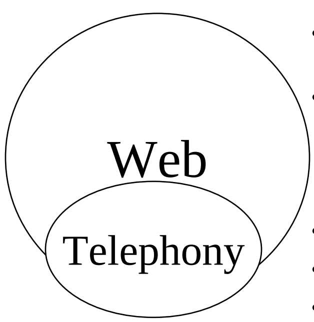
TIPHON/ETSI specified IP Voice to PSTN/ISDN/GSM Interworking



H.323 products are available

- ITSPs initially concentrated on H.323, now moving to SIP
  - H.323 inherits a lot from ISDN Q.931 signaling, coding style is binary.
- MS Netmeeting, Intel Videophone, Netscape Conference are examples of H.323 clients
- H.323 products have been on markets for years
- Gateways and Gatekeepers/Call managers are available
- SIP has been taking the lead over past 4…6 years but takes time (MS Messanger uses SIP)

IETF alternatives to H.323 pursue Integration of Telephony to the Web

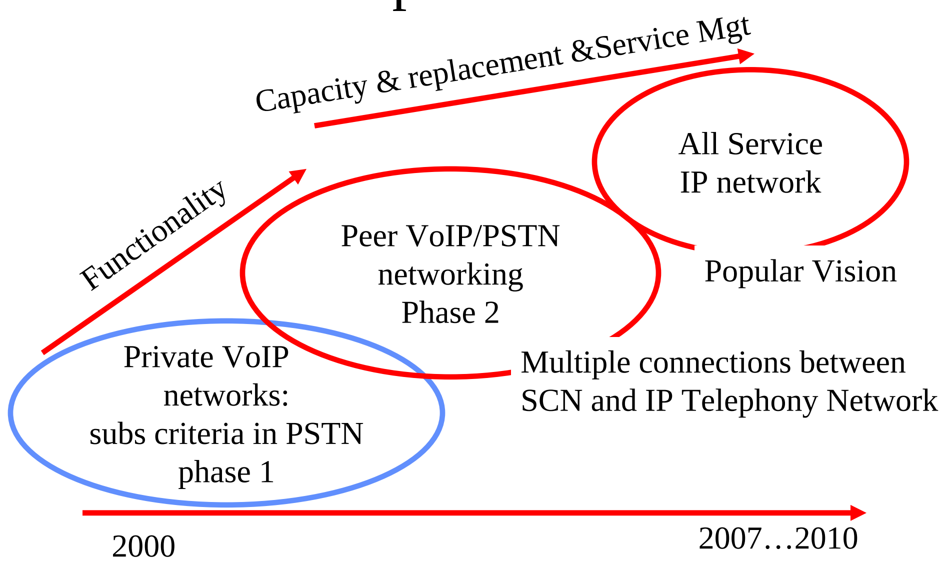


- AVT - Audio Video transport (…RTP)
- PINT worked on Click-to-Dial, Click-to-Fax, Click-to-Fax-Back "www-buttons". The idea is to integrate www to IN
- Mmusic (now SIP group) works on SIP - idea is to use web-technology to absorb signaling
  - SIP has been adopted by 3GPP for 3G packet telephony
- Media Gateway Control (Megaco)
- SIGTRAN works on C7 over IP
- ENUM - numbering info in DNS

Latest move is the emergence of Peer-to-Peer VOIP

- Designers of KaZaA released SKYPE – a peer-to-peer VOIP application
  - No network based servers are needed
  - Has node/supernode structure like KaZaA (one can claim that supernodes are ”network based servers”)
  - The application learns the capabilities of the computer, finding a well connected machine, will become a supernode.
- In p2p, the search of the callee is integrated in the application
- Everything in SKYPE is secret, even monitoring of signaling is forbidden by licence conditions!
- NAT traversal specification was too slow for SIP → opened door to proprietary solutions.

Roadmap to the Future



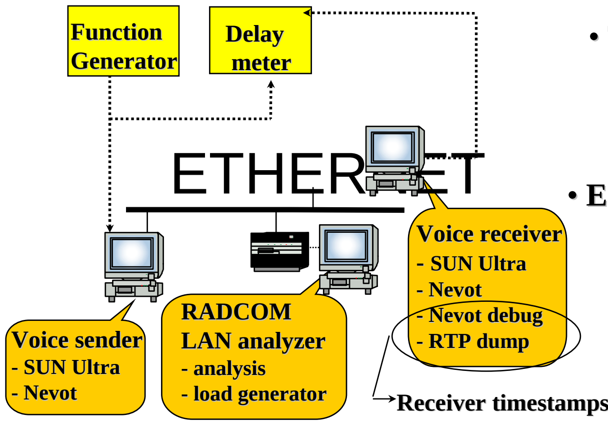
Why does the Introduction of VOIP take so long?

- Business case for an operator?
  - Kills PSTN → compulsion to protect a cash cow.
  - In VOIP it is difficult to maintain time based charging; in BB networks such as the Internet, it follows from the economic theory that flat rate pricing is efficient.
  - Voice is moving to cellular networks, investing in wireline is not attractive.
- QoS can be tolerable only in BB network. E.g in Finland and in Europe BB penetration has been slower than e.g. in Japan and South Korea.
- VOIP is not a driver of change because the business opportunity is poor. But I believe will happen anyway when the time is ripe.

Interoperability Issues

- Signaling and Call control
- Quality of Service
- Telephony Routing and addressing
  - Input Information gathering
  - Alternative routing over IP
- Service Management in the hybrid network

IP Voice in Ethernet - Delay is in the Workstation (IPANA -97)



- **Terminal delay:**

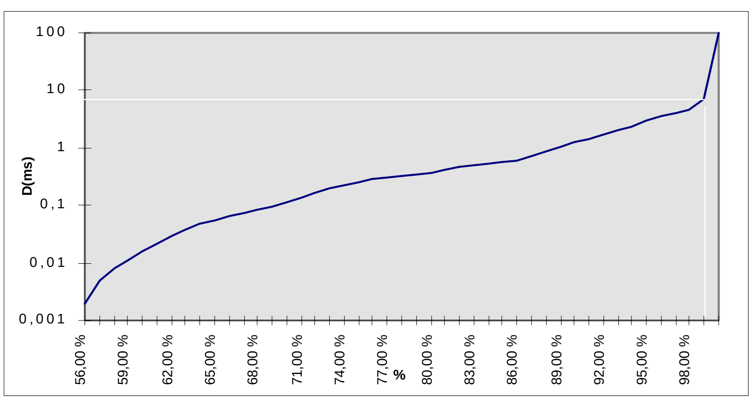
	Delay
HW	**8.9 ms**
VoIP Client	**103.9 ms**

- **End-to-end delay**

Packet length	Delay
0.02 s	**104.5 ms**

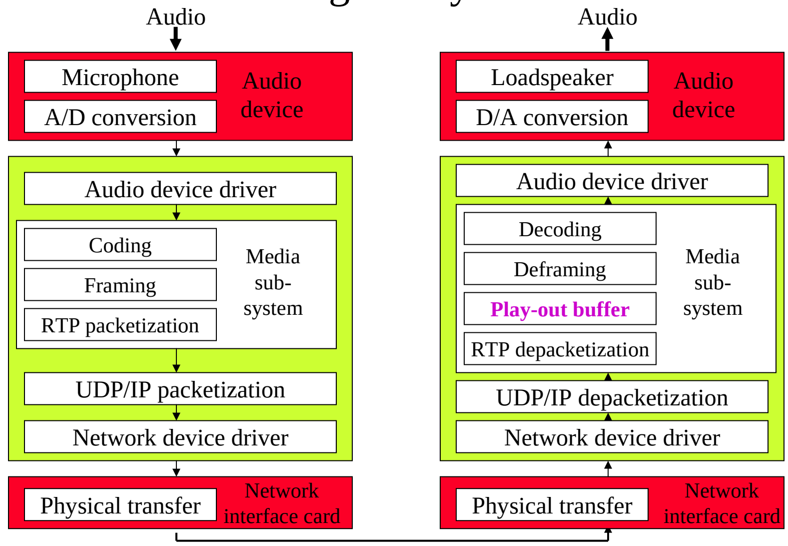
Difference = network delay < 1ms

Packet spacing difference in a campus network



- In the public Internet lack of bandwidth, congested routes/links and underdeveloped charging are blockers to IP Voice.

Media processing path in terminals and media gateways



Delay in practical IP voice systems

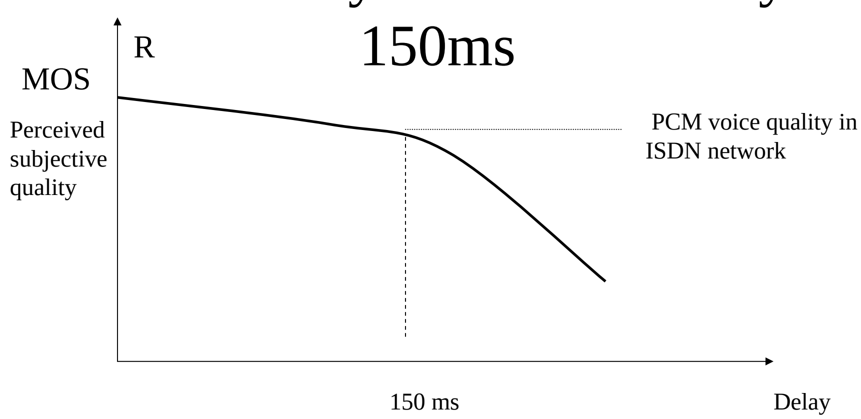
Delay component	ms	Explanation
Audio HW &device driver	0-100	Buffering
Algorithm	20-37.5	Sample length + lookahead time
Operating system	0 - 30	Depends on load and implementation
Coder	<5	Predictable delay in coding algorithm
Decoding	<1	Typically an easy process
Framing and packetization	<1	A small software delay
NIC and device driver	<5	Has some signifigance especially in WLAN
Network	0 - 500	In LAN about 1 ms, Dimensioning Issue!
Play-out buffer	0 - 100	At reception,depends on the state of the network
Synchronization	0 - 30	Audio device requests for data at constant intervals that can not be synchronized with packet arrivals. Avg = half a packet time

Source: M.Sc thesis by Jari Selin

Voice coding for IP networks

- IP networks are characterised by packet loss
  - coders that have dependancies between packets do not perform well (such as Cellular etc.)
  - even 5% packet loss may seriously degrade quality
- Higher than PSTN quality can be targeted:
  - Coding can be done at e.g. 16kHz (not 8 kHz like in PSTN),
  - packets can be variable length
  - in BB environment bit rates can be increased
- E.g. GIPS (Global IP Sound) provides proprietary codecs specifically designed for packet loss networks. E.g. sound quality stays good even at 30% packet loss (at avg 80kbit/s + packet overhead) acc to GIPS www-site.

Voice quality starts to degrade, when one way end-to-end delay > 150ms



Quality can be measured e.g. based on the E-model or using MOS –measurements.
MOS - Mean Opinion Score.

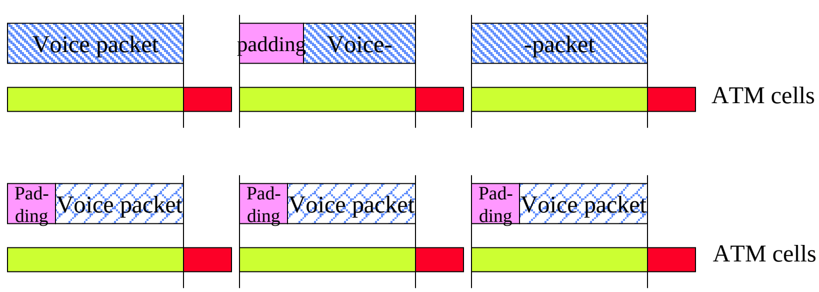
Packetization of voice in IP network

MAC -header	IP header	UDP	RTP	one or more voice frames

- Layer 1: e.g. bit stuffing, synchronisation
- MAC varies in different networks
  - In Ethernet, packets are variable size (delimiters, MAC addresses, checksums, channel identification etc),
  - a frame may have a minimum size (e.g. GE)
  - e.g. 5 bytes per ATM cell + AAL layer overhead.
  - E.g. in ATM also padding is needed.
- IPv4 header is 20 octets per packet. Alternative: IPv6 header is 40 octets!
- UDP is 8 octets per packet
- RTP is 12 octets per packet or more

Red is *header overhead*, green is payload.

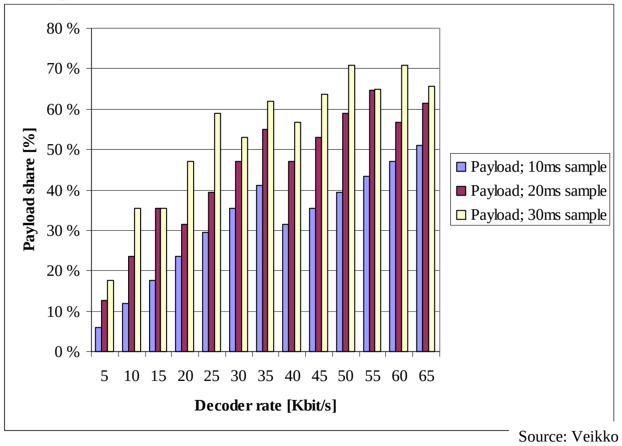
Packetization of voice samples in ATM cells



- Length of voice packet depends on coding method and the length of voice frame
- *Packet overhead* includes ATM headers and padding, which is needed in order not to increase packet delay.
- Generally, Packet overhead = packet header overhead + padding overhead
- We also talk about signaling overhead because users do not pay for signaling but only transmitted voice payload.

Example: ATM transport ->Packet header and padding overhead is significant

Percentage of voice payload when samples are carried over IP, UDP and RTP protocols, and IP is carried over **ATM**



Source: Veikko Brax (Lic thesis)

Packet overhead lowers voice transport efficiency in IP networks

- Voice transport efficiency in IP networks should be compared to voice transport efficiency in PSTN. On a voice channel in PSTN there are just the voice bits. In a PCM system, the overhead is one timeslot out of 32 per 2M – i.e. it its very low. In an IP network the efficiency of voice transport can be measured as the relation of voice bits to total bits tranferred on a link *in the packets that carry voice bits*. In addition to voice bits, packets have layer 1 overhead (small), layer 2 header (and padding) overhead, IP and transport layer overhead.

E.g. voice coding rate is 12 kbit/s and 20 ms samples are used:
voice frame = 20 ms x 12 kbit/s = 240 bits = 30 bytes

If one voice frame per packet already RTP+UDP+IPv4 header overhead is larger than the payload. Layer 2 and 1 overhead makes the situation worse.

Packet overhead – so what

- We can ignore high packet overhead on Broadband links when the share of voice traffic is low or when alternatives are more expensive
  - e.g. a BB connection at home on DL/UL 512kbit/s/512kbit/s and the home has e.g. two IP phones.
  - we can not ignore the overhead in cellular networks, not even in 3G!
- If we increase packet size (e.g. from 20 ms voice samples to 40 ms voice samples), packet loss tends to become a bigger problem for quality of voice and end to end delay increases which also tends to sacrifice quality of voice.
- Packet overhead can be reduced by header compression (IP+UDP+RTP)
  - Can be applied for access networks but not for the core, because IP routing needs at least the IP header and naturally lower layers are needed both in access and the core always.
  - In the core statistical multiplexing works very well, and UDP+RTP compression is not used (could not be used simultaneously with IP+UDP+RTP compression)

Packet header compression

- For IP+UDP+RTP for example in a cellular access network
  - needs to be supported by terminals and some device in the access network prior to IP routing needs to be applied
- The idea is to look to two consecutive VOIP packets, take the difference and replace the packet header overhead in a packet by the difference: overhead+payload → diff+payload
  - this essentially means that from connectionless IP packets we move to virtual connections between the terminal and the header decompressor. Both will need to maintain state of the connection. The state contains the overhead bytes that need to be used to replace the compressed diff bits. Because compression removes IP addresses, it can not be applied in the core.
  - Header compression makes sense in the access, e.g. cellular access. A method that covers as many headers as possible is best in access. But only a method that does not compress IP headers could be used in the core. Only one methods can be used for one packet. So the reasonable choice is to compress also IP headers and not use compression in the core at all.
- Different compression methods need to be compared in terms of how they tolerate errors, e.g. packet loss.

Silence suppression replaces silence by comfort noice that uses less bits than voice

- In practice, each speaker in a telephone conversation speaks less than half of the time. Silence can be detected by a coder and replaced by a low bitrate comfort noice packet stream.
- Functionality resides in terminals: coding at the source and decoding of comfort noice at the receiver
- Improvement of voice transport efficiency:
  - Does not help in terms of maximum bitrates in access because during a talkspurt full coder rate is needed.
  - Helps in the core by reducing the required capacity by almost half due to efficient statistical multiplexing of many voice streams on BB links. (stat. multiplexing is a basic feature of the IP network – IP is efficient in carrying packets from different sources on a single link – see courses on teletraffic theory)

Value Tradeoffs of VOIP

- The fundamental reason for low voice transport efficiency is that IP networks have been designed for data transfer. For data, large packets are used. Due to delay requirements, small packets must be used for voice.
- Low bitrate codecs bring only marginal benefit due to packet overhead (header and padding).
  - the lower bitrate codec is used the more difficult the header overhead problem becomes.
    - total bitrate is reduced but the share of payload goes down.
- It really makes sense to try to provide better than PSTN voice quality for BB customers rather than provide poor quality VOIP at low cost to modem users.
  - e.g. SKYPE

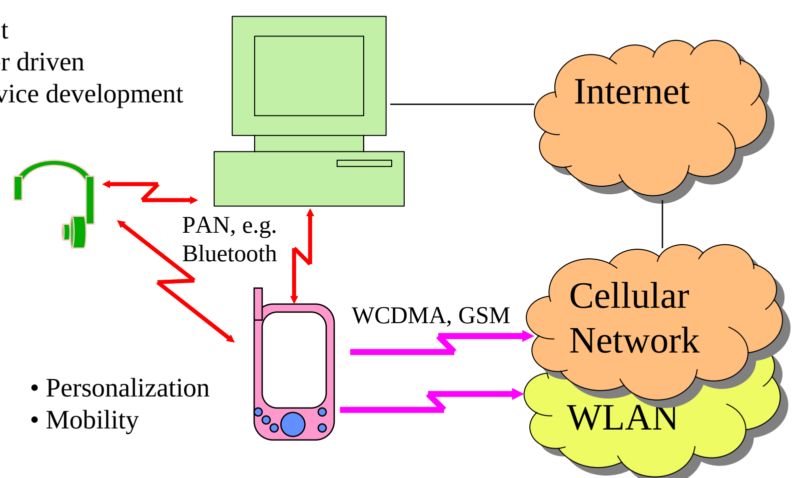
Why (operator) voice over IP, when ISDN/GSM work perfectly well?

NB: Voice brings currently ca. 90% of operator revenues!

- Integration of voice and data networks creates new services.
- Maintaining two networks is expensive.
- Data traffic grows >30%/year, voice ≈ 5%/y, volumes were approximately equal 2002. If trend continues, in 2010 share of voice will be < 10%, data will be 90% of all traffic.
- Cost of transmission is in free fall: xDSL, SDH, WDM - this trend is difficult to take advantage of using circuit switching: only one sample (8 bits) can be switched at a time cmp. E.g. 20 ms sample => 1 Gbit router is less expensive than an exchange with a 1 Gbit switch fabric.
- Terminals can do more -> consumer market economy helps.
- Once you have a BB data network, access speeds in Mbit/s, VOIP can be provided at marginal cost.
- Operator VOIP → NGN – Next Generation Network specification effort ongoing. Bold statements by incumbent operators about move to Voip during this decade (British Telecom is building already).

User view on VOIP

- Cost
- User driven service development



- Personalization
- Mobility

What is still missing in VOIP technology?

- Full telephony replacement capability
  - Emergency calling features
  - Security
  - Self configuration
  - No big deal: easy to solve, if there is a will (e.g. BT)!
- Networking capability
  - VOIP peering among ISPs (preparation under way in Finland.)
  - Operators still use PBX deployment model
- QoS
  - Controlled BE service for malevolent, greedy users?

→We know how to solve all these problems → deployment is a business issue rather than a technical one.

Real time Services in IP

RTP (RFC 1889)

RTCP - “ -

Telephony over IP

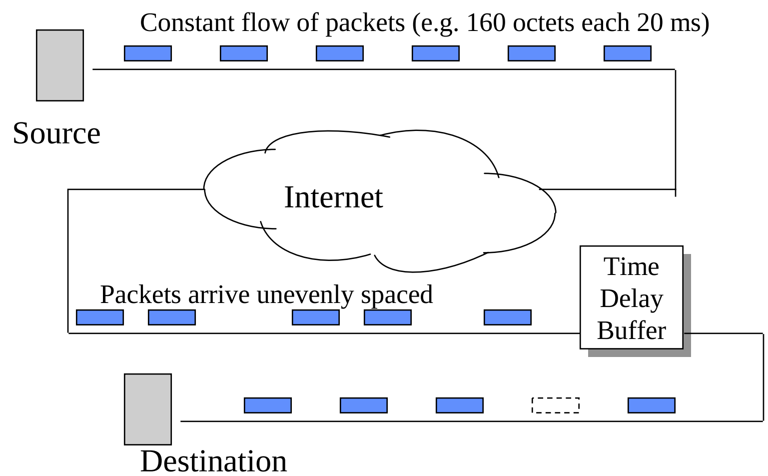
TCP is not suitable for real time services

Applications include

- Audio and video conferencing
- Shared workspaces
- Telephony
- Games
- Remote medicine
- ...

- TCP is point-to-point - not suitable for multicast
- TCP has retransmission for lost segments --> out of order delivery
- No mechanism for associating timing info with segments

Variable delay has to be compensated in the receiving terminal by delay buffer



Packet arrival process is characterised by delay jitter and packet spacing difference

Delay jitter = Maximum variance in packet delay in a session

Example: fastest packet arrive in 1 ms
slowest arrive in 8 ms.
Delay jitter is 7 ms.

Packet spacing difference is measured based on receiver clock only:

$$\text{Spacing difference} = [(t_i - t_{i-1}) - (t_j - t_{j-1})]$$

Unfortunately, jitter and delay do not behave nicely: they exhibit spikes over time. Spike is an abrupt increase in delay (and spacing difference).

End-to-end Delay and Round Trip time

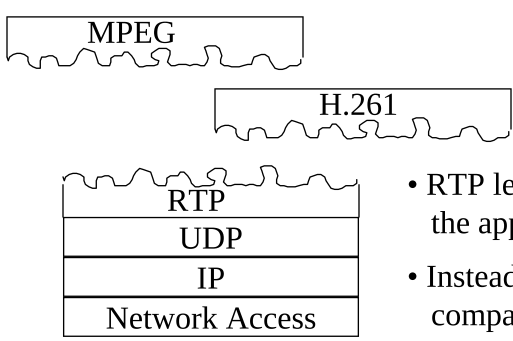
- End-to-end Delay can not be measured by Internet connected hosts directly, because their clocks are not synchronised.
- The sender can place its timestamp into a packet, and the receiver can send it back, so RTT can be measured easily. How the delay is split between the forward and backward paths can not be found out easily.
- For interactive voice quality, delay is important, but since we can not do much about it in a BE Internet, we concentrate on packet spacing difference and packet loss in the RTP design.

Soft real time communications tolerate some loss but need the following

- Low jitter and Low latency
- Ability to integrate real-time and non-real-time services
- Adaptability to changing network and traffic conditions
- Performance for large networks and large nrof connections
- Modest buffering requirements in the network
- Effective capacity utilization
- Low processing overhead per packet

Moore's law helps: The higher the speeds, the less packets spend
On any links and any node.

RTP - Real time protocol is a "sub-layer" library on top of UDP

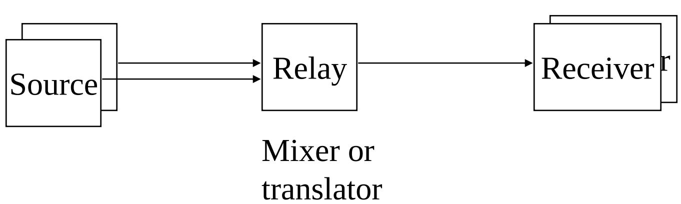


- RTP leaves recovery from loss to the application
- Instead of retransmission e.g. more compact coding may be chosen
- RTP provides sequencing

RTP supports the transfer of real time data among participants of a session

- Session is defined by
  - RTP port number (dest port in UDP header of all receivers)
  - RTCP - Real time control protocol port number
  - Participant IP addresses - multicast address or a set of unicast addresses
- For session set-up e.g H.323 or SIP - Session Initiation Protocol can be used

RTP transport model includes sources, relays and receivers



- A mixer will combine sources - e.g. add voice signals from all conference participants
  - for a two person telephone call no mixer is needed
- A translator may translate from one video format to another
- The relay will mark itself as the synchronisation source

RTP header

V=2 = version

0 4 8 9 16 31

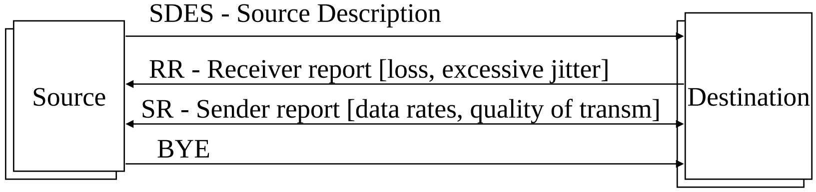
V	P	X	CC	M	Payload type	Sequence number
Timestamp						
Synchronisation source (SSRC) identifier						
Contributing source (CSRC) identifier						
Contributing source (CSRC) identifier						

P - Padding - indicates that last octet of payload = nrof preceeding padding octets
X - Extension - there is an experimental extension header
CC - CSRC count - Nrof CSRC identifiers following the fixed header
M - Marker - e.g. End of video frame, Beginning of talk spurt
Payload type - format of RTP payload.
Seq. nr - each source starts at a random nr and =+1 for each packet - determines order of packets with the same timestamp
Timestamp - value of local clock at source at generation of first octet of payload
SSRC and CSRC identifiers are generated at random

Main RTP functions are ordering of received packets and timely playout

- Sequence number gives the order of packets
  - say one in sequence is missing – when the playout time of the missing packet comes, e.g. the previous packet can be played out again to concel the error
- The order is not enough, the receiver must know the time difference between the playout times of two consequtive packets – timestamp gives exactly this as measured by the source of the packet

RTCP - RTP Control Protocol provides feedback among participants of the session



- RTCP packets may be multicast in parallel to RTP using another UDP port
- RTCP source is identified by plain text
- Few participants: RTCP reports are sent once in 5s Rate of reports is reduced to max 5% of session traffic if there are more participants

RTCP fixed header is

0		4	8	16 ... 31
V	P	RC/SC	PT	Length
SSRC of sender (or CSRC)				

V =2 = version, P - Padding, same as RTP

RC - Reception report block count in SR or RR

SC - Source item count in SDES or BYE

PT - RTCP packet type [RR, SR, SDES, BYE]

Length - lenght of this packet in 32 bit words - 1

SSRC - same as in RTP

Sender Report carries sender info and reception report blocks

Sender information is

NTP timestamp (MS word)	NTP is the wall-clock time when sending this report (used for round-trip time measurement)
NTP timestamp (LS word)	
RTP timestamp	RTP timestamp lets relate this report to RTP stream
Sender’s packet count	Packet and octet counts run from beginning of session
Sender’s octet count	

Reception report block

SSRC_i (SSRC of source)		SSRC identifies source
Fraction lost	Cum nrof packets lost	Fraction lost since last SR or RR, Cum loss is for the whole session
Ext highest seq nr received		16 LS bits= highest RTP seq nr. 16 MS bits= nrof times seq nr has wrapped back to zero
Interarrival jitter		
Time of last sender report		
Delay since last sender report		

SR is sent by party who is both sender and receiver !

Average inter-arrival jitter for a source is estimated as follows

S(i) = Timestamp from RTP data packet i
R(i) = Time of arrival of data packet i in RTP timestamp units
D(i) = (R(i) - R(i -1)) - (S(i) - S(i -1))
J(i) = Estimate of Inter-arrival jitter up to the receipt of RTP packet i

$$J(i) = 15/16 * J(i-1) + 1/16 * | D(i) |$$

- Receivers use the estimate of Jitter to adjust the play-out delay
- According to measurements the above exponential average is not always optimal
- The difficulty is that the jitter is hardly predictable and very unevenly distributed.

RTCP other packets

- RR are made of the fixed header + reception report blocks (see SR format lower part)
- SDES can carry
  - CNAME - Canonical Name
  - NAME - Real user name of the source
  - Email address of the source
  - Phone number of the source
  - TOOL - name of the tool used by the source

How to provide SCN-like QoS over IP?

- Integrated Services (use RSVP to make reservations in routers for each call!) changes Routers into SCN-Exchange - like systems. Does not scale well.
- DiffServ
  - mark voice packets with higher than BE priority at ingress
  - priority queuing in transit nodes
  - QoS determined by bottleneck links
  - QoS must be supported by link technologies as well (WLAN, …)
  - How to prevent voice from blocking BE traffic?
  - How to do Service Management?
  - Voice packets have high overhead - how to minimize?
- Overprovisioning

How is IP Telephony different from Circuit switched telephony?

Circuit Telephony

- Voice sample = 8 bits
- A- and μ -law PCM voice standard
- Reference connection gives network design guidelines => end-to-end delay is under control
- Wire-line telephones are dumb. Cellular phones are pretty smart
- Call control is tied to the voice path - IN is used to add service processing on the side.

IP Telephony

- Voice in 10…40 ms samples, Bits in a sample can be switched in parallel
- No single coding standard
- End-to-End delay is big challenge
- Terminals are intelligent - consumer market economics
- Call control is separate from voice path - first find out whether parties want and can talk, if yes, set-up the voice path

Note: Using todays technology IP Telephony is not less expensive in replacement nor green field investments in Corporate networks!

How realistic is the idea of replacing CSN with VOIP?

- There is more data traffic now than Voice traffic.
  - Data is growing much faster than Voice
  - Voice revenue is still >> ½ of operator’s networking revenue.
  - Voice is fast becoming mobile
  - Case BT: BT provides ADSL etc BB services, DSLAMs are upgraded to support separating Voice from other services, phones may be analogue, but all circuit switches are replaced by NGN (IMS + servers).
- CSN networking product development has stopped. All R&D effort in telephony goes to VOIP telephony
- Replacement Scenario in Finland: PSTN can be replaced, required max link capacities are 2,5…10Gbit/s. Present FUNET upgrade is to links of 10 Gbit/s and FUNET is just the University Network!
- PCs are still lousy phones!

VOIP deployment scenarios

- VOIP as a replacement to PSTN
  - voice is going mobile very quickly, so there may be nothing to replace?
  - if provided at flat price for BB customers, may still fly – a need for home user telephony for long conversations is difficult to fullfill with a cellular service
  - Because BB network (may be in a few years if not today) can accommodate any number of telephone calls, the service will be non-depletable. Consequently, market laws tell us that efficient pricing is flat rate (or taxation). Time based charging is expensive to run and will die at some point in time.
- Voice is a feature of IP based applications such as games, buttons on www-pages to contact a the owner of the www-page.
  - P2P VOIP a'la SKYPE falls under this category
- Mobile VOIP a'la 3G and NGN – next generation network