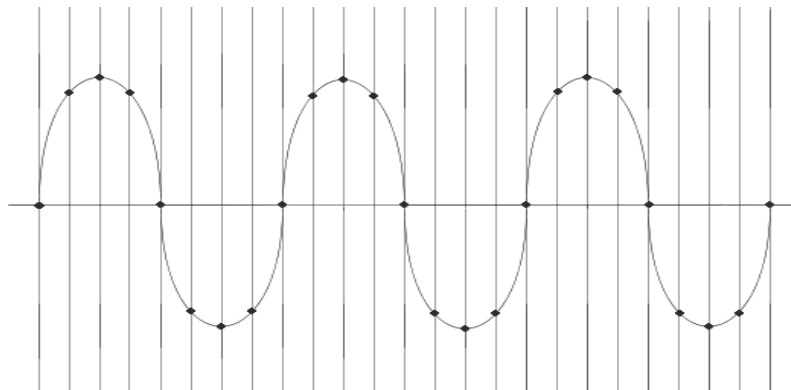


VoIP stands for Voice over Internet Protocol. At its most basic level VoIP simply means that Voice is digitised, short segments of voice data are assembled into larger packets and the packets passed as IP datagrams across an IP network. This is usually achieved using the Internet standard for voice signalling, the Session Initiation Protocol and Real-time Transport Protocol for the media stream data.

The Role of the Codec

In order for analogue waveforms like voice conversations to travel as packets of data over IP networks, they must first be digitised. This is done by slicing up the waveform into short segments at time intervals using a clock source and translating the signal amplitude of each slice into a number in exactly the same way as a CD contains digitised music.



Sampled in this way, a faithful representation of an audio waveform contains a lot of information. CD Quality audio for example contains 44,100 16-bit samples a second (times two for stereo). Even one channel (mono) speech would be nearly 700 kilobits per second of data for one side of a conversation.

Codecs are applied to this raw data in order to compress it to reduce the bandwidth required in a way that preserves the essential characteristics of speech. All codecs reduce the amount of audio information transported and hence the received quality.

The simplest and least destructive ones in widespread use, G711a/u, are logarithmic codecs. These have been in use for many years on the pre-VoIP digital circuits such as ISDN. G711 uses a reduced sampling rate of 8000 samples a second and also reduces the sample size to 8-bits using a logarithmic transformation which preserves both low and high amplitude impulses. The loss of the very highest frequency artefacts is insignificant for speech intelligibility and the resulting calls are therefore perceived as very high quality. The bit rate is however reduced by a factor of more than 10 to 64kbit/s over CD quality audio.

Other much higher compression codecs take this further and use a range of patented techniques to aggressively compress bandwidth requirements down as low as 8kbit/s or even 5.3 kbit/s by removing less important parts of the signal. Necessarily at these compression levels the perceived speech quality is reduced. That is why a mobile phone using the 14kbit/s GSM codec always sounds more “fuzzy” than a landline call which probably uses G711a at 64kbit/s over a digital backbone. G711a/u are still therefore seen as the standard for lossless, landline quality transport of voice over IP networks.

Call Control

Telephone networks do a lot more than simply transport audio signals from A to B. In order to control the setup and progress of calls, control signalling between the originator of a call and the prospective callee via their respective telephone service providers needs to take place. When a call is placed on a conventional analogue telephone, the caller tells their telco the number they want to place a call to via DTMF tones, their telco switch then examines the number and decides how to route it based on cost and connectivity. It signals to the switch at the next telco along that it has a call and asks if it can be connected, in general that next switch doesn't actually know the answer as it has to progress the call through the network. Some time later, assuming a failure like congestion has not been encountered first, the call setup message finds a path through the network to the recipient's switch.

At this point a signal may be sent back saying "Busy" or "Out of service" or "Ringing" etc. Some time later again the recipient may pick up the phone and the call actually completes. The call control protocol also needs to deal with other issues such as accurate agreement on call start time for billing, and routing eventualities such as call forwarding, call waiting etc.

Any implementation of VoIP needs a session control protocol that similarly controls end to end call setup when the calls are proceeding as media streams across a data network.

There are several VoIP session control protocols in widespread use: H.323, IAX, MGCP and SIP.

SIP tends to be the most widely used due to its simplicity, flexibility and Internet Standard status. In the following protocol example, we will use SIP terminology but for the most part the same concepts exist in the other major protocols.

Registration

The first part of call control, particularly for a terminal device such as a phone is usually registration. This is where the user device announces to its configured SIP registrar (PABX or switch) that it is present, signals its current network address and authenticates to prove that it is indeed authorised to receive calls.

As a result of the registration, the switch builds internal state so that it knows how to forward calls to the network of its registered SIP peer.



Invitation

When one of the parties (PABX or phone) wishes to initiate a call, it sends a SIP INVITE request. This request simply tells the other party that it wants to initiate a call and contains information such as the caller and callee address. It usually also contains information via the session description protocol (SDP) in the body about call parameters (codecs and bandwidth) that the requestor is able to support. On receiving the INVITE, the peer decides if it can handle the call and returns a response.

Example responses include:

200 OK – Call is OK to proceed, client then send an ACK and RTP streams are setup

302 Moved – Redirect to some other peer

486 Busy Here – Caller is busy

On receipt of a 200 – OK response, the client will normally send an ACK and then commence sending data on the chosen media stream, the call is now in progress.

Bye!

At the end of the call, one of the parties will hang up. The SIP device connected to the closing party sends a BYE SIP request. The other party responds with a 200 OK to indicate that this has been received and the media streams are terminated.

Bandwidth, Latency and Packet Loss

A key difference between Internet Protocol networks, and the existing TDM digital networks used to transport calls in legacy PSTN networks is the asynchronous nature of data transport and delivery caused by use of IP.

In a TDM system, the voice data is sent in a continuous digital stream from one end of the network to the other, synchronously at a guaranteed bit rate. In an IP system small segments of voice data, usually 10-40ms are assembled into a packet and this is sent to the far end over the network. The far end device then unpacks the data and converts the samples back into an audio stream.

If a packet is lost, or delayed so that it is isn't received by the time it is needed then there is a problem as the receiver has no data to "play" to the user. An intelligent receiver will normally insert silence, or guess at a suitable synthesised sample based on previous data in order to make this packet loss as transparent as possible but it will almost certainly be an audible distraction to the user if it occurs at all frequently.

In this context, network latency itself is not a significant problem as all samples are subject to a constant delay and are received in order at the far end. Jitter on the other hand, where some samples arrive quickly and others take much longer is almost as much of a problem as packet loss because it too can leave the receiver in a position where it needs data, has no sample to play and has to "bluff". Some VoIP systems employ long "jitter buffers" which mean that the receiver delays for some time after receiving a sample before playing it, effectively queueing up samples locally to smooth out jitter. This has the side effect of substantially increasing latency or lag in the conversation.

In practice few reasonably configured small to medium sized LAN environments have significant packet loss, latency, or jitter. Even if they do, these are comparatively easily resolved by strategic upgrades of key components based on modern switches which support traffic prioritisation to ensure VoIP packets are not delayed or lost due to bursts. These factors are more important in congested WAN environments where some form of Quality of Service management is needed to ensure that voice data does not suffer packet loss and jitter as the bandwidth used approaches capacity on an oversubscribed link.



Glossary

Codec - Compression/Decompression algorithm, a mechanism for taking a raw digital representation of a waveform and encoding it to reduce the number of bits required to represent it without destroying its essential audio qualities. Typically done to reduce the amount of bandwidth required when transmitting it over communication links.

H.323 – A session control protocol invented by an International Telecommunications Union committee, originally to control media streams over a LAN. It is used widely by PC multimedia applications such as Netmeeting and as a VoIP interface for legacy telecommunication backbones where its commonality with ISDN signalling simplifies gateway implementation. For many VoIP applications it is seen as both too complex and lacking in flexibility.

IETF – Internet Engineering Task Force – The standards coordination body for Internet protocols.

ISDN – Integrated Services Digital Network – a pre-VoIP standard for transporting voice as a digitised signal all the way to the subscribers premises. In the UK ISDN is still the main available mechanism for connecting digital phone systems to the BT network, where it is available as ISDN2e (2 64kbit/s voice/data channels) and ISDN30e (up to 30 64kbit/s voice/data channels).

IAX – Inter Asterisk eXchange protocol, an efficient simple call control protocol. Less flexible than SIP, it is designed to address the requirements of call trunking between PABX units. It enhances firewall traversal by combining signalling and media stream into a single data flow.

ITU – International Telecommunications Union – Organisation established by UN treaty that oversees technical standards used between national telecommunications operators.

MGCP - Media Gateway Control Protocol- A call protocol used within large, distributed VoIP systems with multiple media gateways. MGCP is designed for the specific purpose of controlling call routing in big, usually multinational networks with lots of gateways to the PSTN.

PSTN – Public Switched Telephone Network

RTP - Real-time Transport Protocol a standard packet format for delivering arbitrary real time audio and video data over IP networks. RTP concerns itself only with transporting the audio or video payload.

SIP – Session Initiation Protocol, the Internet standard for controlling the progress of multimedia streams. Most commonly used today for VoIP “calls”, but also designed video and other streams which are real time and progress on a caller/callee model. Designed to be both simpler and more flexible than H.323. It achieves this using the Internet model of simple text based request/response protocol exchanges and is agnostic about individual client capabilities like codec support.

TDM - Time-Division Multiplexing – A type of telephone data trunking or multiplexing where two or more data streams are transferred apparently simultaneously on one communications channel. Physically the data are taking turns on the channel in strict order and the communications medium has a defined bit-rate. There is therefore a fixed synchronous relationship between when single quanta of data is placed on the channel and when it is received at the far end with negligible jitter.

