Voice Over IP

This lecture describe Voice Over IP (VoIP), the term for telephony over IP networks.
After this lecture you should be able to: differentiate between the type of information carried by SIP and RTP protocols, explain why SIP is carried over TCP and RTP over UDP, compute voice call data rates after adding packetization overhead, compute PCM bit rates based on sampling rate and bits/sample, distinguish between and explain causes of near- and far-end echo, and identify some limitations of VoIP compared to the PSTN.

Introduction

Voice over IP (VoIP) is the use of packet-switched IP networks to carry "telephone" calls.

According to the ITU in 2012 there were about 1.2 billion fixed telephone lines and 2.1 billion active cellular phone subscriptions. In many cases we are interested in connecting calls to this Public Switched Telephone Network (PSTN) that provides near-ubiquitous connectivity.

Signalling Protocols

SIP (Session Initiation Protocol) is an IETF standard defined in RFC 3261 and is probably the most popular VoIP signalling protocol. It has also been adopted by IP-based 4G cellular systems.

SIP is a text-based protocol using client-initiated transactions similar to HTTP and other IP protocols.

Resources (e.g. subscribers) are defined with URIs. For example, sip:ecasas@bcit.ca could be a SIP subscriber identifier.

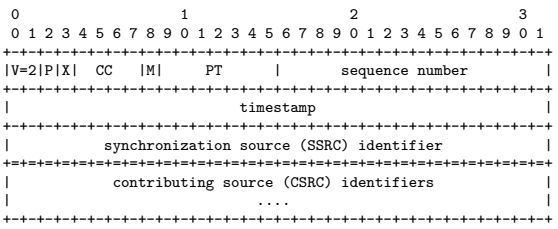
SIP protocol messages do not carry the speech data but instead allows negotiation of (typically) RTP connections between endpoints. In theory this is more efficient but in practice this separation makes it difficult for SIP to operate through NAT firewalls.

There are many other VoIP signaling protocols (H.323, XMPP, IAX, MGCP, Skype). Some are proprietary and some are public.

RTP

Although TCP-based signalling protocols such as SIP are used to set up and terminate calls, speech and video data are transmitted over UDP using different, lighter-weight protocols.

RTP (Real Time Protocol) is a simple protocol used over UDP to ensure that speech (or video) packets can be re-assembled at the destination in correct order and that missing data can be identified. The RTP header is shown below:



The sequence number is the packet number and the timestamp is typically a sample number. The source identifiers are unique, randomly-chosen values that identify a particular media source.

Exercise 1: Why can't the speech and video stream be transmitted using only UDP? Why might we want to avoid transmitting them over TCP?

Packetization Overhead

Each VoIP packet includes IP, UDP and RTP headers in addition to the PCM or vocoder data. The headers can add significant overhead. The minimum RTP header is 8 byte.

Exercise 2: How many bytes of header overhead are added to each packet assuming the smallest possible IP, UDP and RTP headers? If 64 kb/s PCM is being transmitted in 20 ms frames, what is the total data rate, including both headers and speech data? What fraction of that is for headers?

Exercise 3: Assuming the minimum header lengths, which has less overhead, TCP or RTP?

Packet Delay and Packet Loss

Since IP and UDP do not guarantee reliable delivery, packets will not only be delayed but they can be lost, duplicated or delivered out of order.

Since speech samples must be output at a constant rate, if a sample is not available when it's required then it may as well be lost because it cannot be used if it's "lost its turn."

Several techniques are used to cope with IP packet loss and delay.

The simplest technique is simply to use a first-in first-out (FIFO) output sample buffer to allow time for variable packet delays. As each packet arrives, it is placed in the appropriate location in the output buffer based on the sequence number. This is sometimes called a "jitter buffer" because jitter is the term for delay variability.

If a frame of speech data has not been received from the network when its samples are due to be output, various "error concealment" (or loss concealment) techniques can be used. One is to substitute the previous frame for the lost frame with the expectation that the sounds of the two frames are similar and that this will be less noticeable than outputting silence.

Voice Codecs

A codec (coder-decoder) coverts the speech waveform to and from a digital format.

The simplest example is Pulse Code Modulation (PCM, which is not actually a type of modulation). Typically the analog speech signal is compressed using a μ-law or A-law curve, low-pass filtered to less than 4 kHz and sampled at 8 kHz. The signal is sampled with 8 bits per sample and the bits are grouped into packets and transmitted to the destination.

Exercise 4: If the sample rate is 8 kHz and each sample is quantized with 8 bits, what is the bit rate in each direction?

VoIP systems can use speech coding to reduce the data rate. Vocoders (voice coders) exploit the redundancy in speech signals (e.g. periodic components) to reduce the required data rate. For example, the ITU-T G.729 speech coder produces reasonable ("toll quality") speech at data rates of 8 kb/s. However, this comes at the cost of additional complexity. As internet data rates increase, there is less need for speech coding. Today, vocoders are rarely used by commercial VoIP service providers.

VoIP systems that do not need to interface to the PSTN can use "wideband" codecs that use higher bandwidths and sampling rates to provide more natural-sounding speech.

Exercise 5: What is the maximum bandwidth and the bit rate if the sampling rate is 16 kHz and there are 10 bits per sample?

Echos

Echos are a major source of problems for VoIP systems.

There are two common causes of echos: 2-to-4 wire conversion and feedback between the speaker and microphone in a telephone set.

The telephone line to a subscriber is only one pair while inter-office trunks and VoIP calls have to handle the two directions separately. The function that splits the signal into two directions is called a hybrid. If the hybrid does not completely separate the two directions, some of the signal coming in one direction will go out in the opposite direction.

Exercise 6: Why can't trunks be bidirectional?

The other source of echos is simply that some of the sound output from the speaker or earpiece will be picked up by the microphone.

Echos can be a particular problem for VoIP systems because the use of jitter buffers adds additional delay to the signal and makes any echo more noticeable and disruptive.

VoIP systems use adaptive echo cancelers that compare the outgoing signal with the incoming signal to detect echos. If they are found, they can be canceled by subtracting a delayed and scaled replica of the outgoing signal from the incoming signal.

An echo canceler can cancel echos at the source (a "near-end" echo canceler) or echos generated at the far end (a "far end" echo canceler).

The near-end echo canceler would try to eliminate the near-zero-delayed component being transmitted back to the remote end. The far end echo canceler cancels echos coming back from the remote end with a delay of twice the one-way propagation delay plus

the jitter buffer delay (possibly totaling hundreds of milliseconds).

Since the echo in the remote environment and the jitter buffer lengths can change, the echo canceler is always adapting. This can result in occasional artifacts in the audio as the echo canceler adapts.

Customer Premises Equipment

An ATA (Analog Telephony Adapter) is an adapter that interfaces an ordinary POTS phone to an IP network, typically using an Ethernet interface. It has analog electronics to provide BORSCHT functions and a processor that implements the VoIP signalling and media transport protocols. The following figure[1] shows a typical ATA with two POTS and one Ethernet interfaces.

Another widely-advertised example is the "Magic Jack" that bundles an ATA and PSTN gateway service.

The ATA functions are often integrated into "residential gateways" that include a cable, ADSL or PON interface as well as POTS, Ethernet and video interfaces that allow service providers to provide "triple play" service.

Exercise 7: Which customer-facing interfaces provide which of the three services?

VoIP functions can also be included in a "soft phone" which is software that emulates a telephone using the computer's audio interface and graphical user interface. Skype is a popular proprietary example.

[1] From wikimedia.

PSTN gateways

A server or router that includes TDM ports (e.g. T-carrier or SONET) can act as a gateway to the PSTN[2]. This router has to implement both the VoIP and PSTN signalling protocols as well as accounting features to determine which calls are allowed and to collect information for billing.

Limitations of VoIP

There are some limitations of VoIP service compared to conventional PSTN.

Fax and Data Calls

It can be difficult or impossible to transmit fax and data calls over VoIP links. One reason is that when vocoders are used the waveform is distorted. Although it may still be intelligible to people, a demodulator will not be able to recover the transmitted data.

Another reason is that fax and data calls are not as tolerant to delay and signal loss due to lost or delayed packets. A human can ignore the loss of short portions of the waveform but a modem is likely to drop the call. There are work-arounds but this is becoming less of an issue with the decrease in fax and dial-up data calls.

Power Outages

Another issue is that the VoIP equipment will stop working during power outages unless battery power backup is supplied for all of the equipment involved (modems, routers, ATAs,...). This is typically not done unless the ATA is part of the ISP's residential gateway. In contrast, POTS service provides enough battery power from the CO to operate a conventional telephone.

E911

The third problem is that when a subscribers makes a call to an emergency dispatcher (e.g. 911) the telephone company provides location information for

[2] A CO's trunks may actually be implemented using a packet-switched network.

the caller. This is not possible with VoIP calls because there is no mapping available from IP address to street address as there is with phone numbers. VoIP providers for residential customers must therefore provider their own “E911” databases.