Internet Real-time Communication

Text: Christian Schlatter, SWITCH, schlatter@switch.ch

The Internet has revolutionized the way people communicate today. We surf the web to gather information, and e-mail traffic has increased massively since the Internet first came into existence. Real-time applications like videoconferencing and instant messaging further enhance the channels of virtual collaboration that are open to us. In this article, we look at the basic requirements for real-time communication in the Internet, together with the way in which they are implemented and the problems encountered.

What is Real-time Communication?

If we look at the different ways in which people communicate using the Internet, we can distinguish between asynchronous and real-time communication (RTC). E-mail is the most common example of asynchronous communication, while other examples include discussion boards, blogs, video on demand and shared calendars. Examples of RTC include audio / video conferencing, whiteboarding and application / desktop sharing. RTC is conducted in synchronous or "same time – different place" mode, while asynchronous communication can be summed up as "different time – different place".

Both asynchronous communication and RTC have their pros and cons. Asynchronous communication allows people to connect as and when it is convenient for them. Asynchronous tools are useful for sustaining dialogue and collaboration over a prolonged period and providing people with resources and information that are instantly accessible, day or night. But sometimes it is much easier to make a phone call to discuss a problem rather than to use an asynchronous tool. In this case, RTC has the advantage of permitting instant feedback from the participating parties, providing a more natural and efficient way of working together.

Basic Requirements for Internet RTC

Asynchronous communication tools normally rely on client-server architecture. One participant sends the information to the server, and all other participants take the data in question from the server. Most RTC applications, by contrast, are based on peer-to-peer architecture. Content does not have to be stored by an intermediate device, and the media streams ought to flow end-to-end for better scalability and lower latency. RTC thus places different demands on the network infrastructure than asynchronous communication.

Three basic requirements can thus be identified for enabling Internet RTC. First, we have to know the actual IP address of the callee in order to contact him or her. An address resolution service is thus needed which provides a means of obtaining the IP address that goes with an endpoint (EP) address. Asynchronous communication tools normally rely on the well-established Domain Name System (DNS), since only the address of a server has to be known. But EPs are usually mobile and often use dynamically-assigned IP addresses, thus demanding a service that is capable of addressing mobile EPs.

Secondly, there has to be a means of performing the session establishment. This includes the call indication and the negotiation of media parameters such as the media transport addresses and the type of audio codec.

Finally, a protocol is used to transport the real-time data exchanged between the participants of a session. Real-time data often consists of speech or video, but it can also include the pixels of a shared desktop.

These three basic requirements are implemented in different ways by the RTC protocols in use today. In this article, we take a closer look at the most prominent RTC protocols and how they implement these requirements. Interestingly enough, we have pinpointed quite a few analogies in the way in which the requirements are satisfied.

Address Resolution Service

As already mentioned, we do not normally know the actual IP address of our communication partners in advance. Both H.323 and SIP thus rely on name servers that maintain a database with EP-address-to-IP-address mappings. These servers are called Gatekeepers within the H.323 world, while SIP calls them SIP Registrar / Proxy servers. An EP registers with a name server using its EP address and current IP address. Upon initiating a call, it first sends an address resolution request to a name server in order to obtain the IP address of the callee. For scalability and administrative reasons, many organisational units maintain their

H.323 and SIP

If we look at voice and video RTC, there are chiefly two standardized protocol suites in use, namely H.323 published by the International Telecommunication Union (ITU) and the Session Initiation Protocol (SIP) with its accompanying protocols from the Internet Engineering Task Force (IETF). At present, H.323 is the most widespread standard for Voice over IP (VoIP) and videoconferencing. H.323 is not a protocol but a standard which takes in multiple protocols and network components enabling multimedia conferencing over the Internet. It is thus a vertically integrated protocol suite, whereas SIP has a modular or layered design. In order to offer the same functionality as H.323, SIP uses additional protocols such as the Session Description Protocol (SDP) for the determination of the media codecs. SIP can be regarded as a generic RTC protocol that can also be used for instant messaging or for other forms of real-time data exchange.

Because H.323 was based closely on the H.320 standard used for videoconferencing over ISDN, earlier versions of H.323 were not tailored to the Internet's working in packet switched mode. The current version 5.0 can be described as "Internet ready", but each new version of H.323 must be fully backward compatible and hence H.323 has become rather complicated. SIP, which was designed for the Internet right from the start, would seem set to replace H.323 in the future.

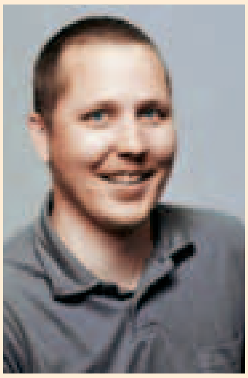
own name server. To allow inter-organisational calling, the challenge is to find the name server that is responsible for a specific EP address, so as to acquire the respective IP address.
Since H.323 was originally designed for LAN environments, it does not include a description of how to find the Gatekeeper belonging to a specific EP address. The Global Dialling Scheme (GDS) was thus developed. The GDS introduces a numbering plan for H.323 EPs that resembles the highly-familiar E.164 numbering plan used in the Public Switched Telephone Network (PSTN). A GDS number can thus be resolved using hierarchically interconnected Gatekeepers, each of which is responsible for a specific number prefix. More information on the GDS can be found at http://www.switch.ch/vconf/gds.html.
SIP, by contrast, uses addresses that look like e-mail addresses, known as SIP Uniform Resource Indicators (SIP URI). An example would be sip:schlatter@switch.ch. To find the corresponding name server, it is necessary to submit a DNS SRV request (a DNS resource record specifying the location of services) and, in return, receive the address of the SIP proxy server responsible for the domain switch.ch. This works in the same way as the familiar e-mail address resolution process using DNS MX records.

Session Establishment

Henning Schulzrinne, a key person in the development of SIP, writes in one of his recent publications that "three pieces of information are needed for establishing a call between two EPs, namely the signalling destination address, local and remote media capabilities, and local and remote media transport addresses at which the EPs can receive the media packets". The signalling destination address is the outcome of the address resolution process described in the previous section. A protocol is thus needed to handle the negotiation of the media parameters used for a specific session. The difficulty here lies in finding a common denominator for the media capabilities of the EPs participating in a session.
In addition, RTC requires all session participants to be virtually present at the same time. This calls for another protocol to signal the start of a session to its participants. An everyday example of this is the phone ringing to tell a user that someone is trying to establish a session.
Whereas H.323 uses two distinct TCP connections and message exchanges for call indication and the negotiation of call media parameters (multistage signalling), SIP combines these two processes into a single message exchange round. SIP session establishment thus needs fewer network resources and is easier to parse by intermediate application level gateways.

Transport of Real-time Data

In the Internet, real-time audio and video data is normally transported using the Real Time Protocol (RTP) published by the IETF. Although RTP is independent of the underlying transport and network layers, RTC applications commonly use RTP over UDP. This is because the backoff and retransmission mechanisms in TCP can lead to an undesirable end-to-end delay. RTP provides payload type identification, sequence numbering, and timestamping, whereas the accompanying Real Time Control Protocol (RTCP) is used for delivery monitoring. Both H.323 and SIP use RTP/RTCP for the real-time data transport.
Although higher bandwidths permit the transmission of real-time data such as high-resolution video, the Internet is not the ideal medium for transporting synchronous data.
Before it can be used for RTC, a communication channel must obey several so-called Quality of Service (QoS) parameters. The most important QoS parameters for RTC are maximum one-way delay, delay variation (jitter), minimum bandwidth, and packet loss. By way of an example, voice RTC requires a maximum one-way delay of 150 ms (ITU-T standard G.114), otherwise a natural conversation cannot take place.
In circuit-switched networks like the PSTN, an end-to-end connection with guaranteed QoS parameters is set up before communication takes place between two entities. The Internet, on the other hand, only offers best-effort services, meaning that no guarantee can be given on the QoS.
At present, it would seem that end-to-end resource reservation in the Internet is not feasible, given that the Internet is a huge collection of administrative domains interconnected almost chaotically by peering contracts. We thus need better troubleshooting resources for end-to-end performance. And RTC applications must improve their inherent mechanisms for adapting to changing network conditions.
Firewalls (FW) and Network Address Translation (NAT) boxes pose another challenge for the transport of real-time data. Because these devices destroy the transparency of the Internet, an EP is not normally reachable via an EP outside the intranet. It is necessary to take special measures to permit UDP based peer-to-peer communication.
It is questionable whether Internet RTC will ever attain the 99.999 % ("5 times 9") reliability of the PSTN. It would seem that we either have to convert the anarchistic packet switching to something resembling the ordered and controlled routing of the PSTN, or have to live with a lower level of reliability. On the other hand, Internet RTC provides us with new possibilities, especially if we look at applications that integrate different communication means such as presence, instant messaging, audio/video conferencing, and data collaboration. Internet RTC will also break the vertically controlled Telco business models to some extent. Another advantage of Internet RTC is that RTC tools can be easily integrated with other Internet services, and this could perhaps lead to truly unified messaging solutions or to services that we have not yet thought of. What is quite definite is that the use of Internet RTC is set to increase in future.
If you have questions or comments just drop me a line at schlatter@switch.ch



Christian Schlatter studied Computer Science at the ETHZ. He is leading SWITCHvconf, a project aiming to improve the means enabling virtual collaboration in the Swiss academic community. He is involved in all aspects of Internet real-time communication, from the evaluation of innovative applications to network protocol analysis.

Glossary

Endpoint (EP)
An endpoint can call and be called. It generates and/or terminates information streams.

Call
Point-to-point multimedia communication between two EPs.

Callee
An EP that gets called.

Caller
An EP that initiates a call.

Session
A set of multimedia senders and receivers and the data streams flowing from senders to receivers. A call can be described as a session with two participating EPs.

Signalling
The information exchange concerning the establishment and control of a session, in contrast to user information transfer.