Integrated-Services Architecture for Internet Multimedia Applications

Zhonghua Yang
Nanyang Technological University, Singapore

Yanyan Yang
University of California, USA

Yaolin Gu
Southern Yangtze University, China

Robert Gay
Nanyang Technological University, Singapore

A HISTORICAL PERSPECTIVE

The Internet has gone from near invisibility to near ubiquity and penetrated into every aspect of society in the past few years (Department of Commerce, 1998). The application scenarios have also changed dramatically and now demand a more sophisticated service model from the network. In the early 1990s, there was a large-scale experiment in sending digitized voice and video across the Internet through a packet-switched infrastructure (Braden, Clark, & Shenker, 1994). These highly visible experiments have depended upon three enabling technologies: (a) Many modern workstations now come equipped with built-in multimedia hardware, (b) IP multicasting, which was not yet generally available in commercial routers, is available, and (c) highly sophisticated digital audio and video applications have been developed. It became clear from these experiments that an important technical element of the Internet is still missing: Real-time applications often do not work well across the Internet. The Internet, as originally conceived, offers only a very simple quality-of-service (QoS), point-to-point, best-effort data delivery. However, for a real-time application, there are two aspects of the problem with using this service model. If the sender and/or receiver are humans, they simply cannot tolerate arbitrary delays; on the other hand, if the rate at which video and audio arrive is too low, the signal becomes incomprehensible. To support real-time Internet applications, the service model must address those services that relate most directly to the time of delivery of data. Real-time applications like video and audio conferencing typically require stricter guarantees on throughput and delay. The essence of real-time service is the requirement for some service guarantees in terms of timing. In response to these demands of real-time multimedia applications, the Internet Engineering Task Force (IETF) has significantly augmented the Internet protocol stack based on the Internet integrated-services model, which is the focus of this article.

THE INTERNET INTEGRATED-SERVICES MODEL

An Internet *service model* consists of a set of service commitments; that is, in response to a service request, the network commits to deliver some service. The Internet is conventionally designed to offer a very simple service model, best effort, providing no guarantee on the correct and timely delivery of data packets. Each request to send is honored by the network *as best as it can*. This is the worst possible service: Packets are forwarded by routers solely on the basis that there is a known route, irrespective of traffic conditions along that route. This simplicity has probably been one of the main reasons for the success of IP technology. The best-effort service model, combined with an efficient transport-layer protocol (TCP [transmission-control protocol]), is perfectly suited for a large class of applications, which tolerate variable delivery rates and delays. This class of applications is called elastic applications.

However, demanding real-time applications require more sophisticated service models beyond the best effort. There has been a great deal of effort since 1990 by IETF to add a broad range of services to the Internet service model, resulting in the Internet integrated service model (Braden et al., 1994; Crowcroft, Handley, & Wakeman, 1999). The Internet integrated services model defines five classes of services that should satisfy the requirements of the vast majority of future applications.

1. *Best effort*: As described above, this is the traditional service model of the Internet.
2. *Fair*: This is an enhancement of the traditional model where there are no extra requests from the users, but the routers attempt to partition network resources in some fair manner. This is typically implemented by adopting a random-drop policy when encountering overload, possibly combined with some simple round-robin serving of different sources.
3. *Controlled load*: This is an attempt to provide a degree of service guarantee so that a network appears to the user as if there is little other traffic, and it makes no other guarantees. The admission control is usually imposed so that the performance perceived is as if the network were overengineered for those that are admitted.
4. *Predictive service*: This service gives a delay bound that is as low as possible, and at the same time, is stable enough that the receiver can estimate it.
5. *Guaranteed service*: This is where the delay perceived by a particular source or to a group is bounded within some absolute limit. This service model implies that resource reservation and admission control are key building blocks of the service.

The level of QoS provided by these enhanced QoS classes is programmable on a per-flow basis, and end-to-end QoS commitment for the data flow is built using a unified multimedia protocol stack and through resource reservation.

THE INTERNET MULTIMEDIA PROTOCOL ARCHITECTURE

The integrated-services Internet offers a class of service models beyond the TCP/IP's (Internet protocol) best-effort service, and thus it imposes strict new requirements for a new generation of Internet protocols. The set of Internet real-time protocols, which constitute the Internet multimedia protocol architecture, represents a new style of protocols. The new style of protocols follows the proposed principles of *application-level framing* (ALF) and *integrated layer processing* (Clark & Tennenhouse, 1990). In this approach to protocol architecture, the different functions are next to each other, not on top of one another. The Internet multimedia protocol architecture is shown in Figure 1.

As shown in Figure 1, the overall multimedia data and control architecture currently incorporates a set of real-time protocols, which include the real-time transport protocol (RTP) for transporting real-time data and providing QoS feedback, the real-time streaming protocol (RTSP) for controlling delivery of streaming media, the session-announcement protocol (SAP) for advertising multimedia sessions via multicast, and the session-description protocol (SDP) for describing multimedia sessions. In addition, it includes the session-initiation protocol (SIP), which is used to invite the interested parties to join the session. But the functionality and operation of SIP does not depend on any of these protocols. Furthermore, the resource-reservation protocol (RSVP) is designed for reserving network resources. These protocols, together with reliable multicast (Handley, Floyd, Whetten, Kermode, Vicisano, & Luby, 2000), are the underlying support for Internet multimedia applications. While all the protocols above work on top of the IP protocol, the Internet stream protocol, version 2 (ST-II), is an IP-layer protocol that provides end-to-end guaranteed service across the Internet.

The Real Time Transport Protocols: RTP and RTCP

The real-time transport protocol, named as a transport protocol to emphasize that RTP is an end-to-end protocol, is designed to provide end-to-end delivery services for data with real-time characteristics, such as interactive audio and video (Schulzrinne, Casner, Frederick, & Jacobson, 2003). Those services include payload-type identification, sequence numbering, time-stamping, and

Figure 1. Internet protocol architecture for real time applications

<table>
<tr><td colspan="3">Multimedia Applications</td><td colspan="7">Multimedia Session Setup & Control</td></tr>
<tr><td colspan="2" rowspan="2">RTP/RTCP</td><td rowspan="2">Reliable Multicast</td><td colspan="2" rowspan="2">RSVP</td><td rowspan="2">RTSP</td><td colspan="4">SDP</td></tr>
<tr><td>SAP</td><td>SIP</td><td>HTTP</td><td>SMTP</td></tr>
<tr><td rowspan="2">ST-II</td><td colspan="3">UDP</td><td colspan="6">TCP</td></tr>
<tr><td colspan="9">IP + IP Multicast</td></tr>
<tr><td colspan="10">Integrated Service Forwarding (Best Effort, Guaranteed)</td></tr>
</table>

nb: For the acronyms, refer to "Key Terms"

Related Content