

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Supporting Technologies for Synchronous E-learning

Juan C. Granda, Christian Uría,
Francisco J. Suárez and Daniel F. García
*University of Oviedo
Spain*

1. Introduction

Recent advances in multimedia technology have enabled new communications possibilities that improve the distance learning experience. Thus, the introduction of multimedia contents in learning materials results in an enhanced learning process. (Steinmetz & Nahrstedt, 1995); (Mishra & Sharma, 2004). In fact, virtual learning environments have an enormous potential (Padmore et al., 2006).

Similarly, enterprise corporations must introduce technical innovations to their production cycle in order to increase their market competitiveness. As a consequence, personal staff must adapt to the changes imposed in the organization to cope with these technical improvements to positively contribute to the corporate success. Thus, it turns essential to design training programs as a part of the corporate policies to keep employees technologically updated. Multiple studies show that interactive training increases knowledge retention and shortens learning curve (Shulman, 1992), and offers great flexibility since training may be provided just-in-time specifically for each user's needs. Interactive learning is even more appropriate to those situations in which an employee manipulates fragile or dangerous equipment. Simulation environments are very suitable to avoid possible misuses and faults that would imply high costs. As an example, in (Mollet & Arnaldi, 2006) the authors propose a virtual reality environment for the training of human staff in a military equipment factory.

Geographical dispersion in multinational corporations constitutes another reason to use e-learning in staff training. Thus, there is no need to maintain a specific training center, so employees do not have to move from their workplace to this center to receive training courses. This implies a cost-saving alternative compared to traditional learning, because displacements are not necessary, so training may be provided just-in-time at the workplace. Nowadays, e-learning has become a widely used learning strategy in higher education and staff training. In fact, e-learning is one of the most suitable tools for developing training plans in large business corporations in order to establish a continuing education offer to their human resources.

Often, e-learning is divided into asynchronous e-learning and synchronous e-learning attending to the spatio-temporal limitations imposed to the learning process. Thus,

asynchronous e-learning allows participants to interact wherever they may be and whenever they may desire. In contrast, attendees to a synchronous e-learning activity must join the virtual session at the same time, so interactions must occur in real time. To accomplish this goal, tools like shared whiteboards, instant messaging, and audio/video conferences are used.

Synchronous e-learning allows multiple remote participants to interact live despite of their geographical location. The instructor and the rest of participants can solve doubts from other participants immediately, providing a sense of co-presence. Therefore, synchronous e-learning promotes learning communities by the interchange of ideas and experiences among learners. In short, synchronous e-learning combines advantages from e-learning such as geographically independence, and benefits from traditional education such as face-to-face interaction.

Audio and video conferencing provide support to synchronous e-learning activities. IETF and ITU-T are organizations that elaborate and promote standard conferencing frameworks and protocols. In this chapter, the main issues of the transport of continuous data over packet networks are explained, especially those related to multicasting and how Real-time Transport Protocol (RTP) copes with the handicaps of this kind of networks (packet loss, latency, jitter, etc.). The next section introduces the characteristics of synchronous e-learning and its advantages over other educational strategies. In section 3, the technological implications of synchronous e-learning are described. In sections 4 to 7, the protocols and recommendations from IETF and ITU-T related to audio and video conferencing are detailed, compared and organized into layers according to their purpose. These protocols are mainly used to establish multimedia sessions between users within a synchronous e-learning session. In section 8, multiple audio and video codecs specifically oriented to synchronous e-learning environments are mentioned. Finally, conclusions are presented in section 9.

2. Synchronous e-Learning

Usually, e-learning systems are classified according to the temporal restrictions imposed on the learning process. Two types of e-learning systems are considered: asynchronous and synchronous systems. The asynchronous e-learning systems do not impose any temporal restriction on the learning process. Thus, there is no need that instructor and learners are connected at the same time, so they can deal with learning materials at their own pace.

In contrast, synchronous e-learning systems imply a temporal synchronization between instructor and learners, but they may be geographically dispersed. The learning experience is guided by the real-time interactions that occur among learners and instructor and among learners themselves. Table 1 compares some relevant characteristics of asynchronous and synchronous e-learning.

Synchronous e-learning presents some features which differentiate it from other learning methodologies:

- Live. Synchronous e-learning activities take place live, that is, they are not previously recorded, but recorded or pre-produced material may be used during the activity.

Asynchronous e-learning	Synchronous e-learning
Intermittent on-demand access	Real-time
Previously recorded or pre-produced	Live
Just in time	Scheduled
Individual or poorly collaborative	Collaborative
Independent learning	Co-presence of learners and instructor
Self-paced	Concurrent learning

Table 1. Differences between asynchronous and synchronous e-learning

- Real-time. Although synchronous e-learning activities may be recorded for a later playback, they are essentially in real time, so they cannot be paused or resumed like self-paced courses.
- Facilitated. Normally, every synchronous session is guided by a facilitator whose mission is to control that interactions between learners are focused on knowledge acquisition. They help to consolidate knowledge.
- Learning-oriented interactions. This differentiates synchronous e-learning from other types of real-time activities like video conferences, or online product demonstrations.

Asynchronous e-learning	Synchronous e-learning
E-mail	Instant messaging
Discussion forums	Online chat
Web-based training	Live webcasting
Podcasting	Audio conferencing
DVD	Video conferencing
Computer aided systems	Web conferencing

Table 2. Examples of asynchronous and synchronous e-learning features

In table 2 the most common features used to provide both asynchronous and synchronous e-learning are listed. The utilization of synchronous e-learning systems has multiple advantages:

- They connect multiple dispersed learners, which results especially appropriate to large multinationals with sites in multiple countries. Corporations may benefit from the ability to train employees at the workplace.
- They allow interactions and collaboration in real time, emulating relations within a traditional classroom. This way, every attendee to the session relates each other more naturally, so the session flows spontaneously. Questions are raised immediately and responses are given directly.
- They provide a sense of immediacy, which is useful to deliver last-minute contents or time-sensitive data. Furthermore, the presence of the instructor is noticeably for

- the learners, so anxieties arisen in the learners regarding the non-personal learning experience can be alleviated.
- Learners can share doubts and experiences. This stimulates the sense of connectedness among them. As a result, these tools permit social interactions between learners, which foster learning communities and create a learning synergy. Long-term effects include better teamwork and collaboration skills.
 - They promote an egalitarian learning experience among learners. Those extroverts do not dominate class dynamics as in traditional classrooms. Anonymous participation in the e-learning sessions make introvert learners feel more comfortable to participate actively. Due to the avoidance of face-to-face interaction, racial or social differences do not affect the flowing of sessions.

Usually, synchronous e-learning systems provide common features such as audio and video conferencing, instant messaging and presence control, shared whiteboard with electronic ink and telepointers, desktop and application sharing, etc. In (García et al., 2007) a functional evaluation of 20 synchronous e-learning tools is presented. In figure 1, a screenshot of a synchronous e-learning tool (Granda et al., 2008) is shown. This tool provides the most common features of synchronous e-learning systems.

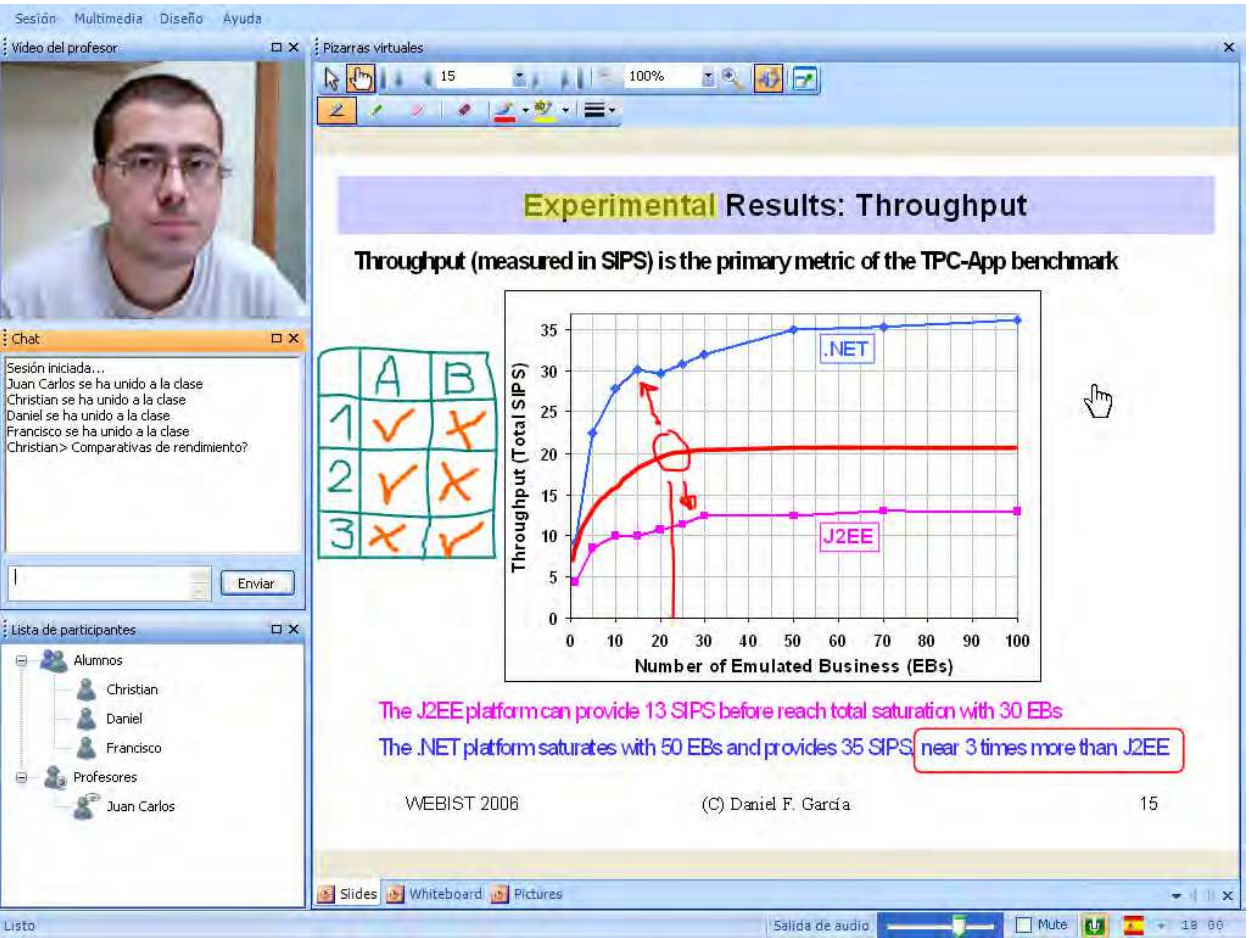


Fig. 1. Screenshot of a synchronous e-learning tool

3. Technological Requirements

In order to provide the functionalities listed in the previous section, synchronous e-learning systems must interchange different types of multimedia data:

- Audio. It conveys the speech from the instructor, which is the most important information in an e-learning session. Thus, it must be prioritized over the rest of data.
- Video. It usually conveys the talking head from the instructor to the learners, so it is less important than audio. Video is used to reinforce the sense of presence of the instructor in the learners to avoid anxieties (Weller, 2007).
- Instant messaging. It allows participants in a synchronous e-learning session to communicate textually. This is very useful for learners to post questions to the rest of the class without abruptly interrupting the class with an oral intervention.
- Presence control. It reinforces the sense of presence of all participants, so the instructor may track the users connected to the session.
- Slide presentation or shared whiteboard. It constitutes the learning materials that support the e-learning session. Formats like PowerPoint and PDF are very common. Most of the tools allow the annotation of the contents by the instructor to provide clarifications.
- Telepointer. It is mainly aimed to point to elements within a slide.

The main goal of a synchronous e-learning tool is to bring instructor's explanations to students and to allow a certain level of feedback from learners to the instructor, so he may provide extra explanations or adapt the pace of the session. Because of this, there must be a two-way traffic among instructor and learners. Furthermore, learners collaborate among them, so traffic turns multipoint, so data generated by a participant must be delivered to the rest of them, increasing significantly the network bandwidth requirements.

However, there is no need to transmit all data within a synchronous e-learning session in a multipoint way. For example, video data is only useful to reinforce the sense of presence of participants, so a multipoint transport may be discarded for video data, and a solution in which only the video from the instructor is delivered to all learners may be adopted. In this situation, video data is transmitted in a one-way fashion from the instructor to all learners. Other types of traffic such as audio data from each participant, instant messages, annotations on the learning contents or telepointers, must be delivered in a multipoint way. Therefore, the speech of a participant must be delivered to the rest of them; the annotations from a participant must be rendered in all session participants' screens, and so on. Nevertheless, this implies higher bandwidth consumption in the underlying network, which is usually the most valuable resource in synchronous e-learning systems. If network multicasting techniques are used, the final bandwidth consumption is reduced significantly. Audio and video data are the most bandwidth-consuming data in a synchronous e-learning session. In fact, the bandwidth of the rest of data in the session is negligible in contrast to the bandwidth needed to deliver audio and video. Hence, in the rest of the chapter, focus is put on protocols and techniques to deliver real-time audio and video through the Internet.

4. Technological Background

The technologies involved in synchronous e-learning can be grouped in different categories depending on their scope of application. Undoubtedly, the transmission of multimedia streams and the management of user sessions are the most relevant fields and hence the widest range of technological alternatives is focused on them. The great variety of available audio and video coding techniques constitute another significant group, but less important than the aforementioned. In fact, most of the technologies that will be considered here are related to the transmission of multimedia information through the Internet and they belong to the extremely populated group of standards and recommendations on which Internet communications are based. There are two main players in the field of the standardization of Internet communications: ITU-T and IETF.

The ITU-T is a standards development organization (SDO) that is one of the three sectors of the International Telecommunications Union (a specialized agency of the United Nations). ITU-T is responsible for generating worldwide "recommendations" (non-binding standards) for telecommunications. The ITU-T is divided into fourteen Study Groups. Study Group XVI is responsible for generating recommendations for data collaboration and video conferencing.

The Internet Engineering Task Force (IETF) develops and promotes Internet standards, cooperating closely with the W3C and ISO/IEC standard bodies, and dealing in particular with standards of the TCP/IP and Internet protocol suite. It is an open standards organization and its work is usually funded by companies or sponsors.

Both organizations contribute to the technological development of Internet developing recommendations that end up being de facto standards, when not official ones. They usually work together to spread and consolidate their developments but, occasionally, industry pressure can hold up their advances or even confront standards from both organizations.

IETF and ITU-T standards differ in their initial conceptions. While the IETF standards follow a client/server philosophy like most of the Internet protocols, ITU-T recommendations are based on ISDN networks. Both organizations propose their respective protocol stacks for multimedia transmissions, as shown in figure 2. ITU-T has adopted some IETF protocols in its stack: IP, TCP, UDP and RTP. So both organizations agree in the protocols to use at the lower layers. On the other hand, their proposals for managing user sessions are quite different.

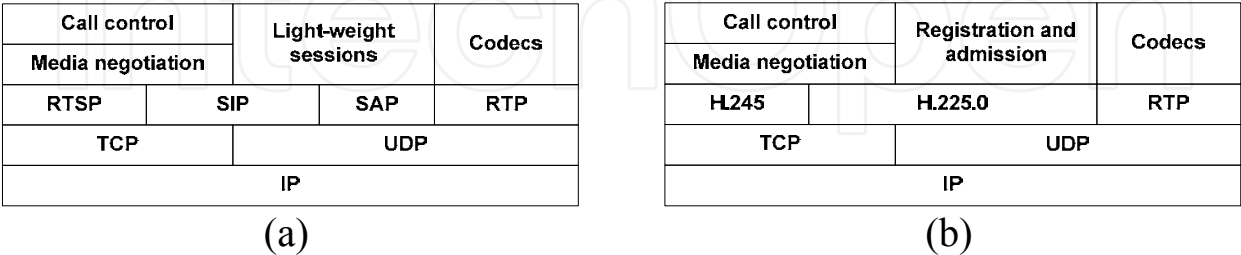


Fig. 2. Multimedia protocol stacks from IETF (a) and ITU-T (b)

5. Network layer

Ideally, multimedia applications should know underlying network characteristics in order to interchange data using the network efficiently. It would be very useful to identify the

type of network and its instant congestion to adapt the behavior of the application to the changing state of the network. Applications that cope with the changing conditions of the underlying network are known as net-aware applications.

However, IP networks provide an abstraction layer which hides to applications all information and topology of the underlying network. Thus, applications must measure different metrics in the network to understand its current state. In (Paxson, 1999) traces from multiple connections among several Internet sites are analyzed, so dynamic behavior of the network is estimated according to packet loss, disordered packets, and other metrics. In (Jain & Dovrolis, 2003) a methodology to estimate the network bandwidth available between two hosts connected to the Internet is introduced.

Usually, IP networks are characterized by packet loss, packet replication, packet corruption, network transit time and maximum transfer unit. Nevertheless, there must be taken into account that measurements may vary from one site to another, that is, they depend on geographical location, time of day, network congestion and so on. As a result, multimedia applications must never consider an ideal behavior of the underlying network, and must be prepared to cope with packet loss, different transit times, etc.

5.1 Packet loss

A packet is considered lost when it does not arrive at the receiver. There are few metrics that can be used to measure packet loss in the network. For example, the average packet loss rate may be used to estimate current network congestion, and packet loss correlation may be used to analyze dynamic behavior of the network. Multiple works have demonstrated that the observed average packet loss rate is not constant across the Internet, and it changes smoothly, but the change might be abrupt in rare occasions (Paxson, 1999); (Bolot & Vega-García, 1999); (Yajnik et al., 1999).

It would be ideal for applications that packet loss occurs as isolated events, so they would be uniformly distributed in time. This way, an application can recover from packet loss easily, in contrast to a scenario in which packet loss occurs in bursts. Unfortunately, measurements show that packet loss occurs most of the time as isolated events, but in few occasions it occurs in bursts, so several consecutive packets are lost. Consequently, applications must be designed to tolerate a small fraction of packet loss without affecting their operation significantly.

5.2 Packet replication

Packet replication happens when the same packet arrives multiple times at the receiver. This implies malfunctioned or misconfigured network equipment. However, this must not be a problem for applications; the simplest solution is to discard replicated packets silently. Anyhow, it might become a real problem if the packet replication rate is too high, because it would imply a waste of network resources.

5.3 Packet corruption

Information flowing across networks could be corrupted by the network or by physical phenomena affecting the transmissions. Wireless networks are more prone to packet corruption than wired ones. The receiver identifies corrupted packets by a failure in the

checksum validation at the transport level in the protocol stack. Usually, a corrupted packet is discarded, so applications consider it as a lost packet.

However, depending on the data being carried in a corrupted packet, an application may be interested in receiving the packet regardless it is corrupted. The application might recover the data within the packet partially or totally. This is a very common scenario in audio and video transmission.

5.4 Packet transit time

Transit time refers to the time in which a packet traverses the network from the sender application to the receiver. It is mainly influenced by the length of the path that the packet has to traverse and the length of routers queues. The former refers to the number of routers in the path instead of the physical length of the path. The latter affects the time that a router spends in forwarding the packet to the next router in the path. The longer the queues of the routers are, the more congested the network is. As a result, the transit time is not constant but changes as network status changes.

It is easier for applications to work with a constant transit time. Unfortunately, because of the dynamics of the network, it changes from packet to packet. The variation in the transit time is known as interarrival jitter, and represents one of the challenges that multimedia applications must deal with. To ensure a proper operation, applications usually estimate jitter and have a buffer arranged to bear its effects.

In (Chong & Matthews, 2004) the authors considers a maximum transit time of 150 milliseconds for a successful audio conference among two partners. The same author estimates a maximum interarrival jitter of 75 milliseconds for maintaining an intelligible conversation.

5.5 Maximum Transfer Unit

Packets sent to the network cannot have an arbitrary size. The network technology imposes a maximum packet size. This is commonly known as the maximum transfer unit (MTU). Typical MTUs are 576 bytes for PPP links and 1500 bytes for Ethernet networks.

If a packet is bigger than the MTU of the network, it is fragmented into multiple pieces and they are sent independently to the destiny. Multimedia applications must not delegate to this mechanism delivery of packets bigger than MTU. If an application does so, the loss of any one fragment will make impossible for the receiver to reconstruct the packet. This results in a loss multiplier effect.

5.6 Multicast and unicast delivery

IP Multicast is a bandwidth-conserving technology that reduces traffic by simultaneously delivering the same data to multiple receivers. In addition to synchronous e-learning tools, other applications such as streaming servers or collaborative environments benefit from multicast delivery techniques. In figure 3, both the unicast and multicast delivery techniques are schematized. As shown in the figure, in the unicast delivery it is necessary to replicate data as many times as the number of receptors, while a single copy of data is sent to the network in multicast delivery. In multicast delivery the network is responsible for sending a copy of data to every receptor with as less replication as possible.

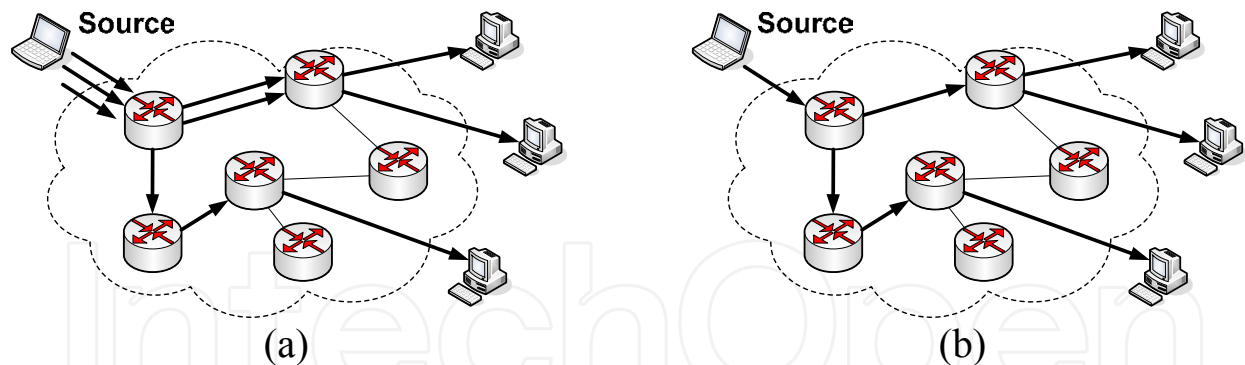


Fig. 3. Unicast (a) and multicast (b) data delivery

IP Multicast is based on the concept of multicast group membership. A multicast group represents multiple hosts that receive the same data stream. This group is not physically or geographically scoped, so receptors may be located wherever in the Internet. A host must register in the multicast group to receive multicast traffic sent to the group. In contrast, any host may send data to the multicast group without registering in the group.

6. Transport layer

Using IP for delivering continuous media has a series of drawbacks as previously explained. In practice, IP is not used directly. Instead, a transport protocol is used over IP to cope with these issues. TCP and UDP are the most commonly used transport protocols in the Internet. TCP is a reliable connection-oriented protocol, while UDP is a connectionless protocol which has the same downsides that IP. Nevertheless, UDP is more suitable to deliver real-time data because it is time-deterministic as opposite to TCP. TCP uses acknowledgments and retransmissions, so the network transit time of a packet may vary significantly.

Today, almost all multimedia tools use UDP for data transport. All advantages of TCP such as flow control, packet loss detection, ordered reception and retransmissions; turn into disadvantages when it is necessary to transmit multimedia streams. In multimedia tools, it is preferable low delays rather than assuring that all data arrive correctly at the receiver. Moreover, IP multicast may be used at the network level with UDP at the transport level, so optimal distribution topologies may be employed.

However, the main problem of UDP is that packets may be lost, reordered or corrupted, exactly as observed for the raw IP service. As a result, mechanisms similar to those in TCP must be implemented in applications to avoid these issues.

6.1 Real-time Transport Protocol

To solve the problems imposed by UDP to the multimedia applications, the IETF has promoted the Real-time Transport Protocol (RTP) (RFC 3550). Although RTP is independent of the transport protocol, it is usually used in conjunction with UDP. It provides services such as packet loss detection and timestamps for timing reconstruction. Multicast delivery may be used if available in the underlying network, so all RTP features are scalable.

The main goal of RTP is to provide end-to-end delivery of time-sensitive data such as audio and video. However, RTP does not guarantee quality of service for real-time services. Therefore, RTP packets may be lost, corrupted or reordered.

RTP was deliberately designed incomplete, so it could be adapted to every application. It offers a set of common functionality instead of specific algorithms; hence it constitutes a framework for developing multimedia applications. The application must implement the specific algorithms to make RTP operate. Thus, in general, RTP is part of the application instead of a separate layer. For this reason, a particular application, in addition to the RTP specification, requires at least the following documents:

- Profile specification. It particularizes different elements that RTP leaves domain-specific and defines extensions to the original RTP specification.
- Payload format specification. It defines the format of the data to be delivered using RTP. For example, the encoding of audio and video and how data must be packetized.

The most common RTP profile is the audio and video profile (RFC 3551). It is intended to support audio and video conferences with multiple participants. In fact, RTP is the de facto standard for voice over IP (VoIP) solutions. RTP is the most appropriate protocol for delivering almost every data in a synchronous e-learning session.

The functionalities of RTP are actually implemented by two protocols, that is, the RTP specification defines two protocols: the transport protocol and the Real-time Transport Control Protocol (RTCP). The RTP transport protocol is responsible for end-to-end data delivery, which includes detection of packet loss, reordered packets and stream timing reconstruction. RTCP monitors quality of service and conveys information about the participants in an on-going session. Moreover, it provides inter-stream synchronization so lip synchronization may be achieved in a video conference. For the operation of RTP two transport addresses are necessary, one for RTP transport packets and the other one for RTCP.

Typically, congestion control is carried out at the RTP level. RTCP packets may be used to estimate network status, because these packets notify the quality of data reception at the receiver side.

Sometimes, especially in low-bandwidth environments when transmitting audio, the RTP packet header has a significant size compared to the audio payload. This implies a high overload that might be unacceptable. In order to reduce the size of the RTP header, and consequently the packet size, RFC 2508 defines a technique to compress RTP/UDP/IP headers, so the final size of the headers is negligible compared to the payload size.

Finally, the IETF has defined a profile for secure RTP transactions (RFC 3711). The secure RTP (SRTP) profile specifies how RTP packets must be encrypted, to guarantee the confidentiality of the communications. Another RTP profile, RFC 4585, provides more feedback from receptors based on RTCP packets, which may be very useful in synchronous e-learning scenarios.

7. Control layer

There are some important tasks that must be carried out in this layer in order to ensure the correct running of a synchronous e-learning activity. The protocols that deal with these issues are signalling protocols because, roughly said, they interchange signals for controlling communications. A signalling protocol must fulfil several tasks:

- Establishment, management and termination of multimedia sessions.
- Handshaking and selection of session features, especially the types and formats of the multimedia information to transmit.
- Management of the session participants, enabling users to log in and log out of a session.
- User localization and address translation; common addresses from different sources like URLs, telephone numbers or e-mail addresses will need to be adapted in order to establish links among them.

The H.323 set of protocols, published by the ITU-T, and the SIP protocol, proposed by the IETF, are the two most widespread signaling protocols. H.323 defines the protocols to provide audiovisual communication sessions on any packet network, but its philosophy is clearly influenced by traditional circuit switching networks. SIP is a lightweight protocol that was designed with the aim of achieving two advantages against H.323: more flexibility and easier implementation. There are countless works that defend, with more or less intensity, one against the other. Some of them are (Schulzrinne & Rosenberg, 1998); (Dalgic & Fang, 1999); (Glasmann et al., 2003). However, although their architectures differ substantially, the latest revisions of both protocols are similar in terms of functionality and the many differences between them have vanished.

H.323 is extremely more complex than SIP. This is because the SIP specification barely imposes any protocol or service to be present in a SIP system. On the contrary, H.323 defines multiple profiles that require the use of several specific protocols and components. From this view, SIP gives much more freedom than H.323 for choosing surrounding technologies. This results in better expansion capabilities. H.323 systems can only handle media formats that have been previously registered by the ITU-T. Thus, trying to use a format unsupported by any H.323 profile can be painful, since ITU-T usually introduces slowly new media formats in its recommendations. Besides, SIP is a text-based protocol, allowing for easy inspection by administrators. On the other hand, inspection of H.323 messages requires specific and more complex tools. It is clear that the high level of specification of H.323 has some drawbacks, but when particular needs match well a specific profile, the deployment and management of an H.323 system can be friendlier than a SIP one, since all elements are perfectly detailed.

The control layer must also serialize the access to shared resources within a synchronous e-learning activity like the audio channel. This task is carried out by floor control protocols.

7.1 Session Initiation Protocol

In contrast to H.323, the multimedia sessions architecture proposed by the IETF is constituted by multiple independent protocols. This architecture includes two Internet specific protocols: SIP and SAP. SIP (RFC 3261) carries out the operations related to session

establishment, modification and termination. The Session Announcement Protocol (SAP) is described in the IETF's RFC 2974; it serves to announce sessions to users or services and was developed for multicast environments. Additionally, IETF and ITU-T worked together to define the H.248 protocol, intended to control the gateways that interconnect networks based on different technologies, e.g. packet switching networks and circuit switching networks. In the nomenclature of the IETF, H.248 is known as the Megaco protocol.

The operation of the SIP protocol is based on request and response messages HTTP-like. It reuses many of the codification rules, error codes and header fields from HTTP. Hence, the call control functions provided by SIP can be deeply and easily integrated in any web infrastructure. In addition, SIP uses the Session Description Protocol (RFC 4566), to exchange the session settings.

The specification of the SIP architecture distinguishes two elements in a SIP network: user agents (UA) and servers. A SIP user agent is a logical network endpoint used to create or receive SIP messages and thereby manage a SIP session. A user agent can perform the role of a User Agent Client (UAC), which sends SIP requests, and a User Agent Server (UAS), which receives the requests and returns SIP responses. These roles of UAC and UAS only last for the duration of a SIP transaction. A SIP transaction consists of a client request message sent to a server and the subsequent interchange of messages until the client receives a definite response from the server. Since SIP servers are only aware of the status of their transactions (i.e. they only know the individual requests and responses that are associated to a transaction), they do not have to keep the status of every call they manage. Therefore SIP systems are stateless and highly scalable.

Although two SIP endpoints can communicate without any intervening SIP infrastructure (servers), this approach is often impractical for a public service as the most advanced functionalities need their presence. The most common duties of servers are user location and address resolution. A server element can perform three different roles that are not mutually exclusive:

- Redirect server. Receives requests from UACs and sends responses directing the client to contact another set of servers.
- Registrar server. Registers the devices at the time they are connected, so it can provide their contact addresses (URIs) when incoming sessions addressed to the devices are received.
- Proxy server. Like HTTP proxies or SMTP mail transfer agents, a SIP proxy server is an intermediary entity that processes transactions on behalf of other user agents. A proxy server primarily plays the role of routing and it is also useful for enforcing policies. A proxy interprets, and, if necessary, rewrites specific parts of a request message before forwarding it.

7.2 H.323 Recommendation

H.323 is an umbrella recommendation of the set of recommendations that the ITU-T elaborated for video conferencing (H.3xx). Of all of them, H.323 is the only suitable for multimedia data transmission through the Internet when there are not any mechanisms to guarantee quality-of-service in the underlying network.

The H.323 system defines several network elements that work together in order to deliver rich multimedia communication capabilities. Those elements are terminals, multipoint

control units (MCUs), gateways, gatekeepers, and border elements. Collectively, terminals, multipoint control units and gateways are often referred to as endpoints. The terminals are IP network clients and constitute the most basic elements in any H.323 system. Gateways are optional devices that enable communication between H.323 networks and other networks, such as PSTN or ISDN networks. A Gatekeeper is an optional component in a H.323 network that provides a number of services to terminals, gateways, and MCU devices. Those services include endpoint registration, address resolution, admission control, user authentication, and so forth. A multipoint control unit is responsible for managing multipoint conferences and is normally composed of two types of logical entities: one multipoint controller and one or more multipoint processors. A border element is a signalling entity that generally sits at the edge of an administrative domain and communicates with other border element located at other administrative domain. H.323 is defined as a binary protocol, which allows for efficient message processing in network elements. The syntax of the protocol is defined in ASN.1 and uses the Packed Encoding Rules (PER) form of message encoding for efficient message encoding on the wire. There are four different types of communication flows in H.323 systems:

- Call signaling. Convey all the information related to connections control and supplementary services according to the protocols detailed in the recommendations H.225.0 and H.450.x.
- Call control. Enable control of the transmission of multimedia information and capabilities negotiation, based on the recommendation H.245.
- RAS signaling. The RAS protocol is detailed in the recommendation H.225.0 and is intended to communicate endpoints with gatekeepers.
- Logical Channel Signaling. These flows are media streams indeed; they convey multimedia data using an underlying RTP session for each separate medium.

In the latest revisions of H.323, all flows use unreliable transport protocols, like UDP, except the call control flows, that require a reliable transport protocol, like TCP. Table 3 shows some of the most relevant recommendations included in the H.323 specification.

Recommendation	Purpose
H.245	Call and media control
H.225.0	Establishment and control of the connection
H.332	Management of densely populated conferences
H.235	Security (privacy, authentication, etc.)
H.246	Interworking with services based on circuit switching
H.450.x	Supplementary services (call transfer and diversion, etc.)
H.26x	Video codecs
G.7xx	Audio codecs
T.120	Data sharing (collaborative tools)
X.680	Abstract Syntax Notation One (ASN.1)
X.691	Specification of packet encoding rules (PER)

Table 3. Recommendations included in the H.323 specification

7.3 Floor Control

A high number of attendees usually participate in a synchronous e-learning session. These participants interchange audio, video and other types of multimedia data in real time. Interactions must be coordinated, so data from multiple participants does not interfere. For example, a limited number of users should speak at the same time, so audio streams from multiple participants do not overlap.

In a conference with two peers floor control may be considered trivial. Both peers naturally agree who can speak anytime, so conversation is intelligible. However, as participant number grows, it is necessary to define floor control mechanisms to serialize the access to the audio channel. In general, this applies to every shared resource in synchronous e-learning sessions, such as the shared whiteboard, the video channel, telepointers, etc. It is also interesting to differentiate roles among participants in the session. Usually, the instructor is responsible for granting and revoking privileges to the learners. All these are issues that a floor control protocol must solve.

There are many works about floor control and they are not restricted to synchronous e-learning. In (Dommel & García-Luna-Aceves, 1997) and RFC 4376 the requirements for floor control protocols in large collaborative environments are enumerated. In (Malpani & Rowe, 1997), the authors introduce floor control techniques in very large web seminars.

The Binary Floor Control Protocol (BFCP) was published by the IETF in RFC 4582 to manage joint or exclusive access to shared resources in conferencing environments. It defines different entities: a floor control server, a floor chair and multiple floor participants. The server is responsible for granting and revoking floors to participants. The floor chair decides whether the server must grant or revoke a floor to a participant.

Attendees to a conference may discover floor existence using SIP media negotiation, and consequently BFCP is integrated in the multimedia conferencing framework proposed by the IETF.

8. Media processing

Usually, multimedia data in a synchronous e-learning session imposes important bandwidth requirements, so it is unworkable to send it uncompressed to the network. This is especially noticeable with audio and video. Therefore, it is necessary to encode data in order to reduce the needed network bandwidth. The encoding process is done by audio and video codecs. Those used in synchronous e-learning sessions should exhibit the following characteristics:

- Low latency, so data transmission fulfills real-time requirements.
- High compression ratio in order to consume as less network bandwidth as possible.
- Packet loss resilience to recover from packet loss.

8.1 Audio codecs

Audio conferencing is one of the most important features in synchronous e-learning sessions. It is based on the transmission of the voice of a participant to one or multiple receptors. Thus, it is necessary to capture participant's voice in real time. The capture process implies a digitalization of the analog audio signal and depends on several parameters:

- Sampling rate. It represents the times per second that the analog voice is sampled during digitalization. The number of samples needed is determined by the Nyquist-Shannon sampling theorem. It proofs that the digitalization of an analog signal must be done with a sampling rate twice the maximum bandwidth of the signal. For voice is typical a sampling rate of 8000 Hertz.
- Number of channels. An audio stream may contain one o more independent signals. That is, audio can be mono, stereo or multi-channel. In synchronous e-learning and VoIP, only mono audio is used as it conveys voice from the participants.
- Sample size. It represents the resolution of each sample. The bigger this size is, the higher resolution is obtained in the digitalization process. Typical values are 8 and 16 bits per sample.
- Capture period. Audio capture devices usually have a capture buffer where they keep samples as the digitalization process is taking place. After a period of time, the device returns all samples digitalized to the capturing application. It is important to use a capture period as short as possible to minimize digitalization latency. Typical values for audio conferencing are 20 and 30 milliseconds.

Codec	Sampling rate (Hz)	Bitrate (kbps)	Packet duration (ms)	Packet size (bytes)	Nominal bandwidth (kbps)
G.711	8000	64	20	160	87.2
G.723.1	8000	5.3	30	20	20.8
	8000	6.4	30	24	21.9
G.726	8000	24	20	60	47.2
	8000	32	20	80	55.2
G.728	8000	16	30	60	31.5
G.729	8000	8	20	20	31.2
GSM	8000	12.2	20	-	23.2
	8000	13	20	33	36.4
iLBC	8000	13.3	30	50	28.8
	8000	15.2	20	30	38.4
Speex	8000	2.15 – 24.6	20	Variable	25.35 – 47.8

Table 4. Voice codecs

Depending on the type of audio to encode, different techniques are used. Audio is usually classified into three categories attending to its sampling rate: narrowband, mediumband and wideband. Narrowband audio refers to voice audio, while wideband audio represents music. In synchronous e-learning human voice is used, so only narrowband codecs are employed to encode audio signals. Particularly, table 4 lists the most commonly used audio codecs in VoIP and synchronous e-learning sessions. This table shows the codec characteristics and the nominal bandwidth, which is the result of considering de bitrate of the audio stream generated by the codec and the RTP, UDP, IP and Ethernet packet headers.

8.2 Video codecs

Video conferencing is another important feature of synchronous e-learning systems. The video from the instructor is delivered to the rest of participants in an e-learning session. Usually a videoconference is point-to-point, but occasionally may be multipoint, so video from multiple participants may be viewed simultaneously. Nevertheless, video is appropriate to reinforce the sense of presence of users; so at least, video from the instructor is delivered to the learners.

Similarly to audio, video is captured in real time and it must be encoded to reduce its bandwidth consumption. Multiple parameters may be configured during video digitalization:

- Image resolution. It represents the dimensions, width and height, of every video frame. The higher the resolution is, the better quality the image has. However, high image resolutions imply high bitrates. A typical image resolution in video conferencing is 320x240 pixels.
- Image format. It specifies the format of every video frame to be captured.
- Frames per second. It represents the temporal resolution of the video. Typical values for video conferencing are 1, 5, 10 or even 15 frames per second.

Audio and video codecs are usually classified into lossless or lossy codecs. The former are used in audio and video edition because they keep the original quality of data. The latter achieve higher compression ratios and they are more suitable to synchronous e-learning sessions. The following are the most widespread codecs for video conferencing:

- MJPEG (Motion JPEG) is a codec in where frames are separately encoded as JPEG images. It is commonly used in IP cameras, but it is not suitable to synchronous e-learning due to its high bandwidth consumption.
- MPEG-4 is a set of standards for audio and video encoding. Similarly to MPEG-1 and MPEG-2, MPEG-4 defines profiles, so different solutions implements different capabilities. Specifically, MPEG-4 part 2 defines the Advanced Simple Profile (ASP) that is very common in video conferencing.
- H.261 is an ITU-T recommendation for the encoding of video signals in ISDN networks. It supports image resolutions of 352x288 and 176x144 with bitrates between 40 Kbps and 2 Mbps.
- H.263 is an evolution from H.261 oriented to low-bandwidth environments. It supports five image resolutions and significantly improves the compression ratio of H.261.
- H.264, also known as MPEG-4 Part 10 AVC, is the result of a joint effort of the MPEG group and ITU-T. It improves ratio compressions of H.263. Thus, it may be used in very-low-bandwidth environments with an acceptable quality. It may also be used for high resolution video. In fact, it is used for encoding video in Blu-ray discs.
- VC-1 is an alternative to H.264 and it is promoted by Microsoft. Its main goal is to provide high bandwidth compression of interlaced video, so it is no necessary to deinterlace it before. In some scenarios it competes with H.264 producing similar results.

9. Conclusions

Audio and video conferencing are two major features of synchronous e-learning systems. The former is used to allow participants to participate orally in the e-learning sessions, while the latter is used to reinforce the sense of presence of users.

IETF and ITU-T are the organizations that promote standards and recommendations for real-time delivery of continuous media over the Internet. IETF has published several RFCs about the Session Initiation Protocol (SIP) in order to define a multimedia conferencing framework in the Internet. ITU-T has developed the H.323 recommendation as an umbrella, so multiple extra recommendations constitute the reference ITU-T framework for multimedia conferencing in the Internet.

However, in spite of these independent conferencing frameworks, both IETF and ITU-T frameworks have in common the Real-time Transport Protocol for the transmission of continuous media in real time through the Internet.

Nowadays, IETF framework based on SIP is more extended than the framework from ITU-T. Almost all software-implemented solutions use SIP and protocols from the IETF, while hardware solutions remain using H.323 recommendation. Nevertheless, both SIP and H.323 are converging to each other, providing the same services, but SIP is still easier to work with than H.323. It is likely that SIP will dominate audio and video conferencing market in the near future.

10. References

- Bolot, J.-C. & Vega-García, A. (1996). Control mechanisms for packet audio in the Internet. *Proceedings of 15th Annual Joint Conference of the IEEE Computer Societies. Networking the Next Generation*, pp. 232–239, ISBN 0-8186-7293-5, March 1996, San Francisco, CA, USA.
- Chong, H. & Matthews, H. (2004). Comparative analysis of traditional telephone and Voice-over-Internet Protocol (VoIP) systems. *Proceedings of the 2004 IEEE International Symposium on Electronics and the Environment*, pp. 106–111, ISBN 0-7803-8250-1, May 2004, Scottsdale, AZ, USA.
- Dalgic, I. & Fang, H. (1999). Comparison of H.323 and SIP for IP telephony signaling. *Proceedings of SPIE Volume 3845*, pp. 106–122, ISBN 0-8194-3438-8, September 1999, Boston, MA, USA.
- Dommel, H.-P. & García-Luna-Aceves, J. (1997). Floor control for multimedia conferencing and collaboration. *Multimedia Systems*, Vol. 5, No. 1, pp. 23–38, ISSN 0942-4962, Germany.
- García, D. F., Uría, C., Granda, J.C. & Suárez, F.J. (2007). A Functional Evaluation of the Commercial Platforms and Tools for Synchronous Distance e-Learning. *North Atlantic University Union International Journal of Education and Information Technologies*. Vol. 1, No. 2, pp. 95–104, ISSN 1109-9577.
- Glasmann, J., Kellerer, W. & Müller, H. (2004). *Service architectures in H.323 and SIP: A comparison*. IEEE Communications Surveys & Tutorials, Vol. 5, No. 2, pp. 32–47, ISSN 1553-877X.
- Granda, J.C., García, D. F., Suárez, F.J., Peteira, I. & Uría, C. (2008). A Multimedia Tool for Synchronous Distance e-Training of Employees in Geographically Dispersed

- Industries. *Proceedings of 7th IASTED International Conference on Internet & Multimedia Systems & Applications*, ISBN 978-0-88986-727-7, March 2008, Innsbruck, Austria.
- Jain, M. & Dovrolis, C. (2003). End-to-end available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput. *IEEE/ACM Transactions on Networking*, Vol. 11, No. 4, pp. 537-549, ISSN 1063-6692.
- Malpanim, R. & Rowe, L. A. (1997). Floor control for large-scale MBone seminars. *Proceedings of the 5th ACM International Conference on Multimedia*, pp. 155-163, ISBN 0-89791-991-2, November 1997, Seattle, WA, USA.
- Mishra, S. & Sharma, R. C. (2004). *Interactive Multimedia in Education and Training*. Idea Group Publishing, ISBN 1-59140-393-6. USA.
- Mollet, N. & Arnaldi, B. (2006). Storytelling in virtual reality for training. *Proceedings of 1st International Conference on Technologies for E-Learning and Digital Entertainment*, pp. 334-347, ISBN 3-540-33423-8, April 2006, Hangzhou, China.
- Padmore, M., Hall, L., Hogg, B. & Paley, G. (2006). Reviewing the potential of Virtual Learning Environments in schools. *Proceedings of 1st International Conference on Technologies for E-Learning and Digital Entertainment*, pp. 203-212, ISBN 3-540-33423-8, April 2006, Hangzhou, China.
- Paxson, V. (1999). End-to-end Internet packet dynamics. *IEEE/ACM Transactions on Networking*, Vol. 7, No. 3, pp. 277-292, ISSN 1063-6692.
- Shulman, R. E. (1992). Multimedia – A high-tech solution to industry's training malaise. *Supermarket business*, Vol. 47, No. 4, pp. 23-24, ISSN 0196-5700.
- Schulzrinne, H. & Rosenberg, J. (1998). A comparison of SIP and H.323 for internet telephony. *Proceedings of International Workshop on Network and Operating System Support for Digital Audio and Video*, pp. 83-86, July 1998, Cambridge, UK.
- Steinmetz, R. & Nahrstedt, K. (1995). *Multimedia: Computing, Communication and Applications*. Prentice-Hall Inc., ISBN 0-13-324435-0, USA.
- Weller M. (2007). The distance from isolation: Why communities are the logical conclusion in e-learning. *Computer & Education*, Vol. 49, No. 2, pp. 148-159, ISSN 0360-1315.
- Yajnik, M., Moon, S. B., Kurose, J. F. & Towsley, D. F. (1999). Measurement and Modeling of the Temporal Dependence in Packet Loss. *Proceedings of 8th Annual Joint Conference of the IEEE Computer and Communications Societies*, pp. 345-352, ISBN 0-7803-5417-6, March 1999, New York, NY, USA.



E-learning Experiences and Future

Edited by Safeeullah Soomro

ISBN 978-953-307-092-6

Hard cover, 452 pages

Publisher InTech

Published online 01, April, 2010

Published in print edition April, 2010

This book is consisting of 24 chapters which are focusing on the basic and applied research regarding e-learning systems. Authors made efforts to provide theoretical as well as practical approaches to solve open problems through their elite research work. This book increases knowledge in the following topics such as e-learning, e-Government, Data mining in e-learning based systems, LMS systems, security in e-learning based systems, surveys regarding teachers to use e-learning systems, analysis of intelligent agents using e-learning, assessment methods for e-learning and barriers to use of effective e-learning systems in education. Basically this book is an open platform for creative discussion for future e-learning based systems which are essential to understand for the students, researchers, academic personals and industry related people to enhance their capabilities to capture new ideas and provides valuable solution to an international community.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Juan C. Granda, Christian Uria, Francisco J. Suarez and Daniel F. Garcia (2010). Supporting Technologies for Synchronous E-learning, E-learning Experiences and Future, Safeeullah Soomro (Ed.), ISBN: 978-953-307-092-6, InTech, Available from: <http://www.intechopen.com/books/e-learning-experiences-and-future/supporting-technologies-for-synchronous-e-learning>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen