Interworking Between SIP/SDP and H.323*

Kundan Singh and Henning Schulzrinne
Dept. of Computer Science
Columbia University
New York, USA
{kns10,hgs}@cs.columbia.edu

May 8, 2000

Abstract

There are currently two standards for signaling and control of Internet telephone calls, namely ITU-T Recommendation H.323 and the IETF Session Initiation Protocol (SIP). We describe how an interworking function (IWF) can allow SIP user agents to call H.323 terminals and vice versa. Our solution addresses user registration, call sequence mapping and session description. We also describe and compare various approaches for multi-party conferencing and call tranfer.

1 Introduction

It appears likely that both the Session Initiation Protocol (SIP) [1, 2], together with the Session Description Protocol (SDP) [3], and the ITU-T recommendation H.323 in its various versions [4, 5] will be used for setting up Internet multimedia conferences and telephone calls. For example, currently H.323 is the most widely used protocol for PC-based conferences, due to the widespread availability of Microsoft's NetMeeting tool, while carrier networks using so-called soft switches and IP telephones seem to be built based on SIP. Thus, in order to achieve universal connectivity, interworking between the two protocols is desirable. This paper describes approaches to achieving this.

The ITU-T Recommendation H.323 [4] defines packet-based multimedia communication systems and is based heavily on previous ITU-T multimedia protocols.

*This work was supported by a grant from Sylantro Corp. An earlier version of this paper appeared in the *1st IP-Telephony Workshop (IPTel'2000)*, Berlin, April 2000.

In particular, H.323 call signaling is inspired by H.320 [6] for ISDN, and call control by H.324 [7] for GSTN terminals. SIP [1], developed in the IETF, builds on a simple text-based request-response architecture similar to other Internet protocols such as HTTP [8] and RTSP [9]. With the exception of conference control, SIP provides a similar set of basic services as H.323 [10, 11].

Interworking between the protocols is made simpler since both operate over IP (Internet Protocol) and use RTP (Real time Transport Protocol [12]) for transferring realtime audio/video data, reducing the task of interworking between these protocols to merely translating the signaling protocols and session description. Since no media data needs to be translated, a single gateway can likely serve thousands of end systems.

Interworking between SIP and H.323 requires transparent support of signaling and session descriptions between the SIP and H.323 entities. We call the server providing this translation a SIP-H.323 *interworking function* (IWF). We refer to the set of terminals speaking H.323 and SIP as the H.323 and SIP *networks*, respectively, even though they are likely to be intermingled on the same IP network. We use the term *native network* to refer to the network used by a particular terminal, while the *foreign network* is the network whose access is mediated by the IWF. For an H.323 terminal, a SIP terminal is in a foreign network.

When addressing a terminal using another signaling protocol, there are two approaches. First, the user can explicitly identify the protocol as part of the address, for example, by inventing some form of H.323 URL[1] such as *h323:alice@columbia.edu*. If, for example, an H.323 URL is used by a SIP terminal, it would then be the responsibility of the SIP terminal to find the appropriate IWF.

Alternatively, a terminal using a particular signaling protocol sees all other terminals as being native, and does not know or care that a particular address refers to a terminal in the foreign network. Indeed, an address could well change between being native and foreign, depending on what equipment the owner of the address happens to be using. This approach is preferable, but requires that user registrations are exported into the foreign network. Depending on the type of information sharing between H.323 or SIP elements and the IWF, different architectures are possible to provide the transparent address resolution and call establishment, as we will discuss below.

1.1 Outline of the rest of the paper

The remainder of the paper is organized as follows. In Section 2, we list the problems in translating SIP to H.323 and vice versa. Section 3 describes and compares

[1] Such a URL scheme was proposed by Cordell [13] in an expired Internet draft.

different approaches to address user registration. In Section 4, we describe a mechanism to map SIP addresses to H.323 addresses. Call sequence mapping between SIP and H.323 is described in Section 5. Section 6 gives an insight into translating multi-party conferencing and call transfer. Finally, we describe our current implementation and future work in Section 8.

2 Background

2.1 Protocol overview

H.323 includes various other subprotocols: H.225.0 [14] for connection setup and media transport (RTP), resource access and address translation, H.245 [15] for call control and capability negotiation, H.332 [16] for large conferences, H.235 [17] for security, H.246 [18] for interoperability with the PSTN, H.450.x [19, 20, 21] for supplementary services like call transfer.

In H.323, a simple call is established as follows. If a user (say Alice) wants to talk to another user (Bob), Alice first sends an admission request to its gatekeeper. The *gatekeeper* acts as a management entity in H.323, which grants access to resources, controls bandwidth and maps user names to IP addresses, among other things. The gatekeeper finds out the IP addresses at which Bob can be reached and informs Alice. After that, Alice establishes a TCP connection to the IP address of Bob. This is followed by ISDN-like *call signaling* procedure. Alice sends a Q.931 [22] SETUP message and Bob responds with a Q.931 CONNECT message. Once the first stage of Q.931 signaling is complete, H.245 takes over. H.245 messages are used to negotiate terminal capabilities, i.e., the support for various audio/video algorithms. The H.245 OpenLogicalChannel procedure is used for opening different unidirectional media channels. A *media channel* is defined as a pair of UDP channels, one for RTP and the other for RTCP. Audio and video packets are encapsulated in RTP and sent from one end system to the other. Depending on the version of H.323, Q.931 and H.245 steps can be combined in various ways.

SIP sets up calls with an INVITE message and a response from the called party. Both INVITE and the response contain a *session description* indicating terminal capabilities, typically, but not necessarily, encoded using SDP. Proxy and redirect servers are responsible for translating between user names and the called party's IP address.

2.2 Call setup translation

Three pieces of information are needed for establishing an call between two endpoints, namely the signaling destination address, local and remote media capa-

bilities, and local and remote media transport addresses at which the endpoint can receive the media packets. In H.323, this information is spread over different stages of the call setup, while SIP conveys it in an INVITE message and its response.

Translating a SIP call to an H.323 call is straightforward. The IWF gets all three pieces of information in the SIP INVITE message and can split it across multiple stages of the H.323 call establishment. However, in the reverse direction, from H.323 to SIP, the different stages of H.323 call establishment have to be merged into a single SIP INVITE message. We describe and compare various approaches in Section 5. The H.323v2 (version 2.0) Fast Connect procedure is a step towards simplifying the multi-stage signaling of H.323. However, it is optional and an H.323v2 entity is required to support the traditional multi-stage signaling. Thus, we describe call setup both with and without Fast Connect.

2.3 User registration

SIP-H.323 translation also has to solve the user registration problem. User registration involves mapping of user names, phone numbers or some other human-understandable identifier such as email addresses to network addresses. By allowing users to be reached by location-independent identifiers, User registration provides personal mobility. For instance, a call destined at *sip:bob@mydomain.com* reaches user Bob no matter what IP address he might currently be using.

In SIP, proxy and redirect servers access a location server, often a registrar that receives user registration information. A server at *mydomain.com* will map all the addresses of the form *sip:xyz@mydomain.com* to the appropriate IP addresses, depending on where *xyz* is currently logged in. In H.323, the same functionality is performed by the H.323 gatekeeper. The IWF should use the user registration information available in both networks to resolve a user name to an IP address. The IWF can contain a SIP registrar server, an H.323 gatekeeper or neither, as discussed in Section 3.

2.4 Session description

An IWF also must map session descriptions between the two signaling protocols. H.323 uses H.245 for session description. H.245 can negotiate media capabilities, provide conference floor control, and establish and tear down media channels. In H.245, media capabilities are described as a set of capability descriptors, listed in decreasing order of preference. A *capability descriptor*, also called a simultaneous capability set, is a set of alternative capability sets, where each alternative capability set contains a list of algorithms, only one of which can be used at any given time. For instance, a capability descriptor $\{[a_1, a_2][v_1, v_2][d_1]\}$ has three alterna-

tive capability sets: $[a_1, a_2]$, $[v_1, v_2]$, and $[d_1]$. It indicates that the terminal can support audio, video and data simultaneously. Audio can use either codec a_1 or a_2, video codec v_1 or v_2, and data format d_1.

SIP can, in principle, use any session description format. In practice, however, SDP is used exclusively. SDP lists media types and the supported encodings for each. Unlike H.245, SDP cannot express cross-media or inter-media constraints, however. For example, SDP cannot indicate that for a particular media type, the other side can only choose subset A or subset B of the listed codecs, but not codecs from both subsets. Similarly, SDP cannot express that certain audio codecs can only be used in conjunction with certain video codecs.

Thus, a SIP media capability can be easily described in H.245, however the reverse is more complicated. One approach is to carry multiple SDP messages in the message body of SIP INVITE requests and responses, using the "multipart" content type. Each SDP message then represents one capability descriptor of the H.245 capability set. In Section 5 we describe how sending multiple SDP messages can be avoided.

2.5 Multi-party conferencing

Ad-hoc conferencing among SIP and H.323 end systems is not possible without modifying one or both of these protocols. Ad hoc conferencing is defined as the one in which the participants do not know in advance whether the call will be point-to-point (two-party) or multi-party. The participants can switch from a point-to-point call to a multi-party conference or vice-versa during the call. It is possible for the participants to invite a third party in the conference or for the third party to join the conference. Both SIP and H.323 individually support ad hoc conferencing. In SIP, conference topology can be a full mesh with every participants having a signaling relationship with every other participant or a centralized bridged conference (star topology) in which every participant has a signaling relationship with the central conference bridge [23, 24]. It is possible to switch from a mesh to a bridged conference. In H.323, conferences are managed by central entity called a *Multipoint Controller* (MC). An MC can be part of an H.323 terminal, gateway, gatekeeper, or MCU (Multipoint Control Unit). H.323 conferences have inherently a star topology with every participant having an H.245 control channel with the MC. The MC is responsible for deciding the common media capabilities for the conference, conference floor control, and other conferencing functions. All the participants are required to obey the media capabilities given by the MC. Because of the difference in the topology of the conferences in the SIP and H.323 (star like in H.323 and full mesh or star like in SIP), the transparent support of multiparty conferencing cannot be achieved without modifying the protocols. However, with some simplifying

assumptions, basic conferences can be set up, as described in Section 6.

2.6 Call services

Advanced call services like call forwarding and call transfer are supported by both SIP and H.323. H.323 uses H.450.x for these supplementary services. SIP has support for blind transfer, operator assisted transfer, call forwarding, call park and directed call pickup [23]. These services are not yet widely deployed, so that translation is not critical at this moment. Section 6 describes some of the issues related to this.

2.7 Security and quality of service

Other problems in SIP-H.323 translation include security and quality of service (QoS). Both, SIP and H.323, individually support these. However, translating from the open architecture of SIP, where security and QoS is independent of the connection establishment, to H.323, where security and QoS go hand-in-hand with the call establishment, remains an open issue.

3 Architecture for user registration

In this section, we describe different architectures for user registration and address resolution. *User registration servers* are the entities in the network which store user registration information. SIP registrars and H.323 gatekeepers are user registration servers. It simplifies locating users independent of the signaling protocol if the IWF has direct access to user registration servers. The user registration server forwards the registration information from one network, to which it belongs, to the other.

3.1 IWF contains SIP proxy and registrar

Our first approach combines an IWF with a SIP registrar and proxy server, as shown in. Fig. 1(a). In this approach the registration information is maintained by the H.323 gatekeeper(s). Whenever the SIP registrar receives a SIP REGISTER request, it generates a registration request (RRQ) to the H.323 gatekeeper, translating a SIP URI into H.323 Alias Address. H.323 users register via the usual H.225.0 procedure. Since the SIP registration information is also available through the H.323 gatekeeper(s), any H.323 entity can resolve the address of SIP entities reachable via the SIP server/IWF. In the other direction, if a SIP user agent wants to talk to another user, who happen to reside in the H.323 network, it sends a SIP INVITE message to the SIP server. The SIP server multicasts H.323 location requests

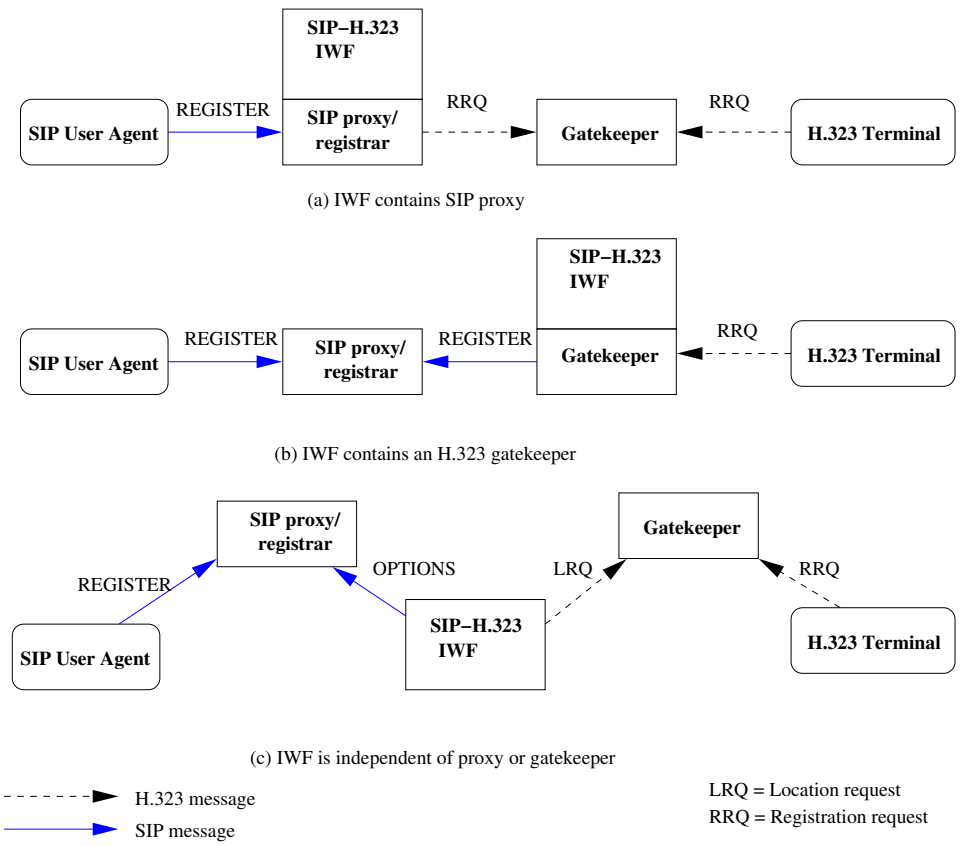


Figure 1: Architectures for user registration

(LRQ) to the H.323 gatekeepers. The gatekeeper to which the H.323 user is registered responds with the IP address of the H.323 user. Once the SIP server knows that the address belongs to the H.323 world, it can route the call to the destination.

One drawback of this approach is that the H.323 gatekeepers are burdened with all the registrations in the SIP network.

This approach only makes those SIP addresses handled by the registrar available to the H.323 zone. Typically, a registrar is responsible for a single domain, e.g., *columbia.edu*. Thus, each H.323 zone would have to have an IWF. If an H.323 user wants to call a SIP terminal, first the H.323 terminal locates, using DNS TXT records, [25, p. 57] the appropriate gatekeeper[2], which in turn uses the registration information conveyed by the IWF to discover that this address is actually located in the SIP network.

[2]It is not clear how widely implemented this approach is.

3.2 IWF contains an H.323 gatekeeper

This architecture, shown in Fig. 1(b) is similar to the previous approach except that the SIP proxy server maintains the user registration information from both networks. Any H.323 registration request received by the H.323 gatekeeper is forwarded to the appropriate SIP registrar, which thus stores the user registration information of both the SIP and H.323 entities.

To the SIP terminal, H.323 terminals simply appear as SIP URLs within the same domain. (See Section 4 on how H.323 addresses are translated to SIP URLs.) If an H.323 entity wants to talk to a user who happens to reside in the SIP network, it sends an admission request (ARQ) to its gatekeeper. The gatekeeper multicasts the location request (LRQ) to all the other gatekeepers. The GK-IWF server captures the request and tries to find out if the address belongs to a SIP user. It does so by sending a SIP OPTIONS request, which does not set up any call state. If the address is valid in the SIP network and the user is currently available to be called, the IWF responds with the location confirmation (LCF), letting the H.323 terminal know that the destination is reachable.

This approach has the similar drawback as the previous approach (Section 3.1) in that the proxy has to store all H.323 registration information.

However, this approach has the advantage that even if some H.323 gatekeepers are not equipped with a IWF, the address resolution works: If an H.323 gatekeeper cannot resolve a called address, it multicasts a location request (LRQ) to the other gatekeepers in the network. As long as at least one H.323 gatekeeper exists with the SIP-H.323 signaling translation capability, the SIP user can be located from the H.323 network. Note that the previous approach (Section 3.1) required that all the SIP registrars/proxy servers must be equipped with IWFs.

3.3 IWF is independent of proxy or gatekeeper

In the third approach, shown in Fig. 1(c), the IWF is not colocated with either an H.323 gatekeeper or an SIP proxy server. User registration is done independently in the SIP and H.323 networks. However, when a call reaches the IWF, the IWF queries the other network for user location. Here, we assume that the IWF is capable of interpreting and responding to the location request (LRQ) from the H.323 network.

The address resolution mechanism works as follows. Suppose the SIP user Sam wants to talk to Henry, an H.323 user. Henry has registered with its own gatekeeper in the H.323 network and the gatekeeper knows Henry's IP address, conveyed via RRQ. When Sam contacts the SIP proxy with Henry's name, the SIP proxy has no registration for Henry, but is configured to contact the IWF in case the called party

is in the H.323 network. The IWF, in turn, multicasts the location request (LRQ) for Henry to all gatekeepers. If there is no positive response from the gatekeepers of the H.323 network within a timeout period, the IWF concludes that the address is not valid in the H.323 network and the branch fails.

In the other direction, Henry sends an admission request (ARQ) to its gatekeeper. Since this gatekeeper does not have the address mapping for Sam, it multicasts the location request (LRQ) for Sam to the other gatekeepers in the network. In addition, the IWF is tuned to receive the LRQ. The IWF then uses the SIP OPTIONS request (as in Section 3.2) to find out if Sam is available in the SIP network and informs the GK if the request succeeds. This is followed by H.323 call establishment between Henry and the IWF and a SIP call between the IWF and Sam.

The IWF should support direct H.323 connections. For instance, a SIP user (Sam) should be able to call an H.323 user (Henry) through the IWF (say *sip323.columbia.edu*) by placing a call to *sip:henry@sip323.columbia.edu*. Similarly, the H.323 user should be able to reach a SIP user (*sip:sam@mydomain.com*) by establishing a Q.931 TCP connection to the IWF and providing the destination address or the remote extension address in the Q.931 SETUP message as *sip:user1@mydomain.com*. The direct connection does not involve user registration and the caller is expected to know that the destination is reachable via the IWF.

4 Address translation

While user registration exports identities into the foreign network, address translation is performed by the IWF to create valid SIP addresses from H.323 addresses and vice versa. In SIP, addresses are typically SIP URIs of the form sip:*user@host*, where *user* names can also be telephone numbers. However, SIP terminals can also support other URLs schemes, for example "tel:" URLs for telephone numbers [26] or H.323 URLs [13]. Generally, SIP terminals proxy calls to their local server if they do not understand the particular URL scheme, in the hope that the server can translate it.

In H.323, addresses (ASN.1 AliasAddress) can take many forms, including unstructured identifiers (h323-ID), E.164 (global) telephone numbers, URLs of various types, host names or IP address, and email addresses (email-ID). Local user names and host names appear to be most common. For compatibility with H.323 version 1.0 entities, the h323-ID field of H.323 AliasAddress must be present.

For SIP-H.323 interoperability, there should be a consistent and unique way of mapping a SIP URI to an H.323 address and vice-versa. Translating a SIP URI to an H.323 AliasAddress is easy: We simply copy the SIP URI verbatim into the h323-ID. The user and host parts of SIP-URI are used to generate an email

identifier, "*user@host*", which is stored in the email-ID field of AliasAddress. The transport-ID parameter is copied from the host part of SIP-URI if the latter is given numerically. The e164 field is extracted from the user part of SIP address if it is marked as a telephone number.

Translating an H.323 AliasAddress to a SIP address is more difficult since multiple representations (e.g., e164, url-ID, transport-ID) need to be merged into a single SIP address. In the easiest case, the alias contains a url-ID with a SIP URI, in which case it is simply copied into the SIP message. Otherwise, if the h323-ID can be parsed as a valid SIP address (e.g., "Alice <sip:alice@host>" or "alice@host") it is used. Next, if the transport-ID is present and it does not point to the IWF itself, then it forms the host and port portions of the SIP URI. Finally, if the H.323 alias has an email-ID, it is used in the SIP URI prefixed with "sip:" URI scheme.

Note that the translated address may not necessarily be valid. On the H.323 side, it may be desirable to configure a gatekeeper to route all calls that are not resolvable within the H.323 network to the IWF, which would then attempt a translation to a SIP URI. This would allow H.323 terminals to reach any SIP terminal, even those not cross-registered.

5 Connection establishment

Once the user knows that the destination is reachable via the IWF, the connection is established. A point-to-point call from Alice to Bob needs three cruicial pieces of information, namely the logical destination address (A) of Bob, the media transport address (T) at which each of the users is ready to receive media packets (RTP/RTCP) and a description of the media capabilities (M) of the parties. Alice should know A, T and M of Bob and Bob needs to know Alice's T and M. The difficulty in translating between SIP and H.323 arises because A, M, and T are all contained in the SIP INVITE request and its response, while H.323 may spread this information among several messages.

5.1 Using H.323v2 Fast Connect

If the H.323v2 Fast Connect procedure is available, the protocol translation is simplified because fast connect establishes a call in a single stage, with a one-to-one mapping between H.323 and SIP call establishment messages. Both the H.323 SETUP message with fast connect and the SIP INVITE request have all three components. If the call succeeds, both the H.323 CONNECT message with Fast Connect, and the SIP 200 response, including the session description, have the

required components (M and T of the call destination).

Since Fast Connect is optional in H.323v2, an H.323 entity must be able to handle calls without the Fast Connect feature for backward compatibility. In particular, the IWF must accept a non-Fast Connect call from the H.323 side. In the other direction, the IWF should try to use H.323v2 Fast Connect, but must be prepared to switch to the multi-stage call establishment procedure if the response from the H.323 entity indicates that this is not supported.

5.2 Call translation without using Fast Connect

Translating a SIP call to an H.323 call is straightforward even without Fast Connect. The IWF uses A, M and T for the Q.931 and H.245 phases. The responses from the H.323 side are collated and forwarded to the SIP side, as shown in Fig. 2.

A multi-stage H.323 call can be translated to a SIP call in a variety of ways. One obvious approach is to accept the H.323 call without informing the SIP user agent. The H.323 call proceeds between the H.323 terminal and the IWF as if the IWF is just another H.323 terminal. The IWF may get the media capabilities of the SIP user agent using the SIP OPTIONS message. Media capabilities of the H.323 terminal are obtained via H.245 capability negotiation. Once the logical channels are established from the IWF to the H.323 terminal, the IWF knows M and T and can place a SIP call by sending an INVITE. The media transport address from the 200 response is conveyed to the H.323 terminal while acknowledging the OpenLogicalChannel requests of the H.323 terminal.

While this approach is pretty simple, it has the disadvantage that the IWF accepts the call without even asking the actual destination, leading to caller confusion if the SIP destination is not reachable.

This problem can be solved if the IWF sends a SIP INVITE without session description or a session description without media transport information when receiving the Q.931 SETUP message from the H.323 terminal. Only after the SIP user agent has accepted the call, the IWF forwards the confirmation (Q.931 CONNECT) to the H.323 terminal. The rest of the call establishment proceeds as before, except that the SIP OPTIONS message is not needed because the 200 response from the SIP user agent describes the media capabilities.

The media capabilities of the H.323 terminal are received in the H.245 TerminalCapabilitySet message and are forwarded to the SIP user agent as part of the ACK message or via an additional INVITE. The media capabilities of the SIP user agent are found in the session description of the 200 response to the INVITE request.

The different interpretations of media capabilities by H.245 and SDP potentially cause problems during the call. In SDP, a receive media capability of G.711

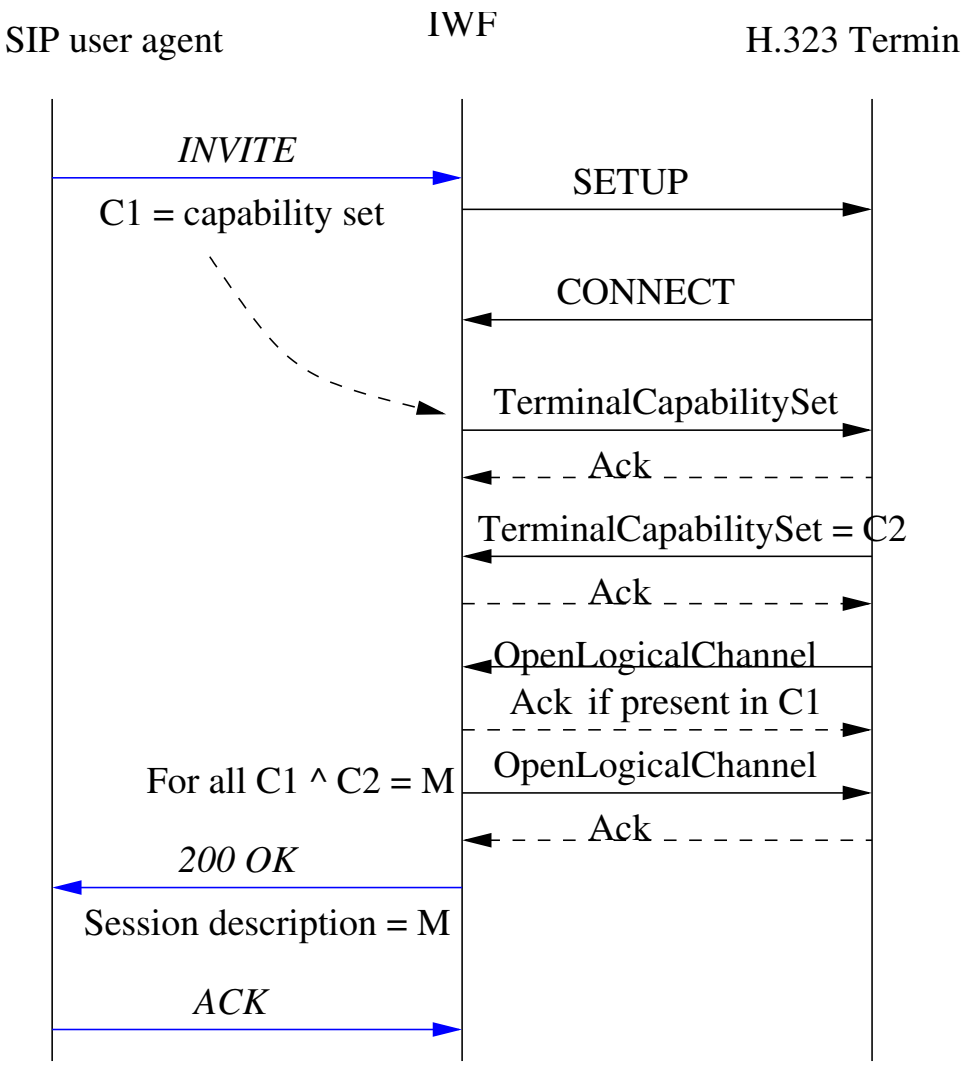


Figure 2: Call from SIP terminal to H.323 terminal without Fast Connect

and G.723.1 means that the sender can switch between these algorithms at any time during a call without explicitly informing the receiver beyond changing the RTP payload type. However, in H.245, the sender chooses an algorithm from the capability set of the receiver and explicitly opens a logical channel for that algorithm. The sender cannot switch dynamically to another algorithm without informing the receiver. The sender has to close the previous logical channel and re-open it with new algorithm. Alternatively, the receiver can use H.245 ModeRequest to request

the sender to use a different algorithm.

This problem can be addressed by having the RTP/RTCP packets from SIP to H.323 be intercepted by the IWF. If the IWF detects a change in coding algorithm, it initiates the required H.245 procedures. However, this approach is not advisable, as it scales poorly.

Another approach limits the media description sent to the SIP side to only one algorithm per media (or per alternative capability set). This can be achieved by maintaining a maximal intersection of the SIP and H.323 terminal capability sets. A maximal intersection of two capability sets is a capability set which is a subset of both the capability sets and no other superset is a subset of those capability sets. The operating mode, that is, the selected algorithms for the call, is derived from the intersection of the two capability sets by selecting one algorithm per alternative capability set. If the SIP side sends additional INVITE requests during the call to change media parameters, the IWF simply recalculates the operating modes.

Finding maximal intersection of capability sets is described in [27]. As an example, let the SIP capability set be {[PCMU,PCMA,G.723.1][H.261]} and H.323 capability set be {[PCMU,PCMA,G.729][H.261]} {[G.723.1][H.263]} (i.e., the SIP user can support PCMU, PCMA or G.723.1 audio and H.261 video, whereas the H.323 user can support either one of the PCMU, PCMA, G.729 audio with H.261 video or G.723.1 audio with H.263 video). The maximal intersection as calculated by the IWF is {[PCMU,PCMA][H.261]} {[G.723.1]}. The IWF derives an operating mode by selecting a capability descriptor from the maximal intersection and selecting one algorithm per alternative capability set (e.g., {PCMU,H.261}). The IWF conveys only the PCMU audio and H.261 video to the SIP user agent. If the SIP side sends additional INVITE with a different capability set ({[G.729,G.723.1][H.261]}, the new maximal intersection becomes {[G.729][H.261]}{[G.723.1]}. The IWF derives a new operating mode ({G.729,H.261}) and initiates the H.245 procedure to change the PCMU audio to G.729.

6 Translating advanced services

Both SIP and H.323 support advanced services like multi-party conferencing and call transfer. In this section we propose possible approaches for translating these services.

6.1 Multi-party conferencing

A transparent support for multi-party conferencing can be achieved by having the IWF mirror the endpoint(s) in each direction. Fig. 4 shows a scenario in which two

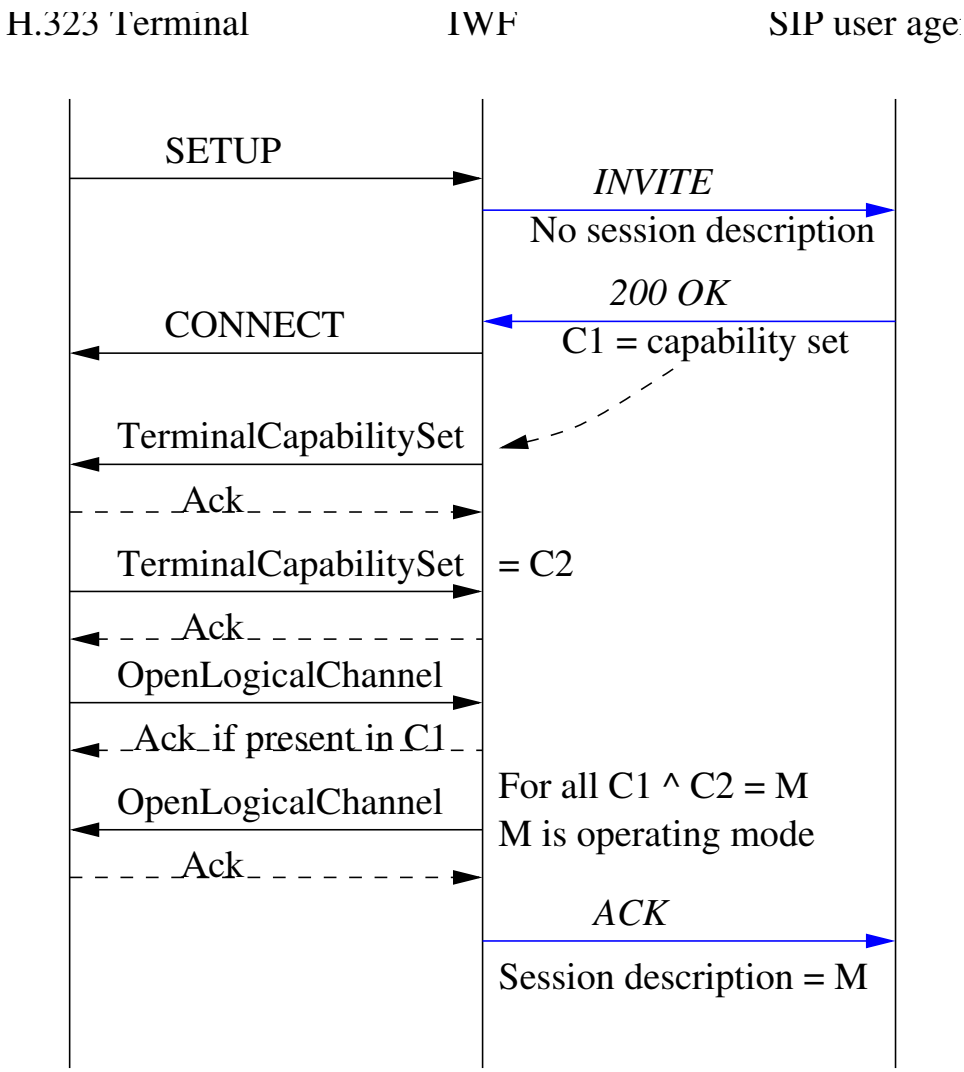
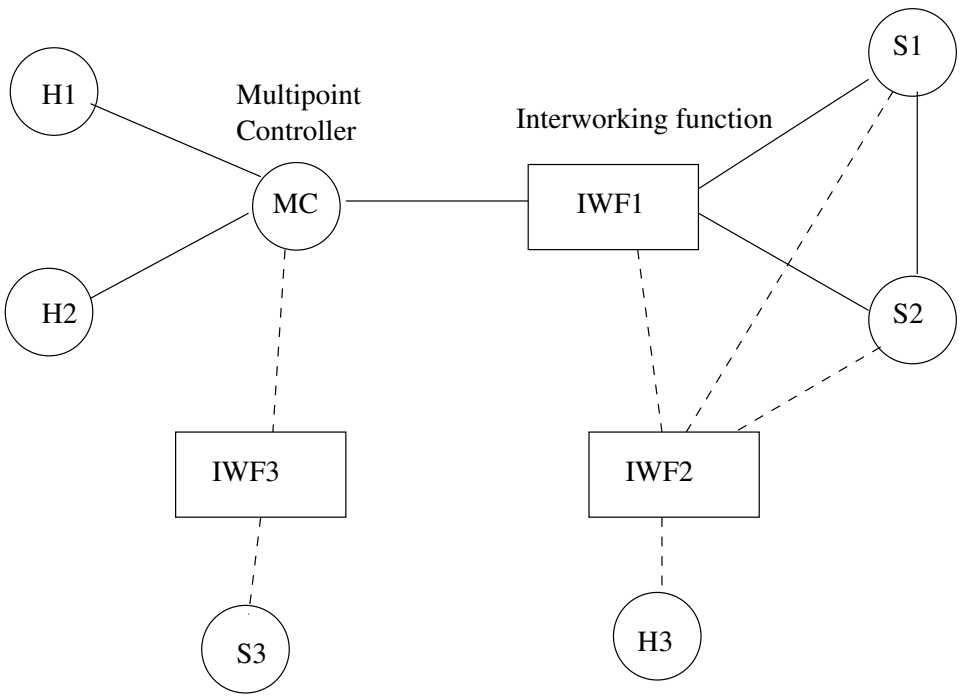


Figure 3: Call from H.323 to SIP terminal call without Fast Connect

H.323 terminals (H1 and H2) and two SIP user agents (S1 and S2) are involved in a conference. From the H.323 side, the interworking function (IWF1) looks like a single H.323 terminal. From the SIP side, the IWF acts as a single SIP user agent.

This approach fails if S1 invites another H.323 user H3 via a different interworking function (IWF2). How will the other participants such as H2 know that H3 has joined the conference? Alternatively, if H1 invites a SIP user, S3, S2 will not know of the presence of S3. One way for the participants to know about the



Convention: Hn : H.323 terminals; Sm : SIP user agents

Figure 4: Ad-hoc conferencing among SIP and H.323 endpoints

existence of the other participants is to rely on the RTP/RTCP packets. This goes against the idea of H.323 conferencing where H.245 messages are used to convey the existence of new participants.

We can solve this problem by forcing all invitations to pass through the IWF. Fig. 5(a) shows a conference managed by an MC where H.323 terminals are directly connected to the MC and SIP user agents are connected through IWFs. A SIP user agent is allowed to only invite other SIP UAs through the IWF, so that the IWF can update the MC state. In a SIP-centric architecture, Fig. 5(b), the H.323 terminals take part in the conference through the IWFs.

We recommend a SIP-centered architecture because the SIP conferencing model is more general, allowing full mesh with distributed control or centralized bridged conferences. In general, translating services is greatly simplified if an operator adopts a primary signaling protocol, with services offered only in that protocol. Terminals using another protocol are restricted to making calls through the IWF.

Supporting H.332 loosely coupled conferences is straightforward, since SDP

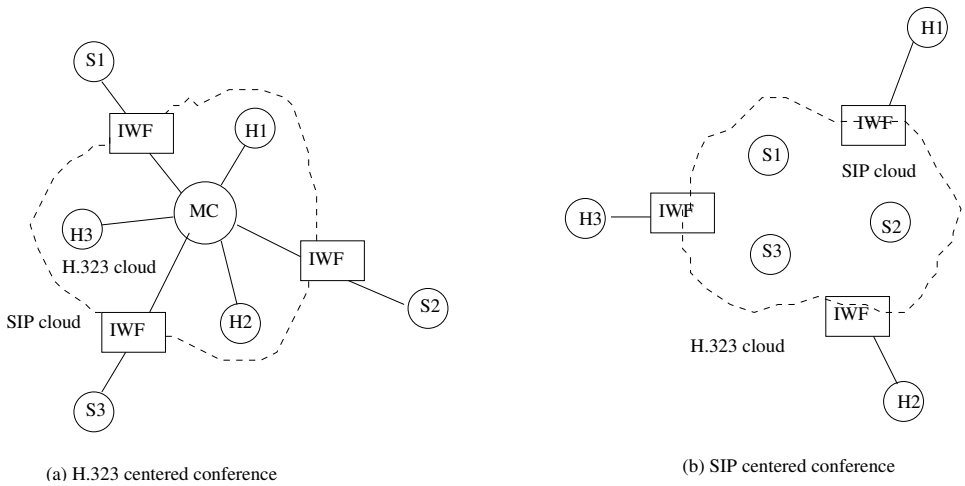


Figure 5: Different conferencing architectures

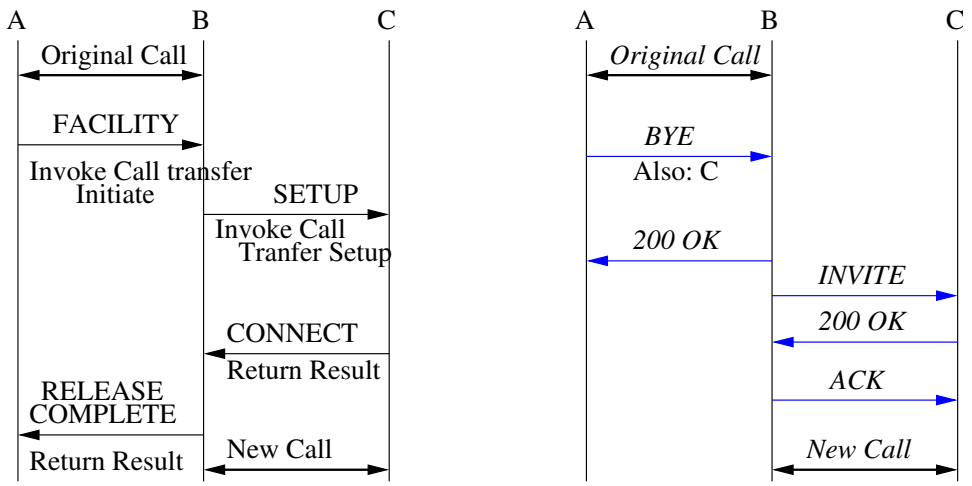
is used in that context.

6.2 Call transfer

Call transfer is one of the many supplementary services needed for internet telephony. The idea is to transfer a call between two entities (say, A and B) to a call between B and C. Fig. 6 shows the message sequence in H.323 and SIP and a possible translation when A and B are H.323 terminals and C is a SIP user agent.

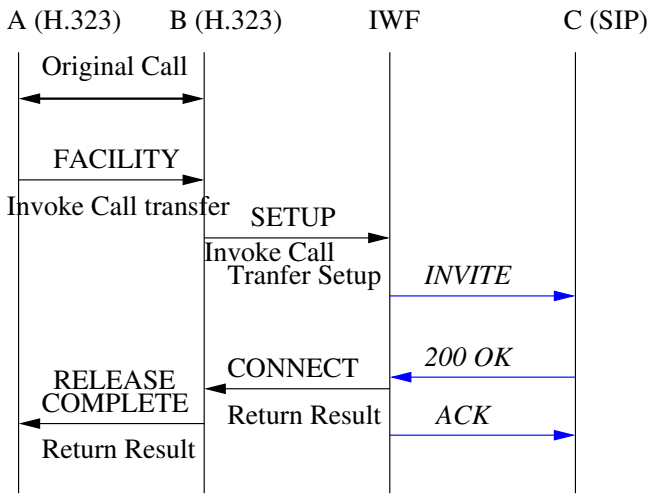
A difference between SIP and H.323 arises because of the different philosophies of protocol extension. H.323 designers identify a supplementary service such as call transfer, call forwarding, call hold and define a new set of messages to accomplish it. This results in different procedures for different advanced services (e.g., H.450.2 for call transfer, H.450.3 for call diversion, H.450.4 for call hold). In SIP, crucial information needed for call services is identified and is encapsulated in new message headers (e.g., Also, Replaces, Requested-By). Different call services are then designed using these building blocks.

A number of open issues remain when translating advanced services, including whether all call parameters can be translated and how security and authentication are to be handled.



(a) Call transfer in H.323

(b) Call transfer in SIP



(c) Call transfer in mixed network. A and B are H.323 terminals and C is a SIP user agent.

Figure 6: An example of call transfer mapping

7 Related work

The problem of interworking between SIP and H.323 has only recently started to attract attention, with ETSI TIPHON and ITU now likely to get involved.

Details of the SIP-H.323 interworking described here can be found in [27]. Agboh [28] and Kausar and Crowcroft [29] address the problem of interworking, but do not solve the issues of registration and media capability translation.

8 Conclusion and future work

We have described a framework for interworking between SIP and H.323. The challenges include call sequence mapping, address translation and mapping session descriptions.

Ad-hoc conferencing among SIP and H.323 participants is not possible without modifying one or both of these protocols. The problem can be made tractable by keeping an IWF aware of all call state changes.

H.323 has picked up a number of features from SIP, such as Fast Connect or, more recently, UDP-based signaling. It is possible that further convergence may occur, although not without fundamental changes to either SIP or H.323.

We have implemented a basic interworking function using the OpenH323 library and a SIP signaling stack developed locally and demontrated a simple audio call setup between SIP user agents and Microsoft NetMeeting.

We have yet to address the issue of multistage translation, where two H.323 users communicate via a SIP gateway. It is not yet clear how common such a scenario would be, given direct network connectivity between the two parties.

9 Acknowledgments

We would like to thank the members of the `sip-h323` mailing list (sip-h323@eGroups.com) for their comments.

References

[1] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, "SIP: session initiation protocol," Request for Comments (Proposed Standard) 2543, Internet Engineering Task Force, Mar. 1999.

[2] H. Schulzrinne and J. Rosenberg, "Internet telephony: Architecture and protocols – an IETF perspective," *Computer Networks and ISDN Systems*, vol. 31, pp. 237–255, Feb. 1999.

[3] M. Handley and V. Jacobson, "SDP: session description protocol," Request for Comments (Proposed Standard) 2327, Internet Engineering Task Force, Apr. 1998.

[4] International Telecommunication Union, "Packet based multimedia communication systems," Recommendation H.323, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Feb. 1998.

[5] J. Toga and J. Ott, "ITU-T standardization activities for interactive multimedia communications on packet-based networks: H.323 and related recommendations," *Computer Networks and ISDN Systems*, vol. 31, pp. 205–223, Feb. 1999.

[6] International Telecommunication Union, "Narrow-band visual telephone systems and terminal equipment," Recommendation H.320, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, May 1999.

[7] International Telecommunication Union, "Terminal for low bit-rate multimedia communication," Recommendation H.324, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Feb. 1998.

[8] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee, "Hypertext transfer protocol – HTTP/1.1," Request for Comments (Draft Standard) 2616, Internet Engineering Task Force, June 1999.

[9] H. Schulzrinne, A. Rao, and R. Lanphier, "Real time streaming protocol (RTSP)," Request for Comments (Proposed Standard) 2326, Internet Engineering Task Force, Apr. 1998.

[10] H. Schulzrinne and J. Rosenberg, "A comparison of SIP and H.323 for internet telephony," in *Proc. International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV)*, (Cambridge, England), pp. 83–86, July 1998.

[11] I. Dalgic and H. Fang, "Comparison of H.323 and SIP for IP telephony signaling," in *Proc. of Photonics East*, (Boston, Massachusetts), SPIE, Sept. 1999.

[12] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: a transport protocol for real-time applications," Request for Comments (Proposed Standard) 1889, Internet Engineering Task Force, Jan. 1996.

[13] P. Cordell, "Conversational multimedia URLs," Internet Draft, Internet Engineering Task Force, Dec. 1997. Work in progress.

[14] International Telecommunication Union, "Media stream packetization and synchronization on non-guaranteed quality of service LANs," Recommendation H.225.0, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Nov. 1996.

[15] International Telecommunication Union, "Control protocol for multimedia communication," Recommendation H.245, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Feb. 1998.

[16] International Telecommunication Union, "H.323 extended for loosely coupled conferences," Recommendation H.332, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Sept. 1998.

[17] International Telecommunication Union, "Security and encryption for H-Series (H.323 and other H.245-based) multimedia terminals," Recommendation H.235, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Feb. 1998.

[18] International Telecommunication Union, "Interworking of h-series multimedia terminals with H-Series multimedia terminals and voice/voiceband terminals on GSTN and ISDN," Recommendation H.246, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Feb. 1998.

[19] International Telecommunication Union, "Generic functional protocol for the support of supplementary services in h.323," Recommendation H.450.1, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Feb. 1998.

[20] International Telecommunication Union, "Call transfer supplementary service for H.323," Recommendation H.450.2, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Feb. 1998.

[21] International Telecommunication Union, "Call diversion supplementary service for H.323," Recommendation H.450.3, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Sept. 1997.

[22] International Telecommunication Union, "Digital subscriber signalling system no. 1 (dss 1) - isdn user-network interface layer 3 specification for basic call control," Recommendation Q.931, Telecommunication Standardization Sector of ITU, Geneva, Switzerland, Mar. 1993.

[23] H. Schulzrinne and J. Rosenberg, "SIP call control services," Internet Draft, Internet Engineering Task Force, June 1999. Work in progress.

[24] H. Schulzrinne and J. Rosenberg, "Signaling for internet telephony," Technical Report CUCS-005-98, Columbia University, New York, New York, Feb. 1998.

[25] O. Hersent, D. Gurle, and J.-P. Petit, *IP telephony*. Reading, Massachusetts: Addison Wesley, 2000.

[26] A. Vaha-Sipila, "URLs for telephone calls," Internet Draft, Internet Engineering Task Force, Dec. 1999. Work in progress.

[27] K. Singh and H. Schulzrinne, "Interworking between SIP/SDP and H.323," Internet Draft, Internet Engineering Task Force, Jan. 2000. Work in progress.

[28] C. Agboh, "A study of two main ip telephony signaling protocols: H.323 signaling and sip; a comparison and a signaling gateway specification," Master's thesis, Unversite Libre de Bruxelles (ULB), Facuts des Science, Dpartment Informatique, Brussels, Belgium, 1999. supervised by Eric Manie.

[29] N. Kausar and J. Crowcroft, "An architecture of conference control functions," in *Proc. of Photonics East*, (Boston, Massachusetts), SPIE, Sept. 1999.