IP Telephony Cookbook

> TERENA REPORT //

> / MARCH, 2004

TERENA
TRANS-EUROPEAN RESEARCH AND EDUCATION NETWORKING ASSOCIATION

The IP Telephony Cookbook was created through the IP Telephony project as a reference document for setting up IP Telephony solutions at university campuses and NRENs. The project started in April 2003 and ran until February 2004. The Cookbook provides an overview of available and future IP Telephony technologies, scenarios for IP Telephony deployment and infrastructures, guidelines on protocols, service set-ups and connection to a global 'dialling plan'. Furthermore, the Cookbook reports on the interoperability of equipment, existing IP Telephony projects and regulatory aspects.

The project was carried out by the University of Pisa, Italy, TZI-University of Bremen and FhG FOKUS, Germany, with contributions from CESNET, GRNET, SURFnet and the University of Graz.

Funding to the project was provided by TERENA, ARNES, CARNet, CYNET, SUNET and UKERNA. Representatives of the funding organisations were members of the Project Review Committee.

ISBN 90-7759-08-6

Production: TERENA Secretariat
Design: Eva de Lange
Printing: GraphicResult

AUTHORS:

Margit Brandl - Karl Franzens - UNI Graz, Dimitris Daskopoulos - GRNET,

Erik Dobbelsteijn - SURFnet, Rosario Giuseppe Garroppo - University of Pisa, Jan Janak - FhG Fokus,

Jiri Kuthan - FhG Fokus, Saverio Niccolini - University of Pisa, Jörg Ott - Universität Bremen TZI,

Stefan Prelle - Universität Bremen TZI, Sven Ubik - CESNET Uni Graz, Egon Verharen - SURFnet

CONTENTS //

1 Introduction //>

> 1.1 Goal

The IP Telephony Cookbook is a reference document addressing technical issues for the setup of IP Telephony solutions. Its goal is to provide the user community with guidelines and information about the IP Telephony world and everything related to it. Since the Cookbook is intended to be a technical document, the main target audience are the network engineers and system administrators at universities and national research and education networks (NREN); however, university students and researchers may find it useful, both for enriching their technological background as well as for finding information about advanced research topics and projects in the European community.

> 1.2 Reasons for writing this document

Members of the NREN community asked TERENA to start an investigation into IP Telephony in September 2001. The response was very positive and suggestions were made to co-ordinate the creation of a cookbook with recommendations for setting up IP Telephony solutions at university- and national-level, with information about protocols and the interoperability of equipment as well as about integration with the existing international hierarchies for IP videoconferencing. For this reason, a number of people in the TERENA community with significant expertise in the area of IP Telephony decided to undertake this task and to compose this document, The IP Telephony Cookbook.

> 1.3 Contents

The IP Telephony Cookbook is divided into chapters, which guide the reader through increasing levels of knowledge of the IP Telephony world. This first chapter contains introductory information and gives details of the contents of the Cookbook, useful tips on how to read this document and techno-economic considerations. Chapter 2 explains the technological background needed in order to understand the topics addressed in the rest of the Cookbook. This chapter describes the basic IP Telephony components and gives an overview of the IP Telephony protocols. Chapter 2 ends with additional considerations on call routing and perspectives about the future. Chapter 3 gives a high-level overview of scenarios a user may face when building an IP Telephony environment. Details are given to explain what a particular scenario is about, what is needed in order to deploy it and what needs it is serving. The next three chapters (Chapter 4, Chapter 5 and Chapter 6) detail how to set up IP Telephony services; those chapters give the reader the chance to learn how to set up basic services, advanced services (still telephony-centric) and value-added services (with respect to classic telephony service). Chapter 7 is about the integration of global telephony, describing the technological solutions available for the integration of global IP Telephony and the successful replacement of classic telephony. Chapter 7 reports on today's situation, as well as migration and future trends. The last chapter contains the regulatory/legal considerations users have to be aware of when moving from classic telephony to IP Telephony. The topics here relate to the regulation of IP Telephony in Europe and in other countries outside the European Union. A large number of legal issues for classic telephony are detailed, from licensing to unbundling, and their mapping to the IP world. Finally, the IP Telephony Cookbook contains two annexes. Annex A lists and describes current and future IP Telephony Projects in Europe. Annex B gives the reader useful information about IP Telephony hardware and software, reporting 'hands on' experience (i.e., how the devices performed, how good tech-support was, what were the workarounds for some of the problems faced, etc).

> 1.4 How to read this document

Since the IP Telephony Cookbook is a technical reference document, it must include guidelines for users who do not want to read the whole document, so that they can find the information they need. In this section, we give the reader tips on how to read the document in order to retrieve the information needed as fast as possible; for a detailed overview of the contents of the Cookbook, please refer to the previous section. To speed up the information retrieval process, each reader should identify himself as belonging to one of the following three groups:

- readers who have no knowledge of IP Telephony;
- readers who have basic knowledge of IP Telephony;
- readers who have advanced knowledge of IP Telephony.

Readers belonging to the first group should, first of all, refer to Chapter 2 to acquire the necessary background to understand the rest of the cookbook. Readers who are interested in setting up an IP Telephony service should read Chapter 3 to have a clear picture of the possible scenarios offered by IP Telephony and target the one best-suited to the needs of their environment. The second group of readers may skip the previously-mentioned chapters, but Chapter 3 may be of some interest to them; the main focus of this group of users is more likely to be in Chapter 4 and Chapter 5 which give tips and help in setting up an operative service. The third group of users is likely to be more interested in the 'value added' services available nowadays with IP Telephony (Chapter 6) or in the integration problems of an IP Telephony architecture that is widely distributed across multiple sites and organisations (Chapter 7). All three groups of users may find useful information in Chapter 8 and European project information in Annex A. Last but not least, the list of products and testing experience reported in Annex B is a must for all users who do not want to risk making the wrong choices in a buying decision.

> 1.5 Techno-economic aspect of moving from classic telephony to VoIP

Many institutions are facing investment decisions with respect to replacing or expanding their existing telephony infrastructure, which currently consists mainly of large PBXs with proprietary phones and interfaces. As is there such a clear trend to replace old-style (TDM) PBXs with IP Telephony ones, it is important that there is a guide on how to attach such an IP Telephony solution to the existing network. IP connectivity can be used as the basis for establishing good communication between scientists that might not use traditional, still relatively expensive, long-distance calls as extensively as they could use IP Telephony. Even where financial constraints are not the driving force, the potential for enhancing IP Telephony with additional services that support scientific co-operation makes IP Telephony an attractive solution.

IP Telephony can provide a number of benefits beyond replacing existing PBX/PSTN telephony:

Enhanced speech quality
The PSTN (and most PBXs) are limited to 3.1 kHz, 8-bit/sample audio. It is likely that future IP phones can provide CD quality and possibly even stereo audio. Even where the additional bandwidth required for this extreme level of quality cannot be provided; modest codecs such as G.722 (7 kHz speech bandwidth) can be used to provide better quality than conventional telephony;

Improved availability
There are many aspects of availability. Lowering the cost can make telephony more available to low-budget activities. Redundancy can provide as good as (or even better) reliability than traditional telephony. Integrating telephony with location-based computing and group-awareness systems can make the communication partners much more 'available', or provide the means to transfer communication to a point in time where it is more appropriate than the usual interrupt-driven telephone call;

Improved coverage
In a similar argument, IP Telephony can be made available in places where traditional phones are often not available in a university, e.g., lab settings (in particular, student labs). Also, many universities still consider the cost of phone installations high enough to force their employees to share phones in a common office, again, not necessary when workstation-based IP Telephony is used;

Improved mobility
It is very easy to move an IP phone to another room. There is no need to deal with ports on the PBX and change dial numbers. Simply plugging it into an ethernet socket in a new room makes it available;

Improved media integration
IP phones can be enabled to add media to an ongoing call as required, e.g., viewing a picture or drawing on a whiteboard. Using workstations themselves as IP phones can facilitate providing this function, whereas the standards are not yet there for coupling traditional phones and workstations;

New services
As IP Telephony evolves, it can be used to provide new services (like user-defined call processing) or to integrate existing concepts, e.g., Presence, Location Awareness or Instant Messaging. Because of the open standards available for these services, they need not to be limited to vendor-specific solutions. In other words, it can be much easier to deal with issues such as CTI (Computer Telephony Integration) and so pave the way to a completely new way of understanding telephony;

Research
As mentioned before, the protocols and standards used for IP Telephony are open and publicly available. This allows research institutions to work on their own services and solutions.

It is important to point out that before introducing IP Telephony into the network of an organisation; several issues unknown to the old telephone system have to be taken into account. A rough, non-exhaustive list may include addressing (special subnet/VLAN for phones), Quality of Service (QoS), security, positioning of gateways, interfacing of firewalls and, last but not least, maintenance of the system (backups, spares, etc., - something not very common in the legacy PBX world).

With regard to the economic aspects, the 'packetisation' of voice using Voice over IP has given rise to new international telecommunications carriers. These carriers have distributed network architectures using the Internet as a platform. VoIP networks have an architecture offering the most efficient way to implement multilateral telecommunications agreements, thus eliminating the need for carriers to engage in hundreds of bilateral traffic agreements as are required between traditional circuit-switched PSTN carriers. Moreover, since packet networks are software driven, they can be configured more dynamically than traditional PSTN networks. For example, with a global voice over packet network, new destinations are available to all users on the network, without the need for constant additional investment.

IP Telephony telecommunications companies may expand the availability of services to a wider audience. IP Telephony technologies can be used to build voice networks more rapidly and at a lower cost than legacy PSTN systems. Easier deployment of Voice over IP networks can bring the benefits of telecommunications to more people in a much shorter timeframe than would be possible with conventional PSTN networks. At the same time, not having to build extensive infrastructure provides the motivation for many companies to migrate to IP Telephony architectures.

2 Technological Background }

This chapter provides technical background information about the protocols and components used in IP Telephony. It introduces the relevant component types, gives detailed information about H.323, SIP and RTP as well as information about media gateway control and vendor -specific protocols.

} 2.1 Components

An IP Telephony infrastructure usually consists of different types of components. This section gives an overview of typical components without describing them in a protocol-specific context.

} 2.1.1 Terminal

A terminal is a communication endpoint that terminates calls and their media streams. Most commonly, this is either a hardware or a software telephone or videophone, possibly enhanced with data capabilities. There are terminals that are intended for user interaction and others that are automated, e.g., answering machines.

An IP Telephony terminal is located on at least one IP address. There may well be multiple terminals on the same IP address but they are treated independently. Most of the time, a terminal has been assigned one or more addresses (see Section 2.1.5), which others will use to dial to it. If IP Telephony servers are used, a terminal registers the addresses with its server.

} 2.1.2 Server

Placing an IP Telephony call requires at least two terminals, and the knowledge of the IP address and port number of the terminal to call. Obviously, forcing the user to remember and use IP addresses for placing calls is not ideal and dynamic IP addressing schemes (DHCP) make this requirement even more intolerable.

As mentioned before, terminals usually register their addresses with a server. The server stores these telephone addresses along with the IP addresses of the respective terminals, and is thus able to map a telephone address to a host.

When a telephone user dials an address, the server tries to resolve the given address into a network address. To do so, the server may interact with other telephony servers or services. It may also provide further call routing mechanisms like CPL (Call Processing Language) scripts or skill-based routing (e.g., route calls to 'WWW-Support' to a list of persons who are tagged to be responsible for this subject).

Finally, a telephony server is responsible for authenticating registrations, authorising calling parties and performing the accounting

2.1.3 Gateway

Gateways are telephony endpoints that facilitate calls between endpoints that usually would not interoperate. Usually this means that a gateway translates one signalling protocol into another (e.g. SIP/ISDN signalling gateways), but translating between different network addresses (IPv4/IPv6) or codecs (media gateways) can be considered gatewaying as well. Of course, it is possible that multiple functionalities exist in a single gateway.

Finding gateways between VoIP and a traditional PBX is usually quite simple. Gateways that translate different VoIP protocols are harder to find. Most of them are limited to basic call functionality.

2.1.4 Conference bridge

Conference bridges provide the means to have 3-point or multi-point conferences that can either be ad-hoc or scheduled. Because of the high resource requirements, conference bridges are usually dedicated servers with special media hardware.

2.1.5 Addressing

A user willing to use a communication service needs an identifier to describe himself and the called party. Ideally, such an identifier should be independent of the user's physical location. The network should be then responsible for finding the current location of the called party. A specific user may define to be reached by multiple contact address identifiers.

Regular telephony systems use E.164 numbers (the international public telecommunication numbering plan). An identifier is composed of up to fifteen digits with a leading plus sign, for example, +1234565789123. When dialling, the leading plus is normally replaced by the international access code, usually double zero (00). This is followed by a country code and a subscriber number.

The first IP Telephony systems used the IP addresses of end-point devices as user identifiers. Sometimes they are still used now. However, IP addresses are not location-independent (even if IPv6 is used) and they are hard to remember (especially if IPv6 is used) so they are not suitable as user identifiers.

Current IP Telephony systems use two kinds of identifiers:
- URIs (RFC2396);
- Numbers (E.164).

Some systems tried to use names (alpha-numeric strings), but this led to a flat naming space and thus limited zones of applicability.

A Universal Resource Identifier (URI) uses a registered naming space to describe a resource in a location-independent way. Resources are available under a variety of naming schemes and access methods including e-mail addresses (mailto), SIP identifiers (sip), H.323 identifiers (h.323, RFC3508) or telephone numbers (draft-ietf-iptel-rfc2806bis-02). E-mail-like identifiers have several advantages. They are easy to remember, nearly every Internet user already has an e-mail address and a new service can be added using the same identifier. The user location can be found with a Domain Name System (DNS). The disadvantage of URIs is that they are difficult or impossible to dial on some user devices (phones).

If we want to integrate a regular telephony system with IP Telephony, we must deal with phone number identifiers even on the IP Telephony-side. The numbers are not well suited for an Internet world relying on domain names. Therefore, the ENUM system was invented, using adapted phone numbers as domain names. ENUM is described in Chapter 7.

{ 2.2. Protocols

{ 2.2.1 H.323

The H.323 Series of Recommendations evolved out of the ITU-T's work on video telephony and multimedia conferencing. After completing standardisation on video telephony and videoconferencing for ISDN at up to 2 Mbit/s in the H.320 series, the ITU-T took on work on similar multimedia communication over ATM networks (H.310, H.321), over the analogue Public Switched Telephone Network (PSTN) using modem technology (H.324), and over the stillborn Isochronous Ethernet (H.322). The most widely-adopted and hence most promising network infrastructure - and the one bearing the largest difficulties to achieve well-defined Quality of Service - was addressed in the beginning of 1995 in H.323: Local Area Networks, with the focus on IP as the network layer protocol. The primary goal was to interface multimedia communication equipment on LANs to the reasonably well-established base on circuit-switched networks.

The initial version of H.323 was approved by the ITU-T about one year later, in June 1996, thereby providing a base on which the industry could converge. The initial focus was clearly on local network environments, because QoS mechanisms for IP-based wide area networks, such as the Internet, were not well established at this point. In early 1996, Internet-wide deployment of H.323 was already explicitly included in the scope, as was the aim to support voice-only applications and, thus, the foundations to use H.323 for IP Telephony were laid. H.323 has continuously evolved towards becoming a technically sound and functionally rich protocol platform for IP Telephony applications. The first major additions to this end were included in H.323 version 2, approved by the ITU-T in January 1998. In September 1999, H.323v3 was approved by the ITU-T, incorporating numerous further functional and conceptual extensions to enable H.323 to serve as a basis for IP Telephony on a global scale and as well as making it meet requirements in enterprise environments. Moreover, many new enhancements were introduced into the H.323 protocol. Version 4 was approved on November 17, 2000 and contains enhancements in a number of important areas, including reliability, scalability, and flexibility.

New features help facilitate more scalable Gateway and MCU solutions to meet the growing market requirements. H.323 has been the undisputed leader in voice, video, and data conferencing on packet networks, and Version 4 endeavours to keep H.323 ahead of the competition.

2.2.1.1 Scope

As stated before, the scope of H.323 encompasses multimedia communication in IP-based networks, with significant consideration given to gatewaying to circuit-switched networks (in particular to ISDN-based video telephony and to PSTN/ISDN/GSM for voice communication).

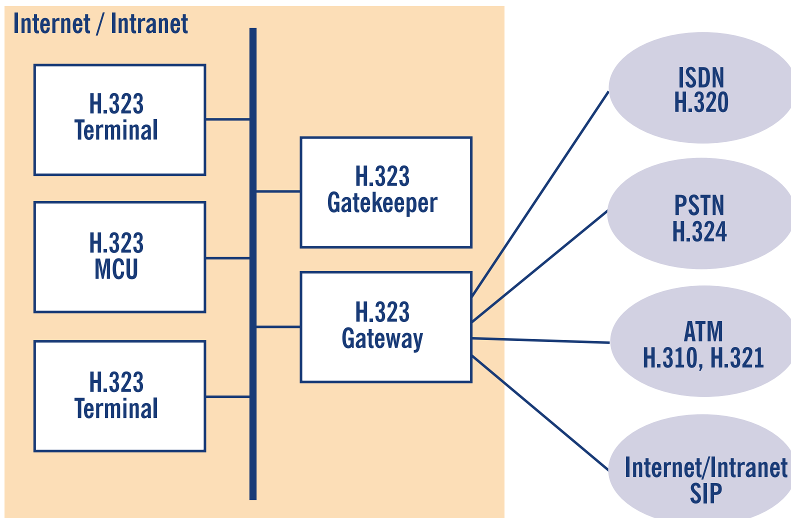

Figure 2.1 Scope and components defined in H.323

H.323 defines a number of functional / logical components as shown in Figure 2.1:

- **Terminal**
  Terminals are H.323-capable endpoints, which may be implemented in software on workstations or as stand-alone devices (such as telephones). They are assigned to one or more aliases (e.g. a user's name/URI) and/or telephone number(s);
- **Gateway**
  Gateways interconnect H.323 entities (such as endpoints, MCUs, or other gateways) to other network/protocol environments (such as the telephone network). They are also assigned one or more aliases and/or telephone number(s). The H.323 Series of Recommendations provides detailed specifications for interfacing H.323 to H.320, ISDN/PSTN, and ATM-based networks. Recent work also addresses control and media gateway specifications for telephony trunking networks such as SS7/ISUP;
- **Gatekeeper**
  The gatekeeper is the core management entity in an H.323 environment. It is, among other things, responsible for access control, address resolution and H.323 network (load) management and provides the central hook to implement any kind of utilisation / access policies. An H.323 environment is subdivided into zones (which may, but need not be congruent with the underlying network topology); each zone is controlled by one primary gatekeeper (with optional backup gatekeepers). Gatekeepers may also provide added value, e.g., act as a

conferencing bridge or offer supplementary call services. An H.323 Gatekeeper can also be equipped with the proxy feature. Such a feature enables the routing through the gatekeeper of the RTP traffic (audio and video) and the T.120 traffic (data), so no traffic is directly exchanged between endpoints. (It could be considered a kind of IP-to-IP gateway that can be used for security and QoS purposes);

- **Multipoint Controller (MC)**
  A Multipoint Controller is a logical entity that interconnects the call signalling and conference control channels of two or more H.323 entities in a star topology. MCs coordinate the (control aspects of) media exchange between all entities involved in a conference. They also provide the endpoints with participant lists, exercise floor control, etc. MCs may be embedded in any H.323 entity (terminals, gateways gatekeepers) or implemented as stand-alone entities. They can be cascaded to allow conferences spanning multiple MCs;
- **Multipoint Processor (MP)**
  For multipoint conferences with H.323, an optional Multipoint Processor may be used that receives media streams from the individual endpoints, combines them through some mixing/switching technique, and transmits the resulting media streams back to the endpoints;
- **Multipoint Control Unit (MCU)**
  In the H.323 world, an MCU is simply a combination of an MC and an MP in a single device. The term originates in the ISDN videoconferencing world where MCUs were needed to create multipoint conferences out of a set of point-to-point connections.

{ 2.2.1.2 Signalling protocols

H.323 resides on top of the basic Internet Protocols (IP, IP Multicast, TCP, and UDP) in a similar way as the IETF protocols discussed in the next subsection, and can make use of integrated and differentiated services along with resource reservation protocols.

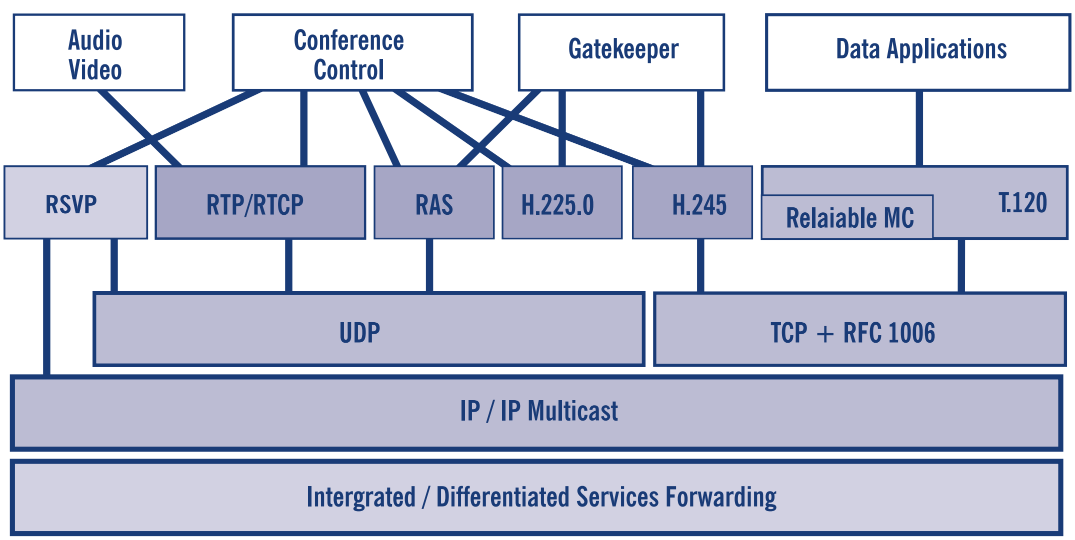

Figure 2.2 H.323 protocol architecture

For basic call signalling and conference control interactions with H.323, the aforementioned components communicate using three control protocols:

- **H.225.0 Registration, Admission, and Status (RAS)**
  The RAS channel is used for communication between H.323 endpoints and their gatekeeper and for some inter-gatekeeper communication. Endpoints use RAS to register with their gatekeeper, to request permission to utilise system resources, to have addresses of remote endpoints resolved, etc. Gatekeepers use RAS to keep track of the status of their associated endpoints and to collect information about actual resource utilisation after call termination. RAS provides mechanisms for user/endpoint authentication and call authorisation;
- **H.225.0 Call Signalling**
  The call signalling channel is used to signal call setup intention, success, failures, etc, as well as to carry operations for supplementary services (see below). Call signalling messages are derived from Q.931 (ISDN call signalling); however, simplified procedures and only a subset of the messages are used in H.323. The call signalling channel is used end-to-end between calling party and called party and may optionally run through one or more gatekeepers (the call signalling models are later described in the 'Signalling models' Section).

  Optimisations: Since version 3, H.225.0 supports the following enhancements:

  - **Multiple Calls** – To prevent using a dedicated TCP connection for each call, gateways can be built to handle multiple calls on each connection.

  - **Maintain Connection** – Similar to Multiple Calls, this enhancement will reduce the need to open new TCP connections. After the last call has ended, the endpoint may decide to maintain the TCP connection to provide a better call setup time for the next call.

  The primary use of both enhancements is at the communication between servers (gatekeeper, MCU) or gateways. While, in theory, both mechanisms were possible before, beginning with H.323v3, the messages contained fields to indicate support for the mechanisms;

- **H.245 Conference Control**
  The conference control channel is used to establish and control two-party calls (as well as multiparty conferences). Its functionality includes determining possible modes for media exchange (e.g., select media encoding formats that both parties understand) and configuring actual media streams (including exchanging transport addresses to send media streams to and receive them from). H.245 can be used to carry user input (such as DTMF) and enables confidential media exchange and defines syntax and semantics for multipoint conference operation (see below). Finally, it provides a number of maintenance messages. Also, this logical channel may (optionally) run through one or more gatekeepers, or directly between calling party and called party (please refer to the 'Signalling models' Section for details).

  It should be noted that H.245 is a legacy protocol inherited from the collective work on multimedia conferencing over ATM, PSTN and other networks. Hence it carries a lot of fields and procedures that do not apply to H.323 but make the protocol specification quite heavyweight.

  Optimisations:
  The conference control channel is also subject to optimisations. Per default, it is transported over an exclusive TCP connection but it may also be tunnelled within the signalling connection

(H.245 tunnelling). Other optimisations deal with the call setup time. The last chance to start an H.245 channel is on receipt of the CONNECT message which implies that the first seconds after the user accepted the call, no media is transmitted. H.245 may also start parallel to the setup of the H.225 call signalling, which is not really a new feature but another way of dealing with H.245. Vendors often call this **Early Connect** or **Early Media**. Since H.323v2, it is possible to start a call using a less powerful but sufficient capability exchange by simply offering possible media channels that just have to be accepted. This procedure, called **FastConnect** or **FastStart**, requires less round-trips and is transported over the H.225 channel. After the **FastConnect** procedure is finished or when it fails, the normal H.245 procedures start.

A number of extensions to H.323 include mechanisms for more efficient call setup (H.323 Annex E) and reduction of protocol overhead e.g., for simple telephones (SETs, simple endpoint types and H.323.Annex F).

{ 2.2.1.3 Gatekeeper discovery and registration

An H.323 endpoint usually registers with a gatekeeper that provides basic services like address resolution for calling the other endpoints. There are two possibilities for an endpoint to find its gatekeeper:

- **Multicast discovery**
  The endpoint sends a gatekeeper request (GRQ) to a well-known multicast address (224.0.1.41) and port (1718). Receiving gatekeepers may confirm their responsibility for the endpoint (GCF) or ignore the request
- **Configuration**
  The endpoint knows the IP address of the gatekeeper by manual configuration. While there is no need for a gatekeeper request (GRQ) to be sent to the preconfigured gatekeeper, some products need this protocol step. If a gatekeeper receives a GRQ via unicast, it must either confirm (GCF) the request or reject it (GRJ).

When trying to discover the gatekeeper via multicast, an endpoint may request any gatekeeper or specify the request by adding a gatekeeper identifier to the request. Only the gatekeeper that has the requested identifier may reply positively. (see Figure 2.3)

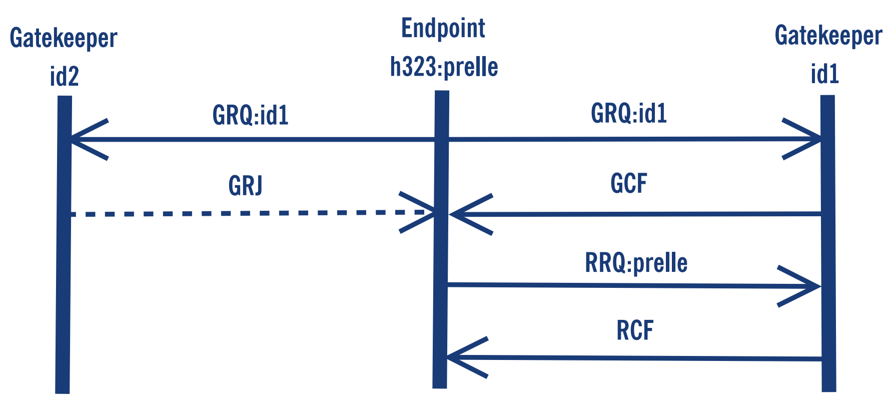

Figure 2.3 Discovery and registration process