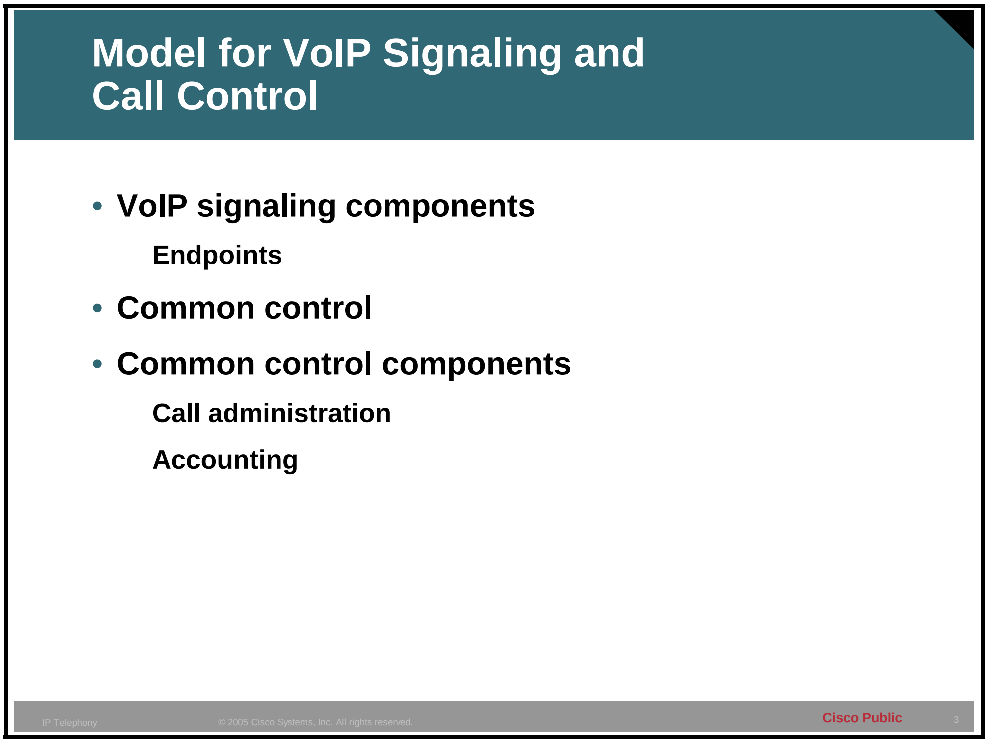
Need for Signaling and Call Control

VoIP Signaling

In a traditional voice network, call establishment, progress, and termination are managed by interpreting and propagating signals. Transporting voice over an IP internetwork creates the need for mechanisms to support signaling over the IP component of the end-to-end voice path. This topic introduces the components and services provided by VoIP signaling.



In the traditional telephone network, a voice call consists of two paths: an audio path carrying the voice and a signaling path carrying administrative information such as call setup, teardown messages, call status, and call-progress signals. ISDN D channel signaling and Common Channel Signaling System 7 (CCSS7) or Signaling System 7 (SS7) are two examples of signaling systems that are used in traditional telephony.

By introducing VoIP into the call path, the end-to-end path involves at least one call leg that uses an IP internetwork. As in a traditional voice call, support for this VoIP call leg requires two paths: a protocol stack that includes RTP, which provides the audio call leg, and one or more call control models that provide the signaling path.

A VoIP signaling and call control environment model includes endpoints and optional common control components, as follows:

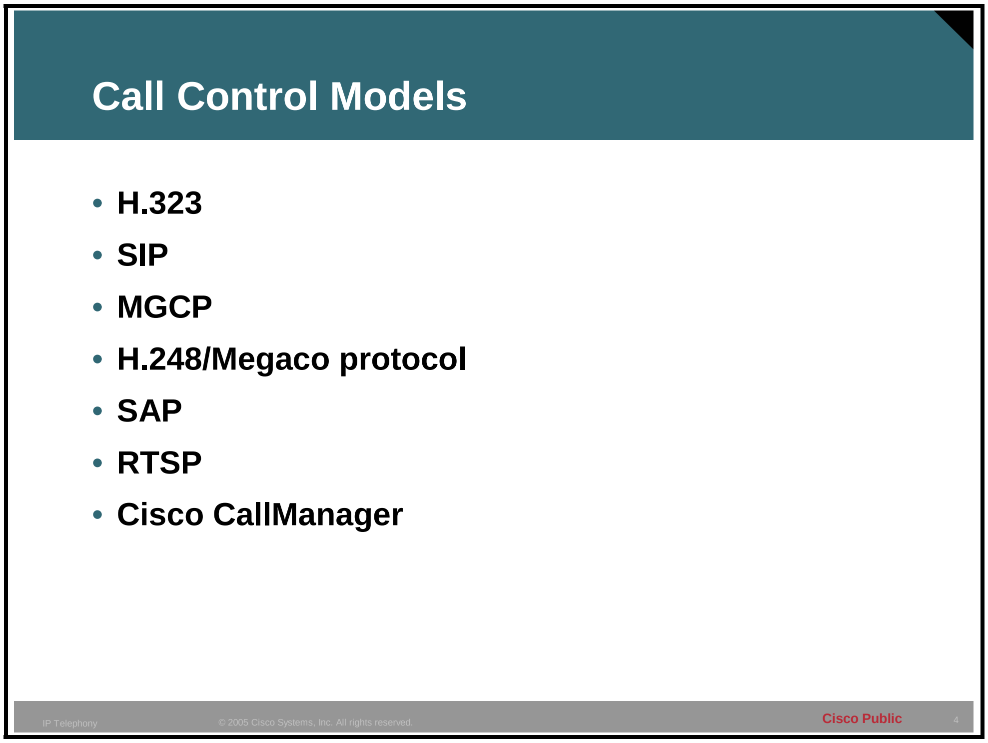
- **Endpoints:** Endpoints are typically simple, single-user devices, such as terminals, that support either a voice process (for example, the Cisco IP Phone application) or a gateway. In either case, the endpoint must be able to participate in signaling with other VoIP endpoints—directly or indirectly—through common control components. The endpoints must also be able to manipulate the audio that is in the audio path. This may involve performing analog-to-digital conversion or converting the format to digital voice so that it takes advantage of compression technology.

  Gateways provide physical or logical interfaces to the traditional telephone network. A gateway that is connected digitally to a service provider central office (CO) switch is an example of a gateway providing a physical interface. A gateway that provides access to an interactive response dialog application is an example of a gateway providing a logical interface.
- **Common control:** In some call control models, the common control component is not defined; in others, it is employed optionally. Common control components provide call administration and accounting. These components provide a variety of services to support call establishment, including the following:
  - Call status
  - Address registration and resolution
  - Admission control

Typically, the services of the common control components are implemented as applications. These services are colocated in a single physical device, or distributed over several physical devices with standalone endpoints and gateways.

Call Control Models

This topic describes several call control models and their corresponding protocols.

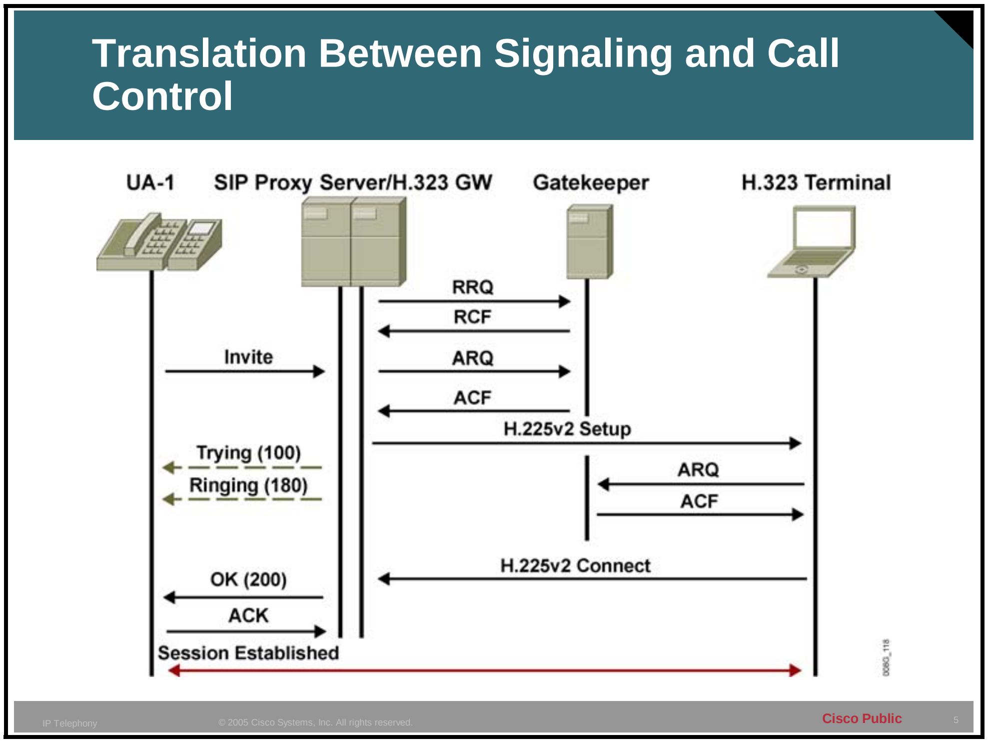


The following call control models and their corresponding protocols exist or are in development:

- **H.323:** International Telecommunication Union Telecommunication Standardization Sector (ITU-T) Recommendation H.323 describes the architecture to support multimedia communications over networks without quality of service (QoS) guarantees. Originally intended for LANs, H.323 has been adapted for IP.
- **SIP:** SIP is an Internet Engineering Task Force (IETF) RFC 3261 call control model for creating, modifying, and terminating multimedia sessions or calls.
- **MGCP:** MGCP (IETF RFC 2705) defines a call control model that controls VoIP gateways from an external call control element or call agent.
- **H.248/Megaco protocol:** The Megaco protocol is used in environments in which a media gateway consists of distributed subcomponents, and communication is required between the gateway subcomponents. The Megaco protocol is a joint effort of IETF (RFC 3015) and ITU-T (Recommendation H.248).
- **Session Announcement Protocol (SAP):** SAP (IETF RFC 2974) describes a multicast mechanism for advertising the session characteristics of a multimedia session, including audio and video.
- **Real Time Streaming Protocol (RTSP):** RTSP (IETF RFC 2326) describes a model for controlled, on-demand delivery of real-time audio and video.
- **Cisco CallManager (“Skinny”):** Cisco CallManager is a proprietary Cisco Systems implementation of a call control environment that provides basic call processing, signaling, and connection services to configured devices, such as IP telephones, VoIP gateways, and software applications.

Translation Between Signaling and Call Control Models

When VoIP endpoints support different call control procedures, the calls between the endpoints require cooperation between the originating and terminating procedures. This topic identifies the need for interworking or translation between call control models.



In the traditional telephone network, the individual call legs contributing to an end-to-end call often involve different signaling systems and procedures. In the graphic, an IP Phone is communicating with its SIP proxy server using the SIP protocol. However, it is also attempting to reach an H.323 endpoint. Because the two VoIP protocols are different, a translation is necessary at the SIP proxy server (namely, an H.323 gateway) to allow the two telephony endpoints to establish a connection.

Example: Call Control Translation

A call between a residential user and an office worker likely involves a signaling system that is unique to the various call legs that exist between the originator and the destination. In this scenario, the sequence of signaling systems includes the following:

- Analog signaling (Foreign Exchange Station [FXS] or Foreign Exchange Office [FXO] loop start) to the CO
- CCSS7 between the COs
- ISDN PRI signaling to the PBX
- Proprietary signaling to the desktop telephone

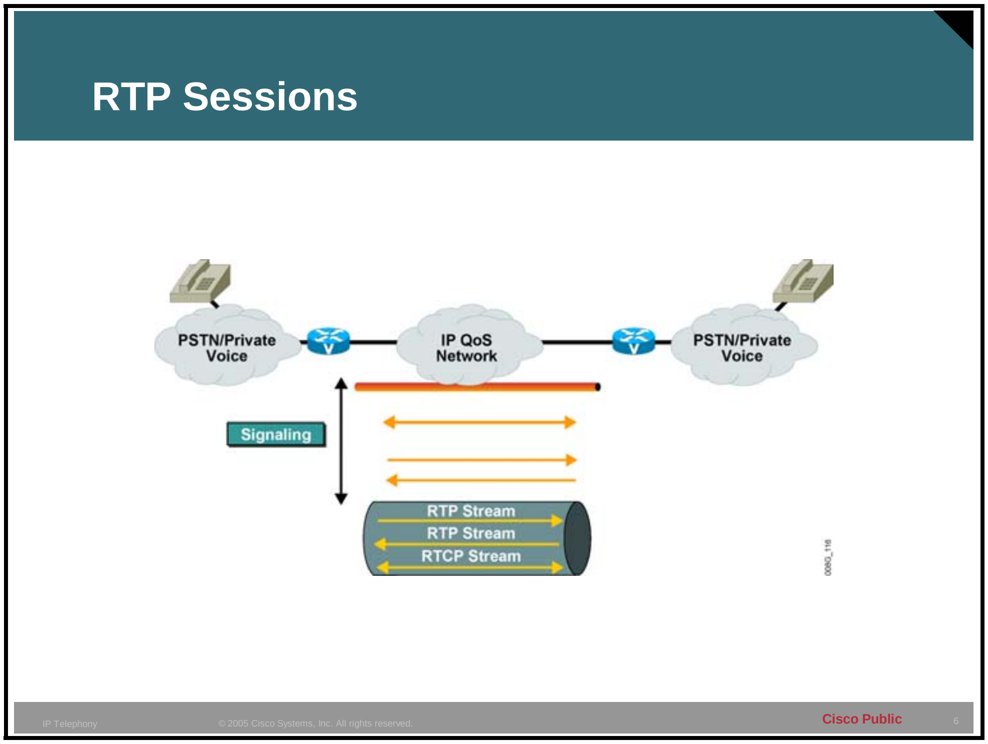
When part of the path is replaced with an IP internetwork, the audio path between the IP endpoints is provided by RTP, and the call control mechanism is based on a call control protocol, such as SIP, H.323, or MGCP.

But what if different call control models represent the endpoints? What if, for example, the originating endpoint uses H.323 and the destination is managed as an SIP endpoint?

To complete calls across the IP internetwork, a call control gateway that recognizes the procedures of *both* call control models is required. In particular, the translating gateway interprets the call setup procedure on the originating side and translates the request to the setup procedure on the destination side. Ideally, this translation is transparent to the endpoints that are involved and results in a single endpoint-to-endpoint audio relationship.

Call Setup

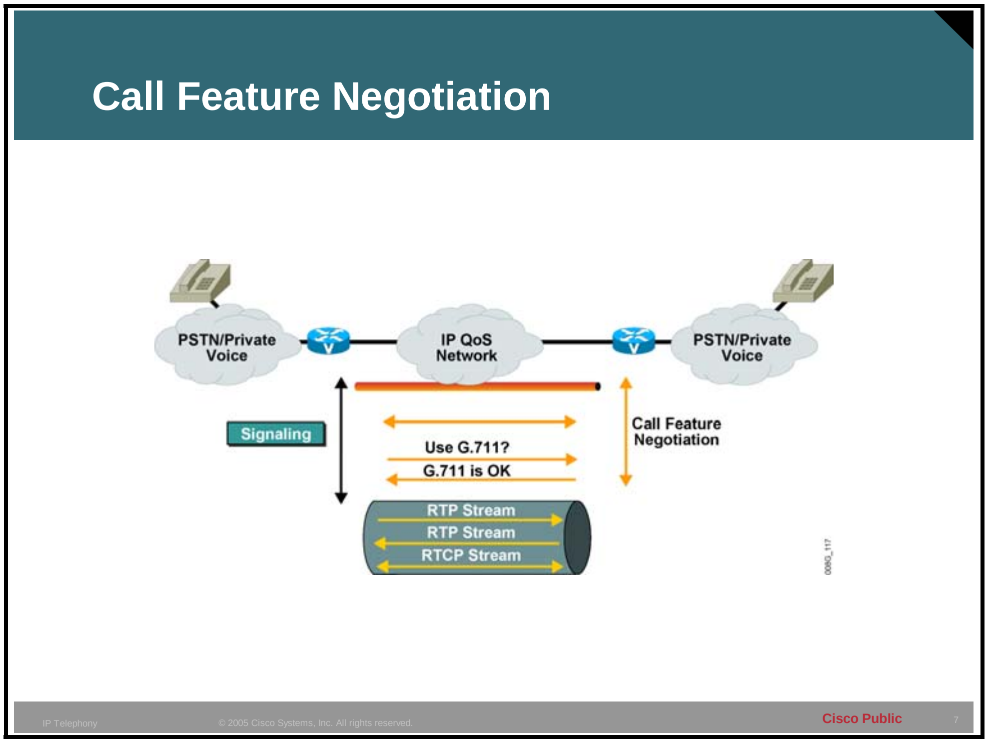
A fundamental objective of VoIP call control is to initiate communication between VoIP endpoints. This topic discusses the role of call control in establishing RTP sessions and negotiating features during the call setup procedure.



An audio path of a VoIP call leg is dependent on the creation of RTP sessions. These RTP sessions transport voice unidirectionally, so that bidirectional voice uses two RTP sessions. (In principle, if voice is needed in one direction only, as in the case of a recorded announcement or voice mail, only one RTP session is required.) The figure shows RTP sessions being created during call setup.

To create RTP sessions, each endpoint must recognize the IP address and User Datagram Protocol (UDP) port number of its peer. In a limited implementation of VoIP, these values are preprogrammed. However, to be truly scalable, the addresses and port numbers must be recognized dynamically and on demand.

During call setup, call control procedures exchange the IP address and UDP port numbers for the RTP sessions.



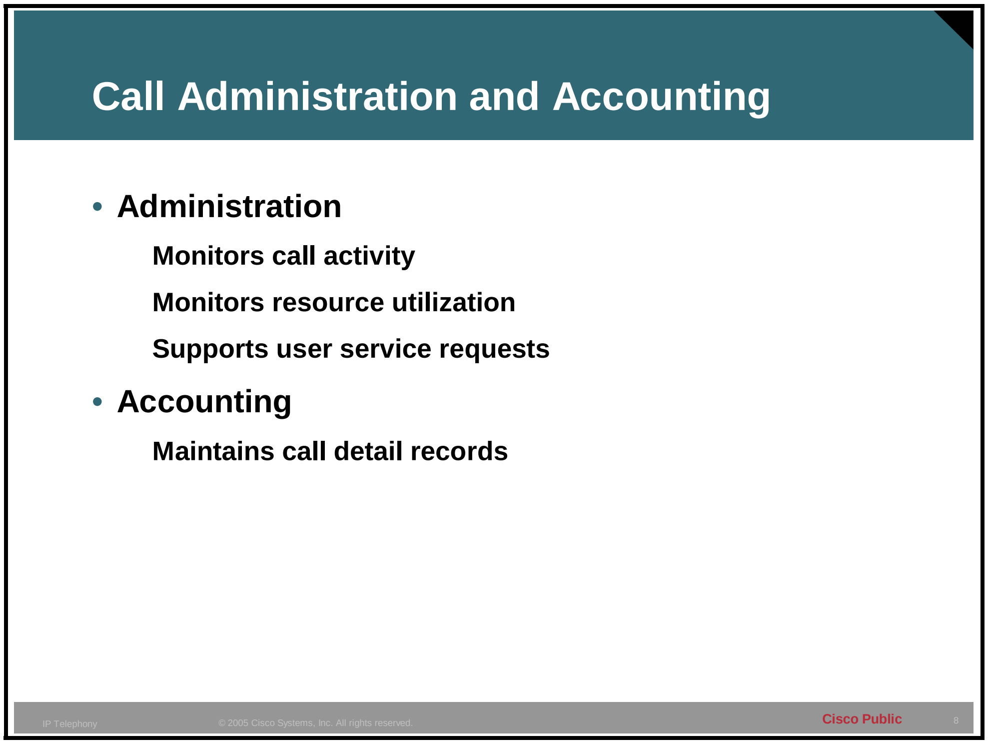
Creating the RTP sessions is not the only task of call control during call setup. The endpoints need to establish a bilateral agreement in which the communicating parties discover acceptable call parameters and then agree on the operating parameters of the call. When agreement is not possible, the call is not completed and is dropped.

Following are some examples of call parameters:

- **Coder-decoder (codec):** Each endpoint must share a common format for the voice, or at least must recognize the opposite endpoint choice for voice encoding. This is an example of a mandatory agreement. Not finding a common format is analogous to calling a foreign land and discovering that you are unable to carry on a conversation because the other party speaks a different language.
- **Receive/transmit:** Based on the application, the voice is one-way or two-way. Some endpoints do not meet the requirement for the session because they are designed to handle receive-only or transmit-only traffic when the call requests two-way communication.
- **Multipoint conferences:** The types of conferences and parameters to join.
- **Media type:** Audio, video, or data.
- **Bit rate:** Throughput requirements.

Call Administration and Accounting

Call control procedures typically provide support for call administration and accounting. This topic discusses administration and accounting capabilities of call control.



Call administration and accounting functions provide optional services for the improved operation, administration, and maintenance of a VoIP environment.

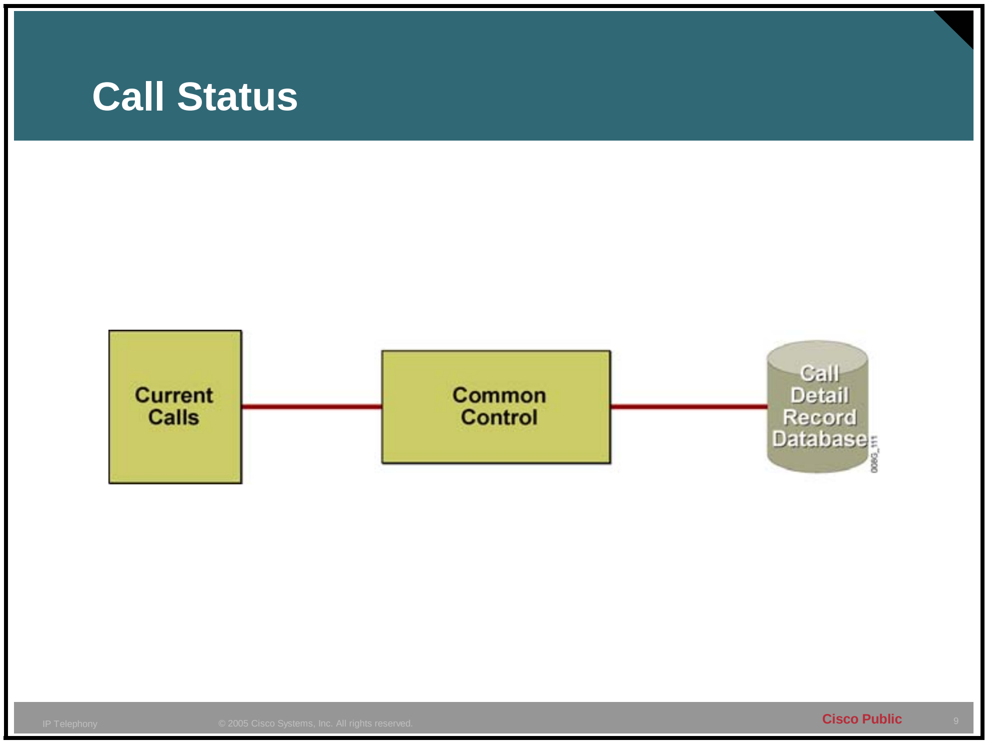
Accounting makes use of the historical information that is usually formatted as call detail records. Call detail records are useful for cost allocation, and for determining call distribution and service grade for capacity-planning purposes.

Call administration includes the following capabilities:

- **Call status:** Monitoring calls in real time
- **Address management:** Supporting users with services such as address resolution
- **Admission control:** Ensuring that resources are being used effectively

Call Status and Call Detail Records

Several call control protocols offer dynamic access to the status of calls within the VoIP network. This topic discusses the benefits of maintaining call status information and describes where and how call status is used.



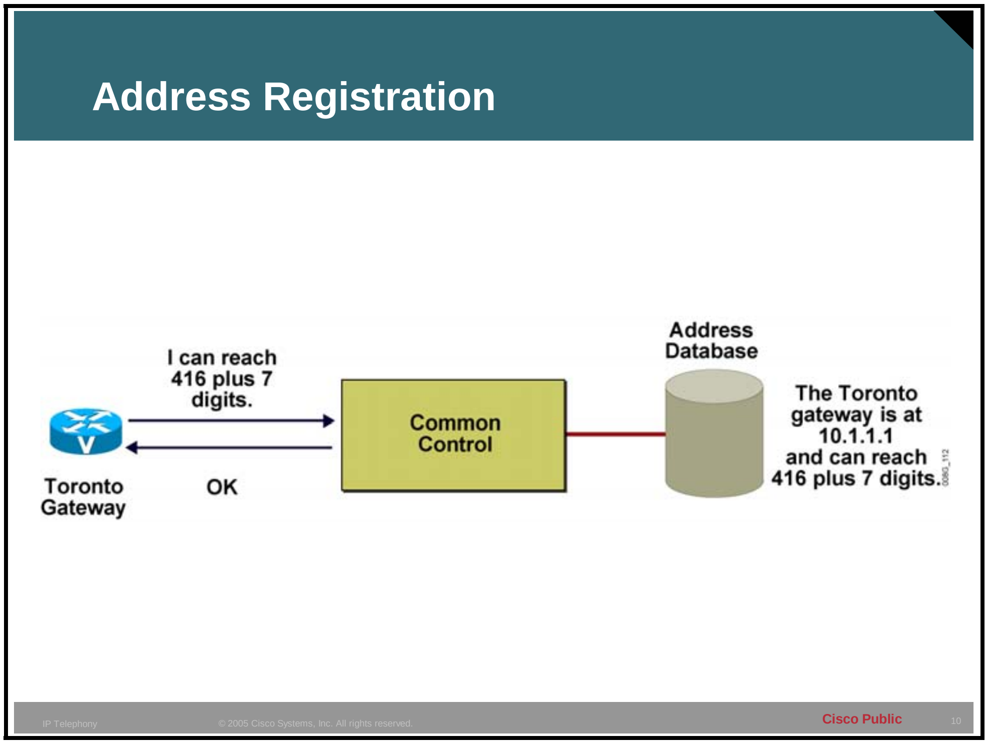
Several of the responsibilities that are assigned to call administration and accounting are dependent on access to current call status information or records of changes in the call status. Call status has both historical and instantaneous (real-time) benefits. Call detail records have consequential benefits in terms of distributing costs and planning capacity.

Call status provides an instantaneous view of the calls that are in progress. This view assists other processes (for example, bandwidth management) or assists an administrator with troubleshooting or user support.

Call detail records include information about a call start time, duration, origin, destination, and other statistics that may be useful for a variety of purposes. This data is collected as a function of call status.

Address Management

As an aspect of call administration, call control maintains a database of endpoints and their identifiers. This topic discusses how endpoints register their addresses and how these addresses are resolved to IP addresses.

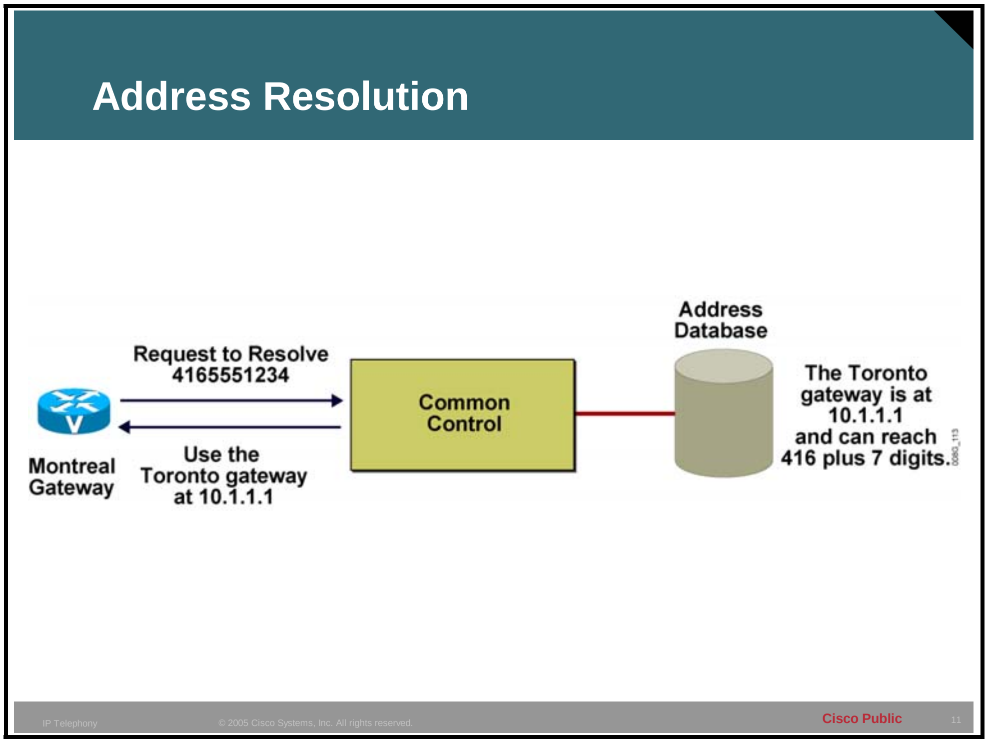


When an endpoint registers with a call control component, it supplies its telephone address or addresses and, if it is a gateway, the addresses of the destinations it can reach. The endpoint provides other information relating to its capabilities. Multiple destinations that are reachable through a gateway are usually represented by a prefix. The use of a prefix allows call control to create a database that associates a telephony-type address, for example, with its corresponding IP address.

In the traditional telephone network, the address of a station is limited to the keys available on a dual-tone multifrequency (DTMF) keypad. In VoIP, an address takes on one of several other formats as well; for example, the address can be a host name or a URL.

Example: Address Registration

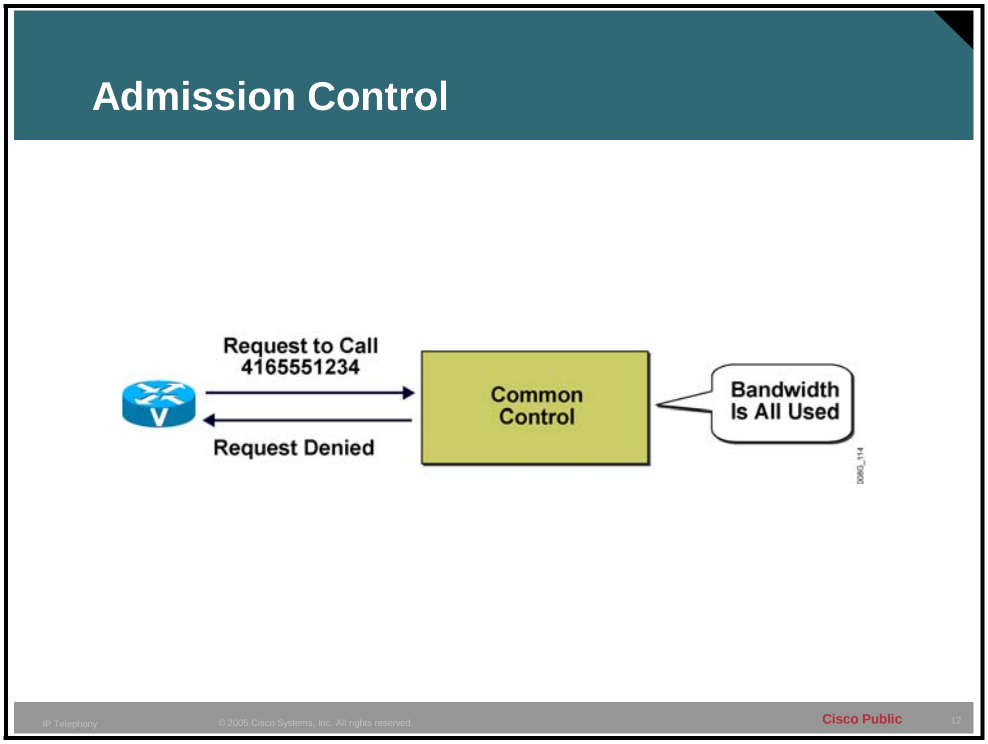
The figure illustrates the Toronto gateway registering its accessibility information. In this example, the gateway informs the common control component that it can reach all telephone numbers in the 416 area. This information is deposited into a database for future reference.



After an address is registered, it can be discovered through address resolution. Address resolution translates a multimedia user address to an IP address so that the endpoints can communicate with each other to establish a control relationship and create an audio path.

Admission Control

This topic discusses how common control and bandwidth management restrict access to the network.



Admission control has at least two aspects: authorization and bandwidth management.

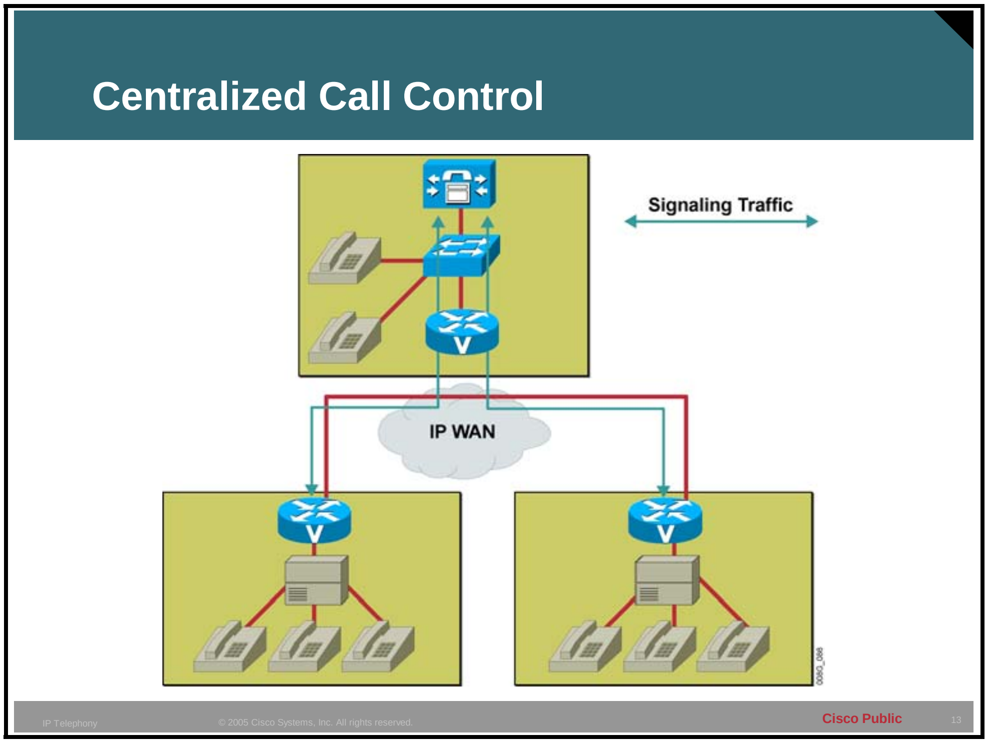
Access to a network should not imply permission to use the resources of the network. Common control limits access to the resources by checking the intentions and credentials of users before authorizing them to proceed.

Bandwidth is a finite resource. Appropriate bandwidth management is essential to maintaining voice quality. Allowing too many voice calls over an IP internetwork results in loss of quality for both new and existing voice calls.

To avoid degrading voice quality, a call control model establishes a bandwidth budget. By using data available from call status, the bandwidth management and call admission control functions monitor current bandwidth consumption. Calls may proceed up to the budgeted level, but are refused when the budget has reached its limit. This process is illustrated in the figure.

Centralized Call Control

This topic describes the function of endpoints in a centralized call control model.



The level of "intelligence" associated with an endpoint classifies call control models. In centralized call control, the intelligence in the network is associated with one or several controllers. Endpoints provide physical interconnection to the telephone network. However, endpoints still require the central controller to dictate when and how to use their interconnect capability. For example, the central controller may require the endpoint to provide dial tone to a telephone and alert the controller when a telephone goes off hook. The endpoint does not make decisions autonomously.

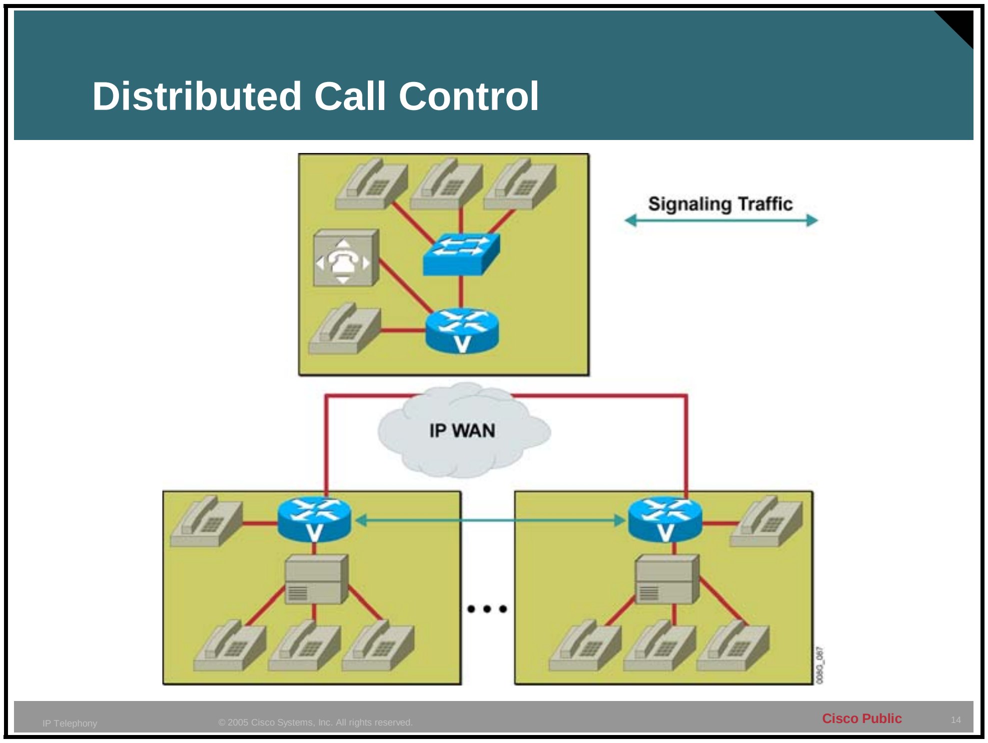
Example: Centralized Call Control

The figure illustrates two gateways interacting with a centralized call control component for call setup and management. Notice that the media path is between the two gateways.

A centralized call control model is similar to the mainframe model of computing and uses centralized common control concepts that are found in circuit-switching solutions.

Distributed Call Control

This topic describes the functions of the endpoints in a distributed call control model.



Endpoints in a distributed call control model have sufficient intelligence to autonomously manage the telephone interface and initiate VoIP call setup. Although the endpoint in a distributed model often depends on shared common control components for scalability, it possesses enough intelligence to operate independently from the common control components. If a common control component fails, an endpoint can use its own capabilities and resources to make routing or rerouting decisions.

Example: Distributed Call Control

In the figure, the two gateways have established a call control path and a media path between themselves, without the assistance of a centralized call control component.

A distributed call model is analogous to a desktop computing model.

Centralized Call Control vs. Distributed Call Control

This topic compares the advantages and disadvantages of the centralized and distributed call control models.

Centralized Call Control vs. Distributed Call Control

Centralized Call Control	Distributed Call Control
Centralized administration	Distributed administration
Ease of dial plan consistency and updating	Dial plan consistency and updating is more difficult
Supplementary services (PBX features)	Supplementary services harder to implement
Difficult to scale; all new features and applications must be implemented on the central controller, central breakpoint, or bottleneck	Scalable: need more applications functions or performance. Add more servers and they can be located anywhere
Difficult to provide resiliency over network failures	Resilient over network failures
Difficult to add new endpoints and applications; elements are tightly associated	Conceptually easy to add new endpoints and applications
WAN inefficient	WAN efficient
Dial delay bears no relation to distance between endpoints	Dial delay roughly proportional to the distance between the endpoints, as in the public switched telephone network (PSTN)
Static endpoint capabilities	Flexible: negotiation of endpoint capabilities per session

The figure shows a comparison of the centralized and distributed call control models. Both models have advantages and disadvantages. Features that are considered advantages of one type of call control model are disadvantages of the other type. The main differences between the two models are in the following areas:

- **Configuration:** The centralized call control model provides superior control of the configuration and maintenance of the dial plan and endpoint database. It simplifies the introduction of new features and supplementary services and provides a convenient location for the collection and dissemination of call detail records. The distributed model requires distributed administration of the configuration and management of endpoints, thus complicating the administration of a dial plan. Although distributed call control simplifies the deployment of additional endpoints, new features and supplementary services are difficult to implement.
- **Security:** Centralized call control requires that endpoints be known to a central authority, which avoids (or at least reduces) security concerns. The autonomy of endpoints in the distributed model elevates security concerns.
- **Reliability:** The centralized model is vulnerable because of its single point-of-failure and contention. It places high demands on the availability of the underlying data network, necessitating a fault-tolerant WAN design. The distributed call control model minimizes the dependence on shared common control components and network resources, thus reducing vulnerability.

- **Efficiency:** Centralized call control fails to take full advantage of computer-based technology that resides in the endpoints. It also consumes bandwidth through the interaction of the call agent and its endpoints. Distributed call control takes advantage of the inherent computer-based technology in endpoints.

Example: Comparing Centralized and Distributed

Because the centralized model increases vulnerability, it mandates the implementation of survivability and load management strategies that involve the replication of the central components. Few implementations of call control are totally distributed. For example, although H.323 and SIP operate in a purely distributed mode, for scalability reasons, both are most often deployed with common control components that give endpoints many of the advantages of a centralized call control environment. Unfortunately, these implementations also inherit many of the disadvantages of centralized call control.

