Network Architecture and Dependability for VoIP

Presentation at Workshop on Dependability of All-IP Networks
May 18-19, 2006 VTT, Espoo, Finland

99999999999999

it's all about nines

Veena B. Mendiratta
Consulting Member Technical Staff

Networking and Network Management Center
Bell Labs, Lucent Technologies
Naperville, IL 60566, USA
Phone: +1 630 979 3872
Email: veena@lucent.com

Lucent Technologies
Bell Labs Innovations

Outline

- Introduction/Motivation
- IMS Architecture
- VoIP Reliability in IMS
- Challenges

Telecommunications Systems for Voice Applications

High availability in telecommunications systems is achieved with architectures that implement redundancy and extensive error detection and recovery mechanisms for effecting fast recovery when failures occur with minimal impact on the application.

Real time systems where continuous operation is of paramount importance.

Goal is high system availability — system is considered available if new calls can be setup; single call setup failures or misrouted calls can be tolerated; partial downtime can be tolerated.

What is Carrier Class Reliability?

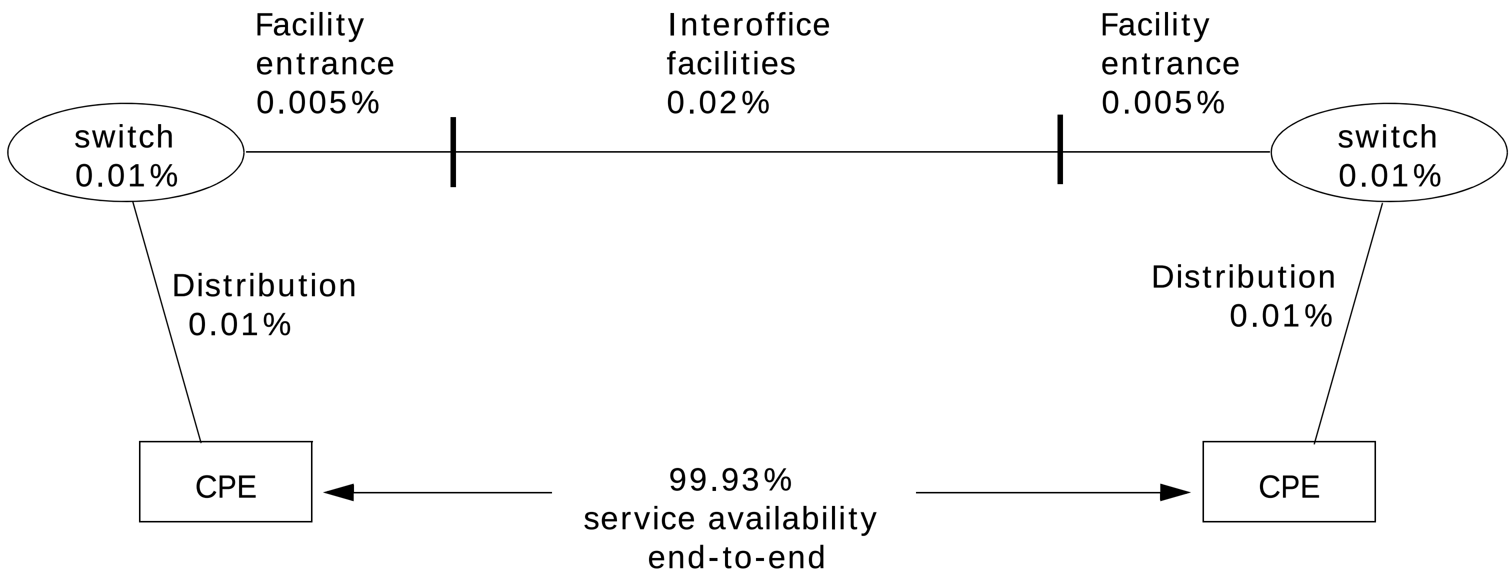
- Reliability
  — characterized by high system availability, annual switch downtime less than 3 minutes
  — characterized by low defect rate for dropped calls
  — annual subscriber downtime less than 20 minutes
  — capability to recover switch automatically from a total outage
  — no single point of system failure
  — outages in a failure group can cause some subscribers to lose service

- Redundancy: both hardware and software
  — all system impacting (system downtime) hardware should be redundant
  — no single failure should cause more than XX subscribers lines to be impacted

- Voice-Grade quality transmission
  — acceptable levels of delay, jitter, packet loss and echo
  — average delay across the PSTN is 30 to 50 milliseconds

- Scalability
  — capability to meet a wide range of performance needs
  — ability to grow subscribers
  — ability to gracefully adjust for varying traffic loads (overload control)

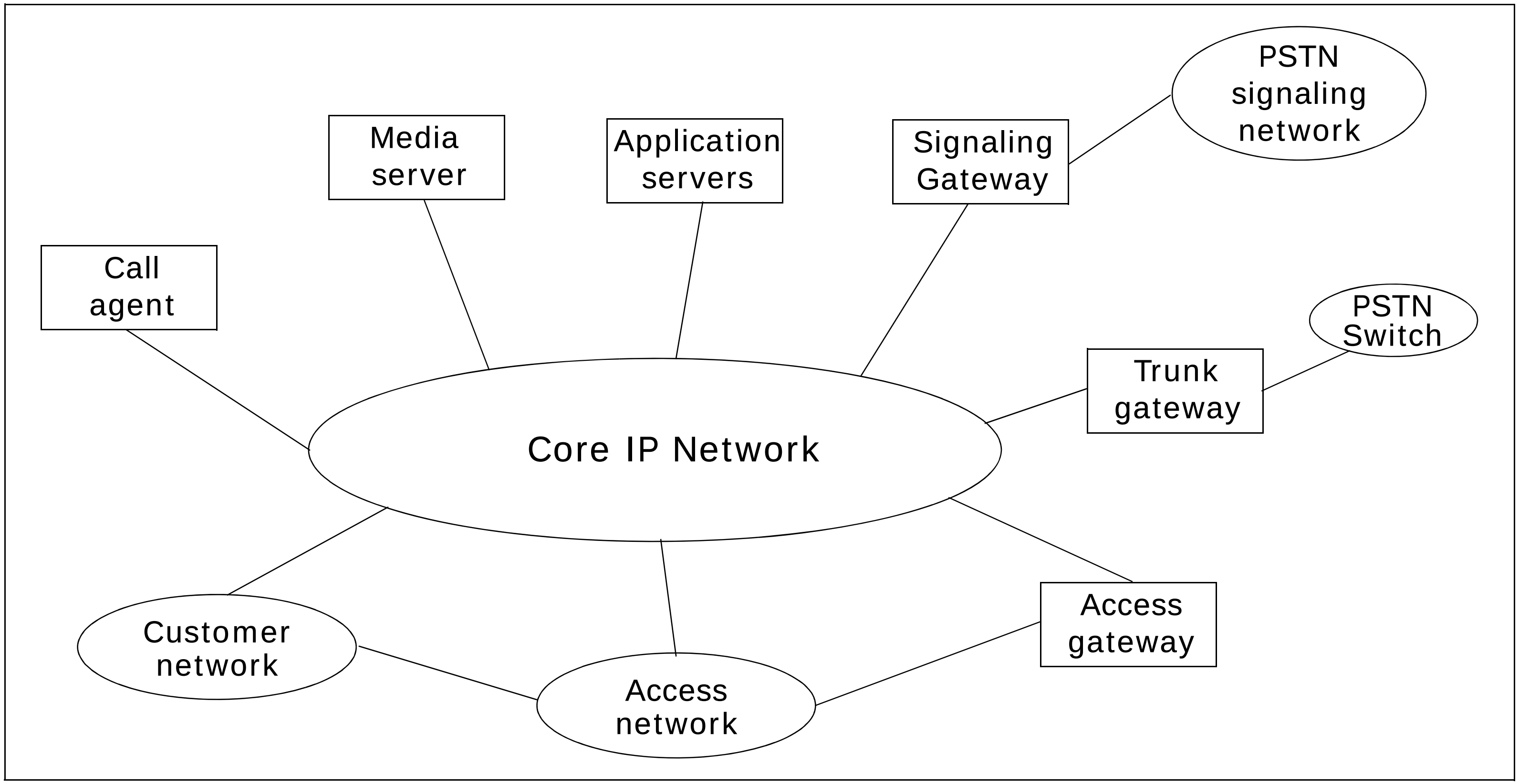
PSTN Reference Connection



Service availability: the availability of a *single path* through the PSTN between two end users.

Source: C. R. Johnson *et al*, "VoIP Reliability: A Service Provider's Perspective," IEEE Communications Magazine, July 2004.

VoIP Functional Reference Architecture (GR-929)



Note that a VoIP network distributes the functionality of a classic Central Office switch over multiple network elements.

Source: GR-929-CORE, "Reliability and Quality Measurements for Telecommunications Systems,"" Issue 8, December 2002.

VoIP Functional Reference Architecture Elements

Access network: provides connectivity between the CPE and the access gateways in the service provider network.

Access Gateway: is located in the service provider network and supports the line side interface to the core IP network for use by phones, devices and PBXs.

Trunk Gateway: supports a trunk side interface to the PSTN and/or IP routed flows in the packet network.

Signaling Gateway: provides the signaling interface between the VoIP and PSTN signaling networks.

Call Agent: located in the service provider network and provides call control and call logic functions, usually maintaining call state information for all active calls.

Application Server: located in the service provider network and provides the service logic and execution for services that are not directly hosted on the Call Agent.

Media Server: located in the service provider network and provides announcements and tones and collects user information.

Core IP Network: provides routing and transport of IP packets between the network elements.

Defining VoIP Service Reliability

- For VoIP, the system consists of: the access and backbone networks and the network elements designed to provide service between end users.

- The service function is comprised of: connection setup, transfer of user information and maintenance of the connection.

- The VoIP service is considered available (service availability) if the following functions are successful:
  connection establishment,
  user information transfer, and
  connection release.

- VoIP service reliability metrics:
  end-to-end downtime
  (equivalent to, for example, 99.93% availability for PSTN reference connection)
  defects per million (DPM)
  (includes blocked calls and cut-off calls due to failures)

IP Multimedia Subsystem (IMS)

What is IMS?

- Standardised Next Generation Networking (NGN) architecture for telecom operators for providing mobile and fixed multimedia services.
- Uses a VoIP implementation based on a 3GPP standardised implementation of SIP.
- Runs over the standard Internet Protocol (IP).
- Existing phone systems (both packet-switched and circuit-switched) are supported.

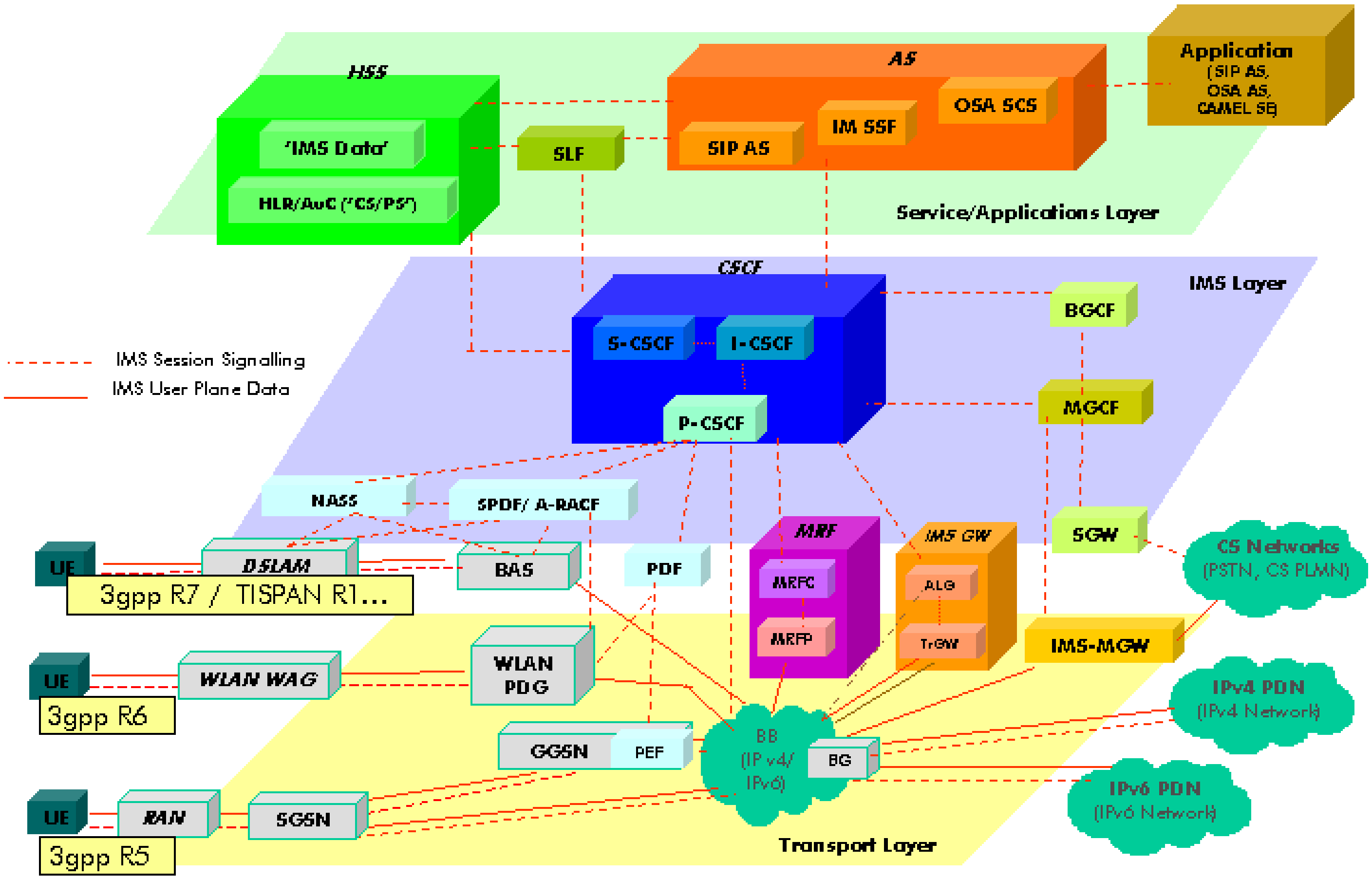
Basic Principles

- Access independence: IMS will eventually work with any network (fixed, mobile or wireless) with packet-switching functions, such as GPRS, UMTS, CDMA2000, WLAN, WiMAX, DSL, cable. Older circuit-switched phone systems (POTS, GSM) are supported through gateways. Open interfaces between control and service layers allow elements and calls/sessions from different access networks to be mixed.
- Different network architectures: IMS allows operators and service providers to use different underlying network architectures.
- Terminal and user mobility: The mobile network provides terminal mobility (roaming), while user mobility is provided by IMS and SIP.
- Extensive IP-based services: IMS should make it easier to offer IP-based services. Examples include VOIP, Push to talk over cellular (POC), multiparty gaming, videoconferencing, Messaging, community services, presence information and content sharing.

IMS Architecture

- IMS is a collection of different functions, linked by standardized interfaces. A function is not necessarily a HW node multiple functions can be implemented on a node. Each node can also be present multiple times in a network for scalability.

IMS Architecture



Source: http://en.wikipedia.org/wiki/Image%3AIms_overview.png

IMS Architecture Elements

Access Network

The user can connect to an IMS network using various methods, all of which are using the standard Internet Protocol (IP). Direct IMS terminals (mobile phones, PDAs, computers, ...), can register directly into an IMS network, even when they're roaming in another network or country (the visited network). The only requirement is that they can use IPv6 (also IPv4 in 'Early IMS') and are running SIP User Agents.
Fixed access (e.g., DSL, cable modems, Ethernet, ...), mobile access (W-CDMA, CDMA2000, GSM, GPRS, ...) and wireless access (WLAN, WiMAX, ...) are all supported. Other phone systems like the POTS, H.323 and non IMS-compatible VoIP systems are supported through gateways.

User Database (Core Network)

The HSS (Home Subscriber Server) is the master user database that supports IMS network entities that are actually handling the calls/sessions. It contains subscription-related information, performs authentication and authorization of the user, and can provide information about the physical location of the user.

Call / Session Control (Core Network)

Several roles of SIP servers or proxies, collectively called **CSCF** (Call Server Control Function), are used to process SIP signalling packets in the IMS.

A **P-CSCF** (Proxy-CSCF) is a SIP proxy that is the first point of contact for the IMS terminal. It can be located either in the visited network (in full IMS networks) or in the home network (when the visited network is not IMS compliant). Some networks might use a Session Border Controller for this function. The terminal will discover its P-CSCF with either DHCP, or it is assigned in the PDP Context (in GPRS).

An **I-CSCF** (Interrogating-CSCF) is a SIP proxy located at the edge of an administrative domain. Its IP address is published in the DNS of the domain (using NAPTR and SRV type of DNS records), so that remote servers (e.g., a P-CSCF in a visited domain or a S-CSCF in a foreign domain) can find it, and use it as an entry point for all SIP packets to this domain. The I-CSCF queries the HSS using the DIAMETER Cx and Dx interfaces to retrieve the user location, and then route the SIP request to its assigned S-CSCF.

A **S-CSCF** (Serving-CSCF) is the central node of the signaling plane. It is a SIP server, but performs session control as well. It is always located in the home network and it uses DIAMETER Cx and Dx interfaces to the HSS to download and upload user profiles.

IMS Architecture Elements (cont'd)

Application Servers

Application servers (AS) host and execute services, and interface with the S-CSCF using SIP.
Examples of services are:

- Caller ID related services (CLIP, CLIR, ...)
- Call waiting, Call holding, Push to talk, Call forwarding, Call transfer
- Call blocking services, Malicious Caller Identification
- Lawful interception
- Announcement and Conference call services
- Voicemail, Text-to-speech, Speech-to-text
- Location based services
- SMS, MMS
- Presence information, Instant messaging

Depending on the actual service, the AS can operate in SIP proxy mode, SIP UA (user agent) mode or SIP B2BUA (back-to-back user agent) mode. An AS can be located in the home network or in an external third-party network. If located in the home network, it can query the HSS with the DIAMETER Sh interface (for SIP-AS and OSA-SCS) or the MAP interface (for IM-SSF).

Media Servers

An MRF (Media Resource Function) provides a source of media in the home network. It is used for playing announcements (audio/video), multimedia conferencing (e.g. mixing of audio streams), text-to-speech conversion (TTS) and speech recognition, and realtime transcoding of multimedia data.

Breakout Gateway

A BGCF (Breakout Gateway Control Function) is a SIP server that includes routing functionality based on telephone numbers. It is only used when calling from the IMS to a phone in a circuit switched network.

PSTN Gateways

A PSTN/CS gateway interfaces with PSTN circuit switched (CS) networks. For signalling, CS networks use ISUP (or BICC) over MTP, while IMS uses SIP over IP. For media, CS networks use PCM and IMS uses RTP.

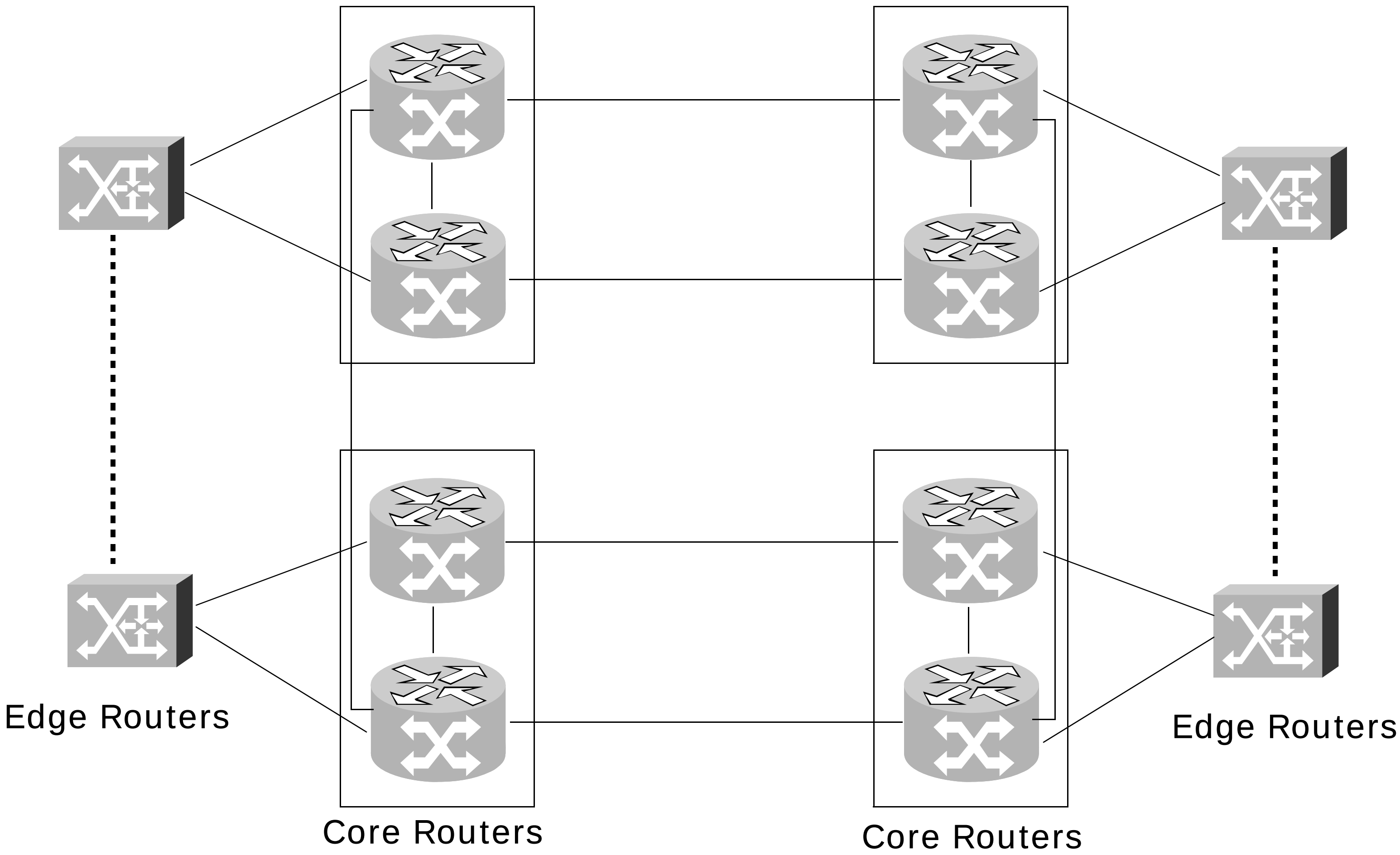
Source: http://en.wikipedia.org/wiki/IP_Multimedia_Subsystem#Architecture

Achieving Reliability for VoIP Services in IMS

- Core IP network
  Typically, backbone IP routers do not have carrier-class reliability (0.99999 availability). Reliability in the core network is achieved through redundancy of core routers to provide multiple paths through the network, dual homing of edge routers to the core routers, and use of protocols such as MPLS for fast path restoration.

- Session control (IMS) layer and Application service layer
  Reliability is achieved by:
  — redundancy of hardware/software within each network element
  — automatic failover or reconfiguration (active/standby or active/active)
  — preservation of stable calls through failover or reconfiguration
  — flexibility in component associations
  — flexibility in assignment of a particular component, e.g.,
    upon registration: P-CSCF
    during a call: MGCF
  — centralized high availability data servers, data separated from applications
  — co-location of IMS servers

Example of a Redundant Core IP Network



System Selection During Registration and Call Setup

- P-CSCF Discovery (VoIP Terminal)
  UE discovers P-CSCF via DNS or DHCP.
  A list of P-CSCF addresses is provided to UE.
  UE will try to request P-CSCF one at a time until one is available.
  If primary DNS server fails to respond, UE queries discovery to the next DNS server.

- I-CSCF Discovery
  I-CSCF discovery by P-CSCF is based on what data is populated in internal IMS DNS.
  When P-CSCF queries with home domain name it can get back results in either network or location.

- S-CSCF Discovery
  For initial registration, HSS will return list of capabilities.
  I-CSCF will have provisioned data to specify what capabilities belong to each S-CSCF, which can be in same or different network/location.
  I-CSCF looks for a match of capabilities and then uses internal IMS DNS with the FQDN to choose a particular S-CSCF IP address.

- AS selection
  S-CSCF contacts AS as directed by filter criteria.
  The AS name provided in iFC is used for DNS query to get IP address of particular AS.

- HSS query
  I-CSCF, S-CSCF or AS selects HSS's Call Function (CF) in load balanced manner.
  CF can select any one of the Data Functions (DFs).

- MGCF/MGW
  MGCF manages multiple MGW at the same or different location.
  If one MGW fails, MGCF can select the one at the other location.

Improving Reliability and Performance by Co-Location of IMS Servers

- Many different SIP Proxies are used for an end-to-end IMS VoIP call.
  - Originating and terminating P-CSCF and S-CSCF, terminating I-CSCF, and application servers
  - This introduces delays (end-to-end signaling delay)
  - This impacts capacity in terms of number of calls handled
  - This impacts reliability as several network elements need to be functioning for successful call completion

- Performance and reliability can be improved by co-location of IMS servers.
  - Magnitude of improvements depends on the percentage of roaming subscribers and the percentage of UEs calling a UE in another IMS network

- Factors determining whether IMS servers can be co-located for the same call:
  - Geographical position of the caller — roaming or not roaming
  - Geographical positions of the originating Application Servers — same as S-CSCF or not
  - Called home network — same as the caller or not
  - Geographical position of the terminating Application Servers — same as S-CSCF or not
  - Geographical position of the called party — roaming or not roaming

Source: T. Bessis, "Improving Performance and Reliability of an IMS Network by Co-Locating IMS Servers," Bell Labs Technical Journal, Vol 10, No. 4, Winter 2006.

Challenges — Building Dependable Architectures for VoIP

- Meeting traditional availability objectives for voice with a large number of elements in the end-to-end path
- Cost of providing reliability in the network
  — the network needs to have sufficient spare capacity to deal with router and link failures, it is not sufficient to have alternate paths available
- Complexity of network
  — not straightforward to determine which network element should be more reliable
- Quality of Service issues
  — jitter, delay, bandwidth, overload
  — these issues are exacerbated when failures occur and the capacity of the remaining working elements is insufficient to meet the offered traffic load

References

[1] C. R. Johnson *et al*, "VoIP Reliability: A Service Provider's Perspective," *IEEE Communications Magazine*, July 2004.

[2] GR-929-CORE, "Reliability and Quality Measurements for Telecommunications Systems,"" Issue 8, December 2002.

[3] http://en.wikipedia.org/wiki/IP_Multimedia_Subsystem#Architecture.

[4] T. Bessis, "Improving Performance and Reliability of an IMS Network by Co-Locating IMS Servers," *Bell Labs Technical Journal*, Vol 10, No. 4, Winter 2006.