北京邮电大学

博士学位论文

NGN移动环境中多媒体业务支持及移动性管理相关问题研究

姓名：徐鹏

申请学位级别：博士

专业：计算机应用技术

指导教师：廖建新

20060201

NGN移动环境中多媒体业务支持及移动性管理相关问题研究

摘 要

下一代网络（NGN，Next Generation Network）的提出最初是为了利用分组交换的优势，改造传统电信固定网络，但是随着技术的发展和人们对未来电信网架构、技术发展趋势和电信业务需求认识的深入，人们对 NGN 的理解也发生了相应的变化。NGN 应该是融合的网络，这个融合不但指电信网、计算机网和有线电视网的融合，而且包括电信固定网络和移动网络的融合，而且后者的重要性日益突出。ITU-T 对 NGN 的定义包含以下内容：能够使用户自由接入不同的业务提供商；能够支持通用移动性，从而向用户提供一致的和无处不在的业务。从中可以看出，固定和移动只是接入方式的不同，都是 NGN 整体架构的一部分。ITU-T 关于 NGN 的定义中明确说明 NGN 支持通用移动性，所以说，移动性是 NGN 中非常重要的部分。NGN 对移动性的支持构成 NGN 移动环境。

NGN 产生和发展背后的推动力量，是人们对以多媒体为主要特征的新业务的需求和运营商对新业务可能带来的高利润的期待。NGN 移动环境中支持多媒体业务需要很多专用设备，媒体服务器是一种在软交换机控制下支持多媒体业务，提供各种媒体资源服务的专用设备。论文的第一部分针对本校网络与交换技术国家重点实验室网络智能研究中心研发的一种基于软交换的媒体服务器相关研究内容进行论述。

移动性管理是 NGN 移动环境中非常重要的内容，NGN 中的移动性管理有着与传统移动网中移动性管理不同的特点，本文后一部分针对 NGN 移动环境中移动性管理的一些相关问题进行了研究。

论文主要创新工作总结如下：

(1) 通过对 NGN 移动环境中多媒体业务特点的研究，以及对各种类似设备系统架构的比较研究，结合中国通信标准化协会提出的《基于软交换的媒体服务器技术要求》所定义的设备规范，提出了基于软交换的集群媒体服务器系统结构。基于一种商业化 SIP（Session Initiation Protocol）协议栈，设计实现了一种可靠、灵活、易扩展的用 SIP 及其扩展控制多媒体会话的方案。根据对 NGN 移动环境中多媒体业务支持的研究，以及《基于软交换的媒体服务器技术要求》的相关要求，详细分析了基于软交换的集群媒体服务器的网管需求，基于简单网络管理协议版本 2（Simple Network Management Protocol, SNMPv2），设计实现了基于软交换的集群媒体服务器网管子系统，并结合 IETF 相关标准及设备自身特点定义了管理信息库（MIB ,Management Information Base）。

(2) 安全问题是NGN中受到广泛关注和热烈讨论的重要问题。为了保证NGN网络和业务的安全性，NGN设备必须充分考虑安全问题。分析了NGN安全问题的产生原因、危害及解决思路，并重点讨论了SIP协议的安全，根据基于软交换的集群媒体服务器的具体情况，对设备可能面临的安全威胁进行了分析，并给出可行的解决方案。

(3) NGN移动环境中，移动性不仅仅指终端移动性，还包括个人移动性、服务移动性、会话移动性、模式移动性等多种移动性模式。相应的，移动性管理也不只是对移动终端所涉及的切换和位置更新的管理，而是要满足上述通用移动性的管理。介绍了NGN中各种移动性模式及主要实现协议，以及这些协议在满足各种移动性管理方面的优缺点。为有效实现多种移动性管理，提出了一种结合移动IP、SIP等各层移动性管理协议的，统一、有机的多层多协议移动性管理方案。

(4) 在移动IP和SIP分别实现网络层和应用层移动性管理的多层多协议移动性管理方案中，当两种协议独立进行AAA操作时，存在缺乏效率的问题。为解决该问题，提出优化方案——"移动IP与SIP集成应用中优化的AAA过程"(OAPIMS)。在新一代AAA协议Diameter环境下，通过移动注册时，对两种协议的操作信令进行优化，减少了信令交互次数，达到提高效率的目的。分析表明，该方法可以明显降低信令开销，减少时延，提高系统性能。

(5) 移动终端通过GPRS (General Packet Radio Service)接入3GPP提出的IP多媒体子系统(IMS，IP Multimedia Core Network Subsystem)时，GPRS和IMS都会对移动终端进行认证和密钥分配，操作过程类似，独立进行两次非常缺乏效率。为解决该问题，结合相关研究成果，提出一种优化集成方法，只进行GPRS的AKA协议操作，IMS则基于GPRS的认证结果实现对移动终端的认证，同时满足了移动终端对IMS网络的认证需求及双方的密钥分配需求。对该方法的可行性及性能进行了分析。

本论文受以下基金项目联合资助：国家杰出青年科学基金（No. 60525110）；高等学校博士学科点专项科研基金资助课题（No. 20030013006）；国家移动通信产品研究开发专项基金项目（下一代移动智能网络的开发及应用）；电子信息产业发展基金重点项目（下一代网络核心业务平台）；电子信息产业发展基金项目（移动通信增值服务平台及应用系统）。

关键词：NGN，移动环境，媒体服务器，SIP，网管，安全，移动性管理，IP多媒体子系统，认证与密钥分配

STUDY ON MULTIMEDIA SERVICES SUPPORT AND MOBILITY MANAGEMENT RELATED PROBLEMS IN NGN MOBILITY ENVIRONMENT

ABSTRACT

When NGN (Next Generation Network) first brought forward, people mainly wanted to rebuild the telecommunications network using benefits of packet switching. With the development of technology and people's in-depth study of the future telecommunication network architecture, the technology development trend and the telecommunication service requirements, the understanding of NGN improved. NGN is a convergent network, not only the convergence of telecommunications network, computer network and cable television network, but also telecommunication fixed network and mobility network, the importance of which now becomes more and more important.The ITU-T definition of NGN includes: it offers unrestricted access by users to different service providers, and it supports generalized mobility which will allow consistent and ubiquitous provision of services to users. We can see from this that different networks are only different from access ways, but they are all parts of the whole architecture of NGN. The ITU-T definition about NGN indicates clearly that it supports general mobility. So we can say mobility support is a very important part of NGN. The mobility support of NGN makes of NGN mobility environment.

The impulses behind NGN are people's requirements of fresh services charactered with multimedia and carriers' thirst for high profits. To provide multimedia services in NGN mobility environment, many special types of equipment are needed. The media server is a kind of special equipment that is controlled by the softswitch to support multimedia services, providing all kinds of media resources. The first part of the dissertation discusses some relative contents about a softswitch-based media server developed by network intelligence research centre of the state key lab of this university.

The main contributions of the work presented in the dissertation are:

(1) By studying the characteristics of multimedia services in NGN mobility environment, comparing architectures of several kinds of similar equipment, according to the *Technical requirements for media server based on softswitch* proposed by CCSA (China Communications Standards Association), proposed the architecture of the Softswitch-based Clustered Multimedia Server. Based on a commercial SIP(Session Initiation Protocol) protocol stack, designed and realized a reliable, flexible, and extensible scheme of using SIP and its extensions to control multimedia session and analyzes the merits and demerits.By studying the characteristics of multimedia services in NGN mobility environment and according to the *Technical requirements for media server based on sofiswitch*, analyzed the network requirements of the softswitch-based clustered multimedia server. Designed and realized the network management subsystem of the softswitch-based clustered multimedia server based on SNMPv2(Simple Network Management Protocol). Defined MIB(Management Information Base) according to relative IETF specifications and the characteristics of the equipment.

(2) Security is an important problem acquiring much attention and arguments in NGN. To make the NGN and NGN services securely, equipment run in NGN environment should take much care of security problem. Analyzed the coming reason, harmfulness and solution way of NGN security issues, and discussed security of SIP protocol specially. Analyzed security threats of the softswitch-based clustered multimedia server and gave a feasible solution.

(3) In NGN mobility environment, mobility is not only terminal mobility any more, but also mobility modes including personal mobility, service mobility, session mobility and mode mobility et al. Correspondingly, mobility management is also not only management of handoff and location updating of terminal while moving, but also management supporting all aforesaid mobility modes. Gave descriptions of kinds of mobility mode and the relative mobility management protocols. Discussed merits and demerits of the protocols. Raised the point that to provide efficient management of all

kinds of mobility mode, it needs to combine protocols in different layer like Mobile IP and SIP to form a uniform and systematic scheme. Presented such a scheme.

(4) In a multi-protocol, multi-layer mobility management scheme, Mobile IP and SIP were used as mobility management protocol in network layer and application layer, respectively. It was very inefficient when performing AAA operations in Mobile IP and SIP independently. To solve this problem, proposed an optimized scheme, named Optimized AAA Procedure of Integration of Mobile IP and SIP (OAPIMS). Under the new AAA protocol, DIAMETER, it optimized signaling procedure of the two protocols while performing mobility registration. So it reduced time of round trips of signaling exchange and improved efficiency. Analysis indicates that this scheme can reduce signaling cost, time delay and improve system performance significantly.

(5) When mobile node (MN) accesses into IMS (IP Multimedia Core Network Subsystem) through GPRS (General Packet Radio Service), the MNshould perform authentication and key distribution operations in protocol AKA (Authentication and Key Agreement) at both GPRS and IMS level. They are very similar, so it is very inefficient to perform twice independently. To solve the problem, an optimized integrated scheme based on some previous work is proposed, in which only GPRS AKA operation is performed and IMS can authenticate MN using the result of GPRS authentication. At the same time, the scheme can meet the requirements of MN authenticating IMS and key distribution. Feasibility and performance is analyzed.

This work is jointly supported by: (1) National Science Fund for Distinguished Young Scholars (No. 60525110); (2) Specialized Research Fund for the Doctoral Program of Higher Education (No. 20030013006); (3) National Specialized R&D Project for the Product of Mobile Communications (Development and Application of Next Generation Mobile Intelligent Network); (4) Development Fund Key Project for Electronic and Information Industry (Core Service Platform for Next Generation Network); (5) Development Fund Project for Electronic and Information Industry (Value-added Service Platform and Application System for Mobile Communications).

KEY WORDS: NGN, MOBILITY ENVIRONMENT, MEDIA SERVER, SIP, NETWORK MANAGEMENT, MOBILITY MANAGEMTN, IMS, AKA

独创性（或创新性）声明

本人声明所呈交的论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京邮电大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

本人签名：_____ 日期：2006.3.31

关于论文使用授权的说明

学位论文作者完全了解北京邮电大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属北京邮电大学。学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。（保密的学位论文在解密后遵守此规定）

保密论文注释：本学位论文属于保密在__年解密后适用本授权书。非保密论文注释：本学位论文不属于保密范围，适用本授权书。

本人签名：_____ 日期：2006.3.31

导师签名：_____ 日期：2006.3.31

第一章. NGN 与 NGN 移动环境

1.1 NGN 的出现

发明于 19 世纪，作为现代文明重要标志的电话，其技术发展已经超过百年，其间经历了人工交换、步进制自动交换、纵横制自动交换，到后来全面数字化、程控化。一直以来这个领域都因为其重要性，受到各个国家的普遍重视。

很长时间里，通信网几乎就是电话网，语音服务占据了通信服务的绝大部分。我国在改革开放伊始，就把通信的发展确定为优先发展的领域，在十几年的时间里，通信产业一直以远远高于国民经济发展速度的速度在发展。国家投入了数千亿元资金，给与电信企业特殊的优惠政策，利用各种最新的通信技术建立了通信基础设施，我们利用后发优势，很快赶了上来。中国的电话用户快速增长，电话在短短的几年里，进入了千家万户。

与此同时，另一个网络也在悄悄的发展着，诞生在 20 世纪 60 年代的因特网，最初仅仅是为了提高安全性，而将一些分散的大型计算机连接在一起。使用的 TCP/IP 协议，实用而不复杂。随着越来越多的机构和个人把自己的计算机和个人电脑连入因特网，因特网迅速扩展到整个世界，并成为人们工作、学习和生活中不可缺少的一部分。因特网是伴随着计算技术和计算机产业的大发展而发展的，尤其是 80 年代以后个人计算机的发展普及，使得越来越多的人可以利用因特网。基于这个网络平台开发了越来越多的应用和服务。人们可以利用 WWW 服务查找资料，浏览丰富多彩的内容；可以利用电子邮件更快捷的联系；可以利用各种聊天工具和认识不认识的朋友即时交谈。新的应用还在不断涌现，人与人的远程交流不再局限于传统电话，虽然电话永远都会是一种不可取代的交流手段。

1.1.1 传统电信网的局限性

传统电信网一般包括接入、交换、传输三大部分。接入是指用户如何连接到网络；交换涉及呼叫在网络中的选路；传输则是如何在网络中传播通信信息。

交换的方式有三种，分别是电路交换、报文交换和分组交换。电路交换是在通信开始之前先建立电路连接，然后一直占有这条通信线路进行通信，通信结束后释放线路；报文交换是接收整个报文后，再根据报文附带的地址信息进行转发；分组交换是把整个报文分成一定大小的分组，每个分组都带有地址信息，独立进行寻址、发送，到达目的地后，再重组成原来的报文。电路交换实现简单，而且

通话安全和服务质量非常有保障。所以一直是公众电话交换网（PSTN，Public Switched Telephone Network）所采用的交换方式。然而使用电路交换，通信使用者在整个通信过程中始终占据通信线路，即使不说话的时候，别的用户也不能使用，造成了资源的极大浪费。资源是稀缺的，这样的状况就显得很不经济。

传统电信网，最重要的设备是程控交换机，程控交换机最复杂的部分是执行呼叫处理的软件。电信业务的处理软件与执行呼叫控制的软件集成在一起，都处于程控交换机内。由于交换机软件系统是由交换机厂商自己开发，遵循的标准也大都是厂商自己的标准，与其他厂商提供的交换机系统不兼容，如果有什么新的业务要提供，就必须由交换机厂商来更改软件，不仅费时，而且费用高昂。若交换机厂商出于某种自身利益的考虑，不愿配合做出更改，则运营商就不能提供新业务。这种现象无疑是对信息时代求新、灵活、开放的巨大挑战。

同时，由于历史的原因，传统电信市场是相对封闭的市场，投资大，政府控制，后来者很难进入。这种现象一方面造成了垄断和竞争不充分，使得消费者不能从技术进步和经济发展中充分享受到实惠，不能以便宜的价格享受到丰富多彩的信息服务；另一方面，性能优越，通过巨大投资建起来的通信网络仅仅能够提供语音服务，充其量有一些应用有限的数据服务，如传真，实在是一种资源浪费。

综上，传统电信网存在着电路交换效率低下，浪费资源；专用设备升级困难，兼容性差，维护费用高，新业务引入周期长；市场准入门槛高，缺乏充分竞争等问题。所有这些与信息时代和因特网发展带给人们的信息化感受都是相矛盾的。

下一代网络（NGN，Next Generation Network）正是在这样的背景下产生的。

1.1.2 下一代网络（NGN）的提出及其含义

NGN是一个很宽泛的概念，它不是要一夜之间取代我们目前广泛使用的基于电路交换的电信基础设施；也不是要利用目前的公共因特网提供电信业务，从而让传统的电信运营商寿终正寝。它的提出，是与时俱进的结果，是信息技术发展带来的必然后果。人们对通信、信息服务的需求是无止境的，这既是信息产业的机遇，同时也是挑战。

NGN从诞生起，定义就有所争议，大家从不同的侧面对它进行描述，却从来也没有一个确定的，得到广泛认可的定义。2004年2月ITU－T SG13会议经过激烈的辩论，终于给出了一个得到较广泛认可的NGN定义：

A Next Generation Network (NGN) is a packet-based network able to provide

Telecommunication Services to users and able to make use of multiple broadband, QoS-enabled transport technologies and in which service-related functions are independent of the underlying transport-related technologies. It enables unfettered access for users to networks and to competing service providers and services of their choice. It supports generalised mobility which will allow consistent and ubiquitous provision of services to users. [1](NGN 是一个分组网络，它提供包括电信业务在内的多种业务，能够利用多种带宽和具有 QoS 能力的传送技术，实现业务功能与底层传送技术的分离；它提供用户对不同业务提供商网络的自由接入，并支持通用移动性，实现用户对业务使用的一致性和统一性。)

我们由此可以给出 NGN 所具有的一些最显著的特征：

1) 分组传输。NGN 接入网和核心网部分都将采用分组交换和传输，相对于传统的电路方式，分组交换传输大大提高资源利用率，并且可以更好地支持数据业务和应用。

2) NGN 采用开放分层的网络架构体系，控制功能按照承载、会话/呼叫、应用/业务分离。现有 PSTN 电路交换网以交换机为核心，交换机对业务的处理是黑箱操作，集中了业务接入，媒体处理，呼叫控制和业务管理的全部功能。这就带来一些负面的影响，对设备制造商来说，升级交换机比较困难；对运营商来说，一旦选择了某个制造商的设备，就被其束缚，后续的业务提供，设备升级都要依靠特定制造商，形成垄断，新业务的引进也很麻烦。NGN 采用开放分层的架构，由上到下分为接入/传输层，媒体层，呼叫控制层和业务应用层，并在它们之间采用标准的协议进行连接，见图 1.1。其特点是将传统交换机的功能模块分离为独立的网络部件，各个部件可以按相应的功能划分，各自独立发展；部件间的协议接口基于相应的标准。部件化使得原有的电信网络逐步走向开放，运营商可以根据业务需要自由组合具备各部件功能的产品来组建网络。

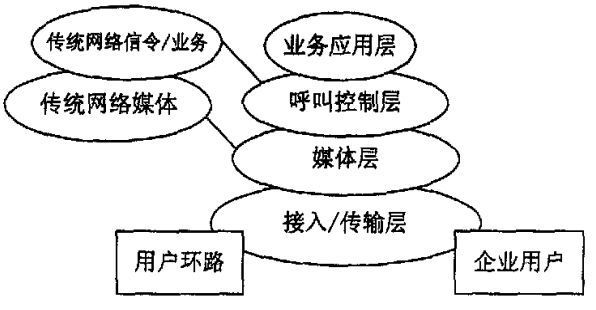

图 1.1 NGN 功能分层

3）**NGN 是业务驱动的网络。**NGN 业务与呼叫控制分离，这样分离的好处是业务可以独立于网络，使得业务的提供可以自由、灵活，业务的来源也可以多样化，不必局限于一些技术要求很高的电信设备供应商。不过电信服务设施毕竟不同于因特网，对于安全有更高的要求，如何让第三方业务提供商进入电信网络，是个需要认真考虑的问题。完全开放运营商的网络并不现实，可行的方法是开发标准的接口，通过可控的接口有限度地开放运营商网络。

1998 年 3 月，Parlay 工作组由英国电信 BT、Ulticom、美国微软（Microsoft）、加拿大北方电信（Nortel）和德国西门子（Seimens）等 5 家企业联合成立，主要研究支持外部应用访问电信网络内部资源的网络接口规范，拓展网络智能，使第三方业务供应商可以通过这样的接口，为无线、因特网或 PSTN（Public Switched Telephone Network）开发新业务。后来，欧洲电信标准协会（ETSI）和第 3 代移动通信合作伙伴计划（3GPP，3rd Generation Partnership Project）都采用 Parlay 作为 NGN 和 3G 网络业务提供的支撑技术，3GPP 据此提出了开放式业务架构（OSA，Open Services Architecture）。鉴于 Parlay API 对于网络智能实现的重要性，ITU-T 在研究智能标准的第 11 研究组（SG11）设立专题研究课题，总结已有的 API，提出 API 的选用建议。这样，第三方业务提供商就有了标准的手段接入到运营商网络，电信应用就有可能像因特网应用一样，在众多参与者的努力下，开发出各种新鲜、有益的应用来。竞争的引入，带来的好处就是业务的丰富多彩和成本的降低。

业务驱动还体现在 NGN 发展围绕如何实现业务提供多样性，方便性展开。传统电信网的业务比较有限，主要是语音业务，和有限的一些数据业务，如传真。随着计算技术的发展，尤其是互联网的普及。一方面使得人们对业务的类型有了更高的要求，尤其是对多媒体数据业务的要求大大提高，另一方面，通过电信网向人们提供丰富的多媒体业务的技术可能性也渐渐变得现实。

4）**NGN 支持多种接入手段，支持通用移动性。**20 世纪 90 年代，信息通信领域有两大亮点，一是因特网的迅速发展，二是移动通信的普及和移动技术的发展。这两大亮点带给人们生活巨大的变化，同时促进了信息技术产业的蓬勃发展。人们很自然的想到将因特网和移动通信结合起来，这样人们的通信和享受信息服务将更加自由。通过移动通信网连接到因特网，一般称为“移动互联网”技术，代表技术是 GPRS（General Packet Radio Service）；通过无线局域网技术连接因特网，一般称为“无线因特网”技术，代表技术是 IEEE 802.11。NGN 中，用户可以自由接入网络，各种移动技术最终将可以实现互联，对用户提供透明的移动接入服务。用

户不论通过何种技术接入，都可以享受到不变、方便的移动通信服务。

1.2 NGN 移动环境与多媒体业务支持

1.2.1 NGN 移动环境

移动通信的广泛应用，使得人们的信息交流更加自由和灵活，脱离了固定通信设备对人们通信场所的约束，带来了通信的革命。移动通信所占的市场份额越来越大，2004 年已经超过固定电话用户，据信息产业部统计，截至 2005 年底，全国移动用户已超过 4 亿户，大大超过固定电话用户数。这充分说明移动通信的优越性。

NGN 最初提出的时候，是从传统电信固网的角度出发的，但是随着人们对 NGN 理解的深入，以及技术和市场环境的发展，人们对 NGN 的理解也发生了相应的变化。NGN 应该是融合的网络，这是 NGN 发展的必然要求，这个融合主要指电信固定网络和移动网络的融合。ITU-T 对 NGN 的定义中提到：能够使用户自由接入不同的业务提供商；能够支持通用移动性，从而向用户提供一致的和无处不在的业务。从中可以看出，固定和移动只是接入方式的不同，都是 NGN 整体架构的一部分。ITU-T 关于 NGN 的定义中明确说明 NGN 支持通用移动性，所以说，移动性是 NGN 中非常重要的部分。

NGN 对移动性的支持构成 **NGN 移动环境**。

1.2.2 多媒体业务支持与基于软交换的媒体服务器

前面提到过，NGN 之所以产生，根本原因还是传统电信网自身存在着弊端。新技术，新概念之所以被人们关注和逐渐接受，背后都有相应的推动力量。NGN 背后的推动力量，主要是人们对新业务的需求，以及电信运营企业在面对因特网竞争时所产生的压力和对新利润来源的渴望。传统电信网可以提供的业务非常有限，主要是语音业务和有限的一些数据业务，如传真。计算机技术和通信技术的发展，尤其是因特网的普及，产生了许多新鲜应用和娱乐形式，如电脑游戏，即时聊天，视频点播，流媒体等。这些应用娱乐形式背后蕴藏着巨大商机，如何将这些事物移植到电信基础设施，并施以合适的运营模式，一定会产生巨大的经济和社会效益。

所以，如何方便地引入新业务，尤其是包含了音频、视频、彩色图像等多种形式的媒体所构成的多媒体业务，是 NGN 成败与否的关键。多媒体业务种类繁多，

是未来电信业务的发展方向。多媒体业务的出现与计算机技术、通信技术的发展紧密相关。传输多媒体信息与传统的电话语音信息相比，需要大得多的传输带宽。解决传输能力不足，可以通过计算机技术领域的压缩技术对信息进行压缩，或者通过通信技术领域新的传输技术去增加传输带宽。此外，多媒体业务对通信系统集成性、交互性、同步性等特性要求也比较高。这就要求从网络的整体架构上去考虑问题。

支持多媒体业务的 NGN 体系，国内外研究者做了大量的研究工作。国际软交换联盟（ISC）提出了基于 SIP 的 NGN 业务功能参考模型[2]，见图 1.2。PARLAY 组织、ETSI 以及 3GPP 共同建议了 OSA/PARLAY 开放业务体系结构。IETF 制定了 NGN 业务/应用层的主要协议，如 SIP，MEGACO(H.248)等。ITU-T SG13 等工作组对 NGN 的框架体系、业务需求、网络功能、互通、服务质量（QoS）、移动性管理、NGN 的演进方式等各方面提出了总体要求并明确了 NGN 业务层的基本框架[3]。尤其 3GPP R5 提出的 IP 多媒体子系统（IMS，IP multimedia subsystem）[4]，为 NGN 中多媒体业务的支持又提供了一个完整的综合平台。

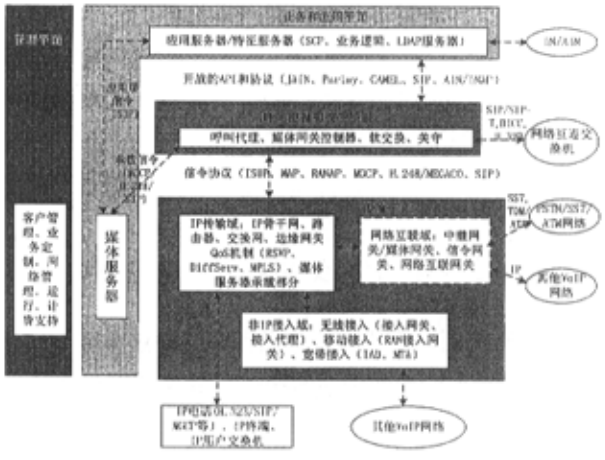

图 1.2 ISC 功能平面参考模型

为使我国在电信产业标准研究中不会再次落后，中国通信标准化协会各技术工作委员会自 1999 年开始，就密切跟踪国际上 NGN 标准的每一步进展，并积极研究制订我国的 NGN 相关标准。

结合国际标准组织研究成果，我国 NGN 标准研究工作者提出了可以支持多媒体业务的软交换系统框架，如图 1.3 所示[5]。这个框架中，接入层的功能是将包括 PSTN 用户、ISDN 用户、移动用户、互联网用户、有线电视用户等的各种用户接

入网络，并在需要时进行媒体格式转换；传输层的功能是进行快速传输；呼叫控制层的功能主要是进行会话与呼叫控制；业务层的功能是在呼叫建立的基础上提供各种业务与网络服务。各层都列出一些关键设备，如接入层的媒体网关，信令网关，以及各种接入设备；传输层的各种传输设备；控制层的媒体网关控制器（又叫软交换）。

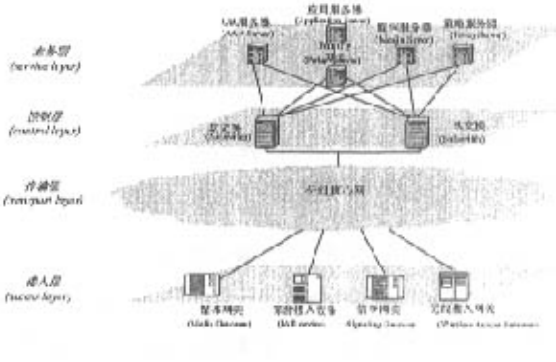

图 1.3 支持多媒体业务的 NGN 系统架构

应用层的关键设备有应用服务器、媒体服务器、AAA 服务器、Parley 网关等。应用服务器是业务层的主体，它提供各种增值业务或智能业务业务逻辑的驻留和执行环境；AAA 服务器提供用户认证（Authentication）、授权（Authorization）、记费（Accounting）服务；Parley 网关，即 Parlay 服务器，对第三方业务提供支持。

媒体服务器是 NGN 应用层中提供多媒体资源功能的关键设备，它在软交换机和/或应用服务器的控制下，为来自电路交换网或分组交换网的各种类型的多媒体业务提供分组化的高级媒体处理服务，是多媒体业务得以实现的关键。这种受软交换机控制的媒体服务器称为基于软交换的媒体服务器。

1.3 NGN 移动环境中的移动性管理

最初移动环境就是指移动通信系统——第一代模拟移动通信系统或第二代 GSM 和 IS95 CDMA 网络。随着无线局域网技术（如 IEEE 802.11）及其他无线通信网技术（如 Ad Hoc 网络技术）的出现，移动环境的概念也发生了变化。目前，NGN 研究逐步由最初定位的主要利用分组传输和软交换技术对电信固定网络进行改造和升级换代，扩展到包含固网和移动网在内的符合 NGN 分组化、开放分层架构等原则的所有通信技术。网络融和成为 NGN 领域的重要研究内容。移动环境成为 NGN 的重要组成部分，NGN 各研究机构对它的关注和研究逐渐增多。

一般来说，在移动环境中，移动性是指当用户和终端位置改变时，能够持续接入服务、继续通信的能力。移动性可划分为两个级别，一个是"游牧"移动性，指用户在移动时能改变其网络接入点，但正在进行的会话将停止，须重新启动；另一个是无缝移动，指用户或终端移动时，能随时改变其网络接入点而不中断正在进行的会话。2004 年 ITU-T SG13 给出的 NGN 定义中，明确指出 NGN 支持通用移动性。对于这个"通用移动性"，可以这样理解，用户在用不同的接入技术接入网络时，可作为同一客户；允许用户跨网络使用和管理他们的业务。用户的移动性需求包括：用户可以作为移动/游牧用户改变接入点或终端；用户可以应用各种接入技术从任何网络接入点接入网络；用户可以得到连续性服务和网络应用，这些业务和网络应用可以由网络运营商、业务提供商或第三方提供。网络根据提供业务的能力，应支持个人移动性、终端移动性以及它们的任意组合。

为支持移动性，网络要进行相应的移动性管理，这些管理包括用户鉴别（Authentication）、授权（Authorization）、计费（Accounting）、位置管理、寻呼、终端地址分配、用户信息的下载、VHE 管理等许多内容。为用户和终端提供有效的移动性管理以保证异构网络间的漫游和业务的无缝移动性是NGN最紧迫的需求之一。

NGN 移动环境中，用户可以通过各种接入方式接入网络，其获取的业务应该没有大的差别。ITU-T FGNGN 2005 年 3 月第五次会议中提出的 Y.NGN-FR《NGN 移动性功能需求》建议草案，描述了 NGN 中的移动性管理需求及其体系结构，将 NGN 的移动性管理分为网间移动性管理、网内移动性管理和接入网内移动性管理三种情况，并从通用、用户和网络三个方面提出了对移动性管理的要求，提出了 NGN 移动性管理的功能体系结构。由于解决问题的着重点不同，这三类对移动性管理的要求也不同。接入网内移动性管理由于其相同的接入技术和同一的运营者而相对简单，对网内移动性管理也同样如此。而网间移动性管理由于涉及不同的接入技术和不同的运营者而相对复杂。

此外，2G 和目前的 3G 网络中，移动性管理主要指的是移动终端的管理，它有两个含义：一是切换管理，一是位置管理。前者指的是当终端移动时，正在进行的会话不会被中断；后者则是指系统跟踪和定位移动终端，并在呼叫到来时，建立与该移动终端的呼叫。随着移动通信技术的发展和用户对移动通信业务需求的发展，人们对未来网络移动性的认识也发生了变化，移动通信不再仅仅指移动终端的移动，还包括个人移动性、服务移动性、会话移动性等多种移动性。相应的，移动性管理也不再只是对移动终端所涉及的切换和位置的管理。实现 NGN 中

所要求的通用移动性管理，必然是实现了多种移动性模式的综合的移动性管理。

1.4 国内外相关研究动态

1.4.1 国际标准组织的研究情况

NGN 内涵非常丰富，研究范围也相当广泛，主要研究内容有：新业务及应用；网络基础设施；网络体系架构；接入技术；移动性管理；IP 网络技术；网络测试技术；网络融合及互通技术；网络管理及运行维护；AAA 技术；网络协议；网络安全技术等。

对 NGN 的研究包括了计算机和电信领域的标准组织，主要的有因特网工程任务组（IETF，Internet Engineering Task Force），国际电信联盟（ITU-T，International Telecommunications Union），欧洲电信标准组织（ETSI，European Telecommunication Standards Institute），以及 3GPP，3GPP2，PARLAY 等

IETF

IETF 工作重点在于业务承载层以及业务层，NGN 中重要的标准 IPv6、SIP、MEGACO 等都是由 IETF 指定的。NGN 需要高带宽、足够的地址空间。目前 IPv6 的研究取得重要进展，IP 安全性和移动性有了很好的解决，并且都成为 IPv6 的核心部分。IPv6 设计了庞大的地址空间，保证 NGN 未来应用的广泛性和通用性。SIP 协议的研究和扩展在和其他相关研究机构的合作中也在不断取得进展。

ITU-T

ITU-T 研究 NGN 的部门主要有 SG13、SG11、SG15、SG16 和 SG19，其主要研究领域分别为：

1）SG13 领导 ITU-T 中的 NGN 研究，完成有关 NGN 体系框架、演进、融合等研究课题，具体包括框架和体系结构、NGN 信令要求、NGN 项目管理协调和计划颁布、NGN 实现方案和应用模型、网络和业务能力、互操作性、IPv6 对 NGN 的影响、NGN 移动性和固定/移动融合、公共数据网络、NGN 安全等。SG13 还在业务需求、框架、演进和 QoS 等方面进行了研究。

2）SG11 主要研究信令和信令控制架构。

3）SG15 研究 NGN 接入技术和 NGN 传输技术。在智能光网络领域，SG15 的主要工作是定义标准的自动交换光网络体系结构。

4）SG16 主要研究多媒体业务。

5）SG19 致力于研究固网和移动网的融合。

为追赶 ETSI 在 NGN 方面研究的领先，加快 NGN 研究的步伐，2004 年 6 月 ITU-T 成立了下一代网络专题组（FGNGN，Focus Group on Next Generation Network），属于 SG13 ，全面领导 ITU-T 对 NGN 的研究。FGNGN 的输出文件提交给 SG13 或他的研究组，经过讨论，通过后形成建议。FGNGN 集中研究 7 个领域的问题：业务需求、功能体系架构和移动性、IPQoS、控制和信令能力、网络安全、网络演进、 未来基于分组的网络（FPBN）。

FGNGN 一共工作了一年半，举行了 9 次会议，最后一次会议于 2005 年 11 月召开，共收到 1206 篇文稿，参加的总人数为 1166 人，可见其研究活动的活跃性。一些重要的领域都完成了版本 1 的工作。同时所有的输出文件都被收集到 ITU－T NGN BIG BOOK 中。

ETSI

ETSI 作为区域性标准组织，一直对国际电信标准制定发挥着重要、积极的影响。继成功主导制定全球 3G 标准之一后，又在 NGN 研究领域投入了巨大的力量。2003 年 9 月成立专门针对 NGN 研究的 TISPAN（Telecommunications and Internet converged Services and Protocols for Advanced Networking）研究组，它由原来从事固网标准化的 SPAN 组织和进行 VoIP 研究的 TIPHON 组织合并而成。TISPAN 分八个工作组，分别关注业务、体系、协议、号码与路由、服务质量、测试、安全和网络管理等方向。TISPAN 主席 Mr. Alain le Roux 曾说：“组建 ETSI TISPAN 的目的是在两年内开发出可执行的 NGN 规范。26 个月的努力工作后，在 TISPAN 第九次会议上，我们将完成并发布 TISPAN_ NGN 版本 1 规范。我们的行业正热切期待这个能够使 NGN 变成现实的规范。”

2005 年 12 月 5 日至 9 日，ETSI 在法国的 Sophia Antipolis 召开会议。会上来自世界各主要电信公司的超过 200 名的代表参加 TISPAN's NGN Realease 1 的官方发布仪式。此次 TISPAN 会议上提交了 20 多项规范等待批准，不久以后相同数量的规范将完成并准备进入批准程序。

TISPAN 将 NGN 架构描述为四个网络子系统，即 IP 多媒体子系统（IMS）、PSTN/ISDN 仿真子系统、基于 RSTP 的流媒体子系统和其他多媒体子系统。后来补充了网络附属子系统（NASS）和资源和接纳控制子系统（RACE）。IMS 为 NGN 接入网和终端提供基于 SIP 的业务，包括多媒体会话业务、集群信息订购等；

PSTN/ISDN 仿真子系统模拟 PSTN/ISDN 网络，使通过网关从 IP 网接入的传统终端使用 PSTN/ISDN 业务时，感觉他们是与 PSTN/ISDN 直接连接的；NASS 提供 IP 地址分配（如利用动态主机配置协议 DHCP）、IP 层认证、基于用户定制信息的网络接入鉴权、基于用户定制信息的接入网络配置、IP 层位置管理等功能；RACS 提供接纳控制和网关控制功能，接纳控制方面主要是依据用户配置信息，运营商运营策略和资源情况进行接纳控制，而网关控制则包括网络地址和端口转换、DSCP 标记等。

TISPAN 认为 IMS 代表了 NGN 网络发展的方向，所以提出基于 IMS 的体系架构应该成为 NGN 的主体架构。所以可以认为它的主要的工作就是研究如何将 3GPP 的 IMS 成果应用到固网当中，通过解决固定接入等问题，使得 IMS 成为固定和移动网络融合的业务控制层面的体系架构。

TISPAN 目前发布的 Realease 1 基于以下几个关键点：对 IMS 是适应而不是修改的原则，QoS 仅限于接入，同时支持 PSTN / ISDN 模拟和仿真业务，并且是两个分离的子系统，支持多种接入方式，包括 xDSL、WLAN、3GPP 和 LAN，主要的业务能力包括会话、呈现、即时消息、内容提供、PSTN / ISDN 演进和因特网业务。

3GPP 与 3GPP2

3GPP（3rd Generation Partnership Project），由 ETSI（欧洲电信标准）、T1（美国 T1 电信标准委员会）、ARIB/TTC（日本无线工商业联盟/电信技术委员会）、TTA（韩国电信技术联合会）、CWTS（中国无线通信标准研究组）组成。中国通信标准化协会（CCSA）成立后，CWTS 在 3GPP 的组织名称更名为 CCSA（China Communications Standards Association）。）。主要发展基于 GSM 的 3G 标准。

3GPP2 主要工作是制订以 ANSI-41 核心网为基础，cdma2000 为无线接口的移动通信技术规范。该组织于 1999 年 1 月成立，由美国 TIA、日本的 ARIB、日本的 TTC、韩国的 TTA 四个标准化组织发起，中国无线通信标准研究组（CWTS）（CCSA 成立后，CWTS 在 3GPP2 的组织名称更名为 CCSA）于 1999 年 6 月在韩国正式签字加入 3GPP2，成为这个当前主要负责第三代移动通信 cdma2000 技术的标准组织的成员。

3G 无线技术可以提供交通工具内 144kb/s，行走中 384kb/s 和室内环境下 2Mb/s 的传输速率。数据服务不能用与语音服务同样的网络技术实现，需要分组交换技术的支持，IP 在因特网中的成功，以及移动通信与因特网连接的内在需求，使得

人们很自然地选择 IP 作为移动通信数据服务的承载协议。于是，3GPP 和 3GPP2 分别定义了基于 IP 网络框架结构，可分别参考图 1.4 和图 1.5[6]。

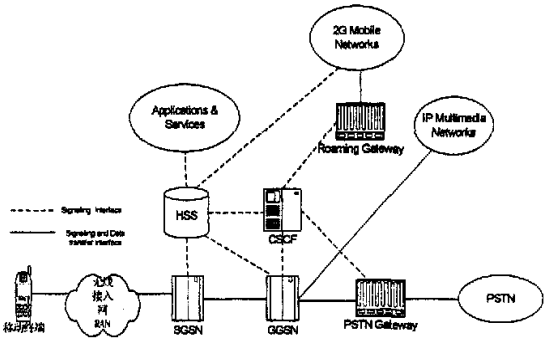

图 1.4 3GPP IP 参考架构

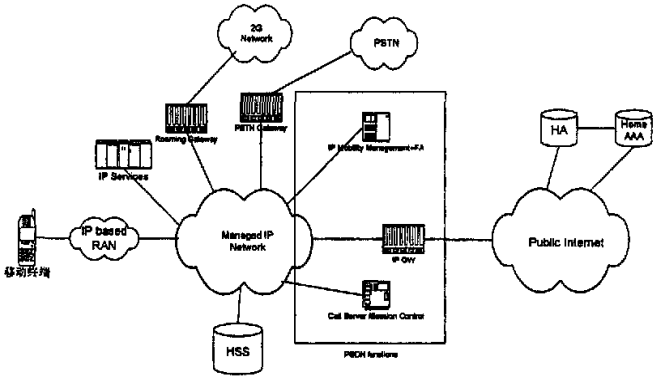

图 1.5 3GPP2 IP 参考架构

在移动网络中实现 NGN，相对于固定网络有更大的困难。3GPP 在 NGN 标准方面的贡献，是在 R5 种推出了 IMS 体系架构。ITU-T，ETSI 接受了 IMS 架构，并加入到对它的改进和完善中。

3GPP R5 于 2002 年 6 月冻结。其主要特点是：第一，为 GPRS/3G 网络设计了基于 IP 传输之上的标准化的多媒体解决方案 IMS，包括安全功能、计费功能、漫游功能和 QoS；第二，基于 IETF 的协议，采用 SIP 和 SIP 扩展；第三，SIP 作为惟一的 IMS “呼叫控制”协议；第四，所有的 IMS 网络模块都用了 IPv6，包括移动终端；第五，注册网络控制，漫游用户由原始注册网络控制；第六，接入独立，IMS 独立于底层的 IP 接入网络。3GPP R6 于 2004 年年底被冻结。其主要特点是：第一，IMS 到 CS 互通，支持 SIP/ISUP 互通和 CS 漫游情况；第二，支持 IMS 到 PS 互通以及同 IP 端点的互通；第三，支持 WLAN 接入方式；第四，支持多种业务，如聊天、IMS 会议，并支持在线功能；第五，支持基于 IP 流的计费。

3GPP2 也接受 3GPP 的 IMS 理念，从而使得未来基于 SIP 的业务可以在两个网络中得到统一，但是双方的网络架构设计有很大不同，最大的一个是 3GPP2 采用 IETF 移动 IP 标准支持终端移动性。3GPP2 的目的是结合 3G 的高速数据传输能力和移动 IP 标准的广泛适用性，提供 IP 支持，增强网络能力。3GPP2 全 IP 网络的目的是提供端到端 IP 连接，分布式的控制与业务，并通过网关与传统网络连接。而采用移动 IP 的优势是可以更好地实现 IP 网络的互联与漫游。

PARLAY

1998 年 4 月，Parlay 工作组成立，创始成员有英国电信、Ulticom 、微软 、北电 、西门子，于 1998 年 12 月出版最初的版本。1999 年 5 月，AT&T、Cegetel、Cicso、Ericsson、IBM 和 Lucent 等 6 个新成员加入，成员总数增加到 11 名。2000 年 1 月 Parlay 2.0 出版，6 个月后，Parlay 2.1 出版。2001 年 2 月，Parlay 工作组成员增加到 37 个，并推出 Parlay 3.0。Parlay API 的版本中比较具有影响力的版本是 2.1 和 3.0，由于 Parlay 标准化组织以后的版本只对版本 3.0 具有向后兼容性而对 2.1 版本没有后相兼容性，因此目前我国行业标准是以 Parlay3.0 为基准。

标准组织的协作

为避免重复劳动，相互借鉴经验和教训，各标准组织在 NGN 研究方面，在保持着一定程度竞争的同时，更加注重彼此的合作。

IMS 由 3GPP 提出，得到了 3GPP2，ITU-T 以及 ETSI TISPAN 的重视，尤其 TISPAN 将 IMS 引入到固网中，希望 IMS 成为固网与移动网融合的平台。对于 IMS 的研究，TISPAN 和 3GPP 的保持密切合作，但是在一些问题上还没有明确的结论，在规范的制定方面还有很多工作要做。TISPAN 和 3GPP 成立了联合工作组研究下一代网络与 IMS 相关问题。目前的焦点问题主要有：用户识别卡问题，在 3GPP 的规范中每个终端都有一个 UICC（Universal Integrated Circuit Card）卡，可以实现多种应用，包括目前 SIM 卡功能和其他应用功能，TISPAN 正在评估下一代网络终端中是否也需要一个 UICC 卡，目前还没有形成统一意见，还要在 TISPAN 内部继续讨论；IMS 中的各种业务引擎问题，TISPAN 将采用 3GPP 中定义的各种业务引擎；Presence（在线业务）的应用标志问题，TISPAN 倾向于除了 SIP 电话以外的呼叫都采用 URI 标志，而 3GPP 规定所有呼叫都使用 URI 标志，这还要和 3GPP 进行协商；组管理与会议问题，TISPAN 正在研究 NGN 中是否要定义 MRFC（多媒体资源功能控制器）和 MRFP（多媒体资源功能处理器）之间的接口，3GPP 已经将其定义为采用 H. 248 协议的接口；关于即时消息（Instant Message）的问题，TISPAN 正在研究是否要在媒体通道上设置 NAT 或防火墙，如果设置则需要消息

中继，3GPP 目前没有相应的规范。

从目前的情况来看，TISPAN 标准推进速度较快，也具有可操作性。并且 TISPAN 和 3GPP 的合作较为紧密，TISPAN 主要在 3GPP 规范的基础上加以扩展来实现固定的特性，并争取与 3GPP 尽量保持一致。

IMS 基于 IETF 制定的 SIP 协议，作为因特网协议，SIP 缺乏电信标准所需要的一些特征，电信标准组织在采用它时，需要进行一些改造和扩展，它们会向 IETF 提出自己的需求，然后配合 IETF 进行 SIP 协议修订和扩展标准的设计。

ITU-T 在 NGN 的研究方面，相对于 ETSI TISPAN 有些落后，所以它在标准制定方面，一方面在 TISPAN 的压力下，促使自己加快进度，另一方面在工作中大量借鉴了 TISPAN 研究的成果和思想。

1.4.2 我国关于 NGN 研究的情况

经过几十年的经济建设，中国已经开始从跟踪世界先进技术进入到自主创新的时期。为使我国在未来信息产业的发展中不再继续落后，在 NGN 的研究上必须紧紧跟随国际上的每一步进展。NGN 技术对中国通信产业的发展非常重要，虽然改革开放以来，我国电信产业取得了长足发展和进步，但我国一直缺乏知识产权，特别是原创性的知识产权，在参与国际竞争时非常被动，因而期望在 NGN 技术上有所创新，取得突破，获取原创性的知识产权，并在某些方面取得优势和领先。

为此，中国通信标准化协会（CCSA）各技术工作委员会密切跟踪国际上 NGN 标准的发展进程，研究和制订我国 NGN 的相关标准，并积累了大量资料和经验，目前已经取得了一些成绩，为我国 NGN/软交换网络的建设，网络演进，设备开发、研制和引进打下了一定基础。在 ITU-T NGN 相关活动中，我国参会人数及提交文稿数越来越多，参加 ITU-T 会议从以前的跟踪转为参与讨论和积极推进。我国相关科研机构和企业承担了不少新建议草案的起草。

我国在 NGN 及软交换领域，目前的研究情况大致是这样的：

1) NGN 相关标准研究：我国正在对 NGN 框架规范、NGN 业务平台/体系、NGN 网络管理技术、NGN 端到端的 QoS、NGN 网络安全、NGN 移动性管理、固网与移动网融合等内容进行积极研究。

2) 软交换相关研究：正在研究软交换网络框架体系、软交换网络安全。基本完成了软交换机、移动软交换服务器、ATM 中继媒体网关、IP 中继媒体网关、综合接入媒体网关、多媒体网关、信令网关、基于软交换的应用服务器、基于软交

换的媒体服务器、综合接入设备、SIP 服务器、软交换业务接入控制设备、PARLY 网关等组网设备的标准制订。

3) 移动通信相关研究：我国对 3G 空中接口，组网，业务接口等方面都有所贡献。基本完成了移动软交换服务器设备技术要求及测试方法的标准，移动媒体网关设备规范、测试方法及接口等方面的标准。

4) IP 与多媒体研究：NGN 将采用 IPv6， 但还不能完全排斥 IPv4。IPv6 是目前各界研究的重点，我国正在启动 CNGI 的项目，目的是组建一个全国范围内的 IPv6 示范网络，推动 IPv6 的产业化。IPv6 相关技术标准的研究制定正在进行中。国内电信制造企业也根据标准开展了 IPv6 路由器等通信产品的生产研制工作，目前已有了 IPv6 产品的相关技术标准，包括 IPv6 基本协议，IPv4 向 IPv6 过渡的相关内容，双协议栈（IPv6+IPv4）的一些设备规范。IMS 研究方面的工作已经开展了很长时间，但目前标准化工作才刚刚起步。CCSA TC5 中的第九工作组是 IMS 标准主要的起草组织，即将完成 IMS 总体的预研报告，2006 年将开始对 IMS 设备和接口的标准化工作。第三组主要负责 IMS 安全相关的研究，目前已完成 IMS 安全的规范。第七组主要是负责和 IMS 未来业务相关的研究，也已经出台了一系列业务标准。

5) 其他方面的研究：接入网方面主要是面向宽带的无线接入技术，如第二代不对称数字用户线，ADSL2＋的技术，综合接入系统的技术要求，甚高速数字用户线等；传输网方面主要是光传送网的体系、设备功能特性，自动交换光网络的结构和要求，自动交换光网络设备技术要求等。

目前看来，我国有可能在国际上提出建议和突破的方向包括以下部分：固定/移动网络融合架构、软交换演进方向（包括 PSTN/ISDN 向基于软交换/CS 的 NGN 演进的策略和场景、基于软交换的组网架构、软交换的未来演进方向等）、IMS 在固定网中的应用、NGN 承载网技术、NGN 业务、NAT 穿越技术等。

1.4.3 NGN 实际应用状况

NGN 网络的试验和商用部署，国外的运营商起步较早。1999 年 5 月，英国电信（BT）利用北电网络的 Succession 解决方案，在西班牙建立了世界上第一个大规模、基于分布式 NGN 结构的电信级长途网，并投入商业运营，揭开了电信业 NGN 演进的序幕。目前，BT 西班牙已拥有近 2000 个大型企业、67000 个中小企业、15000 个居民用户以及大约 90 万个通过其它运营商 PSTN 间接接入的客户群。

美国第一个部署 NGN 的是 Sprint 公司。2001 年 11 月，Sprint 公司的本地电信业务部将其整个电话网改造为基于分组的下一代网络，拉开了进军 NGN 的帷幕。2003 年 5 月，Sprint 宣布已将 C5 局移植到分组交换网络，标志着 Sprint 的电路交换网正式转换到分组交换结构，实现将三个相互独立的 TDM、DSL 和专线网合并为一个可同时提供语音、数据和专线业务的分组网络。Sprint 公司发展 NGN 的首要驱动力是降低网络运营维护和升级成本，简化网络平台，从三个独立的网络转向一个分组网。

美国 Verizon 公司于 2004 年初宣布全面启动 NGN 网络建设，包括长途、本地汇接及本地接入等各环节。Verizon 是美国最大的固网运营商，但是面临着发展的内忧外患。外部环境方面，美国电信运营商多如牛毛，加之 FCC（美国联邦通信委员会）鼓励充分竞争，因此运营商提供的电信业务及服务日渐同质化，只能通过降低运维成本来保证利润率，新兴运营商的 VoIP 业务不断分流客户；公司内部方面，Verizon 拥有多个独立业务网，包括语音网、数据网、传输网、智能网、7 号信令网、视频会议网等，这种分离网络的状况使成本居高不下，同时难以满足未来发展要求。启动了 NGN 是 Verizon 摆脱困境的最佳选择。

此外，德国、加拿大、澳大利亚、韩国等国家都已经有了 NGN 网络的运营项目。

国内方面，在信息产业部等相关部门的大力支持下，中国电信、中国网通等运营商都在进行 NGN 相关研究和组网试验。

中国电信在集团公司统一部署下，从 2001 年开始进行软交换技术的的跟踪。2002 年选用西门子、中兴、阿尔卡特、爱立信、北电、华为六厂家软交换系统，分别在广州、深圳、上海、杭州四个城市建设实验网络，对软交换系统的功能、性能和协议等能力进行了全面测试。2003 年继续在广东、上海和浙江进行了业务试验，进一步验证软交换系统提供业务的能力，与此同时，两北地区开始试商用。2004 年在深圳、肇庆开展试商用，在业务、运维和营销等各方面全面验证软交换网络规模商用的能力。试验证明软交换体系在技术上已经基本成熟，在技术上，窄带软交换产品已经成熟，宽带软交换产品基本成熟。2005 年 5 月份，中国电信长途汇接局 NGN 项目的启动，标志着中国电信的 NGN 规模应用正式启动。

中国网通在建设南方网络的时候，已经充分考虑到在包括交换网、传输网、智能网、数据网等在内的各个网络里可能会采用不同于现有网络的新技术，并为此做了充足的准备。2003 年 8 月，中国网通启动了广东、陕西和山东三地的 NGN 商用实验网，为将来大范围部署和推广 NGN 做准备。2004 年底，江苏网通在全省

范围内建成全新的 NGN 网络，面向企业及个人用户开通一系列新业务。

2001 年，铁通在北京、上海、广州三地建设 NGN 试验网，主要测试分组中继长途话音业务的技术性能，这次测试侧重在话音业务。2002 年 7 月，铁通开始了以软交换为核心的第二次 NGN 网络现场测试。2003 年 7 月，铁通首先在重庆开展了 NGN 本地话音商用试验网的测试和部署。

与大陆主要运营商在 NGN 方面的进展相比，中国香港的电信运营商为了抢占市场、取得竞争优势，在 NGN 方面的步伐较快。2002 年 6 月，中国香港的新兴电信运营商香港宽频宣布采用北电网络 Succession 本地接入分组话音解决方案，在香港本地开展话音业务，迄今用户数已增至 7 万。 2002 年 12 月，中国香港另外一家电信运营商新世界电信也宣布采用北电网络的 Succession 本地话音解决方案，加入香港本地话音业务的角逐，提供分组话音、数据和未来的多媒体新业务。

1.5 研究工作介绍

NGN 移动环境是 NGN 研究领域中日益为人们所重视的研究内容。不仅 NGN 最初所设想的分层结构及业务与控制分离、控制与传输分离的思想扩展到移动通信领域，而且 NGN 的研究也在大量吸收移动通信领域中的思想和标准。由移动通信领域标准组织 3GPP 提出的 IP 多媒体子系统 IMS 日益受到 ITU-T、ETSI 等 NGN 最初参与者的重视和青睐，并构想通过 IMS 实现移动网和固网在 NGN 大背景下的融和。NGN 的最初提出很大程度上是受了因特网的影响，而因特网相对于传统电信网的最大优势就是其业务种类的丰富多彩。从用户的角度来说，NGN 的魅力应该就是极具吸引力的业务，对运营商来说，采用 NGN 的动力也来源于新业务所带来的利润。随着计算机技术和通信技术的发展，多媒体业务是未来业务的发展方向，单纯语音和简单数据业务是没有吸引力的。

在 NGN 移动环境中，提供多媒体业务，需要相应的设备支持，如应用服务器、媒体服务器、PARLAY 网关等，其中媒体服务器是 NGN 中提供多媒体资源功能的关键设备，它在软交换机和/或应用服务器的控制下，为来自电路交换网或分组交换网的各种类型的多媒体业务提供分组化的高级媒体处理服务，是多媒体业务得以实现的关键。媒体服务器提供的资源包括语音、图象、视频和文本。在语音方面，媒体服务器支持 DTMF 信号的采集与解码、信号音的产生与发送、录音通知的发送、语音电话会议、自动语音合成、自动语音识别、文本转语音、录音等功能。其他方面，媒体服务器支持传真信号的编解码、多种视频（图像）格式的编解码、音频视频的集成和同步等。

根据中国通信标准协会提出的《基于软交换的媒体服务器总体技术要求》，针对这种设备在未来 NGN 移动环境中提供多媒体业务的需求，我们设计了其系统结构，功能模块，控制协议，网管及操作维护系统，并实现了原形系统。本论文的第一部分主要介绍基于软交换的媒体服务器的研究与设计的相关内容，及有关其在 NGN 移动环境中应用的内容。

NGN 移动环境中，移动性管理是一个重要的研究内容。而且这里的移动性管理不再只是对终端移动性的管理，还包括业务移动性、个人移动性、模式移动性等移动模式。为了支持 NGN 移动环境中复杂的移动性管理，需要采用多层多协议的管理策略。本文提出一种这样的多层多协议管理架构。

多层多协议管理架构中，移动 IP 和 SIP 协议分别实现网络层和应用层的移动性管理。AAA 是网络应用部署的基础，当各层协议独立进行 AAA 操作时，存在缺乏效率的问题。为解决这个问题，提出优化解决方案——"移动 IP 与 SIP 集成应用中优化的 AAA 过程"（OAPIMS）。并对优化方案进行了分析，分析表明，该方法可以明显降低信令开销，减少时延，提高系统性能。

IMS 对在 NGN 移动环境中提供多媒体业务和进行移动性管理都有非常重要的意义。通过 GPRS 接入 IMS 是目前比较有现实意义的 IMS 接入手段，但是移动终端通过 GPRS 接入 IMS 时，GPRS 和 IMS 都会对移动终端进行认证和密钥分配，操作过程类似，独立进行两次非常缺乏效率。为解决该问题，结合相关研究成果，提出一种优化集成方法，可以只进行 GPRS 的 AKA 协议操作，IMS 基于 GPRS 的认证结果实现对移动终端的认证，同时满足移动终端对 IMS 网络的认证需求及双方的密钥分配需求。对该方法的可行性及性能进行了分析。

参考文献

1. ITU-T's Definition of NGN. http://www.itu.int/ITU-T/ngn/definition.html
2. International Softswich Consortium, Reference Architecture Version 1.2, June 2002. http://www.softswitch.org.
3. 续合元. 下一代网络（NGN）的框架结构. http://viewpoint.cttl.com.cn/
4. 3GPP TS 33.228 V5.13.0—2004, IP Multimedia Subsystem (IMS); Stage 2(Release 5) [S].
5. 赵慧玲,叶华.多媒体软交换技术探讨[J].中国无线电管理, November 2002.

6. Patel, G. and S. Dennett. The 3GPP and 3GPP2 movements toward an all-IP mobile network [J]. IEEE Personal Communications, 2000, 7(4): 62-64.

7. 门汝静.英国电信的NGN之路. http://market.cttl.com.cn

8. 邢燕霞,赵慧玲.基于IMS的网络融合分析[J].电信科学. 2005，3:1-5

9. 王坤，冯波. NGN在国外的发展之路. http://market.cttl.com.cn

10. 刘多.NGN的国际标准化进展. http://viewpoint.cttl.com.cn/

11. 孙元宁.基于全IP核心网络架构的IMS. http://www.ccidcom.com/

12. 王柏.智能网教程[M]. 北京：北京邮电大学出版社,2000.

13. 陈前斌，黄琼，隆克平.下一代网络通用移动性管理技术初探[J].通信学报,2004 25(12):65-70.

14. 万晓榆.下一代网络技术与应用.北京：人民邮电出版社，2003.

15. 陈建亚,余浩.软交换与下一代网络.北京：北京邮电大学出版社，2003.

16. 赵慧玲，解冲锋. ITU在NGN体系结构等方面的标准化进展情况[J].电信技术,2004,10.

17. 赵惠玲,叶华等.以软交换为核心的下一代网络技术.北京：人民邮电出版社，2002.

18. 赵慧玲.标准化工作进展及NGN定义.人民邮电报, 2004-12-30.

19. 赵鹏, 周胜, 望玉梅(译). IMS:移动领域的IP多媒体概念和服务. 北京：机械工业出版社，2005.

第二章. 基于软交换的媒体服务器研究及原型系统设计实现

2.1 引言

NGN 之所以产生，根本原因还是传统电信网自身存在着弊端。新技术、新概念之所以被人们关注和逐渐接受，背后都有相应的推动力量。NGN 背后的推动力量，主要是人们对新业务的需求，以及电信运营企业在面对因特网竞争时所产生的压力和对新利润来源的渴望。传统电信网可以提供的业务非常有限，主要是语音业务和有限的一些数据业务，如传真。计算机技术和通信技术的发展，尤其是因特网的普及，产生了许多新鲜业务和娱乐内容，如电脑游戏，即时聊天，视频点播，流媒体等。高效、经济地提供具有一定服务质量要求的传统电信业务，并且可以提供新的，丰富多彩的多媒体业务，是 NGN 成败的关键。

在支持多媒体业务的 NGN 系统中，媒体服务器是提供多媒体资源功能的关键设备，它在软交换机/应用服务器的控制下，为各种类型的多媒体业务提供分组化的高级媒体处理服务，是多媒体业务得以实现的关键。这里，媒体服务器主要受软交换机或应用服务器控制，处于这种特殊地位的媒体服务器，称为基于软交换的媒体服务器，以区别于其他环境，如因特网环境中的媒体服务器。

根据中国通信标准协会提出的《基于软交换的媒体服务器技术要求》[1]，北京邮电大学网络与交换技术国家重点实验室网络智能研究中心通过两年的努力，自主研发了一个基于软交换的媒体服务器原型系统，并通过了系统测试与功能测试，目前已经进入商用开发阶段。

本章对这种基于软交换的媒体服务器系统的总体设计、系统结构及控制协议进行阐述。我们基于一种商业化 SIP 协议栈，设计实现了用 SIP 及其扩展控制多媒体会话的方案。通过对 IVR（交互语音响应）应用实现的描述，对该方案进行了介绍。

2.2 基于软交换的媒体服务器的总体技术要求

中国通信标准协会提出的《基于软交换的媒体服务器总体技术要求》设备规范于 2003 年年底推出了 1.00 版本。目前，规范的 2.00 版本也进入审核批准阶段。其 1.00 版本规定了以下内容：

1）规定了基于软交换的媒体服务器在软交换体系中的地位与应用

在软交换网络中，媒体服务器结合业务逻辑，提供业务所需的媒体资源，是

业务实现过程中不可或缺的必要组成部分，广泛的应用于包括基本语音、IP Centrex，IP 会议、预付费业务、通知业务、Voice E-mail、统一通信等各种业务类型，可以提供拨号音、忙音、回铃音、等待音和空号音等基本信号音，以及会议、通知等复杂的媒体处理功能。媒体服务器有两种基本的工作方式，一是可以在软交换机控制下，提供业务媒体服务，常用于软交换机直接提供的业务；二是可以在应用服务器的控制下，提供业务所需媒体服务，常用于应用服务器提供的各种增值业务。另外，应用服务器也可以间接地通过向软交换机提出请求，由软交换机控制媒体服务器为应用服务器上的业务逻辑提供所需媒体服务。

2）基于软交换的媒体服务器的功能要求

基于软交换的媒体服务器应具有资源功能、与其它设备通信的功能和对自身资源的管理、维护功能。

资源功能包括 DTMF 信号的采集与解码、音信号的发送、录音通知的发送、会议、语音的合成和音频混合。通信功能指媒体服务器应该具有与网络中的软交换机、应用服务器、媒体网关、IP 智能终端、网管中心等实体的通信能力，通过通信接口发送或接收相关消息，检查消息格式，进行协议转换处理。管理、维护功能是指媒体服务器应该可以以本地或远程两种方式提供对媒体资源及设备本身的维护、管理，包括对媒体资源的编辑、数据配置、状态监控、故障管理等。

3）基于软交换的媒体服务器的接口要求

需要支持 10M/100M/1G Base-T 网络接口，可选支持 ATM STM-1 网络接口。可选支持 TDM 接口。媒体服务器应该具有与维护终端和网管中心的接口。媒体服务器应该具有磁盘、光盘或磁带驱动器，或者可以通过 FTP 连接到其他设备上，以提供备份或软件的导入。

4）基于软交换的媒体服务器的协议要求

需要支持的媒体传输控制协议包括 RTP/RTCP，可以选择支持 RTSP 协议。应该支持的信令控制协议有 H.248/MGCP、SIP。支持 SNMP 网络管理协议。可以支持 FTP 或 TFTP 协议。

5）基于软交换的媒体服务器的网管和操作维护要求

基于软交换的媒体服务器支持本地维护管理，还要求通过内部的 SNMP 代理模块与网管中心进行通信，接受网管中心的远程维护管理。网管系统是维护网络系统及设备正常运行的重要保证。现代网络设备必须具备支持网络管理的能力。

要求设备具有配置管理（包括对信号音 ID、消息 ID、录音通知的内容进行加载、修改以及删除等操作。当修改或补充数据时，不影响设备正常运行）、故障管理（可以在必要时或定期进行在线自检，检测设备的状态和故障，并通过告警系统，对监测异常作出相应的处理，告警系统可以按照故障的严重程度分类）、统计测量与计费管理（应提供业务统计功能，一方面反映本设备的业务负荷信息和运行状况，另一方面提供计费依据。应实现对业务量和协议的统计）、性能管理（通过对网络及网络设备进行性能监测，采集相关的性能统计数据，对它们进行性能分析，了解网络服务质量和运转效率）。

此外，因为基于软交换的媒体服务器处于类似于因特网的分组应用环境中，NGN 又是开放的网络，所以安全问题显得尤为重要，对设备本身的安全，如权限设置，访问控制等有严格要求，另外对于网络安全、协议安全也需要加以研究和防范。

6）对性能指标、系统能力、可靠性等的要求

对电信级应用的媒体服务器，最低处理能力应该为 287 对操作/秒，时延应该控制在 40ms 内。最低容量为：语音能力至少可以提供 2048 个话音信道，可存储的录音通知最短时间不低于 400 小时，视频能力至少可以同时为 128 个速率为 384kbps 的终端提供服务，同时提供 64 组会话。系统应达到 99.999%的可用性,系统无故障工作时间应大于 8500 小时，系统的故障恢复时间应小于 5 分钟，媒体服务器应该具有高可靠性，系统主要部件应该具有热备份冗余，并支持热插拔功能。

2.3 基于软交换的媒体服务器的系统结构

2.3.1 因特网媒体服务器

因特网或局域网环境中存在着不少媒体服务器，有代表性的是提供流媒体信息的流媒体服务器[2-5]。深入考察后我们发现，现存的流媒体服务器在系统结构和功能与基于软交换的媒体服务器的要求有很大差别，主要表现在：

1) 功能上比较单一，如流媒体服务器大都只提供视频媒体的下载和播放。而对于基于软交换媒体服务器，在为新型业务提供多媒体资源和媒体操作环境的同时，还必须以新的媒体格式和新的传输方式为传统电信业务提供媒体支撑。这些传统业务包括基本呼叫和补充呼叫业务、预付费业务、传真业务、IVR、语音信箱业务、录音通知业务等。

2) 系统扩展能力不足，因特网媒体服务器往往用于小型局域网环境，或者作为新闻网站主体内容的补充，很少出现扩容改造的问题。而基于软交换的媒体服务器则不同，它是用于电信运营的设备，用户量大，用户对业务质量的要求高，为了保护投资，需要根据用户量的变化，逐渐扩容，所以对设备的可扩展性要求较高，这需要在系统结构设计上予以支持。

3) 可靠性不高，由于使用环境的不同，因特网媒体服务器可靠性要求不高。基于软交换的媒体服务器因属于电信设备，对可靠性要求极高，一般要达到99.999%，即5个9的可靠性，这需要在系统结构设计中采用多种提高可靠性的设计。

4) 用户与服务器交互方式不适用，因特网媒体服务器大都采用“客户/服务器”模式，用户直接与服务器之间进行信令交互控制服务器的行为。基于软交换的媒体服务器不能这样。NGN移动环境是一种可运营可管理的，业务控制行为由软交换或应用服务器上的业务逻辑执行，基于软交换的媒体服务器只是在软交换或应用服务器的指示下为用户提供业务所需的媒体资源。基于软交换的媒体服务器与用户之间不存在控制信令的直接交互。

因特网媒体服务器的系统结构大致可分为以下两种：集中式结构和分布式集群结构。

集中式结构的媒体服务器基于传统的共享内存、多处理器的高端服务器系统。其系统结构具有以下特点：采用紧耦合多处理器系统，在多个处理器之间共享内存；由于直接沿MPP结构高性能机[21]，或针对流媒体服务对其中的某些子系统进行专门优化，因此技术成熟；由于使用专用硬件设计，性能提高的同时增加了系统成本。

分布式结构的流媒体服务器使用通用工作站，通过高速局域网连接成集群系统。分布式媒体服务器一般采用松耦合集群结构。这种集群结构有很多优势：a) 低成本，高性能。集群系统往往由普通工作站或服务器通过某种方式连接起来的多机系统，相对于大型计算机系统，成本较低，但是通过应用分布式计算技术，及负载均衡等手段，可以大大提高系统的性能，与具有相同性能的大型计算机系统相比，集群系统往往具有更好的性价比；b）高可用性。集群系统因为是多机系统，在分布式管理软件的管理下，能够在某处出现故障时，由系统中其它设备接替工作，使正在进行的服务不致中断，提高了系统可用性；c）可扩展性好。集群由相对独立的计算机构成，可以通过一定的配置手段，增加或删除计算节点，灵活扩充集群计算能力。

2.3.2 基于软交换的媒体服务器的服务特征

在 NGN 移动环境中，基于软交换的媒体服务器应具有以下服务特征：

1） 大容量存储系统。多媒体数据往往数据量很大，如播放一小时的视频信息，RealMedia 格式的媒体文件一般在 80MB 左右，而 MPEG-1 格式的媒体文件可达 1GB。而且为了满足多媒体业务的丰富多彩，要求媒体服务器有能力存储和访问大量的媒体数据，这就要求配备大容量的存储系统。

2） 宽带传输能力。多媒体数据具有密集性和间歇性特性[3]，在多媒体实时业务中，对多媒体数据传输的时间要求较高，这不仅需要传输网络的支持，而且需要服务器提供必要的带宽处理能力。根据编码格式的不同，多媒体业务所需带宽也有所不同，RealMedia 格式的媒体文件一般需要几百 Kb/s 的带宽，MPEG-1 格式文件需要 1.5Mb/s 的带宽，而 MPEG-2 格式文件的带宽需求可达 6 Mb/s。

2）控制方式多样、灵活。基于软交换的媒体服务器有两种基本控制方式，一是在软交换的直接控制下提供业务所必须的媒体资源，如图 2.1-（1）所示，常用在由软交换直接提供的业务中，例如基本呼叫和补充呼叫业务；二是在应用服务器的直接控制下提供资源，如图 2.1-（2）所示，常用在应用服务器提供的业务中，例如视频会议、Voice Email 和统一通信。此外，应用服务器可通过软交换提出媒体要求，由软交换控制媒体服务器为应用服务器上执行的业务逻辑提供媒体资源服务。

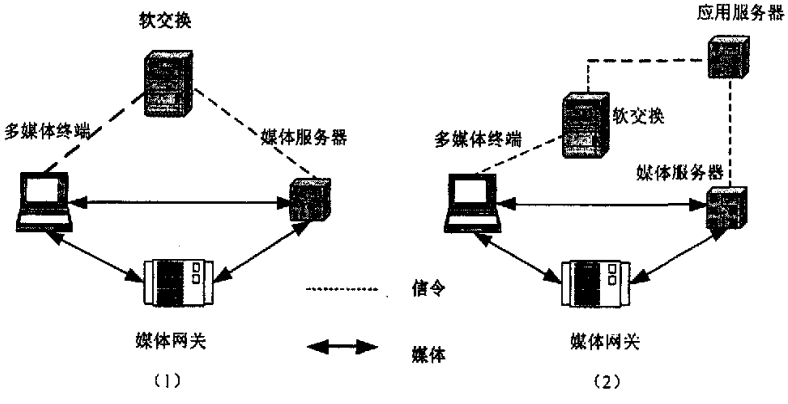

图 2.1 基于软交换的媒体服务器接受多种控制方式

4） 多种控制协议。媒体服务器由软交换机或应用服务器控制时，应支持 SIP 协议、H.248 协议或 MGCP 协议，SIP 协议应作为优先选择。因此，在媒体服务器中需同时支持多种控制协议，并根据实际需要，灵活选择合适的协议。这种多协

议环境，增加了基于软交换的媒体服务器的呼叫模型在设计和实现时的复杂性。

5） 支持业务多样性。基于软交换的媒体服务器应为多种业务类型提供资源保证，这些业务既有传统电信业务又有新型多媒体业务。传统电信业务包括：基本呼叫和补充呼叫业务、预付费业务、传真业务、IVR、语音信箱业务等，新型的多媒体业务例如多媒体会议、彩铃、彩话、信息服务、视频、Voice Email 和统一通信等。不同的业务类型有着不同的资源要求和不同的资源处理方式。

2.3.3 基于软交换的集群媒体服务器的系统结构

为满足基于软交换的媒体服务器的系统要求及其在NGN移动环境中的服务特征，结合《基于软交换的媒体服务器技术要求》，我们设计了基于软交换的集群媒体服务器（SCMS，Softswitch-based Clustered Multimedia Server）的系统结构[6]。SCMS 继承了现有因特网媒体服务器的一些优点的同时，可以满足各方面对它更高要求。

SCMS 的系统结构如图 2.2 所示。系统主要由两大部分组成，一是资源控制节点（RCN，Resource Control Node），它是整个媒体服务器的全局资源管理者和整个系统的接入点，维护一个保存着各种资源信息的数据库，并采用合理的分配算法对资源进行分配。RCN 中实现了媒体服务器要求的多个控制协议，接受和处理从业务控制点（软交换或应用服务器）发来的控制请求；另一个是资源处理节点（RPN，Resource Process Node），它是多种类型的资源处理设备的总称，包含基本语音资源处理点、IVR 资源处理点、会议资源处理点和视频资源处理点等专门的资源处理部件，为提供特定资源处理功能。

在 SCMS 中，由于媒体信息直接在 RPN 与用户之间传送，所以不存在内部网络带宽瓶颈，系统性能具有随系统规模增长而线性扩展的能力。同时，用户和 RPN 之间没有控制信令的交互，所有的控制信令都按照如下的路径传送：用户←→软交换/应用服务器←→RCN，RCN 通过内部信令与 RPN 交互，实现软交换/应用服务器对业务资源的控制。

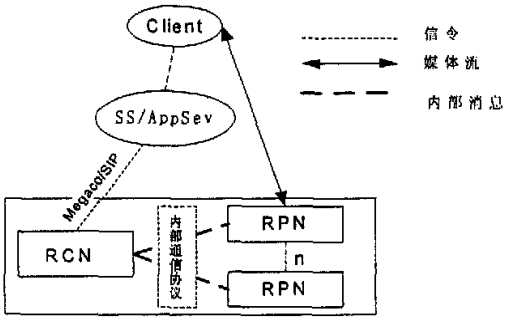

图 2.2 SCMS 的系统结构

2.3.4 SCMS 软硬件结构

SCMS 的硬件结构如图 2.3 所示。分前台处理机和后台处理机两个部分。前台处理机实现了 RCN 的功能，是全局资源的管理和分配者。后台处理机为各类型业务提供专用的资源处理功能。监控台监管整个集群系统并负责对设备故障的报警。全部设备通过局域网相连，局域网通过高性能路由器连接到 IP/ATM 骨干网络上。为了满足 SCMS 的高可用和实时性约束，SCMS 为前台处理机和 LAN 分别采用了热备份方案。系统中具有相同功能的后台处理机分别配置了多台，它们之间以负荷分担的方式工作。

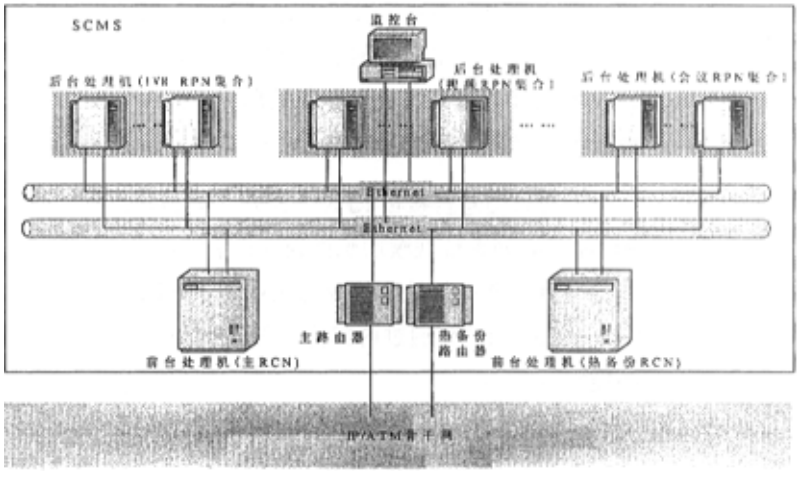

图 2.3：SCMS 的硬件结构

SCMS 的软件结构如图 2.4 所示。RCN 包含了 6 个软件模块和 1 个资源信息数据库。SIP 协议栈、H.248 协议栈和 MGCP 协议栈分别实现了各控制协议的协议

功能。资源控制模块抽象了多媒体呼叫的呼叫模型，并实现了对 RPN 的负载均衡策略、媒体流的 QoS 控制等功能。在 RCN 和 RPN 之间定义了对资源进行操纵的内部消息格式，内部消息接口模块提供了对内部消息进行包装和解析的 API。

SCMS 中的通信中间件为 RCN 和 RPN 提供了通信手段，它包含多个组件，驻留在 RCN 和 RPN 上的只是其中的一部分。通信中间件具有以下功能：对通信各方通信进程的管理功能、对通信进程之间的消息的自动路由与分发功能、流量控制功能、性能监测与告警功能、控制台接入功能等。

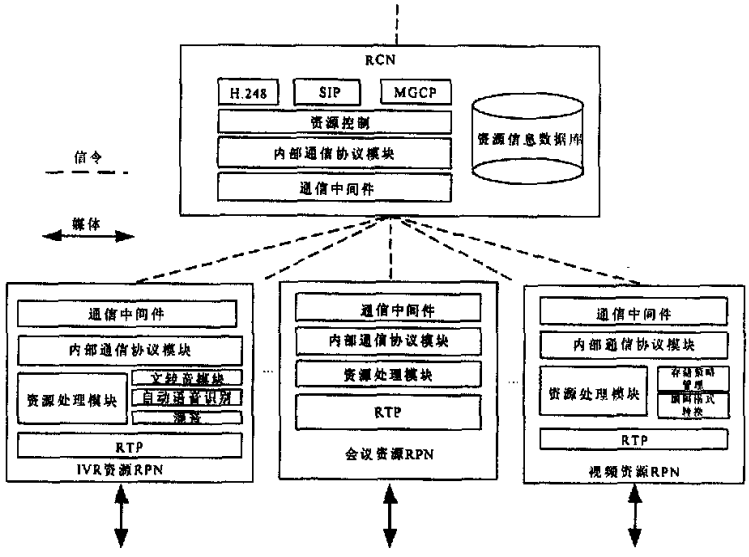

图 2.4 SCMS 的软件结构

资源信息数据库存储了 RPN 的媒体内容信息和资源处理能力信息。RCN 利用这些信息来获知哪些 RPN 是可用的，每一个 RPN 上有多少个资源可用，以及与用户通信采用的媒体格式、带宽等信息。RCN 对媒体内容的使用情况和 RPN 的系统负载情况进行跟踪，从而确认 RPN 的处理能力。

RPN 包含了多种类型的资源处理设备，例如会议资源处理设备、IVR 资源处理设备和视频资源处理设备等，为各种业务类型提供资源支持。从软件模块上看，各种类型的 RPN 都包含了一个通信中间件、一个资源处理模块和一个 RTP 协议栈模块。通信中间件与 RCN 中的通信中间件是相同的。RTP 协议栈模块实现 RTP 的协议功能，使媒体数据以 RTP 的包格式进行传输。资源处理模块接受来自 RCN 的内部控制消息、判断是否接纳新的业务请求，如果业务请求被接纳，资源处理模块将协调系统的软硬件设备，实现 RCN 的控制意图。

2.4 基于软交换的媒体服务器中的控制协议

根据《基于软交换的媒体服务器技术要求》的要求、建议以及我们对各种备选协议的认识了解，我们选择 H.248 协议和 SIP 协议作为控制协议。

2.4.1 H.248 协议

H.248[7]也叫 Megaco[8]，是 ITU-T 和 IETF 联合提出的媒体网关控制协议，在 ITU-T 称为 H.248，在 IETF 称为 Megaco，二者略有差异。H.248 的最新版本是 2002 年 3 月的出版的 ITU-T H.248.1: Gateway control protocol。其发展过程如图 2.5 所示。

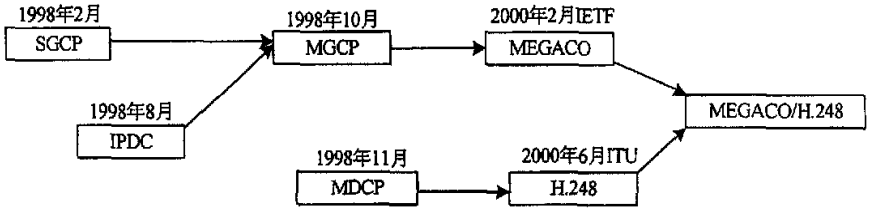

图 2.5：H.248 的发展过程

2.4.2 H.248 协议内容

2.4.2.1 H.248 协议的连接模型

H.248 被设计为媒体网关（驻地网关、中继网关、综合接入网关等）的承载连接行为进行控制的协议，为了对媒体网关内部可被媒体网关控制器描述和控制的对象进行抽象和描述，H.248 提出网关连接模型的概念。模型基本构件有两个：终结点（Termination）和关联（Context）。

终结点是媒体流的源或宿。一个终结点可以发起或终结一个或多个数据流的逻辑实体。终结点可以对应一个物理实体，也可以对应一个虚拟实体。虚拟终结点对应信息流，例如一个 RTP 信息流，它依附呼叫，呼叫结束时，即消亡，又叫瞬时终结点。

终结点一般有 4 类属性，分别是性质（Property）、事件（Event）、统计（Statistics）信号（Signal）。性质属性分终结点状态特性和媒体流相关特性，前者表示终结点所出的服务状态（如正常工作、故障或测试），后者主要表示短时终结点相关的媒体信道属性（如只发/只收终端、媒体类型、编码格式、编码参数等）；事件属性表

示该终结点应对那些事件进行监视并向媒体网关控制器报告，典型事件如摘机、挂机、收到被叫号码等；统计属性指示终结点应采集并上报给媒体网关控制器的统计数据；信号属性表示应向终结点施加的信号，典型的如拨号音、DTMF 信号、录音通知。H.248 协议用“描述符”（descriptor）这一数据结构描述终结点属性，并针对终结点公共属性，分门别类地定义了 19 个描述符，一个描述符通常只包含某一类终结点属性。

关联用于描述终结点之间的关系。任何终结点都存在于某个关联中，如果一个关联中包含两个以上终结点，关联将会描述它们之间的拓扑关系和媒体混合/交换参数。协议定义了一种特殊的关联——空关联（NULL Context），它包含所有未与其他终结点联系的终结点。关联域有若干属性，如关联域标识、拓扑、优先级和紧急呼叫指示语等。

图 2.6 是 H.248 协议采用的连接模型的一个具体示例。

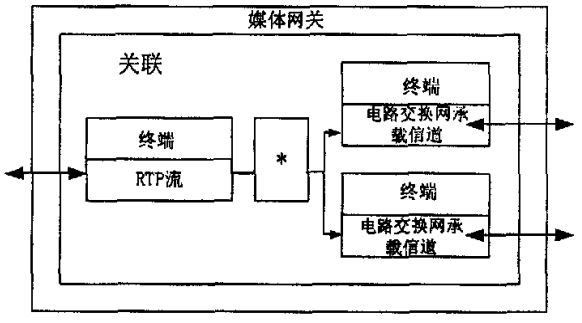

图 2.6：双边连接模型示例

2.4.2.2 H.248 协议的命令和描述符

H.248中定义了八个命令，它们分别是Add，Modify，Subtract，Move，AuditValue，AuditCapability, Notify和ServiceChange。命令的简单说明见表2.1。命令的参数由描述符进行封装。

一个描述符包含一个名字和一些属性，属性可以有不同的值。在ITU-T H.248.1中一共列出了包括Modem 、Mux 、Media 、TerminationState 、Stream、Local、Remote、LocalControl、Events、EventBuffer、Signals 、Packages 、DigitMap 、ServiceChange、ObservedEvents在内的23种描述符。

表2.1: H.248命令及其用途

命令名称	用途
Add	使用Add命令可以向一个关联添加一个终结点。
Modify	使用Modify命令可以修改终结点的特性，事件和信号，如通知MG设备收号码、放音等。
Subtract	使用Subtract命令可以删除一个终结点与它所在的关联之间的联系。
Move	使用Move命令可以自动地将一个终结点从一个关联转移到另一个关联。
AuditValue	使用AuditValue可以获取有关终结点的当前特性，事件，信号和统计信息。
Audit capabilities	使用该命令可以获取媒体网关所允许的终结点的特性、事件和信号的所有可能值的信息。
Notify	MG使用Notify命令可以向MGC报告MG中所发生的事件，如摘机事件。
ServiceChange	MG使用该命令向MGC报告一个终结点或者一组终结点将要退出服务或者刚刚进入服务或者注册，MGC使用该命令通知MG将一个终结点或者一组终结点进入或者退出服务。

2.4.2.3 H.248 协议的通信机制

H.248 的消息结构包含一个消息头部和紧跟的一系列事务（transaction）。消息头包含了协议的版本信息和消息发出者的标识信息，标识信息是接收者做消息认证和消息回复时的依赖信息。事务由一个或多个动作（action）组成，而一个动作由在一个关联中使用的一系列命令（command）组成。它们的关系如图 2.7 所示。

事务分为事务请求、事务回复、事务回复确认和事务未决四种类型。事务请求与事务回复一一对应，当收到请求但未能及时回复时，接收者应该发事务未决指示消息。收到事务回复后应该发事务回复确认。事务交互保证对命令的有序处理，即：在一个事务交互中命令是顺序执行的，但并不保证各个事务之间的有序处理。在事务的交互过程中，如果某一部分执行失败，则该事务中的所有命令都停止执行。

动作包含了对某个关联使用的一系列命令。动作的开头部分包含对关联特性进行操作的信息，其后是一系列命令。每一条命令针对该关联中的特定终结点。命令是协议中的最小执行单位。命令的内容被细分为一级或多级描述符（descriptor），每一类描述符都携带了相应的协议功能。这些描述符中最重要的一个是 Stream 描述符，一个 Stream 描述符对应于终结点中的一个流。

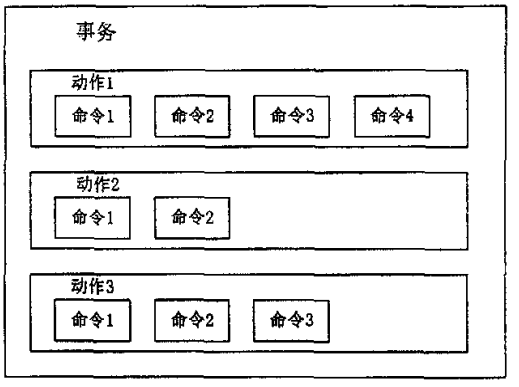

图 2.7：H.248 消息中事务结构

H.248协议消息可以在TCP或UDP上传输，或者同时支持两者。如果对等实体没有提供相应的通信端口，无论是TCP还是UDP，指令将被送到默认端口上，对于文本编码的操作，端口号应当使用2944；对于二进制编码操作，则端口号使用2945。

2.4.2.4 H.248 协议的扩展机制

由于应用的多样性和技术的不断发展，新的终端和属性不断出现。为此，H.248协议定义了一种终端属性描述的扩展机制，即控制包（package）机制。未在基础协议的描述符中定义的终结点属性可以根据需要增补定义相应控制包，包中定义的属性用{PackageID，属性ID}标识，其中特性ID可以是PropertyID、EventID、SignalID和StatisticID。控制包经批准并得到IANA（因特网号码分配署）分配的号码后，即成为H.248协议的正式附件，可以在相关的H.248命令中使用。

2.4.3 SIP 协议介绍

会话初始化协议 SIP（Session Initiation Protocol）是合并最初由 Mark Handley 和 Eve Schooler 提出会话邀请协议 SIP（Session Invitation Protocol）和由 Henning Schulzrinne 提出的简单会话邀请协议 SCIP（Simple Conference Invitation Protocol），并由 IETF（Internet Engineering Task Force）多方多媒体会话控制（MMUSIC, Multiparty Multimedia Session Control）工作组经过草案讨论，而于 1999 年 2 月作为 RFC2543[9]公开发表的。2002 年 7 月,SIP 协议由 RFC3261[10]进行了更新。

SIP的提出，是为解决邀请特定用户加入某个会话的问题。正如IETF一贯倡导的那样，它所制定的协议，应该是为丰富IETF工具箱而制定，也就是说是在现有协议无法满足所需功能的情况下制定。协议设计应该简洁，优雅，功能制定应该刚好能充分满足某种需求。SIP恰恰是这样的产物，它就是为建立会话而制定，它

将建立一个会话和描述一个会话的功能分离。对于描述一个会话，它可以借助于其他的协议，比如会话描述协议（SDP, Session Description Protocol）[11]。这样的好处，一是SIP协议本身可以比较简单，明了；二是当需要描述其他会话时，可以利用其他的会话描述协议，而不改变SIP建立会话的方式。

2.4.3.1 SIP 的优点

SIP 由 IETF 制定发布，所以 SIP 协议有着因特网协议的明显特征，SIP 设计中借鉴了不少因特网协议的优点，比较典型的两个是 HTTP[12]协议和 SMTP[13]协议，和它们一样，SIP 信令用文本方式描述，易读，易描述，也容易调试。作为 IETF 工具包中的一个工具，SIP 协议执行自己的功能并能够利用其他的因特网协议处理另外的任务，这提供了很大的灵活性，因为 SIP 协议和其他因特网协议一样可以以一种模块化的方式升级。例如，当一种新的鉴别协议由 IETF 提出，使用 SIP 的协议就可以很快利用这种机制，而无须对 SIP 协议进行修改。此外，SIP 公认的一些主要优点是：

1）可扩展性好 从诞生之日起，SIP 协议就一直处于不断地丰富和完善中。SIP 设计者在保持其核心部分简洁的同时，为 SIP 设计了强大的扩充机制。SIP 在扩展时，须注意一些原则。首先，不能破坏工具包方法，即 SIP 是 IETF 多媒体工具包的一部分，SIP 扩展不应该扩大 SIP 的使用范围，使之应用于其它因特网协议可以处理得更好的地方；其次，SIP 将会话建立与会话描述区分，这一分离原则在扩展中应该维持；第三，不能改变原有方法（命令）的语义，即保持原有方法目的的单纯性，语义的清晰性。

SIP 扩展可以从消息类型（即命令类型）、消息头、消息体三方面入手。增加消息类型的例子有 INFO[14]，用于在会话存续期间在通话双方传递信息。REFER[15]用于呼叫转移。COMET 用于会话建立过程中检查资源状况。增加消息头，也是常用的 SIP 扩展方式，因为 SIP 消息中，包含了很多的消息头，可以携带需要的信息，例如，REFER 消息中，通过消息头 referred-by 来指示发起转移的一方，referred-to 指示会话被转移到的地方。SIP 消息体可以携带 MIME 信息，所以 SIP 消息体可以携带的类型非常多，如音频、图像、Java 小程序等。在需要扩展消息体携带类型时，可定义新的 MIME 类型，如 application/isup 就能携带 SS7 的 ISUP 信令，可以在 PSTN 网络通过 SIP 网络时传输信令。

2）支持移动性 SIP 协议的动态注册机制，使支持用户移动变得十分方便。用户接入 SIP 网路时，首先向注册服务器发出注册请求，将自己的 SIP URL 和 IP 地址通知注册服务器。用户变换到其他位置时，可以再次向注册服务器注册，这样

当有其他用户呼叫该用户时，首先通过服务器了解到用户的当前位置，然后可以将呼叫路由到用户当前位置，实现用户的移动。

3）多媒体业务支持能力强 SIP 利用很多因特网协议好的特性，使得 SIP 成为一种使用户可以方便地使用多种服务的理想协议，它与 HTTP 和 SMTP 之间的相似性，使得它可以很容易地将目前最成功的因特网服务（WWW 和 e-mail）与多媒体结合起来。SIP 代表一种集成服务的解决方案，它以一种直接的方式整合了 Web 浏览器、e-mail、语音电话、视频会话、在线服务与即时消息。

SIP 的扩展能力也体现在对新业务的支持能力上。SIP 可以使用同一框架提供新的业务。例如，某业务提供商建立一个 SIP 系统提供 VoIP 业务，但是过了一段时间后，有其他新的业务产生，比如互动游戏，这时，业务提供商不必对网络做大规模改造，只需将 SIP 消息体中描述会话的协议由 SDP 改为游戏会话描述协议。此外，IETF 为基于 SIP 的应用设计了几种业务生成机制，对于不可信任者，如终端用户，可以用呼叫控制语言（CPL，Call Processing Language）生成会话处理逻辑；对于可信任用户提供 SIP-CGI 和 SIP Servlet，前者类似于因特网网站的 HTTP-CGI，独立于编程语言，为复杂程序处理提供一个开放接口，后者用 Java 语言实现，类似 WWW 服务器中的 Java Servlet，可以在需要时调用，指示 SIP 服务器处理并响应消息。

2.4.3.2 SIP 系统组成

SIP 系统一般由用户代理（User Agent）、代理服务器（Proxy Server）、重定向服务器（Redirect Server）、注册服务器（Registrar）等功能实体组成。用户代理又分为客户端（UAC，User Agent Client）和服务器端（UAS，User Agent Server），前者发起呼叫，后者响应呼叫。在用户终端，必须同时有 UAC 和 UAS。

代理服务器接收用户呼叫，根据网络策略代替用户将请求发送给相应的服务器，这可能是个接续的过程，直到到达UAS，然后再根据处理情况，返回给用户响应。代理服务器可以是有状态的，也可以是无状态的，可以对呼叫用户不进行干预，也可以对进行的呼叫积极介入。

重定向服务器根据用户呼叫，查询位置服务，以获得用户位置信息，然后将被叫用户的位置信息转给呼叫用户代理，由呼叫用户代理重新发起呼叫。

注册服务器接受用户的注册并将用户的最新位置信息保存到位置服务器中。位置服务器存储位置信息并提供位置信息的查询服务，位置服务器可以用各种方式实现，不列入SIP功能实体。

2.4.3.3 SIP 消息

SIP消息有两类：从客户端到服务器端的为请求消息（Request），从服务器端到客户端的为响应消息（Response）。这两类消息都遵从RFC 822[16]定义的通用消息格式，即由起始行（start-line）、一个或多个头域（header）、一个标志消息头结束的空行（CRLF）和可选的消息体构成，格式举例如下：

Generic-messge = start-line
*message-header
CRLF
[message-body]

起始行分请求行（Request-line）和状态行（Status-line）两种，其中请求行用于请求消息的起始行，状态行用于响应消息的起始行。

SIP请求消息

SIP请求消息格式为：

Request = Request-line
* （general-header
|request-header
|entity-header）
CRLF
[message-header]

请求行由方法、Request-URI、SIP协议版本号顺序组成，中间用空格隔开。格式如下：

Request-line = Mehod Request-URI SIP-version CRLF

SIP核心规范定义了六种方法。

1）INVITE　INVITE方法邀请其它用户加入会话，相应的消息体是对会话的描述，常采用会话描述协议SDP[9]。内容包括主叫可接收的媒体类型、可发送的媒体类型以及相关参数。INVITE也可用来修改一个已经存在的会话的参数。

2）ACK　ACK是UAC在发出INVITE，并收到最终响应时发出的确认命令，ACK只能配合INVITE使用，不能单独使用。

3）BYE　当用户代理准备释放一个呼叫时发送BYE请求。主被叫双方都可以发出这一请求，相当于电话系统中的挂机操作。

4）CANCEL　用这个命令可以取消一个尚未完成的请求。用户代理客户机和代理服务器都可以发出CANCEL请求。虽然CANCEL请求可以取消除ACK和CANCEL以外的其它所有请求类型，但它通常只用来取消INVITE请求。

5）REGISTER 用户代理通过REGISTER请求消息将To域中的地址注册在位置服务器中。注册服务器中的每个注册记录都有一个对应的有效期，一旦到期，记录就被删除。

6）OPTIONS　用于查询用户代理服务器的能力。用户代理服务器在响应消息的Allow域中指明它能支持的方法。代理服务器和重定向服务器只将该请求消息前转而不指示它们自己的能力。

SIP响应消息

SIP请求消息格式为：

Response = Status-line
　　*（general-headers
　　|response-headers
　　|entity-headers）
　　CRLF
　　[message-headers]

SIP响应消息的起始行为状态行，状态行由SIP协议版本号、数字表示的状态码和自然语言原因短语顺序构成，中间用空格隔开。格式如下：

Status-line = SIP-version Status-Code Reason CRLF

状态码由三个阿拉伯数字组成，用于计算机识别具体响应的结果，原因短语为帮助SIP用户理解响应消息而设计的。目前只定义了六类状态码，由第一位数字来区分。

- 1xx：提示信息，表示请求已收到，正在处理。
- 2xx：请求被成功接收并理解。
- 3xx：重定向，为完成请求功能需进一步执行动作。
- 4xx：客户端错误，如请求消息语法错，服务器不能完成操作。
- 5xx：服务器端错误，服务器不能处理明显合法的请求。
- 6xx：全局错误，请求不能在任何服务器中处理。

SIP消息体

六种请求消息中，ACK, INVITE和OPTIONS请求的消息体一般为会话描述。对于响应消息，请求方法和响应状态码决定了消息体的类型和对消息体的理解。所有的响应消息都可以包含消息体。在描述会话方面，一般使用SDP。

SDP仅仅规定了会话描述的格式，不关心媒体类型参数等的协商，而且在制定过程中尽量向通用性努力，因此它可以被SIP，SAP，RTSP，HTTP等应用协议所采纳。SDP包括会话名称和目的、会话活跃时间、组成会话的媒体及接收媒体的相关信息（地址、端口、格式等）。

SDP描述由许多文本行组成，文本行的格式为<类型>=<值>。<类型>是一个字母，　　<值>是结构化的文本串，其格式依<类型>而定。合法的类型和顺序如下：

- v=（协议版本）；
- o=（会话创建者和会话标识）；
- s=（会话名称）；
- i=（会话信息）；
- u=（含有会话描述的URI）；
- e=（获得会话信息的E- mail地址）；
- p=（获得会话信息的电话号码）；
- c=（连接信息，如果在所有媒体中都已包含则此处不需要）；
- b=（带宽信息）；
- t=（会话激活时间）；
- z=（时区调整）；
- k=（密钥）；
- a=（零个或多个会话属性行）；

2.5 基于 SIP 控制协议的系统实现

2.5.1 SIP 协议中会话及其控制

2.5.1.1 SIP 会话（Session）及对话(Dialog)与事务(Transaction)

SIP 协议中有几个重要的逻辑概念，分别是会话（Session）、对话(Dialog)与事务(Transaction)。

SIP 会话是指若干个媒体发送者和接收者及他们彼此间的媒体流。描述一个会话通常需要涉及具体媒体类型，编码方式，连通地址等细节问题。描述会话的一种最基本的方式就是 SDP[5]协议。会话参与者需要通过适当的方式，协商会话参数，在 SIP 中，利用的是“Offer/Answer”[17]模式。UAC 在 INVITE 命令或 UAS 在第一个非失败可靠响应 2xx 的消息体中携带 Offer。在 INVITE 消息体中携带有 Offer 的情况下，UAS 必须在响应 2xx 的消息体中携带 Answer；在 UAS 发送 Offer 的情况下，UAC 则在最后的 ACK 消息体中携带 Answer。经过这样的交互后，通信双方了解到对方能够接收媒体信息的类型，编码格式，地址及端口，之后就可以通过媒体连接进行通信。

对话是 SIP 中另外一个重要概念，它代表两个 SIP 用户代理之间一个 P2P（peer-to-peer）的，并且存续一定时间的关系。一个 UA 发送 INVITE 后，在接收到 2xx 后，建立对话。它代表交互 SIP 消息的一个上下文，使得两个用户代理之间的消息序列管理及消息路由都变得方便。对话由 Call-ID、From 域的 tag 参数和 To 域中的 tag 参数唯一确定。

SIP 是基于事务的协议，一个 SIP 事务通常由一个请求和针对该请求的所有响应（有临时 Response，也有最终 Response）构成。唯一例外是 INVITE 会话中，如果最终 Response 不是 2xx，即正确响应，就包括 ACK 命令。

SIP 事务分客户端事务和服务器端事务，每个事务都对应一个有限状态机，利用这种方式，可以使协议更加清晰，便于理解和实现。

2.5.1.2 会话属性的修改与会话控制

会话进行过程中，如果需要修改会话属性，可以通过重新发送 INVITE 命令，即 re-INVITE 来进行修改。通过 re-INVITE，可以修改一个对话中会话的属性，主要是媒体流的一些属性，用的同样是[17]中定义的“Offer/Answer”模式。在消息体中，需要携带完整的新的会话描述（例如 SDP 描述），而不仅仅是要修改的会话

属性，通信任何一方都可以通过这种方式，修改媒体连接的地址、端口，添加新的媒体流，删除旧的媒体流。

SIP 核心标准[6]中，没有定义在会话进行中控制会话的方法，SIP 扩展[10]定义了 INFO 命令，可以使通信双方在会话进行中，通过信令链路，交互应用层信令，对会话进行控制，或在应用层传递与正在进行的 SIP 会话相关的信息。

2.5.2 资源控制节点 RCN 软件结构

在前面关于基于软交换的集群媒体服务器（SCMS）的系统结构的介绍中，提到资源控制节点 RCN，它是整个媒体服务器的全局资源管理者和整个系统的接入点，维护一个保存着各种资源信息的数据库，并采用合理的分配算法对资源进行分配。RCN 中实现媒体服务器控制协议，接受和处理从业务控制点（软交换或应用服务器）发来的控制请求。对控制协议 H.248 和 SIP 的处理都是 RCN 的任务。在接收到相应的外部命令后，RCN 通过内部通信命令与资源处理节点 RPN 通信，指示后者进行媒体处理，提供服务。

RCN 用 SIP 协议实现控制功能的资源控制节点的软件结构如图 2.8 所示：

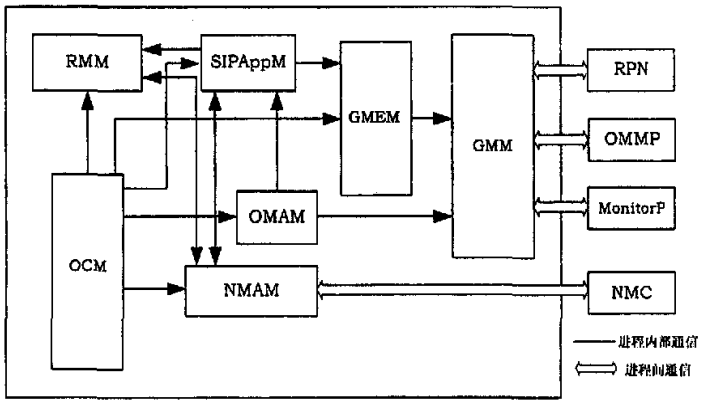

图 2.8 RCN 软件结构图

图 2.8 右边与 RCN 进行交互的外部系统或进程分别是操作维护管理进程（OMMP）、网管控制中心（NMC）、资源处理节点（RPN）、系统守护进程（MonitorP）。

资源控制节点软件共 7 个模块：资源管理模块（RMM）、操作维护代理模块（OMAM）、网管代理模块（NMAM）、SIP 应用模块（SIPAppM）、通用

消息模块（GMM）、消息封装模块（GMEM）以及总控模块（OCM）。各模块分别说明如下：

- OCM 是调度模块，负责初始化其他模块，对监控输入消息并分发处理，此外还负责向外部监控程序 MonitorP 定时发送例检消息。

- SIPAppM 是资源控制节点软件的核心模块，它负责实现 SIP 的协议功能，发出到资源处理节点（RPN）的指令。该模块需要处理两类消息，一类是由底层 SIP 协议栈送来的与呼叫相关的 SIP 消息，另一类是 RCN 与 RPN 之间的内部消息。模块根据 SIP 消息，建立会话，并根据处理要求，随时通过内部通信消息向相应 RPN 节点发出指令。SIPAppM 包含各种与呼叫和协议相关的数据结构，能在媒体服务器的外部控制设备的控制下对呼叫进行管理。

- RMM 实现对媒体服务器的资源的有效管理。系统初始化时，该模块根据资源配置文件将媒体服务器所能使用的资源全部加载到系统中，SIPAppM 将从该模块获得对资源的各种描述（例如根据 toneId 获得语音文件的名字，获得 rtp 端口号，获得 rtp 端口的编解码能力等），从而有根据地对外提供媒体服务功能。

- NMAM 收集整个媒体服务器的可管理信息，包括资源信息、统计信息等，通过 SNMP 协议响应网管控制中心发出的网管指令，或向网管控制中心发送报警信息。

- OMAM 响应 OMMP 发送的命令消息，实现用户对资源控制子系统的配置、监视和管理。

- GMEM 负责将 RCN 与 RPN 之间的内部消息封装成内部通用消息格式，再通过 GMM 进行发送到 RPN，或反向操作，接收内部通用消息并解包成 RCN 与 RPN 之间的内部消息，传给 RCN。

- GMM 提供 RCN 与 RPN 通信的内部通信协议编程接口。使用该编程接口，可以方便实现 RCN 与 RPN 之间的多点异步通信机制。

2.5.3 SIP 协议控制交互语音应答（IVR）的实现方案

通过 SIP 协议的控制，SCMS 可以实现多种多媒体业务应用，包括交互语音应答（IVR）、多方多媒体会议等，这里以 IVR 业务的实现为例，对我们用 SIP 协议作为控制协议，实现软交换机/应用服务器控制 SCMS 实现多媒体业务的方法进行

说明。实现基于美国惠普（hp）公司的商业化 SIP 协议栈。

交互语音应答（IVR，Interactive Voice Response），是自动与用户进行交互式操作的业务。用户可以通过电话等通讯终端拨号呼叫 IVR 平台，根据 IVR 平台的语音提示进行进一步操作，从而完成交易、娱乐等业务。比较典型的 IVR 应用场景有企业的客户服务系统、电话银行等。NGN 环境中，在软交换机/应用服务器的控制下，可以用媒体服务器来提供 IVR 应用。用户无论从 PSTN 还是 IP 网络发起呼叫，首先接入到软交换设备或应用服务器中，然后在它们的控制下，和媒体服务器建立媒体连接，接受媒体服务器的语音服务。以下说明中，均以应用服务器为例，并用 APPS 来代表，相应的媒体服务器用 MS 代表。

实现方案分三个部分，一是会话控制，包括会话的建立、修改、终止。这需要一系列的信令交互；二是语音资源的定义和组织；三是 IVR 中各种具体操作及参数的传送、解析与执行。下面针对这三个部分分别进行说明和讨论。

2.5.3.1 会话控制与信令交互

IVR 会话一般作为一种业务应用的辅助部分，当 APPS 进行业务操作时，在需要提供语音提示并接收用户按键选择时，通过第三方控制信令建立 MS 与呼叫用户之间的媒体连接，由 MS 提供 IVR 服务。呼叫建立的信令流程如图 2.9（1）~（6）所示。（1）APPS 发送 INVITE 信令到 MS，要求同 MS 建立呼叫。（2）MS 对呼叫请求进行排队，并用可靠中间响应的方式向 APPS 发送 182（排队）响应。（3）APPS 发送 PRACK[18]予以确认，告诉 MS 收到了临时相应消息 182。（4）MS 用 200 OK 响应 PRACK 信令。（5）MS 准备好提供服务时，发送对应于 INVITE 的 200 OK 。（6）APPS 发送 ACK 消息，至此会话建立结束。值得注意的是，虽然信令交互是在 APPS 和 MS 之间进行的，但是媒体连接是建立在用户终端和 MS 之间，这是通过设置 SIP 信令的 SDP 消息体中相关项实现的。

IVR 服务的主要操作有三个，分别是放音（PA, PlayAnnouncement）、放音收号（PC, PlayCollect）、放音录音（PR, PlayRecord）。PA 表示放一段语音，PC 表示放提示语音，然后从媒体流中收取 DTMF 拨号音信号，PR 表示放提示语音，然后对媒体流进行录音。

用户终端与 MS 之间的媒体连接建立后，APPS 需要指示 MS 进行 IVR 的具体操作，这需要用到 SIP 扩展 INFO 信令。INFO 信令的功能是在 SIP 会话建立后，在通信双方传递信息，但不改变会话状态。

图 2.9 中（7）~（10）表示 PA 操作的信令。（7）APPS 通过 INFO 信令，指

示 MS 进行 PA 操作，具体的操作命令和参数需要在 SIP 消息体中定义。（8）MS 中 RCN 指示 RPN 准备放音，并向 APPS 回复 200 OK，对 INFO 信令进行确认。（9）MS 中的 RPN 放音结束，通知 RCN，RCN 通过 INFO 信令，向 APPS 报告操作结果。（10）APPS 发 200 响应。PA 操作结束。

图 2.9 中（11）~（14）表示 PC 操作的信令。（11）APPS 通过 INFO 信令，指示 MS 进行 PC 操作，操作命令和参数在 SIP 消息体中定义（可参见 2.5.3.3 节介绍）。（12）MS 中 RCN 指示 RPN 准备放音收号，并向 APPS 回复 200 。（13）MS 中的 RPN 放音收号结束，通知 RCN，RCN 通过 INFO 信令，向 APPS 报告操作结果，结果中包括收号结果。（14）APPS 发 200 响应。PC 操作结束。

图 2.9 中（15）~（18）表示 PR 操作的信令。（15）APPS 通过 INFO 信令，指示 MS 进行 PR 操作，操作命令和参数在 SIP 消息体中定义。（16）MS 中 RCN 指示 RPN 放音，并做录音准备，然后向 APPS 回复 200 。（17）MS 中的 RPN 放音，并对用户终端发送的媒体流进行录音操作，操作结束后，通知 RCN，RCN 通过 INFO 信令，向 APPS 报告操作结果，结果中包括录音文件位置。（18）APPS 发 200 响应。PR 操作结束。

图 2.9 中（19）~（20）表示 IVR 操作结束后，APPS 发送 BYE，结束与 MS 的会话连接。

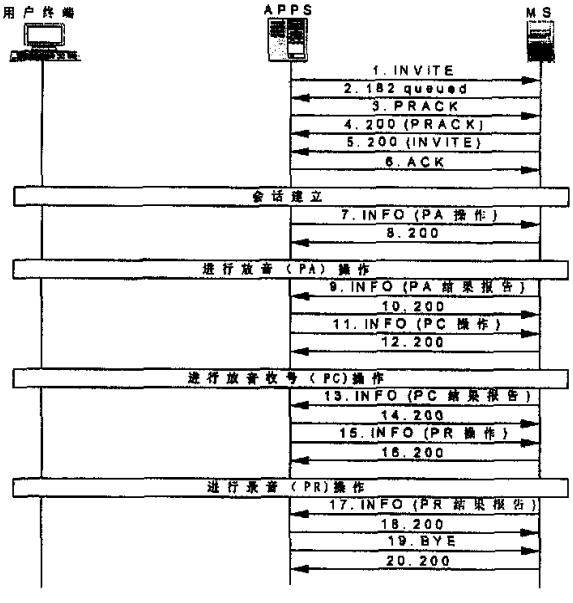

图 2.9　IVR 信令交互

2.5.3.2 语音资源的定义与组织

IVR 操作中涉及大量的语音资源，这些语音资源有些是简单的语音片断，有些则是较为复杂的包含了语音变量（如数字，地点，人名等）的语音结构，还可以通过选择符选择不同的语种，如中文、英文等，或方言，如普通话、粤语等。为了提供不同业务的提示语音，还需要按照具体的业务类型对语音资源进行定义和组织。具体提供语音资源时，根据指示，构造业务需要的语音片断，进行播放。

MS 中的语音资源按照服务类型、功能语音变量、语种等分类方法存放在 RPN 节点上。RCN 节点则通过配置文件的形式，记录语音资源在 RPN 上的具体位置。

RCN 上的配置文件，采用 XML 语言进行组织。例如预付费业务充值服务语音资源的配置信息组织如下：

```
<service prepaidcard_chongzhi>
    <mediafile directory=/opt/nms/services/prepaidcard_chongzhi >
        <welcome type=audio auid=tone101 aufilename=welcome.vce />
        <input_cardnum type=audio auid=tone102 aufilename=input_cardnum />
        <input_phonenum type=audio auid=tone103 aufilename=input_phonenum />
        <success type=audio auid=tone104 aufilename=success.vox />
        <fail_notenoughmoney type=audio auid=tone105 aufilename=fail_notenoughmoney.vox />
        <fail_wrongcardnumber type=audio auid=tone106 aufilename=fail_wrongcardnumber.vox />
        ... ...
    </mediafile>
        ... ...
</service prepaidcard_chongzhi>
```

虽然需要增加相应的解析工作，但由于 XML 语言强大的功能和结构化的组织方法，可以更好地管理复杂的语音资源配置信息，所以是值得采用的。

RCN 根据 APPS 的要求，将相应的语音需求转换为对应的语音 ID 和参数，然后通过 RCN 和 RPN 之间的内部通信协议传递给 RPN，由 RPN 进行实际的语音播放及其他操作。

2.5.3.3 IVR 操作中的消息及参数的定义、传送

RFC 2897[19]详细定义了用于 IVR 操作的消息，信号以及相应的参数，格式。协议中定义了 5 种事件，分别是 pa（ PlayAnnouncement，放音）、pc（PlayCollect，

放音收号）、pr（PlayRecord，放音录音）、es（EndSignal，正常结束）、oc（OperationComplete，操作成功）、of（OperationFailed，操作失败）。

对于每个事件，都有若干的参数，在 RFC 2897 中都有详细介绍，比如 pa 参数 an，给出要播放的语音文件的位置；pc、pr 都有的 ip 参数，给出在收号或录音前播放的提示音片断的位置；pc 的 dp 参数，给出收号的可接受格式等。

这些消息及参数以一定格式形成文本，可以放在 SIP INFO 信令的消息体中传递。例如图 2.9（11）中，APPS 通过 INFO 信令，向 MS 发出命令，要求 MS 向用户终端进行 PC 操作，INFO 消息体中可以按如下格式定义：

<aupc>
ip=tone88;rp=tone99;ni=false;dp=[1-2]xxx;

这段消息体的含义是 APPS 要求 MS 进行 PC 操作，操作参数 ip 给出需要播放的提示语音片断；rp 给出如果没有收到号码时需要播放的语音片断；ni 表示播放的语音片断是否允许打断，取值为 true 或 false，取 true 表示不可以打断，取 false 表示可以被打断；dp 给出判断收取号码合法性的数字映像。

图 2.9（13）是 MS 结束 PC 操作后，通过 INFO 信令将操作结果返回，INFO 消息体中的内容举例如下：

<auoc>
dc= “1234”；rc=100；na=2；

该消息体的含义是，返回 auoc，表示是成功响应，是否成功也可以看返回参数 rc 的值，100 表示成功，300-399 表示失败；dc 指示收到的数字号码；na 表示用户输入时，进行了两次尝试。

语音消息的解析在 RCN 中处理，RCN 通过配置文件，将所需操作及参数转换一下，并重新用与 RPN 的内部通信协议进行编码，发往 RPN。这样做的一个原因是 RCN 管理着若干 RPN，需要根据系统负载情况在 RPNs 之间进行负载均衡调度。语音资源则存放在共享存储器中，所以对它的访问须 RCN 来管理。

2.6 本章小结

通过对 NGN 移动环境中多媒体业务特点的研究和了解，以及我们对各种类似设备系统架构的比较研究，基于《基于软交换的媒体服务器技术要求》设备规范，提出了基于软交换的集群媒体服务器（SCMS）系统结构。接着对 SCMS 的控制协

议，H.248 与 SIP 协议进行了阐述。SIP 协议在 NGN 移动环境中有着非常重要的作用，也是基于软交换的媒体服务器中最主要的控制协议，基于一种商业化 SIP 协议栈，设计了用 SIP 及其扩展控制多媒体会话的方案，并给出 IVR（交互语音响应）的具体实现过程。该方案具有可靠、灵活、易扩展的特点。

参考文献

1. 中国通信标准化协会.基于软交换的媒体服务器技术要求[S]，2003
2. 邓玉辉，张江陵，冯丹．系统带宽可动态扩展的流媒体系统体系结构研究[J].小型微型计算机系统，Vol.25, No.7, July 2004.
3. Guang Tan, Hai Jin, Song Wu.Clustered multimedia servers: architectures and storage systems[R]. Annual Review of Scalable Computing, Vol.5, 2003, pp.92-132.
4. 吴松．高性能集群流媒体服务器系统结构及存储系统研究[R].华中科技大学博士学位论文,2002.
5. D H C Du, Y J Lee. Scalable server and storage architectures for video streaming[C]. In Proc: IEEE Int. Conf. Multimedia Computing and System, June 1999, pp.62-67.
6. 吴乃星，廖建新，徐鹏，朱晓民.一种基于软交换的集群媒体服务器的系统结构[J].电信科学，2004, 7：11-15.
7. ITU-T. H.248.1: Gateway control protocol [S]. March 2002.
8. N Greene, A Rayhan et al. "Megaco Protocol Version 1.0". IETF RFC 3015, November 2000.
9. M. Handley, H. Schulzrinne et al. “SIP: Session Initiation Protocol”. IETF RFC 2543, March 1999.
10. Rosenberg J， Schulzrinne H, Camarillo G. SIP: session initiation protocol. IETF RFC 3261, 2002.
11. M Handley, V Jacobson. "SDP: Session Description Protocol". IETF RFC 2327, April 1998.
12. R Fielding, J Gettys et al.,"Hypertext Transfer Protocol -- HTTP/1.1". IETF 2068, January 1997.

13. Jonathan B Postel."SIMPLE MAIL TRANSFER PROTOCOL". IETF RFC 821, August 1982.

14. S Donovan. "The SIP INFO Method". IETF RFC 2976, October 2000.

15. R Sparks. "The Session Initiation Protocol (SIP) Refer Method". IETF RFC 3515, April 2003.

16. David H Crocker. "STANDARD FOR THE FORMAT OF ARPA INTERNET TEXT MESSAGES". IETF RFC 822，August 1982.

17. Rosenberg J, Schulzrinne H., "An Offer/Answer Model with the Session Description Protocol (SDP)", RFC 3264, IETF, June 2002.

18. J. Rosenberg, H. Schulzrinne, "Reliability of Provisional Responses in the Session Initiation Protocol (SIP)", IETF RFC 3262, 2002.

19. D. Cromwell, "Proposal for an MGCP Advanced Audio Package", IETF RFC 2897, August 2000.

20. 杨景,王晓庆.软交换业务交易平台:下一代电信网络业务控制体系结构[J].电信科学,2002,7.

21. Rajkumar Buyya编，郑纬民等译. 高性能集群计算:结构与系统（第一卷）[M]. 北京:电子工业出版社，2001年6月.

22. 中华人民共和国信息产业部.基于软交换的应用服务器技术要求[S]，2003.

第三章. 基于软交换的媒体服务器网管系统的研究与实现

3.1 引言

网络管理系统是维护网络系统及设备正常运行的重要保证。现代网络设备必须具备支持网络管理的能力。网管子系统是基于软交换的集群媒体服务器（SCMS）的重要组成部分[1]。该子系统基于简单网络管理协议版本 2（SNMPv2，Simple Network Management Protocol version 2）设计，结合 IETF 相关标准及设备自身特点定义了管理信息库（MIB ,Management Information Base），可以满足网管中心对 SCMS 进行远程控制管理的需求。

本章以下内容是这样安排的，第 3.2 节简单介绍基于 SNMP 协议的网络管理及用于开发网管设备代理的软件包 Snmp++与 Agent++；第 3.3 节分析 SCMS 网管需求；第 3.4 节介绍 SCMS 网管子系统的设计与实现；第 3.5 节进行分析总结。

3.2 基于 SNMP 协议的网络管理及开发工具

3.2.1 基于 SNMP 协议的网络管理

SNMP 协议最初作为 TCP/IP 网络管理的临时解决方案被提出，它的优点是结构简单、容易实现，可以满足 TCP/IP 网络管理的基本需求，所以在实践中迅速得到广泛应用，成为 TCP/IP 网络管理的事实标准。随着 IP 网络的不断发展，尤其是近年下一代网络以 IP 为核心融合传统网络，建设全 IP 分组交换网络的思想，使得作为 IP 网络管理事实标准的 SNMP 协议的地位不断提高，越来越多的网络通信设备提供 SNMP 接口。

用 SNMP 协议进行网络管理的模型（见图 3.1）包含网管中心、网管代理、MIB、SNMP 协议栈。被管系统的可管理资源被抽象为一个个的管理对象，用抽象语法记法 1（ASN.1，Abstract Syntax Notation 1）[2-9]进行定义，管理对象的集合构成 MIB。基于 SNMP 协议的网管中心采用轮询方式向被管设备请求信息，被管设备中的网管代理负责接收网管中心发来的 SNMP 请求消息，根据 MIB 库的定义取得相应信息，再用 SNMP 消息回复给网管中心。网管代理也可以主动向网管中心发送 TRAP 消息，报告一些紧急事件。

国际标准化组织（ISO）定义了网络管理的 5 大功能域，包括性能管理、故障管理、配置管理、安全管理、计费管理。这些功能域规范了网管系统的实现，被

管系统可以根据这些功能域的要求建立 MIB 库，网管中心则根据这些功能域的要求建立管理应用，从而对网络进行有效的管理。下面，按照这 5 大功能域，分别讨论 SCMS 对网管的需求。

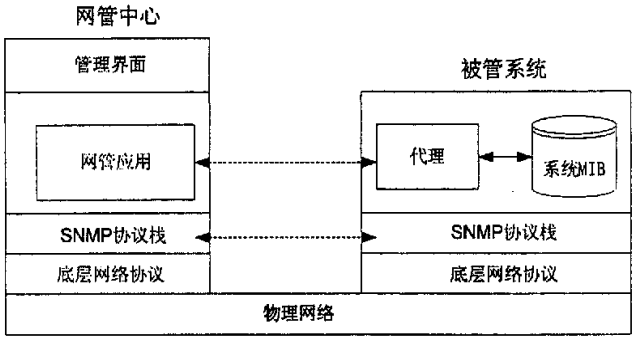

图 3.1 SNMP 网管系统模型

3.2.2 Snmp++与 Agent++软件包及其使用

SNMP++是一种基于面向对象技术，用C++语言开发，实现了SNMP协议栈，用于开发网管应用的应用编程接口（API，Application Programming Interface），具有易用、安全、可移植性与扩展性强的优点。SNMP++主要用来支持SNMP网管中心的开发，Agent++通过扩展SNMP++，可以同时支持SNMP网管中心和设备侧SNMP网管代理的开发。

SCMS的网管子系统主要是利用Agent++软件包，开发支持SNMP的网管代理，使管理者可以通过网管中心对SCMS系统进行远程管理。

使用 Agent++软件开发网管代理，首先需要获得 Agent++软件包[10]，并在开发平台上编译生成 snmp++和 agent++库；接着需要使用 agentgen(属于 agentpp 开源软件组工具——AgenPro 的一部分)编译 MIB 文件，生成代理存根（STUB）；然后在生成的代理存根中添加管理对象与实际系统资源的接口。最后将管理代理存根以及生成的 snmp++和 agent++库一起编译生成网管代理可执行文件。

3.3 SCMS 网管需求

3.3.1 SCMS 的故障管理需求

故障管理是指网管系统对网络中的问题或故障进行定位并修复的过程，一般

包括三个步骤，发现故障，定位故障，修复故障。通过故障管理，网络管理者可以及时地发现网络中的异常或故障，确定其所在位置，并采取正确的手段排除问题。作为电信运营级的设备，SCMS 有比较高的故障管理需求。

SCMS 应当在必要时或定期进行在线自检，检测设备的状态和故障，并通过告警系统，对监测到的异常情况（如硬件故障、系统资源紧张、通信状况不良、传输质量下降等）做出反应。告警系统应该按照故障的严重程度分类，一般至少应分为两大类，即紧急告警和非紧急告警。利用 SNMP 的 TRAP 消息，可以实现报警功能。

3.3.2 SCMS 的配置管理需求

配置管理的目的是掌握网络设备的构成和配置。配置管理应该提供设备的资源清单管理功能、资源供给功能、业务提供功能和网络拓扑服务功能。通过配置管理，可以对网络中设备的配置信息进行收集、修改和维护等管理活动，可以增强网络管理者对网络配置的控制。配置管理是通过对网络设备的配置数据进行访问实现的。

SCMS 的配置管理包括三部分，即设备资源的配置、业务资源的配置、系统自身正常工作的配置。设备资源的配置，包括系统的板卡及板卡资源；业务资源的配置，包括系统收号资源、录音资源、各种语音文件的配置；系统自身正常工作的配置，包括 RCN 和 RPN 相互通信管理的一些配置信息。

3.3.3 SCMS 的安全管理需求

安全管理通过控制对信息的访问权限，保护计算机网络中的敏感信息不被任意访问。在网络中存在众多的网络设备和主机，网络管理既要保证设备与主机彼此可以相互通信、访问资源，又要保护重要数据和个人信息不被非法访问。安全管理与其他管理功能有着密切的关系。安全管理的实施有赖于配置管理的信息，而安全管理在发现安全问题时，需要通过故障管理向网管中心告警，通知网管中心采取必要措施。

SCMS 应对管理员的访问权限做严格规定。管理员登录时要求帐户和密码，系统对每次访问做记录。根据管理的需要，系统可以根据权限大小对管理员进行分类，如系统管理员、配置管理员、维护管理员等等。

3.3.4 SCMS 的性能管理需求

性能管理的目的是通过对网络及网络设备进行性能监测，采集相关的性能统计数据，对它们进行性能分析，了解网络服务质量和运转效率。性能管理功能通常包括系统的性能监测功能、性能分析功能和性能管理控制功能。

SCMS 需要提供业务统计功能，以反映本设备的业务负荷信息和运行状况。还应具有业务量测量和记录功能以及协议统计功能。媒体服务器应该可以自动、实时地监视各种专用资源设备以及资源的状态，并对其使用情况进行统计。

3.3.5 SCMS 的计费管理需求

计费管理记录网络资源的使用，目的是控制和监测网络操作的费用和代价。一方面计费管理是实现网络经营者投资收益的保证，另一方面通过计费管理可以合理的规划网络资源，提高网络运营的效率。

计费管理的基础是对网络操作的详细记录，网管中心根据每个用户对网络资源的占用情况，对用户分别计费。因为性能管理已经对网络使用情况进行了详细的记录，SCMS 作为被管对象，不需要为计费管理做更多的工作。

3.4 网管子系统的设计与实现

3.4.1 SCMS 网管子系统系统结构

SCMS 的网管子系统涉及资源控制进程的资源管理模块(RMM)和网管信息记录模块(NMIRM)、网管代理进程、资源记录数据库、网管信息数据库等部分，见图 3.2。

资源控制进程是 RCN 的核心进程，它接受外部控制信令，根据系统资源情况进行资源分配，并控制 RCP 对外提供相应的媒体服务。网管代理进程处理网管中心的网管请求消息，根据请求返回相应数据。资源控制进程随时监视系统运行，出现意外或故障时，及时通知网管代理进程，后者通过 SNMP TRAP 消息通知网管中心。

资源记录数据库保存系统资源配置，包括 RCN 和 RPN 的数目以及它们之间的控制关系，系统语音资源，收号资源，所支持的媒体类型等，可以提供信息给配置管理，性能管理，计费管理。

网管信息数据库保存系统运行的动态数据，包括系统吞吐量、试呼次数、系统占用时间和次数、系统各种服务和功能的成功与失败记录等信息，可以提供信息给故障管理，性能管理，计费管理。

RMM 模块负责系统资源的分配。系统启动时，RMM 从资源记录数据库获得系统资源信息，系统运行时，RMM 跟踪资源的使用情况，并实时更新资源记录数据库。这样，网管代理进程可以通过资源记录数据库获得最新的系统资源情况。RMM 模块在分配资源时，会检查资源使用情况，当出现资源紧张或其他影响系统正常运行的情况时，就通过进程间通信的方法通知网管代理进程，后者通过 SNMP TRAP 消息通知网管中心。

NMIRM 模块在系统执行各种操作时，随时进行记录和统计，并将统计数据及时保存在网管信息数据库，供网管代理进程访问。

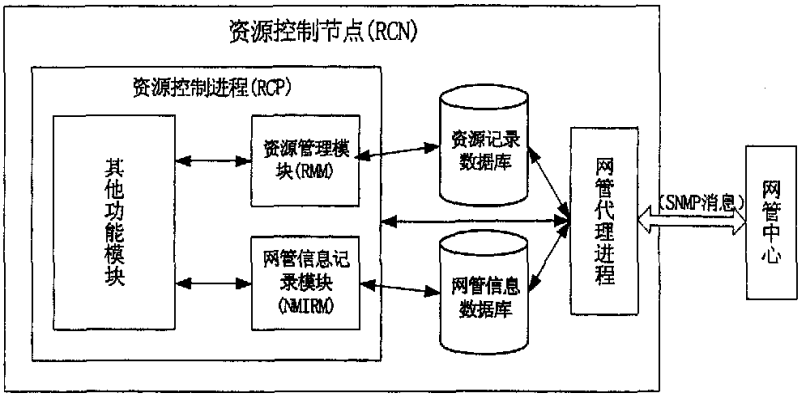

图 3.2 SCMS 网管子系统结构

3.4.2 SCMS 网管子系统 MIB 库设计

SNMP 规范中的管理信息架构(SMI, Structure of Management Information)[11]给出了定义和构造 MIB 的通用框架。它规定了 MIB 中数据的表示方法、组织方法、可以使用的数据类型等内容。

MIB 结构与对象都用 ASN.1 形式化地加以描述，说明 MIB 库的制定者、所引用的其他 MIB 库，所定义对象的标识符、数据类型、取值范围以及与其他对象的关系等内容。MIB 库中所有对象组织成一个树型结构。为了使 MIB 具有简单性和良好的可扩展性，MIB 可以使用的数据类型非常简单，只有 INTEGER, OBJECT INDENTIFIER, OCTET STRING, NULL, SEQUENCE, SEQUENCE OF 等几种。

根据需求，SCMS 系统 MIB 库设计为两个部分，一个是系统资源配置信息，

另一个是系统运行统计信息。系统资源配置信息满足配置管理的需求，将系统主要资源，分类存放在 MIB 库的多个表中，根据资源属性和管理的方便，这部分 MIB 信息，组织成 4 个表，分别是 medSvVoiceResTable（存放系统语音文件信息）、medSvResBoardsTable（存放系统资源板卡信息）、medSvDigitmapsTable （存放用于收号的 Digitmap 信息）、medSvRecordsTable（存放系统录音通知信息）。随着系统功能的扩充，这部分内容可以继续增加。

系统运行统计信息记录系统运行过程中的各种数据，这些数据本身很原始，基于这些数据，管理中心可以实现性能管理、计费管理等网管功能。这些统计信息定义在 MIB 库表 medSvStatsTable （存放系统全部统计信息）中。

设计好 MIB 库，按照 ASN.1 语法，写成文件 media-server-mib.txt。

3.4.3 网管系统的实现

3.4.3.1 RMM 模块的实现

资源控制进程运行时，凡是涉及资源调度和分配时，都需要调用 RMM 模块的功能函数。这种调用发生位置很多，为了保证系统资源分配的一致性，采用面向对象中单健模式[12]的方法定义类 TRmm，即整个进程中保留一个 TRmm 实例。

资源控制进程启动时，生成 TRmm 类的实例，同时调用 TRmm 类的启动函数 startup()，这个函数负责从资源记录数据库的文件形式（普通文件或数据库文件）中，将系统配置信息读入资源记录数据库对应的共享内存部分中。

TRmm 类的成员函数主要是负责分配各种系统资源的，包括各种标识符 ID（如 H.248 协议中的 Context ID，Transaction ID, Termination ID）的分配与回收，以及 RPN 上的各种媒体资源。

3.4.3.2 NMIRM 模块的实现

NMIRM 模块需要收集系统各部分的运行数据，所以被调用的位置很多，又因为对网管信息的统计需要一致性，所以 NMIRM 模块中最主要的类 TGetRcsInfo 也采用单健模式的方法定义。TGetRcsInfo 根据系统 MIB 定义，给出获取相应信息的接口函数。

TGetRcsInfo 类的唯一实例同样在资源控制进程启动时生成，实例指针可以被资源控制进程的所有函数访问，当资源控制进程执行特定操作时，调用相应的接口函数，将这些统计信息写入网管信息数据库对应的共享内存中，网管代理进程

可以访问这块共享内存，获取需要的统计信息。

3.4.3.3 资源记录数据库与网管信息数据库的实现

资源记录数据库采用外部文件和共享内存相结合的方法实现，外部文件可以是普通文件，也可以采用数据库管理的方式。目前实现的 SCMS 原型系统采用普通文件的形式实现。系统启动时，共享内存初始化，资源控制进程中的 RMM 模块将资源记录数据库从相应文件读入对应的共享内存，此时，该数据库内容完全存在于共享内存中。系统运行过程中对数据库内容的访问与实时更新都针对共享内存中信息。网管代理进程可以访问到共享内存信息，所以当网管中心请求关于系统资源状况的 MIB 信息时，网管代理进程可以通过访问这块共享内存，获得最新系统资源信息。

网管信息数据库也可以采用外部文件和共享内存相结合的方法实现，在系统故障或即将停止运行时，数据库内容可以从共享内存保存到外部文件中保存，供系统分析与诊断使用。不过，目前 SCMS 原型系统只采用共享内存的形式保存网管信息数据库内容。系统启动时，初始化网管信息数据库对应的共享内存。NMIRM 模块中的 TGetRcsInfo 类实例在需要记录的系统行为发生时，访问共享内存，更新相应的统计信息。网管代理进程在接到网管中心请求时，访问共享内存，获得相应的统计信息，返回给网管中心。

3.4.3.4 网管代理进程的实现

网管代理进程的实现基于 Agent++软件包。从[5]得到源代码后，需要在所用的开发平台上编译链接生成对应于该平台的 snmp++和 agent++库；接着要用 agentgen 工具编译 MIB 文件 media-server-mib.txt 生成 MIB 库和代理存根文件 media_server_mib.cpp 、 media_server_mib.h；然后需要修改代理主程序 agent.cpp，例如增加初始化资源记录数据库与网管信息数据库对应的共享内存的代码。

接着修改存根代码 media_server_mib.cpp 和 media_server_mib.h，需要给每个 MIB 表对应的类增加 update() 函数，在其中添加与资源记录数据库和网管信息数据库数据相对应的 MIB 对象的接口代码。

最后，将 *agent.cpp*、*media_server_mib.cpp* 与 *snmp++*库、agent++库一起编译链接在一起生成网管代理进程。

3.5 本章小结

用上述方法实现 SCMS 网管子系统后，运行系统，并启动网管代理，网管中心可以通过对象标识符取得所需数据，数据无误，正确反应了系统即时运行状态。

目前设计的 MIB 库，可以满足网络管理各管理域的基本需求。系统测试时，网管中心通过实时监测与系统性能相关的数据，可以及时了解系统性能的各项指标；对于故障管理，目前实现的不是很充分，对于一些明显的故障，可以产生报警；配置管理的数据，网管中心可以随时获得，对于配置数据的动态更新，可以实现，但是对于目前的原型系统，测试还不是很充分；安全管理需要和媒体服务器整体的安全解决方案相配合，所以目前只实现了简单的安全机制；对于计费管理，可以给出基本的统计信息。

综上，SCMS 网管子系统的设计与实现可以满足目前原型系统的网管需求，随着系统本身的进一步完善，在目前实现的基础上，网管子系统也会进一步改进，完善。

参考文献

1. 中华人民共和国信息产业部.基于软交换的媒体服务器技术要求[S]，2003
2. ITU-T Rec. X.680 (2002). Abstract Syntax Notation One (ASN.1) Specification of Basic Notation
3. ITU-T Rec. X.681 (2002).Abstract Syntax Notation One (ASN.1) Information Object Specification.
4. ITU-T Rec. X.682 (2002). Abstract Syntax Notation One (ASN.1) Constraint Specification.
5. ITU-T Rec. X.683 (2002). Abstract Syntax Notation One (ASN.1) Parameterization of ASN.1 Specifications.
6. ITU-T Rec. X.690 (2002). ASN.1 encoding rules Specification of Basic Encoding Rules (BER), Canonical Encoding Rules (CER) and Distinguished Encoding Rules (DER).
7. ITU-T Rec. X.691 (2002).ASN.1 encoding rules Specification of Packed Encoding Rules (PER).

8. ITU-T Rec. X.692 (2002).ASN.1 encoding rules Specification of Encoding Control Notation (ECN).

9. ITU-T Rec. X.693 (2001).ASN.1 encoding rules XML encoding rules (XER).

10. Agent++ - SNMPv1/v2c/v3 Agent API for C++. http://www.agentpp.com.

11. J. Case，K. McCloghrie et al., Structure of Management Information for Version 2 of the Simple Network Management Protocol (SNMPv2), RFC 1902, 1996.

12. Cay Horstmann著.张琛恩译.面向对象的设计与模式.北京:电子工业出版社, 2004.

第四章. NGN 安全与基于软交换的媒体服务器的安全问题

4.1 引言

随着技术的发展，以及人们物质、文化需求的日益增长，电信服务也朝着广度和深度发展。NGN 因为其独特的优越性，能够适应需求的发展，在电信业未来的发展中有着举足轻重的作用。但是因为电信业务日益开放的趋势和 NGN 本身特殊的网络架构，开放的网络环境和允许第三方业务提供商接入等特点，使得 NGN 面临着比传统电信服务大得多的安全威胁。

安全问题从 NGN 概念提出之日起，就是一个受到关注和热烈讨论的重要问题[1][2][3][4]。NGN 概念源于互联网，而安全问题一直是困扰互联网，使互联网难以发挥更好作用，甚至在某种意义上威胁互联网未来的一个大问题。互联网的优势何在？在于其开放性，在于其端到端智能，以及其应用的多样性。可是恰恰因为开放性，使得参与者众多且不可控，带来了安全隐患；端到端智能，更是颠覆了传统电信网“网络智能化，终端傻瓜化”的理念，使得业务控制智能和网络控制智能决定于终端，终端用户隐秘而复杂，也带来安全隐患；应用多样性，应用开发者不再仅仅局限于少数人，少数大型企业，理论上所有的网络用户都可以开发自己的应用，这使得应用本身的缺陷，以及恶意使用者故意制造的不良程序在网络中散步，这就带来更大的安全隐患。

NGN 的提出，为的是借鉴因特网成功经验，但是“鱼与熊掌”不可兼得，如上所述，因特网的优点与其带来的安全隐患几乎就是结伴而来。想利用因特网的成功之处，就不得不面对其先天所具有的安全隐患。所幸的是，正如因特网安全问题在人们的努力下，已经取得一些成功经验，NGN 安全问题一方面因为借鉴因特网安全研究取得的成功而有所解决，另一方面 NGN 因为不同于因特网网络架构的一些方面，也使得 NGN 中的安全问题的解决有着因特网不可比拟的一些优势。解决的思路是：1）NGN 虽然开放，但又不能完全开放，NGN 必须是可运营的网络，运营商对网络要加以有效控制，核心网只能通过一定的控制手段，如 API 方式开放；2）终端虽然智能，但是网络也不能完全变傻，智能在网络与终端必须合理分配，取得平衡；3）业务多样性虽然重要，但是业务却不会随意进入网络，业务提供商准入是受限的。

为了保证 NGN 网络和业务的安全型，NGN 设备必须充分考虑安全问题。本章分析 NGN 安全问题，并重点针对 SIP 协议的安全问题进行了分析，结合 SCMS 情况，对 SCMS 可能面临的安全威胁进行分析，并给出可行的解决方案。

4.2 NGN 安全问题

4.2.1 NGN 安全威胁

NGN 安全对不同角色都很重要：1）对电信用户而言，他们需要对使用的网络和网络提供的业务产生信任感，否则网络再好，也无人敢用。此外，网络应该是可靠的，公平竞争（不被垄断）的，并且用户的隐私应受到保护。2）运营商、业务提供商以及接入服务商需要安全手段保护他们的商业利益，比如对提供服务的收费、防止盗用等。他们需要安全服务去尽他们对用户和社会应尽的义务。3）对国家机构而言，不安全的电信服务无异于灾难，国家需要用立法和行政的手段保障安全，惩罚犯罪。

NGN 安全问题的来源有这么几个方面：1）NGN 的开放，目的是提供用户选择业务和业务提供商的自由，最终，一个网络应用可能基于多个提供者的服务，为了更好地使用业务，用户必须有一定途径对网络/业务的管理工具进行监视，控制，并对自己的业务和使用的网络资源进行裁剪。即用户必须有进入网络管理设施的接口。允许用户接入管理设施，相应的安全管理非常重要，既要保护用户数据的安全，又要保证网络和应用的安全，可靠，稳定。2）黑客和内部破坏，因特网环境活跃着黑客，他们利用各种工具发现计算机系统的漏洞，非法进入系统，获得机密内容或进行其他破坏活动。运营商和业务提供商内部人员良莠不齐，或者出于故意，或者出于操作不当，可能带给网络安全问题。3）协议开放，使得更多的人有机会了解网络的实现和运行细节。和早期的电信设备及服务不同，因特网及NGN协议都是开放的，有一定技术背景的人都可以轻松获得并了解协议细节，借助一些容易获得的工具，很容易成为新的黑客，这就增加了系统受威胁，并遭受入侵的可能性。4）新业务对安全的更高要求，传统电信业务主要是语音业务，保护用户通信主要是防止窃听，而 NGN 中的业务主要是数据业务，包括对数据安全性要求特别高的涉及金融的业务，这些新业务对安全有着比语音通信更高的要求。5）NGN 移动环境中，业务和用户都涉及移动问题，不再局限于某一个地方，这也给维护网络和应用的安全以及整个系统的一致性带来了挑战。

通过对网络进行进行威胁分析和风险评估，是采取正确的安全措施的前提，从攻击发生的层次分，可以分为对底层协议的攻击和对高层协议的攻击[5]：

对底层网络协议的攻击：针对从物理层到传输层的网络攻击，如，针对链路层、IP、TCP、UDP 或 SCTP 协议等的攻击。攻击者可以采取破坏物理传输链路，干扰传输线路，以及其他如 IP 劫持，TCP 劫持等主动攻击方法，也可以采取分析

网络流量，分析协议报头等被动攻击手段，威胁网络正常运行。

对高层应用协议的攻击：这是针对 NGN 应用协议的攻击，如 SIP、H.323、MEGACO、COPS，也包括因特网传统的如 HTTP、SMTP 等协议。这些攻击针对特定目标协议的。

从攻击的具体形式上看，可以有以下这些类型[4][6]：

拒绝服务及放大流量攻击 拒绝服务攻击（DoS, Denial of Service）的目的是使得某个网络部件失去作用，从而不能对外提供服务。分布式拒绝服务攻击(DDoS,Distributed DoS)是攻击者通过某种手段，使得大量的主机同时向被攻击主机发送大量数据，造成被攻击者资源耗尽，丧失服务能力。

窃听 指对通话内容进行截取，从而了解通信双方正在进行的会话，并掌握用户身份、进行的服务，网络拓扑等信息。

伪装 攻击者通过一定手段，伪装成某个用户的身份，或者伪装成服务器，从而窃取用户秘密信息，或进行其他违法活动。

非授权访问 访问网络设施必须被严格限制并符合安全策略，如果攻击者获得非授权访问，可以进行各种破坏行为，非授权访问可以说是多数攻击行为的一个阶段目的和进一步攻击的起点。

信息篡改 信息被篡改后，服务将不可用，结果是系统完整性受到破坏，对于用户授权范围内的数据操作应该是允许的，但是也要有必要的防范措施，防止用户的不当操作。

否认参与 参与会话的一个或多个用户否认参与过会话的全过程或部分过程，他/他们可能拒绝承认传输过某信息，访问或修改过某个数据。从服务提供者的角度，这种行为会导致利润流失，失去用户信任从而失去用户。对客户而言，这会带来很多法律纠纷和商业利益的损失。

4.2.2 安全服务

针对 NGN 的安全问题，有相应的安全服务。提到安全服务，一般包括机密性（confidentition）保护、完整性（integrity）保护、认证（authentication）、不可否认性（non-repudiation）、访问控制（Access Control）等几个方面[6]。

机密性保护确保被传输的数据不被非法访问。机密性一般通过加密的方法获得。加密算法有两类，一类对称密钥算法，加密解密密钥相同，优点是运算较快，

但是密钥分发不易；一类是非对称密钥算法，加密解密的密钥不同，用其中一个加密，用另一个解密，它的优点是一个密钥可以公开，使密钥分发的问题得到解决，但运算速度较慢。

完整性保护确保接收到的消息如同发送的消息一样，没有冗余、插入、篡改、重排序或延迟。完整性一般通过报文加密、计算报文鉴别码（MAC，Message Authentication Code）、用散列函数计算消息摘要的方法进行验证。报文加密是用双方都知道某一个密钥 K，一方用密钥 K 对某个信息加密，另一方解密，假设密钥不会泄露，那么可以用 K 解密，就证明该密文是由知道 K 的一方产生，未被他人修改。报文鉴别码（MAC）使用特殊的 MAC 函数和密钥，产生定长数据分组，并附着在报文中，接受方收到 MAC 后，用同样的 MAC 函数和密钥，重新计算 MAC 值，如果该 MAC 值与报文中附着的 MAC 值相同，则可以相信报文未被非法修改。与加密报文不同，MAC 函数一般是单向函数，而加解密函数必须是可逆的，这使得 MAC 函数比加解密函数更不易破解。散列函数是 MAC 函数的一个变种，它以整个报文 M 作为输入，产生定长的散列值 H（M），也叫消息摘要，作为输出。散列值是报文所有比特的函数值，并有差错检测能力，因为报文中任一比特的改变都会使产生的散列码发生改变。所以通过传输散列值（可加密），接受方通过计算比较，可以确定消息是否被修改。

认证确保通信是可信的，是对通信参与者身份的确认。单个消息情况下，认证服务的功能是使接受者能够相信消息来自其所申称的源。在双工通信的情况下，认证服务一方面在连接发起时，保证通信双方可信（即双方确是彼此所申称的实体）；另一方面认证服务确保连接不被干扰，第三方不能假冒通信双方中任何一个参与到通信中。

不可否认性是指通信双方不能对发送或接收的消息进行抵赖。收到信息，接受方确认发送信息的是某发送方，发送方不能不承认，接受方收到信息，予以确认后，事后也无法抵赖。实现不可否认性，需要用到数字签名（digital signature）技术，公钥加密系统中，有两个密钥，一个是公钥 PK，一个是私钥 SK，PK 可以公开得到，SK 则保密，发信者在发送消息时，用 SK 对整个报文或报文摘要进行加密，接受者收到这个加密信息后，用 PK 解密，如成功，则可确认报文来自发送者，因为除发送者外，无人知道 SK，也就相当于发送者对发送信息进行了签名。

访问控制是限制和控制经通信链路对主机系统或应用程序进行访问的能力。为取得访问权限，每个试图访问的实体需首先进行身份认证，再根据授权取得一定权限的访问能力。

这些安全服务，往往组合起来，根据系统的安全需求，制订一定的安全策略，获得一定的安全性。值得注意的是，任何安全服务都要付出一定的代价，而取得安全性的强弱往往与代价成正比，所以根据具体的环境，根据安全性要求和代价取得折中是安全分析需要重点考虑的问题。

4.3 协议安全问题与 SIP 安全性

攻击往往针对协议，SIP 协议是 NGN 会话管理中最重要的协议，3GPP 也采用 SIP 作为 IP 多媒体核心网（IMS）的会话控制协议，SIP 协议也是基于软交换的媒体服务器的主要控制协议之一，所以对 SIP 安全性的研究非常重要。

SIP 消息包含的信息，有些是用户和服务器希望保密的。SIP 消息头常常会泄露有关通信模式、通信者个人信息等私密信息。SIP 消息体中也会包含诸如媒体类型、编码方式、地址、端口之类的敏感信息。所以保护 SIP 消息头和消息体是非常重要的，否则信息泄露不但会给通信者带来隐私泄漏的损失，也有可能使正常通信遭到攻击。

针对 SIP 协议的安全服务可以分两大部分[7][8]，一类是端到端（end-to-end），一类是点到点（hop-to-hop）。端到端安全服务由主被叫用户代理实现，利用的是 SIP 自身提供的安全手段。点到点则不然，它保护的是相邻 SIP 实体（主要是出于通信链路中间的代理服务器、重定向服务期、注册服务器等）之间的通信安全，SIP 协议自己没有提供相应的安全手段，需要利用网络层的 IPSec 和传输层的 TLS 等安全服务来实现。因为 SIP 通信中，中间服务器起着重要的路由作用，所以端到端安全不能施加于中间服务器读取的消息部分，否则消息将不能被正确路由。

下面分别从端到端安全和点到点安全来说明 SIP 安全服务的实现。

- 端到端安全

SIP 端到端认证，可以通过类似于 HTTP 摘要认证（Digest Authentication）的摘要方法实现，SIP 摘要认证使服务器或 UAS 在收到 INVITE、REGISTER 等请求时，可以质询（Challenge）请求发送者，在对方提供响应（Response）并得到身份证实后，会话或操作可继续进行。举例如下（图 4.1）：

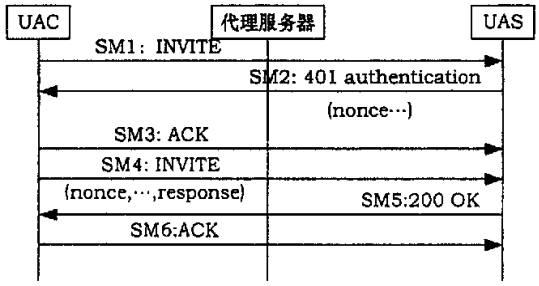

图 4.1 SIP Proxy 认证

SM1:INVITE sip:bob@sip.bupt.edu SIP/2.0

…

SM2: SIP 2.0 401 Unauthorized

…

WWW-Authenticate: Digest

realm="sip.bupt.edu.cn",

qop="auth,auth-int",

nonce="dcd98b7102dd2f0e8b11d0f600bfb0c093",

opaque="5ccc069c403ebaf9f0171e9517f40e41"

SM4:INVITE sip:bob@sip.bupt.edu SIP/2.0

…

Authorization: Digest username="bob",

realm="sip.bupt.edu.cn",

nonce="dcd98b7102dd2f0e8b11d0f600bfb0c093",

uri="sip:bob@sip.bupt.edu.cn",

qop=auth,

nc=00000001,

cnonce="0a4f113b",

response="6629fae49393a05397450978507c4ef1",

opaque="5ccc069c403ebaf9f0171e9517f40e41"

UAC 发送 SM1 发起呼叫，UAS 通过 SM2 401 响应，要求 UAC 进行认证，头域 WWW-Authenticate 中的信息含义是，Digest 指示使用摘要认证，realm 指出要求认证的域，nonce 是服务器端给出的“现时”，防止重放攻击，opaque 域的值，在 UAC 发回认证信息时，须保持不变。UAC 收到 401 响应后，根据 WWW-Authenticate 头域中的相关指示，用摘要算法（缺省为 MD5）计算包含 username, 用户 password, nonce，SIP 方法等信息的散列值，构造新的 INVITE 消

息(SM4),其中包含头域 Authorization,该头域中有用户名 username、认证域 realm,nonce 等变量，还有一个 response 变量，它的值就是所计算的散列值。然后发送到 UAS，UAS 收到包含新的 INVITE 消息后，根据用户名，用户口令（具体实现中也可不保存口令，而保存某个中间计算值）等信息用同样的方法计算散列值，与 Authorization 头域中的 response 值比较，如果相等，用户认证则通过。通过这样的方法，服务器确认了 UAC 的身份，同时避免传递密钥可能造成的泄漏。

S/MIME(Secure/Multipurpose Internet Mail Extensions，安全多用途因特网邮件扩充),基于RSA技术,提供一种发送/接收安全MIME数据的方式。由IETF S/MIME Mail Security 组负责制定，包括一系列的标准。 S/MIME 可以提供的安全服务有：1）加密数据，利用选定的对称密钥加密算法加密数据，同时利用接收者的公钥加密密钥，连同加密者的公开密钥一起形成 envelopedData；2）签名数据，用散列函数，MD5 或 SHA-1，计算数据报文的摘要，然后用发送者的私钥对摘要签名，再对原始内容和签名都用 RADIX-64 编码；3） 清澈签名数据（Clear-Signed），即形成报文摘要的数字签名后，只有数字签名部分使用 radix-64 编码，内容保持原始状态；4）签名并加密数据，即加密的数据可以签名，签名或清澈签名的数据可以被加密的服务形式。

S/MIME 应用于 SIP 协议时，扩展了 message/sip 类型。隧道 SIP 方式用来提供 SIP 头域的完整性和机密性。这时，整个 SIP 报文用 message/sip 格式进行加密。中间设备需要访问的头域信息，如 From、To、Via 有相应的外部副本，且可以被修改，接受方验证时，若这些头域出现内外不一致时，也可以接收。利用这种方式 S/MIME 可以提供一定程度的端到端认证和加密。使用 S/MIME 的问题是目前缺乏有效部署的公钥基础设施，很多情况下，需要报文自己携带证书，增大了中间人攻击的机会。

➢ 点到点安全

SIP 中没有相应的点到点安全机制，实行点到点安全需要利用网络层的 IPSec[9][10][11]或传输层的 TLS[12]来实现。

IPSec 提供局域网、专用和公用广域网及因特网上安全通信的能力。它能够加密和/或认证 IP 层的所有通信量。IPSec 在传输层以下，对应用程序透明，使得应用在获得安全性的同时不必做大的修改。使用 IPSec 可以实现认证和加密的安全服务。IPSec 由两个子协议构成，一个是认证首部（AH，Authentication Header）[10]，一个是封装安全有效载荷（ESP，Encapsulating Security Loading）[11]。前者可以提供认证服务，后者可以提供认证和/或加密服务。TLS 提供传输层安全服务，实现

加密服务或报文完整性检查服务。它只能用于可靠传输层协议，如 TCP，不能用于 UDP。

用于 SIP 环境时，IPSec 通常用于 SIP 实体之间已经有预先建立静态安全关联（SA）的情况，而 TLS 用于需要动态建立安全关联的情况。另外，在利用 IPSec 或 TLS 进行点到点安全服务时，只能保证两点之间的安全，不能保证信令传输其余部分的安全。

4.4 SCMS 安全分析与解决方案

4.4.1 SCMS 安全分析

以下分析在涉及的控制协议为 SIP 协议。

通过对 SCMS 使用环境的分析，SCMS 可能遇到的安全威胁有：

1）非授权访问。SCMS 受软交换机/应用服务器控制，在 SIP 协议中，只能根据 FROM 头域中的申明来确定请求者的身份，这就给了攻击者伪装成控制设备的机会。攻击者可以构造假的请求消息，在 FROM 域中填写 SCMS 可以接受的控制设备的 URI，从而使 SCMS 受骗。防止这种攻击，需要 SCMS 具备对控制设备的认证机制。在 SCMS 与控制设备之间应预先分配密钥，在控制设备控制 SCMS 时，SCMS 应对控制设备进行摘要认证。

2）注册劫持。SIP 协议中，管理域中的各个设备都必须通过 Registrar 进行登记，只有这样，才能被其他设备访问到。注册时，用 REGISTER 命令，FROM 中是注册消息发送者，TO 中是登记的名字（如 name@host），CONTACT 中是所登记名字的当前联系地址。SIP 允许第三方注册，因此存在这样的危险：攻击者伪装成有权利进行第三方登记的身份，将 SCMS（或其他设备或用户）的真实地址修改为伪地址，使得服务无法提供，或提供假服务；攻击者也可以破坏原有注册信息，使得正常的位置服务瘫痪。防止这种攻击，需要建立安全通道，并对服务方进行认证。

3）伪装 SCMS 欺骗。控制设备请求建立会话时，需在 Request-URI 中指明 SCMS 的 URI，或 IP 地址，如果是 URI，则请求在到达 SCMS 所在域时，域边界的 Proxy，会询问本地位置服务器，获得 SCMS 的具体地址。如果攻击者利用上诉的注册劫持，将 SCMS 的注册信息篡改成伪装的 SCMS，则控制设备请求的将是伪装的 SCMS。攻击者也可以伪装成域中的重定向服务器（Redirector），当控制设备通过

重定向模式访问 SCMS 时，重定向服务器返回 301（Moved Permanently），同时在 Contact 域中返回自己想要指定的地址（视攻击目的而定），也可以使伪装的 SCMS 代替真实设备工作。对付这种攻击，需要控制设备具有对 SCMS 进行认证的能力，也就是说通信双方应该有双向认证的能力。

4）篡改消息体。通常情况下，控制设备和 SCMS 之间会有一些中间设备（Proxy 或 Redirector）。这些中间设备一般是可信的，但并不希望它们检查甚至修改消息体。消息体中可能有通信双方协商的会话密钥，而解密通信会话并不是通信双方希望中间代理服务器做的事情。如果代理服务器本身存在恶意，它甚至可以修改会话密钥，或其他所请求的安全属性。避免这种攻击，可以使用端到端的消息体加密服务。

5）会话异常终止。SIP 协议中会话建立后，可通过 re-INVITE，修改会话属性，会话需要结束时，通信任一方可通过发送 BYE 命令结束会话。控制设备与 SCMS 建立会话过程后，需要防止攻击者伪造 re-INVITE 命令，修改会话参数，从而降低会话安全性，或发送伪造来自某一方的 BYE 信令，使会话异常中止。反击这类攻击，有两个思路，一是对建立会话之后所发送的命令进行有效认证，确认 BYE 或 re-INVITE 的发送者是正确的对象；二是通过加密会话参数，使得攻击者失去伪造消息的可能。

6）拒绝服务攻击。SCMS 要服务大量用户，而 SIP 协议存在一些潜在的可被利用来进行分布式拒绝服务攻击（DDoS，Distributed Denial of Service，是攻击者通过某种手段，使得大量的主机同时向被攻击主机发送大量数据，造成被攻击者资源耗尽，丧失服务能力。）的机制。例如攻击者可以构造这样一种 SIP 请求消息，使用一个假的源地址，并在 Via 头域中给出 SCMS，然后将这样的请求消息发给大量的主机，这些主机都会回复该请求，并且都会路由到 Via 中给出的主机，从而造成该主机大量的流量，难以提供有效服务。类似的，攻击者也可以利用 Route 头域和 Record-Route 头域达到同样的目的。攻击者也可以通过注册大量的信息，耗尽 Registrar 的内存，实现拒绝服务。对付这类攻击，一方面也要借助加密和认证手段，另一方面，还要借助于一些策略定义的方法。

4.4.2 安全解决方案

针对以上对 SCMS 可能遇到的安全威胁的分析，及对 SIP 协议可利用的各种安全服务的研究，我们设计了 SCMS 安全解决方案。该安全解决方案主要包括这样几个部分：

1）控制设备与 SCMS 处于同管理域

SCMS 所在域注册服务器（Registrar）保存有域内设备，包括 SCMS、应用服务器，代理服务器的安全密钥。SCMS 加电启动时，向所在的管理域的注册服务器（Registrar）进行注册，注册方法是 SIP Digest 认证，避免注册劫持的发生，从而避免服务器伪装的发生。同样，域内的应用服务器也要向注册服务器进行 Digest 认证。在代理服务器（每个域至少有一个出口代理服务器，Outbound Proxy）与注册服务器、应用服务器、SCMS 之间建立 TLS 安全连接。这样可以保证各服务器之间的 SIP 信令交互的保密性。安全服务的示意图见图 4.2。

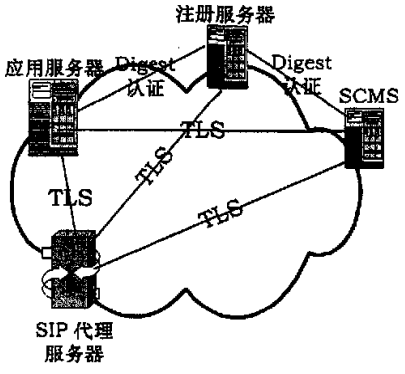

图 4.2 SCMS 安全解决方案(1)

2）控制设备与 SCMS 不在同一管理域

为使 SCMS 提供媒体服务具备更大灵活性，提高效率，控制 SCMS 的控制设备（如应用服务器）与 SCMS 可能不在同一个管理域（不属于同一运营商或一地），这时对应用服务器的认证要借助其所在域的代理服务器，而各域出口代理服务器之间必须建立 TLS 安全连接。在 SCMS 与所在域的出口代理服务器之间也存在 TLS 安全连接，从而保证 SIP 信令的安全。安全服务的示意图见图 4.3。

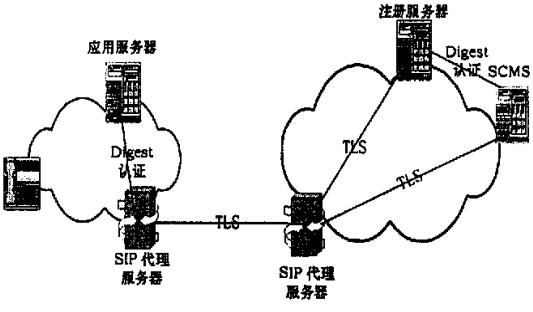

图 4.3 SCMS 安全解决方案(2)

3）其他安全措施

对于可能发生的拒绝服务攻击，可以在出口代理服务中实现部分应用网关防火墙的功能，或在出口代理服务器前增加专用的防火墙，设置相应的策略，在请求异常增多的情况下，采取措施，不再转发呼叫。

4.5 本章小结

NGN 环境中，安全至关重要。本章对 NGN 安全问题发生的原因，NGN 可能遇到的安全威胁及相应的安全服务进行了说明。因为安全问题的发生大多针对具体的协议，因为 SIP 协议是重要的 NGN 协议，尤其是，SIP 是 SCMS 最重要的控制协议，所以针对 SIP 协议的安全进行了分析说明。在此基础上，详细分析了 SCMS 所面临的各种安全威胁，并给出了可行的安全解决方案。安全问题的攻防是动态变化的一对矛盾。对 NGN 安全性的研究和相关设备安全方案的改进仍将是我们工作的重点。

参考文献

1. 蒋林涛.下一代网安全技术的研究. http://technology.cttl.com.cn.
2. 徐鹏，廖建新.以软交换为核心的下一代网络的发展前景及其在发展中可能面临的问题[C].第九届全国青年通信学术会议论文集.北京：电子工业出版社,2004.755-759.
3. 魏亮，惠亮. NGN面临的安全威胁与应对原则[J].通信世界,2004, 126.
4. B. Gamm, B. Howard, et al. Security features required in an NGN [J]. Alcatel Telecommunications Review, 2001, 2nd Quarter.
5. 孙文建.NGN网络安全问题分析和对策.现代通信,2005,3.
6. 杨明，胥光辉等译.密码编码学与网络安全：原理与实践（第二版）[M].北京：电子工业出版社, 2001.
7. Rosenberg J，Schulzrinne H, Camarillo G. "SIP: session initiation protocol". IETF RFC 3261, 2002.
8. Salsano S, Veltri L, Papalilo D. "SIP security issues: the SIP authentication procedure and its processing load" IEEE Network 2002, 16(6):38-44.
9. S Kent, R Atkinson. "Security Architecture for the Internet Protocol". IETF RFC 2401. November 1998.

10. S Kent, R Atkinson. "IP Authentication Header". IETF RFC 2402. November 1998.

11. S Kent, R Atkinson. "IP Encapsulating Security Payload (ESP)". IETF RFC 2406. November 1998.

12. T Dierks, C Allen."The TLS Protocol Version 1.0". IETF RFC 2246. January 1999.

13. 3GPP TS 33.102 v5.4.0, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3G Security; Security Architecture (Release 5), 2004.

14. A. Niemi, et.al., Hypertext Transfer Protocol (HTTP) Digest Authentication Using Authentication and Key Agreement (AKA), IETF RFC 3310,2002.

第五章. NGN 多层多协议移动性管理框架研究

5.1 引言

移动通信的广泛应用，使得人们的信息交流更加自由和灵活，脱离了固定通信设备对人们通信场所的约束。但是，传统的移动通信网和传统的固定电信网一样，基于电路交换技术。因为电路交换固有的资源利用率低，成本高，不灵活等问题，在 NGN 中，移动通信同样会逐步过渡到完全基于分组交换技术[1][2]。

从发展的角度来看，在 NGN 中，固定电话、移动电话、有线电视、各种数据业务将融合形成统一的 IP 核心网，它们的区别只是接入方式不同。NGN 中的移动也不仅仅是 2G、3G 移动网的升级，而是包括了多种接入方式，多种移动服务模式的含义更丰富的移动环境。事实上，移动环境已经成为 NGN 研究中的重要领域[3][4][5][6]。移动性管理是移动通信中最重要的研究内容，在 NGN 中，移动性管理与以往有很大的不同。

2G 和目前的 3G 网络中，移动性管理主要指的是移动终端的管理，它有两个含义：一是切换管理，一是位置管理。前者指的是当终端移动时，正在进行的会话不会被中断；后者则是指系统跟踪和定位移动终端（处于空闲状态），并在呼叫到来时，和该移动终端建立呼叫。随着移动通信技术的发展和用户对移动通信业务需求的发展，人们对未来 NGN 移动性的概念也发生了变化，移动性管理不再仅仅指移动终端的移动，还包括个人移动性、服务移动性、会话移动性、模式移动性等多种移动形式。相应的，移动性管理也不再只是对移动终端所涉及的切换和位置的管理，而是要满足上述广泛移动概念的管理。

本章后续内容安排如下，5.2 节详细介绍 NGN 中各种移动性，包括个人移动性、服务移动性、会话移动性及其他移动类型；5.3 节介绍实现各种移动性的技术手段——移动 IP 和 SIP，并讨论了它们各自的优缺点；5.4 节提出观点，只有结合移动 IP、SIP 协议，实行分层管理的方法才能有效实现多种移动性的管理，并给出一个多协议多层次的移动性管理框架。5.5 节是结论。

5.2 NGN 中的移动性

5.2.1 个人移动性

个人移动性指的是用户拥有一个逻辑标识，通过这个逻辑标识，可以定位到

该用户的若干个不同的通信终端，比如固定电话、移动电话、电子邮箱和语音信箱等。逻辑标识到通信终端的映射可以是一对多的（一个标识对应多个终端），也可以是多对一的（多个标识对应一个终端）[3]。用户可以根据不同时间，不同地点设定想要收到呼叫的通信终端。例如用户 A 有一个移动电话、一个工作电话和一个笔记本电脑。利用 SIP 派生代理（fork proxy）技术，用户可以设定呼叫同时到达所有注册终端，也可以设定根据时间段将呼叫接入到不同终端，如在白天工作时间呼叫接入到工作电话，晚上 8 点后，呼叫接入到笔记本电脑的 SIP 电话上，其它时间呼叫接入到移动电话上。当有人呼叫用户的"逻辑标识"时，网络可以根据用户的设定自动转移到不同的终端设备，如图 5.1 所示。

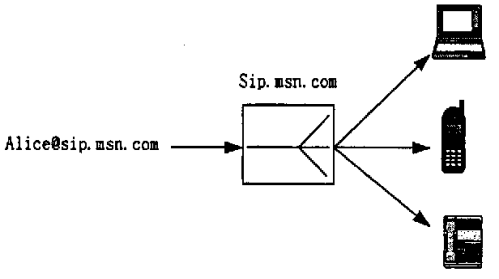

图 5.1 用户移动性示例

5.2.2 服务移动性

服务移动性指的是当用户的通信设备发生改变或用户移动到其他服务提供商的运营网络时，仍然可以使用自己在归属网络中定制的各种业务，并且所定制的业务属性和用户信息（电话本、好友列表、显示及声音设置等）保持不变[7]。

服务移动性的一个解决方案是让用户携带服务信息，比如目前用在 GSM 手机中的 SIM 卡，可以保存部分用户信息。但这种方法存在以下缺陷：一是携带信息的容量受到移动终端中存储设备存储容量的限制，二是当用户使用不同的终端设备时，必须要求所有类型的终端对信息存储设备可以兼容（例如都必须支持 SIM），三是当相同的用户信息存在于不同的终端时，难以保证这些信息之间的同步更新，因此，难以保证信息的一致性。

另一个解决方法是，通过网络存储，将用户个人信息保存在网络中。网络存储解决了用户终端存储中存在的问题，是一种实现服务移动性的更好的途径。

5.2.3 会话移动性

会话移动性可以保证用户在移动过程中，甚至在变换终端的过程中，仍然保持正在进行的会话。会话移动性更加强调用户将正在进行的会话切换到其他的通信终端上继续进行。如果正在进行的是多媒体会话，这将包括将多媒体会话的不同部分，如视频、音频从一个设备切换到不同设备上。

5.2.4 其他移动性

目前，与 NGN 相关的移动性研究还包括：Ad Hoc 移动性，以及从 Ad Hoc 网到有专用移动通信基础设施支撑的移动网的"模式移动性"[8]。

Ad Hoc 网是一种没有固定网络基础设施支持的移动网络，网络完全由移动主机构成，由于不依赖于任何外部设施，Ad Hoc 网在可靠性方面具有较强优势。Ad Hoc 网最初主要应用于军事领域，随着研究的不断进展以及无线通信和终端技术的快速发展，Ad Hoc 网络目前已经开始应用到民用领域，在移动通信基础设施难以建设的环境中，搭建 Ad Hoc 网络可以使人们迅速获得移动通信带来的巨大方便和高效。

Ad Hoc 网中，移动主机既是通信终端，也是路由器，必要时需承担寻找路由和转发报文的工作。当需要通信的两台移动主机在彼此的通信覆盖范围内时，通信可直接进行，否则需要通过它们之间的其他移动主机进行信息转发来实现。

"模式移动性"主要是指在有专用移动通信基础设施支撑的移动网和 Ad Hoc 网之间完成移动切换。在模式移动性中，当缺乏移动通信基础设施支撑时，移动设备组成 Ad Hoc 网，相互通信，或借助可与外界连接的移动设备与外界通信，当移动设备移动到有移动通信基础设施的环境中时，恢复使用无线基础设施，在这个过程中，正在进行的会话不应该被中断。

基于上述讨论，可以将 NGN 中所涉及的移动性进行分类，见表 5.1。

表 5.1 NGN 移动性分类

名称	简单描述
终端移动性	用户可以带着终端移动，保持正在进行的会话和连接。
会话移动性	用户可以在一个个人区域网络范围内转换终端的同时，保持正在进行的会话不被中断。
个人移动性	通过一个唯一的用户 ID，用户可以在世界上任何地方被访问，可以用任何一个认证过的终端发起或接收一个会话。

服务移动性	即使用户连接到一个外部网络，也可以获得一个订购过，而且个性化了的业务。
其他	Ad hoc 移动性，指在 Ad hoc 网中，移动终端之间的通信没有固定的设施，靠彼此互相路由信息的移动性。模式移动性指在不同的移动模式，比如移动 IP 网移动模式和 Ad hoc 网移动模式之间切换，而用户的正在进行的会话不受影响。

5.3 移动支持协议

5.3.1 移动 IP 协议

2G 系统中的移动性管理只涉及简单的终端移动性管理，这种移动性管理主要在链路层实现。所以它只适用于特定的系统，比如 GSM 网络或 CDMA 网络。在下一代网络中，底层异构的网络将在网络层实现统一的协议覆盖，从而形成一个全 IP 的网络。在这个全 IP 的网络中，用网络层协议实现移动性管理是很自然的选择。

在 IP 协议（IPv4 或者 IPv6）中，IP 地址同时具有两个标识作用，一个是唯一的标识终端，另一个是标识终端所处的位置。这种情况下，如果终端移动，离开自身 IP 地址所标识的位置，则所有发往该终端的信息将无法送达终端，通信将无法继续。为解决这个问题，IETF 提出了移动 IP 协议，目前已经推出了 MIPv4[9] 和 MIPv6[10]两个版本。在移动 IP 中，移动终端可有两个地址，一个是家乡（归属）地址（Home Address），当移动终端没有移动的时候，家乡地址既是终端的标识，又是终端所在位置的标识。当终端移动到其他管理域的时候，终端会获得一个新的地址，这个地址被称为转交地址（Care of Address）。在获得转交地址后，移动终端随即向家乡网络中的家乡（归属）代理（Home Agent）注册这个转交地址。与移动终端进行通信时，通信对端所发信息总是先到达移动终端的家乡网络，家乡代理将代替移动终端接收发往移动终端家乡地址的数据包，并用隧道的方式将数据包转发到移动终端的转交地址。移动 IP 的操作过程见图 5.2。

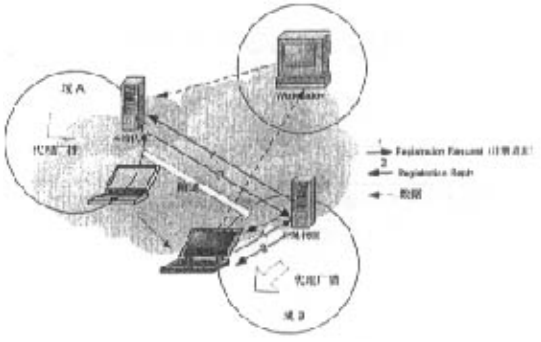

图 5.2 移动 IP 操作

5.3.1.1 移动 IP 可以支持的移动性

移动 IP 协议有效地解决了 IP 地址双重身份在移动终端的网络位置发生改变时产生的问题，在 IP 层实现了终端移动性。但是,移动 IP 协议会产生三角路由的问题，导致在处理终端切换时，效率低下。因此，研究人员又提出了许多补充协议[11][12][13][14]来解决这个问题。

移动 IP 也可以有效支持 Ad Hoc 网中主机与其他网络的互连。在 Ad Hoc 网内部，主机自己创建并维护动态的路由信息。在 Ad Hoc 网与其它网络衔接的边缘，有些主机具有双重身份，一方面它们是 Ad Hoc 网中的主机，可以利用 Ad Hoc 路由协议与其他主机通信，另一方面它们还拥有其他网络的网络地址（如 IP 地址），可以被其他网络的通信主机所寻址。通过这些双重身份的移动主机， Ad Hoc 网主机与 IP 网主机之间的互通得以实现。具备双重身份的主机可以在相邻的两个网络中移动，即在移动 IP 和 Ad Hoc 两种模式之间移动，实现前文所述的"模式移动"。移动主机在 Ad Hoc 网络与 IP 网络之间实现模式移动示意如图 5.3 所示。

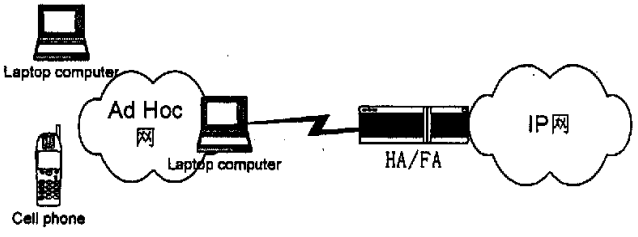

图 5.3 Mode 移动

5.3.1.2 移动 IP 难以支持的移动性

移动 IP 在网络层通过 IP 地址的转换实现了对移动性的支持。但是，值得注意的是，在上文提到的各种移动性中，个人移动性的实现需要依赖独立于底层网络的用户标识——应用层标识，服务移动性则建立在统一的网络存储基础上，而会话移动性很难单独在网络层协议的支持下得以完成[3] 。因此，对于不关注网络层以上的通信机制的移动 IP 来说，这些移动性都无法实现。

5.3.2 SIP 对移动性的支持

SIP 协议[15]是为建立多方参与的多媒体会话（包括语音、视频、游戏等）而设计的应用层协议。最初设计时，并未考虑对移动的支持，Wedlund 和 Schulzrinne 在文献[16]中提出了用 SIP 支持移动性的原理与方法，认为它可以更有效率地支持

实时移动通信。此后，他们在[7]中 SIP 支持多种移动性模式的能力和方法进行了分析。SIP 灵活的的路由方法及会话管理，如会话建立和会话描述分离，允许会话参与者或者第三方建立会话等特点，使得通过 SIP 协议可以实现多种移动性的管理。

SIP 中，用 SIP URL 标识用户。SIP URL 的格式类似于 E-MAIL 地址，如 Alice@sip.msn.com。用户如想被人找到，必须向注册服务器注册自己当前的位置，当用户变换位置时，必须向注册服务器重新注册，这个变换位置和重新注册的过程使得 SIP 具备了对移动性进行管理的能力。

5.3.2.1 SIP 对终端移动性的支持

SIP 可以支持终端移动性，根据移动发生的时间有所不同。呼叫前的终端移动性管理是指发生移动的用户并未处于任何 SIP 呼叫之中，这时，用户移动到新的位置，获得了新的地址，需要向自己的注册服务器进行注册。这样，当某通信对端的呼叫到达该用户归属网络时，重定向服务器通过注册服务器可以获知该用户终端的当前位置，通过信令响应告知通信对端，通信对端可以通过新位置信息重新发起呼叫，如图 5.4 所示。

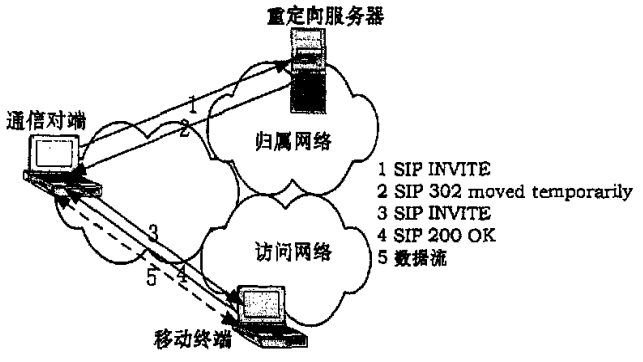

图 5.4　SIP 终端移动性管理—呼叫前移动

呼叫中的终端移动性管理是指当移动终端发生移动时，有正在进行的 SIP 呼叫，当移动终端获得新的地址后，发送 re-INVITE，在消息体的会话描述中，给出新的通信地址，这样，通信对端就会将通信数据发送到新地址，会话将继续进行，见图 5.5。但是在通信对端收到 re-INVITE，并将数据发送到移动终端新地址之前，会有数据丢失发生，对此，可采用一些提高移动性能的改进手段。

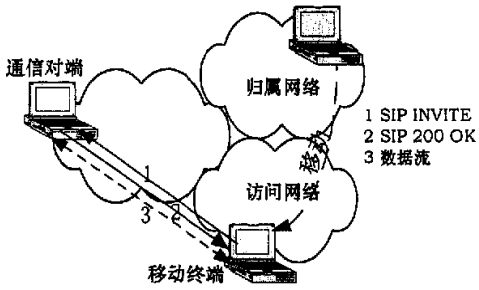

图 5.5 SIP 终端移动性管理—呼叫中移动

5.3.2.2 SIP 对会话移动性的支持

SIP 目前有两种方法支持会话移动性，一种是利用 SIP 的第三方会话控制，另一种是利用 SIP 的扩展命令 REFER[17]。

以图 5.6 为例阐述第一种方法。图中，A 的移动电话与 B 的移动终端之间存在一个正在进行的会话，A 想要将该会话从移动电话转移到笔记本电脑上，这种转移可以通过以下流程来实现：A 的移动电话先向 A 的笔记本电脑发送 INVITE 信令，在 INVITE 信令中 Call-ID 使用已经存在的会话的 Call-ID，但不携带会话描述消息体[11]，A 的笔记本电脑随后返回 200 OK 信令，并携带自己的会话描述信息，A 的移动电话再向 B 的移动终端发 INVITE 信令，信令中携带 A 笔记本电脑的会话描述信息，B 的移动终端在返回的 200 OK 信令中携带自己的会话描述信息。A 的移动电话在返回给笔记本电脑的确认 ACK 信令中携带 B 发来的会话描述信息。A 移动电话接着向 B 发送 ACK 信令进行确认。至此，原本在 A 的移动电话和 B 的移动终端之间进行的会话被转移到 A 的笔记本电脑和 B 的移动终端之间继续进行，成功地实现了会话的移动。采用第三方会话控制实现会话的移动性存在一定的局限性，具体表现在当会话实现移动后，会话的终止仍然需要由初始的会话参与方进行控制，例如，图中移动后的会话仍然需要由 A 的移动电话进行终止。

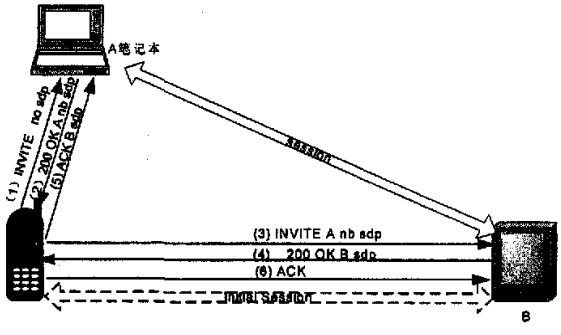

图 5.6 通过 SIP 第三方会话控制会话移动

对第二种方法的描述如图 5.7 所示。图中 A 同样要把与 B 之间的会话从移动电话转移到笔记本电脑上。采用第二种方法时按照以下信令流程进行：A 的移动电话向 B 的移动终端发送 REFER 命令，要求转接已经存在的会话，B 的移动终端以 202 Accept 进行响应，随后，B 的移动终端发送 INVITE 信令给 A 的笔记本电脑，在 INVITE 的头域 Referred-By 中指明本会话由 A 的移动电话转接而来，A 的笔记本发送 200 OK 进行响应，B 的移动终端回送 ACK 信令确认，此时，B 的移动终端通过 Notify 通知 A 的移动电话，会话已经转接成功，A 移动电话响应 200 OK 后，发 BYE 信令终止与 B 的移动电话之间已经存在的会话。第二种方法中，A 的移动电话在会话转接完成后就可以退出，避免了第一种方法的不足。

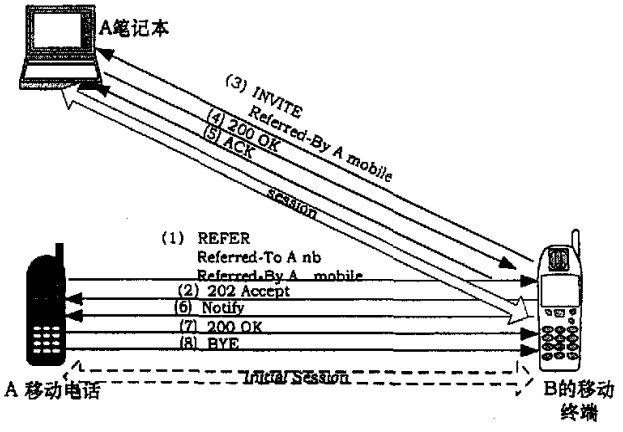

图 5.7 通过 REFER 命令实现会话移动

5.3.2.3 SIP 对个人移动性的支持

SIP 对个人移动性的支持是通过以下方式实现的。

SIP 协议中，通常是用类似于电子邮件地址的 SIP 统一资源定位器(URL)来标识用户的。URL 通常由用户名和域名构成，如：在 sip:byxp@bupt.edu.cn 中，byxp 为用户名，bupt.edu.cn 为域名。用户所在的域中存在一个注册服务器（Registrar），用户向注册服务器及时登记自己当前联系方式，这些联系方式可以是固定电话、移动电话、SIP 电话、电子邮件地址或者它们的组合。当用户被呼叫时，用户域的代理服务器 Proxy（或转接服务器 Redirector）会向注册服务器询问用户的当前联系方式，在获得所需信息后，呼叫被转接。通过这种机制，不管用户移动到什么位置，使用什么终端，只要及时地向注册服务器登记，都可以接到呼叫，从而实现个人移动性。用户在注册服务器进行注册时，可以设定联系方式的采用时间等

条件，使得转接在条件满足时发生。

5.3.2.4 SIP 对服务移动性的支持

服务移动性的实现需要两个前提，第一应该有一个统一的服务属性定义，包括属性的范围，参数；第二应该有“合适”的位置存储属性定义。

SIP 扩展[19]及 IETF 标准[20][21][22]定义了 SIP 用户代理的能力描述，格式及协商框架；用户代理都有自己所归属的本地网络，这个网络的域名与用户 URL 相联系，该域的注册服务器及 DNS 服务器可以作为存储服务属性的位置。

用户周期性的，或当自身地址、服务属性发生改变时，通过 REGISTER 命令，向注册服务器进行注册更新。在 SIP 消息的 Contact 头域中可以对服务属性进行描述，例如，如果用户服务存在下列属性：用户代理当前位置是 example.edu.cn，用户通话语言是“Chinese, English”、会话采用的媒体类型是“audio, video, application/chat”，本会话属于固定电话、用于工作目的。则 Contact 应该设定如下：

Contact: doctor1<sip: doctor1@example.edu.cn> ; language=”cn,en”
; media=”audio,video,application/chat”
; mobility=”fixed”
; class=”business”

因为属性位置可知，当用户移动到其他运行商服务区域时，新的运营商可以从用户的归属网络获得服务属性信息，继续为用户提供定制的服务。

5.3.2.5 SIP 支持终端移动性和模式移动性

从理论上讲，SIP 是可以支持终端移动性的，当移动终端移动到新的位置，可以发送 re-INVITE 给通信对端，在消息体中给出新的会话信息描述，提供新的终端地址、端口等参数，通信双方可以据此建立新的连接，继续双方的会话。

但是，因为 SIP 处于应用层，是应用层协议，它无法实现对终端移动的及时检测，所以当终端连续移动时，切换延迟较大，效率很低，这将严重影响通话的质量。因此在支持终端移动性方面，SIP 并不是最理想的选择。

与 Ad Hoc 网络互联的最佳层次是网络层，处理与 Ad Hoc 网之间的模式移动的理想位置也是网络层，所以 SIP 也不适用于支持模式移动性。

5.4 多层多协议移动性管理策略

从上面的分析可以看出,单靠某一层的协议机制很难实现对 NGN 移动性的全

面支持，必须将各层移动性管理协议有机的结合起来，才能既实现 NGN 中不同类型的移动性管理，又保证效率和可靠性。

移动 IP 可以屏蔽底层网络的异构性，对上层协议与应用提供一个统一的网络环境，是实现终端移动性和模式移动性的理想层次。为了提高效率，层次型移动 IP 是更好的方案。对于层次型移动 IP 管理，分为宏移动性管理和微移动性管理。在宏移动性管理中，使用移动 IP，而在微移动性管理中，则使用 HMIP[11]，HAWAII[13]，Cellular IP[14]等协议。采用微移动性管理时,由于在小的管理范围内的移动，不需要向家乡代理进行注册，提高了移动性管理的效率。在实现网络层的移动性管理时,充分利用不同链路层协议提供的移动检测、软切换等功能可以进一步提高移动切换的效率，对于个人移动性，服务移动性和会话移动性，与应用层关系密切，用 SIP 协议及相关扩展予以实现是最佳的选择。

由认证（Authentication）、授权（Authorization）、计费（Accounting）组成的 AAA 问题，是各种网络与应用得以大规模部署的基础。IETF 新一代的 AAA 协议是 Diameter 协议，Diameter 由基本协议和扩展协议构成，对于移动 IP 和 SIP，IETF 分别定义了 Diameter 应用扩展，为它们提供有针对性的 AAA 服务。

这样，在多协议多层次的环境下实现 NGN 的移动性管理，可以构成一个多协议多层次 NGN 移动性管理架构。见图 5.8。

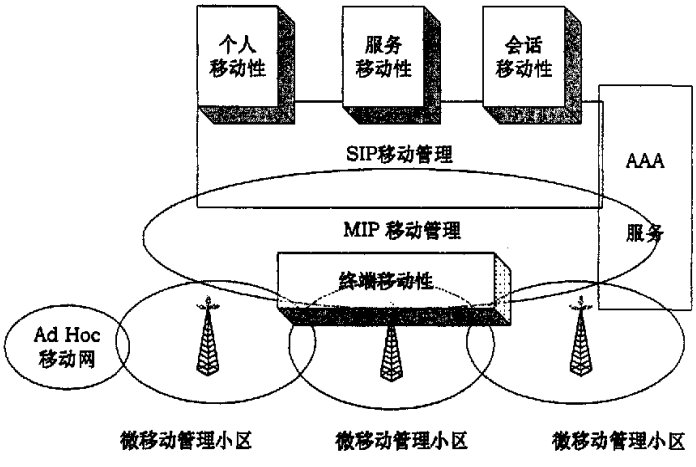

图 5.8 多协议多层次移动性管理

5.5 本章小结

NGN 中，除传统的终端移动性管理之外，移动性管理还将包括个人移动性、

服务移动性、会话移动性等多种移动性的管理。网络层移动性管理协议——移动 IP 及其扩展协议可以很好的满足终端移动性管理的要求，但是对于其他类型移动性的管理，则显得能力不足。SIP 协议具备在应用层实现移动性管理的能力，SIP 扩展性也很强，SIP 及相关扩展可以很好地满足个人/服务/会话等移动性的管理。将移动 IP 及微移动性管理协议和 SIP 协议进行有机地组合，形成多协议多层次的下一代移动性管理方案是 NGN 移动性管理的最佳选择。本章提出了在这种方案下的管理架构。

参考文献

1. Patel, G. and S. Dennett. The 3GPP and 3GPP2 movements toward an all-IP mobile network [J]. IEEE Personal Communications, 2000, 7(4): 62-64.
2. Faccin, S. M., P. Lalwaney, et al. IP multimedia services: analysis of mobile IP and SIP interactions in 3G networks [J]. IEEE Communications Magazine, 2004, 2(1): 113-120.
3. 赵慧玲.标准化工作进展及NGN定义.人民邮电报, 2004-12-30.
4. 邢燕霞,赵慧玲.基于IMS的网络融合分析[J].电信科学. 2005, 3:1-5
5. 孙元宁.基于全IP核心网络架构的IMS.通信产业网, http://www.ccidcom.com/.
6. ITU-T FGNGN."ITU-T NGN FG proceedings PartI&PartII".2005.
7. Schulzrinne, H and E Wedlund. Application-Layer Mobility using SIP [J]. ACM Mobile Comp. and Commun. Rev. , 2004, 4(3): 47-57.
8. Qi Wang, Mosa Ali Abu-Rgheff. Next-Generation Mobility Support [J]. Communications Engineer, 2003, 1(1):16–19.
9. C Perkins. "IP Mobility Support for IPv4". IETF RFC 3344, August 2002.
10. D Johnson, C Perkins, J Arkko. "Mobility Support in IPv6". IETF RFC 3775, June 2004.
11. Hesham Soliman. Hierarchical Mobile IPv6 mobility management (HMIPv6). draft-ietf-mipshop-hmipv6-03, October, 2004.
12. Rajeev Koodli. Fast Handovers for Mobile IPv6. draft-ietf-mipshop-fast-mipv6-03, October 2004.

13. Ramjee, R., K. Varadhan, et al. HAWAII: a domain-based approach for supporting mobility in wide-area wireless networks [J]. IEEE/ACM Transactions on Networking, 2002, 10(3): 396-410.

14. A.Valkó, Cellular IP: A New Approach to Internet Host Mobility [J]. ACM SIGCOMM Comp, Commun. Rev., 1999, 29(1):50-65.

15. Rosenberg J，Schulzrinne H, Camarillo G. SIP: session initiation protocol. IETF RFC 3261, 2002.

16. Wedlund E, Schulzrinne H. Mobility support using SIP [A]. Proceedings of Second ACM International Workshop on Wireless Mobile Multimedia (WOWMOM) [C]. 1999.

17. R. Sparks. The Session Initiation Protocol (SIP) Refer Method. IETF RFC 3515, 2003.

18. 白建军，彭晖，田敏等译. SIP揭密[M]. 北京：人民邮电出版社, 2003年.

19. J. Rosenberg, et al. Caller Preferences for the Session Initiation Protocol (SIP). IETF RFC 3841, 2004.

20. Rosenberg, J., Schulzrinne, J., and P. Kyzivat. Indicating User Agent Capabilities in the Session Initiation Protocol (SIP). IETF RFC 3840, August 2004.

21. Klyne, G. Protocol-independent Content Negotiation Framework. IETF RFC 2703, 1999.

22. Holtman, K., Mutz, A., and T. Hardie. Media Feature Tag Registration Procedure. IETF RFC 2506, 1999.

23. P Calhoun, H Akhtar, J Arkko, E Guttman et al. "Diameter Base Protocol", RFC 3588, September 2003.

24. Pat R Calhoun, Tony Johansson and et al. ,"Diameter Mobile IPv4 Application", IETF Internet Draft,draft-ietf-aaa-diameter-mobileip-20.txt, August 2004.

25. M.Garcia-Martin, M Belinchon et al. ,"Diameter Session Initiation Protocol (SIP) Application" , IETF Internet Draft, draft-ietf-aaa-diameter-sip-app-07, March 2005.

26. 马建等.IPv6原理及在移动通信中的应用.北京：科学出版社，2004.

第六章. SIP 与移动 IP 在移动性管理中的协同问题研究

6.1 引言

在上一章 NGN 移动性管理中谈到，为有效的实现 NGN 移动性管理，必须采用多层多协议的移动性管理方案。多层就是指移动性管理要利用链路层，网络层和应用层各自的长处来实现，多协议是指各层对移动性管理要依靠不同的协议，这些协议都有移动性管理的能力，但是长处各有不同，在移动性管理中，既需要利用它们各自的长处，也需要它们有机的配合。在这些协议中，移动 IP 和 SIP 是最重要的两个。

与传统 IP 相同，移动 IP 是网络层协议，它的功能就是支持终端在保持通信连接不中断情况下进行移动的协议。传统 IP 协议中，IP 地址既是主机的标识，又是数据包寻址的依据，所以一旦主机移动，就不能根据 IP 地址找到它。当移动主机(MN, mobile node)移动到访问域（visited domain），将获得一个临时地址，并将该地址通知归属域（home domain）相关设备，与之通信的终端仍将数据包发往 MN 原地址——归属域地址（home address），归属域截获数据包，并根据 MN 临时地址进行转发，从而不影响其正在进行的通信。

SIP 最初设计时，并未考虑对移动的支持，Wedlund 和 Schulzrinne 在文献[1]中提出了用 SIP 支持移动性的原理与方法，认为它可以更有效率地支持实时移动通信。此后，他们在另一文献[2]中提出了下一代网络中的多种移动性模式的概念。文献[3]讨论了在下一代网络中支持多种移动性的问题，该文献与文献[4][5][6]都建议，为支持多种移动性模式，并提高效率，应采用多层多协议的思路进行移动性管理。这些建议的核心都是如何让网络层的移动 IP 协议和应用层的 SIP 协议更有机的集成。

[4]提出的模型着眼于利用移动 IP 和 SIP 各自的长处，减少移动切换的时间，[5]设计在注册和移动切换时，避免 SIP 注册，以减少信令交互，[6]则是设计专门的边界设备来协调不同的协议和移动模式。这些文献[3~6]所提的技术方案各有所长，也都有一定的适应环境，但在这些方案中，都未涉及在未来商用环境中最关键的，由认证（Authentication）、授权（Authorization）、计费（Accounting）组成的 AAA 问题。AAA 问题，是各种网络与应用得以大规模部署的基础。IETF 新一代的 AAA 协议是 Diameter 协议，Diameter 由基本协议和扩展协议构成，针对移动 IP 和 SIP，IETF 分别定义了 Diameter 应用扩展。在包含二者的集成应用中，如果让它们各自独立地进行 AAA 操作，会存在缺乏效率的问题。

本章提出一种优化方案——“移动 IP 与 SIP 集成应用中优化的 AAA 过程”（OAPIMS，Optimized AAA Procedure of Integration of Mobile IP and SIP），可以很好地解决这个问题，分析表明，它可以大大的优化性能。

本章后续内容如下安排，第 6.2 节分别描述移动 IP 和 SIP 进行注册时的 AAA 过程；第 6.3 节是优化的集成注册方案；第 6.4 节是对优化方案的性能分析；第 6.5 节进行小结。

6.2 AAA 环境下的移动 IP 及 SIP 注册过程

6.2.1 AAA 协议——Diameter

目前来说，应用最广泛的 AAA 协议是 RADIUS[7]，它的制定和推广使用对互联网的发展起过很重要的作用，但是因为协议本身的一些缺陷，如没有定义统一的故障恢复机制，无法保证消息的可靠传输，缺乏能力协商机制等，使得 RADIUS 在面对新应用新环境时，无法满足需要，新的 AAA 协议 Diameter 应运而生。

作为新一代 AAA 协议，Diameter 与 RADIUS 有很大的不同。Diameter 克服了 RADIUS 的不足，并在设计时，充分考虑了协议的可扩展性。Diameter 协议分基本和扩展两大部分。基本部分[8]定义协议的基本结构，元素及扩展原则，扩展部分则在基本部分的支持下处理专门的应用需求，如针对网络接入服务的接入应用扩展（Diameter Network Access Server Application），针对移动 IPv4 的应用扩展（Diameter Mobile IPv4 Application ）[9]，针对 SIP 协议的应用扩展（Diameter Session Initiation Protocol Application）[10]等。

Diameter 中所有数据都以 AVP（Attribute Value Pair）的格式定义和传输。在基本协议中定义了一些 AVPs，在各个应用扩展中，根据具体应用需求，可以定义更多的 AVPs。

下面的讨论基于这样的场景，带有移动 IP 和 SIP 协议栈的移动主机移动到访问域，获得新的 IP 地址后，移动 IP 和 SIP 都要求向家乡域进行注册，同时还要进行相关的 AAA 操作。根据移动 IPv4 和 SIP 的 Diameter 协议应用扩展，注册与 AAA 操作是同时进行的。

6.2.2 Diameter 环境下移动 IP 注册过程

实际可运营的移动 IP 系统中，当 MN 离开归属域，到达并接入一个访问域时，该访问域需要对接入到本域的 MN 进行接入控制，包括认证接入者的身份，授权

使用本网络资源，同时要向接入者的归属域管理器报告接入者在本域内的资源使用情况，以便管理与计费。

对于这里的 AAA 需求，文献[11，12]进行了分析，并给出了相应的模型（见图 6.1）。移动到访问域的 MN 通过接入服务点（Attendant）接入访问域。接入服务点可以由本地代理（FA，foreign agent）担任。接入服务点要求 MN 提供认证证书（credential），然后与访问域 AAA 服务器（AAAL）联系，请求认证。一般情况下，AAAL 没有足够的信息认证移动节点，需要联系 MN 归属域中的 AAA 服务器（AAAH），进行认证。Diameter 移动 IP 应用扩展根据这样的需求分析及应用模型定义了 4 条命令，分为 2 对，如表 6.1 所示。

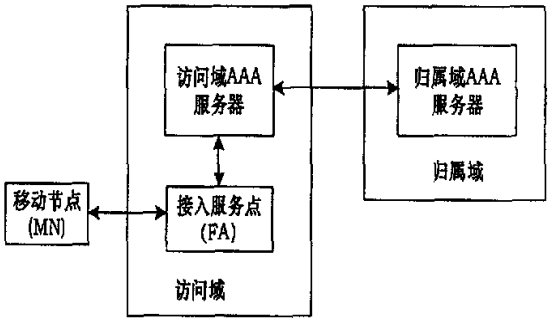

图 6.1 跨域 AAA 模型

表 6.1：Diameter 移动 IP 应用扩展中的命令

命令	缩写	功能
AA-Mobile-Node-Request	AMR	Attendant(FA)收到来自 MN 的注册请求后，构造此消息，发送到 AAAH，进行认证。
AA-Mobile-Node-Answer	AMA	AAAH 利用此消息将认证结果返回到 Attendant(FA)。
Home-Agent-MIP-Request	HAR	由 AAAH 发送到 MN 的家乡代理，用于注册。
Home-Agent-MIP-Answer	HAA	由 MN 的家乡代理发送到 AAAH，作为 HAR 的响应。

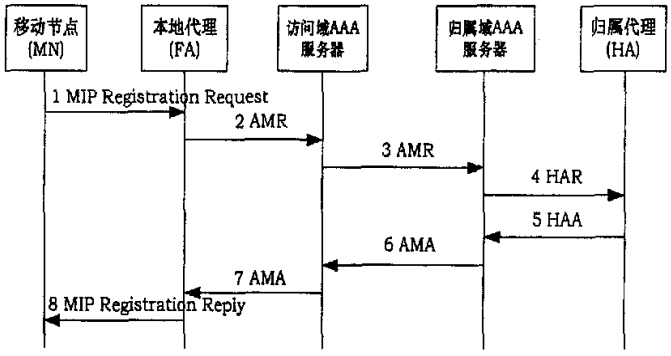

图 6.2 移动 IP 注册认证消息

MN 移动到访问域，通过某种方式（如 DHCP）获得临时转交地址 CoA（care of address），然后通过 FA，向归属域中的归属代理（HA，home agent）进行注册，同时进行 AAA 操作，不考虑失败情况下，步骤如下（见图 6.2）：

1) MN 发送移动 IP 注册请求消息到 FA；

2) FA 根据注册请求消息中的信息，构造相应的 AVPs，如 MIP-Mobile-Node-Address AVP（根据 MN 家乡地址）、MIP-Reg-Request AVP（根据移动 IP 注册请求消息扩展）、MIP-Home-Agent-Address AVP（根据 MN 归属代理地址）等，形成 AMR 消息，发送到本地域 AAA 服务器 AAAL（diameter server）；

3) AAAL 将 AMR 消息转发到 MN 归属域的 AAA 服务器 AAAH（diameter server）；

4) AAAH 取出相应的 AVPs，对 MN 进行认证；

5) 如果认证通过，AAAH 发送 HAR 消息到 HA，消息中包含 MIP-Reg-Request AVP，进行 MN 注册请求；

6) HA 注册结束，通过 HAA，向 AAAH 发送响应；

7) AAAH 构造 AMA 消息并发送到 MN 所在访问域的 AAAL，AAL 将 AMA 消息转发到 FA；

8) FA 提取认证、授权、计费信息，并根据相关的 AVPs，构造移动 IP 注册响应消息，发送到 MN。

6.2.3 Diameter 环境下 SIP 注册过程

SIP 协议对移动性的支持中，当移动节点（MN）移动到访问域，通过某种方式（如 DHCP）获得新的 IP 地址后，需要用 REGISTER 命令向归属域进行注册，同时也要进行认证、授权、计费的 AAA 操作。各域通常都有 SIP Proxy，对消息进行转发。下面的讨论中，归属域的 SIP Proxy 不涉及所讨论问题，简单起见，予以省略。

Diameter SIP 应用扩展定义了 12 条命令，分为 6 对，这里介绍相关的 4 条（见表 6.2）。

表 6.2：Diameter SIP 应用扩展中的 4 条相关命令

命令	缩写	功能
Multimedia-Auth-Request	MAR	SIP server 中的 Diameter client 发到 Diameter server，请求对某用户进行认证，并授权其使用 SIP server 提供的服务
Multimedia-Auth-Answer	MAA	对 MAR 消息进行响应，通过 Result-Code AVP 通知 SIP server 认证处理的状态，以便其进行下一步处理
Server-Assignment-Request	SAR	SIP server 中的 Diameter client 发到 Diameter server，通知认证结束，并请求记录自己为认证用户的服务器，也可用于请求对应用户的用户定制信息。
Server-Assignment-Answer	SAA	对 SAR 的响应，可携带用户定制信息。

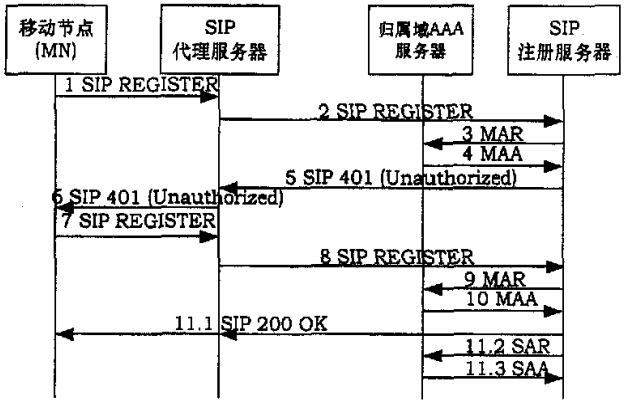

图 6.3 SIP 协议下 AAA 注册消息

MN 移动到访问域，获得新的 IP 地址后，向归属域的 SIP 注册服务器（SIP Registrar）进行注册，同时通过归属域的 AAA 服务器 AAAH（diameter server），进行认证操作，不考虑失败情况下，步骤如下（见图 6.3）：

1) MN 发送 SIP 注册命令 REGISTER 到访问域的 SIP Proxy；
2) 访问域的 SIP Proxy 将 SIP REGISTER 转发到 MN 归属域，最后到达 SIP Registrar；
3) SIP Registrar 中的 Diameter client 向域中的 Diameter server 发送 MAR 消息；
4) Diameter server 记下 MN 与 SIP Registrar 的对应关系，发送 MAA 消息到 SIP Registrar，Result-Code AVP 设置为 DIAMETER_MULTI_ROUND_AUTH，并在 SIP-Authentication-Scheme AVP 给出所使用的认证策略，如 HTTP-Digest，在 SIP-Authenticate AVP 中则给出认证所需参数。
5) SIP Registrar 发送 401 响应到 SIP Proxy，在 WWW-Authenticate 头域中给出 Challenge 信息，如 nonce，opaque 等；
6) SIP Proxy 将 401 消息转发到 MN；
7) MN 根据消息中的 Challenge，计算响应凭证(Credential)，放在头域 Authorization 中，重新生成 REGISTER 消息，发送到访问域的 Proxy；
8) 访问域的 Proxy 将新的 REGISTER 发送到 MN 归属域，最后到达 SIP Registrar；
9) SIP Registrar 向 Diameter server 发送 MAR 消息，将 MN 的认证信息发送到 Diameter server；
10) Diameter server 对 MN 进行认证后，通过 MAA 向 SIP Registrar 发回响应；
11) SIP Registrar 向 MN 发回 200 响应，同时通过 SAR 消息，向 Diameter server 报告认证结束信息，并请求 Diameter server 将自己登记为 MN 服务者，Diameter server 通过 SAA 进行响应。

6.3 Diameter 环境下优化的集成注册过程

6.3.1 优化的注册过程

从上面对移动 IP 和 SIP 注册过程的描述中，我们可以发现，当 MN 移动到访问域时，如果独立地进行移动 IP 和 SIP 的注册与 AAA 操作，需要在访问域和归属域之间进行三个往返的信令交互，效率很低，尤其当 MN 距离归属域很远时，这种操作的低效就更加难以接受。

我们分析，移动 IP 的 Diameter 应用与 SIP 的 Diameter 应用，都是基于 Diameter 的应用扩展，具有相同的应用模式，所不同的主要是消息所带的 AVP 参数。

这里，我们根据上述特点，建议一种优化的移动 IP 和 SIP 集成应用的注册认证方法，即"移动 IP 与 SIP 集成应用中优化的 AAA 过程"（OAPIMS）。这种方法中，当 MN 移动到访问网络后，要确保先进行移动 IP 的注册认证操作。具体实现及步骤描述如下（参考图 6.6）：

1) 在 MN 进行移动 IP 注册的请求报文[13]（见图 6.4）中，设置保留标记 x（本应清 0，一般处理中忽略），表示移动节点具有 SIP 协议栈，下一步需要进行 SIP 注册，发送请求报文到 FA。

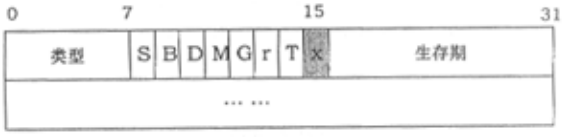

图 6.4 移动 IP 注册请求报文格式

2) 将 FA 和 SIP Proxy 的功能实现在同一个设备中，记为 FA/SIP Proxy，当它收到设置了 x 标记的移动 IP 注册请求报文时，在生成发往本地 Diameter server(AAAL)的 AMR 消息时，设置 Diameter 消息头[8]中命令标记字段的第 7 位（亦为保留字段，应清 0，并忽略处理。见图 6.5），用于指示归属域 Diameter server 在返回消息中提供 SIP 认证 challenge 信息，之后将 AMR 消息发到本地 Diameter server(AAAL)，AAAL 将消息转发到归属域 Diameter server(AAAH)。

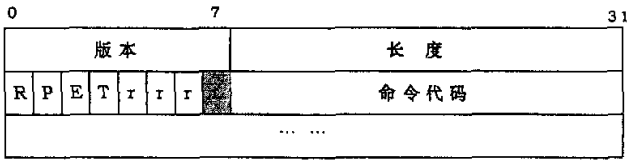

图 6.5 Diameter 消息头格式

3) 当归属域 Diameter server(AAAH)收到设置了相应标记的 AMR 消息时，就可以知道后续还要处理 SIP 注册消息。在发送 HAR 消息到 HA 进行注册，并收到返回的 HAA 后，需要向位于访问域的 FA/SIP Proxy 发送 AMA 消息，这时在其中添加 SIP-Authentication-Scheme AVP（值为"Digest"）与 SIP-Authenticate AVP（包含可以生成 HTTP Digest 认证中生成 WWW-Authenticate 头域和 Proxy-Authenticate 头域的信息,如 nonce, opaque 等），然后发送此 AMA 消息到 AAAL。

4) AAAL 处理后，将 AMA 发到 FA/SIP Proxy。

5) FA/SIP Proxy 构造移动 IP 注册响应消息，并发送到 MN，完成移动 IP 注册过程。

6) MN 发起 SIP 注册，发送注册命令 REGISTER 到访问域的 FA/SIP Proxy。

7) FA/SIP Proxy 根据从 SIP-Authenticate AVP 中得到的数据，构造 Proxy-Authenticate 头域，发送 407 响应到 MN。注意，这里把第一次失败注册信令交互限制在本地。

8) MN 根据 Proxy-Authenticate 头域中的数据，进行 Digest 认证计算，放在新的 REGISTER 请求的 Proxy-Authorization 头域中，再次发送到 FA/SIP Proxy；

9) FA/SIP Proxy 将新的 REGISTER 发到 MN 归属域的 SIP Registrar；

10)SIP Registrar 向 Diameter server 发送 MAR 消息，将 MN 的认证信息发送到 Diameter server；

11)Diameter server 对 MN 进行认证后，通过 MAA 向 SIP Registrar 发回响应；

12)SIP Registrar 发送 200 响应到访问域的 FA/SIP Proxy；同时，通过 SAR 消息，向 Diameter server 报告认证结束信息，并请求 Diameter server 将自己登记为 MN 服务者，并为 MN 保存一些服务信息，Diameter server 通过 SAA 响应。至此，完成了 SIP 注册。

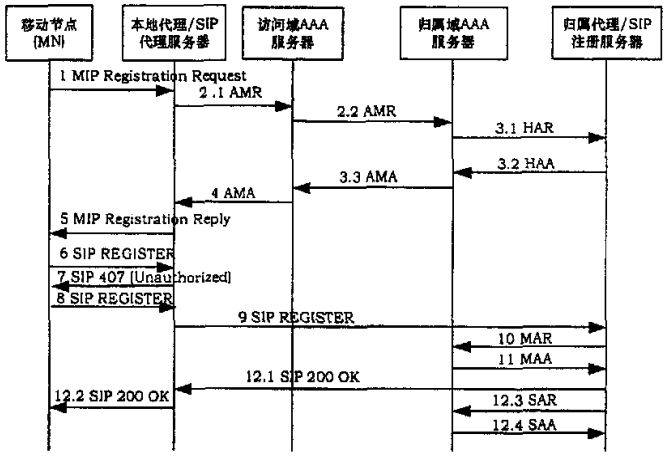

图 6.6 优化注册过程

6.3.2 优化注册过程的性能分析

在这一节，对于上述的 OAPIMS 优化方案进行性能分析，以说明该方法相对于独立进行移动 IP 与 SIP 注册操作的性能改进程度。我们从信令开销和时延的角度分别考察不同方法下的性能表现。定义信令开销为网络节点之间交换的信令流量。

我们定义 A、B 两点之间一次信令传输的信令开销为 $C_{A\text{-}B}$，定义移动 IP 注册信令开销为 C_{MIP}，SIP 注册信令开销为 C_{SIP}，非优化方案的集成开销为 C_{INT}，OAPIMS 方案下开销为 $C_{O\text{-}INT}$，则根据图 6.2、6.3、6.6，可以得到：

$$C_{MIP} = 2C_{MN\text{-}FA} + 2C_{FA\text{-}AAAL} + 2C_{AAAL\text{-}AAAH} + 2C_{AAAH\text{-}HA};$$

$$C_{SIP} = 4C_{MN\text{-}PROXY}^{注1} + 4C_{PROXY\text{-}REGISTRAR}^{注2} + 6C_{AAAH\text{-}REGISTRAR};$$

$$C_{INT} = C_{MIP} + C_{SIP};$$

$$C_{O\text{-}INT} = 6C_{MN\text{-}FA} + 2C_{FA\text{-}AAAL} + 2C_{AAAL\text{-}AAAH} + 2C_{AAAH\text{-}HA} + 4C_{AAAH\text{-}REGISTRA} + 2C_{PROXY\text{-}REGISTRAR};$$

注 1 PROXY 代表 SIP proxy；注 2 REGISTRAR 代表 SIP Registrar

我们将访问域与归属域之间的一次信令传递的开销归一化为 1，则域内部进行一次信令开销的代价为 α（$0 \le \alpha \le 1$）。

于是我们得到：

$C_{MIP} = 2 + 6\alpha$，$C_{SIP} = 4 + 10\alpha$

$C_{INT} = 6 + 16\alpha$

$C_{O\text{-}INT} = 4 + 14\alpha$

我们定义 S 为 OAPIMS 方案相对于非优化方案的性能优化比：

$$S = \frac{C_{INT} - C_{O\text{-}INT}}{C_{INT}} = \frac{2+2\alpha}{6+16\alpha} \quad (1)$$

表 6.3：α 取值与性能优化比 S 对应表

α	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
S	0.333	0.297	0.273	0.255	0.241	0.231	0.222	0.215	0.209	0.204	0.182

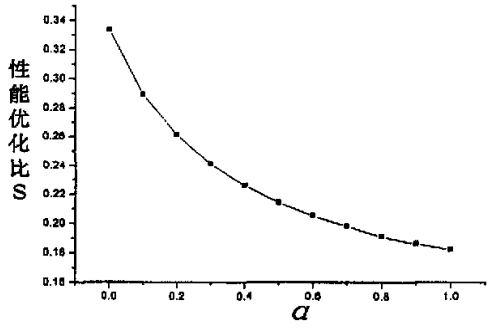

图 6.7：OAPIMS 相对于非优化方式的性能优化比

计算在α不同取值下的 S 值（表 6.3），并画出曲线（图 6.7），可以看出，根据α取值的不同，OAPIMS 优化方案可以提高性能 18.2% 到 33.3%。性能优化比 S 与α取值成反比关系。几点特殊说明，α取 0 时意义是，相对于域间信令开销，域内信令开销可以忽略不计，这时性能优化比 S 最大，达到 33.3%；α取 1 的意义是，域间信令开销与域内信令开销相同，这时 OAPIMS 优化方案对性能的优化最小，主要来自减少信令交互次数，性能优化比 S 可以达到 18.2%。

类似的，我们定义 A、B 两点之间一次单向信令传输的传输时延为 $T_{A\text{-}B}$，定义移动 IP 注册过程中，AAA 服务器认证处理和 HA 注册处理的总处理时延为 T_p，定义 SIP 注册处理中，AAA 服务器认证处理和 Registrar 注册处理的总处理时延为 T_q，定义移动 IP 注册时延为 T_{MIP}，SIP 注册时延为 T_{SIP}，非优化集成方案下两种协议都完成注册的总时延为 T_{INT}，OAPIMS 方案下总时延为 $T_{O\text{-}INT}$，则根据图 6.2、

6.3、6.6，可以得到：

$T_{MIP} = 2T_{MN\text{-}FA} + 2T_{FA\text{-}AAAL} + 2T_{AAAL\text{-}AAAH} + 2T_{AAAH\text{-}HA} + T_p$；

$T_{SIP} = 4T_{MN\text{-}PROXY} + 4T_{PROXY\text{-}REGISTRAR} + 4T_{AAAH\text{-}REGISTRAR} + T_q$；

$T_{INT} = T_{MIP} + T_{SIP}$；

$T_{O\text{-}INT} = 6T_{MN\text{-}FA} + 2T_{FA\text{-}AAAL} + 2T_{AAAL\text{-}AAAH} + 2T_{AAAH\text{-}HA} + 2T_{AAAH\text{-}REGISTRA} + 2T_{PROXY\text{-}REGISTRAR} + T_p + T_q$；

在 T_{INT} 和 $T_{O\text{-}INT}$ 中，T_p 与 T_q 基本不受影响，可以认为是常数，定义二者之和为 T_w。我们将访问域与归属域之间的一次单向信令传输的传输时延归一化为 1，域内部进行一次单向信令传输的传输时延则为 t（$0 \le t \le 1$），相应的 T_w 表示为 w（$w \ge 0$）。

可以得到：

$T_{INT} = 6 + 14t + w$

$T_{O\text{-}INT} = 4 + 12t + w$

我们定义 K 为 OAPIMS 方案相对于非优化方案的时延优化比：

$$K = \frac{T_{INT} - T_{O\text{-}INT}}{T_{INT}} = \frac{2+2t}{6+14t+w} \qquad (2)$$

表 6.4：t, w 取值与时延优化比 K 对应表

K	w=0	w=0.25	w=0.5	w=1	w=3	w=6	w=8
t=0.0	0.333	0.320	0.308	0.286	0.222	0.167	0.143
t=0.1	0.297	0.288	0.278	0.262	0.212	0.164	0.143
t=0.2	0.273	0.265	0.258	0.245	0.203	0.162	0.143
t=0.3	0.255	0.249	0.243	0.232	0.197	0.160	0.143
t=0.4	0.241	0.236	0.231	0.222	0.192	0.159	0.143
t=0.5	0.231	0.226	0.222	0.214	0.188	0.158	0.143
t=0.6	0.222	0.218	0.215	0.208	0.184	0.157	0.143
t=0.7	0.215	0.212	0.209	0.202	0.181	0.156	0.143
t=0.8	0.209	0.206	0.203	0.198	0.178	0.155	0.143
t=0.9	0.204	0.202	0.199	0.194	0.176	0.154	0.143
t=1.0	0.200	0.198	0.195	0.190	0.174	0.154	0.143

对 t 和 w 取不同数值，计算 K 值（见表 6.4），并划出相应的变化曲线（见图 6.8）。从表 6.4 数据和图 6.8 曲线可以看出，当 w 固定时，随着 t 逐渐增大，K 逐渐递减，K 与 t 成反比关系。t 取 0 时，意义是相对于域间传输时延，域内传输延迟可以忽略，这时，OAPIMS 对时延优化最大，当 t 取 1 时，意义是域间传输时延与域内传输时延相同，此时，OAPIMS 对时延优化最小；w 取值的变化反映的是

注册认证处理时延相对于传输时延大小，对 OAPIMS 优化程度的影响，根据表 4 数据及图 8 曲线可以看出，随着 w 取值逐渐增大，曲线下移，表明优化性能整体下降，这很好理解，OAPIMS 优化方案主要的优化思路是减少信令交互，所以当信令传输对时延的影响相对下降时，优化程度也会相应下降。

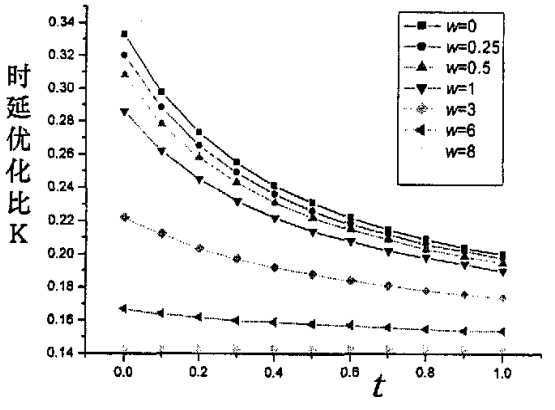

图 6.8：OAPIMS 相对于非优化方式的时延优化比

w 取 0 时的意义是相对于传输时延，认证处理时延可以忽略，这时 OAPIMS 对时延性能的优化最好，w 取值逐渐增大，OAPIMS 对时延的优化逐渐减弱，当 w 取 8 时，由⊙可以得到 K 取常数 1/7 ，其意义是认证处理造成的时延成为主要矛盾，域间或域内传输时延的相对大小成为次要矛盾。这种情况下，OAPIMS 方案的优化意义就不是很大了。

6.4 本章小结

未来移动应用中，为了满足终端移动性、个人移动性、会话移动性、业务移动性等多种移动性管理的需求，需要采用多层多协议移动性管理的方案。移动 IP 与 SIP 协议可以分别满足网络层和应用层的移动性管理需求。在一个有机的系统中，异层协议间必须有很好的协同与配合。由认证、授权、计费组成的 AAA 过程是各种移动应用得以部署的必要条件，IETF 针对移动 IP 和 SIP 分别定义了 Diameter 应用扩展，当移动主机移动到访问网络时，如果移动 IP 和 SIP 各自独立进行注册认证，从信令开销的角度看，很缺乏效率。本文提出一种优化解决方案 OAPIMS，详细描述了实现过程，并分析了该方案相对于原来两种协议独立进行 AAA 操作的方法，在信令开销和时延方面的优化程度。分析表明，这种优化方案可以明显改善性能。

参考文献

1. E.Wedlund, H.Schulzrinne, "Mobility Support Using SIP", Proceedings of Second ACM International Workshop on Wireless Mobile Multimedia (WOWMOM), August 1999.

2. Schulzrinne, H. and E. Wedlund (2000). "Application-Layer Mobility using SIP." ACM Mobile Comp. and Commun. Rev. vol. 4(no. 3): pp. 47-57.

3. Q Wang, M A Abu-Rgheff, et al. (2004). Design and evaluation of an integrated mobile IP and SIP framework for advanced handoff management. Communications, 2004 IEEE International Conference on.

4. Jung, J.-W., R. Mudumbai, et al. (2003). Performance evaluation of two layered mobility management using mobile IP and session initiation protocol. Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE.

5. Lee, H., S. W. Lee, et al. (2003). Mobility management based on the integration of mobile IP and session initiation protocol in next generation mobile data networks. Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th.

6. Q Wang, M A Abu-Rgheff, et al. (2004). Design and evaluation of an integrated mobile IP and SIP framework for advanced handoff management. Communications, 2004 IEEE International Conference on.

7. C Rigney, S Willens, A Rubens et al. "Remote Authentication Dial In User Service (RADIUS)", RFC 2865, June 2000.

8. P Calhoun, H Akhtar, J Arkko, E Guttman et al. "Diameter Base Protocol", RFC 3588, September 2003.

9. Pat R Calhoun, Tony Johansson and et al. "Diameter Mobile IPv4 Application". IETF Internet Draft,draft-ietf-aaa-diameter-mobileip-20.txt, August 2004.

10. M.Garcia-Martin, M Belinchon et al. "Diameter Session Initiation Protocol (SIP) Application". IETF Internet Draft, draft-ietf-aaa-diameter-sip-app-07, March 2005.

11. Perkins, C. E. (2000). "Mobile IP joins forces with AAA." Personal Communications, IEEE 7(4): 59-61.

12. S. Glass, S. Jacobs, C. Perkins, "Mobile IP Authentication, Authorization, and Accounting Requirements". RFC 2977. October 2000.

13. Perkins, C., "IP Mobility Support for IPv4", RFC 3344, August 2002.

第七章. IMS 移动性管理中认证操作的优化研究

7.1 引言

3GPP R5 中，提出基于 SIP(Session Initiation Protocol)协议的 IMS 子域[1]，基于 IMS，可以为用户提供丰富多彩的多媒体业务。IMS 可采用多种接入手段，GPRS 是其中一种。通过 GPRS 接入时，移动终端(MN，Mobile Node)通过 UTRAN(UMTS Terrestrial Radio Access Network)接入到 SGSN，通过 SGSN 的认证，然后由 GGSN 为其分配 IP 地址，并进行 PDP(Packet Data Protocol) Context 激活，从而获得 IP 连接[2]。

IMS 通过 CSCF(Call Session Control Function)来进行会话管理和控制。CSCF 有三种，分别是 P-CSCF(Proxy-CSCF)，I-CSCF(Interrogating-CSCF) 和 S-CSCF(Serving-CSCF)。P-CSCF 是 MN 接入到 IMS 的入口，一般位于访问网络(Visited Network)，I-CSCF 为 MS 选择提供服务的 S-CSCF，S-CSCF 位于归属网络(Home Network)，具体为 MS 提供注册及各种多媒体服务。

GPRS 中，当 MN 发送位置更新请求(Location Update Request)，附着请求(Attachment Request)等消息到 SGSN 时，SGSN 需要对 MN 进行认证。认证过程中，SGSN 向认证中心 HSS/AuC(Home Subscriber Server/Authentication Center)请求用户的身份信息，HSS/AuC 根据移动用户标识 IMSI(International Mobile Subscriber Identity)，返回一组认证向量 AVs，每一个向量由一个五元组(随机数 RAND，期待响应 XRES，认证令牌 AUTN，加密密钥 CK，完整性密钥 IK)构成。SGSN 选择下一个未使用的向量，使用 AKA 协议对 MN 进行认证。并将 CK 和 IK 发送到 UTRAN，MN 则利用 RAND 及与 HSS/AuC 的共享密钥等信息计算得到同样的 CK 和 IK，它们分别作为 UTRAN 与 MN 之间的加密和完整性保护密钥。

MN 若通过 GPRS 认证，并获得与 UTRAN 通信密钥之后，如果请求 IMS 服务，还需要进一步与 IMS 进行相互认证，并建立相应的安全通信手段。这里使用的同样是 AKA 协议。

在 IMS 应用中，两个认证过程往往前后进行，它们存在很多类似甚至完全相同的操作。为提高效率，有改进和优化的必要。文献[3]对此进行了分析研究，并给出一种优化方法，将规范[4]建议的 IMS 两回合认证过程优化为一回合认证过程。但[3]中方法有如下一些问题：(1)所提出的优化方法单纯考虑了对认证过程的简化，却忽略了在 GPRS 与 IMS 认证过程中，与认证同样重要的密钥分配问题，所提方

案中缺少了 MN 与 P-CSCF 之间建立安全连接所需要的密钥，缺少了这个密钥，整个 IMS 系统的安全性会受到很大威胁；(2)该方法缺少 MN 对 IMS 网络的认证；(3)该文对优化方法的分析中，将 P-CSCF，S-CSCF 和 I-CSCF 合在了一起，但它们的功能在本问题中有不同的作用，合在一起讨论并不合适。本文在分析 IMS 系统安全性及 GPRS 与 IMS 认证过程的基础上，参考[3]所提方案，提出一种新的优化方案，该方案在优化认证过程，提高认证效率的基础上，兼顾了 IMS 注册认证过程中的密钥分配及 MN 对 IMS 网络认证的需求。

本章后续内容是这样安排的，第 7.2 节描述 IMS 安全架构；第 7.3 节描述 GPRS 与 IMS 注册认证及密钥分配过程；第 7.4 节描述优化方法；第 7.5 节对优化方案进行可行性及性能分析；第 7.6 节是结论。

7.2 IMS 安全架构

2G 中，MN 到基站的无线接入有加密保护，但是核心网缺乏标准的安全解决方案。因为 2G 核心网相对封闭，且接入其中的点有限，安全问题不是很严重。3G 中，引入 IP 协议，并开放网络，使得核心网可能受到的安全威胁增大，核心网安全问题在 3G 网络管理中成为非常重要的内容[5]。3GPP 定义了一系列安全规范，尤其对核心网中的 IP 数据流的安全，用 NDS/IP (Network Domain Security)规范[6]来解决。

在 NDS 中，核心网划分为一个个安全域。一个安全域通常由一个运营商的网络组成，在安全域边界设置专门的安全设备——安全网关(SEG, Security Gateways)。每个安全域会有若干 SEG，进入或离开安全域的数据流都必须通过 SEGs。在 SEGs 上实施域的对外安全策略。安全域内部不同设备之间的安全接口由参考点 Zb 定义，安全域之间的接口则是参考点 Za。

在 IMS 网络安全方案中，接入网安全机制是单独定义的。Gm 参考点定义 MN 与 IMS 之间的接口，其安全基于 IPsec，规范[4]定义了 MN 与 IMS 网络如何相互认证的机制，同时定义了算法、密钥等的协商机制。

P-CSCF 是 MN 接入 IMS 的入口，在 MN 和 P-CSCF 之间需要建立 IPSec 安全关联(SA，Security Association)。这需要有共享的密钥。规范[6]6.3 节中，这个密钥使用 S-CSCF 在 401 响应中用 WWW-Authenticate 消息头发送给 P-CSCF 的 IK。IMS 网络安全域及各参考点位置如图 7.1 所示。

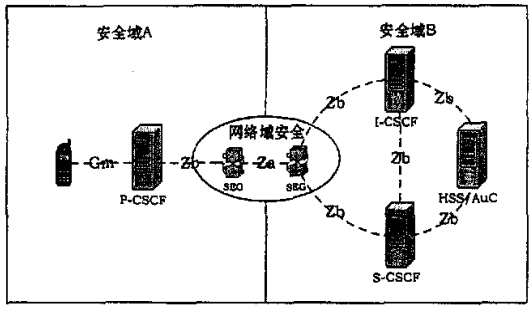

图 7.1 IMS 网络的安全域

7.3 GPRS 与 IMS 认证的过程

7.3.1 GPRS 注册认证及密钥分配

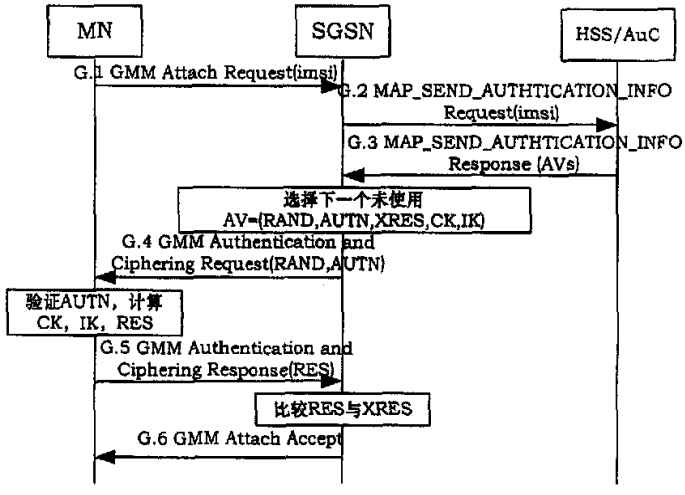

图 7.2 GPRS 注册及认证过程

MN 发起接入认证时，发送附着请求到 SGSN，引发 GPRS 认证过程。过程包含三个实体，MN，SGSN 和 HSS/AuC。MN 与 SGSN 之间使用 GPRS 移动性管理(GMM，GPRS mobility management)协议，SGSN 与 HSS/AuC 之间是 7 号信令的移动应用部分(MAP，Mobile Application Part)。步骤描述如下（参见图 7.2）：

G.1) 假设 MN 具有 IMSI 值 imsi，开始 GPRS 认证时，发送 GMM Attach Request 到 SGSN。

G.2) 如果 SGSN 需要从 HSS/AuC 获得认证向量 AVs，发送 MAP_SEND_AUTHENTICATION_ INFO Request 到 HSS/AuC，携带参数 imsi。

G.3) HSS/AuC 用 imsi 检索数据库，取得 MN 的用户数据，基于与 MN 的共享密钥 K 和序列号 SQN，计算一组 AVs，然后通过 MAP _SEND _AUTHENTICATION_INFO Response 到 SGSN。

G.4) SGSN 选择下一个没有使用的 AV，用 GMM Authentication and Ciphering Request 发送 AV 中的参数 RAND 和 AUTN 到 MN。

G.5) MN 通过检验收到的 AUTN，对网络进行认证，然后利用 RAND 和 AUTN 以及与 HSS/AuC 的共享密钥 K 和 SQN，计算加密密钥 CK，完整性保护密钥 IK，并计算 RES，最后，通过 GMM Authentication and Ciphering Response 将 RES 发回 SGSN。SGSN 比较 RES 和 AV 中的 XRES，如果相等，MN 通过认证。

G.6) SGSN 发送 GMM Attach Accept 消息到 MN，附着过程结束。

完成 GPRS 认证之后，MN 进行 PDP Context 激活，获得 IP 连接。

7.3.2 IMS 注册认证及密钥分配

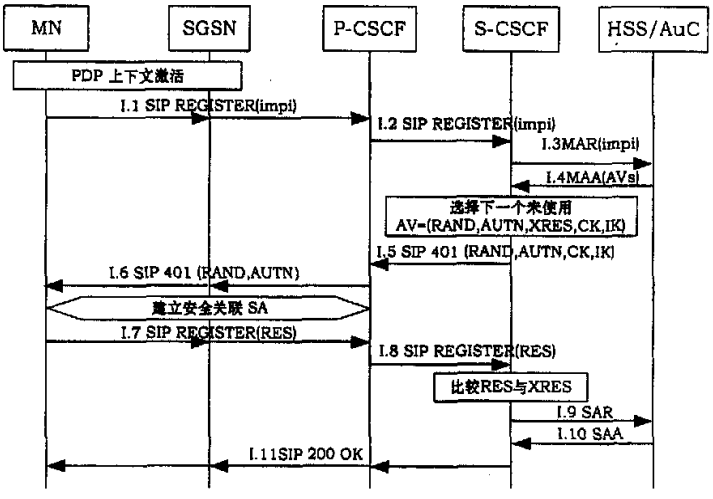

图 7.3 IMS 注册认证及密钥分配

IMS 中，通过 IMPI(IP multimedia private identity)来唯一识别用户，在 HSS/AuC 中，根据用户的 IMPI 值保存用户信息。

PDP Context 激活后，如 MN 请求 IMS 服务，需要进行 IMS 注册认证和密钥分配过程。具体步骤描述如下（参见图 7.3）：

I.1) 假设 MN 的 IMPI 值为 impi，MN 发送 SIP REGISTER 消息到 P-CSCF，

消息经过 SGSN。

I.2) P-CSCF 将消息发送到 I-CSCF，再到 S-CSCF。

I.3) S-CSCF 发送 Diameter 消息 Multimedia Authentication Request(MAR)到 HSS/AuC,消息中携带参数 impi。

I.4) HSS/AuC 通过参数 impi 检索数据库，取得 MN 的用户数据，基于共享密钥 K 和序列号 SQN_{IMS}，计算得到一系列 AVs，然后通过 Multimedia Authentication Answer(MAA)发送到 S-CSCF。

I.5) S-CSCF 选择下一个未使用的 AV(由五元组 AUTN，RAND，XRES，IK，CK 构成)，去掉 XRES，通过 401 响应将包含 AUTN，RAND，IK，CK 的消息返回到 P-CSCF。

I.6) P-CSCF 去掉 IK，CK，将 401 消息经由 SGSN 发送给 MN。IK 用于后面 P-CSCF 与 MN 建立安全关联(SA，Security Association)。

I.7) MN 通过共享密钥 K 和 SQN_{IMS} 检验 AUTN，如果校验通过，则网络认证通过。随后，MN 基于 K、RAND 和 AUTN，计算密钥 IK，作为和 P-CSCF 建立 SA 的基础，并计算 RES。MN 和 P-CSCF 通过 IPSec 消息建立 SA，通过安全通道，MN 发送新的 SIP REGISTER 消息发送到 P-CSCF。

I.8) P-CSCF 在验证消息没有被修改的基础上，将消息发送到 I-CSCF，再到 S-CSCF。S-CSCF 比较 RES 与 XRES，如果一致，则 MN 通过认证。S-CSCF 进行 SIP 注册操作。

I.9) S-CSCF 发送 Server Assignment Request (SAR)到 HSS/AuC，请求 HSS/AuC 保存 MN 与 S-CSCF 的服务关系。

I.10)HSS/AuC 发送 Server Assignment Answer (SAA)到 S-CSCF，作为 SAR 的响应。

I.11)S-CSCF 发送 200 响应，经 P-CSCF、SGSN，到达 MN。

通过这个过程，MN 和 IMS 网络互相认证了对方，在 MN 和 P-CSCF 之间分配了密钥。保证了后续通信的安全性。当 S-CSCF 有正确的未使用认证向量 AV 时，I.3)与 I.4)可跳过。

7.4 优化认证过程

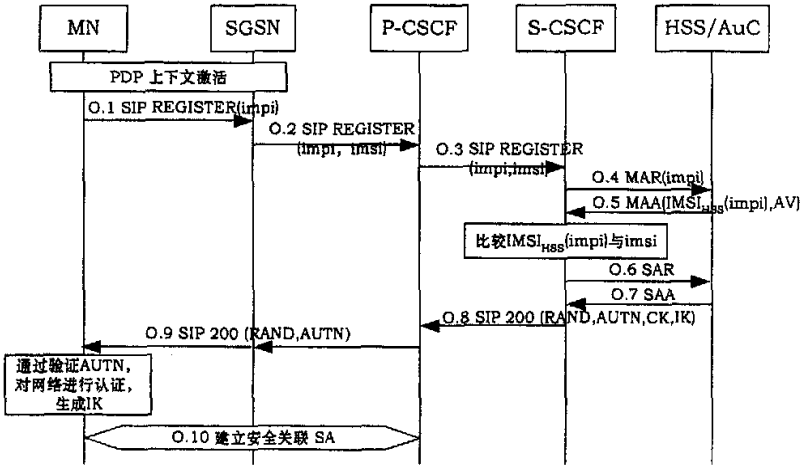

图 7.4 优化的 IMS 认证注册过程

如前所述，MN 在通过 GPRS 认证后，继续进行 IMS 注册认证和密钥分配，仍需要进行两回合的交互。我们建议的方法，只进行一个回合的认证，实现 MN 与 IMS 网络的双向认证，并正确分配密钥。在 SGSN 中需要实现一个 SIP 功能模块，处理 SIP 协议信令。IMS 对 MN 的认证采用与[3]介绍类似的方法，此外，增加了 MN 对网络的认证和 MN 与 P-CSCF 之间的密钥分配。步骤介绍如下（参见图 7.4）：

O.1) MN 发送 SIP REGISTER 消息到 SGSN，携带参数 impi。

O.2) 经过 GPRS 认证与 PDP 上下文激活，SGSN 了解 MN 的 IMSI 值(imsi)。它将该 MN 的 IMSI 值 imsi 添加到 REGISTER 消息中，然后发到 P-CSCF。

O.3) P-CSCF 将 SIP REGISTER(impi,imsi)消息发送到 S-CSCF(要经过 I-CSCF，此处省略)。

O.4) S-CSCF 保存(impi, imsi)对应关系。这里用 $IMSI_{HSS}$(impi)表示 IMPI 值为 impi 的 MN 的 IMSI 值。如本地没有记录 $IMSI_{HSS}$(impi)及认证向量 AVs，则发送 Diameter Multimedia Authentication Request(MAR)消息，携带 impi 参数，到 HSS/AuC，请求 IMSI 值。

O.5) HSS/AuC 基于共享密钥 K 和序列号 SQN_{IMS}，计算得到一系列 AVs，连同 $IMSI_{HSS}$(impi)，通过 Diameter Multimedia Authentication Answer (MAA)发送到 S-CSCF。S-CSCF 检查 $IMSI_{HSS}$(impi)与 imsi 是否相同，如果相同的话，MN 通过

认证。S-CSCF 为 MN 进行注册。

O.6) S-CSCF 发送 Diameter Server Assignment Request(SAR) 消息到 HSS/AuC。通知 MN 通过认证；请求 HSS/AuC 记下 MN 与为之提供服务的 SIP Server 的对应关系，需要时请求 MN 的用户信息。

O.7) HSS/AuC 发送 Diameter Server Assignment Answer(SAA)消息到 S-CSCF，作为 SAR 的响应。

O.8) S-CSCF 选择下一个未使用的 AV 向量，去掉 XRES，将其中的参数 RAND，AUTN，CK，IK，放在 200 响应中，发送到 P-CSCF。

O.9) P-CSCF 去掉 CK 和 IK，再将消息发送到 MN。MN 利用共享密钥 K 验证 AUTN，从而对 IMS 网络进行认证，同时计算 IK，作为和 P-CSCF 建立安全关联的密钥。

O.10) MN 与 P-CSCF 建立安全关联，作为后面通信的基础。

S-CSCF 在发送 MAR 请求 $IMSI_{HSS}$(impi)和 AVs 后，将进行本地保存，因此 O.4)、O.5)两步并非每次认证都需进行，不必要时可跳过

7.5 可行性证明及性能分析

7.5.1 IMS 网络对 MN 认证的正确性证明

第 7.4 节中 S-CSCF 通过检查 $IMSI_{HSS}$(impi)与 imsi 是否相同来对 MN 进行认证。本节进行证明。

UMTS 中，每个 MN 都存储 IMSI，IMPI 及 MN 与 HSS/AuC 的共享密钥 K。将 MN 的这三个属性设置为一个集合，即 S_{MN} = {IMSI, IMPI, K}。定义三个函数 $IMSI_{MN}$，$IMPI_{MN}$，K_{MN}，假设 MN 的 IMSI=imsi，IMPI=impi，K=k，对于任意 $x \in S_{MN}$，有：

$$IMSI_{MN}(x) = imsi, \quad (1)$$

$$IMPI_{MN}(x) = impi, \quad (2)$$

$$K_{MN}(x) = k, \quad (3)$$

类似的，对于任何一个 MN，设在 HSS/AuC 中，以集合的方式保存三元组 S_{HSS}={IMSI, IMPI, K}。对于合法的 GPRS 与 IMS 用户，$S_{MN} = S_{HSS}$。定义函数

$IMSI_{HSS}$，$IMPI_{HSS}$，K_{HSS}，并设 MN 的 IMSI=imsi，IMPI=impi，K=k，对于任意 $x \in S_{HSS}$，有：

$$IMSI_{HSS}(x) = imsi, \quad (4)$$

$$IMPI_{HSS}(x) = impi, \quad (5)$$

$$K_{HSS}(x) = k, \quad (6)$$

根据规范[4]和[6]，在 GPRS 和 IMS 中，对用户的认证，基于定理 1。

定理 1：假设 MN 申明其 IMSI 值为 imsi，IMPI 值为 impi，则有：

1)MN 是合法的 GPRS 用户，当且仅当 $K_{MN}(imsi) = K_{HSS}(imsi)$ ；

2)MN 是合法的 IMS 用户，当且仅当 $K_{MN}(impi) = K_{HSS}(impi)$ ；

在 AKA 协议认证中，避免直接检验密钥，采用计算摘要进行比较的方法，基于如下事实：

事实 1：

1)若一个 MN 申明其 IMSI=imsi，则当 XRES = RES 时，有 $K_{MN}(imsi) = K_{HSS}(imsi)$ ；

2)若一个 MN 申明其 IMPI=impi，则当 XRES = RES 时，有 $K_{MN}(impi) = K_{HSS}(impi)$ ；

证明当 $IMSI_{HSS}(impi) = imsi$ 时，MN 通过 IMS 认证，基础是 MN 已经通过 GPRS 认证，根据定理 1 和事实 1，就是满足了 $K_{MN}(imsi) = K_{HSS}(imsi)$，要证明 $K_{MN}(impi) = K_{HSS}(impi)$。将条件和结论进行整理，于是有以下定理。

定理 2：假设有以下条件：

1)IMSI 值为 imsi 的 MN 通过了 GPRS 认证，既有 $K_{MN}(imsi) = K_{HSS}(imsi)$；

2)MN 申明其 IMPI 值为 impi；

3)通过 impi 值从网络中得到相应的 IMSI 值是 imsi，即 $IMSI_{HSS}(impi) = imsi$；

则有：$K_{MN}(impi) = K_{HSS}(impi)$

证明：根据条件 1)，MN 具有 IMSI 值 imsi，则有 $imsi \in S_{MN}$，根据条件 2)，

MN 申明自己的 IMPI 值为 impi，则有 $impi \in S_{MN}$，根据(3)，

有 $K_{MN}(imsi) = K_{MN}(impi)$　　(7)

由条件 1)和式(7)有：

$K_{MN}(impi) = K_{HSS}(imsi)$　　(8)

由条件 3)，$IMSI_{HSS}(impi) = imsi$，集合 S_{HSS} 的定义及式(6) 有：

$K_{HSS}(imsi) = K_{HSS}(impi)$　　(9)

由 式(8)，(9)可得：

$K_{MN}(impi) = K_{HSS}(impi)$ 得证。

7.5.2 MN 对 IMS 网络的认证及密钥分发可行性证明

通过 IMS 认证和密钥分发，要达到的目的是 MN 与 IMS 网络相互得到认证，在 MN 与 IMS 网络接入点 P-CSCF 之间建立安全连接。这里我们通过分析证明本文介绍的优化方法可以实现上述目的。

S-CSCF 通过比较 $IMSI_{HSS}(impi)$ 与 imsi 对 MN 进行认证之后，选取下一个未使用的 AV 向量，将 XRES 之外的 AUTN、RAND、IK、CK，放在 200 响应中，可利用头域 WWW-Authenticate，举例如下：

```
SIP/2.0 200 OK
WWW-Authenticate: Digest realm="sip.bupt.edu.cn",
        nonce=ae2ikclKDIdeK093lslckewKurol,
        algorithm=AKAv1-MD5,
        ik="012345abdediowlelllee9888kekkfa",
        ck="983ldkkeooldfajkeillele0123l4555"
```

其中的 nonce 由 AUTN、RAND 及可选的服务器专用数据构成。当消息经过 P-CSCF 时，P-CSCF 从 WWW-Authenticate 头域中去掉 IK 和 CK，并保存，其它内容不变，发送到 MN。MN 接收到 200 响应后，从 nonce 中取出 AUTN(由三部分构成，如图 5 所示)、RAND，利用与 HSS/AuC 共享的密钥 K 及序列号 SQN，同[7][8]中介绍的 IMS AKA 协议中 MN 所进行的计算一样（如图 7.5 所示），计算得到 XMAC、RES、CK、IK。然后比较计算所得的 XMAC 与 AUTN 中的 MAC，如果相等，则说明 AUTN 来自可信任的网络，从而实现对 IMS 网络的认证。与此

同时，MN 得到了 CK、IK，同样的密钥也保存在 IMS 接入点 P-CSCF 中。这样就实现了 MN 与 IMS 网络入口 P-CSCF 之间的密钥交换。利用其中的 IK，MN 与 P-CSCF 可以建立起 IPSec 的 SA，从而为后边的通信建立安全通道。

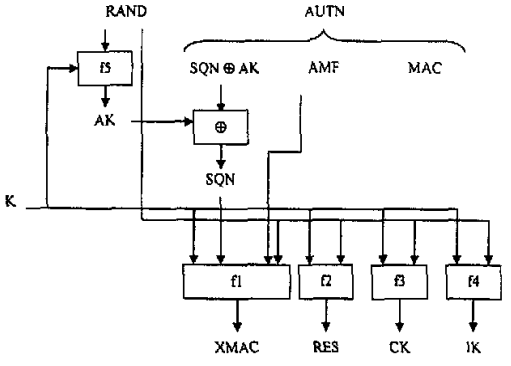

图 7.5 MN 对 IMS 网络认证及密钥生成算法

7.5.3 优化方法的性能分析

为方便分析，将SIP信令从MN出发，经由SGSN、P-CSCF、I-CSCF到达S-CSCF的信令开销作为一个整体考虑，并将一次信令传输开销归一化为 1，则 S-CSCF 与 HSS/AuC 之间的一次信令传输开销为 β。因 S-CSCF 与 HSS/AuC 通常处于同一域，而 MN 常常漫游到其它域，且 SGSN 常位于 MN 的接入域，所以可以认为 β 取值范围为 $0 \le \beta \le 1$。

对于非优化方案的 IMS 认证，定义 C_n 为总的信令开销，根据上边定义及图 7.3，可得到：

1) 当 I.3 与 I.4 执行时，$C_{n,1}=4+4\beta$

2) 当 I.3 与 I.4 跳过时，$C_{n,2}=4+2\beta$

设一次请求，分配 n 个 AV 向量，则有：

$$C_n=\left(\frac{1}{n}\right)C_{n,1}+\left(\frac{n-1}{n}\right)C_{n,2}=4+\left(\frac{n+1}{n}\right)2\beta \quad (10)$$

类似的，对于优化方案的 IMS 认证，我们定义 C_O 为总的信令开销，根据定义及图 7.4，可以得到：

1) 当 O.4 和 O.5 执行时，$C_{O.1}=2+4\beta$

2) 当 O.4 和 O.5 跳过时，$C_{O.2}=2+2\beta$

同上，设一次请求，分配 n 个向量，有：

$$C_O=\left(\frac{1}{n}\right)C_{O.1}+\left(\frac{n-1}{n}\right)C_{O.2}=2+\left(\frac{n+1}{n}\right)2\beta \qquad (11)$$

定义优化方案相对于非优化方案的改进 P 为：

$$P=\frac{C_n-C_O}{C_n}=\frac{n}{2n+(n+1)\beta} \qquad (12)$$

对 n 和 β 取不同数值，计算 P 值(表 7.1)，并根据数据画出曲线(见图 7.6)。可以看出，性能优化比 P 与 β 成反比关系，随着 β 值增大，P 值减小。当 n 为 1 时，若 β=0，含义是 S-CSCF 与 HSS/AuC 之间的一次信令传输开销相对于 MN 到 S-CSCF 的一次信令开销可以忽略不计，此时优化效果最好，达到 50%；β=1，此时含义是 S-CSCF 与 HSS/AuC 之间的一次信令开销与 MN 到 S-CSCF 的一次信令开销相当，因为优化减少的是 MN 到 S-CSCF 的信令开销，此时优化效果最差，但也可以取得 25%的优化。

从数据和图中还可以看出，一次取得的 AVs 向量数 n 对优化性能也有一定的影响。随着 n 值增大，曲线向上移动，优化作用变好。说明在优化方式下，AV 向量个数对 S-CSCF 与 HSS/AuC 之间信令开销的减少作用也大于非优化方式。

表 7.1 β, n 取值与性能优化比 P 对应表

P	n=1	n=3	n=5	n=7	n=9	n=12
β=0.0	0.500	0.500	0.500	0.500	0.500	0.500
β=0.1	0.455	0.469	0.472	0.473	0.474	0.474
β=0.2	0.417	0.441	0.446	0.449	0.450	0.451
β=0.3	0.385	0.417	0.424	0.427	0.429	0.430
β=0.4	0.357	0.395	0.403	0.407	0.409	0.411
β=0.5	0.333	0.375	0.385	0.389	0.391	0.393
β=0.6	0.313	0.357	0.368	0.372	0.375	0.377
β=0.7	0.294	0.341	0.352	0.357	0.360	0.363
β=0.8	0.278	0.326	0.338	0.343	0.346	0.349
β=0.9	0.263	0.313	0.325	0.330	0.333	0.336
β=1.0	0.250	0.300	0.313	0.318	0.321	0.324

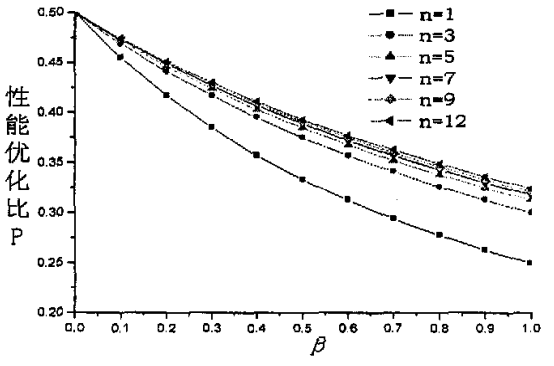

图 7.6 性能优化曲线图

7.6 本章小结

移动终端通过 GPRS 接入 IMS 时，先后通过 GPRS 的注册认证过程和 IMS 网络注册认证过程。两次注册认证过程先后进行，且使用的都是 AKA 协议，有优化的可能和需要。在相关研究成果的基础上，本文提出一种新的优化集成方法，可以只进行 GPRS 的完整 AKA 协议操作。IMS 网络可以基于 GPRS 的认证结果实现对移动终端的认证，同时满足移动终端对 IMS 的认证需求和双方的密钥分配需求。本文详细描述了该方法，对优化方法的可行性进行了分析和证明。最后进行了性能改进分析。

参考文献

1. 3GPP TS 33.228 V5.13.0—2004, IP Multimedia Subsystem (IMS); Stage 2(Release 5) [S].
2. 赵鹏,周胜,望玉梅(译).IMS:移动领域的IP多媒体概念和服务.北京：机械工业出版社，2005.
3. Lin Yibing, Chang Mingfeng, Hsu Mengta et al. One-pass GPRS and IMS authentication procedure for UMTS[J], Selected Areas in Communications, IEEE Journal on, 2005, 23(6): 1233- 1239.
4. 3GPP TS 33.203 V5.9.0—2003, 3G security; access security for IP-based services[S].
5. 詹志强，邱雪松，孟洛明. 第三代移动通信网的网络管理体系结构及实现技术[J]. 北京邮电大学学报，2003,26(2):9-13.

6. 3G TS 33.210 V5.5.0—2003, 3G security; network domain security; IP network layer security[S].

7. 3G TS 23.060 V4.9.0—2004, General packet radio service (GPRS); service description; stage 2[S].

8. 姚惠明，隋爱芬，杨义先.3GPP 网络AKA 协议中若干算法的设计[J].北京邮电大学学报，2002，25（3）：98-102.

结束语

NGN 是融合的网络，包括固定网和移动网的融合，也包括各种独立于电信网发展起来的无线组网技术，如 WLAN、Ad hoc 等，ITU-T 关于 NGN 的定义中明确说明 NGN 支持通用移动性，移动性是 NGN 中非常重要的部分。NGN 对移动性的支持构成 NGN 移动环境。人们对 NGN 的研究日益深入和广泛，这主要是由于 NGN 可以更好、更有效地提供以多媒体为主要特征的电信新业务，在 NGN 移动环境中提供多媒体业务，需要很多专用的设备，其中的媒体服务器是一种在软交换机控制下支持多媒体业务，提供各种媒体资源服务的专用设备。北京邮电大学网络与交换技术国家重点实验室网络智能研究中心，基于有关标准研发了基于软交换的媒体服务器。本论文第一部分针对其体系结构、控制协议、网管系统、安全防护等相关研究内容进行了论述。

NGN 移动环境，是一个研究内容非常丰富的领域，这一方面是因为移动通信本身魅力与 NGN 理念相结合产生出很多丰富多彩的构想，另一方面也是因为 NGN 移动环境有着巨大的包容性，它可以将现有的包括无线局域网、无线自组网、移动通信网、无线市话小灵通网在内的各种无线技术融合在一起。3GPP 提出的 IMS 架构是一种非常好的方案，在这个体系架构下，各种无线技术都可以作为一个部分接入到 IMS 核心网，可以用一种统一的方式接受 IMS 多媒体业务。

论文对 NGN 移动环境涉及到的安全威胁、接入安全性都进行了研究探索，基于这些研究，可以继续在以下方面进行研究：

1）安全是 NGN 是否能够大规模商用的一个至关重要的问题，NGN 的安全可以借鉴因特网安全研究的成果，但是二者还是有很多不同之处，与因特网安全相比，NGN 在可靠性、数据完整性、访问控制方面都有着更加严格的要求，如何在借鉴因特网安全研究成果的基础上，有针对性地解决 NGN 相关安全问题是一个很好的研究方向。

2)移动 IP 是解决 IP 层宏观移动性问题的重要协议，在新一代 IP 协议——IPv6 中，移动支持已经是成为协议的一个部分，但是目前 IPv4 无论在因特网中，还是在移动互联网中都还占据着统治地位。3GPP 规定 IMS 只支持 IPv6，而 3GPP2 在倾向于既支持 IPv6，也支持 IPv4，关于移动 IP 和 SIP 协议在移动性管理中协调配合问题，论文主要集中在 IPv4 与 SIP 的配合上，对于 IPv6 与 SIP 的配合是下一步的研究内容。

3）IMS 在 NGN 移动环境中的地位和作用日益增强，目前 IMS 又被看作实现

未来移动与固网融合的重要工具，对于 IMS 认证与安全接入，业务安全等内容可以进一步深入研究。

作者将以本论文取得的成果为基础，进一步研究探索相关的学术问题。由于研究内容涉及的知识面较广，限于本人的学识和精力，论文中存在不少不足之处，敬请批评指正。

攻读博士学位期间发表的论文

1. 徐鹏，廖建新，朱晓民，武家春. 移动IP与SIP集成应用中优化的AAA过程[J]. 通信学报，2006，27(1):1-7.
2. 徐鹏，廖建新，朱晓民等. GPRS与IMS认证与密钥分发的一种集成优化方法[J]. 北京邮电大学学报（EI 源刊）. 已录用.
3. Peng Xu, Jian-Xin Liao, Xiao-Ping Wen, Xiao-Min Zhu. Optimized integrated registration procedure of mobile IP and SIP with AAA operations. IEEE AINA2006 (EI source), Accepted.
4. 徐鹏，廖建新，吴乃星，马旭涛. SIP协议控制多媒体会话的研究与应用[J]. 计算机工程（EI 源刊），已录用（将发表于2006年第8期）.
5. Xu Peng, Liao Jianxin. Multi-protocols, Multi-layer Mobility Management Scheme[C]. ICCI2005 (EI source), Beijing, China, Pages: 589-593.
6. 徐鹏，廖建新，吴乃星，马旭涛. 基于软交换的集群媒体服务器的安全问题及其解决方案[J]. 计算机应用研究，已录用（将发表于2006年第6期）.
7. 徐鹏，廖建新，吴乃星，朱晓民. 基于软交换的集群媒体服务器网管系统的设计与实现[J]. 计算机系统应用. 2005，9：18-21.
8. 徐鹏，廖建新，吴乃星. 下一代网络中的移动性管理[J]. 计算机应用研究 ，已录用（将发表于2006年第4期）.
9. 徐鹏，廖建新. 以软交换为核心的下一代网络的发展前景及其在发展中可能面临的问题[C]. 第九届 全国青年通信学术会议论文集，重庆，2004.
10. 徐鹏，廖建新，吴乃星. 媒体服务器在NGN中的应用[J]. 电信网技术，已录用.
11. 吴乃星，廖建新，徐鹏，朱晓民. 一种基于软交换的集群媒体服务器的系统结构[J].电信科学, 2004, 20(7):11-15.
12. 吴乃星，廖建新，徐鹏，朱晓民. 一种基于软交换的集群媒体服务器的性能评价PETRI网模型[J]. 通信学报, 2005, 26(5): 73-79.
13. 朱晓民，廖建新，吴乃星，徐鹏等. 基于软交换的媒体服务器的研究与设计，计算机工程，第31卷，第13期，2005年7月，pp201-203 (EI indexed)

14. 吴乃星，廖建新，杨孟辉，徐鹏. 一种可用于基于软交换的集群媒体服务器的NDQAS接纳控制算法[J]. 北京邮电大学学报，已录用.

15. 马旭涛，朱晓民，徐 鹏，廖建新. 基于Parlay API的呼叫禁止业务的设计和实现[J].中国数据通信，2004，6(10):68-71.

16. 吴乃星，廖建新，徐鹏，马旭涛，基于软交换的媒体服务器的高可用系统建模与分析[J].通信学报，正在审稿.

已申请的专利：

由北京邮电大学及东信北邮信息技术股份有限公司资助提交一项国内专利：

“一种基于软交换的媒体服务器”

序列号. 200510002059.5

申请时间 2005年1月24日

致谢

在此，首先向导师廖建新教授表达深深的谢意，感谢他在我近四年的读博生活中，对我在学习、生活和工作各方面的指导和帮助。廖老师宽广的胸襟、儒雅的风度、渊博的知识和勤奋的工作态度时刻感染和激励着我，使我受益匪浅。

感谢实验室老师朱晓民博士对我的研究工作和论文写作给与的无私帮助。感谢实验室王晶、武家春、李炜等老师。

感谢吴乃星、杨孟辉、张奇支、马旭涛、雷正雄、徐童、焦利、郭伟等同学在学习生活中给与的友谊与帮助。感谢樊利民、杨波、张昊、刘军、黄亮、黄俊飞、杨戈、王敬宇等实验室师兄弟，和他们相处的日子十分愉快，与他们的交流、讨论使我受益良多。

感谢我亲爱的妈妈，妈妈无微不至的关怀、全心全意的支持和鼓励，是我在遇到困难时，能够坚强起来，坚持完成学业的主要力量。感谢爸爸、姐姐和所有爱我的人，他们的爱和支持永远是我生命中宝贵的财富。

感谢我永远的的母校——北京邮电大学！

感谢各位评委老师在百忙之中审阅我的论文！

徐 鹏

2006年1月31日

NGN移动环境中多媒体业务支持及移动性管理相关问题研究

作者： 徐鹏
学位授予单位： 北京邮电大学

相似文献(5条)

1.期刊论文 夏竞辉 NGN在融合与创新中演进 -中国电信业2007(4)

NGN的目标:能够提供多种业务的融合网络

融合可以向用户提供各种形式的业务和一站式的服务,使用户不管是在固定环境中还是在移动环境中都能享受同样的服务;融合还给运营商带来增加收入的机会,减少引进新业务的风险,特别是适合全业务的经营.而NGN的目标就是这样一个能够提供多种业务的融合网络.

2.期刊论文 雷震洲 未来移动通信的定位与应用 -移动通信2006,30(1)

1未来移动通信的定位

从服务的角度看,移动通信最初是为了在移动环境中打电话而发明的,但是21世纪的移动通信绝不是单单为了打电话,它将成为下一代网(NGN)的重要组成部分.下一代网只是手段,下一代服务才是目的.下一代的宽带服务与应用要比具有100多年历史的电话服务复杂得多,甚至是包罗万象的.通信方式不仅是人与人之间,而且还包括人机之间和机-机之间的通信.未来移动通信在社会进步中将起非常重要的作用,主要体现在向我们提供宽带多媒体、全球性、个性化和无所不在的服务.

3.学位论文 张淼 SCTP在NGN中的应用与性能研究 2006

随着移动通信网络在经历了2代的GSM、2．5代的GPRS，发展到目前以TD-SCDMA、WCDMA和CDMA2000为主的第三代系统标准，逐步显现出了对多媒体业务、以及无线传输带宽的巨大需求，这一切都推动了下一代网络的发展，为了使SCTP在NGN中有更为广泛的应用，SVWGZE作组不断发布新的扩展协议，但是目前这些扩展协议都只是停留在探讨阶段，对其在NGN中的具体传输性能和应用的研究仍很缺乏。因此，本论文着眼于新型传输协议流控制传输协议SCTP在下一代网络中的应用与性能研究。

本文从理论上分析了SCTP最新的两种扩展协议：移动SCTP和SCTP非可靠扩展，这是针对SCTP移动性访问所面临的问题专门提出的，扩展草案中提出了利用SCTP进行传输层的位置管理和切换管理的设想，本文通过对这个草案及相关文献的研究，提出了结合SIP、MobileIP和RserPool的位置管理和切换管理方法和流程，从而便于SCTP应用到移动环境中。

为了能够公平比较传输协议的差异性，仿真采用了目前NGN中使用很广泛的SIP协议作为应用层协议，SCTP性能的分析主要是通过与现有的两个传输层协议UDP和TCP传输时延的比较得出的，文中分别采取了仿真图示和数理统计的方法进行了比较归纳，得出结论基本符合理论分析结果。

移动Internet的业务将以多媒体业务为主，目前主要依赖传输()协议TCP来保证可靠传输，但是SCTP虽然是为了传输信令提出的，其先进的特性使其应用的前景十分广阔。结合SCTP的扩展协议，本文构建了基于SCTP的移动Internet的三种传输模型，分析了SCTP特性在移动性访问接入性能上带来的提高，文章最后总结归纳了SCTP在NGN中其他的一些应用。

4.期刊论文 雷震洲.Lei Zhenzhou 未来移动通信的定位与应用 -世界电信2005,18(9)

未来移动通信的定位

从服务的角度看,移动通信最初是为了在移动环境中打电话而发明的.但是21世纪的移动通信绝不是单单为了打电话,它将成为下一代网(NGN)的一个重要组成部分.下一代网只是手段,下一代服务才是目的.下一代的宽带服务与应用要比具有100多年历史的电话服务复杂得多,甚至是包罗万象的.通信方式不仅是人与人之间,而且还包括人-机之间和机-机之间的通信.未来移动通信在社会进步中将起到非常重要的作用,主要体现在向我们提供宽带多媒体、全球性、个性化和无所不在的服务.

5.学位论文 魏彬彬 基于SIP的无线数字签名系统的研究与应用 2007

随着移动通信和互联网逐渐成为信息产业的两大支柱，无线通信技术在银行、证券、商务、贸易、办公、教育等各方面的需求越来越多，无线通信的安全性也显得日益重要，因而WPKI技术也渐渐发展起来，它为解决移动环境下的安全认证奠定了基础。

信息安全是电子商务的重要组成部分，而数字签名在信息安全中占有举足轻重的地位，它能保证信息的完整性、有效性和防抵赖，随着电子签名法的实施以及无线电子商务的迅速发展，简单方便的用户名加口令的方式已不再安全，传统的数字签名方式也未能很好的满足需要，为了保证移动性和安全性，无线数字签名成为了无线电子商务中一个必不可少的要求。

RFC3261中规范的SIP协议是NGN中一个重要的应用层控制协议，可以用来建立、修改和终止多媒体会话，实现基于IP的网络中的软交换；WPKI是传统的PKI技术应用于无线环境的优化扩展，用于有效建立安全和值得信赖的无线网络环境。

本文设计与模拟实现了一个基于SIP的、适用于基于IP的无线网络中的无线数字签名系统，它以Java，为基础平台，包括SIP子系统、CA子系统以及安全传输子系统等，使用SIP协议进行呼叫与控制，典型应用有在手机上进行的无线支付、证券交易和文档签署等，可以看作是WPKI在实施方式上的一种创新。

无线数字签名的应用前景非常广阔，市场潜力也非常巨大。本文设计的基于SIP的无线数字签名系统既能作为当前电信运营商的一项增值业务，也可以作为独立的服务提供商或集成商运作，这也为不少公司打开了一扇进入电信领域之门。