



信息通信领域产学研合作特色期刊

第三届国家期刊奖百种重点期刊 | 中国科技核心期刊

ISSN 1009-6868

CN 34-1228/TN

中兴通讯技术

ZTE TECHNOLOGY JOURNAL

www.zte.com.cn/magazine

2015年4月 • 第2期

专题：移动云计算和云服务



《中兴通讯技术》第6届编辑委员会委员名单

主 任 钟义信 (北京邮电大学教授)

副主任 侯为贵 (中兴通讯股份有限公司董事长) 糜正琨 (南京邮电大学教授)

编委 (按姓氏拼音排序)

艾 波 中国联通学院党委书记, 北京大学兼职教授, 北京邮电大学兼职教授

曹淑敏 中国信息通信研究院院长、教授级高工

陈建平 上海交通大学电子信息与电气工程学院电子工程系教授

陈 杰 中兴通讯股份有限公司高级副总裁

陈前斌 重庆邮电大学通信与信息工程学院执行院长、教授

陈锡生 南京邮电大学通信与信息工程学院教授

程时端 北京邮电大学网络技术研究院教授

高 文 中国工程院院士, 北京大学信息科学技术学院教授

葛建华 西安电子科技大学通信工程学院副院长、教授

顾晓仪 北京邮电大学信息光子学与光通信研究院教授

郭云飞 解放军信息工程大学副校长、教授

管海兵 上海交通大学电子信息与电气工程学院副院长、教授

侯为贵 中兴通讯股份有限公司董事长

何士友 中兴通讯股份有限公司执行副总裁

洪 波 中兴发展股份有限公司总裁

纪越峰 北京邮电大学信息光子学与光通信研究院执行院长、教授

江 华 中兴通讯股份有限公司副总裁

蒋林涛 中国信息通信研究院科技委主任、教授级高工

雷震洲 中国信息通信研究院科技委副主任、教授级高工

李红滨 北京大学信息科学技术学院教授

李建东 西安电子科技大学副校长、教授

李乐民 中国工程院院士, 电子科技大学通信与信息工程学院教授

李融林 华南理工大学电子与信息学院教授

李少谦 电子科技大学通信与信息工程学院院长、教授

李 星 清华大学电子工程系教授, CERNET 网络中心副主任

马建国 天津大学电子信息工程学院院长、教授

孟洛明 北京邮电大学职称评审委员会主任, 学术委员会副主任, 网络与交换技术国家重点实验室主任、教授

糜正琨 南京邮电大学通信与信息工程学院教授

庞胜清 中兴通讯股份有限公司高级副总裁

史立荣 中兴通讯股份有限公司总裁

孙枕戈 中兴通讯股份有限公司副总裁

孙知信 南京邮电大学物联网学院院长、教授

谈振辉 北京交通大学电子信息工程学院教授

唐雄燕 中国联通网络技术研究院首席专家、教授级高工

田文果 中兴通讯股份有限公司执行副总裁

童晓渝 中电科软件信息服务有限公司董事、副总经理

王 京 清华大学教授, 清华大学信息科学与技术国家实验室副主任

王文东 北京邮电大学软件学院党委书记兼副院长、教授

王晓明 中兴通讯微电子研究院院长

王育民 西安电子科技大学通信工程学院教授

韦乐平 中国电信集团公司科技委主任、教授级高工

卫 国 中国科学技术大学信息科学技术学院教授

郭贺铨 中国工程院原副院长、院士

吴克利 香港中文大学电子工程系教授

徐安士 北京大学信息科学技术学院教授

须成忠 美国韦恩州立大学电子与计算机工程系主任、教授, 中科院深圳先进技术研究院首席科学家、云计算中心主任

续合元 中国信息通信研究院通信标准研究所总工、教授级高工

薛一波 清华大学 CPU&SOC 中心副主任, 信息技术研究院研究员

杨义先 北京邮电大学信息安全中心主任, 灾备技术国家工程实验室主任、教授

杨 震 南京邮电大学校长、教授

尤肖虎 东南大学信息科学与工程学院院长、教授

乐光新 北京邮电大学信息与通信工程学院教授

张宏科 北京交通大学下一代互联网国家工程实验室主任、教授

张 平 北京邮电大学网络技术研究院执行院长、教授

张同须 中国移动通信集团设计院副院长兼总工

赵厚麟 国际电信联盟 (ITU) 副秘书长、俄罗斯电信科学院院士

赵慧玲 中国电信股份有限公司北京研究院总工程师、教授级高工

赵先明 中兴通讯股份有限公司执行副总裁

郑纬民 清华大学计算机系高性能计算研究所所长、教授, 中国计算机学会理事长

钟义信 北京邮电大学计算机学院教授

朱近康 中国科技大学信息科学技术学院教授



第三届全国期刊奖百种重点期刊
中国科技核心期刊
工信部优秀科技期刊
中国五大文献数据库收录期刊
ISSN 1009-6868
CN 34-1228/TN
1995年创刊

办刊宗旨

以人为本,荟萃通信技术领域精英;
迎接挑战,把握世界通信技术动态;
立即行动,求解通信发展疑难课题;
励精图治,促进民族信息产业崛起。

Contents 目次

中兴通讯技术 总第121期 第21卷 第2期 2015年4月

专题:移动云计算和云服务

- 02 关于移动云计算的几点思考 何宝宏
04 按需供给的移动云计算动态安全服务 陈小华,董振江,金怡爱
10 移动云存储服务关键技术研究 崔勇,赖泽祺,缪葱葱
14 面向5G的分布式移动云计算协同架构与管理机制 李宏佳,陈鑫,周旭
20 移动云计算与运营商自有业务云化探讨 符刚,薛淼,唐雄燕
25 移动云计算——移动增强现实技术和服务 吕强,黄成,刘明
30 网络功能虚拟化及其标准化 薛海强,张昊
35 云计算安全框架分析 陈清金,陈存香,李晓宇

专家论坛

- 39 卫星平台虚拟化——满足多种卫星移动通信需求的必然选择 王京,赵明,高镇
44 ITU-R 卫星移动通信标准——整合移动业务的动力 金秀英

技术广角

- 48 基于中断概率的多输入多输出异构网络资源分配 栾天祥,高飞飞
53 基于MPI和OpenCL多层次并行图像卷积算法设计 王继刚,刘惠,姜滨

企业视界

- 56 M-ICT时代的业务智能化发展趋势 董振江,赵培,杨勇

综合信息

中国电信将构建第三张全国骨干网 (09) 传感器应用领域不断扩大 迎来大好发展 (34) 国家集成电路基金宣布 2015 年将向芯片业投资 200 亿 (38) 2014 年全球光纤出货量达 2.96 亿芯公里 同比增长 13% (52)

期刊基本参数:CN 34-1228/TN*1995*b*16*64*zh*P*¥ 20.00*15000*13*2015-04

Contents 目次

ZTE TECHNOLOGY JOURNAL Vol. 21 No. 2 Apr. 2015

Special Topic: Mobile Cloud Computing and Cloud Service

- 02 Reflections on Mobile Cloud Computing HE Baohong
- 04 On-Demand Mobile Cloud Computing
Security Service CHEN Xiaohua, DONG Zhenjiang, JIN Yiai
- 10 Key Techniques in Mobile Cloud Storage Services CUI Yong, LAI Zeqi, MIAO Congcong
- 14 Cooperative Management Architecture and Mechanism
of 5G-Oriented Distributed Mobile Cloud Computing LI Hongjia, CHEN Xin, ZHOU Xu
- 20 Mobile Cloud Computing and Cloud-Based
Operator Services FU Gang, XUE Miao, TANG Xiongyan
- 25 Mobile Augmented Reality Technology and Service
in Mobile Cloud Computing Lü Qiang, HUANG Cheng, LIU Ming
- 30 Network Function Virtualization and Standardization XUE Haiqiang, ZHANG Hao
- 35 Cloud Security Architecture CHEN Qingjin, CHEN Cunxiang, LI Xiaoyu

Expert Forum

- 39 Virtualizing the On-Board Platform for Multi-Purpose Mobile
Satellite Communication Systems WANG Jing, ZHAO Ming, GAO Zhen
- 44 Standardization of Mobile Satellite Service Technology
in ITU-R: Driving Force of Integrated Mobile Services KIM Sooyoung

Technology Perspective

- 48 Outage-Probability-Based Resource Allocation
for MIMO Heterogeneous Networks LUAN Tianxiang, GAO Feifei
- 53 Design of Multi-Level Parallel Image Convolution Algorithm
Based on MPI and OpenCL WANG Jigang, LIU Hui, JIANG Bin

Enterprise View

- 56 Business Intelligence in the Era of M-ICT DONG Zhenjiang, ZHAO Pei, YANG Yong

敬告读者

本刊享有所发表文章的版权,包括英文版、电子版、网络版和优先数字出版版权,所支付的稿酬已经包含上述各版本的费用。

未经本刊许可,不得以任何形式全文转载本刊内容;如部分引用本刊内容,须注明该内容出自本刊。

邮购须知

本刊常年办理邮购订阅业务,欢迎订阅。订阅方法:从邮局汇款至编辑部,在汇款单上将订阅者的详细地址、收件人姓名及联系电话填写清楚,并在汇款单附言栏注明所购杂志期次及数量。



唐雄燕

中国联通网络技术研究院首席专家、教授级高工，中国联通科技委固网专业副主任，北京邮电大学兼职教授、博士生导师，工业和信息化部通信科技委委员兼传送与接入专家咨询组副组长，中国通信标准化协会 TC10 副主席；主要专业领域为宽带网络、光纤通信、IP 技术、下一代网络、物联网、未来网络等；已发表技术论文 150 余篇，出版专著 6 部。

专题导读

终端的移动化、服务的云化和网络的宽带化已成为信息通信业的基本趋势，移动互联网、物联网、云计算和大数据正在驱动新一轮的信息通信产业大变革、大融合和大发展。移动通信与云计算的结合为打造全新的信息通信商业模式创造了技术条件，随着移动智能终端的日益普及和移动宽带的加速发展，基于移动终端的移动云服务必将成为移动互联网的主流形态，有着广阔的市场前景和创新空间。今天我们看到无论是个人用户还是政企用户对移动云服务的需求正在快速增长，不但互联网服务提供商的大量创新应用（如视频、音乐、游戏、商务、金融、医疗等）以移动云服务的方式提供，电信运营商的通信网络和业务也在日趋云化。对于移动云服务，不但终端是移动的，数据与应用也可移动，“端、管、云”在技术和产业链方面的协同对于移动云服务至关重要，应用开发商、云计算基础提供商、终端开发商和电信运营商在移动云服务产业中都扮演着不可或缺的角色。“端、管、云”的协同不仅能为用户提供无处不在的优质用户体验，也带来了更高效的资源利用和更灵活的业务创新。

移动云计算既是新兴的 ICT 技术，也是新兴的业务应用，在发展中还面临技术标准、商业模式、服务质量保障和安全等问题。本期专题重点关注移动云计算与云服务相关的技术、应用和安全问题，在本期专题中读者既能见到行业权威专家对移动云计算的宏观思考，也有多篇有关具体解决方案的文章。网络和信息安全应该是移动云服务产业发展中广受关注的问题，本专题将其作为重点之一，收录了多篇有关云计算安全的文章。业务应用和商业模式创新是移动云服务发展的出发点和落脚点，本专题收录了移动云存储、移动增强现实以及移动通信业务云化等移动云应用方面的文章。云计算的理念和技术也在向通信网络领域渗透，云化的网络架构和网络功能虚拟化就是这方面的典型，本专题还收录了面向 5G 的移动云计算协同架构和 NFV 方面的文章。希望本专题能帮助读者更全面深入地了解移动云计算与云服务领域的技术进展和最新动态，为进一步的研究和应用提供有价值的参考。

唐雄燕

2015 年 3 月 10 日

2015 年第 1—6 期专题计划

1

自组织异构小基站网络

张平 北京邮电大学网络技术研究院 执行院长

2

移动云计算和云服务

唐雄燕 中国联通网络技术研究院 首席专家

3

移动互联网安全技术

刘建伟 北京航空航天大学电子信息工程学院 教授

4

软件定义光网络

迟楠 复旦大学信息科学与工程学院 教授

5

虚拟运营业务和网络

续合元 中国信息通信研究院通信标准研究所 总工

6

移动群智感知和协同计算

王文东 北京邮电大学软件学院 教授

关于移动云计算的几点思考

Reflections on Mobile Cloud Computing

何宝宏/HE Baohong

(中国信息通信研究院, 北京 100083)
(China Academy of Information and
Communications Technology, Beijing
100083, China)

中图分类号: TN929.5; TP393 文献标志码: A 文章编号: 1009-6868 (2015) 02-0002-002

摘要: 认为云计算正在从炒作进入理性发展阶段, 但仍然面临不少有待解决的问题。基于此, 探讨了发展移动云计算所面临的一些关键问题, 比如对移动性的理解和分工, 信任体系如何建立, 存储介质的寿命, 软件定义可靠性以及商业模式的未来等。指出对这些问题的探讨可以加深云计算未来发展的理解。

关键词: 移动互联网; 云计算; 可信云; 存储介质; 软可靠性; 商业模式

Abstract: From all the initial hype, cloud computing has developed into something quite substantial. This paper discusses key issues in cloud computing, including how to understand mobile management and cooperation in cloud, how to build a trusted cloud system, the influence of short-lived storage media, software-defined reliability, and the future business model of cloud service. Such discussion can deepen our understanding of future cloud development.

Key words: mobile internet; cloud computing; trusted cloud; storage media; software defined reliability; business model

1 关于移动性

互联网最初的典型终端是计算机, 采用键盘、鼠标和视窗模式, Wintel 架构, 有线连接。终端的能耗不是问题。移动互联网的典型终端是智能手机和平板电脑, 或只具有无线局域网通信能力(如平板电脑)再通过本地网关连接到互联网, 或直接具有无线广域网通信能力(如智能手机), 但从传输控制协议/网间协议(TCP/IP)的角度看, 本质上都是 PC^[1]。

移动互联网的终端、应用甚至数据都具有移动性, 在带来诸多优势的同时, 新的限制开始出现了。主要利用电池供电, 使得能耗上升为最关键的限制因素; 小屏幕使得 I/O 系统也无法继续沿用鼠标键盘做 I/O 了; 处理性能与 PC 的差距使得通用平台型的 Web 受限。

这些因素导致了节能型芯片 ARM 的崛起, I/O 系统融合为采用多点触控技术, PC 互联网 10 多年前流行的客户机/服务器(C/S)模型(即 APP Store 模式)获得新生。移动终端无法胜任的很多工作, 尤其是大部分

存储和部分计算工作, 通过无线和应用商店(APP Store), 被巧妙地转移到了云平台上。这种转移本应该是彻底的, 是超文本链接标记语言 5 (HTML5) 方式的, 但因为与 PC 互联网的性能“代差”, 因为操作系统厂家的利益, 目前主流还是采用 Native App 方式。多点触控技术直接改变了用户使用习惯, 使得操作系统革命成为可能, 于是面向移动和面向多点触控的新一代操作系统 iOS 和 Android 发展起来, 一个闭源一个开源, 和当年 Windows 对应 Linux, DOS 对应 UNIX 一样的格局。

如果说移动互联网是从用户侧发起的一场技术革命, 那么云计算就是从服务侧发起的另一场技术革命, 两者遥相呼应。APP Store 模型让移动互联网软硬一体化和终端服务化, 云计算则是让服务侧也开始了软硬一体化和服务化, 面向互联网的大规

模服务器定制(以 Google 和 Facebook 为代表, 中国以 BAT 牵头的天蝎整机柜服务器为代表)和软件定义网络(SDN)技术, 像水电那样的提供公共计算资源。

使用云计算服务的用户, 可以使用像 PC 和电视机那样的固定终端, 也可以是平板或智能手机那样的移动终端, 甚至可以是专门设计生产的云终端(比如虚拟桌面业务的瘦客户端)。

从 ISO/OSI 网络模型看, 移动性也是要分层要分工的。移动性可以是链路层的, 可以是网络层或 TCP 层的, 也可以是应用层的甚至数据层的。移动性的分工, 不仅会发生在垂直层面, 水平层面的终端、网络、应用/数据等之间也有移动性如何分工协作问题。

在移动互联网环境下, 在云计算的支持下, 目前市场上移动性如何分

收稿日期: 2015-01-03
网络出版时间: 2015-02-22

层和分工非常混乱,都想获得更多的控制权和利益,都想进入对方的领域,经常以一种很笨拙的方式支持了移动性。即使不区分终端,云计算自身也已经具有很强的应用移动性(如SaaS)和数据移动性(如同步功能)。长期看,上层的移动性支持会越来越重要,三层以下的移动性支持会逐步沦落为“移动比特”或者说“哑移动”。

2 关于信任

云计算本质上是一项数据托管服务,所有托管服务成立的基础是信任,是一种“信托”。多项市场调研^[2]均表明,云计算需要赢得用户更多的信任。当把用户数据保存到云服务商处时,将会因为托管而产生一些新的安全顾虑。这一幕也是提供金钱托管服务的银行业早期也曾经遇到过的困境。

数据是21世纪的石油,具有资产的明显属性。经过400多年的发展,银行业根据潜在的风险,制订了一系列的体系、规范和制度,对资金和IT的风险管控都已经积累了比较丰富的经验。多项调查均表明,用户对云服务的信任程度依然不高。围绕云服务的合规性,建立起一整套云计算和数据的标准和规范,是一种必然的选择。

云计算服务的规范性至关重要。云服务商必须以事先约定的规范方式,公开一些用户最关心的问题(如数据安全、服务质量和责权利),并对照做出相应的承诺。这是赢得用户更多信任度的关键。

数据控制权就是数据财产的安全。现代金融业发展到今天,资金风险的管控已经相对完善。而对于云服务,目前还很初级,还处于如何取得信任过程中。

2014年,云计算发展与政策论坛与数据中心联盟推出的“可信云服务”标准与认证^[3],得到了业界绝大多数云服务商的支持,做出了一些有益

的探索。

3 关于可靠性

云计算是一种“软可靠”技术,即假设低层的硬件不可靠,通过软件和算法来改善整个系统的可靠性。电信业普遍采用的是一种“硬可靠”的方法,比如设备的双备份、网络的环形拓扑和保护倒换,电力的双路供电和备用柴发等,当然还有7×24小时待命的运维人员。冗余是解决可靠性的基本套路,无论是软件、硬件还是人员。冗余的代价是成本大增,效率大降。

随着相关技术的进步,以及设备和零配件的标准化,以云计算为代表的技术,现在越来越多地通过“网络+软件+服务”的新模式实现^[4],或者说是“软件定义的冗余”。很明显,这种冗余和要求数据必须是“可漫游”的。传统的硬件冗余追求的是单机设备的零故障,而软冗余追求的系统性冗余,允许单个硬件故障的发生,但用服务和软件的方式加以解决。软冗余相比硬冗余的优势明显,建设和运营成本都会降低。

在TCP/IP的设计中,假设互联网终端是可靠的。但互联网终端(比如服务器)出现宕机的问题,不在当时的设计考虑中。解决互联网服务侧可靠性的思路,之前主要还是靠“硬冗余”(多机备份、集群和切换等)。现在服务器的可靠性,也从之前的集群等硬冗余技术,转向软件方式,就是云计算技术。云计算将互联网技术的“软冗余”思路,从IP技术发展到了应用,从网络设备发展到了终端设备。当TCP/IP网络发生故障时,IP技术将流量迁移到其他路由上去;当计算(主要指服务器设备)出现问题时,云将计算服务迁移到其他IT设备甚至其他数据中心去。

未来必将会有更多的“软冗余”技术和应用场景出现^[5]。数据中心就是一台计算机,制冷和配电应该与计算、存储和网络一样池化,也是可以

统一调度的资源,但现在还缺少一个面向数据中心的操作系统。

4 关于商业模式

70年代前软件业的主要生存模式是“买计算机送软件”,软件还不是商品,也没有第三方市场,是寄生的。这时软件的来源主要有3个,一是计算机制造商买机器时赠送的通用软件和工具软件,二是计算机制造商协助其客户开发的软件,三是客户自己开发的应用软件。

PC产业兴起后,计算机架构和操作系统趋于开放,面向个人用户的软件需要爆发式增长。这时无无论是计算机厂家还是用户自身,都已无力面对,迫切需要一个通用软件的市场,将计算机制造商和客户从重复而繁重的软件开发任务中解放出来,这是社会化分工的基本要求。

软件产业的独立,首先是商业模式的独立。需要用“软件思维”创造出一种新的商业模式,正如今日要用“互联网思维”做互联网一样。比如,石油、汽车等工业无法形成全球性的垄断企业,因为随着规模的进一步扩大边际效应是递减的,而软件业却是边际效应递增的。软件业独立运动的结果,是发明了“版本控制”、“许可控制”、“二进制代码”、“API”以及Wintel生态系统建设等典型商业模式,这是“软件思维”创新的结果。

开源软件是1990年以后新兴的开发和商业模式。Linux是最早也是最成功的开源软件,是基于互联网开发复杂软件“集市模式”的成功典范,也是今日Android的基础,是所谓“国产”操作系统的基础。现如今,开源也成了软件业市场营销,解决“产能过剩”的一种手段,全球至少1/3的软件市场与开源软件相关。

但无论是闭源还是开源,核心还是产品,服务还是附属地位。软件是工具,服务才是目的。如果用户可以直接得到服务,为什么还要去购买软

➡下转第43页

按需供给的移动云计算动态安全服务

On-Demand Mobile Cloud Computing Security Service

陈小华/CHEN Xiaohua
董振江/DONG Zhenjiang
金怡爱/JIN Yiai

(中兴通讯股份有限公司, 深圳 518057)
(ZTE Corporation, Shenzhen 518057, China)

瑞星公布的数据显示, 2014年1—6月, 瑞星系统共截获新增病毒样本3 032万个, 病毒总体数量比2013年同期增长85.67%, 呈现出爆发式增长趋势, 其中大部分是木马病毒。电子邮件、即时通信工具、网上下载的电影和MP3、网页上被植入木马等, 只要移动终端接入互联网, 就会立刻面临木马病毒的包围。大部分普通用户在感染了木马病毒后, 根本没有感觉, 不懂得如何通过检查注册表、可疑进程等信息进行分析并上报。普通的病毒分析工程师, 每天最多能分析的病毒有限, 面对成几何级数爆炸增长的新木马病毒, 反病毒公司难以承担如此严峻的挑战。如果依然沿袭以往的反病毒模式, 安全厂商将被淹没在木马病毒的汪洋大海中。

云计算改变了IT设施的部署与使用模式, 提供了计算存储等资源的弹性供给模式, 效率更高使用更灵活。安全业界已经开始利用云计算技术, 在移动云计算环境中, 安全服务已经作为一项新的业务模式, 受到

中图分类号: TN929.5; TP393 文献标志码: A 文章编号: 1009-6868 (2015) 02-0004-006

摘要: 提出一套按需供给移动云计算安全服务解决方案, 提升了云计算环境中安全资源的使用率, 充分保障云服务的安全。该方案针对移动互联网环境, 基于云计算提供服务的业务, 满足不同用户、不同业务对安全的复杂、多变、多样化的安全需求, 能够动态、实时地感知并分析安全态势、调整安全执行策略, 尽可能地为用户降低安全支出成本, 为不同用户、不同安全状态下的移动互联网业务提供差异化、高效的安全防护机制。

关键词: 云计算安全; 按需供给; 服务; 移动互联网

Abstract: This paper presents on-demand mobile cloud security solutions. These solutions guarantee security for cloud computing business and improve utilization of security resources in the cloud environment. The program focused on the cloud-based services business in the mobile Internet environment. It can meet complex, varied security needs of different users and services and can dynamically analyze the security situation and adjust security policies. It is as much as possible for users to reduce security costs for different users in different security state to provide differentiated and efficient security mechanisms in mobile Internet services.

Keywords: cloud computing security; on-demand; service; mobile internet

了广泛的关注, 发展迅速。典型的安全企业都已经相继推出移动云安全服务, 相对传统杀毒技术, 移动云安全服务使得杀毒效果获得显著提升。例如奇虎的“云查杀”、瑞星的“云安全”等就是众多移动云安全服务中的明星产品。

然而, 随着移动云计算的发展, 在边界并不是很清晰的环境中, 本文认为移动云安全服务需构建完善的安全体系, 不只是关注移动终端的安全, 将云计算中的资源(包括安全资源)利用到用户及企业的安全防护的全流程中; 另一方面, 本文认为更重要的是, 移动云计算安全服务不仅要关注利用云对外提供服务, 更要考虑利用外部或者其他途径来获取得到信息情报, 从而更好的保证云计算平

台自身的安全, 提高安全效率, 确保用户在云中的资产的机密性、完整性、可用性等安全保障机制。

本文认为, 移动云计算安全服务包括两个方面功能。第一, 它将致力于保障云计算本身的安全, 即保证云中的业务、云中的数据、云中的用户是安全的, 保障安全资源的利用率和效果更高。通过将安全资源集群和池化、组件化, 对已有安全能力理分配, 集中管理安全资源, 根据安全情报弹性管理和使用这些安全资源, 达到高效、低成本的目的。第二, 它作为网络安全领域在云计算环境中的应用和拓展, 形成了一种新的技术和业务模式, 从而提升网络安全(包括访问控制、DDOS防护、病毒和恶意代码的检测和处理、网络流量的安全检

收稿日期: 2015-01-06
网络出版时间: 2015-02-22
基金项目: 国家科技重大专项
(2012ZX03002003)

测和过滤、邮件等应用的安全过滤、网络扫描、Web等特定应用的安全检测、网络异常流量检测等)的能力。总之,移动云安全服务在最小化成本下,为互联网业务、移动终端,通过提供便捷、按需可伸缩的安全防护,用户不需要自身对安全设施进行维护和管理。

后者的研究比较成熟,业界推出了许多研究成果,可供参考借鉴,如云安全联盟(CSA)SecaaS白皮书。前者的研究在云计算提供商中得到越来越多的重视,移动云安全服务在成本、高效、提高用户安全体验等具有明显的优势,成为各大云服务提供商研究和实践的热点。

中兴通讯自主研发的Cocloud云计算平台率先推出了移动云计算安全服务,充分保障了云计算平台自身的安全,提供按需供给云计算安全服务,产品已经商用^[1-15]。

1 系统架构

针对云计算的安全威胁,业界已经研究并推出许多针对性的安全技术,具备一定的安全防护效果。但是,当前的安全方案,对所有的云计算业务提供统一的安全机制,它们被一致对待,且一旦设定,就不能随意更改,这将导致云计算服务在动态扩展时,安全能力无法及时动态调整,无法满足不同时间、不同用户之间的动态安全调整需求。这不但不能实现对用户提供精确安全保障,而且造成安全资源的浪费。

移动云计算平台需要提供一种可以动态化的、差异化的云安全服务,可以实时调整、按需供给。中兴通讯的移动云安全服务方案就是针对上述需求专门设计的。图1为云安全服务示意图。

提供动态的、差异化的移动云安全服务,并不是提供新的安全能力,而是对现有的安全能力进行整合,根据用户预设不同的策略和安全模块上报实际数据,进行动态的调整和执

行。例如:用户在移动云安全服务系统订购对应的安全能力,并且设定了对应的执行策略。移动云安全服务系统就将该策略下发给安全模块执行具体的安全能力。在安全模块执行安全能力时感知到特定阈值异动时,将该异动数据及时上报给移动云安全服务系统,移动云安全服务系统根据预定的策略分析,确定是否需要下发新的安全策略。

系统在设计逻辑上分为四大部分:用户层、业务系统层(管理层)、能力中心层、云计算服务层。移动云安全服务系统逻辑层次如图2所示。

(1) 用户层

移动云安全服务系统的用户类型包括云计算服务的租户和系统管

理人员。

(2) 管理层

管理层主要提供系统的安全管理功能,供用户与安全能力中心的交互。为了充分利用现有系统能力,移动云安全服务系统可以和IROS集成,通过开放接口方式,将IROS中已有的基础数据,同步给移动云安全服务系统。业务系统层主要包括如下功能模块:

- 能力管理:该模块主要负责安全能力中心所提供能力的注册和管理功能。只有在本系统注册的安全能力,才能提供给租户进行订购。租户可取消已经订购的安全能力。

- 策略管理:策略就是指租户订购了哪些安全能力,以及该安全能力

图1 移动云安全服务

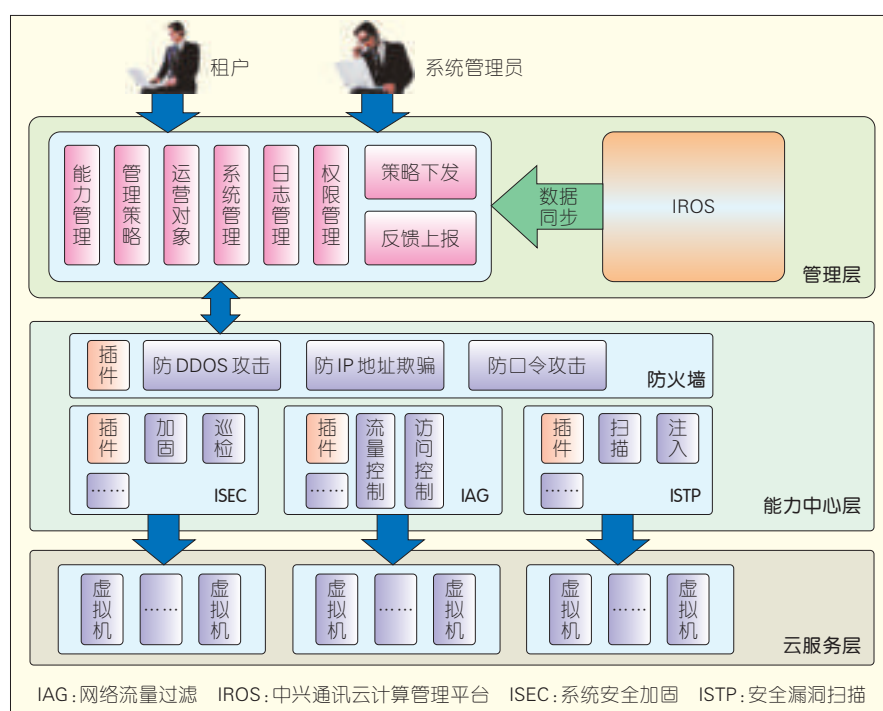
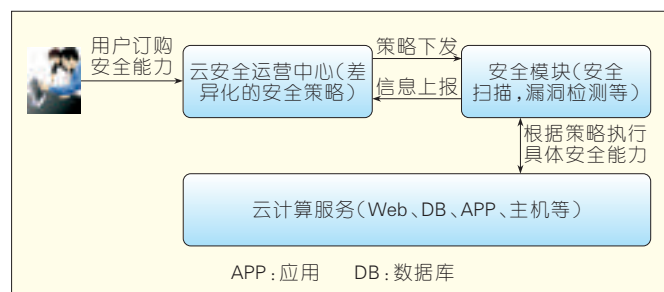


图2 移动云安全服务系统的逻辑层次

在何种情况下触发执行。例如:租户默认订购了中高版本的安全能力,但在默认情况下只需要执行低版本的安全能力。因此租户需要设定当现场环境状态改变后,安全运营中心能及时发现,并且动态的调整安全能力等级,自动调整至中高版本的安全能力防护。从而在平时可以节约安全能力资源降低成本,同时在出现异常情况时,又能获得高强度的安全防护能力。

- 运营对象管理:运营对象指安全能力所执行的对象的统称,运营对象的管理主要是对这些设备、业务管理和维护。运营对象包括如下:物理主机、虚拟主机、Web 服务器、DB 服务、APP、FTP 服务等。

- 策略下发:该模块主要负责将租户预设的策略下发给对应的安全能力执行模块。

- 反馈上报:该模块主要负责接收安全能力执行模块上报过来的数据信息,例如:策略的执行情况,所监控服务的实时状态灯。同时负责将上报的数据进行解析,通过解析的结果和预设的策略变动阈值进行匹配,以确定移动云安全服务系统是否达到新的策略触发条件,下发新的安全策略。

- 日志管理:该模块主要负责记录策略下发模块和反馈上报模块的所有操作。方便租户和管理对安全能力的监控。

- 插件:插件是一个独立小型代理程序,主要负责与反馈上报模块和策略下发模块之间的通信交互。插件需要根据不同安全模块提供对应插件,目前需要支持有系统安全加固(ISEC)插件、网络流量过滤(IAG)插件、安全漏洞扫描(ISTP)插件和防火墙插件。其中防火墙插件在该项目的一期暂不实现。

- 系统支持与 IROS 集成,充分使用 IROS 中已经管理的基础数据,例如:租户信息、设备信息(IP 地址、设备类型、操作系统等),也可通过

IROS 获取到移动互联网终端的安全情报信息,例如业务对象的状态等安全数据。

(3)能力中心层

能力中心层主要包括系统所支持的安全模块,以及各模块所支持的安全能力,包括 ISEC、IAG、ISTP 和防火墙。具体的安全能力举例如图 3 所示。

各安全能力组合对应的等级划分举例如图 4 所示。

具体安全能力的等级划分如图 5 所示,其中安全能力不同等级可以根据能力提供者的实际情况自行定义。

(4)云计算服务层

云计算服务主要是指云计算平台所能提供的所有云计算服务。云

计算的安全防护由安全模块进行负责执行。

2 关键技术

2.1 动态差异化的服务能力

(1)日志采集和安全情报提取

系统由插件获取安全能力日志和报告并上传到安全运营中心,以此海量的数据作为基础进行析解,可以提取供安全策略数学模型进行决策的一系列安全情报参数,例如 IAG 安全能力报告的访问量大小、IAG 安全能力报告的被攻击事件、防篡改安全能力报告的被篡改事件、ISTP 安全能力报告的漏洞多少。

(2)安全策略数学模型

图 3
安全能力举例

ISEC	ISTP	IAG	防火墙
安全加固	安全扫描	访问控制列表	防 DDOS 攻击
安全巡检	页面木马检测	流量控制	防 IP 地址欺骗
密码破解防御	端口检测	HTTPS/SSL	防口令攻击
异地登陆提醒	页面漏洞检测	防御 SQL 注入	IDS
页面防篡改		跨站脚本请求攻击	协议过滤
		录遍历请求攻击	反垃圾邮件

DDOS:分布式拒绝服务攻击
HTTPS:安全超文本传输协议
IAG:网络流量过滤
IDS:入侵检测系统

ISEC:系统安全加固
ISTP:安全漏洞扫描
SQL:结构化查询语言
SSL:安全套接层

图 4
安全能力组合举例

低	小低	小高	高
安全巡检	安全回固	密码破解防御	防 DDOS 攻击
安全扫描	页面木马检测	异地登陆提醒	防 IP 地址欺骗
	端口检测	流量控制	防口令攻击
	页面漏洞检测	防御 SQL 注入	IDS
			协议过滤

DDOS:分布式拒绝服务攻击
IDS:入侵检测系统

SQL:结构化查询语言

图 5
安全能力等级

能力名称	等级一	等级二	等级三
安全巡检	无	无	
安全扫描	1月	一周	一日
端口扫描	1月	一周	一日
木马扫描	1月	一周	一日
页面漏洞检测	1月	一周	一日
流量控制/TPS	1 000	1 000	3 000
密码防御	无	32 位	64 位

安全策略数学模型通过输入参数输出安全等级等策略以供下发到插件执行,每一个参数的变动都通过各自的权值影响到最终输出安全等级的高低,安全等级的高低又最终影响到各安全能力的具体执行控制,例如 IAG 安全能力报告访问量增加最终导致 IAG 增加相应带宽和处理能力, IAG 安全能力报告的被攻击事件最终导致防篡改扫描范围和扫描力度的加大等等。

按需供给移动云安全服务策略计算模型可以分为3个层次,即用户配置和系统感知层次、差异化安全策略生成和控制层次以及安全功能模块执行策略层次。按需供给移动云计算安全服务策略计算层次模型如图6所示。

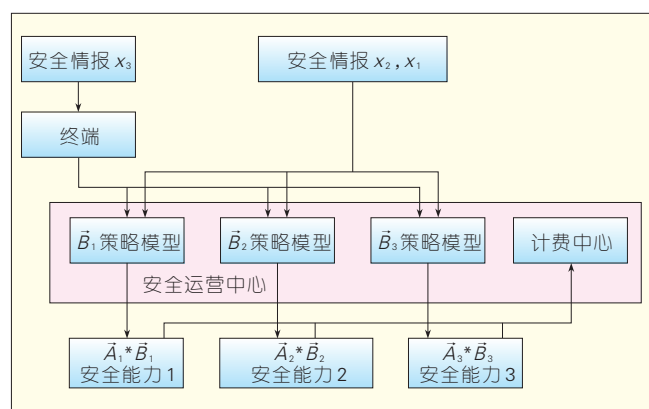


图6 按需供给移动云计算安全服务策略计算层次模型

如图6所示,假设向量 \vec{A} 是安全元集合, \vec{A}_1 是安全能力1安全元集合, \vec{A}_2 是安全能力2元集合, \vec{A}_3 是安全能力3安全元集合; \vec{A}_1 、 \vec{A}_2 和 \vec{A}_3 都是 \vec{A} 的子集,即 $\vec{A}_1 \subset \vec{A}$ 、 $\vec{A}_2 \subset \vec{A}$ 、 $\vec{A}_3 \subset \vec{A}$ 。 \vec{B}_1 、 \vec{B}_2 、 \vec{B}_3 是以上3个安全能力安全元集合的参数向量,该参数向量的关系式为: $\vec{B}_i(x_1, x_2, x_3)$,其中, x_1 代表安全情报类型一, x_2 代表全情报类型二, x_3 代表全情报类型三。

那么,用户需要的按需供给移动安全服务解决方案即为 $\vec{A}_1 * \vec{B}_1 \oplus \vec{A}_2 * \vec{B}_2 \oplus \vec{A}_3 * \vec{B}_3$,即3个安全能力安全解决方案的集合。其中, \oplus 是连接符号,表明按需供给移动安全服

务需要 $\vec{A}_1 * \vec{B}_1$ 、 $\vec{A}_2 * \vec{B}_2$ 和 $\vec{A}_3 * \vec{B}_3$ 的多方作用。

需要指出的是, \vec{A}_i 中的安全元,是系统设备在研发时已经在云计算平台上实现的各种安全元功能,如:加密功能、认证功能、完整性功能等。

$\vec{A}_1 * \vec{B}_1$ 、 $\vec{A}_2 * \vec{B}_2$ 和 $\vec{A}_3 * \vec{B}_3$ 是安全运营中心要实现的评估模型,由“数学模型+必要的策略”实现。在该模型下,安全管理员需要获取的参数是 x_1 、参数 x_2 、和参数 x_3 。这些参数只与业务、平台和业务场景相关,一旦业务、平台及场景确定下来,这3个参数也就确定下来了。

(3) 插件的动态扩展

插件与安全运营中心之间采用统一的通信方式和消息结构以支持动态扩展。插件与安全运营中心之

置,日志分析结果上传可以采用标准的系统日志(Syslog)方式,插件获取的策略可以采用由传输控制协议/网间协议(TCP/IP)传送具体的安全等级值,再由插件内部解析成相应具体的安全策略以调用执行。

2.2 共享数据与安全运营中心独立数据并存

安全运营中心数据库中基础信息同时存在两种数据:(1)租户通过Web模块添加的数据,包括租户、设备和系统信息;(2)安全运营中心的数据库同步模块从IROS同步的设备和租户信息,并由租户自己补充设备的系统信息,两者采用字段标识进行区分,两者的运行效果是相同的,只有数据库同步模块才根据字段标识取出相应的IROS同步设备和租户信息进行新增修改删除操作,并同时租户自己补充的相应设备的系统信息维持原状或进行删除操作。除租户、设备为代表的基础信息外,安全运营中心还拥有自己独特数据以供业务运行,如安全能力信息、安全策略信息、运营对象信息,这些信息是IROS中所没有的,需要由租户通过Web模块进行维护,如日志信息是由插件从各安全能力采集来的。数据库数据来源如图7所示。

2.3 扩展性的设计考虑

移动云安全服务系统的扩展性

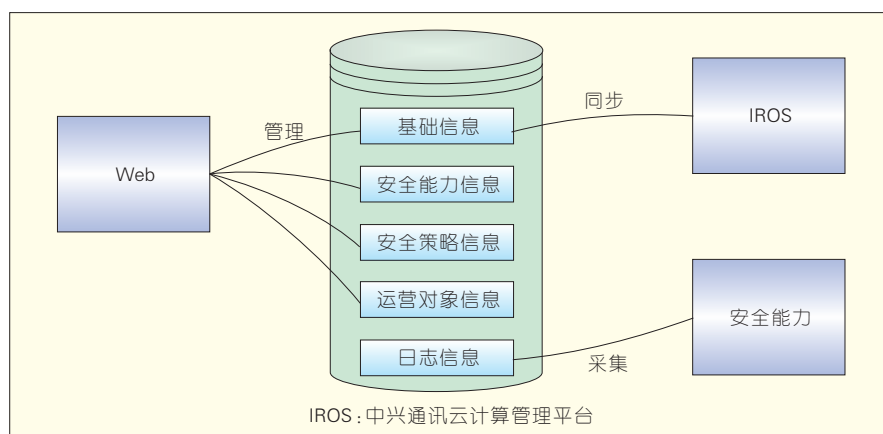


图7 数据库数据来源

逻辑如图8所示。

移动云安全服务系统对安全服务能力的支持基于插件的动态部署和系统管理员在Web管理平台进行的能力管理配置,可以通过开发新插件的方式为系统增加新的安全服务能力。

3 实践

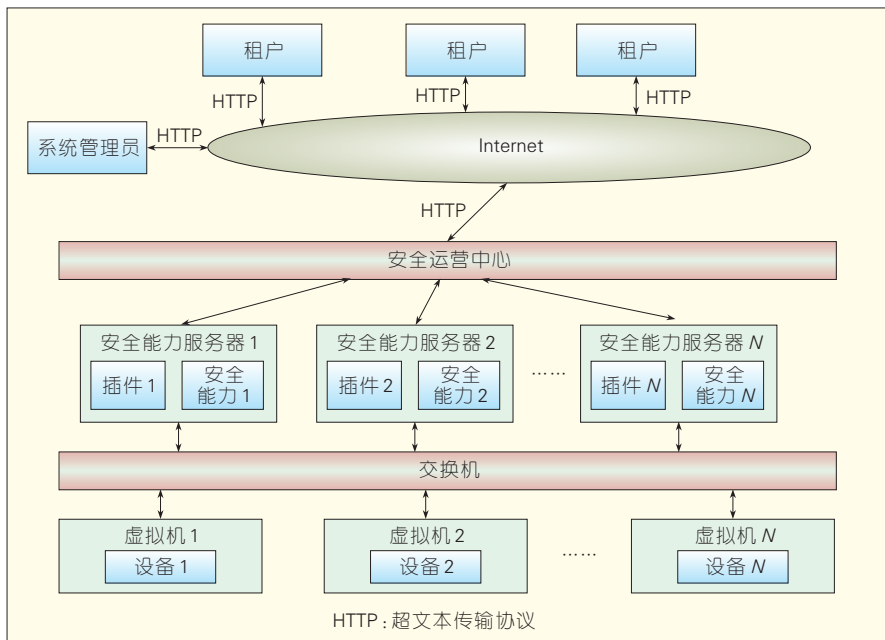
移动云安全服务系统结构如图9所示。图中,各相关系统部署在以太网上,移动云安全服务系统通过安全模块对云计算服务进行安全防护。云计算服务都部署在对应虚拟服务器上,这些服务器有:文件服务器、邮件服务器、电子商务服务器、代理服务器、数据库服务器、及时通讯服务器、内容管理服务器等等。

3.1 漏洞扫描

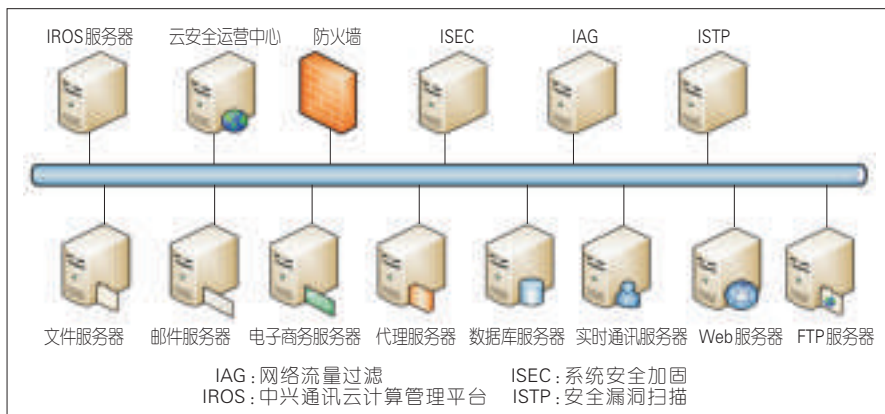
假设ISTP在默认情况下执行低等级的安全防护,当IAG、ISEC、防火墙等捕捉到云计算服务受到安全攻击,并记录下来。安全插件通过轮询IAG、ISEC、防火墙等的记录信息,感知到云计算受到安全攻击时,安全插件将该信息反馈给安全运营中心,反馈上报模块分析对应的反馈信息,并和该运营对象所预设的策略进行匹配。确定是否需要触发新的策略等级,并将新触发的等级策略下发给安全插件,由安全插件触发ISTP执行新的安全能力。对应安全能力流程如图10所示。

3.2 流量过滤

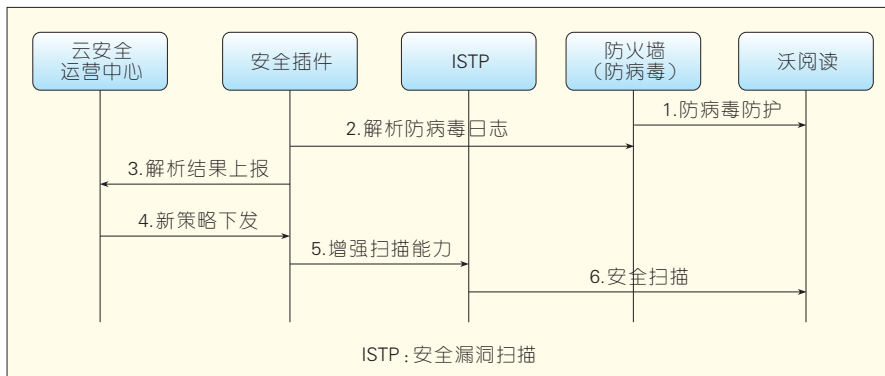
假设IAG在默认情况下执行的流量控制位1 000 TPS,当安全插件通过分析IAG记录的云计算服务的访问量数据,并将该数据上报给移动云安全服务系统,移动云安全服务系统将该数据和预设的流控最大阈值进行匹配,发现该访问量在租户的允许范围内,但需要网络流量过滤增加对该服务的流控阈值。因此,移动云安全服务系统将新的流控阈值通过测量



▲ 图8 移动云安全服务系统的扩展性逻辑



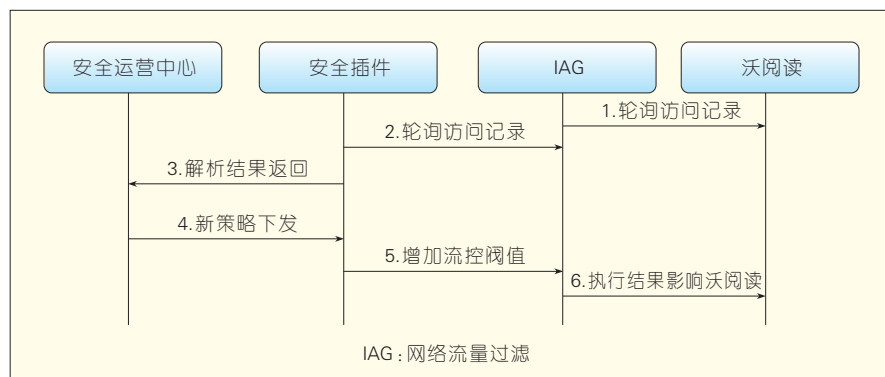
▲ 图9 系统结构



▲ 图10 ISTP动态改变安全能力流程

下发模块,下发给安全插件,安全插件调用网络流量过滤对应接口,从而

网络流量过滤对该云计算服务的流控阈值进行修改。对应流程如图11



▲ 图 11 IAG 流量控制流程

所示。

4 总结

本文介绍了按需供给中兴通讯移动云安全服务系统架构及关键技术,并在中兴通讯 Cocloud 云平台上进行了商用,用户反映效果良好。后期还将针对安全能力进行系统扩展,增强安全情报大数据采集和分析能力,充分发挥云平台、移动互联网的优势。移动云安全服务可以有效地保证云中的业务、云中的数据、云中的用户安全,安全资源的使用率得到提高,可以提供动态实时性安全防护保障。

参考文献

- [1] 冯登国. 开启云计算安全时代 [R]. 司局长论坛, 2011
- [2] 黎春武, 巫建文. 基于等级保护的信息系统综合监控及评估系统设计 [J]. 信息网络安全, 2012, 12(05): 92-95
- [3] 李国琴. TNC 在等级保护中的应用研究 [J]. 信息网络安全, 2012, 12(04): 89-93
- [4] CSA. Cloud Controls Matrix Version 3.0 [EB/

- OL]. [2015-01-01]. https://downloads.cloudsecurityalliance.org/initiatives/ccm/CSA_CCM_v3.0.xlsx
- [5] LI J, KROHN M, MAZIERES D, et al. Secure Untrusted Data Repository (SUNDR) [C]// Proceedings of the 6th conference on Symposium on Operation Systems Design and Implementation, Berkley: USENIX Association, 2004: 9
- [6] GOH E, SHACHAM H, MODADUGU N, et al. SiRiUS: Securing Remote Untrusted Storage [C]// Proceedings of the 10th Network and Distributed Systems Security (NDSS) Symposium, San diego, 2003: 151-159
- [7] WIEBELITZ J, BRENNER M, KUNZ C. Early Defense – Enabling Attribute-Based Authorization in Grid Firewalls [C]// Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing Pages. 2010: 336-339
- [8] MODI C, PATEL D, BORISANYA B, PATEL A. : A Novel Framework for Intrusion Detection in Cloud [C]// Proceedings of the Fifth International Conference on Security of Information and Networks Pages, 2012: 67-74
- [9] ADAM B, BENJAMIN M, MASOUD V, KEVIN R. Towards secure provenance-based access control in cloud environments [C]// Proceedings of the CODASPY, 2013: 277-284
- [10] 王莹. 网络隔离技术 [J]. 微计算机应用, 2003, 24 (3): 133-136
- [11] NIST SP800-125, Guide to security for Full virtualization Technologies [EB/OL]. [2015-

- 01-01]. <http://csrc.nist.gov/publications/nistpubs/800-125/SP800-125-final.pdf>
- [12] NIST whitepaper, Challenging Security Requirements for the US Government Cloud Computing Adoption [EB/OL]. [2015-01-01]. http://www.nist.gov/customcf/get_pdf.cfm?pub_id=912695
- [13] NIST sp500-292 NIST Cloud Computing Reference Architecture [EB/OL]. [2015-01-01]. http://www.nist.gov/customcf/get_pdf.cfm?pub_id=909505
- [14] CSA. Top Threats to Cloud Computing [EB/OL]. [2015-01-01]. <http://www.cloudsecurityalliance.org/topthreats>
- [15] CSA. Security Guidance for Critical Areas of Focus in Cloud Computing v2.1 [EB/OL]. [2015-01-01]. <https://cloudsecurityalliance.org/guidance/csaguide-cn.v2.1.pdf>

作者简介



陈小华, 中兴通讯股份有限公司高级技术预研系统工程师; 主要从事网络安全、密码学、云计算及大数据安全技术标准研究; 已参与 5 项基金项目。



董振江, 中兴通讯股份有限公司战略与技术专家委员会业务专家组组长、云计算及 IT 研究院副院长; 主要研究方向为云计算与大数据、新媒体、移动互联网等技术; 已主持基金项目 10 余项, 发表学术论文 10 余篇, 出版专著 1 部。



金怡爱, 中兴通讯股份有限公司高级技术预研系统工程师; 主要从事移动互联网、云计算以及安全等技术的研究。

综合信息

中国电信将构建第三张全国骨干网

中国电信云计算公司副总经理徐守峰在“2015 中国通信行业云计算峰会”上表示, 中国电信 2015 年将会在 ChinaNet、CN2 两张全国骨干网的基础之上, 打造第三张全国性骨干网络, 主要用于数据中心之间的节点互联, 助力电信打造面向未来的云服务提供商。

徐守峰说, 现在已经处于 IT 和数据大爆炸的时代, 人们的工作还是生活都已经离不开 IT。而这个趋势所

面临的最核心问题, 就是业务的不确定性, 要有海量的存储、密集计算、服务不间断、信息安全、全网部署等很多要求, 怎样去提升或者构建 IT 基础架构, 将是 IT 从业者一个非常好的选择的方向。

按照规划, 中国电信首先将会在一些核心数据节点构建 VDC, 然后通过 VDC 将数据中心和第三张网扁平化的连接起来, 打造一个二层网络。

(转载自《中国信息产业网》)

移动云存储服务关键技术研究

Key Techniques in Mobile Cloud Storage Services

崔勇/CUI Yong
赖泽祺/LAI Zeqi
缪葱葱/MIAO Congcong
(清华大学, 北京 100084)
(Tsinghua University, Beijing 100084, China)

近年来,个人云存储服务^[1-2]占有的市场份额越来越大,已成为用户个人信息存储领域不可分割的一部分。伴随着移动互联网^[3-7]的发展,轻巧便捷的移动设备受到用户广泛青睐,移动终端的人均持有量快速上升,随之而来的是移动终端应用的迅猛增长,这促使移动终端逐渐成为新的应用平台。用户对终端的存储空间以及终端资源的在线共享等要求越来越高,使得移动云存储服务成为移动端信息存储领域的研究热点。与传统的移动硬盘、U盘等存储设备相比,移动云存储服务提供云端超大容量的数据中心,并能够统一高效管理用户多个终端上的零散数据。用户不仅可以通过移动互联网跨平台、跨终端将个人数据同步至云端,也可以使用任意终端随时随地存储、同步、获取并分享云端数据。移动网络下用户比较注重数据同步流量和同步效率,服务商也提供了多种优化技术提升移动端服务性能,但目前全球学者对移动云存储服务的技

收稿日期: 2015-01-05

网络出版时间: 2015-02-22

基金项目: 国家自然科学基金重大国际合作项目(61422206、61120106008)

中图分类号: TN929.5; TP393 文献标志码: A 文章编号: 1009-6868 (2015) 02-0010-004

摘要: 着眼于移动应用数据共享,探讨了整合移动应用产品线的需求、实现方法和现状,研究了移动云存储服务主流架构、文件同步协议和优化技术这些关键技术。从优化同步效率、节约同步流量着眼,设计了测试床,评估了目前主流移动云服务文件分块、捆绑、冗余消除、增量同步和数据压缩5项优化技术的实现情况,测量了文件传输时各服务的协议开销和同步完成时间,用于评估各服务的移动客户端性能。实验结果显示,移动云存储服务的优化技术对移动客户端服务性能有决定性的作用。

关键词: 移动云存储服务; 整合移动应用产品线; 关键技术; 移动客户端性能

Abstract: The motivation of the research in this paper is the demand of data sharing. We discuss the demands, implementation and current situation of integrating product line, leading to the study of key technologies including mainstream architecture, file synchronization protocol and capabilities for optimization. To make synchronization more efficient and economize synchronization traffic, we design a test bed that evaluates the implementation of five capabilities—chunking, bundling, de-duplication, delta encoding and compression—and measures the performance of mobile clients in terms of protocol overhead and synchronization completion time. The results show the capabilities significantly affect mobile client service performance.

Keywords: mobile cloud storage service; integrating product line; key technologies; mobile client performance

术研究并不深入。

本文着眼于移动应用数据共享,探讨了整合产品线的需求,实现方法和发展现状,研究了移动云存储服务主流架构、文件同步协议和优化技术这些关键技术。实验了设计测试床,测试目前主流云服务的优化技术实现情况和移动客户端性能。结果显示,移动云存储服务的优化技术对移动客户端性能产生重大影响。目前主流的移动云存储服务各有优缺点,也还有提升的空间。

1 移动应用产品线

移动应用产品线由一类功能近

同,顾客群与渠道类同的移动应用程序组成。整合移动应用产品线通过云端将此类应用连接起来,所有应用数据在云端共享。

1.1 整合产品线需求

随着移动互联网已经渗透到人们社会和生活各个领域的各个角落,基于移动终端的应用程序也越来越多,应用程序产生的数据量正呈现出爆炸式增长。但是各应用间的数据均分布在各个孤立的服务器端,数据间没有互相连通,每个应用均是一个“数据孤岛”^[8-9]。而随着移动应用程序快速增长,现有的应用数据存储方式难

以满足用户的需求。如用户需要在多应用上发布相同信息时,他们需要在每个应用上重复操作,浪费宝贵的时间。移动云存储服务提供云端超大容量的数据中心整合移动应用产品线,用户在使用移动应用时产生的数据均在云端共享。整合产品线不仅方便了用户对数据的管理,而且还提升了服务商的竞争力。当各服务商提供的移动云存储服务功能相似时,用户更加注重服务的体验质量(QoE)^[10-12]。服务商提供的移动云服务功能越丰富,产品线中的产品种类越多,其统一管理的数据量越庞大,越能满足人们日常办公和生活的需求,产生更好的用户体验,容易在竞争中脱颖而出。

1.2 整合产品线方法

百度、谷歌已成为用户搜索入口,安智市场、App Store已成为用户下载新应用入口,与它们相似,移动云存储服务正逐渐成为共享移动用户数据的入口。图1显示了社交应用数据、文件应用数据、游戏应用数据、其他移动应用等数据通过统一的数据入口接入云端,云端共享这些应用的所有数据。移动云存储服务作为移动应用程序数据入口平台,提供应用集成服务。应用集成服务的核心是连接“数据孤岛”,共享应用数据。整合产品线不仅方便用户统一管理多应用程序的数据,而且提供多应用数据云端共享,挖掘数据的潜在价值。用户在移动端使用应用程序

产生的数据均共享至云端,云端服务器充分利用应用数据,实时分析并决策,智能地实现不同应用间数据的流通。为了更好的吸引移动应用接入云端,移动云存储服务商通过开放云存储应用编程接口(API),向开发者提供了面向云端应用数据集成解决方案,开放应用程序入口,共享应用程序数据。

1.3 整合产品线现状

整合产品线主要的方法是通过云端连接多个移动应用程序,共享应用数据。目前市场上,整合产品线主要分为两种方式,即整合单一服务商的应用数据和整合多个服务商的应用数据。

整合单一服务商应用数据一般适用于应用产品丰富的服务商。这类服务商有较大的市场规模和一定的用户基础,旗下应用产品比较丰富,涵盖全面。它们更倾向于提供自己的云存储服务,聚合用户使用旗下产品产生的数据。如百度提供百度云盘聚合百度地图、百度搜索、百度贴吧、百度百科等应用的数据,腾讯提供腾讯微云聚合QQ、微信、腾讯微博等应用的数据,Google提供Google Drive聚合谷歌翻译、谷歌照片、谷歌搜索等应用的数据。这类服务商在整合产品线时将所有应用的API以统一的格式,整合在一个平台中。用户在使用服务商旗下应用时产生的数据会自动同步至云端,通过云服务推送至其他应用。用户使用一个账

号便可畅游该类服务商旗下的所有应用。如腾讯用户在QQ空间发送的状态不仅会同步至云端,也会推送至腾讯微博等其他应用。

上述移动云存储服务只是整合单一服务商旗下应用数据,但无法统一管理用户多个服务商应用的数据。数据在各服务商之间并未真正实现连通,各服务商提供的云端是一个更大的“数据孤岛”。为解决各服务商之间数据不连通的问题,业界产生了一批集成平台服务公司,包括IFTTT、CloudWork、MuleSoft CloudHub、SnapLogic等。该类公司提供集成平台服务,主要面向不同类型、不同服务商提供的应用程序,它们通过API提供了多个服务商数据连通平台,各服务商之间的应用都可以通过该平台共享数据,真正实现所有应用数据共享。集成平台服务主要有以下几个特点:

(1) 预先集成/连接在线服务供应商。用户通过该服务可以直接访问其他服务商。

(2) 任务执行的自动化。用户通过该服务对应用程序的操作会自动同步至其他应用。

(3) 服务供应商选择多样化。用户可以针对性的选择服务供应商,满足用户需求。

(4) 无需软件开发。用户只需学会使用该服务,无需任何开发代价。

MuleSoft推出的CloudHub是多云融合的代表,它提供了全球性、多用户、可伸缩的集成云,目标要像Facebook连接全球人群一样,做到所有应用程序数据的融合。CloudHub提供应用集成和连接,共享所有应用数据,能让应用进行互动、分享和共同工作的平台,满足用户各种各样的需求。

2 移动云存储服务架构和同步协议

移动云存储服务连通了各应用程序,使促进应用数据的交流。本章



图1 移动云存储聚合平台

通过对多个服务商提供的移动云存储服务进行研究,总结出同步协议,如图2所示。其架构主要由云端存储服务器、控制服务器和移动客户端3部分组成。

2.1 移动云存储架构

移动云存储架构包括:存储服务器、控制服务器、移动客户端等。

(1) 存储服务器

存储服务器主要存储文件内容。在服务器中,所有的文件并不是单独存在,而是拆分成多个数据块存储。存储服务器只存储文件内容,不包含文件名称、大小、版本号等信息。

(2) 控制服务器

控制服务器主要负责与移动端交换控制信息,主要分为通知流、元数据流两个部分:

(a) 通知流。通知流是通知信息的载体,用于保持移动端和云端文件的一致性。移动端向控制服务器发送请求,监测其他终端是否修改同步文件。若云端同步文件发生改变,云端返回“已更新”通知信息,移动端同

步云端文件最新版本。若云端文件在一定时间内未发生改变,云端返回“未更新”通知信息。当移动端同步完毕或者接收到“未更新”通知信息,移动端会再次向控制服务器发送请求,进行新一轮检测。

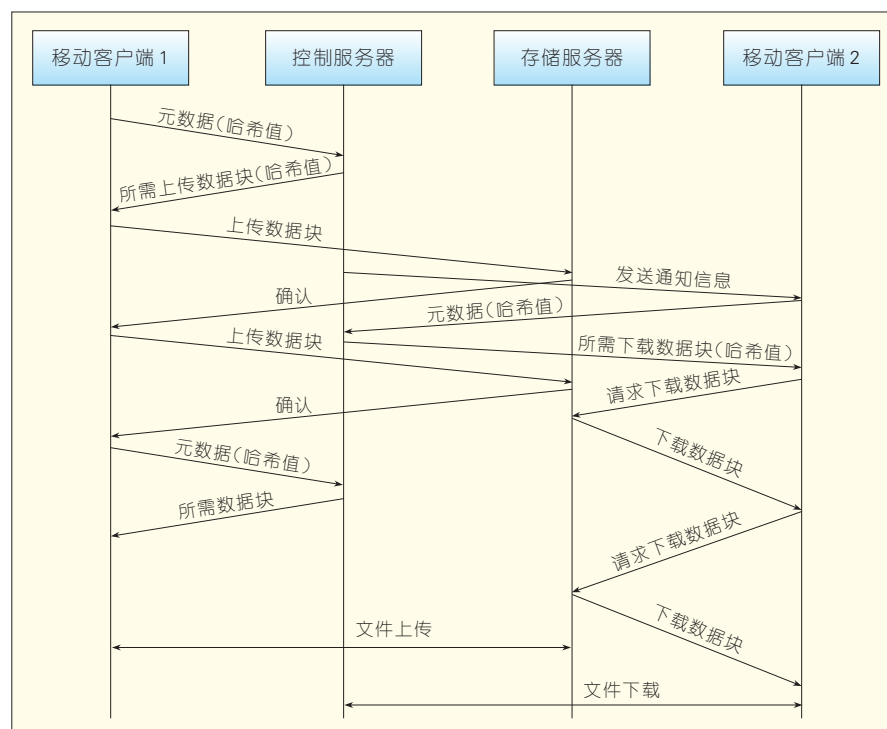
(b) 元数据流。元数据流通常传输文件元数据信息。元数据信息包含哈希列表、服务器文件日志等信息。哈希列表存放数据块的哈希值,哈希值是数据块的唯一标识符。服务器文件日志存放文件的ID、数据块列表等信息。

(3) 移动客户端

移动客户端是由分块器、索引器、通知器和缓存器4个部分组成。分块器将本地的文件切分成固定大小的数据块。索引器存放文件的元数据信息。通知器通常向云端发送请求,检测云端文件是否更新。缓存器缓存云端文件,所有对缓存文件的操作会自动同步至云端。

2.2 同步协议

当多个移动终端共享一个云盘



▲ 图2 移动云存储服务同步协议

账户时,在任一终端上传的文件均会同步至其他终端。从图2可以看出用户同步文件时的协议流程,分为文件上传、文件下载两部分。

(1) 文件上传。移动端1上传文件时,元数据流上传文件元数据信息至云端控制服务器。控制服务器将其与云端元数据对比,并返回所需上传数据块的哈希表。移动端1上传所需数据块至云端,每个数据块上传成功时都会返回一个确认信息。当所有数据块上传成功时,移动端1再次发送元数据信息进行核对,确保所有文件上传成功。

(2) 文件下载。移动端2下载云端更新文件保持数据一致性。当云端文件元数据信息发生改变时向移动端2返回通知流,移动端2向云端发送元数据信息。云端返回所需下载数据块,移动端2发送数据块下载请求,云端向移动端2发送数据块。移动端2每下载完一个数据块均会再次发送下载请求,直到所有数据块下载完成。

3 实验评估

相对于网络稳定、带宽较大的传统PC终端,用户在带宽有限、流量有限和信号强度不稳定的移动环境下更加注重文件的同步效率和同步流量^[13-15]。实验对目前流行的百度云盘、360云盘、腾讯微云和Dropbox 4个云服务进行评估,评估目标分为两部分:优化技术实现情况;通过同步协议开销和同步完成时间评估云服务移动端性能。

3.1 测试床设计

实验所用测试机是谷歌三星Galaxy Nexus手机,版本号是Android 4.3.1。在数据传输过程中,实验跟踪上传或者下载数据包的IP地址、端口号等信息,再通过标识符(UID)分别滤出上述厂商提供的云服务所传输的数据包。

文章设计了两个评估实验,对于

第一个目标,实验针对每个优化技术分别设计了测试方法测试其实现情况。对于第二个目标,实验设计多种文件同步方式,包括数量、大小、类型,使其覆盖各式各样的同步场景。为防止文件同步时被压缩,实验使用随机的二进制数据块填充文件。同步场景包括:1个100 kB的文件;1个1 MB的文件;1个10 MB的文件;10个100 kB的文件。

3.2 优化技术实现情况

文件分块、捆绑、冗余消除、增量同步和数据压缩5项优化技术大大提高了移动端文件传输性能。文件分块是指在传输大文件时,文件被切成多个数据块上传,避免了网络中断时重复上传整个文件,降低流量开销。文件捆绑是指批量上传小文件时,文件被捆绑成一个数据块上传,云端返回一个确认完成传输,减少了云端反复确认时延和控制开销。冗余消除针对于云端已经存在的文件,客户端无需重复上传相同文件,不仅节约了移动端上传流量,还节约了运营商使用带宽。对于云端反复修改的文件,增量同步只上传文件修改部分,大大减少了移动端上传流量。数据压缩技术以压缩时间为代价,消除文件冗余信息,减少网络流量和存储空间,同时也方便文件管理。

本文通过设计实验对目前流行的云服务进行测试。结果显示,百度

云盘和Dropbox在上传文件时将大文件切成4 MB数据块传输,360云盘将大文件切分成0.5 MB传输,腾讯微云不分块。所有的云服务均实现了冗余消除技术,而均未实现捆绑、增量同步和数据压缩技术。对于未实现的优化技术,服务商可能正在开发之中。

3.3 移动客户端性能评估

针对移动云存储客户端性能,实验在测试床中测量了移动客户端同步文件时的协议开销和同步完成时间。协议开销是用户同步文件时的数据量与文件数据量的差值,同步完成时间是移动客户端上传第一个数据包至移动端接收到最后一个数据包的确认所经历的时间。移动客户端性能评估结果如图3所示。

在同步文件时,不同服务的移动客户端性能存在很大的差异。对于上传任意大小的单个文件,腾讯微云均只开启一条传输控制协议(TCP)连接传输文件,且其产生的数据量最少,均为0.66 kB。除了Dropbox外,其他服务均开启10条TCP连接传输10×100 kB文件,产生的协议开销均是上传100 kB产生的协议开销的10倍,而且同步完成所用的时间均比上传1 MB大,这是因为云端的反复确认造成的传输时延和控制开销。在上传10 MB文件时360云盘开启20条TCP连接传输文件,产生大量的协

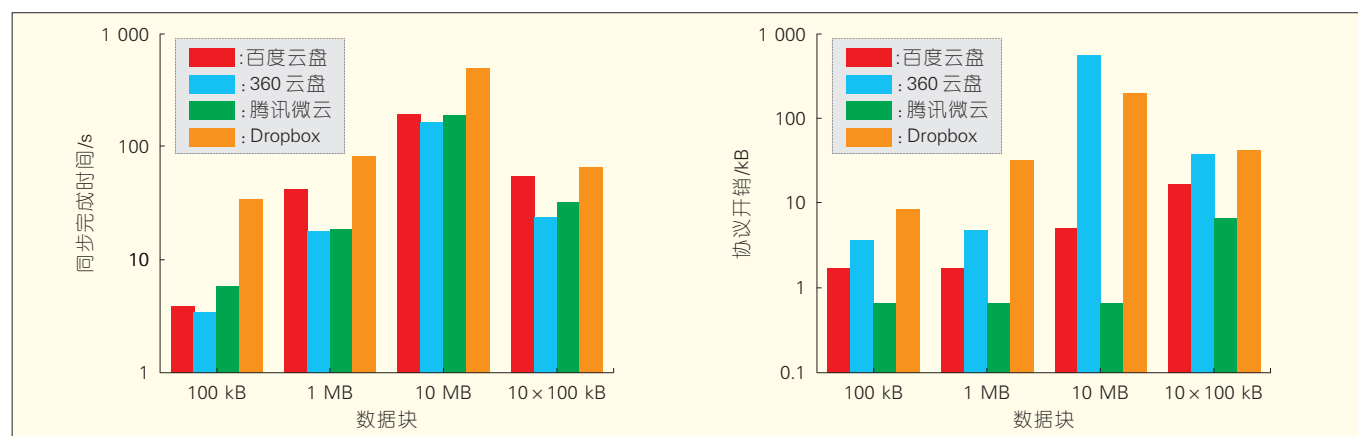
议开销,是同等条件下Dropbox的2.8倍,腾讯微云的846倍。而在同步其他文件时,Dropbox的协议开销和同步完成时间都是最大的,实验根据Trace中的IP地址得到其云端服务器地址在美国东部(Ashburn, Virginia),其他服务器地址均在中国。Dropbox在上传文件建立传输连接时,往返时延(RTT)比较大。中国云存储服务在传输单个文件时,360云盘的协议开销均是最大的,而同步完成时间均是最小的。在传输时,360云盘将文件拆分成多个0.5 MB块,每个数据块均开启一条TCP连接抢占带宽资源,减少总的文件传输时间。

对于中国的3种云存储服务,它们有各自的优点。360云盘在同步文件过程中并行建立多条TCP流来抢占带宽资源,以增加协议开销为代价减少了同步完成时间。而百度云盘和腾讯微云只建立一条TCP传输,减少了协议开销。

4 结束语

本文分析了移动云存储服务对移动应用生态系统产生了重大影响,探讨了服务商整合产品线的需求。本文更深入的研究了移动云存储服务架构、文件同步协议和优化性能等关键技术,并对目前流行云存储服务进行了评估。目前流行的移动云存储服务各有优缺点,也还有提升的空

➡下转第19页



▲图3 移动客户端性能评估

面向 5G 的分布式移动云计算协同架构与管理机制

Cooperative Management Architecture and Mechanism of 5G-Oriented Distributed Mobile Cloud Computing

李宏佳/LI Hongjia

陈鑫/CHEN Xin

周旭/ZHOU Xu

(中国科学院, 北京 100093)
(Chinese Academy of Sciences, Beijing
100093, China)

随着移动互联网与智能终端技术的发展, 智能终端及运行于其上的虚拟现实、健康监测、游戏、社交、多媒体内容识别与检索等多元应用的使用成为现代人日常生活的一部分。尽管智能终端不断的推陈出新, 但是, 现在以及未来的智能终端仅凭借其自身处理能力很难完全满足虚拟现实等未来应用和最佳用户体验的需求。具体而言, 作为智能终端能量源的电池技术的发展速度(电池容量每 7 年提升 1 倍)远远落后于体现智能终端处理能力的处理器芯片的发展速度^[1]。另外, 由于人类本身对于信息获取、感知与认知的能力越来越高, 例如, 低灯光条件下人类对于环境与危险识别的时间小于 700 ms^[2], 对于人脸的识别时间在 370 ~ 620 ms 之间, 对于人类声音的识别时间小于 4 ms^[3], 因此现有智能终端的计算能力仍然无法满足

收稿日期: 2015-01-15

网络出版时间: 2015-02-22

基金项目: 国家自然科学基金(61302108)

中图分类号: TN929.5; TP393 文献标志码: A 文章编号: 1009-6868 (2015) 02-0014-006

摘要: 提出了一种分布式移动云计算协同架构以及端到端终端计算卸载服务质量保证机制。该架构将云计算资源嵌入移动通信网络基站等网元, 形成近端低时延云计算能力; 引入分布式移动云计算协同控制器, 以及基于其的终端、移动网络与分布式云服务器协同的感知与计算卸载决策机制。该架构可以降低终端在计算卸载中的网络交互信令开销, 以及终端驻留决策进程的 CPU 占用率与能耗; 可以保证移动云计算的端到端服务质量; 可以规避传统移动云计算给广域网(WAN)带来的巨大负载量与能耗。

关键词: 移动云计算; 服务质量; 智能终端; 协同; 时延; 能耗

Abstract: In this paper, we propose a new distributed mobile computing over 5G networks and a cooperative management framework for guaranteeing end-to-end quality of service of terminal computation offloading. Specifically, we design cloudlets that have only a small fraction of the computational ability of the cloud computing center. These cloudlets are associated with eNodeBs to reduce transit delay through the WAN. A cooperative controller is introduced to determine the computation offloading policy according to the key factors measured from smart terminals, the RAN and cloudlets. Compared with the classical mobile cloud computing, the proposed structure and cooperative management framework reduces overhead, CPU occupancy rate, and energy consumption of smart terminals before and during the computation offloading. It also guarantees end-to-end QoS of distributed mobile cloud computing and avoids huge overhead and energy consumption.

Keywords: mobile cloud computing; QoS; smart terminal; cooperation; delay; energy consumption

高用户体验的要求。再者, 随着 5G 网络时代的到来, 将出现超过 500 亿个链接的万物互联服务, 例如, 谷歌眼镜等可穿戴计算、车联网服务、在线医疗服务、军事指挥等, 智能终端的形式将呈现多元化, 服务计算复杂越来越高, 对于服务质量的要求越来越高。

移动云计算, 即移动智能终端将

计算量与能耗密集的任务通过移动互联网卸载到云计算中心, 通过云计算满足移动智能终端应用计算服务的需求, 成为云计算服务的重要应用之一。其中, 如何确定需要卸载移动云计算的任务是研究领域主要重点关注的问题之一。依据智能终端与云服务器间的数据交互协议, 已有的计算卸载可以分为 3 类:

(1) 用户-服务器机制

智能终端应用通过预先安装在智能终端与云服务器上的服务程序交互数据并决策计算卸载任务,已有研究成果如 Spectra^[4]、Chroma^[5] 和 Cuckoo^[6]等。

(2) 虚拟化机制

采用虚拟机承载分割的终端应用,有终端或终端与移动云服务器共同决定如何分割应用任务,并以虚拟机迁移的方法将已分割的终端应用任务迁移到云计算服务器执行,已有研究成果如 MAUI^[7]、CloneCloud^[8] 和 MobiCloud^[9]。

(3) 移动代理机制

以第三方移动代理与智能终端协商如何分割与卸载终端应用任务,已有研究成果如 Scavenger^[10]。

依据终端应用的分割方式,已有的计算卸载可以分为3类:

(1) 静态方式

在终端应用开发过程中预先设定卸载计算任务的分割。

(2) 半动态方式

在终端应用执行时,根据终端、网络与云计算服务器状态,协商决定计算卸载任务。

(3) 动态方式

在终端应用执行过程中,实时根据感知终端、网络与云计算服务器状态,动态协商决定计算卸载任务,这种方式目前真正意义的相关研究尚属起步阶段。

尽管移动云计算借力云计算能力大大增强了其计算能力,以及终端能耗开销,但是,由于移动智能终端与云计算中心的端到端网络传输时延与带宽具有不稳定,文献[11]对于260个Amazon EC2实测得到最优接入点的循环时间(RTT)为74 ms,而广域网(WAN)中往返传输时间(RTT)呈现长尾分布,即,大量网络传输的RTT分布于几百到几千毫秒之间,因此,移动云计算的通信通道传输时延无法保证。其次,由于影响移动云计算服务质量的主要因素在

应用执行过程中是在不断动态变化,如,智能终端中的电池荷电量(SOC),CPU与内存占用率,以及无线信道状态、用户的位置与移动速度等因素;移动网络中的负载与无线资源占用率等因素;互联网中的负载与带宽等因素;云计算中的虚拟机占用率等因素,因此,有效应用任务的分割与卸载决策需要终端、网络与云计算多方协同。再者,目前计算卸载决策主要由终端主动感知与协商完成,频繁的终端感知与协商网络大大增加额外的终端能耗与网络开销。

鉴于上述问题,本文结合4G、5G网络架构特点,提出了一种分布式移动云计算协同架构,以及端到端终端计算卸载服务质量保证机制。具体而言,将云计算资源嵌入移动通信网络基站等网元,形成近端低时延云计算能力;引入分布式移动云计算协同控制器(下文简称为协同控制器),以及基于其的终端、移动网络与分布式云服务器协同的感知与计算卸载决策机制。文章所提出的架构与机制首先可以降低终端在计算卸载中的网络交互信令开销,以及终端驻留决策进程的CPU占用率与能耗;可以保证移动云计算的端到端服务质量;可以规避传统移动云计算给WAN带来的巨大负载量与能耗。

1 面向5G的分布式移动云计算网络协同架构

图1结合4G、5G网络无线接入网架构特点给出了分布式移动云计算协同架构示意图,主要包括4个功能主体。

(1) 智能终端

作为移动云计算的发起端,由于其设计更加倾向于更轻的质量,更小的尺寸,更加美观与更好的散热性,因此,其计算能力与电池容量有限,例如,与服务器相比,目前它们之间处理器计算速度最大相差27倍。在应用执行过程中,其周期地通过移动通信网络上传其自身状态感知信息,

并接收协同控制器计算相关卸载分割决策。

(2) 移动网络

为发起计算卸载请求的智能终端提供无线接入与传输,由于其负载在时间具有波动性与空间具有差异性^[12],因此,具体提供服务的基站周期向协同控制器发送负载与频谱资源占用率等感知信息,并且接收协同控制器决策信息预留无线资源。

(3) 分布式移动云计算服务器

部署于移动接入网络侧的小型服务器或多小型服务器集群,由于其只需以虚拟机形式承载其所连接单个基站或多个基站服务终端的计算卸载任务,因此,部署成本低,通过近端服务有效克服了WAN RTT的长尾效应。但是,其自身的负载状态与虚拟机计算能力将直接影响终端计算卸载任务的处理时间,从而影响移动云计算的服务时延。因此,其负载状态与虚拟机计算能力感知信息将周期上传到协同服务器,并接收协同控制器决策信息为计算卸载任务预留虚拟机资源。

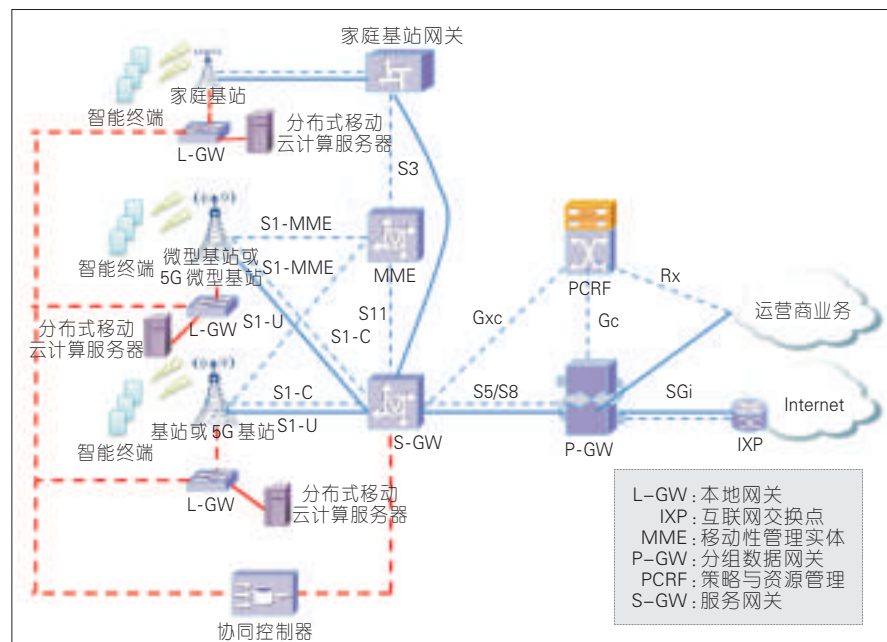
(4) 协同控制器

收集上述三者感知信息,并生成计算卸载分割决策下发给智能终端,以及资源预留决策下发给服务基站与分布式云计算服务器。协同控制器的形式可以是运行于分布式云计算服务器中的实例或虚拟机,也可以是共存于运营商其他网元,例如,服务网关(S-GW)与策略与资源管理(PCRF)等。

尽管目前5G移动网络技术仍处于预研过程中,但是,众多5G网络可以预见的潜在技术为文章提出的分布式移动云计算提供了有效的可行性支撑,潜在技术有:

(1) 超密集蜂部署

延续以提高网络覆盖质量与网络容量为目的的LTE 4G网络微型基站部署的思路,如毫微微基站(Femto cell)基站与微微基站(Pico cell)等。在5G网络微型基站的部署将更加密



▲图1 分布式移动云计算协同架构

集,在同频干扰与无线资源复用问题有效解决的前提下,这种无线组网方式将能够承载更多的移动网络传输,并且接入点与智能终端间的无线传输平均距离的拉近,可以有效降低移动云计算中智能终端无线数据传输的能耗与时延^[13]。

(2)大规模MIMO

作为多输入多输出(MIMO)^[14]技术的衍生增强,大规模MIMO(Massive MIMO)通过在发送端及/或接收端部署更大数量(几十甚至成百)的天线单元,从可实现通过简单的线性预编码算法,获得空间波束赋形增益的大幅度提升,显著增强点对点/多点链路的通信可靠性,链路吞吐量,频谱效率以及能量效率,从而能够以高速率承载更多的并发移动云计算传输。

(3)毫米波回传

随着基站数量的激增,连接宏基站、微型基站与移动交换节点间的回传网络成为网络性能保证的关键。目前,业界已经开始将毫米波(包括71~76、81~86、92~95 GHz频段等)作为4G回传网络的承载方式^[15],其也必将在5G网络中进一步得到增

强。微波回传将为5G网络中包括宏基站与微型基站的回传带宽与时延提供保证,特别是能够有效解决微型基站ADSL回传链路的时延抖动与拥塞问题,从而以低时延、高速率支撑分布式移动云计算协同结构中智能终端与基站向协同控制器的感知上传,以及协同控制器向终端与基站的决策下发。

2 智能终端-移动网络-分布式云服务器协同机制

2.1 协同机制概述

智能终端-移动网络-分布式云服务器协同机制如图2所示。在协

同机制中感知哪些参量与如何生成有效的计算卸载决策与资源预留决策是整套机制的核心问题。时延和终端能耗是衡量移动云计算质量的两项关键指标,为了保证以上两项指标,需要对包括终端、移动网络与分布式移动云计算服务器端到端的状态进行感知,并基于感知内容高效地生成决策。

2.2 智能终端感知

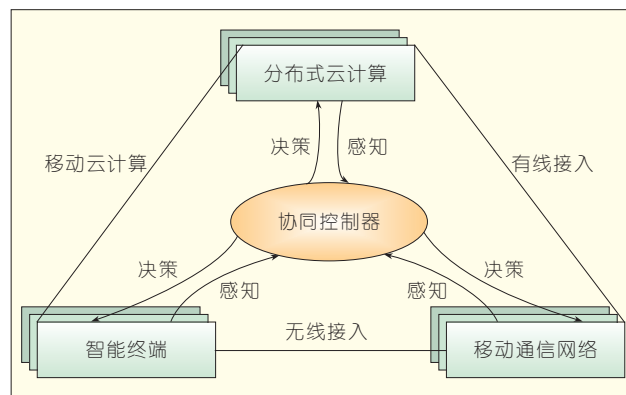
智能终端感知包括对于其应用程序架构与单元信息(如,调用函数)的获取,这是进行计算卸载中应用分割的基础,程序架构与单元信息可以根据调用函数关系以有向图表示。在协同管理中,协同控制器引入终端应用调用函数自由度对终端应用计算进行分割,例如,以二进制因子 $x_i \in \{0,1\}$ 表示调用函数 i 的自由度, $x_i=1$ 表示调用函数 i 卸载到分布式云计算服务器执行,否则,本地执行。

每个调用函数卸载需要无线传输的数据数量(以 A_i 位表示),计算量(以 B_i cycle表示)以及每个调用函数所需CPU计算速率(以 C_i cycle/s表示)是直接计算卸载中终端能耗与服务时延相关的因素。根据已有研究与实测,CPU能耗近似与其占用率成正比,因此,通过这些感知参数,协同控制器估计出调用函数 i 在智能终端本地执行的估计能耗为:

$$P_F \times (C_i/C_F) \times (B_i/C_i) = P_F \times (B_i/C_F) \quad (1)$$

其中, C_F 为最大处理能力,以cycle/s

图2 智能终端-移动网络-分布式云服务器协同机制



为单位; P_F 为 CPU 满载功耗, 以 mW 为单位。如果调用函数 i 卸载到分布式云计算服务器, 则智能终端无线数据传输的能耗为:

$$\frac{A_i \cdot p_i}{B_i \log(1 + |h|^2 / \Gamma(BER) \sigma_n^2)} \quad (2)$$

其中, p_i 为终端发射功率; h 为无线信道参数; σ_n^2 为噪声功率; $\Gamma(BER)$ 为满足 BER 的信噪比余量因子, BER 为目标误码率。

智能终端实际的能耗是反映在其对能量源——锂离子电池的消耗量, 但是, 锂离子电池具有复杂非线性特性^[16], 例如:

(1) SOC-电流非线性特性

电池载荷量(SOC)与电池放电速率之间存在非线性单调关系。

(2) 电压-电流非线性特性

电池端电压并不是恒定的, 它随着放电电流变化而非线性变化。

(3) 自恢复特性

电池在放电后, 则电池容量会自恢复, 并且电池的自恢复能力取决于 SOC 与电池停止放电的时间。

(4) 温度效应

电池在不同温度下放电速度不同, 能够放出的电量也不同。

因此, 感知智能终端锂离子电池必要的状态参数(例如, 温度、电压等)以及良好的电池模型, 是准确估计与预测应用程序及每个调用函数真正能耗的基础, 本文作者在研究成果文献^[16]中给出了单体锂离子电池的精确建模与估计方法, 为移动云计算的提供了基础。

2.3 移动通信网络感知

在采用一系列从物理层到应用层技术保证通信链路的可靠性的情况下, 响应时间已经成为影响移动应用用户体验以及终端能耗的重要因素^[17], 而通过无线通信环境进行计算卸载的移动云计算的性能对响应时间的依赖程度进一步增大。

移动云计算最初的设计初衷即

通过终端的无线带宽资源换取终端计算资源, 降低由高复杂度计算任务导致的计算处理单元高负荷长时间运转而带来的终端能耗^[18]。然而, 当响应时间过长时, 终端的通信单元将长时间处于发送/接收状态, 同样将带来大量的能耗^[19], 可能降低甚至消除移动云计算所期望的终端性能上的增益。

因此, 对终端无线通信环境的响应时间的感知是实现面向移动云计算的良好的协同控制, 自适应选择最佳的协同方式的重要基础。而响应时间决定于包括移动通信网络负载以及无线资源占用率在内的多方面因素, 具体包括:

(1) 移动通信网络接入网的负载状况

包括与终端相联的宏基站/微型基站做覆盖小区的业务复杂量, 以及基站回传网络负载。

(2) 移动通信网络核心网的负载状况

包括核心网中各链路的流量负载状况, 以及核心网中各节点吞吐量负载状况。

(3) 移动通信网络空口无线资源使用情况

包括终端所在小区(或微小区)的无线频谱资源使用情况, 基站信道与功率资源使用情况, 小区间干扰情况等。

上述信息可通过在网络节点处部署探针形成拨测系统, 或引入软件定义网络(SDN)^[20]技术实现感知, 协同控制器在对上述信息进行感知与综合分析情况下, 形成对终端响应时间的准确感知与有效预测, 以支持对移动云计算过程的良好协同控制。

2.4 分布式云计算感知

终端通过移动云计算卸载的计算任务最终需要在分布于移动接入网中的分布式云计算服务器中的虚拟机中完成的。因此, 包括各分布式计算节点的负载, 以及各节点内可承

载的虚拟机数量以及各虚拟机的计算能力在内的分布式云计算感知, 是选择计算卸载节点的重要依据, 对移动云计算性能具有重要影响。

由上所述, 分布式云计算感知可分为两个层次进行, 即服务器层次(即整个节点层次)以及虚拟机层次。

(1) 服务器层次

分布式云计算感知将采集整个计算节点服务器(或服务器集群)的负载情况, 包括:

- 服务器吞吐量(若为服务器集群则为每台服务器吞吐量, 其他指标下同), 服务器并发通信状态, 服务器计算资源使用情况(即服务器的 CPU 占用率), 服务器存储资源占用率等。

- 当计算节点为服务器集群且在集群内部采用了虚拟化技术实现整个集群的虚拟化资源池时, 服务器层次感知信息应充分考虑虚拟资源池的整体情况。

(2) 虚拟机层次

分布式云计算感知将对整个节点内虚拟机状态信息进行采集, 信息包括:

- 虚拟机的数量、各虚拟机所占有的计算与存储资源量以及各虚拟机等效产生的吞吐量。

- 值得注意的是, 为了实现对虚拟机的承载, 服务器需提供一定的承载资源(包括计算处理资源与存储资源)来构建承载虚拟机的系统环境, 分布式云计算感知同时需要掌握承载资源的相关状态信息。

目前主流服务器产品均支持对上述信息的软件化采集与存储, 通过高效的接口与信息共享机制, 实现协作控制器的分布式云计算高效感知。

2.5 协同决策

协同控制器通过智能终端感知、移动通信网络感知、分布式云计算感知, 充分掌握终端、移动通信网络(包括接入网与核心网)以及各分布式云计算节点的状态信息, 通过综合分析产生协同决策并分别下发至移动终

端、移动通信网络网元以及分布式云计算节点,各部分按照决策执行各自相应的动作实现协同的分布式云计算,为终端应用提供用户体验保障。

分布式移动云计算协同决策如图 3 所示。协同控制器根据前述感知信息为移动终端、移动通信网络网元以及分布式云计算节点生成决策,每个部分获得相应的决策信息。

(1) 移动终端部分决策信息

- 根据终端目前的电池、能耗及计算资源状态,决定相应移动应用的计算子任务划分,并进一步根据终端的无线带宽资源状态,规划本地计算任务及卸载计算任务。

- 当终端可接入多基站时,决策同时涉及对每个卸载计算子任务指定相应的接入基站,此时终端应用需对乱序返回的计算子任务结果正确重组。

(2) 移动通信网络部分决策信息

- 根据终端与各接入基站间的无线带宽状态,以及各基站的回传网络拥塞状态,为卸载计算任务数据及返回结果数据分配相应接入点完成接收与发送,具体包括:当终端应用想远端发起计算卸载请求时,通知相应基站负责接收来自终端的包含相应远端计算子任务的空口数据;当分布式云计算节点返回子计算任务结果时,指派相应基站向指定终端发送包含返回子计算任务结果的空口数据;考虑到用户移动性,相同子计算任务的请求及结果的发送与接收基站需要分别动态指派。

- 当卸载计算任务数据到达移动网络核心网,根据核心网承载网络状态(拓扑、链路流量、节点负荷等),为各卸载计算子任务数据选择路由,实现各卸载计算子任务数据到指定的分布式云计算节点的高效转发,降低核心网网络负荷及时延;反之,当分布式云计算节点返回各卸载计算子任务结果时提供向指派接入网节点的路由转发;该过程可依托于软件定义网络技术实现路由转发方面的动态适配。

(3) 分布式云计算节点决策信息

- 根据各分布式云计算节点的状态信息,为各卸载计算子任务指派相应分布式云计算节点完成计算承载任务。

- 决策每个分布式云计算节点承载相应卸载计算子任务所需虚拟化资源,生成虚拟机完成计算任务。

- 多个分布式云计算节点间的计算进程实现并行协同。

3 结束语

移动云计算将终端计算任务卸载到云端,在降低终端能耗的同时,提供了在终端有限资源情况下实现大计算量移动应用能力,必将成为构建未来移动互联网创新服务的必不可少的关键技术。

然而,目前移动云计算现阶段技术发展对于移动设备到云计算中心的端到端网络时延与带宽不稳定,以及影响移动云计算服务质量的主要因素动态性等问题仍然缺乏有效

的对抗手段,同时在计算卸载由终端感知与协商决策方面缺少有效的机制设计。

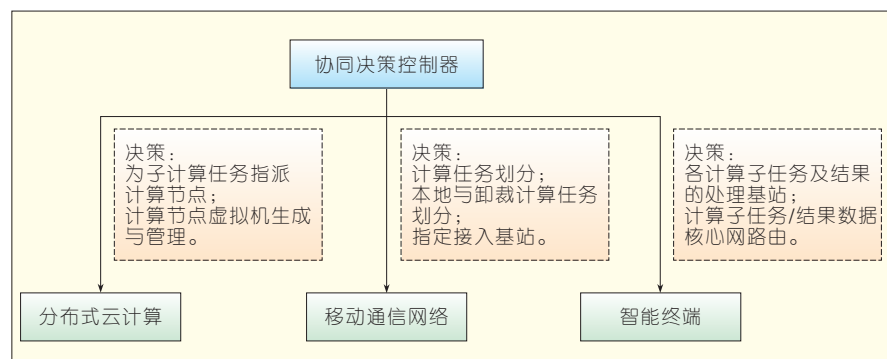
本文结合 4G、5G 网络无线接入网技术的特点,提出了一种分布式移动云计算协同架构,以及端到端终端计算卸载服务质量保证机制。

文章所提出的架构与机制可以降低终端在计算卸载中的网络交互信令开销,以及终端驻留决策进程的 CPU 占用率与能耗;可以保证移动云计算的端到端服务质量;可以规避传统移动云计算给 WAN 带来的巨大负载量与能耗。

可以预见,在用户、智能设备商和移动网络运营商的多方推动下,未来分布式移动云计算将成为 4G 与 5G 网络的主流技术之一,成为移动运营商的基础网络设施并为其提供新的业务增长点,为智能终端应用提供更佳的用户体验。

参考文献

- [1] PALACIN M R. Recent advances in rechargeable battery materials: a chemist's perspective [J]. Chemical Society Reviews, 2009, 38(9): 2565–2575
- [2] LEWIS M B, EDMONDS A J. Face detection: Mapping human performance [J]. Perception–London, 2003, 32(8): 903–920
- [3] RAMON M, CAHAREL S, ROSSION B. The speed of recognition of personally familiar faces [J]. Perception–London, 2011, 40(4): 437
- [4] FLINN J, PARK S Y, SATYANARAYANAN M. Balancing performance, energy, and quality in pervasive computing [C]//Proceedings of the 2002 IEEE International Conference on Distributed Computing Systems, July 2–5, 2002, Vienna, Austria. Piscataway, NJ, USA: IEEE, 2002: 217–226
- [5] BALAN R K, SATYANARAYANAN M, PARK S Y, et al. Tactics-based remote execution for mobile computing [C]//Proceedings of the 2003 ACM International Conference on Mobile systems, applications and services, May 5–8, 2003, San Francisco, CA, USA. New York, NY, USA: ACM, 2003: 273–286
- [6] KEMP R, PALMER N, KIELMANN T, et al. Cuckoo: a computation offloading framework for smartphones [J]. Mobile Computing, Applications, and Services, 2012, 16(4): 59–79
- [7] CUERVO E, BALASUBRAMANIAN A, CHO D, et al. MAUI: making smartphones last longer with code offload [C]//Proceedings of the 2010 ACM International Conference on Mobile systems, applications, and services, June 15–18, 2010, San Francisco, CA, USA.



▲ 图 3 分布式移动云计算协同决策

- New York, NY, USA: ACM, 2010: 49–62
- [8] CHUN B G, MANIATIS P. Augmented Smartphone Applications Through Cloud Execution [C]//Proceedings of the 2009 IEEE Workshop on Hot Topics in Operating Systems, May 18–20, 2009, Monte Veritá, Switzerland. Piscataway, NJ, USA: IEEE, 2009: 8–11
- [9] HUANG D, ZHANG X, KANG M, et al. MobiCloud: building secure cloud framework for mobile computing and communication [C]//Proceedings of the 2010 IEEE International Symposium on Service Oriented System Engineering, June 4–5, 2010, Nanjing, China. Piscataway, NJ, USA: IEEE, 2010: 27–34
- [10] KRISTENSEN M D, BOUVIN N O. Scheduling and development support in the scavenger cyber foraging system [J]. Pervasive and Mobile Computing, 2010, 6 (6): 677–692
- [11] LI A, YANG X, KANDULA S, et al. CloudCmp: comparing public cloud providers [C]//Proceedings of the 2010 ACM SIGCOMM conference on Internet measurement, Vol 15. Nov. 1–3, 2010, Melbourne, Australia. New York, NY, USA: ACM, 2010: 1–14
- [12] LI H, HU D, CI S. High-Resolution Cell Breathing for Improving Energy Efficiency of Ultra-Dense HetNets [C]//Proceedings of the IEEE Wireless Communications and Networking Conference, March 9–12, 2015, New Orleans, LA, USA. Piscataway, NJ, USA: IEEE, 2015: 1470–1475
- [13] LI H, XU X, HU D, et al. Clustering strategy based on graph method and power control for frequency resource management in femtocell and macrocell overlaid system [J]. Communications and Networks, 2011, 13 (6): 664–677
- [14] LARSSON E G, EDFORS O, TUFVESSON F, et al. Massive MIMO for next generation wireless systems [J]. Communications Magazine, 2014, 52(2): 186–195
- [15] LITTLE S. Is microwave backhaul up to the 4G task? [J]. Microwave Magazine, 2009, 10 (5): 67–74
- [16] LIN N, CI S, LI H. An enhanced circuit-based battery model with considerations of temperature effect [C]//Proceedings of the IEEE Energy Conversion Congress and Exposition (ECCE), Sept. 14–18, 2014, Pittsburgh, PA, USA. Piscataway, NJ, USA: IEEE, 2014: 3985–3989
- [17] HASSAN M A, CHEN S. Mobile mapreduce: Minimizing response time of computing intensive mobile applications [J]. Mobile Computing, Applications, and Services, 2012, 16(4): 41–59
- [18] KUMAR K, LU Y H. Cloud computing for mobile users: Can offloading computation save energy? [J]. Computer, 2010, 43(4): 51–56
- [19] BARBERA M V, KOSTA S, MEI A, et al. To offload or not to offload? The bandwidth and energy costs of mobile cloud computing [C]//Proceedings of the 2013 IEEE International Conference on Computer Communications, April 14–19, 2013, Turin, Italy. Piscataway, NJ, USA: IEEE, 2013: 1285–1293
- [20] JAIN R, PAUL S. Network virtualization and software defined networking for cloud

computing: a survey [J]. Communications Magazine, 2013, 51(11): 24–31

作者简介



李宏佳, 中国科学院信息工程研究所助理研究员、博士; 研究领域为宽带无线通信、移动云计算、自组织异构蜂窝网络及绿色信息通信; 已主持和参加基金项目 11 项; 已发表学术论文 40 余篇, 其中被 SCI/EI 检索 20 余篇。



陈鑫, 中国科学院声学研究所博士后; 研究领域为无线资源管理、绿色通信及 SDN 关键技术。



周旭, 中国科学院信息工程研究所副研究员、博士; 研究领域为移动云计算、移动内容分发。

←上接第 13 页

间。但如何做到移动云存储服务最优化, 还需权衡文件同步效率与移动端计算开销和能量开销之间的关系。

参考文献

- [1] DRAGO I, BOCCHI E, MELLIA M, et al. Benchmarking personal cloud storage [C]//Proceedings of the 2013 conference on Internet measurement conference. ACM, 2013: 205–212
- [2] DRAGO I, MELLIA M, MUNAFO M, et al. Inside dropbox: understanding personal cloud storage services [C]//Proceedings of the 2012 ACM conference on Internet measurement conference. ACM, 2012: 481–494
- [3] 崔勇, 张鹏. 无线移动互联网: 原理、技术、应用 [M]. 北京: 机械工业出版社, 2012
- [4] KIM H W, CHAN H C, GUPTA S. Value-based adoption of mobile internet: an empirical investigation [J]. Decision Support Systems, 2007 43(1): 111–126
- [5] BUCHANAN G, FARRANT S, JONES M, et al. Improving mobile internet usability [C]//Proceedings of the 10th international conference on World Wide Web. ACM, 2001: 673–680
- [6] CLEMENTS P, NORTHROP L. Software product lines: practices and patterns [M]. Reading: Addison-Wesley, 2002
- [7] BOSCH J. Design and use of software architectures: adopting and evolving a product-line approach [M]. Pearson Education, 2000
- [8] SPEISER S, HARTH A. Taking the LIDS off data silos [C]//Proceedings of the 6th International Conference on Semantic Systems. ACM, 2010: 44
- [9] BIZER C. The emerging web of linked data [J]. Intelligent Systems, IEEE, 2009, 24(5): 87–92
- [10] JAIN R. Quality of experience [J]. IEEE MultiMedia, 2004, 11(1): 96–95
- [11] HARMAN G. The intrinsic quality of experience [J]. Philosophical perspectives, 1990, 19(3): 31–52
- [12] FIEDLER M, HOSSFELD T, TRAN-GIA P. A generic quantitative relationship between quality of experience and quality of service [J]. Network, IEEE, 2010, 24(2): 36–41
- [13] KUMAR K, LU Y H. Cloud computing for mobile users: Can offloading computation save energy? [J]. Computer, 2010, 43(4): 51–56
- [14] YOO J W, PARK K H. A cooperative clustering protocol for energy saving of mobile devices with WLAN and bluetooth interfaces [J]. Mobile Computing, IEEE Transactions on, 2011, 10(4): 491–504
- [15] YUAN W, NAHRSTEDT K. Energy-efficient soft real-time CPU scheduling for mobile multimedia systems [J]. ACM SIGOPS Operating Systems Review, 2003, 37(5): 149–163

作者简介



崔勇, 清华大学计算机系教授、博导, 清华大学网络技术研究所副所长, 国家优秀青年科学基金和教育部新世纪人才获得者, 国际互联网标准化组织 IETF IPv6 过渡工作组主席; 研究方向为 IPv6 与下一代互联网、无线移动计算等; 已发表学术论文 100 余篇, 出版学术著作 3 部。



赖泽祺, 清华大学计算机系在读博士研究生; 研究领域为移动互联网、个人云存储服务。



缪葱葱, 清华大学计算机系在读博士研究生; 研究领域为个人云存储服务。

移动云计算与运营商自有业务云化探讨

Mobile Cloud Computing and Cloud-Based Operator Services

符刚/FU Gang
薛淼/XUE Miao
唐雄燕/TANG Xiongyan

(中国联通网络技术研究院, 北京 100048)
(Network Technology Research Institute of
China Unicom, Beijing 100048, China)

移动网络的快速发展及智能终端的普及催生了移动互联网的繁荣。近年来,移动设备上的应用(业务)呈现爆炸式发展,功能涵盖娱乐、健康、游戏、商业、金融、旅游、社交及新闻等多个方面。但是由于移动终端在计算、存储等方面的能力限制,业内寻求通过云计算的方式为移动终端提供云业务,从而减少其在计算、存储方面的弱势,同时基于共享的业务数据和集中的用户数据发掘新的移动业务和应用。移动云计算/移动云业务为基于移动网络的业务发展提供了新的思路 and 手段。

而随着4G网络的部署,移动网络的带宽得到大幅提升,移动互联网迎来了新的发展机遇。与此同时,由于4G网络的扁平化,运营商自有业务(包括语音和消息等业务)也迁移到全IP的IP多媒体子系统(IMS)网络,不仅具备提供更丰富业务的能力,也为运营商构建自有业务云提供了基础。

本文聚焦于运营商自有业务与

收稿日期: 2015-01-21
网络出版时间: 2015-02-23
基金项目: 国家科技支撑计划
(2012BAH06B01)

中图分类号: TN929.5; TP393 文献标志码: A 文章编号: 1009-6868 (2015) 02-0020-005

摘要: 结合移动云计算和网络功能虚拟化(NFV)等产业背景,对4G时代运营商自有业务网的网络云化和业务云化进行探讨;基于对网络云化步骤和方式的分析,给出了业务云化的基本架构和典型业务场景;在自有业务云化基础上,通过控制转发分离/软硬件分离技术,提出了运营商云化自有业务网络架构,以提供高度可扩展和平滑演进能力的网络。

关键词: 移动云计算;网络功能虚拟化;自有业务云化;融合通信

Abstract: Considering the industry background of Mobile Cloud Computing (MCC) and Network Function Virtualization(NFV), we investigate a cloud-based operator network and services in the 4G era. In this paper, we analyze the evolving steps and approaches of a cloud-based operator network and picture the cloud-based operator service architecture and classic service scenarios. Using the control-forwarding separating/software-hardware decoupling technologies, we design a cloud-based operator service network to provide scalability and seamless evolution.

Keywords: mobile cloud computing; network function virtualization; cloud-based operator's services; unified communication

云计算结合,分析4G时代运营商自有业务网络建设,探讨基于云的运营商自有业务发展。

1 产业背景

1.1 移动云计算

移动云计算^[1-3]是移动计算和云计算技术的结合,其同时具备移动和云的特性。因此移动云计算可以定义为:集中的(虚拟的)数据中心向移动设备提供移动应用和业务。

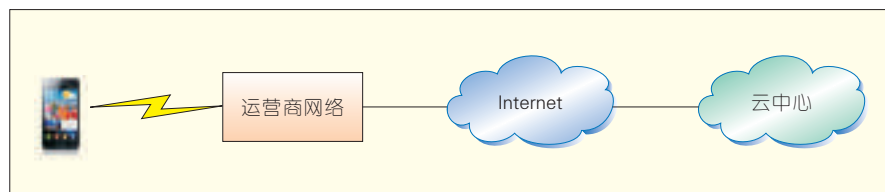
移动云计算允许移动业务使用移动设备外部的弹性云资源。这样移动设备不需要具有很大的存储空间和很强的CPU计算能力,数据处理集中在位于云中的应用平台完成。一方面,移动云计算可以减少移动终端成本;另一方面,移动业务开发者

可以充分利用云中心的资源池,减少开发成本,实现业务迅速迭代。

移动云计算架构如图1所示。移动云计算架构包括移动终端、运营商网络、Internet网络、云中心。其中,移动终端包括智能手机、平板等通过移动宽带或者Wi-Fi等无线技术接入网络的移动设备;运营商网络为移动终端提供信号覆盖、移动性管理,保证移动终端任何时间任何地点能够接入网络;Internet提供移动业务数据流的路由;云中心为移动业务提供平台,其可以是基础设施即服务(IaaS)、软件即服务(SaaS)或者平台即服务(PaaS)。

目前,移动云计算主要包括3种类型:

(1)移动业务,如新闻、旅游等内容服务,在资源丰富的云中心运行,



▲图1 移动云计算架构

移动终端通过移动宽带或者Wi-Fi进行接入。

(2)将其他移动设备作为资源提供者,从而组成移动P2P网络。这样附近移动设备的资源以及其他远端固定设备的资源可以被集合起来统一使用。

(3)移动终端将部分存储、计算功能云化。对于需要复杂计算、大容量存储的业务,移动应用(APP)分割部分计算、存储功能到云中心,从而减少对终端的能力需求,如网盘、在线游戏等。

移动云计算不仅可以缓解移动终端资源紧缺的问题,通过用户数据的集中,如用户位置信息,也可以开发新的应用。

1.2 网络功能虚拟化

网络功能虚拟化(NFV)^[4-5]将传统电信网功能网元软硬件解耦,使得功能软件化,硬件通用化。NFV复用虚拟化技术将网络功能统一部署到通用服务器、交换机和存储平台上,从而能够在硬件资源池上部署/删除/移动网络功能。

欧洲电信标准组织(ETSI)NFV组是NFV相关的主要标准制订组织,其定义了NFV的基本框架。NFV充分复用了IT虚拟化和云计算技术,同时提出了适用于电信网元管理的功能需求,如网络业务(NS)的自动化管理、运营支持系统/商业支持系统(OSS/BSS)对接等。

NFV技术能够提高运营商网络的灵活性和可扩展性,减少业务上线时间,通过自动化管理提高运营效率,降低运营成本。NFV的出现,一方面简化了运营商网络基础设施,统

一了底层硬件;另一方面,虚拟化技术的引入必然导致运营商网络向云化演进。

随着LTE的部署及VoLTE的逐步商用,运营商业网全面转向IP化。运营商在考虑云化组网的同时,也可以充分利用移动云计算,发展运营商自有业务云。

2 网络云化

运营商自有业务网络的发展正沿着设备集中、资源集约、业务集成的方式演进。这里自有业务网络主要指IMS网络,也包括演进的分组核心网(EPC)及其他业务平台。

传统的自有业务网络建设思路是以省为中心。但是随着自有业务向IMS网络迁移,自有业务网络具有了更大范围集中的可能。更大范围集中的自有业务网络具备几个优势:可以集中建设通用化的基础设施,从而借助NFV构建云化的业务网络;自有业务云实现运营业务的统一管理,灵活快速业务迭代,统一业务刚性落地;计费、运营支撑系统全网统一集中建设;用户数据、业务数据集中,大数据发掘创新业务。

自有业务网络的云化^[6]需要重点考虑网元自身属性的区别和网元部署位置的区别两个方面的影响,其决定了不同网元虚拟化部署的先后顺序。控制面功能网元主要占用计算资源和存储资源,具备先行实施虚拟化的条件;数据面功能网元需要提供高效转发能力,目前这种高效转发能力主要基于专用芯片,采用x86架构承载数据面功能,其转发性能仍需要经过现网检验。

运营商自有业务网可以认为主

要分为如下几个组成部分IMS、EPC、2/3G PS、2/3G CS、HSS 5个主要组成部分,如表1所示。

从表1中可以看出,IMS网络中的网元IT化程度最高,控制面网元往往在省或大区中心集中部署,便于向云化过渡,这也是各厂家优先推出IMS虚拟化产品的原因。未来新建IMS网络可以考虑采用虚拟化的方式。IMS云化部署如图2所示。

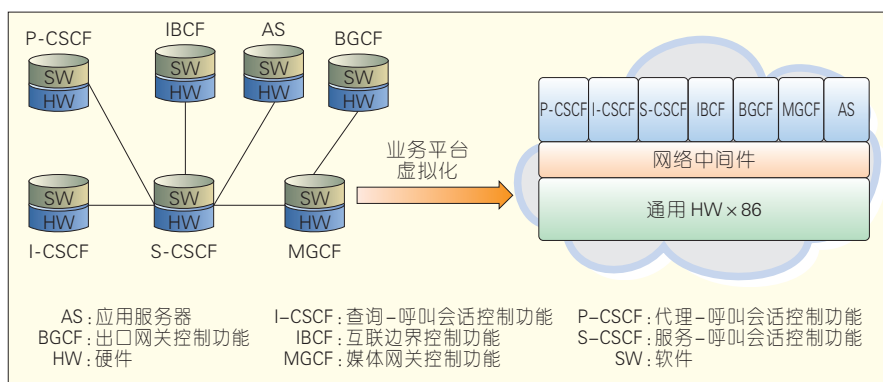
EPC网络中,移动管理实体(MME)、Diameter路由代理(DRA)和PCRF与IMS控制网元类似,都对计算资源有较大需求,可以集中部署并向虚拟化过渡。但是网关具有控制面和转发面双重属性,虽然转发面功能目前仍采用集中的方式部署在省或大区中心,但是随着汇聚流量的增加以及语音、车载网等Peer-to-Peer业务的部署,数据面功能部署位置可能需要更靠近用户。2/3G PS域中的SGSN和GGSN面临SGW/PGW相同的问题。而转发面功能虚拟化目前主要有两种实现方式:采用专用转发面芯片及板卡处理数据转发;采用通用x86架构,但是需要加速网卡和加速软件包。无论哪一种情况,数据面功能都仍需要再特定硬件上部署,难以做到自由迁移。此外,在LTE与3G混合组网的情况下,为了保证LTE和3G互操作,往往把MME和SGSN合设,此时需要综合考虑MME-SGSN的虚拟化,即控制面和转发面都需要考虑,这也提高了MME虚拟化的难度。分组域虚拟化部署如图3所示。

2/3G CS域MSC可以仅作为控制面网元,不做数据转发,因此其具有过渡到虚拟化的条件。虽然2/3G CS域的一些网元具备向虚拟化过渡的条件,但实际上NFV并没有相关的建议用例,这也是充分考虑到LTE部署以后,2/3G CS网元将长期退网的发展趋势。

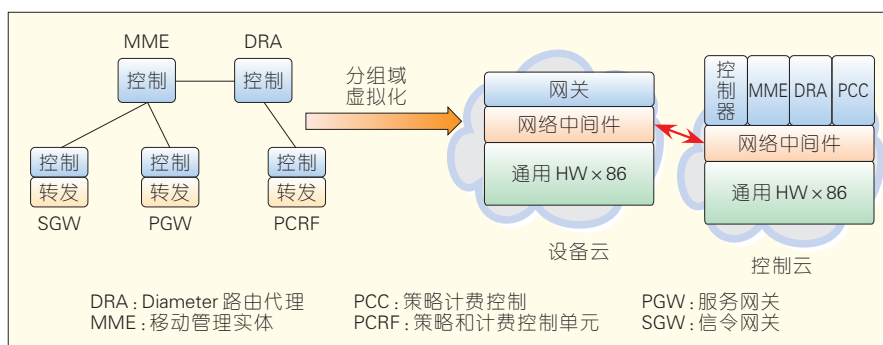
归属用户服务器/归属位置寄存器(HSS/HLR)主要用于存储用户数据。早期HLR设备IT化程度不高,

▼表 1 自有业务网网元特征

业务网组成部分	IMS						EPC				2/3G-CS		2/3G-PS		HSS	
	P/I/S-CSCF	IBCF/MGCF/BGCF	AS	IM-MGW	SBC/P-CSCF 合设	SBC	SGW/PGW	DRA	MME	PCRF	MSC	MGW	GGSN	SGSN	HSS	HLR
功能	控制	控制	控制	转发	控制+转发	转发	控制+转发	控制	控制	控制	控制	转发	控制+转发	控制+转发	控制	控制
设备形态	ATCA	ATCA	服务器	ATCA	ATCA	ATCA	ATCA	ATCA	ATCA	ATCA	ATCA 或私有平台	ATCA 或私有平台	ATCA 或私有平台	ATCA 或私有平台	ATCA 或服务器	小型机和私有平台
接口类型	全 IP	全 IP	全 IP	非全 IP	全 IP	全 IP	全 IP	全 IP	全 IP	全 IP	非全 IP	非全 IP	全 IP	全 IP	全 IP	非全 IP
部署方式	省或大区集中	省或大区集中	省或大区集中	地市	省或大区集中	地市	省集中或地市	省集中或大区集中	省集中	省集中	省集中或地市	地市	省集中	省集中或地市	省集中或大区集中	地市
AS: 应用服务器 ATCA: 先进的电信计算平台 BGCF: 出口网关控制功能 CS: 电路域 DRA: Diameter 路由代理 EPC: 演进的分组核心网 GGSN: 网关通用分组无线业务 HLR: 归属位置寄存器 HSS: 归属用户服务器 I-CSCF: 查询-呼叫会话控制功能 IBCF: 互联边界控制功能 IM-MGW: IM 网关 IMS: IP 多媒体子系统 MGCF: 媒体网关控制功能 MGW: 媒体网关 MME: 移动管理实体 MSC: 移动交换中心 P-CSCF: 代理-呼叫会话控制功能 PCRF: 策略和计费控制单元 PGW: 服务网关 PS: 分组域 S-CSCF: 服务-呼叫会话控制功能 SBC: 会话边界网关 SGW: 信令网关 SGSN: 服务通用分组无线业务支持节点																



▲图 2 IMS 虚拟化部署示意



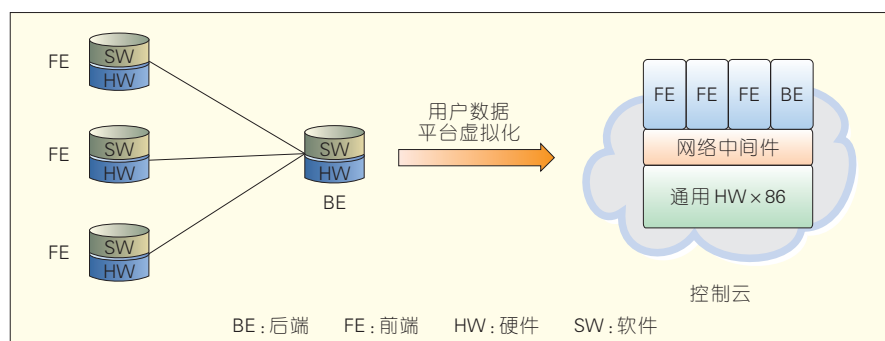
▲图 3 分组域虚拟化部署示意

一些接口仍没有 IP 化；而且随着 LTE 及 VoLTE 的部署，早期 HLR 面临过渡到融合 HSS 的需求，因此不具备虚拟化的客观需求。而 HSS 融合数据库

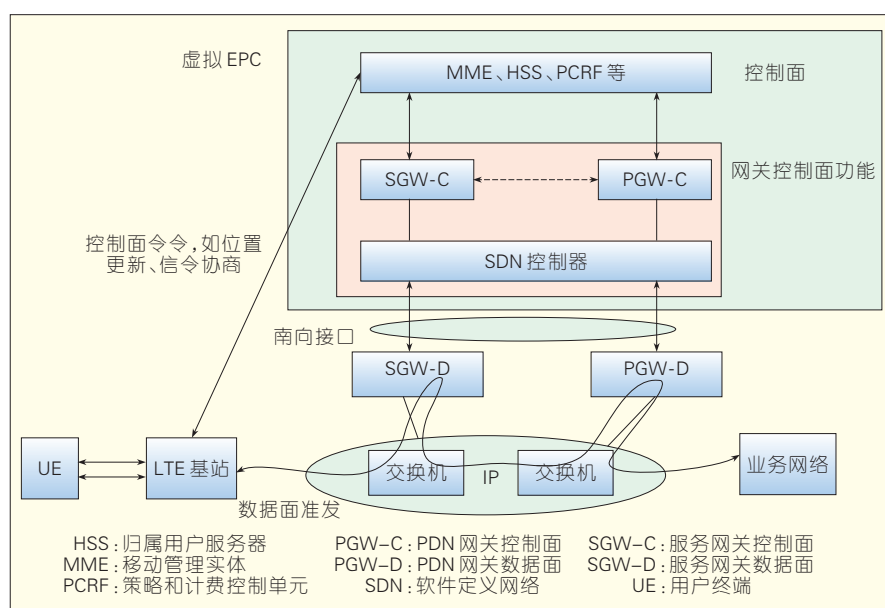
则大多采用用户数据集中（UDC）架构，后端（BE）作为数据库存储拥护数据，具有较大的存储能力，往往在省或大区中心集中部署；前端（FE）

对用户数据请求进行响应，也具备虚拟化实施条件。HSS 虚拟化部署如图 4 所示。

基于 SDN 的控制转发分离的 EPC 架构如图 5 所示，为分组域另一种云化方式。图 5 中将原 EPC 网关的控制面和转发面解耦，从而实现网关中 GPRS 通道协议-用户面（GTP-U）隧道的管理和 GTP-U 隧道数据传输的分离。如图 5 网关解耦后的控制面网元为服务网关控制面（SGW-C）、PDN 网关控制面（PGW-C）、SDN 控制器，SGW-C 和 PGW-C 保留与 MME/HSS/PCRF 的 3GPP 定义接口；转发面网元为服务网关数据面（SGW-D）、PDN 网关数据面（PGW-D），需要支持与 SGW-C/PGW-C 的南向接口。业务网关/PDN 网关控制面（xGW-C）和 SDN 控制器可以合设为同一网元。xGW-C 负责 GTP 隧道的建立/删除/更新，相关参数通过 SDN 控制器南向接口分发给业务网关/PDN 网关数据面（xGW-D）设备，从而建立 GTP-U 隧道的逐主段关联。SGW-D/PGW-D 对 GTP-U 数据包封装/解封装并完成转发操作，并对数据包数量进行计数进而辅助生成话单。在控制转发分



▲ 图4 HSS虚拟化部署



▲ 图5 基于SDN的控制转发分离的EPC架构

离的EPC架构中,数据面网元更专注于数据转发,控制面可以以逻辑集中的方式部署。

总之,业务网络云化成为运营商CT向IT转型,提升核心网络能力的一种有效手段。

3 业务云化

4G时代,随着VoLTE的逐步商用及IMS的部署,运营商提供的业务不仅限于语音和短信,还包括其他如文件传输、图片/视频共享等RCS^[7-8]业务。IMS提供了融合通信的网络基础。用户可以借助与原生终端的业务界面完成各种融合通信的业务,而无需下载其他OTT的APP。运营商提供融合通信业务,可以控制通信入

口,提高用户的黏性,便于运营商推动加载新的业务。

融合通信可以在网络云化的基础上进行业务云化。业务云化支持多租户,实现资源共享;提供灵活的多媒体服务;具备灵活的业务开通及能力开放。业务云化由运营商网络

分担融合通信业务的对移动终端计算、存储能力的要求,例如为文件共享提供缓存和云盘服务。

业务云化平台可以利用运营商靠近用户的优势,开发新的自有业务,作为增值业务或者公共服务。同时,运营商也可以与OTT合作,形成共享或者代理关系。运营商自有业务云化的架构如图6所示。

业务云化典型场景包括:

(1)云端地址簿。用户将地址簿保存在云端,防止终端丢失或者发生误删除。

(2)基于位置服务。根据用户的注册位置,提供社交或者公共服务,如判断交通状况、人群集会会产生拥挤告警等。

(3)云端存储。为用户共享的图片及文件提供缓存,为用户开辟与融合通信界面合一的云盘存储。

(4)移动商业。运营商提供高优先级接入保障,同时长时间保留交易记录。

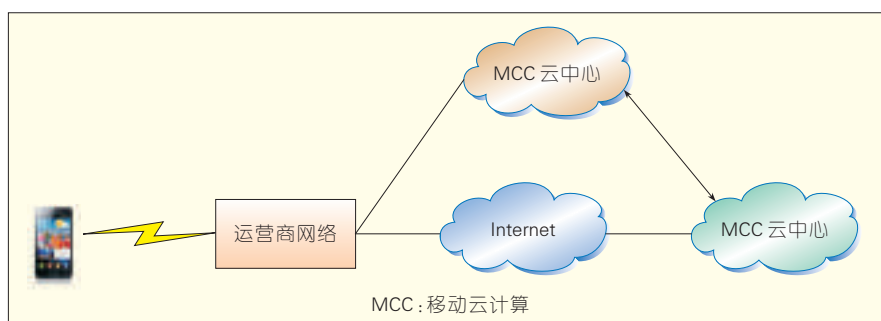
(5)健康监测。根据移动终端传感器,长期记录用户的活动健康状况,提供健康服务。

(6)灾难救援。将用户传输的图片信息云化存储,集中分析,协助救援方案制订。

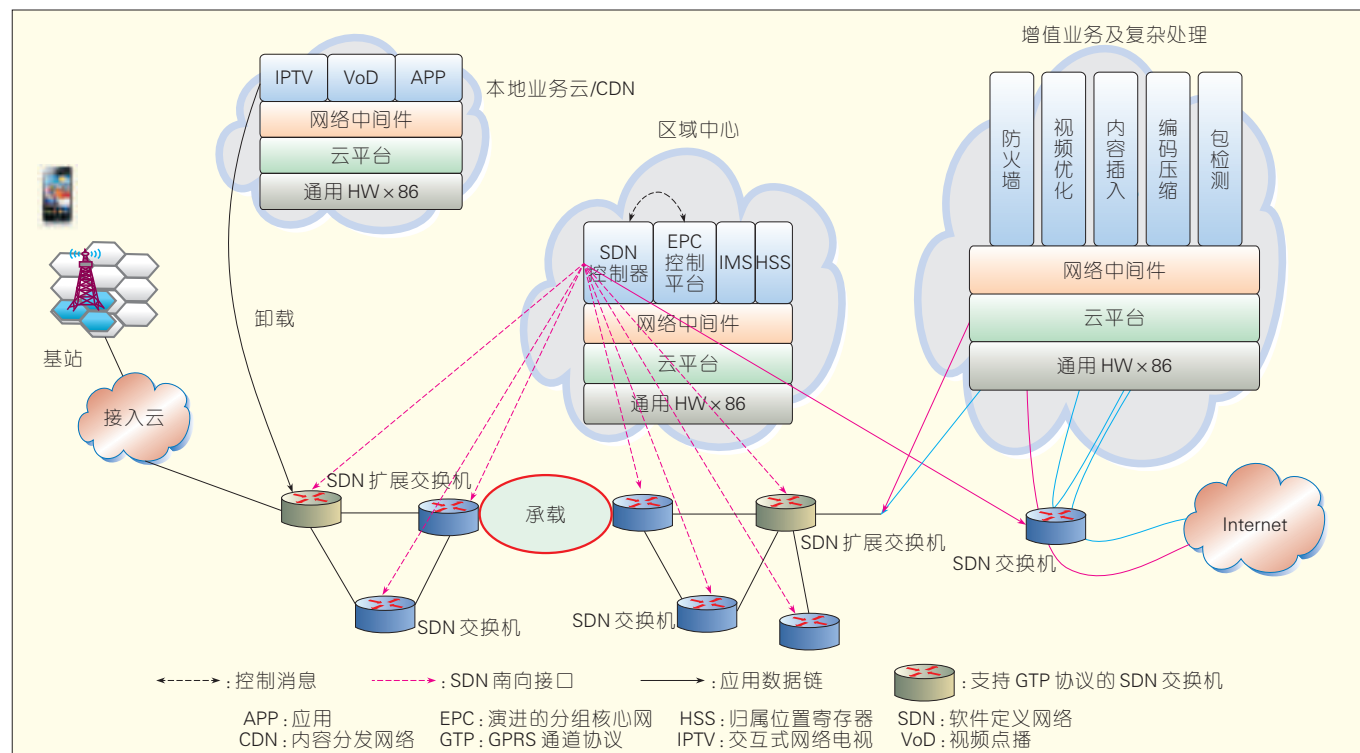
运营商自有业务云化不仅能够拓宽运营商自有业务范围,同时也一定程度上提升了运营商参与公共事务的地位。

4 云化自有业务网络架构

云化自有业务网络架构如图7所示



▲ 图6 运营商自有业务云化的架构



▲图7 云化自有业务网络架构

示。云化自有业务网络架构一方面从“水平面”进一步解耦网关类设备的控制面和转发面；另一方面，从“垂直面”解耦网元功能软件和硬件。云化自有业务网络架构提供集中的网络控制面及自有业务云化，同时提供数据增值业务云化及内容本地化，从而使得全IP的运营商网络具有更好地可扩展性，同时具备向未来网络及代级网络平滑演进的能力。

5 结束语

随着4G时代的到来，运营商的自有业务网络也将全面数据化/IP化，这对运营商既是一个挑战，但也是一个机会。借助于IP化和云计算，运营商网络正在向着IT迅速靠近。本文在总结移动云计算和NFV等产业背景的基础上，对4G时代运营商自有业务网的网络云化和业务云化进行分析和探讨，并给出了云化自有业务网络架构。运营商构建云化自有业务网是大势所趋，但是每个运营商又有各自的实施路线和方案。数

据时代的到来，运营商的网络架构将更加差异化，业务也更将差异化，如2/3G时代统一的网络结构的时代将彻底改变。

参考文献

- [1] FERNANDO N, LOKE S W, RAHAYU W. Mobile cloud computing: A survey [J]. Future Generation Computer Systems, 2013, 29(1): 84-106
- [2] KHAN A U R, OTHMAN M, MADANI S A, KHAN S U. A Survey of Mobile Cloud Computing Application Models [J]. IEEE Communications Surveys and Tutorials, 2014, 16(1): 393-413
- [3] YANG L, CAO J N. Computation Partitioning in Mobile Cloud Computing: A Survey [J]. ZTE Communications, 2013, (12): 8-17
- [4] ETSI. ETSI ISG NFV. Network Functions Virtualization: An Introduction, Benefits, Enablers, Challenges & Call for Action [S]. ETSI, 2013
- [5] ETSI. ETSI ISG NFV. Network Functions Virtualization (NFV): Network Operator Perspectives on Industry Progress [S]. ETSI, 2014
- [6] 薛淼, 符刚, 朱斌, 李勇辉. 基于SDN/NFV的核心网演进关键技术研究 [J]. 邮电设计技术, 2014, 26(3): 16-22
- [7] ONF. Software-Defined Networking: The New Norm for Networks [S]. ONF, 2012
- [8] GSMA. Rich Communication Suite 5.2 Advanced Communications Services and

Client Specification Version 5.0 [S]. GSMA, 2014

作者简介



符刚, 中国联通网络技术研究院高级工程师; 主要从事移动通信核心网络新技术、业务网架构演进等研究工作; 已发表学术论文10余篇, 申请专利40余项。



薛淼, 中国联通网络技术研究院工程师、博士; 主要从事下一代网络架构演进、核心网新技术研究及核心网标准工作; 已发表学术论文10余篇, 申请专利10余项。



唐雄燕, 中国联通网络技术研究院首席专家, 北京邮电大学兼职教授、博士生导师, 工业和信息化部通信科技委委员兼电信传输专家咨询组副组长, 中国通信标准化协会TC10副主席; 主要专业领域为核心网、宽带通信、光纤传输、IP技术、下一代网络、泛在网络等; 发表技术论文150余篇, 出版专著6部。

移动云计算 ——移动增强现实技术和服务

Mobile Augmented Reality Technology and Service in Mobile Cloud Computing

吕强/Lü Qiang
黄成/HUANG Cheng
刘明/LIU Ming

(中兴通讯股份有限公司, 深圳 518057)
(ZTE Corporation, Shenzhen 518057, China)

中图分类号: TN929.5; TP393 文献标志码: A 文章编号: 1009-6868 (2015) 02-0025-005

摘要: 提出了中兴通讯移动增强现实解决方案。方案针对增强现实目标的识别技术采用客户机/服务器(C/S)架构,以充分利用云的海量存储和并行计算能力,进行样本图像的训练以及识别样本库的生成和训练,从而实现快速的增强现实目标的迅速识别;针对海量增强现实内容信息和训练样本数据高效存储的需求,采用分布式的非结构化存储机制以保存图像等信息,对于兴趣点信息采用结构化存储机制以保存兴趣点各种增强现实元数据信息;通过在结构化存储机制中建立特征索引等快速查询机制,以达到高效数据查询的目的。此外,为了分担增强现实业务请求负载,在靠近用户的地方部署若干用户接入云,缓存热点的增强现实信息,减轻增强现实业务处理机的负载压力。

关键词: 移动增强现实; 增强现实系统; 目标识别; 云计算

Abstract: In this paper, we present ZTE's Mobile Augmented Reality system. To take advantage of mass storage and parallel computing power of the cloud, we choose the C/S architecture. This architecture trains sample images training and generates sample library for rapid AR target recognition. To meet the efficient storage requirements of massive AR Content information and related training sample, we save the images based on the distributed unstructured storage and save the augmented reality metadata, such as POI information, based on structured storage. We establish indexing in our structured storage to improve the efficiency of data queries. A number of access clouds are deployed near end users to cache the hot augmented reality information with the aim of reducing traffic load on our augmented reality system.

Keywords: mobile augmented reality; augmented reality system; object recognition; cloud computing

增强现实(AR)是在虚拟现实(VR)技术基础上发展起来的一种综合了计算机视觉、图形学、图像处理、多传感器技术、显示技术的新兴计算机应用和人机交互技术。增强现实技术利用计算机产生的虚拟物体或文字信息对用户所观察的真实环境进行融合,拓展和增强用户对周围世界的感知。典型的移动增强现实系统如图1所示。

随着 IOS、Android、Windows Mobile 等移动智能终端平台的相继推出以及移动互联网技术的高速发展,曾经局限于实验室的增强现实技术开始走进大众视野,一大批以终端定位、图像识别技术为基础的移动互联网 AR 应用开始涌现,称为移动增强现实应用。在移动增强现实的应用中,用户利用移动终端(譬如智能手机)在拍摄周围景物的同时结合用户终端地理指纹信息(经纬度、朝向等),通过终端上运行的增强现实应用程序捕获增强现实业务提供商部

署的增强现实目标,譬如与地理位置相关兴趣点(POI),并根据用户选择偏好获取相对应的增强现实内容,包括:文字、图片、语音、视频等,将获取的增强现实内容与摄像头拍摄实景在屏幕融合显示,将数字世界和现实世界以用户的上下文(如位置、时间、视线、偏好等)属性有机地结合起来,以自然的方式智能的触发互联网服务,真正实现随时随地个性化信息服务。移动增强现实业务具有实

时交互、虚实融合、全三维的独特性。AR 技术的普及为传统的业务领域拓展出新的交互和展现方式,同时促进移动互联网在教育、游戏、促销和购物、社交网络、商业统计、旅游等业务的创新。

国际上移动增强现实技术研究广泛,诺基亚公司、西门子公司、索尼公司、德国的 Lunatic 公司、IBM 公司、高通公司,以及美国、英国、日本、德国、奥地利、瑞典、新西兰、荷兰、澳

收稿日期: 2015-01-15
网络出版时间: 2015-02-22
基金项目: 国家科技重大专项(2012ZX03002004); 国家高技术研究发展(“863”)计划(2013AA013904)

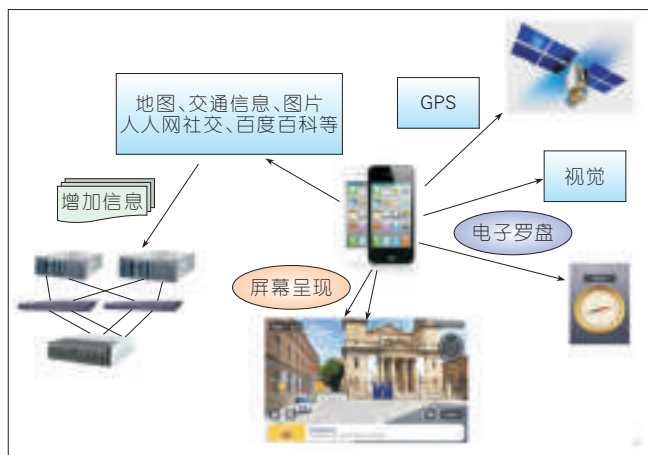


图1
移动增强现实系统

大利亚等国的大学及研究机构都在从事移动增强现实技术的研究工作。国际上移动增强现实市场发展迅猛,荷兰的手机浏览器Layar、奥地利实景导航Wikitude、美国的Yelp iPhone、谷歌公司的Google glass、Nokia公司的Point&Find,都是通过摄像头辨识信息,通过网络得到相关信息,最终达到虚拟和真实的结合。国际一些有实力的增强现实厂商,譬如:Layar、Metaio等,通过发布各自增强现实应用客户端执行环境——增强现实浏览器,为用户提供增强现实Web应用体验,为开发者提供一整套增强现实应用开发工具、应用编程接口(API)和业务部署平台,吸引个人和中小应用开发商^[1-11]。

中国增强现实技术研发和市场推广活动也日趋频繁,中国中小型创业公司(Total immersion、LayAR、触景科技、梦想人)不断涌现,从服务、软件、芯片和商业模式方面不断创新,市场规模不断增长,产业规模正在形成。中兴通讯与高校合作开展相关研究工作,在移动增强现实技术的探索上取得了一定的成绩。中兴通讯借助在IPTV产品、视频监控、会议电视等产品在媒体领域具有的雄厚实力,在增强现实领域的研究处于中国领先水平。代表中国积极推动增强现实技术国际标准制订,并承担了国家重大专项及“863”增强现实领域的专项项目。中兴通讯在云计算领域

具有完整的端到端解决方案,并在中国得到广泛应用;通过领先的增强现实与云计算技术相结合,推出了移动增强现实业务服务平台解决方案。中兴通讯与高校合作研发了中国第一套移动增强现实系统、基于视频透视技术和人工标志物的增强现实技术的博物馆展示系统。

1 移动增强现实关键技术

开发移动增强现实需要解决复杂场景下目标识别技术、实时鲁棒的跟踪注册技术、真实高效的渲染技术、移动终端和云端多核工作站之间不同计算能力设备间的任务协同技术等一些关键技术。

1.1 复杂场景下目标识别技术

增强现实目标识别包括:目标特征点提取(特征点检测、特征点描述)和目标特征点匹配两个阶段。在实现特征提取与匹配时,一般是在图像上搜索特征点,并用强特征描述,使特征点对环境变化如旋转、尺度、光照变换具有一定适应能力。

目前应用于图像检索的计算机视觉特征主要可以分为全局特征和局部特征两类,全局特征主要是指颜色特征和轮廓特征,局部特征主要是指点特征。颜色特征识别率普遍不高但是耗时较短。锥体方向梯度直方图(PHOG)是目前比较常用的轮廓特征。PHOG特征对复杂场景下的识

别正确率低,并且随着目标数目的增加识别率降低比较明显,无法满足海量目标下高识别率的要求。

局部特征主要是指点特征,点特征在保留图像图形重要特征的同时,可以有效地减少信息的数据量,提高信息含量,加快计算速度,相较于全局特征有利于图像的可靠匹配,使得实时处理成为可能。

目前主要的特征点检测算法有两类,一类是基于边缘的特征点检测算法,一类是基于灰度的特征点检测算法。基于边缘的检测算法的基础是要检测出图像的边缘,边缘检测的质量是这类算法检测特征点的基础。这种方法过多的依赖于图像分割及边缘检测的效果,如果在边缘检测的过程中出现问题,对特征点的提取会产生很大的影响,甚至得到错误的特征点,而且图像分割复杂,计算量大,不适于实时处理。而基于灰度的特征点检测算法只依据图像的灰度信息对图像进行特征点检测,而不依赖于其他局部特征,具有算法速度快,更加符合实时性要求,目前比较常用的基于灰度的特征点检测算法有Harris、SUSAN、FAST、SIFT和SURF算法,每种方法都有其特有的一些优势,在实际应用中需要根据具体情况进行选择,下表对各个算法的性能进行列表总结,如表1所示。

最简单的图像识别算法是穷举法或暴力匹配法,这种方法的优点是不需要进行任何的数据预处理,操作也很简单,有效识别率很高,但是其只适用于样本集比较小的情况,其耗时会随着样本图像的增多而增多,在海量样本数据时检索效率比较低。业界普遍采用的方法是词汇树算法来进行海量图像的识别,该算法的主要优点就是其识别时间比较快速,并且不会随着样本数目的增加而发生较大变化。算法分为离线训练和在线检索两个阶段,离线训练阶段存在的问题是训练时系统内存占用较高和运行时间较长,在线检索阶段主要

▼表 1 各特征点提取算法性能比较

算法	准确性	定位性	尺度不变性	旋转不变性	实时性
Harris	一般	L型一很好,其他一较差	不支持	一般	一般
SUSAN	一般	较差(模糊图像时)	最好	一般	较好
FAST	很差	一般	一般	不支持	最好
SIFT	很好	最好	较好	一般	较差
SURF	最好	较好	一般	最好	一般

的问题是其查全率不高,需要对其进行优化调优。

1.2 实时鲁棒的跟踪注册技术

跟踪注册是将计算生成的虚拟物体和真实环境中景象“对齐”的过程,主要通过计算摄像机和目标物体的相对位置和姿态来实现的。现有跟踪注册技术主要分为硬件跟踪注册技术和视觉跟踪注册技术两类,硬件跟踪技术由于传感器的精度低、抖动和漂移等问题,无法实现稳定、精确的跟踪注册结果,而视觉跟踪注册技术又由于计算量大而无法满足用户实时性的问题,如何获得实时、稳定、精确的跟踪注册效果是研究增强现实技术的难点之一。采用混合跟踪注册技术,以传感器获得的数据作为视觉跟踪注册算法初始值,通过迭代使得视觉跟踪注册算法结果快速收敛,对传感器获得的数据进行修正,获得实时、稳定、精确的跟踪注册效果。

目前视觉跟踪注册采用无标记跟踪定位方法,比较常用的是基于场景平面的增强现实跟踪注册和基于模型和关键帧的注册方法。基于平面结构的跟踪注册算法是早期研究人员提出的一种有效的实时无标记增强现实跟踪注册算法,具有算法初始化过程简单,实时性较好,包括给定区域内的特征点提取、匹配、单应矩阵的计算等。但是该算法鲁棒性差。基于模型和关键帧的注册方法利用场景的离线和在线信息完成增强现实的三维实时跟踪注册,对相机的视点变化,环境光照以及物体的遮挡等都保持了较好的鲁棒性,是一种

较为优秀的增强现实跟踪注册算法。

1.3 真实高效的渲染技术

移动终端上的图像渲染是相对独立的技术。利用渲染技术对获取的虚拟的3D模型、图像、视频、超文本文件增强内容进行展示,体现出虚拟场景和现实场景的无缝融合效果。图形渲染引擎以移动终端的摄像头视频流作为背景画面,把虚拟的增强内容渲染叠加在摄像头视频上对应的位置。针对增强内容的渲染技术,可采用跨平台的开源 OpenGL ES 技术,同时针对移动终端的特点,做简化操作,简化冗余的传输数据,优化渲染流程,设置简单基本图元渲染,设计定点模型数据结构,改进帧缓存操作流程,提高渲染的效率。对于复杂3D模型或者3D动画等增强内容,可采用专业的渲染引擎来处理,如 Unity 3D、Ogre3D等。

1.4 任务协同技术

对于移动终端和云端多核工作站之间不同计算能力设备间的任务协同技术。增强现实的效果展示需要进行大量的复杂运算,包括对用户位置的跟踪、渲染、绘制等,这些复杂的运算任务要在移动终端和系统的服务器中进行协调,把一部分工作放在手机端由手机实现,而将另外一部分复杂运算放在服务器中实现。高性能多核图形工作站的运算与处理能力远远优于移动设备,同时3G/4G网络的高效数据传输使得大数据量的网络实时通信成为可能,移动设备与高性能工作站的高效协同工作集成环境是系统设计的关键。采用客

户/服务器(C/S)架构实现移动设备与服务器之间的优势互补。具有大运算量的初始化、标定、三维建模等工作将由具有多核服务器完成,构建一个基于云计算技术的支持大规模图像训练和快速检索和高并发的增强内容访问服务的综合服务平台,以达到系统高并发、高可用、高扩展的特性;而图像采集、特征提取、虚实渲染、实时注册等工作将由移动设备来完成。

综上,移动增强现实关键技术研究集中在:云端对终端对象的精确识别;终端对场景对象快速捕捉、注册跟踪;最终在终端显示设备上高效地虚实渲染呈现出来。

2 移动增强现实云服务系统方案

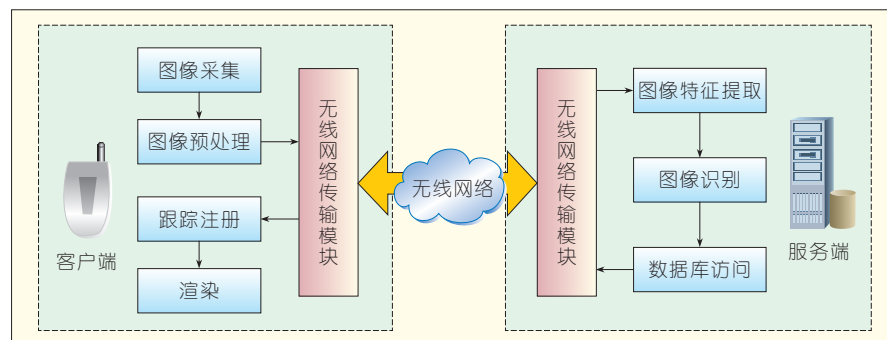
2.1 移动增强现实系统构成

典型的移动增强现实系统如图2所示。基于计算机视觉的C/S架构移动增强现实系统由图像采集、图像预处理、无线网络传输、特征提取、图像识别、跟踪注册、渲染等模块组成,系统通过摄像头获取图像,并进行预处理。为了减少网络传输流量对图像进行压缩,经过通信模块上传到服务端。服务端进行特征提取和图像识别。服务端将根据识别结果返回的目标标识,查询获得相应的增强内容,通过通信模块将目标标识、增强内容返回给客户端,客户端利用跟踪注册模块得到虚拟信息叠加到真实场景中的动态位置,通过渲染模块使叠加虚拟的增强内容信息到客户端摄像头实时界面上。

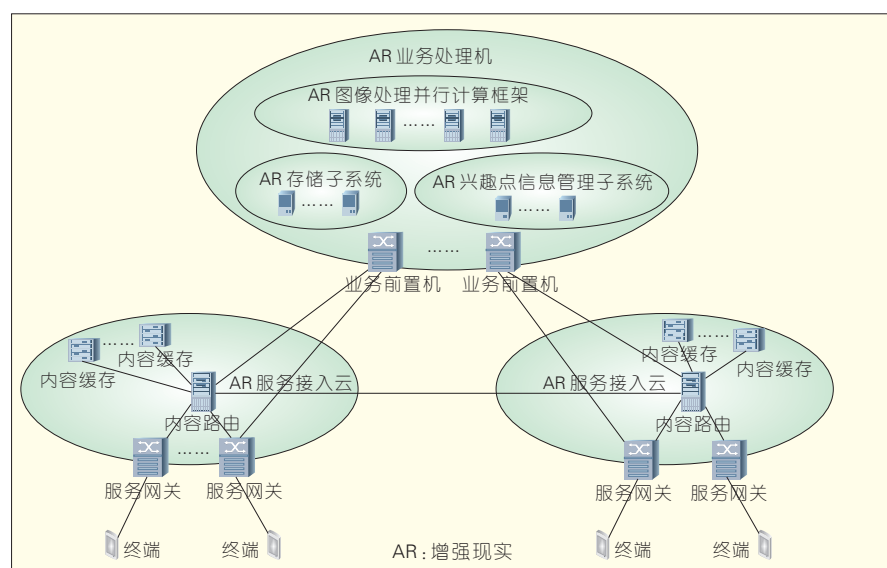
2.2 中兴通讯移动增强现实服务云平台系统架构

中兴通讯移动增强现实服务云平台组网模型如图3所示,实现方法如下:

(1)增强目标的识别技术充分利用云的海量存储和计算能力,进行图



▲ 图2 移动增强现实系统



▲ 图3 中兴通讯移动增强现实服务云平台组网模型

像识别的样本生成和训练,从而实现快速的目标识别。将目标识别的关键算法当作应用进行加载,并可以对算法的实例和参数进行灵活控制。

(2) 针对海量增强现实信息和训练样本数据高效存储的需求,分别研发增强现实存储子系统和增强现实兴趣点信息管理系统。

(3) 针对大规模的图像处理,采用基于 Map/Reduce 的并行计算框架,提高样本训练和目标识别的效率。

(4) 为了分担增强现实业务请求负载、缓存热点数据、提升用户体验,将在靠近用户的地方部署若干用户接入云。该服务接入云将接收用户的各类请求,缓存热点的增强现实信息,减轻增强现实业务处理机的负载压力。

(5) 针对直接影响 AR 用户体验的定位技术,采用组合定位的技术方案提供更加高效的定位技术,将基站定位、WLAN 定位、GPS 定位技术相结合,充分利用服务端的信息存储和计算能力,提高定位的速度和精度。

中兴通讯移动增强现实系统架构如图 4 所示。中兴通讯移动增强现实应用平台架构由 4 部分构成:移动增强现实客户端(MobAR-C)、移动增强现实综合服务平台(MobAR-S)、移动增强现实能力(MobAR-E)、移动增强现实运营管理平台(MobAR-M)。

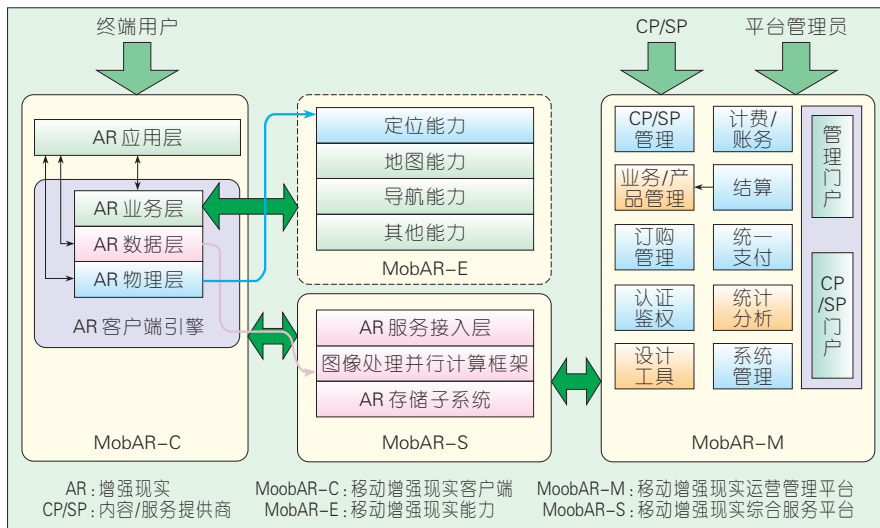
MobAR-C AR 客户端引擎为 AR 应用提供了统一的运行环境,负责与移动终端硬件的交互,完成目标的图像获取,配合服务端进行目标识别,

增强内容解析和渲染,并对上层应用提供 API 调用接口。AR 客户端引擎分为 4 个层次,即物理层、数据层、业务层和应用层。智能终端通过 AR 物理层获取各类运动传感器提供的运动数据、GPS 坐标、以及摄像头采集到的视频,并将其上传至数据层。数据层是增强现实系统的关键层次,包含数据存储及图像处理(包含特征提取、跟踪注册、增强渲染)等模块。AR 客户端业务层主要对客户端引擎提供的包含增强现实服务、定位服务、地图服务等各种服务能力的进行能力封装提供外部调用方法。

MobAR-M 为内容/服务提供商(CP/SP)提供了业务注册和产品发布以及认证鉴权功能,并且为用户提供了 AR 应用的发现渠道,用户可查找 AR 应用并购买、下载和使用。

MobAR-E 为 AR 应用提供特殊的能力平台,比如 GIS 地图、导航、定位等等。

MobAR-S 基于云计算的服务架构,分别提供 AR 服务就近接入,云存储和图像并行计算服务。AR 业务服务端利用并行计算解决大规模快速离线图像训练和在线识别问题;设计增强现实图像并行处理框架,该框架采用分布式架构,基于 Map/Reduce 计算方式,并行地实现图像检索、样本训练等增强现实业务,以达到系统高并发、高可用、高扩展的特性。在图像特征抽取上:首先利用预先训练出来的融合多种影响图像检索因素的特征点筛选模型进行特征点选择,然后通过主成分分析(PCA)降维方法获得较低维度的特征,最后通过预先训练出来的高斯混合模型(GMM)根据不同的要求生成不同大小的费舍尔向量,并对其进行二值化得到描述子的最终紧凑表示可伸缩压缩费舍尔向量(SCFV)。在图像检索上:采用比较简单的线性查找的方法进行图像检索,该方法更加有利于进行分布式处理,内存占用较低,并且支持增量训练,云端新加入的参考图像可



▲ 图4 中兴通讯移动增强现实系统架构



▲ 图5 移动增强现实示范“房产360”应用

以快速的加入到索引中,可以快速的应用于云端目标识别系统中。利用云存储,解决AR信息的非结构化和结构化存储问题,对海量AR信息元数据管理以及快速索引。在增强现实业务应用环境中,将会存储海量的增强现实信息数据,以及对这些数据快速检索。采用分布式的非结构化存储机制以保存图像等信息,采用结构化存储机制以保存各种增强现实元数据信息;通过在结构化存储机制中建立特征索引等快速查询机制,以达到高效数据查询等目的。

中兴通讯基于此增强现实服务云平台,设计并实现面向移动智能终端的增强现实典型示范应用“房产360”,如图5所示。通过实际应用,迭代验证相关理论及关键技术、逐步完善系统架构和关键技术。房产360应用基于移动定位、可视化搜索等增强现实技术,充分利用手机图像采集、GPS、电子罗盘等设备的感知能

力,为用户提供周边房产搜索导航、租售价联系电话等房产详情、用户评价等全方位的房源信息。房产360应用信息丰富,能够很好帮助有租售房屋需求的客户,系统图像识别快速准确,叠加信息虚实融合效果显著。

3 结束语

尽管移动增强现实系统已经显示出了在诸多领域中广泛的应用前景,但是由于技术条件和成本的限制,和大规模的推广应用尚有一定的距离。一方面增强现实中的目标识别与跟踪注册关键技术的精确性和实时性还有很大的提升空间,并且受光照遮挡等复杂环境因素影响非常大。另一方面,增强现实需要展示大量的3D模型,内容制作成本巨大,这些不可控的因素都制约了移动增强现实技术的应用。

参考文献

- [1] SOUPTIK D. Approximate Distributed K-

- Means Clustering over a Peer-to-Peer Network [J]. IEEE, 2009,21(10):2351-2368
- [2] YAN T X. Distributed Image Search in Camera Sensor Networks [D]. Department of Computer Science University of Massachusetts, Amherst, MA 01003
- [3] ASSAD M, ARMICHAEL D J, CUTTING D. AR phone: Accessible augmented reality in the intelligent environment [C]//Proceedings of the OZCHI, 2003
- [4] NISTER D, STEWENIUS H. Scalable Recognition with a Vocabulary Tree. Computer Vision and Pattern Recognition [C]//Proceedings of the 2006 IEEE Computer Society Conference, 2006
- [5] JEFFREY D, SANJAY G. MapReduce: Simplified Data Processing on Large Clusters [C]//Proceedings of the OSDI, 2004
- [6] PAPAOGLOU M P, TRAVERSO P, DUSTDAR S, et al. Service-oriented Computing: State of the Art and Research Challenges. IEEE Computer (COMPUTER), 2007,45(6):2066-2079
- [7] ROBISON S. The Next Wave: Everything as a Service [EB/OL]. [2015-01-10]. <http://www.hp.com/hpinfo/execteam/articles/robison/08eaas.html>
- [8] WEI Y, BLAKE M B. Service-oriented Computing and Cloud Computing: Challenges and Opportunities [J]. IEEE Internet Computing, 2010,22(2):1096-1109
- [9] 陈靖, 王涌天, 郭俊伟, 等. 基于特征识别的增强现实跟踪定位算法 [J]. 中国科学:信息科学, 2010,40(11):1437-1449
- [10] 杜振鹏, 李德华. 基于KD-Tree搜索和SURF特征的图像匹配算法研究 [J]. 计算机与数字工程, 2012,40(2):96-98+126
- [11] 刘红光, 魏小敏. Bag of Words 算法框架的研究 [J]. 舰船电子工程, 2011,31(9):125-128

作者简介



吕强,中兴通讯股份有限公司云计算及IT研究院移动增强现实项目经理;主要研究方向为增强现实、定位业务及应用。



黄成,中兴通讯股份有限公司云计算及IT研究院多媒体系统标准项目经理;主要研究方向为可视化搜索、增强现实、OTT视频与流媒体。



刘明,中兴通讯股份有限公司云计算及IT研究院工程师;主要研究方向为智能媒体分析、模式识别等相关技术。

网络功能虚拟化及其标准化

Network Function Virtualization and Standardization

薛海强/XUE Haiqiang
张昊/ZHANG Hao

(中国移动研究院, 北京 100053)
(China Mobile Research Institute, Beijing
100053, China)

电信网络依赖的技术与互联网基
本一致, 所提供业务也有一定
相似性。近年互联网快速发展, 能够
提供传统电信网所提供的业务, 对电
信网带来前所未有的挑战。技术方
面, 互联网的技术对电信行业也有较
强的借鉴意义, 云计算能够在IT业取
得巨大成功, 可以应用到电信领域
中, 优化电信网络的业务提供方式和
网络运维模式。2012年10月, 中国
移动、AT&T、英国电信、NTT等13家
运营商联合发布了网络功能虚拟化
(NFV)白皮书, 提出了电信网络与云
计算技术结合的构想和要求^[1]。

1 NFV介绍

2013年1月, 欧洲电信标准组织
(ETSI)正式成立NFV ISG工作组, 下
设基础设施、软件架构、管理和维护、
可靠性、性能、安全6个工作子组。
2013年10月发布了第一个版本,
2014年底发布第二个版本。

1.1 NFV是虚拟化技术与电信网络 的结合

NFV ISG认为电信网络设备, 包
括软交换、服务通用分组无线业务支

收稿日期: 2015-01-10
网络出版时间: 2015-02-22

中图分类号: TN929.5; TP393 文献标志码: A 文章编号: 1009-6868 (2015) 02-0030-005

摘要: 基于欧洲电信标准组织(ETSI)网络功能虚拟化(NFV)架构、功能, 从运营商
的视角分析了运营商引入NFV的切入点以及多种虚拟化技术路线, 并给出了技术路
线选择和演进方向建议。NFV在引入了灵活性的同时也带来了管理、运维方面的问
题和挑战, 通过抛砖引玉的方式对NFV面临的问题和挑战进行了分析, 期望推进相
关问题的解决, 加速NFV的商用部署。

关键词: 网络功能虚拟化; 云计算; 高可靠性

Abstract: Network function virtualization is the direction of development for
telecommunication technology. Based on the introduction about architecture, function for
ETSI NFV, this paper analyzes in detail the breakthrough point of NFV and
technology route from the point of operator and gives some suggestion about the
choice of technology route and evolution direction. NFV brings some issues and
challenges to the network management and maintenance in addition to flexibility.
This paper suggests industry should work together to solve this issue to accelerate
the commercial deployment of NFV in the mobile network by analyzing questions and
challenges.

Key words: network function virtualization; cloud computing; high reliability

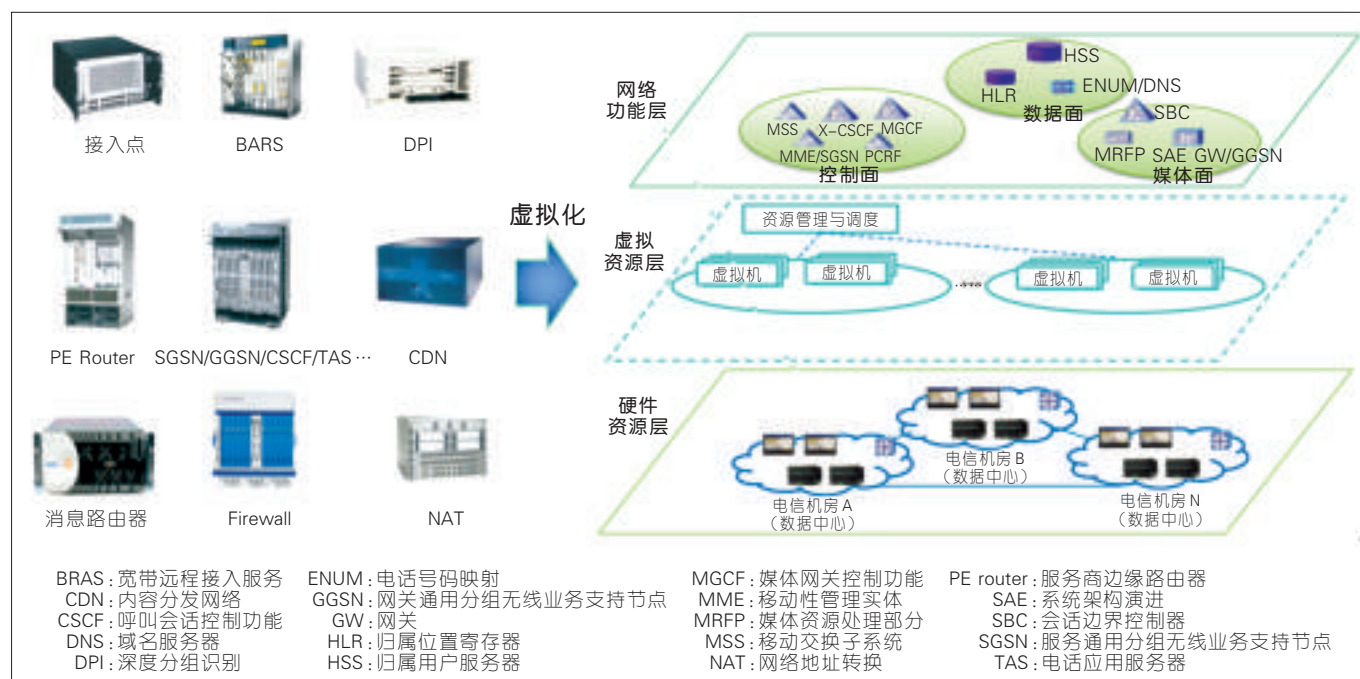
持节点(SGSN)、网关通用分组无线
业务支持节点(GGSN)、呼叫会话控
制功能(CSCF)、归属用户服务器
(HSS)、接入设备、防火墙、深度分组
识别(DPI)、宽带远程接入服务
(BRAS)、PE路由器、内容分发网络
(CDN)设备等, 可以演进为虚拟功能
网元(VNF), 部署在通用硬件设备
上。整个网络可以分为硬件层、虚拟
资源层、网络功能层3层, 硬件可以
部署在传统电信机房, 也可以部署在
数据中心, 如图1所示。

网络功能虚拟化并非简单的在
设备中增加虚拟机, 其中重要特征在
于引入虚拟化层之后, 虚拟功能网元
(VNF)与硬件完全解耦, 改变了电信
领域软件、硬件绑定的设备提供模
式。虚拟机对上层应用屏蔽硬件的

差异, 虚拟功能网元可以部署在虚拟
机上, 进而允许运营商对电信系统的
硬件资源实行统一管理和调度, 能够
大幅提升电信网络的灵活性、缩短业
务的部署和推出时间、提升资源的使
用效率。同时, 网络功能虚拟化之
后, 电信设备演进为虚拟功能网元,
这些网元的开发和实现将不再依赖
于特定的硬件平台, 可以降低电信设
备(虚拟功能网元)的开发门槛, 促进
电信设备制造产业链的开放, 加速新
业务的推出。

1.2 网络功能虚拟化架构介绍

众所周知, 电信网络对设备的可
靠性、性能有严格的要求, 对设备的
可维护性要求也比较高。网络功能
虚拟化之后, 对于硬件资源、虚拟资



▲ 图1 NFV网络

源、虚拟功能网元如何进行有效管理,是电信网络重点关注的内容。ETSI NFV ISG 定义了端到端的架构,描述了网络功能虚拟化的架构,如图2所示^[2]。

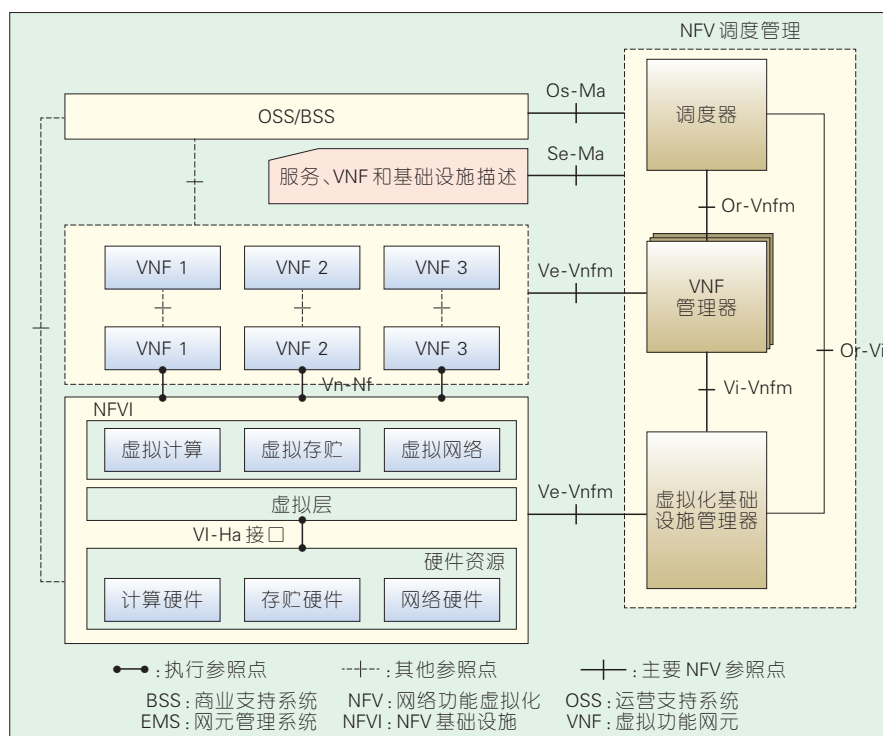
网络功能虚拟化(NFV)架构中包括硬件资源、虚拟资源、虚拟功能网元、运营支持系统/商业支持系统(OSS/BSS)、虚拟化基础设施管理器(VIM)、VNF 管理器(VNFM)、NFV 调度器(NFVO)。

硬件资源:分为计算资源,存储资源和网络资源3部分,计算资源是指本地通用物理服务器,通用物理服务器包含CPU、内存、本地磁盘和网卡等,也可以包含加速的硬件(如硬件加解密、包交换、包转发加速)。存储资源是指外接用于存储的磁盘阵列、或者分布式存储。网络资源通常是指交换机和路由器等网络通信连接设备。

虚拟资源:虚拟资源主要体现形式为虚拟机。虚拟机包含虚拟计算资源,如虚拟CPU(vCPU),虚拟存储资源如虚拟内存,虚拟磁盘,以及虚拟网络资源,如虚拟网卡等。虚拟机

可以有不同规格,虚拟机规格由资源模板描述,虚拟机规格可配置、可管理。虚拟机由虚拟机管理器(Hypervisor)在硬件资源中的通用物理服务器上提供,Hypervisor 将通用

物理服务器与上层软件应用分开,多个虚拟机可以在同一个物理服务器上运行,最大化的利用硬件资源,即一个物理服务器的硬件资源可以被多个虚拟机共享。Hypervisor 可以与



▲ 图2 网络功能虚拟化架构

云管理系统交互实现对虚拟机的创建、删除等操作以及故障管理、性能管理等功能。

虚拟网元功能:虚拟网元功能是传统电信设备在网络功能虚拟系统中的展现形式,VNF可部署在一个或多个虚拟机(VM)上,提供电信系统所需要的功能要求,VNF所提供的网元功能与非虚拟化时的网元功能应保持一致,与其他网元实体的接口与非虚拟化时的接口应保持一致。

EMS:EMS通过北向接口与网管系统相连,提供配置管理、告警管理和性能管理等功能。

VIM:负责虚拟化基础设施管理,主要功能是实现对整个基础设施层资源的管理和监控。包括硬件资源的管理和虚拟资源的管理两大类。

硬件资源的管理:配置并管理机框等设备,监控机框电源、风扇等关键部件状态;配置并管理路由器、交换机,防火墙和负载均衡器等设备,包括添加、删除、更改和查询路由器、交换机、防火墙和负载均衡器等设备信息;监控路由器、交换机、防火墙和负载均衡器的运行状态及使用情况;自动或手动识别、配置并管理物理服务器,包括添加、删除、更改和查询物理服务器等设备信息;监控物理服务器CPU、内存、磁盘以及网卡等关键部件的状态及CPU利用率、内存利用率、网络入口带宽、网络出口带宽、磁盘读取速率、磁盘写入速率、CPU温度等信息;接入并管理外接磁盘阵列(包括IP-SAN等),包括增加、删除、更改和查询外接磁盘阵列等设备信息;监控磁盘的状态及容量使用情况;采集硬件资源的告警信息,并能够上报到NFVO。

虚拟资源的管理:配置并管理虚拟机,包括虚拟机的创建、删除和查询等;虚拟机镜像文件的管理,包括添加、修改、删除和查询等;监控虚拟机的运行状态,以及虚拟机的vCPU占用率,虚拟内存使用率,虚拟磁盘占用率,虚拟网卡的吞吐率等;VIM

可选支持虚拟机迁移;支持采集虚拟资源告警信息后能够上报到NFVO。

VNFM:负责VNF实例的管理,VNFM包括以下功能:

VNF实例的生命周期管理,包括实例化、删除、查询、扩容/缩容、终结等,VNFM应提供基于业务容量模型的VNF自动部署和手动部署能力,能够自动或手动完成VNF的实例化。VNFM支持VNF软件包管理,VNF软件包包括VNFD、GUEST OS镜像文件以及VNF软件镜像文件。VNF包管理包括VNF包的上载,更新和删除。

VNFM应根据VNF的资源利用情况,发起扩容/缩容等操作;VNF所用虚拟资源,以及虚拟资源的性能数据/事件的采集;VNF业务所使用虚拟机故障信息的采集。

NFVO:网元功能虚拟化协调,负责提供硬件资源和虚拟资源的视图,硬件资源和虚拟资源的监控,性能统计和故障管理,并控制VNFM实现VNF软件包的管理,以及VNF实例的创建,更新,终止和弹性伸缩。提供管理接口供操作员进行云管理系统的本地维护。支持网络服务的加载、部署,并可与OSS协同,完成对网络服务的管理。

1.3 网络功能虚拟化的应用领域

ETSI NFV ISG在定义网络功能虚拟化的同时,也梳理了网络功能虚拟化应用9个重点应用场景^[3]。

● 场景1: NFV基础设施即服务(NFVIaaS),允许服务提供商在提供给终端用户的独立管理的NFV基础设施(NFVI)上进行服务的提供保证及收费。

● 场景2: VNF软件即服务(VNFaaS),VNF通常被认为在网络运营商的私有云模型上执行,VNFaaS能够提供远程网络功能。

● 场景3: 虚拟网络平台即服务(VNPaaS),VNPaaS模型提供给用户更宽泛的控制,用户能够进行多个VNF实例的配置。

● 场景4: VNF转发图(VNFFG),VNFFG在VNF之间提供了逻辑连接,可定义数据包的传输路径。

● 场景5: 移动核心网的虚拟化,能够降低网络复杂性,减少费用,并提供更高的网络使用率及服务质量,主要虚拟化目标是EPC网络功能和IMS网络功能的虚拟化。

● 场景6: 移动通信基站的虚拟化,将部分的无线接入网(RAN)节点虚拟化,采用标准IT硬件设备,主要虚拟化目标是传统RAN节点,能够获得更(低的能源消耗,更便捷的管理操作,更快的投入市场。

● 场景7: 家庭环境的虚拟化,会大量减少现今家庭服务中的后端系统设备,只需要简单的物理连接,带来更低的设备成本,主要虚拟化目标是家庭网关(RGW)和机顶盒(STB)。

● 场景8: CDN的虚拟化(vCDN),现今的CDN缓存节点通常是专用物理设备,这会带来很多弊端,虚拟化的目标是CDN的所有组件,但会优先考虑CDN缓存节点。

● 场景9: 固定接入网络功能的虚拟化,会首先应用到光纤到交换箱(FTTcab)、光纤到分配点(FTTdp)这类混合DSL节点,最终会形成一个单独的平台,来服务不同的应用,用户和租户。

以上9个方面应用中,前3个是分别映射到云计算领域的基础设施即服务(IaaS)、平台即服务(PaaS)、软件即服务(SaaS),后续的几个应用场景分别对应着电信网络的IMS核心网、分组域核心网、流量处理、接入网领域,目前IMS核心网、分组域核心网是业界虚拟化研究的重点。下面将介绍网络功能虚拟化在移动核心网中的应用。

2 移动核心网虚拟化

移动核心网通常包括电路域(CS)、分组域(PS)、IMS域3类。电路域主要承载传统的语音业务,是2G、3G网络时代的主要业务支撑实

体,但是随着4G业务的发展,传统的语音业务将逐步萎缩或者最终退出服务。移动网络的分组网在2G,3G时代由于无线接入能力有限;在4G时代,由于无线接入能力大幅提升,将面临着新的发展机遇。IMS业务是传统语音业务的升级,融合通信可以大幅提升传统电信业务的用户体验,IMS域同样面临着发展机遇。

2.1 移动核心网虚拟化分析

根据设备的功能特征,核心网设备可以分为控制面设备、用户数据设备、媒体转发面设备。控制面设备包含CSCF、电话应用服务器(TAS)、MME、SCP、MSC等,主要处理信令消息、控制呼叫和业务过程;用户数据面设备主要包含HSS、PCRF,主要存储用户的业务数据、签约数据、认证和鉴权数据,对于数据的可靠性、读写速度要求较高;媒体转发面设备主要包括会话边界控制器(SBC)、P/S-GW、MGW等,主要处理数据包的转发、编解码转换、数据流的加密等,对数据包的转发性能要求较高。

网络功能虚拟化是指在传统的电信设备中,引入了虚拟机,实现电信应用与硬件的解耦,由于引入虚拟机,不可避免的带来了一定的性能损失。由于电信网络对性能、可靠性的要求较高,这对虚拟网元的性能实现带来了较大的挑战。虚拟化之后,如果不加优化,虚拟机的I/O处理能力下降比较严重,难以满足媒体转发面大量包处理的要求。针对这种情况Intel等厂商提出了单根IO虚拟化技术(SRIOV)、数据平台开发套件(DPDK)技术优化虚拟机的I/O处理性能,取得了显著的成果,但主流厂商的商用产品尚未成熟,同时对于数据流的加解密处理,相比传统的实现方式,仍需提升性价比。用户数据面设备有较多数据读写操作,通常这部分功能依赖于少数厂商提供的高性能数据库和磁盘阵列,这些年厂商在去IOE方面做了许多工作;在HSS中,大

多数厂商已经能够做到不依赖于专用数据库,但数据库虚拟化之后的性能、稳定性、可靠性仍有待验证。相对而言,控制面设备主要是控制用户业务的状态机,业界现有的虚拟化技术经过优化之后基本能够满足要求,厂商的产品成熟度也比较高。综合上述分析,移动网络的网络功能虚拟化可以优先从IMS控制面开始,逐步扩展到用户数据面和分组域。

2.2 移动核心网虚拟化的技术路线分析

网络功能虚拟化使得电信设备的网络功能软件和硬件解耦,运营商第一次可以实现硬件资源和网元功能的分开采购,对硬件资源、虚拟资源可以统一管理和调度,赋予运营商前所未有的灵活性。但是,我们也应该看到,传统电信设备的高可靠性、高性能是由电信厂商保证的。网络功能虚拟化之后,谁来保证电信设备的高可靠性、高性能,系统故障之后,如何做故障定位,将是运营商面临的重大挑战。

对于电信运营商来说,虽然从技术上来说,硬件层、虚拟资源层、虚拟功能网元层可以解耦,但从网络实施的角度来看,是否要实现三层的完全解耦需根据实际情况分析。从实现的层面看,存在4种实现方案。

第一种方案,硬件、虚拟软件、虚拟功能网元由同一厂家提供,其特点在于软件、硬件均由同一厂商提供,与传统电信设备的提供形式相同,由厂商负责系统的可靠性和性能,最大程度上降低了运营商的风险和实施单独;仅在单厂商系统内部引入了虚拟化资源层,可以实现单厂商内部的资源共享和协同调度;但这种方案本质上仍然是封闭的系统,不利于形成开放的产业链,也不利于充分发挥虚拟化的优势。第二种方案,将硬件和虚拟资源软件由一个厂家提供,虚拟网元功能软件由另外一个厂家提供;此方案能够实现虚拟资源的共享,系

统调度灵活,能够发挥虚拟化的优势,形成开放的产业链;但这种方案将虚拟功能网元和虚拟资源解耦,势必会带来虚拟功能网元的故障定位和性能优化问题,由于虚拟功能网元和虚拟资源层分别由不同的厂商提供,如果系统出现故障,如何协调两个厂商,快速定位并解决问题,将是运营商面临的巨大挑战。第三种方案是将虚拟资源层和虚拟功能网元层打包在一起,由同一厂商提供,硬件由运营商选择其他厂商提供;由于对硬件的故障定位相对于虚拟机要简单,此方案的故障定位和性能优化问题将大幅简化,对运营商的技术要求相对低一些,在现阶段可行性更高。第四种方案是将硬件、虚拟资源层、虚拟功能网元完全解耦,可以看作是第二种和第三种方案的叠加;此方案可以获得最大限度的灵活性,可以对硬件资源、虚拟资源实现快速的调度和灵活管理,同时也带来最复杂的故障定位和性能优化问题,此方面业界有一些运营商在尝试和探索,但距离商用尚有距离。

综合来看,网络功能虚拟化的技术方案各有优缺点,不同的运营商根据自身情况可能会有不同的选择。从作者的观点来看,现阶段第三方案有比较高的可行性,同时,选择第三种方案之后,如果业界虚拟化技术进一步成熟,故障定位和性能优化问题完善之后,也容易向三层完全解耦的方式演进。

2.3 网络功能虚拟化的管理和维护问题

网络功能虚拟化改变了传统电信设备的设备形态,实现了硬件和软件的完全解耦;同时,这也对运营商的运维和管理体系带来了新的挑战。下面将从故障定位、网管和资管系统、采购和建设模式方面说明。

传统电信设备的故障定位在单台设备范围内进行,厂商为设备的性能、可靠性、业务指标负责,故障定位

简单。虚拟化之后,故障定位将涉及到虚拟功能网元、虚拟层、硬件层,尽管针对不同的技术路线,会有所区别,但故障定位都不在局限于一台物理设备范围之内,对运维人员的技术能力提出了更高的要求。

传统的设备的网管系统和资管系统通常合并在一起,由OSS统一管理。虚拟化之后,设备将运行在虚拟机之上,网管系统将不能直接看到硬件设备的信息,网管系统是否需要修改?该如何修改?硬件与虚拟网元功能解耦之后,硬件可以统一管理,但如何统一管理?归属传统网管系统?还是和数据中心的基础设施统一管理?这都是有待结合运营商的具体网络情况详细分析,制订相关策略。

在网络规划和建设方面,运营商通常根据网络的业务的发展情况,采购对应的网络设备提供业务;虚拟化之后,硬件资源、虚拟资源和虚拟功能网元解耦,硬件资源、虚拟资源可以预先统一采购和建设,业务需要时候,直接从现有资源中调配相应资源供虚拟功能网元使用,便于业务的快速部署和建设,但这将对运营商的现有采购流程和采购模式产生影响,运营商需探索新的建设和采购模式。

3 NFV 标准化现状

网络功能虚拟化无疑是近期电信领域的热点,除了上文中介绍的ETSI NFV ISG在开展相关的研究和

标准化工作之外。在2014年5月的3GPP SA5会议上^[4],中国的研究课题,得到27个运营商和厂商积极支持,目前正在引导业界研究虚拟化网络的网管问题,预计2015年6月完成研究工作,后续则会转入标准化工作阶段。

在电信管理论坛(TMF)也成立了ZOO的项目,研究虚拟化网络的网管问题,对于网络服务的组合与调度涉及较多。在国际电信联盟(ITU)也有相关的研究课题研究网络功能虚拟化。

在标准化工作开展的同时,2014年9月30日,中国移动、AT&T等运营商与Linux Foundation合作,联合发起成立了NFV开放平台项目(OPNFV)的开源组织,旨在通过开源组织的力量,开发符合NFV需求的虚拟资源层软件,预计2016年中期将有第一个版本推出。

4 结束语

网络功能虚拟化能够将云计算技术引入到电信网络中来,能够大幅提升网络的灵活性,有利于新业务的开发和部署,能够提升网络的管理和维护效率。从2012年业界提出开始,就受到了广泛的关注,成为电信领域的研究热点。目前,经过两年的研究,网络功能虚拟化已经成为电信网络的发展方向,业界的主流设备制造商也开发出了相关的设备,主流运营商都已经开展了相关的测试和试

点验证工作。尽管目前网络功能虚拟化的标准尚未成熟,还面临着故障定位、管理维护等方面的挑战,随着厂商和运营商的持续投入和推动,预计在近两年之内将有大规模的商用,后续将进入快速发展阶段。

参考文献

- [1] NFV White Paper [EB/OL]. [2015-01-3]. <http://portal.etsi.org>
- [2] Network Function Virtualization (NFV) Management and Orchestration [EB/OL]. [2015-01-3]. <http://portal.etsi.org>
- [3] Network Function Virtualization (NFV) Use Cases [EB/OL]. [2015-01-3]. <http://portal.etsi.org>
- [4] Study on Network Management of Virtualized Networks [EB/OL]. [2015-01-3]. <http://portal.3gpp.org>

作者简介



薛海强,中国移动研究院资深研究员、博士;研究领域为网络功能虚拟化、云计算、大数据应用、核心网架构、IMS网络、车联网等。



张昊,中国移动研究院研究员;研究领域为网络功能虚拟化、云计算、大数据应用、核心网架构、IMS网络、车联网等。

综合信息

传感器应用领域不断扩大 迎来大好发展

随着技术研发的持续深入,成本的下降,性能和可靠性的提升,在物联网、移动互联网和高端装备制造快速发展的推动下,传感器的典型应用市场发展迅速。亚太地区将成为最有潜力的市场。

英泰诺咨询公司指出,未来几年亚太地区市场份额将持续增长,预计2016年将提高至38.1%,北美和西

欧市场份额将略有下降。

交通、信息通信成为市场增长最快的领域。据英泰诺咨询公司预测,2016年全球汽车传感器规模可达419.7亿欧元,占全球市场的22.8%;信息通信行业至2016年也可达421.6亿欧元,占全球市场的22.9%,且有可能成为最大的单一应用市场。

(转载自《中国信息产业网》)

云计算安全框架分析

Cloud Security Architecture

陈清金/CHEN Qingjin
陈存香/CHEN Cunxiang
李晓宇/LI Xiaoyu

(中国联通云数据有限公司, 北京 100048)
(China Unicom Cloud Data Co Ltd, Beijing
100048, China)

中图分类号: TN929.5; TP393 文献标志码: A 文章编号: 1009-6868 (2015) 02-0035-004

摘要: 结合已有的安全架构, 提出了一种以用户域和服务域两个层面为关注角度、以软件即服务(SaaS)、平台即服务(PaaS)、基础设施即服务(IaaS)3种服务模式为基准的通用云安全架构。该架构不同层面及不同服务模式中的安全侧重点有所不同, 云服务提供者和云服务使用者可根据自身需要进行定制化开发或使用。针对重要的安全服务, 进行了进一步的细化以便保证安全架构的完整性。

关键词: 云计算; 安全; 架构; 服务模式

Abstract: We present a general cloud security architecture based on SaaS, PaaS, and IaaS service models. It has a user domain and service domain, both of which focus on different areas. Cloud users and providers can customize their options according to their own needs. To ensure the integrity of the security architecture, we refine the security service further.

Keywords: cloud computing; security; architecture; service model optimization

云计算是并行计算、网格计算、效用计算、分布式计算等传统计算机技术和网络技术发展融合的产物。对于这种新兴的计算模型, 目前国际上还没有公认的定义。美国国家标准与技术研究院(NIST)^[1]认为云是一种由一群互联的虚拟化计算机组成的并行分布式系统, 基于服务提供者和消费者间的服务契约动态提供统一的计算资源。维基百科^[2]认为云计算是一种将动态伸缩的虚拟化资源通过互联网以服务的方式提供给用户的计算模式, 用户不需要知道如何管理那些支持云计算的基础设施。中国网格计算与云计算专家刘鹏定义云计算^[3]为将计算任务分布在由大量计算机构成的资源池中, 使各种应用系统能够根据需要获取计算能力、存储空间和各种软件服务。虽然定义有所差异, 但是核心思想都是通过将多个成本低廉的计算实体整合成一个具有强大计算能力的系统, 网络用户借助于SaaS、PaaS、IaaS^[4-5]等服务模式便可获得较强大的计算和存储能力。便利、廉价、灵活的服务促使云计算用户和

服务内容的迅猛发展, 但同时云计算的开放性和动态性也使得云计算的安全问题引起了多方的关注。

2010年著名调研机构IDC^[6]发布了报告《New IDC IT Cloud Services Survey: Top Benefits and Challenges》。报告分析了云计算面临的主要挑战以及用户的关注点, 其中安全问题是各界人士的关注重点, 也是云计算应用面对的主要挑战。

由于云计算中用户及资源的高度集中, 由此带来的安全风险及安全后果也比传统方式高出很多。自云计算商用以来, 各大云计算厂商都曾遭到安全事件带来的损失。2014年, Dropbox、三星、Adobe、Internap、Joyent、Microsoft、Amazon、Autotask、腾讯等公司均发生了不同程度的安全事件。如2014年8月19日微软云计算服务Azure的主要组件发生全球大范围宕机; 2014年9月30日 iCloud 出

现暂时宕机; 2014年11月2日, 腾讯云服务器出现了6分钟的访问故障。这些事件的发生加剧了用户对云计算安全问题的担忧。因此, 安全问题成为了制约云计算应用和发展的重要因素。Gartner调查显示, 截至到2015年, 云服务将提供10%的IT企业级安全产品功能。这些服务在推动市场的同时也会对主要的安全技术领域产生一定的影响, 包括安全邮件以及安全网关、远程漏洞评估还有身份和访问管理(IAM)。预计到2016年, 基于云的安全服务市场产值将达到42亿美元。因此不论从用户的角度还是市场的角度, 云计算的安全问题将是影响未来云计算发展的一个关键因素。

1 云计算安全概述

2013年云计算安全联盟(CSA)发布了报告《The Notorious Nine: Cloud

收稿日期: 2014-12-26
网络出版时间: 2015-02-22

Computing Top Threats in 2013》(9大祸害:2013年云计算威胁排名)^[7],报告中列出了与云计算相关的安全威胁。

(1) 数据泄露

数据泄露是排在首位的安全威胁。分布式的开放云计算服务加重了这种威胁。在多租户环境下,多个用户对网络和计算资源实现共享。用户无法直接控制数据计算与存储的风险,甚至无法知道数据的存储位置。同时云平台很难提供与单独客户环境相同的资源隔离等级和保障措施。因此有效保障云服务商自身内部的安全管理和职责分离体系、安全审计,避免云计算环境中多客户共存带来的潜在风险,都成为云计算环境下用户的重大安全顾虑。

(2) 数据丢失

数据丢失的产生原因很多,譬如黑客故意删除或盗走、客观灾难(如火灾、洪水等)。由此造成的损失或带给用户的麻烦那是不可估量的。

(3) 数据劫持

CSA认为账号或服务流量劫持已经成为了一个新的威胁。黑客在窃取了企业的登录信息后,就会进而窃取用户的服务资料,窃听相关活动或交易,甚至操纵数据、返回虚假信息。要抵御这种威胁,一方面用户需要保护并及时更新登录信息,一方面云计算服务提供商需要采用高安全性的加密措施。

(4) 不安全接口

应用编程接口(API)是服务管理者或者维护者对云服务进行配置、管理、协同和监控的接口。在开放的云环境下,用户也可以根据自身需求利用这些接口进行自主开发。虽然API接口优化了管理以及开发模式,但一定程度上增加了安全风险性。

(5) 拒绝服务攻击

分布式拒绝服务(DDoS)一直是互联网发展中的一大威胁。在云计算时代,DDoS攻击会剥夺用户访问资源的权利,加剧网络延迟,或者在不消耗大量资源的情况下利用网络

或服务弱点攻击Web服务器、数据库和其他云资源。由此引起的服务停止一方面会降低用户的感知度,另一方面也会给以时间或磁盘空间付费的用户造成一定的损失。

(6) 恶意的内部人员

恶意内部人员的风险是安全策略中需要考虑的内容。由于用户数据虚拟化到云平台上,恶意的内部人员在拥有较高的访问权限情况下,可以轻而易举的获取用户数据。因此云服务提供商对于内部人员的管控十分重要。

(7) 滥用云服务

大众化的云服务可以对任何人提供计算资源,因此就不能排除有些人会采用一些特殊手段来获取所需资源,如破解密码、发动分布式拒绝服务攻击、传播恶意软件、共享盗版软件等。这就需要云服务提供商对云服务进行规范化管理。

(8) 审查缺失

对于云服务使用者来说,需要审查云服务提供商的系统环境、相关能力以及相关风险。对于云服务提供商而言,则需要对云计算的相关技术、营运责任等审查。在完成风险的规避后,才能保证服务的可靠性。

(9) 共享隔离问题

云服务提供商将基础设施、平台和应用程序以灵活扩展的方式来交付服务。多租户的共享信息就需要采用隔离技术来达到风险的规避。

2 云计算安全体系架构

合理、完备的体系架构可有效部署各种安全关键技术,以满足业务提供商、运营商、安全厂商、用户构成的云生态系统的需求,从容应对云环境下各种安全风险。尽管对于安全架构国际上还没有达成统一共识,但是各大机构都纷纷投入到了安全架构的研究中。

2.1 基于Eucalyptus的多维安全模型

Eucalyptus是加州大学开发出来

的开源云平台^[8]。基于Eucalyptus的多维安全模型主要从环境维、用户维、数据维和业务维4个角度构建。基于Eucalyptus的多维安全模型如图1所示。

环境维从物理环境以及操作系统环境角度分析,主要包括设备安全、运行安全和网络安全。用户维从用户的安全性角度分析,主要包括访问控制、单点登录和信任管理。数据维主要对云平台中的各种数据进行安全保障,包括数据备份,数据迁移和数据加密。业务维从云平台可提供的业务角度进行分析,主要包括日志管理、资源监控、任务控制和负载均衡。

2.2 SOA通用安全架构

面向服务的体系结构(SOA)^[9]将应用程序的不同功能单元通过服务之间定义好的接口和协议联系起来。基于SOA通用架构是一种安全即服务的安全架构,用户可通过定制化服务将不同的安全功能模块进行整合,从而提供完整的功能要求并满足用户的安全需求^[10]。图2是IBM给出的SOA云计算通用安全架构^[11]。安全架构依据云计算的3种服务模式(IaaS、PaaS、SaaS^[12-14])构建了相应的安全组件。依据这个架构,用户可选择不同的安全服务组件,形成独立的安全服务体系。

2.3 云安全整体架构

中国科学院冯登国教授提出了一种云计算安全的参考框架^[16]。主要涉及到技术体系、服务体系和支撑服务体系。3个体系之间相互关联,相互影响。技术体系主要以数据安全与隐私保护服务为目的,分析软件即服务(SaaS)、平台即服务(PaaS)、基础设施即服务(IaaS)不同层次的技术需求。云安全服务体系由一系列云安全服务组成,根据不同的层次可划分为云安全应用服务、云安全基础服务和云基础设施服务。云计算

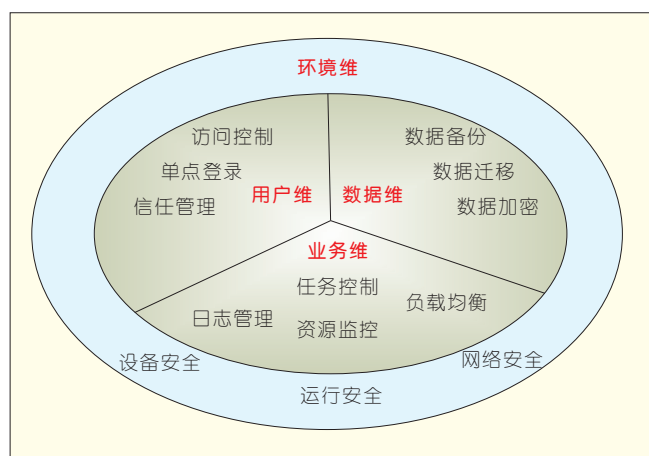


图1
基于 Eucalyptus 的
多维安全模型

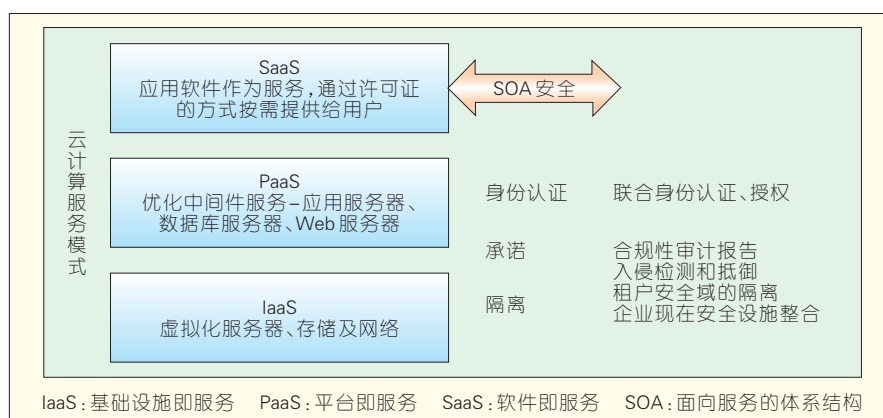


图2 IBM的SOA通用安全架构

安全支撑服务体系主要为安全服务体系提供技术与支撑。云安全架构如图3所示。

3 通用云安全架构

目前,国际上对云计算中的安全框架还没有统一。不同机构从不同的出发点给出了不同的体系架构。本文结合冯登国给出的安全架构和SOA通用架构构造轻量级的云安全平台。

通用云安全架构如图4所示。安全框架分别从云服务域和云用户域两个层面进行构建,分别面向云服务提供者和云服务使用者。云服务域从多角度对其中涉及的安全技术进行剖析。一方面从业务层级(IaaS、PaaS和SaaS)角度考虑,一方面从业务安全的适用层级角度分析。对于通用于各个层次的安全策略可简称

为通用安全,反之称为个性化安全。个性化安全从3种服务模式入手,将安全需求进行层次细化。SaaS层主要针对应用程序安全 and 多租户安全。PaaS层主要针对分布式数据安全、数据库安全和接口安全。IaaS层主要针对物理安全、主机、安全、网络

安全、虚拟化安全和接口安全。其中应用安全包括访问控制、身份鉴别、安全审计、剩余信息防护等。多租户安全包括数据隔离、物理隔离等。数据安全包括数据备份、数据完整性、安全认证、数字签名等。物理安全包括物理位置的选择、物理访问控制、防火防水防静电、温湿度控制、电力供应、电磁防护。网络安全主要包括结构安全、访问控制、安全审计、边界完整性检查、入侵防护、恶意代码防护、网络设备防护、身份鉴别等。主机安全包括访问控制、安全审计、剩余信息保护、入侵防护、恶意代码防护、资源控制。虚拟化安全包括软件虚拟化安全、服务器虚拟化安全。通用安全是每个层次都需要用到的安全服务,主要包括加密与密钥管理、身份识别与访问控制、灾备与业务连续性、数据隔离、安全审计、政策法规等。云用户域主要考虑用户身份安全和终端设备安全。这一层面的安全内容需要用户自行进行安全策略的制订。譬如用户身份安全可包括安全加密、密钥管理、身份识别与访问控制等。终端设备安全可根据实际需要部署一定的安全策略。

4 结束语

云计算的开放性、灵活性促进了市场的蓬勃发展,但同时也带来了一定的安全隐患。各大公司出现的安全问题也让用户对云计算的安全性

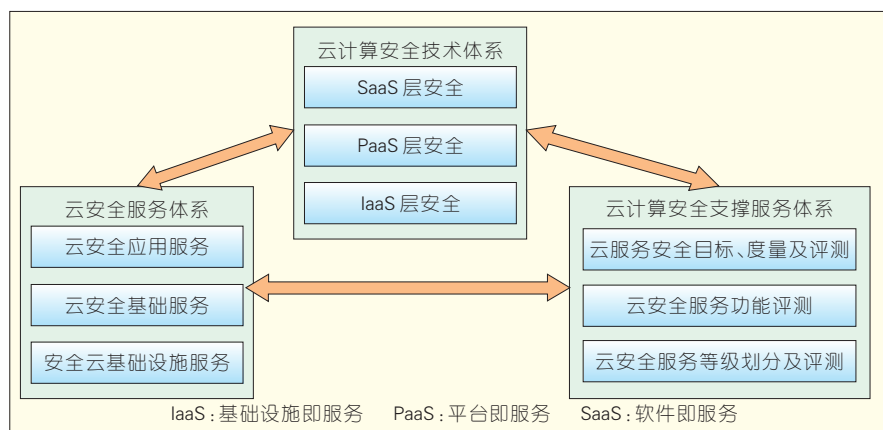
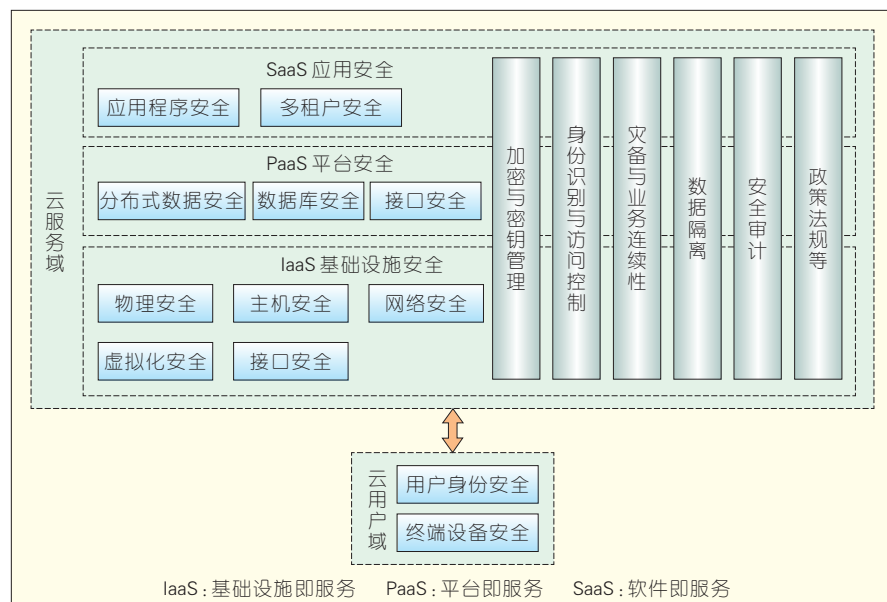


图3 云安全架构



▲ 图4 通用云安全架构

产生了质疑。通过对近年来云计算面临的威胁分析,明确了安全问题的重要性。实施安全措施的首要任务就是构建安全架构。结合主流的安全架构,本文构建了以用户域和服务域两个层面为关注角度、以SaaS、PaaS、IaaS 3种服务模式为基准的通用云安全架构,并对其中的关键因素进行了分析。云服务使用者和云服务提供商可分别根据自身需求来选择相应的安全服务。安全问题的逐步深入解决必将更大程度地带动云计算市场的扩大。

参考文献

- [1] MELL P, GRANCE T. The NIST Definition of Cloud Computing [R/OL]. [2014-12-20]. <http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc>

- [2] 云计算-维基百科 [EB/OL]. [2014-12-20]. <http://zh.wikipedia.org/wiki>
- [3] 刘鹏. 云计算 [M]. 北京: 电子工业出版社, 2010
- [4] Security guidance for critical areas of focus in cloud security computing V3.0 [EB/OL]. [2014-12-20]. <http://www.cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf>
- [5] Top Threats to Cloud Computing, V1.0, Cloud Security Alliance [EB/OL]. [2014-12-20]. <https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf>, 2010
- [6] GENS F. New IDC it cloud services survey: top benefits and challenges [J]. IDC exchange, 2009,18(6):1-5
- [7] The Notorious Nine: Cloud Computing Top Threats in 2013 [EB/OL]. [2014-12-20]. <http://www.cloudsecurityalliance.org/topthreats>, 2013
- [8] DANIEL N, RICH W, CHRIS G. The Eucalyptus Open-source Cloud-computing System [C]//Proceedings of the 9th IEEE/ACM Symposium on Cluster and the Grid, 2009:124-131
- [9] SOA [EB/OL]. [2014-12-20]. http://baike.baidu.com/link?url=33Ze2juBINIC5PWJN8bMxeZ_50H2VWB81z-

- b2iYelYnjWQ50tUDdal-nO8-MCcfhS2a6Rp_3hyQe-8VV0JrCHnINWo4baFzxwTFqwoiPhxKy.2014
- [10] BELL M. SOA Modeling Patterns for Service Oriented Discovery and Analysis [M]. Wiley, 2009
- [11] IBM. IBM Point of View: Security and Cloud Computing [R]. White paper, 2009
- [12] GODSE M, MULIK S. An Approach for Selecting Software-as-a-Service (SaaS) Product Cloud Computing [C]//Proceedings of the IEEE International Conference on Cloud Computing, 2009:155-158
- [13] LIU F, GUO W P. SaaS Integration for Software Cloud [J]. Cloud Computing (CLOUD), 2010,20(3):402-409
- [14] ZHANG Y, LIU S J, MENG X X. GridSaaS: A Grid-Enabled and SOA-Based SaaS Application Platform [J]. Services Computing, 2009,12(2): 521-523
- [15] 冯登国, 张敏, 张妍, 等. 云计算安全研究 [J]. 软件学报, 2011,22(1):71-83

作者简介



陈清金, 中国联通云数据有限公司云计算与产品研发部副经理、博士、高级工程师, 中国电子学会、中国计算机学会高级会员; 主要负责云计算及增值业务研发管理工作; 已主持3项基金项目, 发表学术论文60余篇, 拥有专利10余项、软件著作权10余项。



陈存香, 中国联通云数据有限公司工程师、博士; 主要研究领域为云计算安全策略、云存储策略; 已发表学术论文14篇, 拥有专利3项。



李晓宇, 中国联通云数据有限公司工程师; 主要从事云计算、增值业务和运营支撑系统相关规划研究和咨询设计工作。

综合信息

国家集成电路基金宣布 2015年将向芯片业投资200亿

2015年3月14日,在中国集成电路产业领袖峰会上,国家集成电路基金运营公司华芯投资总裁路军表示,国家集成电路基金在2015年将会在半导体芯片领域投资200亿元,希望借此奠定中国集成电路产业链和建设产业生态的基础。

路军透露,国家集成电路基金在未来的投资中将优先考虑股权投资,也会参与中国芯片行业的投资、并购和整合。未来,还有可能拿出一部分资金投资一些优秀的创业团队,或者参与地方政府的集成电路基金。路军表示,国家集成电路基金会针对芯片行业开展长期大额的投资。 转载自《中国信息产业网》

卫星平台虚拟化——满足多种卫星移动通信需求的必然选择

Virtualizing the On-Board Platform for Multi-Purpose Mobile Satellite Communication Systems

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2015) 02-0039-005

摘要: 认为虽然中国卫星移动通信市场巨大但卫星资源紧缺, 因此充分利用有限的卫星平台, 最大限度满足不同军民通信应用需求并支持多种传输体制, 是中国发展卫星移动通信系统必须要解决的重要问题, 而卫星平台虚拟化是解决这个矛盾的有效途径。基于此, 提出了一种基于快速傅里叶变换(FFT)的虚拟化设计方案, 为推动卫星平台虚拟化研究提供参考。

关键词: 虚拟化; 星载平台; 卫星移动通信

Abstract: The market for mobile satellite communication is huge in China, but satellite resources are very limited. A key issue in the development of mobile satellite communication in China is the efficient utilization of the satellite platform to satisfy the requirements of different military and public applications. Virtualization of the on-board platform is an effective way of addressing this issue. In this paper, we propose a virtualization design based on FFT. This design can be a reference for on research on-board platform virtualization.

Key words: virtualization; on-board platform; mobile satellite communication

王京/WANG Jing¹
赵明/ZHAO Ming¹
高镇/GAO Zhen²

(1. 清华大学, 北京 100084;
2. 天津大学, 天津 300072)
(1. Tsinghua University, Beijing 100084, China;
2. Tianjin University, Tianjin 300072, China)

- 卫星移动通信星载平台应适应多种应用场景和传输体制
- 未来卫星移动通信星载平台应具备灵活的空/时/频资源调配能力
- 基于快速傅里叶变换处理的虚拟化设计是提高星载平台综合利用率的有效途径

卫星移动通信系统具有覆盖区域广、受地形地貌影响小和不受自然灾害影响等特点, 非常适合偏远地区通信、特种行业应用、边海防防务、境外通信和重大灾害中的应急通信等应用^[1-6]。卫星移动通信系统的这些优势对于中国显得尤为重要。首先, 中国幅员辽阔, 但地区发展不均衡, 很多中西部地区由于人口密度低, 并没有被地面移动通信网络所覆盖。第二, 近年来中国突发事件与自然灾害频发, 突发事件及灾情发生地

点往往由于地面网络设施的破坏不能进行有效的通信覆盖, 从而影响事件处理和救灾抢险。第三, 中国有广阔的领空和领海, 而近年来与之相关的领土纷争愈演愈烈, 因此加强这些区域的通信覆盖对于领土安全、保障边防公安正常执法、保护渔业与海上石油开采作业的安全运行都十分重要。另外, 卫星移动通信可与地面系统相辅相成, 共同构建天地互联一体的移动通信系统, 逐步形成一个战略新兴高技术产业, 并支持移动互联网乃至物联网这一新兴战略产业的发展, 支持和推动国家经济社会的持续健康、和谐发展。

鉴于上述卫星移动通信系统对

中国的重要性, 中国从“十二五”开始加大了卫星移动通信系统的研发力度, 并在综合考虑技术储备、系统建设成本、社会经济效益及轨位资源等因素的基础上, 确定了优先发展静止地球轨道(GEO)卫星系统的战略。目前, 中国正沿着“军民结合”的思路, 开始第一代GEO卫星移动通信系统的建设以及下一代GEO卫星移动通信系统的关键技术攻关。

如上所述, 卫星移动通信系统建立以后, 将对中国各个方面的全面发展产生重要的推动作用。但是问题也随之而来, 不同应用场景, 如偏远地区民用覆盖、海上通信服务、航空通信服务、军队作战通信服务(包括

收稿日期: 2015-01-10

网络出版时间: 2015-02-22

基金项目: 国家高技术研究发展(“863”)计划(2012AA01A502)

陆军、空军、海军、二炮)等,对同一卫星移动通信系统的需求各不相同。由于这些应用的体制不同,需各自配备专门的信关站设备,但是卫星资源非常有限,多个系统往往需要共用同一个卫星平台。

另一方面,GEO卫星的寿命通常可达10~15年,从目前IT技术的发展速度来看,这期间一定会有很多新技术被发明和推广到地面移动通信领域。作为未来天地一体化网络中的重要组成部分,卫星移动通信只有紧跟地面通信系统的发展趋势,才能充分发挥天地一体化的优势。为此,未来卫星通信星载平台应该具备灵活的资源重构能力,以使星载平台的处理任务能在必要的时候适应地面系统的发展。

为了使卫星平台同时满足不同应用需求,并能够适应未来天地一体化的发展趋势,本文提出卫星平台虚拟化的概念,并提出一种基于快速傅里叶变换(FFT)处理的实现方式。

1 卫星平台虚拟化的基本概念

卫星平台虚拟化的核心就是“空/时/频”资源的灵活利用。“时/频”资源处理主要是指数字信道化技术^[7-10],即从某个时间内的宽带信号中分离出子信道,以及将多个窄带子信道合成一个宽带信号;空间资源利用主要是指多波束覆盖。

目前主流的GEO卫星移动通信系统大都采用透明转发式的卫星平台。这类系统本质上已经具备一定意义的“虚拟化”特性,即卫星平台不进行具体与体制相关的信号处理,例如调制/解调、编码/解码、复用/解复用等,但是数字信道化过程往往受目标传输体制的带宽等参数限制,多波束覆盖也相对固定。从这个角度看,这些系统只是实现了部分“虚拟化”。

本文的目的就是通过一种基于FFT处理的平台构架,将卫星平台“虚拟化”的程度进一步加深,从而大

大提高星载平台的灵活性及整个系统的综合利用率。

2 国际卫星平台虚拟化的发展情况

21世纪以来,数字信道化技术陆续在一系列商用和军事卫星移动通信系统中得到应用,并成为这些卫星有效载荷设计中的关键技术。国际上采用数字信道化技术的卫星移动通信系统主要包括ACeS、Thuraya、MUOS和WGS等系统。

ACeS系统的数字信道化器主要完成FDMA-TDMA星载交换^[11]。星载平台接收到上行GSM信号后,首先将其下变频到合适的中频,经过采样和模数(A/D)转换后依次完成频分复用(FDM)解复用、TDMA处理和FDM复用。由于每个用户的GSM信号带宽相同,即子信道带宽相同,ACeS星载平台需要解决的是均匀带宽的窄带信号交换问题,因此使用了基于多级法的信道化方案。

与ACeS类似,Thuraya系统^[12]也是兼容GSM的卫星移动通信系统,因此星载数字信道化实现的也是均匀带宽的窄带信号交换。

MUOS系统^[13]是基于超高频(UHF)频段WCDMA体制的卫星移动通信系统,并需要兼容之前的特高频后继星卫星通信系统(UFO)系统,因此星载数字信道化器主要完成UHF系统与UFO系统的信号交换功能,也属于均匀带宽的窄带交换应用场景。

宽带全球卫星通信(WGS)^[14]系统是美军的新一代卫星移动通信系统。它使用X波段和军用Ka波段资源,完成X频段战术通信服务,扩充GBS Phase II系统功能,同时提供Ka波段双向通信。每颗WGS卫星的系统带宽为4.875 GHz,分成1 872个2.6 MHz的子信道,每个子信道都可以实现从任意波束到任意波束的路由,包括频段内(Ka或者X频段内)和频段间(Ka与X波段间)的交换。

综合以上分析可以看出,大部分

系统的星载数字信道化器实现的都是均匀带宽的窄带信号交换功能,只有WGS系统可实现带宽可变的宽带信号交换功能。但是由于WGS是军用系统,且尚未部署完成,因此几乎搜集不到有价值的参考资料。

在多波束覆盖方面,前面提到的国际主要卫星移动通信系统中,除了WGS系统是通过喇叭天线形成波束,其余系统都配备星载大天线以及S波段/L波段(S/L)大规模馈源阵列,并使用了数字波束成型技术。Thuraya和ACeS等系统采用的都是星上数字波束成型技术。由于兼容地面GSM系统,子信道宽度较窄,因此子带内测波束成型系数可认为是一致的,不存在宽带信号的波束成型问题。另外,Thuraya系统支持星上波束成型系数可调,从而实现热点优化覆盖。为了进一步提高多波束覆盖的灵活性,最新的TerreStar和SkyTerra系统都使用了地基波束成型技术(GBBF),将波束成型系数生成和处理完全转移到了地面信关站,大大提高了处理效率。由于SkyTerra支持LTE的传输体制,因此需要解决宽带信号的波束成型问题。

3 中国卫星平台虚拟化的发展情况与需求分析

中国目前正处在第一代卫星移动通信系统的建设时期,以及下一代卫星移动通信系统的关键技术攻关时期,因此在星载平台的虚拟化设计方面还没有实践经验积累。

星载平台数字信道化的理论研究方面,清华大学的杨志明博士针对宽带卫星通信系统提出一种新型数字信道化方法及相应的数字信道化器结构,并提出了精确重构原型滤波器的有效设计方法(VNS-LMSE);解放军信息工程大学李冰博士重点研究了基于Goertzel方法的单通道信道化技术及高效实现,以及基于调制滤波器组的多通道信道化技术及实现结构优化。此外,解放军理工大

学、南京电信技术研究所、东南大学、上海交通大学以及西安空间无线电技术研究所等单位也对星上数字信道化技术开展了研究,基本上都是对传统的几种信道化方法进行理论和实现分析,包括频域滤波(或FFT滤波)方法、多相/离散傅里叶变换法和信号分析方法。

星上数字波束成形技术的原理较为简单,这方面的理论研究国内开展的也比较多,内容涉及多波束天线设计、快速成型技术研究、多波束天线校正技术研究等。研究单位主要包括中科院上海微系统所、西安电子科技大学、哈尔滨工业大学、国防科技大学等。另外,南京航空航天大学 and 重庆大学还专门研究了针对空域目标或者多飞行器测控的数字波束成型技术,但主要是应用于地面对天检测,而不是卫星对地观测。

虽然中国在卫星平台虚拟化技术方面实践经验积累较少,但是对卫星平台虚拟化的需求很迫切。在民用通信方面,正在建设的第一代卫星移动通信系统将采用TDMA体制,主要支持窄带通信应用;而下一代卫星移动通信系统拟采用基于OFDMA和SC-FDMA的LTE体制,拟灵活支持从1.4 MHz到20 MHz的多种信号带宽;未来天地一体化移动通信系统(包括5G系统)将支持海量低速率数据终端。各种应用场景在通信带宽和波束覆盖方面都对卫星移动通信系统提出不同需求,单一而固定的卫星平台很难满足。

4 基于FFT处理的卫星平台虚拟化设计

卫星平台虚拟化的主要目的是使星载平台能够根据需要灵活调整子信道的时间尺度(帧长/突发长度)、频域尺度(带宽)以及空间尺度(波束),以满足多种应用与体制对卫星移动通信系统的传输需求。考虑到基于FFT的频域滤波信道化技术的灵活性,本文提出一种以FFT处理

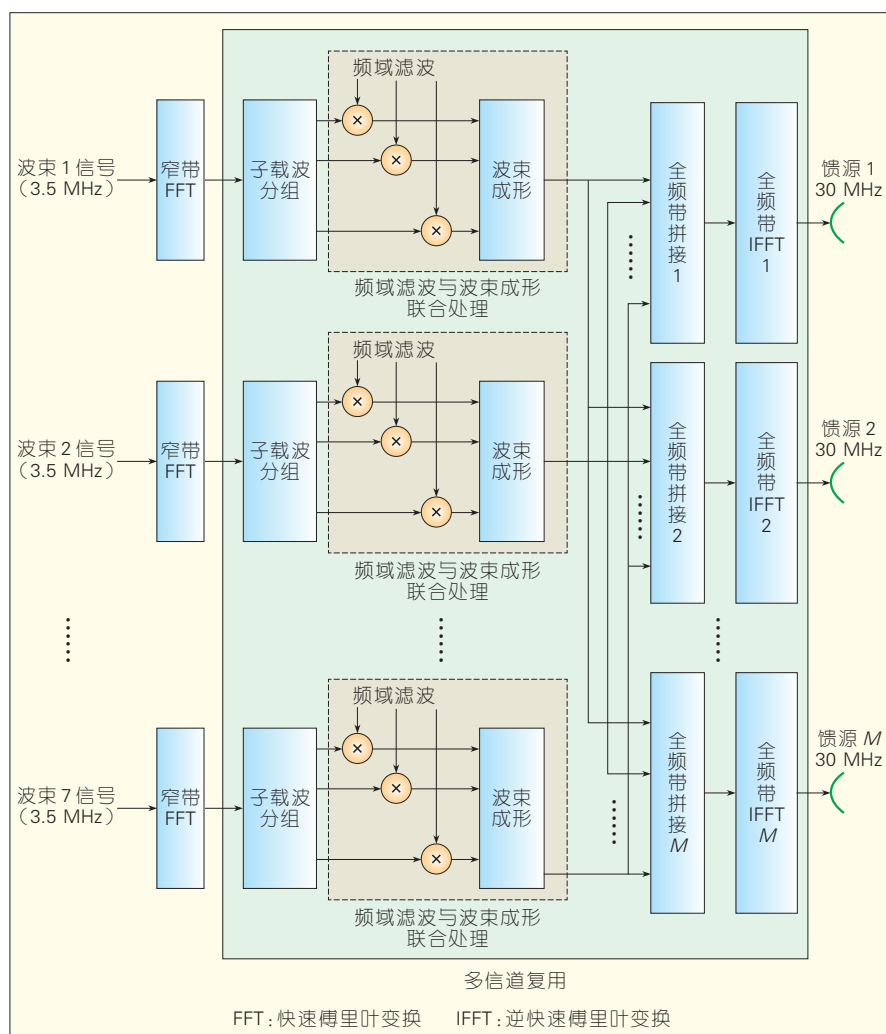
为核心的星载平台虚拟化设计方案。由于可分别对每个子载波做单独处理和路由,这种方式可以适应各种信号带宽的灵活波束设置。该平台可应用于中国基于透明转发的第一代系统以及基于星上数字信道化的新一代系统。

4.1 基于透明转发模式的星载平台虚拟化技术

针对基于透明转发的卫星移动通信系统而言,所设计的虚拟化处理主要完成基于FFT的波束成形,适用于多种子信道带宽信号。

前向链路FFT波束成形实现原理如图1所示(省略了前端的数模转换、上变频和高功率放大器)。以

ITU-R卫星移动通信系统需求为例,系统带宽为30 MHz,若采用7色复用,则每个波束信号的最大带宽约为3.5 MHz^[5]。由于卫星平台的功率受限,每个波束的实际信道带宽 B 会小于3.5 MHz。对应每个波束的前向信号首先经过FFT转换到频域,然后经过子载波分组和频域滤波,获得若干子带,每个子带内的信号可共享一组波束成形系数。每个波束信号的波束成形器输出 M 组带宽为 B 的信号,然后在波束相关联的 M 个馈源通道处进行频域信号拼接,之后通过逆快速傅里叶变换(IFFT)转换为时域信号,分别送至与该波束相关的 M 个馈源上。其中频域滤波与波束成形处理可以联合进行。更多波束将会复



▲图1 前向链路基于FFT的宽带信号波束成形原理

用这7组3.5 MHz(全频30 MHz)的带宽,基本处理流程与图1一致,只是需在同频波束共享的馈源通道入口处先做频域信号叠加,然后再对叠加信号进行全频IFFT操作。图1中方框内的部分定义为多信道复用过程(MCM)。

反向链路基于FFT的宽带信号波束成形流程如图2所示(省略了前端的低噪放、下变频和数模转换操作)。每个馈源接收的全频带信号首先经过FFT转换到频域,然后按照频率复用规则划分为若干波束带宽信号(如7个3.5 MHz信号);同频信号叠加后先经过自带分组和频域滤波,形成可以共享波束成形系数的子带,并经过波束成形区分同色小区信号(如波束1、8、15等);最后经过波束带宽IFFT恢复对应每个波束的时域信号。图2中方框内的部分定义为多信道解复用过程(MCD)。

由于使用了基于快速傅里叶变换和反变换技术,以上宽带波束成形的处理复杂度大大降低。但是由于基于快速傅里叶变换得到的结果仅能表示输入信号段与滤波器冲击响应的循环卷积,而我们需要的是线性卷积结果,因此需要用重叠保留法或者重叠相加法进行处理^[16]。

4.2 基于星上处理模式的星载平台虚拟化技术

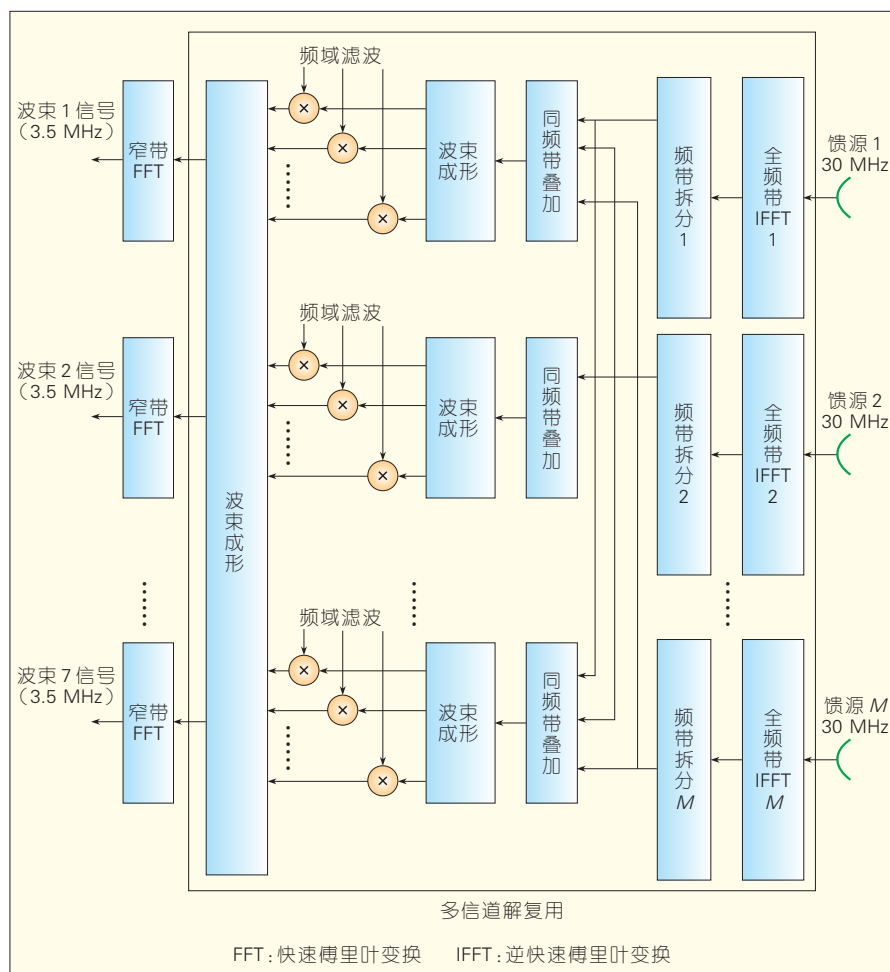
基于星上交换的数字信道化基本过程如图3所示,其中MCD为图2中所示的反向链路中的解复用过程,MCM为图1中所示的前向链路中的复用过程,两者之间为以频域信号为基础的“时/频/空”交换矩阵。可见,数字信道化器的结构就是将透明转发平台复用/解复用过程相组合,并加入星上交换功能,因此主要功能模块是一致的,这构成星载平台虚拟化设计的基础。

5 结束语

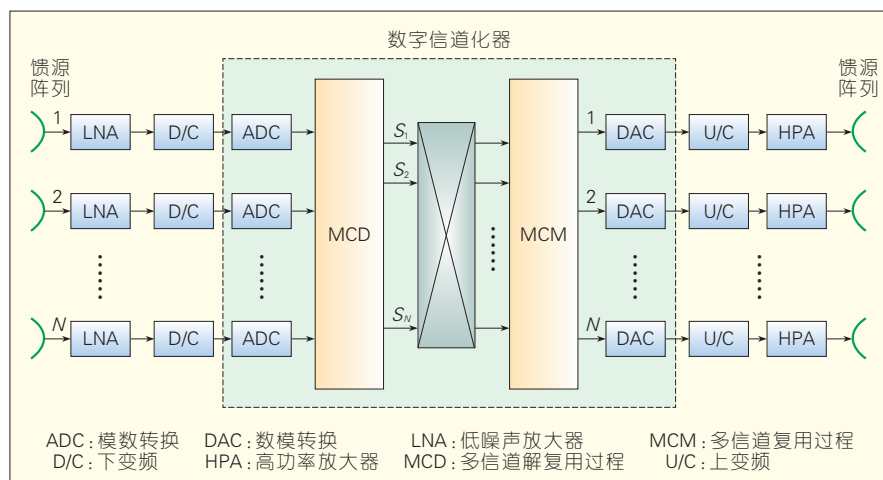
中国的卫星移动通信市场很大,

除了农业、林业、勘探、海监等特殊民用行业外,陆军、海军、空军等多兵种也对卫星移动通信服务有很大需求。由于卫星资源紧缺,如何充分利

用有限的卫星平台,最大限度满足不同军民通信应用需求,是中国发展卫星移动通信系统必须要解决的问题,而卫星平台虚拟化则是应对这



▲图2 反向链路基于FFT的宽带信号波束成形原理



▲图3 基于FFT的星上数字信道化

一挑战的必然选择。本文提出的基于快速傅里叶变换处理的虚拟化设计方案为实现这一目标提供了一种有价值的参考。

参考文献

- [1] 张更新, 甘仲民. 浅论中国卫星移动通信系统的发展思路和策略 [J]. 数字通信世界, 2005 (7):24-27
- [2] 何平江. 静止轨道卫星移动通信系统的设计考虑 [J]. 数字通信世界, 2008(8):18-21
- [3] 何家富, 李广侠. 卫星移动通信的应用与发展 [C]//第四届卫星通信新业务新技术学术年会, 2008:17-24
- [4] 张健. 发展中国卫星移动通信的问题及建议 [C]//第四届卫星通信新业务新技术学术年会, 2008:60-65
- [5] 张更新, 甘仲民, 李广侠. 对发展中国卫星移动通信的有关思考 [J]. 卫星与网络, 2010(5):32-34
- [6] 王京, 赵明, 晏坚. 关于发展卫星移动通信系统的一些思考 [C]//第六届卫星通信新业务新技术学术年会论文集, 2010:241-244
- [7] ANGELETTI P, GAUDENZI R D. From "Bent Pipes" to "Software Defined Payloads": Evolution and Trends of Satellite Communications Systems [C]//Proceedings of the 26th AIAA International Communications Satellite Systems Conference, San Diego, CA, 2008
- [8] ANDERSON P R, BARTAMIAN L. Growth trends in communication satellites and the impact of satellite system architecture [C]//Proceedings of the 26th AIAA International

- Communications Satellite Systems Conference, San Diego, CA, 2008
- [9] ANZALCHI J, HARVERSON M. Generic flexible payload technology for enhancing in-orbit satellite payload flexibility [C]//Proceedings of the 25th AIAA International Communications Satellite Systems Conference, Seoul, South Korea, 2007
- [10] BUTASH T C, MARSHALL J R. Leveraging Digital On-Board Processing to Increase Communications Satellite Flexibility and Effective Capacity [C]//Proceedings of the 28th AIAA International Communications Satellite Systems Conference, Anaheim, California, 2010
- [11] WIEMANN K. The ACeS digital channelizer - The next generation in regional digital satellite telephone communications [C]//Proceedings of the 19th Digital Avionics Systems Conference, Philadelphia, PA, 2000
- [12] ETSI. GMR-1 05.002 (ETSI TS 101 376-5-2). GEO-Mobile Radio Interface Specifications; Part 5: Radio interface physical layer specifications; Sub-part 2: Multiplexing and Multiple Access; Stage 2 Service Description [S]. ETSI, 2000
- [13] CURLES W W, CURRY L T. Mobile User Objective System Program Overview [C]//Proceedings of the MILCOM, 2002
- [14] KUMAR R, TAGGART D, MONZINGO R, et al. Wideband gapfiller satellite (WGS) system [C]//Proceedings of the IEEE Aerospace Conference, El Segundo, California, 2005
- [15] ITU-R. ITU-R Report M.2176-1. Vision and requirements for the satellite radio interface (s) of IMT-Advanced [S]. ITU-R, 2012
- [16] 胡广书. 数字信号处理——理论、算法与实现

(第二版) [M]. 北京: 清华大学出版社, 2000

作者简介



王京, 清华大学教授, 国家科技重大专项“新一代宽带无线移动通信网”总体专家组专家, 工信部“IMT-2020 (5G) 推荐组”专家组专家, 国家“863”计划 5G 移动通信重大项目总体专家组专家, 国家“863”计划卫星移动通信关键技术重大项目专家组专家, 中国通信学会常务理事, 北京通信学会副理事长; 已发表论文 100 余篇, 申请国家发明专利 30 余项。



赵明, 清华大学教授, 国家“863”课题“可重构卫星通信平台”负责人, 卫星移动通信领域国家“863”课题 (2012AA01A502) 负责人; 从事卫星移动通信、无线通信及软件无线电方面的研究。



高镇, 天津大学副教授; 从事卫星移动通信及容错信号处理方面的研究; 已发表 SCI 和 EI 论文 20 余篇, 申请国家发明专利 6 项。

上接第 3 页

件安装在自己购买的硬件上呢? 软件本是第三产业, 却被错当成了第二产业, 是到了软件回归服务业的时候了。软件即服务 (SaaS) 的商业模式不再以产品为核心了, 而是按用户的使用量和定制程度收费, 是向以用户为中心的商业模式迈出的重要一步。

同样的使用量和同样的数据, 对不同用户的价值是不一样的; 对于同样的用户, 不同数据的价值也是不同的。商品的价格不是由生产成本决定的 (这是软件产品的商业模式), 也不是使用量决定的 (这是 SaaS 等云计算的商业模式), 而是用户认可的使用价值来衡量的。

未来的云计算商业模式, 需要继续向以用户为中心前进, 发展到用户价值为核心的商业模式^[6]。对于用户的数据, 使用量只是定价的基础, 用

户自己的估值和报价, 会成为核心商业模式。这就好比是邮局的包裹邮寄服务, 邮局的主要定价方式是重量和体积, 而用户还可以自己定义价值, 据此选择保价并支付额外费用。

5 结束语

因为移动性的需要和移动终端的限制, 移动互联网天然依托云计算发展, 云计算也需要来自移动互联网的支持。移动云计算系统如何分工协作移动性, 云服务商如何保护用户财产权和隐私以赢得更多的信任, 技术上大规模、低成本和长寿命存储介质的问题, 如何向软件定义的可靠性迁移, 以及商业上向用户价值发展等, 都是值得进一步探讨的问题。

参考文献

- [1] 何宝宏. 移动互联网是第三代互联网 [J]. 中兴通讯技术, 2009, 15(4):35-38

- [2] 工信部电信研究院. 云计算白皮书 [R]. 工信部电信研究院, 2012
- [3] 黄舍予. 云计算发展无信不立 [EB/OL]. [2015-01-02]. http://www.cnii.com.cn/technology/2014-07/16/content_1402383.htm
- [4] 何宝宏. 软件定义的世界 [EB/OL]. [2015-01-02]. http://www.cnii.com.cn/index/content/2012-10/23/content_1013866.htm
- [5] 何宝宏. 软件定义的可靠性 [EB/OL]. [2015-01-02]. http://www.cnii.com.cn/wlkb/rmydb/content/2013-06/27/content_1171909.html
- [6] 何宝宏. 软件商业模式的变迁 [EB/OL]. [2015-01-02]. <http://news.163.com/14/1125/09/ABSTSQ7400014AED.html>

作者简介



何宝宏, 中国信息通信研究院 (原工业和信息化部电信研究院) 互联网研究领域首席、通信标准所副所长、原互联网中心主任, 云计算发展与政策论坛秘书长, 数据中心联盟常务副理事长, 中国通信标准化协会 (CCSA) IP 与多媒体工作委员会副主席; 长期从事互联网相关技术、标准、政策和发展战略等的研究, 目前主要致力于可信云计算、大数据、数据中心设备、绿色节能等研究工作。

DOI: 10.3969/j.issn.1009-6868.2015.02.010

网络出版地址: http://www.cnki.net/kcms/detail/34.1228.TN.20150222.1727.007.html

ITU-R 卫星移动通信标准 ——整合移动业务的动力

Standardization of Mobile Satellite Service Technology in ITU-R: Driving Force of Integrated Mobile Services

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2015) 02-0044-004

摘要: 认为国际电信联盟无线电通信部门(ITU-R)关于卫星移动通信的标准化新进展将在全球移动通信设备和系统无缝连接的实现过程中起到重要作用,卫星和地面的集成/混合的天地一体化系统将充分发挥原系统各自的优势,未来多数的卫星移动业务(MSS)都将基于这种天地一体化系统。指出 IMT-Advanced 系统的卫星部分和 ITU-R 标准化无线接口是集成 MSS 系统的关键技术,将成为激活 MSS 产业的新的驱动力。

关键词: 卫星移动通信;天地一体化;标准化;国际电信联盟无线电通信部门

Abstract: Well-developed standards will play an important role in facilitating seamless interworking of equipment and systems globally. Most future MSS services will use integrated and/or hybrid satellite-terrestrial networks that leverage both satellite and terrestrial networks. Satellite components of IMT-Advanced systems and ITU-R standardized radio interfaces provide key technology in integrated satellite and terrestrial networks, and this will be a new driving force behind MSS industries.

Key words: mobile satellite communications; integrated and/or hybrid satellite and terrestrial networks; standardization; ITU-R

金秀英/KIM Sooyoung

(韩国全北国立大学,韩国全州市
561756)
(Chonbuk National University, Jeonju
561756, Korea)

- 国际移动通信系统的卫星无线接口已成功修订并全新开发
- 混合与集成式网络架构将在未来高效卫星移动业务中发挥重要作用
- 天地一体化系统将成为激活卫星移动业务产业新动力

1 ITU-R 标准

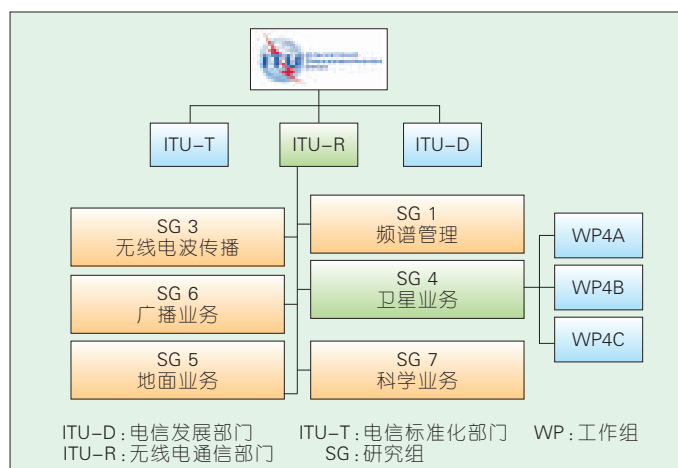
国际电信联盟(ITU)是主管信息与通信技术的联合国专门机构。在 ITU 中,政府及私营机构就全球电信网络和业务进行协调。ITU 包括 3 个部门:无线电通信部门(ITU-R)、电信标准化部门(ITU-T)以及电信发展部门(ITU-D),其组织架构如图 1 所示。三大部门的工作涵盖电信领域的各个方面,从制订标准促进全球系统设备的无缝对接,到为日益增长的大量无线业务和设计计划采

纳运营程序,从而改善发展中国家的电信基础设施。

ITU-R 最重要的任务就是分配无线频谱和卫星轨道,这些都属于有限

收稿日期: 2014-12-29
网络出版时间: 2015-02-22

图 1
ITU 组织架构



的自然资源。ITU-R 由 6 个研究组 (SG) 组成, 每个研究组包括若干个工作组 (WP)。每个工作组都有自己的研究项目, 工作组制订出技术基础, 提交世界无线电通信大会 (WRC) 进行决策。另外, 工作组以建议书、报告和手册的形式, 制订无线电通信的全球标准。ITU-R 第 4 研究组负责制订卫星业务的技术建议书。该研究组包括 3 个工作组: 4A 工作组研究卫星固定业务 (FSS) 和卫星广播业务 (BSS) 的高效轨道或频谱利用; 4B 工作组研究卫星固定业务、卫星广播业务以及卫星移动业务 (MSS) 的系统、空中接口、性能和可用性目标, 包括 IP 应用和卫星新闻采集; 4C 工作组研究卫星移动业务和卫星无线电测定业务 (RDSS) 的高效轨道或频谱利用。ITU-R 第 4 研究组始终在制订相关标准和规则, 确保卫星轨道和频谱资源的有效管理。世界无线电通信大会已实施无线电规则 (RR), 并已更新区域协议。第 4 研究组制订的大量标准化建议书和报告可为卫星通信系统设计人员提供重要信息。虽然建议书或报告非强制性实施, 但其包含的信息可通过高效利用频谱资源而用于设计和实施各种卫星通信系统^[1]。

ITU-R 的建议书和报告一般命名为 ITU-R X.Y-Z 建议书 (或报告), 其中 X 代表整个相关部分, 称为一个系列, Y 代表序列号, Z 代表修订次数。例如, S 系列为卫星固定业务建议书和报告; M 系列为移动、无线电测定、业余及相关卫星业务建议书和报告, 因此, 卫星移动业务技术标准都包含在 M 系列建议书和报告内。例如, ITU-R M.2176-1 报告是与卫星移动业务技术相关的报告, 序列号为 2176, 经过一次修订。ITU-R 的大部分标准都用作频率共享和高效频谱利用的协作研究。

2 卫星移动业务技术建议书

表 1 列出了近期制订的一些主要

M 或 S 系列建议书和报告, 涉及卫星移动业务性能和传输技术方面的内容, 但未罗列有关干扰分析和频率共享研究的建议书和报告。卫星移动业务的主要建议书和报告如表 1 所示。表 1 所列建议书和报告的详细信息将在后续章节进行说明。

ITU-R 报告通常提供卫星系统设计和运行的背景信息, 这部分通常对于卫星系统的设计和运行来说并非强制性的, 但有时可用作对应建议书的补充材料。例如, ITU-R S.2173-1 报告和 S.2222 报告分别为 ITU-R S.1878 建议书和 S.1897 建议书提供了背景信息。另外, M.2176-1 报告提供了 ITU-R S.2047 建议书的制订所需的全部基础信息和必要信息, 而 ITU-R M.2279 报告描述了建议书制订过程中的全部流程。

3 卫星移动业务的无线接口

3.1 IMT-2000 卫星无线接口

国际移动通信系统-2000 (IMT-2000, 即第三代移动通信) 的全部无线接口都包含在 ITU-R S.1457 建议书中^[2]。IMT-2000 系统卫星组件包括 8 个卫星无线接口 (SRI)。由于建议

书一直在修订, 因此 8 个卫星无线接口构成的卫星部分在 ITU-R M.1850 建议书^[3]中单独阐述, 而地面部分在 M.1457 建议书^[4]中阐述。由于卫星部分单独包含在 M.1850 建议书中, 因此必须为卫星部分制订 IMT-2000 终端全球流通新标准, 即 ITU-R M.2014 建议书。

在 ITU-R M.1850 建议书中, SRI-A 族、C 族与 G 族之间存在一个协调工作, 因为其在 3GPP 地面 WCDMA 无线接口方面有很大程度的共性。SRI-A 族与 C 族接口利用了卫星独有的特性, 如功率控制算法、同步码和上行随机接入等, 因此在性能方面具有优势; 而 SRI-G 族接口拥有与地面部分完全兼容的优势。因此, ITU-R S.1850-1 建议书 (第一次修订) 包括了增强型 SRI-G 族接口^[5], 该族接口采纳了 SRI-A 族与 C 族接口的全部优势。增强型 SRI-G 族接口与 SRI-A 族、C 族和 G 族接口的对比如表 2 所示。从表 2 可以看出增强型 SRI-G 族接口与原先 SRI-A 族、C 族和 G 族接口相比所具有的一些特性。

最近的第二次修订 ITU-R S.1850-2 已经过批准, 以反映 SRI-H 族中最新版本静止轨道移动无线接

▼表 1 卫星移动业务的主要建议书和报告

序列号	标题	日期
M.2175 报告	MSS 系统中利用交叉极化消除的双线性计划传输技术	2010 年 7 月
S.1878 建议书	卫星系统的多载波传输技术	2010 年 12 月
M.2149-1 报告	自然灾害和类似紧急情况下救灾行动的卫星移动服务系统的使用和例证	2011 年 9 月
S.2222 报告	基于 IP 的混合卫星地面网络的跨层 QoS	2011 年 9 月
S.1897 建议书	基于 IP 的卫星地面混合网络中的跨层 QoS 配置	2012 年 1 月
M.1850-2 建议书	IMT-2000 卫星组件的无线电接口的详细规范	2014 年 9 月
M.1854-1 建议书	卫星移动业务在灾害应对和救灾中的应用	2012 年 1 月
M.2176-1 报告	IMT-Advanced 卫星无线接口的愿景和要求	2012 年 9 月
M.2014 建议书	IMT-2000 卫星终端的全球流通	2012 年 3 月
M.2047 建议书	IMT-Advanced 卫星无线接口的具体规范	2013 年 12 月
M.2279 报告	IMT-Advanced 卫星进程 (4 到 7 阶段) 的评价、建立共识和决策的结果, 包括 IMT-Advanced 卫星无线接口的特性	2013 年 10 月
S.2306 报告	卫星通信多维信号映射技术	2014 年 7 月
S.2173-1 报告	卫星系统基于多载波的传输技术	2014 年 9 月
MSS: 卫星移动业务 QoS: 服务质量		

▼表2 增强型SRI-G族接口与SRI-A族、C族和G族接口的对比

项目	增强型 SRI-G	原先的 SRI-G	原先的 SRI-A	原先的 SRI-C	
多址接入	WCDMA	WCDMA	WCDMA	WCDMA	
双工模式	FDD	FDD	FDD	FDD	
带宽	5 MHz	5 MHz	2.5 MHz 和 5 MHz	5 MHz	
最大数据率	384 kb/s	384 kb/s	144 kb/s	384 kb/s	
与地面规范的兼容性	完全兼容	完全兼容	略有修改	略有修改	
主要卫星特性	基于 SRI-A 的寻呼信道, 基于 SRI-A 的下行数据扰频, 基于 SRI-C 的随机接入, 基于 SRI-C 的帧结构, 基于 SRI-C 的功率控制, 基于 SRI-C 的同步		物理层无修改	额外的寻呼信道、下行数据扰频、帧结构、功率控制、同步、随机接入	随机接入、帧结构、功率控制、同步
FDD : 频分复用 SRI : 卫星无线接口 WCDMA : 宽带码分多址					

口-1(GMR-1)的关键特性,包括高效多播实现、灵活波束覆盖、新分组业务信道(PDTCH)变量以及控制信道的实现^[6]。

3.2 IMT-Advanced 卫星无线接口

IMT-Advanced 卫星无线接口的详细规范制订始于ITU-R M.2176报告的发布,其中包括IMT-Advanced卫星无线接口的愿景、技术性能相关要求以及SRI技术评估指导。制订过程的最终结果见ITU-R M.2047建议书^[7]。该建议书主要由中国和韩国制订。接口采用基于正交频分复用技术(OFDM)的无线接口,分别称为BMSat和SAT-OFDM^[8]。

两种技术规范都考虑了对地静止卫星轨道(GSO)卫星系统,并且都基于3GPP LTE技术,从而能与地面规范拥有最大程度的兼容。BMSat基于3GPP LTE-Advanced规范最新版本(Release 10)开发^[9],而为了提高卫星系统的稳定性,SAT-OFDM基于Release 8规范开发^[10]。

卫星系统具有较长的往返通信延迟,其信道时间延展不足,很难通过多天线技术获得分集增益。因此,卫星系统对LTE技术的直接应用可能会引发严重的性能下降和效率降低。SAT-OFDM与BMSat均针对卫星系统的这些特性开发一些新的传输方案设计。这些规范的详细标准化过程和技术特征可参见文献[11-13]。

现有卫星无线接口(SRI)在将来可能发生演进,或者出现新的系统结构,这些变化都将体现在新的卫星无线接口版本中。政府和私营机构都会将相关研究成果提交4B工作组会议决策,而卫星无线接口的更新则应在4B工作组和第4研究组进行全面讨论和决策后,并反映在ITU-R 2047建议书的修订版本中。地面和卫星部分的3G无线接口的演进也采用同样的ITU-R程序得以阐述。

4 卫星移动业务的性能增强技术与集成

4.1 多载波和多极化传输

ITU-R S.1878建议书提出卫星系统基于多载波的传输技术,其背景信息详见ITU-R S.2173报告。该建议书由韩国和加拿大联合制订,最近日本修订了报告,在其中添加了一个新的应用举例。该建议书为卫星固定业务(FSS)和卫星移动业务(MSS)对多载波码分多址(MC-CDMA)和载波干涉正交频分复用(CI-OFDM)技术的利用提供了指导,分别针对MC-CDMA和CI-OFDM技术阐述了卫星系统的配置。

ITU-R M.2175报告描述了MSS系统自适应双极化复用模式,以及一个数字交叉极化干扰消除方法,并提供了一些模拟结果。另一方面,ITU-R S.2306报告提出了高级极化传输技

术。该报告描述了多极化复用(PPM)和极化域调制(PDM)模式,而非简单的双线极化模式,并说明了PDM与PPM的基本概念和性能评估结果。

4.2 卫星IP技术

ITU-R S.1879建议书为跨层设计方案提供指导,从而能改善单机或混合卫星网络的多媒体应用性能。ITU-R S.2222报告提供了该建议书的背景信息,并描述了各种形式的混合卫星和地面网络。该报告还提供了跨层协议交互的相关信息。

4.3 MSS在灾害应对和救灾中的应用

由于卫星系统独立于本地的基础设施、拥有广泛的覆盖范围,并且易于部署,因此,是为公共和救灾行动提供即时通信服务的最有效的方式之一。为了在该领域高效利用MSS系统,ITU-R M.1854建议书提出“MSS在灾害应对和救灾中的应用”,ITU-R M.2149报告列举了大量单机或集成网络配置的例子,从而在救灾中能应用MSS系统。MSS系统如表3所示。表3基本包含了现有的可提供救灾通信应用的MSS系统。

5 MSS系统未来的工作

ITU-R定义了一种集成MSS系统,包括卫星部分和地面部分,地面

▼表3 MSS系统

系统名称	轨道类型	覆盖范围
ACeS	GSO	区域3的部分地区
AUSSAT	GSO	澳大利亚州
DBSD North America	GSO	北美洲
Globalstar	非GSO	全球
Inmarsat	GSO	全球
Iridium	非GSO	全球
SkyTerra	GSO	北美洲和中美洲
Solaris Mobile	GSO	区域1的部分地区
Terrestar	GSO	北美洲
Thuraya	GSO	区域1和区域3
GSO:对地静止卫星轨道		

部分与卫星部分相互补充,并且作为MSS系统的一部分运行。在该系统中,地面部分由卫星资源和网管系统控制。地面部分采用与对应的卫星移动系统相同的MSS频带。

在未来的移动通信场景中,单靠卫星系统为用户提供各种多媒体服务将面临诸多限制。集成MSS系统和混合卫星/地面系统拥有较高的频谱利用率,可为未来各类公共和私人移动通信服务提供较好支撑。因此,4B工作组正就“集成MSS系统场景和性能”的建议书和报告的制订开展研究。工作文件包括系统架构和部署场景,以及可能的频段和复用方案。目前已有众多性能评估结果,包括干扰和吞吐量分析结果,其他结果预期将在后续会议中添加。为了将来能够提供高效的服务,考虑了许多提高频谱效率的先进技术,如结合干扰管理的分数频率复用技术和地基波束成形(GBBF)技术。

虽然雨衰对于频带低于6 GHz(针对便携或手持用户终端)的卫星系统而言不会造成重大影响,但对于更高频带的MSS系统则需要采取应对措施。MSS系统正在进行并且还将持续进行的另一项工作是,综合雨衰及用户移动导致的衰落影响,针对自适应传输技术的高效实现及其对卫星系统性能目标的影响开展研究。

目前大多数针对一般卫星系统的性能目标标准都包含在ITU-R S.1062建议书中^[14-15]。这些目标表述为时间百分比函数,反映了卫星流量的性能要求。一般而言,误差性能目标视服务要求而定,目前确定误差性能目标很大程度上取决于前向纠错(FEC)模式^[16]。该建议书在传输链路设计方面给卫星系统设计人员提供了指导。

长期性能反映了系统可达到的性能水平,不会因为衰落或干扰而受到影响,而短期性能必然会因为这些因素而下降。容许下降水平取决于卫星连接传输的流量类型。性能下

降的严重程度决定了容许的性能下降持续时间。与FSS系统相比,MSS系统的短期性能目标可能发挥更大的作用,因为其信道条件变化更为频繁,影响更大。目前正制订短期误差性能目标标准,预期将很快完成,且标准中将更多考虑MSS系统特性。

6 结束语

本文讨论了关于MSS无线接口及性能的ITU-R建议书以及正在开展的相关研究。一些主要的建议书涵盖了IMT-2000和IMT-Advanced系统的卫星部分无线接口、多载波和多极化传输、卫星IP系统、及救灾通信相关的MSS系统。最后,还讨论了有关MSS技术的一些未来工作。

ITU-R运用学术界和产业界的先进技术、理念和实验数据制订并改进MSS标准和规范。ITU-R定义的标准将在全球设备和系统无缝连接的实现过程中起到重要作用。卫星和地面的集成/混合天地一体化系统充分利用原来两种网络的优势,展望未来,多数的MSS业务都将基于这种天地一体化系统。IMT-Advanced系统的卫星部分和ITU-R标准化无线接口是集成MSS系统的关键技术,将成为激活MSS产业的新驱动力。

参考文献

- [1] KIM SY, KOTA S S. ITU-R Standardization of Fixed Satellite Services (FSS) [C]// Proceedings of the International Workshop on Satellite and Space Communications 2007, Salzburg, Austria, 2007:247-252
- [2] ITU-R. Recommendation ITU-R M.14570-0. Detailed specifications of the radio interfaces of International Mobile Telecommunications-2000 (IMT-2000) [S]. ITU-R, 2000
- [3] ITU-R. Recommendation ITU-R M.1850-0. Detailed specifications of the radio interfaces for the satellite component of International Mobile Telecommunications-2000 (IMT-2000) [S]. ITU-R, 2010
- [4] ITU-R. Recommendation ITU-R M.1457-11. Detailed specifications of the radio interfaces of International Mobile Telecommunications-2000 (IMT-2000) [S]. ITU-R, 2013
- [5] ITU-R. Recommendation ITU-R M.1850-1. Detailed specifications of the radio interfaces for the satellite component of International Mobile Telecommunications-2000 (IMT-

2000) [S]. ITU-R, 2012

- [6] ITU-R. Recommendation ITU-R M.1850-2. Detailed specifications of the radio interfaces for the satellite component of International Mobile Telecommunications-2000 (IMT-2000) [S]. ITU-R, 2014
- [7] ITU-R. Recommendation ITU-R M.2047. Detailed specifications of the satellite radio interfaces of International Mobile Telecommunications-Advanced (IMT-Advanced) [S]. ITU-R, 2013
- [8] KIM S Y, KIM H W, KANG K S, CHANG D I. Satellite Radio Interfaces for IMT-Advanced System [C]//Proceedings of the WCSP 2014, Hefei, China, 2014
- [9] 3GPP. 3GPP TS 36.211 V10.0.0. 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) Physical Channel and Modulation (Release 10) [S]. Jan. 2011
- [10] TTAT. TTAT.3G-25.212. Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding [S], TTAT, 2009
- [11] MALAGUTI N S, KIM S Y, DAVID W. Standardization of the satellite component of IMT-Advanced in ITU-R [J]. will be appeared at International Journal of Satellite communications and Networking, 2015
- [12] HE Y Z, LIU S Y, GAO Z, CUI G F, WANG W D. The Evaluation of BMSat System Performance [J]. will be appeared at International Journal of Satellite communications and Networking, 2015
- [13] KIM H W, KANG K S, KU B J, CHANG D I, KIM S Y. SAT-OFDM: A Satellite Radio Interface for IMT-Advanced System [J]. will be appeared at International Journal of Satellite communications and Networking, 2015
- [14] ITU-R. Recommendation ITU-R S.1062-4. Allowable error performance for a satellite hypothetical reference digital path operating below 15 GHz [S]. ITU-R, 2007
- [15] ITU-R. Recommendation ITU-R S.614-1. Allowable error performance for a satellite hypothetical reference digital path in the fixed-satellite service operating below 15 GHz when forming part of an international connection in an integrated services digital network [S]. ITU-R, 2005
- [16] KIM S Y, YEO S M, SHIN D S, SONG S S. Estimation of the error performance objectives for satellite communication systems [J]. International Journal of Satellite communications and Networking, 2008, 26 (6):475-488

作者简介



金秀英,韩国全北国立大学教授,ITU-R的4B工作组韩国国际标准化专家,IEEE GLOBECOM和ICC技术委员会成员,2008年IEEE VTC会议卫星系统组联合主席;研究方向为无线通信系统的编码MIMO技术和迭代软检测和解码技术;已发表学术论文100余篇。

基于中断概率的多输入多输出异构网络资源分配

Outage-Probability-Based Resource Allocation for MIMO Heterogeneous Networks

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2015) 02-0048-005

摘要: 为充分提高通信网络的频谱效率,异构网络采用多个微小基站同时在相同频段上支持用户数据传输的方式;为保障系统通信效能,需要有效控制不同数据接口上信号之间存在的相互干扰。提出了采用资源分配方法利用通信中断概率作为衡量不同微小基站数据分支通信性能的准则,通过将中断概率约束条件有效近似,从而使原复杂矩阵优化问题转换为基于标量的几何规划问题,进而极大地降低计算复杂度。数值仿真结果验证了所提方法的有效性。

关键词: 异构网络;多输入多输出;中断概率;资源分配

Abstract: To maximize spectral efficiency in the heterogeneous network, multiple femto base stations concurrently operate on the same frequency bands for terminal-user communication. Outage probability is the performance criterion for transmission optimization, which is necessary to manage interference between different femto tiers. The complicated matrix-value optimization problem is reduced to a scalar-value geometric programming problem by relaxing the outage-probabilistic constraint. Numerical results are provided for our studies.

Key words: heterogeneous network; multiple-input multiple-output; outage-probability; resource allocation

栾天祥/LUAN Tianxiang
高飞/GAO Feifei

(清华大学, 北京 100084)
(Tsinghua University, Beijing 100084, China)

通过减小网络规模、在主干网络中增加微小基站或中继节点等方式能够实现多种通信制式的微小网络协同通信,有效地提高通信系统频谱效率^[5]。然而从运行成本的角度考虑,支持用户端的微小网络往往是由用户自行配置,因此资源分配不受主网络的统一协调控制^[6-7]。同时,由于可供有效通信的无线频谱资源十分稀缺,已有的静态频谱资源分配方法已经不能适应日益增长的通信业务量需求,因此支持同频资源接入的动态频谱管理方法是优化微小网络效能的重要手段^[8]。由于不能应用集中式的资源分配策略,针对用户端的合理资源分配及不同微小基站链路间的干扰抑制是异构网络面临的主要困难^[9]。

为了实现安全可靠的频谱共享以及动态频谱接入,认知无线电技术应运而生^[10]。在异构网络中,需要解决不同微小网络链路间信号干扰问题。基于多天线传输的多输入多输出(MIMO)技术已被用于层间的干扰抑制问题。利用波束成形、干扰避免、天线及用户选择等策略,在链路信道信息已知或部分已知时能够有效地保障层间干扰被控制在指定范

1 研究背景

在4G及未来蜂窝网通信中,为了保障移动用户、特别是网络边缘用户的服务质量(QoS),物理层传输需要采用频带更宽、频率更高的无线电频谱资源,例如TD-LTE(U.S.),采用频段包括2 496~2 690 MHz。在这种高频传输过程中信号衰减成为限制通信质量的重要因素,将极大地影响通信系统的服务能力,例如网络

的有效吞吐量、网络覆盖范围等^[1]。另一方面,随着智能终端系统的普遍应用,施加在主干通信网络上的数据流越来越多地来自室内无线通信设备。这些新兴的无线数据业务应用在已有的网络系统中往往由于业务切换频繁、数据量分布不均匀等原因造成网络拥塞严重、通信网络效能剧减等后果。为了应对网络资源分配日益困难的问题,采用微小基站、拥有更小网络规模的多种网络混合的自适应异构网络技术成为一种经济实用的选择^[2-3]。通过动态的资源管理、分配策略,它能够更有效地提高频谱利用效率,降低网络功耗,增强系统通信效能^[4]。

收稿日期: 2014-11-05

网络出版时间: 2014-12-26

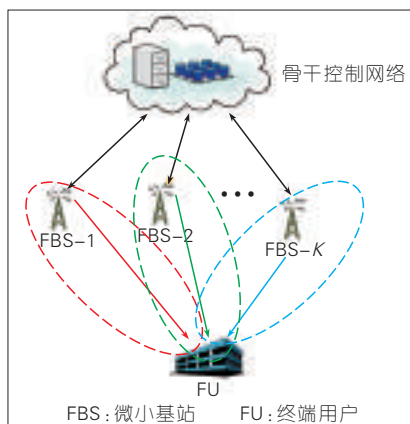
基金项目: 国家重点基础研究发展(“973”)规划(2013CB336600); 国家科技重大专项(2013ZX03004007); 国家自然科学基金(61422109、61201187)

围内^[11-13]。另外,基于正交频分多址(OFDMA)技术的资源分配策略也被用于优化微小网络传输,利用频带的正交性,可以有效避免层间干扰。但OFDMA技术需要较多的频谱资源支持,当可用的无线频段带宽受限时传输策略的性能将被严重影响^[14-15]。已有的资源分配和传输优化方法均依赖于通信链路的信道信息,然而由于异构网系统中往往不存在集中式的资源管理中心,因此不同的微小网络间并无直接协作的通信机制,这使得传统基于同步方式的信道信息估计方法不能有效应用。因此在基于微小基站的异构网络资源分配中,考虑信道信息不完全已知时,多种微小网络通信链路间的干扰抑制及性能保障对于整个系统的性能优化具有重要意义。

本文针对基于多天线传输的多输入多输出异构网络,考虑布设多个微小基站时对多个微小网络进行资源分配。由于不同微小网络间不存在主动协作关系,因此在对某个微小网络进行资源分配时,其他干扰链路的信道信息均不完全已知。论文的关键技术点在于:为了有效抑制频谱共享所造成的链路间干扰、确保网络整体性能的均衡性,采用了通信链路中断概率作为资源分配的约束条件,并提出了基于中断概率分析的异构网络资源分配策略;分析了不同微小网络链路中断概率的上下界,对原概率约束条件进行了合理松弛,将原矩阵优化问题转化为基于标量的几何规划问题;最后,文章通过数值仿真验证了所提方法的有效性。

2 系统模型及问题描述

包含K个微小基站的异构网络下行链路用户数据传输模型如图1所示。考虑异构网络下行链路传输。网络中包括由K个微小基站(FBS-1, FBS-2, ..., FBS-K)支持的微小网络簇,它们各自承载不同的数据业务类型,相互间不存在主动协作关



▲图1 包含K个微小基站的异构网络下行链路用户数据传输模型

系,且使用相同频带资源。它们可以各自把支撑异构网络正常运行所必须的控制参数通过特定控制信道传送到骨干控制网络节点,以便于网络整体资源分配。网络中每个终端用户(FU)具有认知无线电节点的频谱接入能力^[10,16],可以同时接收、处理上述多个微小网络的用户数据信号。

网络中每个节点均为多天线通信节点, FBS- $k(k=1,2,\dots,K)$ 的天线数为 N_k ,FU的天线数为 N 。设第 k 个微小基站至FU的基带信道矩阵为 $H_k \in \mathbb{C}^{N \times N_k}$, $k=1,2,\dots,K$ 。由于不同的微小基站间不存在主动协作,不会进行集中式的传输碰撞避免机制,因此在数据传输时相互间将产生干扰,因此我们需要在传输优化策略中妥善控制这些干扰对链路通信性能的影响。对于第 k 个微小基站,FU接收到的信号为:

$$y_k = H_k C_k s_k + \sum_{n=1, n \neq k}^K H_n C_n s_n + n_k \quad (1)$$

其中 $s_k \in \mathbb{C}^{N_k \times 1}$ 为归一化的第 k 个微小基站至FU的信息码元, $C_k \in \mathbb{C}^{N_k \times N_k}$ 代表相应的波束成形矩阵, $n_k \sim \mathcal{CN}(0, \sigma_k^2 I_{N_k})$ 表示接收端的噪声。对于第 k 条通信链路,它的性能可以利用瞬时信干噪比(SINR)来衡量:

$$\text{SINR}_k = \frac{\text{tr}\{H_k C_k C_k^H H_k^H\}}{\sum_{n=1, n \neq k}^K \text{tr}\{H_n C_n C_n^H H_n^H\} + \sigma_k^2} \quad (2)$$

在公式(2)中对于FBS- k 而言,由于各个微小网络间不存在主动协作,因此它不能够获得各个链路信道信息的瞬时值,只能得到信道信息的统计特性,所以这里假设公式(2)中 $H_n, n \neq k$ 对于FBS- k 是未知的,且 H_n 中各个元素独立同分布为 $\mathcal{CN}(0, \sigma^2)$ 。网络整体的资源分配目标是保障各个微小网络链路性能在有效范围内的同时,力争最小化网络整体能量消耗。

为了有效地保障链路质量,并对网络进行合理的资源分配,我们采用通信中断概率作为衡量准则,可以得到如下异构网络传输优化问题:

$$\begin{aligned} \min_{\{C_k\}} & \sum_{k=1}^K \frac{u_k}{\gamma_k} \text{tr}\{C_k C_k^H\} \\ \text{s.t.} & \text{tr}\{C_k C_k^H\} \leq P_k \\ & \text{tr}\{H_k C_k C_k^H H_k^H\} \geq \nu_k \\ & \Pr\{\text{SINR}_k \leq \gamma_k\} \leq \varepsilon_k, \forall k \end{aligned} \quad (3)$$

其中目标函数为各个微小网络链路功率消耗加权; u_k 表示第 k 条链路功率消耗权重,该值越大所分配的功率趋向于越小; P_k 是第 k 条链路的发射功率限制; ν_k 表示第 k 条链路传输效能的最低值; γ_k 代表第 k 条链路SINR的期望值,该值越大所分配的功率趋向于越大; ε_k 是第 k 条链路为保障自身QoS所能容忍的通信中断概率上限值。

3 资源分配优化方法

从优化问题公式(3)中不难看出,对于整个网络而言,通信资源分配的核心是如何有效地在网络的整体消耗及各个通信链路的性能之间进行折衷。通过求解该优化问题,可获得各个多天线通信节点上最优的波束成形矩阵 $\{C_k\}$ 。

3.1 中断概率上界分析

记 $Q_k = C_k C_k^H$,并记 Q_k 的秩为 r_k ,其特征值分解为 $Q_k = U_k D_k U_k^H$ 。其中 $U_k \in \mathbb{C}^{N_k \times N_k}$ 为酉矩阵, D_k 为半正定对角矩阵,记其对角元素为 $d_{k,1} \geq \dots \geq d_{k,N_k} \geq 0$ 。记

$\varphi_k = \text{tr}\{H_k C_k C_k^H H_k^H\}$, $\tilde{H}_k = H_k U_k$, 那么则有:

$$\varphi_k = \text{tr}\{\tilde{H}_k D_k \tilde{H}_k^H\} = \tilde{H}_k^H \text{vec}(\tilde{H}_k D_k) = \tilde{H}_k^H \tilde{D}_k \tilde{H}_k \quad (4)$$

其中 $\tilde{H}_k = \text{vec}(\tilde{H}_k)$, $\tilde{D}_k = D_k^T \otimes I_{N_k}$ 为对角矩阵, 记其对角元素为 $\{\tilde{d}_{k,j} = d_{k,\lfloor \frac{j}{N_k} \rfloor}, j=1, \dots, N_k N\}$ 。同时, 不难

发现, 由于 U_k 为酉矩阵, \tilde{H}_k 的各个元素也为复高斯独立同分布 $CN(0, \sigma^2)$ 。那么, 第 k 条通信链路的中断概率为:

$$\begin{aligned} \Pr\{SINR_k \leq \gamma_k\} \\ = \Pr\{\tilde{H}_k^H \tilde{D}_k \tilde{H}_k \leq \gamma_k \sigma_k^2 + \sum_{n=1, n \neq k}^K \tilde{H}_n^H \tilde{D}_n \tilde{H}_n\} \quad (5) \\ = \Pr\{\sum_{i=1}^{NN_k} \tilde{d}_{k,i} Z_{k,i} \leq \gamma_k \sigma_k^2 + \gamma_k \sum_{n=1, n \neq k}^K \sum_{j=1}^{NN_n} \tilde{d}_{n,j} Z_{n,j}\} \end{aligned}$$

其中 $Z_{k,i} = \|\tilde{h}_k(i)\|^2$, $Z_{n,j} = \|\tilde{h}_n(j)\|^2$ 。容易得知 $Z_{k,i}$ 及 $Z_{n,j}$ 服从指数独立同分布, 概率密度函数记为 $\lambda e^{-\lambda t}$, $t \geq 0$, $\lambda = 1/\sigma^2$ 。

定义如下随机事件:

$$\begin{aligned} A^{(k)} &= \{\sum_{i=1}^{NN_k} \tilde{d}_{k,i} Z_{k,i} \leq t\} \\ \tilde{A}^{(k)} &= \{\sum_{i=1}^{NN_k} Z_{k,i} \leq t \sum_{i=1}^{NN_k} \frac{1}{\tilde{d}_{k,i}}\} \quad (6) \end{aligned}$$

显然有 $\tilde{A}^{(k)} \supseteq A^{(k)}$, 亦即 $\Pr\{\tilde{A}^{(k)}\} \geq \Pr\{A^{(k)}\}$ 。定义 $\theta_k = \sum_{i=1}^{NN_k} 1/\tilde{d}_{k,i}$, 那么结合公式(5)可以得到:

$$\begin{aligned} \Pr\{SINR_k \leq \gamma_k\} \\ \leq \Pr\{\sum_{i=1}^{NN_k} Z_{k,i} \leq \gamma_k \sigma_k^2 \theta_k + \gamma_k \theta_k \sum_{n=1, n \neq k}^K \sum_{j=1}^{NN_n} \tilde{d}_{n,j} Z_{n,j}\} \quad (7) \end{aligned}$$

容易验证 $X_k = \sum_{i=1}^{NN_k} Z_{k,i}$ 满足 NN_k 阶 Erlang 分布, 即:

$$\Pr\{X_k \leq t\} = 1 - \sum_{m=0}^{NN_k-1} \frac{(\lambda t)^m}{m!} e^{-\lambda t} \quad (8)$$

定义 $Y_k = \sum_{n=1, n \neq k}^K \sum_{j=1}^{NN_n} \tilde{d}_{n,j} Z_{n,j}$, 由式(7)可以得到:

$$\begin{aligned} \Pr\{X_k \leq \gamma_k \sigma_k^2 \theta_k + \gamma_k \theta_k Y_k\} \\ = \int_0^\infty \Pr\{X_k \leq \gamma_k \theta_k (\sigma_k^2 + t)\} d(Y_k \leq t) \\ = \int_0^\infty \left(1 - \sum_{m=0}^{NN_k-1} \frac{[\lambda \gamma_k \theta_k (\sigma_k^2 + t)]^m}{m!} e^{-\lambda \gamma_k \theta_k (\sigma_k^2 + t)}\right) d(Y_k \leq t) \\ \leq 1 - \sum_{m=0}^{NN_k-1} \frac{[\lambda \gamma_k \theta_k \sigma_k^2]^m}{m!} e^{-\lambda \gamma_k \theta_k \sigma_k^2} \int_0^\infty e^{-\lambda \gamma_k \theta_k t} d(Y_k \leq t) \quad (9) \end{aligned}$$

另外, 式(9)最后端的积分可以计算如下:

$$\begin{aligned} \int_0^\infty e^{-\lambda \gamma_k \theta_k t} d(Y_k \leq t) \\ = E\{\exp(-\lambda \gamma_k \theta_k \sum_{n=1, n \neq k}^K \sum_{j=1}^{NN_n} \tilde{d}_{n,j} Z_{n,j})\} \\ = \prod_{n=1, n \neq k}^K \prod_{j=1}^{NN_n} E\{e^{-\lambda \gamma_k \theta_k \tilde{d}_{n,j} Z_{n,j}}\} \\ = \prod_{n=1, n \neq k}^K \prod_{j=1}^{NN_n} \frac{1}{1 + \gamma_k \theta_k \tilde{d}_{n,j}} \quad (10) \end{aligned}$$

结合公式(9)和公式(10)可以得到第 k 条通信链路中断概率的一个上界:

$$\begin{aligned} \Pr\{SINR_k \leq \gamma_k\} \\ \leq 1 - \sum_{m=0}^{NN_k-1} \frac{[\lambda \gamma_k \theta_k \sigma_k^2]^m}{m!} e^{-\lambda \gamma_k \theta_k \sigma_k^2} \prod_{n=1, n \neq k}^K \prod_{j=1}^{NN_n} \frac{1}{1 + \gamma_k \theta_k \tilde{d}_{n,j}} \\ \approx 1 - \prod_{n=1, n \neq k}^K \prod_{j=1}^{NN_n} \frac{1}{1 + \gamma_k \theta_k \tilde{d}_{n,j}} \quad (11) \end{aligned}$$

3.2 中断概率下界分析

根据公式(5)可以得到:

$$\begin{aligned} \Pr\{SINR_k \leq \gamma_k\} \\ = \int_0^\infty \Pr\{\gamma_k \sigma_k^2 + \gamma_k \sum_{n=1, n \neq k}^K \sum_{j=1}^{NN_n} \tilde{d}_{n,j} Z_{n,j} \geq t\} d(\varphi_k \leq t) \quad (12) \end{aligned}$$

定义如下随机事件:

$$\begin{aligned} B^{(k)} &= \{\gamma_k \sum_{n=1, n \neq k}^K \sum_{j=1}^{NN_n} \tilde{d}_{n,j} Z_{n,j} \geq t - \gamma_k \sigma_k^2\} \\ \tilde{B}_{n,j}^{(k)} &= \{\gamma_k \tilde{d}_{n,j} Z_{n,j} \geq t - \gamma_k \sigma_k^2, \forall n, j\} \quad (13) \end{aligned}$$

显然有 $B^{(k)} \supseteq \bigcap_{n,j} \tilde{B}_{n,j}^{(k)}$, 亦即 $\Pr\{B^{(k)}\} \geq \Pr\{\bigcap_{n,j} \tilde{B}_{n,j}^{(k)}\}$ 。那么由公式(12)

可以得到第 k 条通信链路中断概率的一个下界:

$$\begin{aligned} \Pr\{SINR_k \leq \gamma_k\} \\ \geq \int_0^\infty \prod_{n=1, n \neq k}^K \prod_{j=1}^{NN_n} \Pr\{\gamma_k \tilde{d}_{n,j} Z_{n,j} \geq t - \gamma_k \sigma_k^2\} d(\varphi_k \leq t) \\ = \int_0^\infty \prod_{n=1, n \neq k}^K \prod_{j=1}^{NN_n} \exp(-\lambda \frac{t - \gamma_k \sigma_k^2}{\tilde{d}_{n,j} \gamma_k}) d(\varphi_k \leq t) \\ = \int_0^\infty \exp(-\lambda \frac{t - \gamma_k \sigma_k^2}{\gamma_k} \sum_{n=1, n \neq k}^K \theta_n) d(\varphi_k \leq t) \\ = \exp(\lambda \sigma_k^2 \sum_{n=1, n \neq k}^K \theta_n) \int_0^\infty \exp(-\frac{\lambda}{\gamma_k} \sum_{n=1, n \neq k}^K \theta_n t) d(\varphi_k \leq t) \\ = \exp(\lambda \sigma_k^2 \sum_{n=1, n \neq k}^K \theta_n) E\{\exp(-\frac{\lambda}{\gamma_k} \sum_{n=1, n \neq k}^K \theta_n \varphi_k)\} \\ = \exp(\lambda \sigma_k^2 \sum_{n=1, n \neq k}^K \theta_n) \prod_{i=1}^{NN_k} \frac{1}{1 + \tilde{d}_{k,i} \sum_{n=1, n \neq k}^K \theta_n / \gamma_k} \quad (14) \end{aligned}$$

3.3 传输优化算法

利用式(11)对原传输优化问题

(3)中的概率约束条件进行向上松弛, 即当下面的条件满足时, 有 $\Pr\{SINR_k \leq \gamma_k\} \leq \varepsilon_k$:

$$1 - \prod_{n=1, n \neq k}^K \prod_{j=1}^{NN_n} \frac{1}{1 + \gamma_k \theta_k \tilde{d}_{n,j}} \leq \varepsilon_k, \forall k \quad (15)$$

其等价于:

$$\prod_{n=1, n \neq k}^K \prod_{j=1}^{NN_n} (1 + \gamma_k \theta_k \tilde{d}_{n,j})^N \leq \frac{1}{1 - \varepsilon_k}, \forall k \quad (16)$$

对上式左右取对数, 得到:

$$\sum_{n=1, n \neq k}^K \sum_{j=1}^{NN_n} \log(1 + \gamma_k \theta_k \tilde{d}_{n,j}) \leq \frac{1}{N} \log(\frac{1}{1 - \varepsilon_k}), \forall k \quad (17)$$

在各微小网络链路进行数据传输时, FBS- k 获得自身链路信道信息 H_k 后, 为了最优化链路性能, 可根据对角化波束成形方法^[17]确定最佳的传输方向 U_k^* 。令 $H_k^H H_k = \tilde{U}_k \Lambda \tilde{U}_k^H$, 则 $U_k^* = \tilde{U}_k$ 。此时原矩阵变量传输优化问题转化为如下具有实数标量的功率分配问题:

$$\begin{aligned} \min \sum_{k=1}^K \sum_{l=1}^{N_k} u_k \sum_{i=1}^{N_k} d_{k,l} \\ s.t. d_{k,l} \geq 0, \forall k, l \\ \sum_{l=1}^{N_k} d_{k,l} \leq P_k, \sum_{l=1}^{N_k} d_{k,l} \geq v_k, \forall k \\ \sum_{n=1, n \neq k}^K \sum_{j=1}^{NN_n} \log(1 + \gamma_k \theta_k \frac{d_{n,j}}{\tilde{d}_{k,i}}) \leq \frac{1}{N} \log(\frac{1}{1 - \varepsilon_k}), \forall k \end{aligned} \quad (18)$$

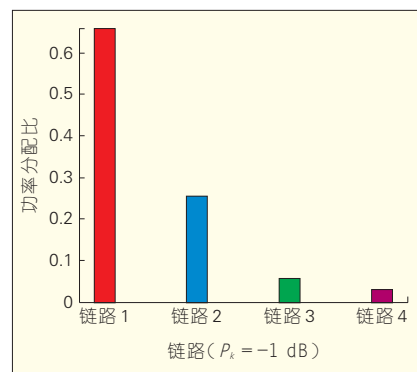
该问题实际上是一类几何规划 (Geometric Programming) 问题^[18], 此类问题可以用标准方式转化成凸优化问题进行求解。由于各通信节点配备的天线数量有限, 标量优化问题, 公式(18)很容易利用数值优化算法求解, 如内点法、基于罚函数的梯度优化算法、凸松弛优化方法等^[19]。

4 数值仿真结果

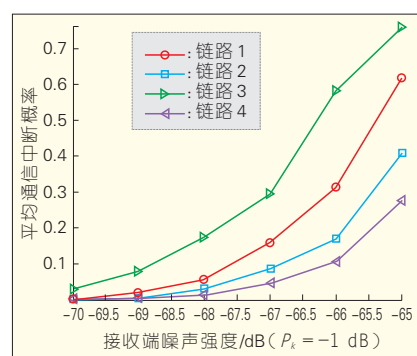
本小节对文中所提传输优化方法进行数值仿真验证, 考虑文中所提的基于微小网络簇的多天线异构网络下行链路传输。微小网络链路个数 $K=4$, 各个微小基站上的天线数 $N_k=2$, FU 上的天线数 $N_u=4$, 信道信息矩阵 $\{H_k\}$ 中各元素分布为 $CN(0, 1)$, 其他相关参数会在仿真实例中给

出。所有仿真数据均取1 000次独立蒙特卡罗试验结果的平均值。

图2和图3分别给出了在复杂应用要求约束条件下,经过网络资源分



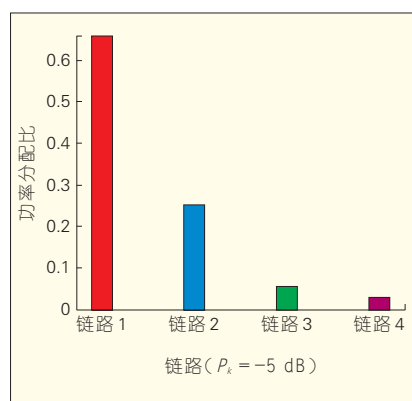
▲图2 各微小网络链路功率分配比例



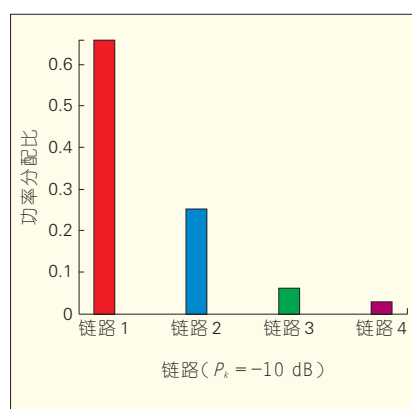
▲图3 各微小网络链路通信中断概率随接收端噪声强度变化曲线

配后各个微小网络通信链路上的资源配置情况及相应的性能体现。仿真中假定各条链路由于承载业务的重要性不同而存在差异,定义相关参数如下:优化目标函数权重为 $u_1=0.1$, $u_1=0.2$, $u_3=0.3$, $u_4=0.4$,各链路所期望的SINR门限值为 $\gamma_1=-15$ dB、 $\gamma_2=-20$ dB、 $\gamma_3=-25$ dB、 $\gamma_4=-30$ dB,微小基站传输功率限制为 $P_k=-1$ dB,资源分配效能限制为 $\nu_k=P_k/100$,各链路中断概率门限值为 $\varepsilon_1=0.2$, $\varepsilon_2=0.3$, $\varepsilon_3=0.4$, $\varepsilon_4=0.5$ 。从图2中功率分配的结果可以看出,微小基站1至4被分配的传输功率顺次递减。值得注意的是,此处资源分配的结果仅仅是异构网络根据各个链路对通信质量的需求以及网络整体资源消耗这两个性能衡量标准间的折衷结果,分配

较高功率的链路并不代表其性能就一定高于其他链路,因为在传输优化中,为了达到整体的效能最优仅需保证最低的性能约束即可。这也恰恰被图3中各个链路的实际平均中断概率曲线所验证。在图3中,被分配最大传输功率的链路1的中断概率并不是整体最低,反而是被分配较低传输功率的链路4具有中断概率最低的整体表现。同时可以看出,随着FU接收端噪声强度的增加,各个链路的平均中断概率都处于上升状态。这说明,随着接收信道状态的恶化,网络整体的通信效能也处于不断降低之中。这种由于外部通信条件所带来的性能恶化最终将导致所提传输优化方法失效,即原优化问题不存在可行解,这体现了所提方法的性能极限。因此如果始终能够保持接收端良好的信道传输条件,即使各个微小基站间不存在直接协助关系,



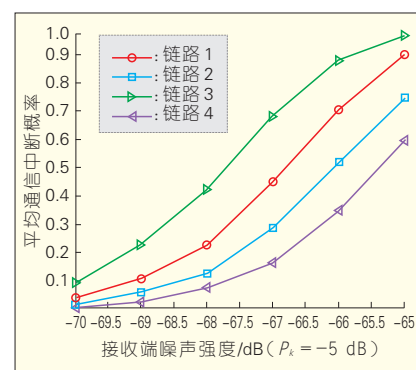
▲图4 各微小网络链路功率分配比例



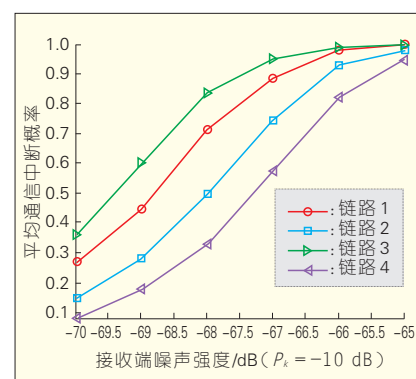
▲图6 各微小网络链路功率分配比例

所提方法仍然能够有效地利用有限的信道不确定性帮助异构网络获得较好的资源分配效果,实现网络性能和资源分配之间的折衷。

图4、图5及图6、图7分别给出了各个微小网络链路传输特性进一步恶化后的性能表现。此时微小基站传输功率限制分别降低为 $P_k=-5$ dB和 $P_k=-10$ dB。可以看出,在功率分配上各个链路获得的资源配比基本没有变化,但是在中断概率中,各个链路的性能曲线都发生了较大变化。从图5和图7可以发现,此时各个链路对接收端噪声强度的敏感性提高了,即和图3相同噪声强度条件下,异构网络的中断概率提高了,网络性能有了较大的下降。这是因为微小基站传输功率限制与接收端噪声强度一样,当传输功率限制苛刻时,会造成所提传输优化问题不存在可行解,从而彻底阻碍所提方法在



▲图5 各微小网络链路通信中断概率随接收端噪声强度变化曲线



▲图7 各微小网络链路通信中断概率随接收端噪声强度变化曲线

网络性能和资源分配间进行折衷,致使部分微小网络链路业务中断。

5 结束语

本文研究了多输入多输出异构网络中的资源分配及传输优化问题。为了充分利用有限的频谱资源,在异构网络中布设多个微小基站,组成微小网络,它们能够同时操作相同频段,并为异构网络中终端用户提供不同类别的数据传输服务。为了降低异构网络的复杂度,不同的微小网络间不存在主动协作,这使得传统的优化方法很难直接应用。为此,本文提出了基于通信中断概率分析的多天线传输优化方法。在考虑场景中,各个微小网络不能够获得其他微小网络的完整信道信息,只能利用信道衰落的统计特性进行传输参数的优化设计;同时在网络整体资源分配中,要确保各个微小网络通信性能不因链路干扰而恶化。所提方法能够有效地在通信链路性能最优化和网络资源分配均衡化之间进行折衷,并且通过对中断概率的上下界分析,对原优化问题中的概率约束条件进行了合理松弛,得到了高效的资源分配途径。最后,文章给出了所提方法的数值仿真结果,较好地验证了所提方法的有效性。

参考文献

- [1] SCHILLER J H. Mobile Communications [M]. Addison Wesley, 2003
- [2] GANZ A, KRISHNA C M, TANG D Y, et al. On Optimal Design of Multitier Wireless Cellular Systems [J]. IEEE Commun. Mag., 1997, 35(2):88-93
- [3] CAVALCANTI D, AGRAWAL D P, CORDEIRO C, et al. Issues in Integrating Cellular Networks WLANs and MANETs: A Futuristic Heterogeneous Wireless Network [J]. IEEE Wireless Commun., 2005, 12(3): 30-41
- [4] 牛志升, 周盛, 周世东, 等. 能效与资源优化的超蜂窝移动通信系统新架构及其技术挑战 [J]. 中国科学: 信息科学, 2012, 42(10):1191-1203
- [5] SANJABI M, RAZAVIYAYN M, LUO Z Q. Optimal Joint Base Station Assignment and Beamforming for Heterogeneous Networks [J]. IEEE Trans. Signal Process., 2014, 62(8): 1950-1961
- [6] CLAUSSEN H, HO L T W, SAMUEL L G. An Overview of The Femtocell Concept [J]. Bell Labs Technical Journal, 2008, 13(1):221-245
- [7] CHANDRASEKHAR V, ANDREWS J G, GATHERER A. Femtocell Networks: A Survey [J]. IEEE Commun. Mag., 2008, 46(9): 59-67
- [8] LIN P, ZHANG J, CHEN Y J, et al. Macro-Femto Heterogeneous Network Deployment and Management: From Business Models to Technical Solutions [J]. IEEE Wireless Commun., 2011, 18(3):64-70
- [9] PEREZ D L, GUVENC I, de la ROCHE G, et al. Enhanced Inter-Cell Interference Coordination Challenges in Heterogeneous Networks [J]. IEEE Commun. Mag., 2011, 18(3):22-30
- [10] ZHANG P, LIU Y, FENG Z Y, et al. Intelligent and Efficient Development of Wireless Networks: A Review of Cognitive Radio Networks [J]. Chinese Science Bulletin, 2012, 57(28):3662-3676
- [11] CHANDRASEKHAR V, KOUNTOURIS M, ANDREWS J G. Coverage in Multi-Antenna Two-Tier Networks [J]. IEEE Trans. Wireless Commun., 2009, 8(10):5314-5327
- [12] WONJAE S, NAMYOON L, WONJONG N, et al. Hierarchical Interference Alignment for Heterogeneous Networks with Multiple Antennas [C]//Proceedings of the 2011 IEEE Int. Conf. Commun. Workshops (ICC), 5-9 June, 2011, Kyoto, Japan: IEEE, 2011:1-6
- [13] TRI M N, QUEKI T Q S, HYUNDONG S. Opportunistic Interference Alignment in MIMO Femtocell Networks [C]//Proceedings of the 2012 IEEE Intl. Symposium on Inf. Theory (ISIT), 2012: 2631-2635
- [14] CHANDRASEKHAR V, ANDREWS J G. Spectrum Allocation in Tiered Cellular Networks [J]. IEEE Trans. Commun., 2009, 57(10):3059-3068
- [15] CHEUNG W C, QUEK T Q S, KOUNTOURIS M. Throughput Optimization, Spectrum Allocation, and Access Control in Two-Tier Femtocell Networks [J]. IEEE J. Sel. Areas Commun., 2012, 30(3):561-574
- [16] DILLINGER M, MADANI K, ALONISTIOTI N. Software Defined Radio: Architectures, Systems, and Functions [M]. John Wiley & Sons Ltd., 2003
- [17] YU Y, HUA Y B. Power Allocation for A MIMO Relay System with Multiple-Antenna Users [J]. IEEE Trans. Signal Process., 2010, 58(5):2823-2835
- [18] BOYD S, VANDENBERGHE L. Convex Optimization [M]. Cambridge University Press, 2008
- [19] PALOMAR D P, ELDAR Y C. Convex Optimization in Signal Processing and Communications [M]. Cambridge University Press, 2010

作者简介



栾天祥, 工作于清华大学自动化系信息技术研究所, 博士; 主要研究领域为信息论、信号处理、机器学习及优化理论, 以及它们在无线通信系统中的应用。



高飞, 清华大学自动化系信息技术研究所副教授、博士生导师, IEEE Senior Member, 中国电子学会信息论分会委员, 中国通信学会青年工作委员会协作委员; 主要研究领域为信号处理在无线通信网络中的各类应用; 已主持和参加基金项目近 10 项; 已发表 SCI 检索 70 余篇, 其中 EI 检索 90 余篇, SCI 他引 500 余次。

综合信息

2014 年全球光纤出货量达 2.96 亿芯公里 同比增长 13%

根据市场研究机构 CRU 公布的统计数据: 2014 年全球光纤出货量达 2.96 亿芯公里, 同比增长 13%。这一数据略低于此前预期的超过 3 亿芯公里。CRU 认为, 主要增长区域是新兴市场, 比如非洲和东南亚。这也直接导致了光缆贸易量相比全球市场增长了 29%。

中国不再被划归到新兴市场范畴。经过连续 6 年的高速增长, 中国光纤市场已经十分成熟, 2014 年出货

量占到全球近一半。分析人士指出, 中国市场仍将保持高位运行, 2015 年将在 2014 年庞大出货量的基础上小幅增长。

展望 2015 年, 中国电信和中国联通势必将增加资本开支, 以迅速扩张其 FDD 网络, 除了基站互联直接产生对光纤的需求外, 同时也带动骨干网、传输网的新建和升级, 这都将维持中国光纤市场的景气度。

(转载自《中国信息产业网》)

基于 MPI 和 OpenCL 多层次并行图像卷积算法设计

Design of Multi-Level Parallel Image Convolution Algorithm Based on MPI and OpenCL

中图分类号: TN919.8 文献标志码: A 文章编号: 1009-6868 (2015) 02-0053-003

摘要: 通过对图像卷积算法的分析,发现算法在对图像处理的过程中具有很高的并行性。提出了一种结合异构开发框架开放运算语言(OpenCL)和并行开发库消息传递接口(MPI)的算法,在支持图形处理器(GPU)的异构集群环境下设计并实现了图像卷积算法的多层次并行实现,使得算法在处理速度上有了显著的提升。

关键词: 异构集群; 消息传递接口; 开放运算语言; 图像卷积

Abstract: The image convolution algorithm has high parallelism in image processing. This paper proposes an algorithm based on heterogeneous development framework open computing language (OpenCL) and parallel development libraries MPI(Message Passing Interface). In the support of the graphics processor (GPU) in heterogeneous cluster environment, the algorithm designs and realizes the parallel implementation of multi-level image convolution, significantly increasing processing speed.

Key words: heterogeneous cluster; message passing interface; open computing language; image convolution

王继刚/WANG Jigang¹

刘惠/LIU Hui²

姜滨/JIANG Bin³

(1. 中兴通讯股份有限公司, 四川 成都 610041;

2. 西安电子科技大学软件学院, 陕西 西安 710071;

3. 北京趣拿软件科技有限公司(去哪儿网), 北京 100080)

(1. ZTE Corporation, Chengdu 610041, China;

2. School of Software, Xidian University, Xi'an 710071, China;

3. Qunar.com, Beijing 100080, China)

卷积运算在图像处理领域有着广泛的应用,图像卷积^[1]运算实质是一种矩阵运算,其特点是运算量大,并且数据复用率高,用软件计算图像卷积很难达到实时性的要求^[2]。

目前,全球许多学者专家对提高图像卷积速度进行了研究,提出的提高卷积速度的方法主要有:

(1) 利用图形处理器(GPU)进行加速^[3-4]。该方法主要利用GPU相对于CPU对图像处理的优点,利用GPU加速图像计算的过程,但是该方法并未考虑到GPU处理能力的极限会影

响图像卷积的加速。

(2) 利用多核特性进行并行加速。如果处理器具有多核,那么可以利用处理器的多个核并行处理加速卷积计算,但是处理器核的数量限制了并行计算的效率^[5]。

(3) 利用单指令多数据流(SIMD)技术加速^[6]。SIMD处理机的数据缓存系统数据极高,能够解决图像卷积中的数据复用带来的CPU重复访问主存储器的问题,从而加速图像卷积。

本文提出结合异构开发框架开放运算语言(OpenCL)^[7-8]和并行开发库消息传递接口(MPI)^[9]相结合的方法,在异构集群上实现了图像卷积算法,利用异构集群平台的处理能力和

GPU在处理图形图像上的优势来提升算法的性能。OpenCL支持多种层次并行,可以高效映射到同构或异构的体系结构上,比如单个或者多个CPU、GPU或其它类型的设备等^[10-11], OpenCL提供了基于任务分割和数据分割的并行计算机制。MPI标准定义了一组具有可移植性的编程接口,各个厂商或组织遵循这些标准接口实现自己的MPI软件包,调用这些接口,连接相应平台上的MPI库,就可以实现基于消息传递的并行计算。

1 图像卷积算法分析

在图像卷积算法中,每个像素点利用邻接像素点的信息修改自身的值。一个卷积滤波器描述了每个像素点利用它相邻的点改变自身值的方法。例如,一个图像模糊化滤波器将每个像素点设置为其相邻点的加权平均值,从而减小像素之间的差

收稿日期: 2014-05-02

网络出版时间: 2014-10-12

基金项目: 国家自然科学基金(61100075、61272456); 中兴通讯产学研基金(CON1209270011)

值。对相同的源图像,仅仅通过改变滤波器,就可以得到锐化、模糊、边缘增强、浮雕等效果。

根据数字图像卷积定理可以推导出卷积的数学模型。设卷积滤波器为对称 $N \times N$ 矩阵。每像素为二进制顶点数,则每一个像素点的数学模型为:

$$V_{x,y} = C \sum_{i=1}^N \sum_{j=1}^N k_{ij} a_{x+i-\frac{N-1}{2}, y+j-\frac{N-1}{2}} \quad (1)$$

公式(1)中, $a_{x,y}$ 表示输入图像的像素点, k_{ij} 表示滤波器中的像素点, C 表示常数。

对于图像卷积,就是将卷积模板放在图像点阵的左上角,则卷积模板必与图像点阵左上角的分割矩阵重合,把重合项对应相乘后求和,便得到了第一个结果点。然后,再将卷积模板右移一列,即可求出第二个结果点。卷积模板在图像点阵中遍历一遍,就可以求出一帧图像的卷积。

经过对算法的研究发现,用滤波器处理源图像上每个像素的过程中,对每个像素的处理是独立的,可以采用单指令多数据流(SIMD)并行编程模型并行实现来提高处理速度。

2 算法设计

2.1 基于 OpenCL 的并行算法设计

在异构集群中,对于那些配置有 GPU 计算设备的计算节点,部署到其上的应用软件不但应该充分利用 CPU 计算能力,更应该充分挖掘 GPU 强大的并行能力。故本节首先阐述对于图像卷积应用,如何对其进行 GPU 上的并行优化设计与实现。

OpenCL 为程序员提供了控制并行计算设备的一些接口以及一些控制运算单元行为的类 C 编程语言。利用 OpenCL 接口可以开发并行度高、运行在 GPU 或者其他处理设备上的程序。OpenCL 的工作模型如图 1 所示。

根据卷积算法的特点,使用

OpenCL 实现卷积算法可分为以下几个步骤。算法执行的前提是上下文(Context)和命令队列(Queue)都已经创建好,源图像、输出图像和滤波器均已在主机端初始化,每个图像都有 Width 和 Height 值。

步骤 1: 创建 OpenCL 的存储器对象,即图像(Image)对象和缓存(Buffer)对象

OpenCL 的存储器对象分为两种类型,一种是缓存(Buffer)对象,另一种是图像(Image)对象;缓存(Buffer)对象存放的是一组一维的元素集;而图像(Image)对象存放一个二维或三维的纹理或帧缓存或图像;Buffer 对象中的数据可以通过一个指针任意获得;而 Image 对象的数据则是需要通过相应的 API 进行访问;存储器对象通过作为对内核函数的实参被 OpenCL 内核函数操作;

创建 Image 对象的格式描述符 `cl_image_format`, `cl_image_format` 用来定义每一个像素数据的大小、类型、通道和通道顺序等。图像中每个像素都存储在 4 个通道中,每个通道分别代表 RGBA (代表 Red (红色) Green (绿色) Blue (蓝色) 和 Alpha 的色彩空间)中的一位。如果图像的每个像素有 4 个值,则可以使用 `CL_RGBA` 作为通道的顺序,如果每个像素只需要一个值(比如灰度图或者矩阵的元素),这些数据只需要一个通道就可以了,通道顺序可以采用 `CL_R` 表示。

创建 Image 对象的格式描述符后,通过 `clCreateImage2D()` 函数创建一

个 Image 对象,此外还创建一个滤波器所需要的 Buffer 对象,这个 Buffer 对象最终将采用常量内存。

步骤 2: 写输入数据

通过提供一个三维的偏移和范围并调用函数 `clEnqueueWriteImage()`, 将一个 Image 对象拷贝到设备端,利用 `clEnqueueWriteBuffer()` 函数将滤波器也拷贝到设备端。

步骤 3: 创建采样(Sampler)对象

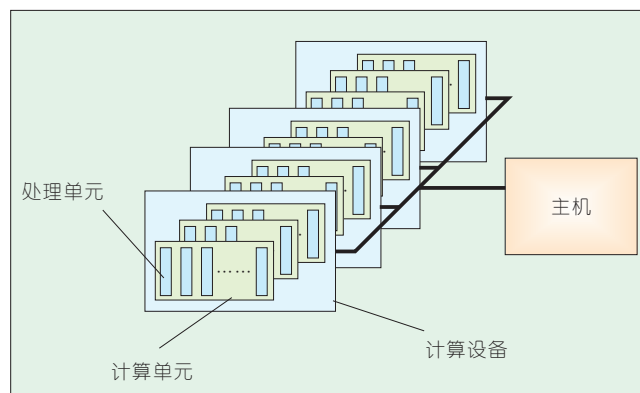
通过函数 `clCreateSampler()` 创建 Sampler。Sampler 描述了如何访问 Image 对象,还通过参数 `CL_TRUE` 将坐标系设置为标准坐标系,并设置越界访问的处理方式为 `CL_ADDRESS_CLAMP_TO_EDGE`, 当访问越界时返回图像的边界值。当访问位置处在多个坐标值之间的时候,滤波器模式可以访问最近的一个坐标值,也可以利用周围的坐标值进行插值。

步骤 4: 编译并执行内核(Kernel)函数

Kernel 特指那些在 GPU 上执行的程序段或函数。创建 kernel 对象,设置 kernel 参数,并编译 kernel 函数。创建和图像像素相同数量的 work-item, 每一个 work-item 都会根据数据类型从原始 Image 读取数据并执行该卷积的数学表达式。

在本文实现中,使用的是 `read_imagef()` 函数,数据类型声明为 `float4` 类型。写入输出 Image 对象也使用一个相应的函数 `write_imagef()`, 数据类型也为 `float4` 类型,写入不支

图 1
OpenCL 工作模型



持越界访问。在 NDRange 的任何一个维度,如果 work-item 数量可能超出像素的数量,则必须在写入输出数据之前进行边界检查。

由于所有 work-item 在每一次迭代中都访问滤波器的同一个元素,因此可以将滤波器放置在常量内存中。

步骤 5:写回结果

调用 clEnqueueReadImage() 函数,将结果写回主机端。

2.2 结合 MPI 和 OpenCL 的并行实现

除了在单 GPU 节点上的并行优化,本文在异构集群上进一步对图像卷积应用进行并行优化。本文使用 MPI^[8-9]库来实现异构集群中各个节点之间的进程间通信。本文所使用的集群为异构集群,集群中有部分节点带有 GPU。

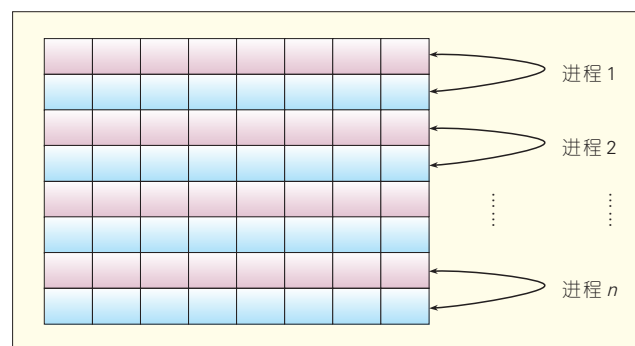
经过对图像卷积算法过程进行分析可知,由于是对每个像素点进行相同的计算,可以对源图像进行分块操作。在纵向上按照进程数将源图像分割,在集群的每个计算节点上并行进行图像卷积算法,且在具有 GPU 的节点上将对应的图像块进行进一步的异构并行优化。异构并行处理过程如图 2 所示,进程数量为 n ,进程 1,2,..., n 分别处理一些图像块,这些图像块在异构节点上基于 OpenCL 进行并行计算。

当处理完成后,除主进程外的其余进程将处理后的图片信息通过 MPI 接口传递给主进程,主进程负责将最终的全部处理后的图片信息汇总并写回硬盘。

3 性能分析

本文所述算法实现的实验环境为 Rocks^[12]+Torque+GPU+CPU 异构集群环境。实验中,软件平台采用 Rocks 操作系统以及 TORQUE 集群管理软件^[13-14]。硬件平台包括一个主节点,若干计算节点,节点间通过局域网连接。集群中的计算节点都配置有高性能的 GPU 卡。其中,TORQUE

图 2
基于 MPI 和 OpenCL 的并行设计原理



集群管理软件是一个控制批作业运行和分配计算节点的集群管理与调度软件,用于管理和监控集群中每个计算节点的工作负载,工作状态,作业的调度等;它是开放源代码软件 OpenPBS 集群管理软件的改进版。TORQUE 的内部调度模块采用 Maui 作业调度器^[15],Maui 作业调度器根据事先定义的集群调度策略决定等待队列中的任务在什么时间和哪些节点上运行,能够在满足不同用户需求和最大化整个集群利用率之间达到动态平衡。TORQUE 和 Maui 的结合可以大大提高作业调度的性能。

在测试算法性能时,本文采用 3 张分辨率分别为 600×400、2 560×1 920、3 456×5 184,位深度均为 8 的图片,分别测试在串行(常规算法)、仅使用 OpenCL(单 GPU 节点上并行图

像卷积算法)和使用 OpenCL+MPI(异构集群多级并行图像卷积算法)3 种处理方式下的运行时间,得到时间对比,具体收集的实验数据对比如图 3 所示。

测试 3 幅图片在 3 种处理方式下的执行时间,其中图片一的像素为 600×400,位深度为 8;图片二的像素为 2 560×1 920,位深度为 8;图片三的像素为 3 456×5 184,位深度为 8。算法执行时间如表 1 所示。

由实验结果可知,随着待处理图像像素数量的增加,并行加速的效果明显增加。当不考虑 MPI 时,OpenCL 并行代码加速比分别为 13.04、25.58、26.92,从数据中可以看出,当数据量越大加速比也就越明显。因为在 OpenCL 实现中,将数据由主机端传

►下转第 62 页

图 3
实验数据对比

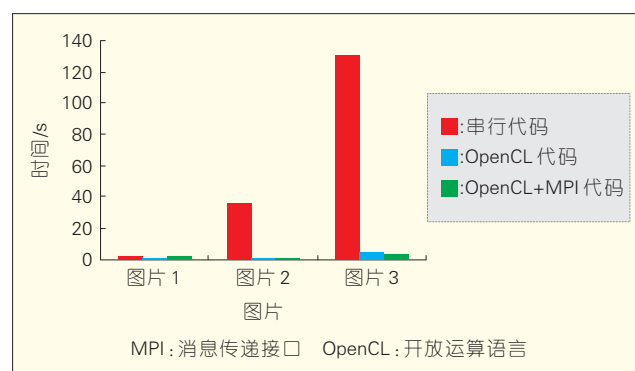


表 1 算法执行时间

图片	串行/ms	OpenCL/ms	MPI+OpenCL/ms
图片一	1 826	140	1 765
图片二	36 317	1 420	1 560
图片三	131 142	4 872	3 465

M-ICT时代的业务智能化发展趋势

Business Intelligence in the Era of M-ICT

董振江/DONG Zhenjiang
赵培/ZHAO Pei
杨勇/YANG Yong

(中兴通讯股份有限公司, 深圳 518052)
(ZTE Corporation, Shenzhen 518052, China)

云计算和政企业务领域在 M-ICT 时代面临大变局, 全面变革时机已经到来。业务领域整体呈现出智能化的大趋势, 这得益于: 智能终端性价比飞速提升与日益普及; 传感器及网络遍布, 并成为智能终端的标配, 产生丰富的传感数据, 社交网络贡献了海量社交数据, 各种网络和系统运行生成的数据, 与各类科学数据一起催生了大数据; 云计算使得计算和存储资源的成本大降, 为大数据存储与计算奠定了基础; 带宽持续增加, 使得数据传输时延和费用不再成为“瓶颈”; 大数据技术日益成熟, 使大数据从专业研究人员的领地, 转变为企业和个人日常工作。各种技术相互促进, 对传统业务与服务的提供产生了日益深远的影响, 甚至是颠覆, 并带动了相关行业的转型^[1-11]。

1 云计算基础设施

1.1 从软件定义存储、计算到软件定义数据中心

云计算数据中心成为新一代 IT

收稿日期: 2015-02-05
网络出版时间: 2015-02-22
基金项目: 国家科技重大专项
(2012ZX03002004-2-1)

基础设施, 据 IDC 预测到 2020 年, 超过 50% 的中国公司一半以上的 IT 资产将在第三方的软件定义数据中心, 超过 60% 的企业级存储能力将转移到云服务提供商。目前这一趋势还

在高速增长中, 且呈现如下趋势:

(1) 软件定义数据中心提升了运营的效率和灵活性

不同企业对数据中心的需求不同, 且需随业务变化而动态调整, 由

中图分类号: TN915; TP393 文献标志码: A 文章编号: 1009-6868 (2015) 02-0056-007

摘要: 认为在 M-ICT 时代, 在云计算和政企业务领域总体趋势为: 首先, 发展方向从软件定义的存储, 到软件定义计算, 到软件定义的数据中心; 其次, 微模块正成为数据中心部署的主流模式, 而 OpenStack 主导数据中心的的管理方式; 第三, 云计算技术正在重塑容灾备份方案, 私有云与公有云融合的混合云正成为政企云建设的重要模式, 而分布式块存储引领企业存储。认为因为云计算的发展和成熟, 使得大数据和人工智能得以快速发展, 从而进一步促进了 ICT 融合业务的创新: 首先, Hadoop 与 Spark 开源大数据平台成为业界主流, 内存和流计算成为热点, 针对特定场景的专有计算方式不断涌现; 其次, 智能问答成为人工智能的热点, 实用性强前景广阔, 而机器智能与行业应用深度结合使基于统一基础平台开发定制成为趋势。认为基于 WebRTC 的轻量级的企业协同走向成熟, Web 技术的发展促使 ICT 融合新业务发展格局正在不断改变并持续前行。

关键词: 软件定义存储; 软件定义数据中心; 网络功能虚拟化; 人工智能; 大数据; 基于网页的实时通信

Abstract: In the M-ICT era, cloud computing and enterprise businesses are undergoing great changes. The overall trends are described as follows. First, with the development of software-defined computing and storage, the software-defined data center is developing rapidly. Second, the micro-data-center model is becoming the mainstream in IDC deployments, and OpenStack is leading cloud management. Finally, cloud computing is reshaping solution related to disaster and recovery. Hybrid cloud is becoming an important mode of government and enterprise business, and distributed block storage is leading enterprise storage. At the same time, it is precisely because of the development and maturation of cloud computing, big data and artificial intelligence that innovative ICT services are being integrated. Open-source platforms such as Hadoop and Spark are becoming mainstream, and which focus on memory and stream computing, while a lot of proprietary methods of calculation for a particular scene are emerging greatly. Second, artificial intelligence is becoming more popular because of its wide practical prospects. Deep combination of machine intelligence and industry business based on a unified platform is becoming a new trend. In addition, with the mature of WebRTC in enterprise collaborative application, the new web technology is promoting the rapid change and development of the new ICT convergent service, and moving it forward continuously.

Keywords: software-defined storage; software-defined datacenter; network function virtualization; artificial intelligence; bigdata; Web real-time communication

此软件定义的数据中心应运而生,使得云计算数据中心从粗放的静态架构转至灵活配置的动态架构。数据中心的网络、计算和存储等各个方面都可被软件定义,快速地调整规模,动态组合网络、存储、计算等各种资源,为最终用户提供灵活定制的虚拟数据中心,达到资源利用的最优化。

自动化部署、动态调度、自服务运维和灵活调试是保障软件定义数据中心的关键,也是影响数据中心运营效率和灵活的核心,将数据、计算、存储与网络调度调整到最佳,需要持续的技术和实践积累。软件定义网络/网络功能虚拟化(SDN/NFV)是网络虚拟化的关键,可真正为数据中心提供高效灵活的网络部署。数据中心通过丰富地应用编程接口(API)调用对外提供云服务。

对外服务的形态类似于虚拟专用交换机对传统用户交换机的替代,不仅提供了更大的灵活性,而且简化了用户的运维成本,扩展性和动态性得到根本的提升。这会成为未来企业数据中心建设的主流方向之一。其中安全的保障是关键。

(2) 软件定义服务器

传统的服务器硬件配置相对固定,通过软件虚拟化对外提供服务。服务器发展的一个重要趋势是:软件定义的服务器,采用分离I/O架构,将CPU计算资源、内存资源、存储资源、网络资源集中,通过高速总线互联,可灵活地将CPU、内存、网卡等资源动态组成各种规格的服务器单元,完成硬件层次的虚拟化组合,相对现有虚拟化技术性能大为提升,这将成为高端服务器的新形态。

(3) 软件定义存储打破传统存储管理的限制

软件定义存储(SDS)通过存储硬件和存储控制软件分离,打破传统存储管理的局限,解决传统存储存在的共享困难、难以扩展、系统性能无法统一优化等根本问题,支持面向应用的、异构存储资源的统一管理,提供

高效数据服务。

存储技术的发展趋势如图1所示。软件定义存储的数据和控制分离,控制平面可集中管理各种异构存储功能,具有如下特点:

- 专业面向应用、基于策略的存储服务,满足不同应用的需求。
- 提供一致、自动化的管理方法,清晰的应用边界,保证服务质量。
- 弹性扩展,实现最优化的资源利用。
- 简化异构存储设备管理,便于扩展支持新的存储类型和应用。

软件定义存储使得异构设备统一使用,可充分利旧,设备不在局限于特定供应商。适应中国国情的软件定义存储供应商将会在这次浪潮中胜出,占得先机。

1.2 新型设备发展趋势

(1) 基于分布式块存储的 Server SAN 融合存储崛起,带动企业外部存储的变革

当前政企客户的存储系统大多基于专用磁阵、存储区域网络(SAN)设备构建,成本高、系统扩展性有限,计算和存储之间采用专用网络连接,易出现性能访问“瓶颈”。基于分布

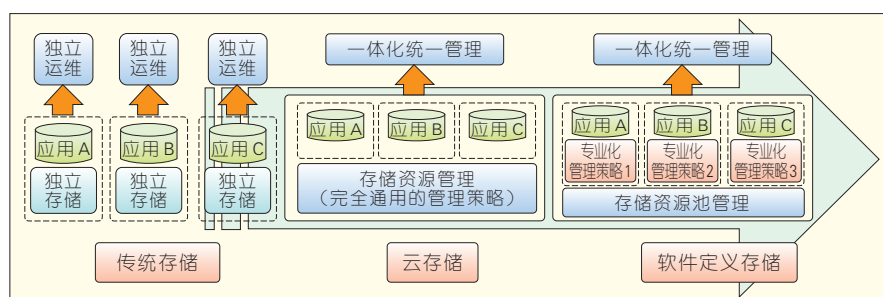
式块存储的服务器存储区域网络(Server SAN),具有弹性高、部署灵活、性价比高等优势,正成为企业存储的主流选择,这是中国自主存储发展的绝佳机会,有利于打破传统存储国际垄断的格局。

基于计算服务器内置磁盘和固态硬盘(SSD)存储,构建弹性块存储 Server SAN,融合计算与存储,系统性能随节点规模线性扩展,降低了系统整体总体拥有成本(TCO)和能耗。据 Wikibon 预测 2013 年全球企业外部存储市场超过 330 亿美元,2017 年 Server SAN 市场份额将超越传统直接附加存储/存储区域网络/网络连接式存储(DAS/SAN/NAS)。不同存储类型优劣势比较如表1所示。

(2) 定制服务器与通用服务器并存,并成为服务器的重要发展分支

采用CPU架构通用服务器适用性强,能满足大部分场景需求,尤其是复杂逻辑控制的场合。但在大数据分析、数据仿真、深度学习等特定应用场景下,效率并不高。

在大量多级浮点计算和逻辑判断的复杂运作的场合,如并发码流格式转换、大数据仿真等,采用GPU计算能力是CPU的几十倍,性能大为提



▲ 图1 存储发展趋势

▼ 表1 不同存储类型优劣势比较

存储类型	优势	劣势	容量
DAS	简便,性能较高	数据共享困难	10 TB
SAN/NAS	性能较高,数据共享方便	扩展受限、成本高	100 TB
云存储	扩展性高,成本低	传统企业应用支持一般,主要用于数据备份	PB ~ EB 级别
企业 ServerSAN	扩展性高,成本低	功能尚待丰富和加强	TB~PB 级别
DAS: 直接附加存储 NAS: 网络连接式存储 SAN: 存储区域网络			

升。深度学习场景需要大量简单并发运算,计算复杂度不高、计算量大,如矩阵计算采用FPGA的服务器,单位功耗计算能力为GPU的20倍,效率更高。针对上述典型场景,采用图像处理器/现场可编程门阵列(GUP/FPGA)的专用服务器效率要比通用服务器高得多。在大数据、机器学习/深度学习等应用日益普及的今天,研发采用GPU/FPGA的专用服务器前景广阔。

(3)全闪存阵列取代磁盘阵列将是未来的趋势

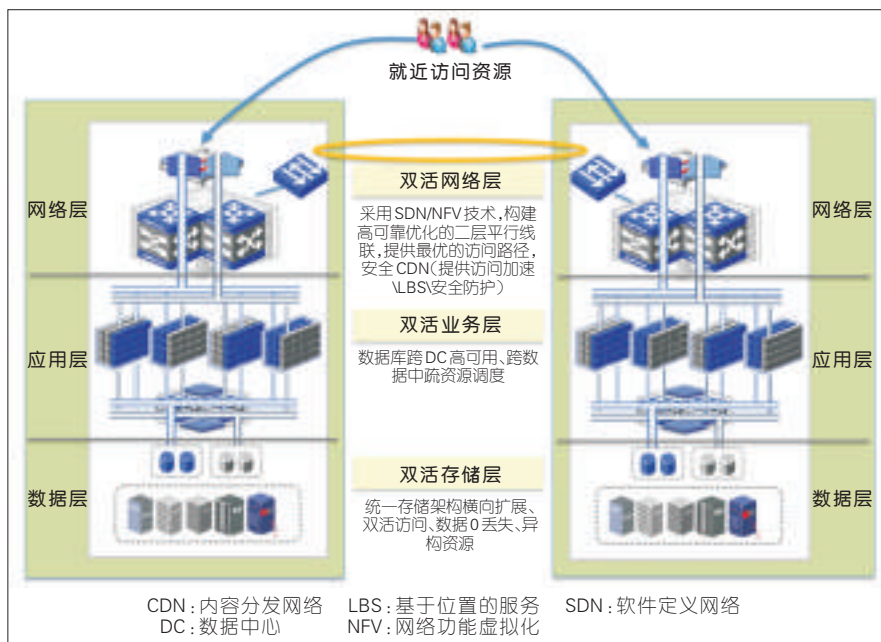
SSD硬盘最大的优势就是IO性能,随机读写时无磁盘的寻道开销,平均性能超过硬盘7~8倍多。在全闪存阵列引入重复数据删除技术,将数据分块,采用哈希算法形成数据指纹,将这些数据指纹映射到内部物理地址存储,在线重复数据删除,能大大降低存储空间。一个域内经过数据重删,存储空间可以缩小为原来的1/8左右,即一个30TB的SSD磁阵,可实现超过250TB有效容量,价格与同容量的普通磁阵相当。这样,全闪存阵列取代普通磁盘阵列就有了可能,将很快大规模进入数据中心。

1.3 云计算双活数据中心保障系统的高可用性

数据中心中不同的服务对可靠性的要求不同,传统的电信域是建立在高可靠的小型机等硬件的基础上,以主备或者热备的小型机自身的可靠性来保障系统的可靠性,这种方式成本高、扩展性能差。在云计算环境下全新的思路是构建双活数据中心,通过引入SDN/NFV等技术构建双活网络、跨数据中心构建双活业务、统一存储构建统一的双活存储,能很好地提升服务高可靠性,满足金融、政府、互联网等行业的高端高可靠性服务需求。

双活数据中心架构如图2所示。图2中各层介绍如下:

- 双活网络层,保障网络层次的



▲图2 双活数据中心架构

高可用性。在网络层移植SDN/NFV技术,提供高可靠优化的二层互联,保障最优的访问路径;同时安全CDN提供访问加速、负载均衡与安全防护功能,构建双活网络层。

- 双活业务层,保障应用的高可用性。数据库跨数据中心高可用,业务层跨数据中心进行资源调度,保障应用的就近访问,从而构建双活的业务层。

- 双活存储层,保障存储的高可用性。双活数据中心采用统一的存储架构,Scaleout架构保证系统根据需求动态横向扩展,支持双活访问,确保数据0丢失高可用,系统支持异构存储资源,统一管理和调度。

本地双活数据中心加上异地中心互为备份,可构建高可靠性高容灾的数据中心。这一模式正改变传统容灾备份中心的建设方式,在金融领域已开始实用,将逐步成为未来数据服务高可靠性高可用性的新模式。

1.4 云数据中心部署与管理的趋势

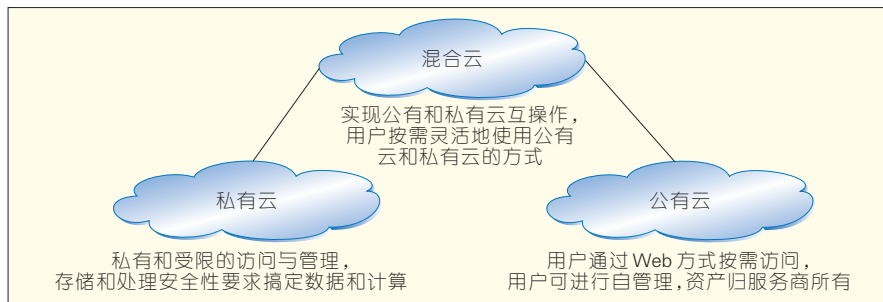
(1)创新的微模块正引领数据中心的建设模式

传统数据中心建设一次投入多、

周期长,风险大,业界正在探索新的数据中心建设模式。其中新型一体化微模块数据中心成为典型。微模块数据中心是将数据中心的机架、空调、消防、布线、配电等子系统集成设计与封装,集成高压直流、行间制冷与能效管理,最简化与最小化工程实施方案,使得安装调试极为简捷,交付周期可由1~2年缩短为2~3个月,对场地要求低:仓库、车间、板房均可用作机房,可整体搬迁再利用,不同模块化数据中理上完全独立,可根据实际需求分期建设,降低一次投资成本与风险,降低了TCO。中兴通讯与中国联通、腾讯建设了亚洲最大的微模块数据中心,正成为数据中心建设新模式。

(2)面向政企的混合云方案

混合云方案如图3所示。在安全技术和安全立法都不够完备的当下,混合云会成为企业利用数据中心的主流形式之一,而且会长期存在。企业用户建设私有云用于核心数据和业务的存储与处理,将安全性要求适中的放到公有云;同时从安全性考虑可以利用公有云做双活容灾,合理的构建私有云与公有云结合的混合云,



▲图3 混合云方案

是未来企业云发展的趋势和热点。

混合云的架构、私有云与公有云间的弹性扩展、按需资源定制、灵活的管理与部署、开放的 API、动态可测量等是混合云方案成功的关键。

针对政企市场，混合云解决方案还需解决企业现有设备的利旧，方便实现异构资源池管理，应对不同的应用场景。

(3) 云端协同空间大

云计算作为基础设施成为计算和存储的中心，终端侧呈现出多样性，随着带宽的扩展，如何协调端与云的协同计算，提供更好的业务体验，需要重点解决如下技术问题：

- 解决跨设备的应用程序的可移植性。
- 数据与应用在多种设备的同步，在协同数据与计算时，要做好数据传输与计算之间成本的平衡。
- 一致性体验。
- 云端协同另一个新的发展趋势是充分利用终端侧的计算能力，将计算分散到不同的终端。

(4) OpenStack 引领数据中心的管理模式

开源 OpenStack 引导了数据中心管理的发展趋势，已成事实标准，其核心模块日渐成熟，管理的范围变宽——从基础设施即服务(IaaS)层向平台即服务(PaaS)、软件即服务(SaaS)层延伸。目标是实现资源自动调配，提升云数据中心服务的透明度、运维效率和交付的敏捷性与控制力。发展重点将是：

- 开放、异构、跨数据中心统一

的云服务与调度管理。资源调度策略不仅体现在资源创建阶段的灵活分配，而且实现资源使用过程中的动态调度。

- 资源管理从虚拟资源拓展到裸机的部署与管理。

- NFV 持续支持，如非统一内存访问 NUMA、数据层开发套件 DPDK、资源预留、性能优化等。

- 应对更大规模的分布式部署场景，保障企业级的高可靠性、高可用性和高扩展性，提升系统性能。

- 可信与安全加固。

- 易用性提升。自动化部署、在线升级与监控、简捷的企业级系统集成，帮助客户简化管理、降低成本，顺应云化和数据流动的需求。

- 从架构、功能、流程等方向不断适应混合云的发展方向，实现跨私有云、公有云和混合云服务生命周期管理，支持统一存储。

2 智能大数据处理发展趋势

数据年复合增长率超过 50%，每年计算成本以 33%、存储成本以每年 38%、传输成本以每年 27% 的速度在快速下降，为实施数据处理打下了坚实的基础。通用数据只有 34% 有用，7% 被标识，1% 被分析，数据价值远未发挥出来，数据挖掘分析的需求在快速增长中，呈现 4 个技术趋势。

2.1 开放开源的 Hadoop/Spark 架构成为大数据平台的主流且专有计算涌现

开源 Hadoop 成为大数据计算处

理的事实标准，其核心思想是分布式与经济性。Hadoop2.0、Spark、Storm 和 AMPlab 等相融合，结合云计算，构建了大数据生态系统，成为业界主流。

- 混合计算模式。Hadoop 与多种计算模式共存的混合计算模式是多样化大数据处理常态，分别应对准实时批处理、非实时批处理、在线分析处理(OLAP)和在线事务处理(OLTP)应用四类应用。采用统一的架构适应不同的场景需求。

- 内存计算与流计算。数据价值与时间成反比，时间越长价值越低，数据处理的实时性至关重要，随着内存容量扩大与成本的持续降低，内存计算成为实时大数据处理的主要手段。Spark 是内存计算开源主导者，生态环境相对完善，国际上 Cloudera、Hortonworks 等公司与中国的中兴通讯等提供通用的服务。

- 新计算模式将涌现。Spark 并非大数据平台发展的全部，业界也在涌现针对新的计算模式，他们针对流计算等上面具有理念上的革新，值得关注和跟踪研究。

2.2 分布式数据库发展迅猛

针对海量数据的分析，特别是非结构化数据的分析，以开源的 NoSQL 为主，已经成了业界的共识和方案的必选，取得了非常好的效果。

海量事务型处理，面临大变革。Oracle/DB2 等数据库是传统事务型数据库的领导者，在大数据场景面临巨大挑战：海量数据超出其处理能力，水平扩展面临瓶颈；小型机+数据库的配置扩展价格昂贵。结合 MySQL、MariaDB、PostgreSQL 等集群的分布式数据库应运而生，他们运行在普通 PC 服务器之上，较好地解决了水平扩展与事务型需求，性价比高，在业界获得较为广泛的应用。但在使用过程中需要解决好如下的问题：

- 支持 SQL 语法，满足事务型处理的需求，并保障已有业务的平滑迁移，改动尽可能少。

- 算法优化,保证事务型处理的效率。

- 分布式事务处理,容量及性能线性可平滑扩展,做到业务无关。

- 集群高可靠与高可用。这是难点和重点,在实现时需根据不同的应用场景采用不同的技术手段,在一致性、可用性、分区容忍性上找到平衡点,找到最合适的技术实现方案。

- 提高通用性,降低与业务的耦合性。目前以阿里为代表的事务型数据库应用,充分证明了该方案的可行性,但存在针对其特定应用场景专用性的问题。系统需要在分布式事务架构上做创新,做好柔性事务机制,将业务层处理分布式事务抽象到中间件层实现,对外接口保持与传统数据库的兼容性,降低与业务的耦合性,适应不同的应用场景,尽量降低业务层改写及开发代价。

- 提升可维护性,与传统数据库相当。当前很多互联网厂商的分布式数据库,针对自身的应用定做,依靠众多的专家队伍来维护,普通用户无法配置庞大的专家队伍,要求高。专有厂商需在这方面大力提升,充分降低维护难度。

2.3 面向媒体大数据的新架构

大数据时代的媒体数据呈现出明显的多源、多维、多媒体的特征,如何针对这些跨媒体大数据进行分析成为业界研究的热点。跨媒体数据的关联分析能更好地反映出世界的本来特征,发现更大的价值。跨媒体大数据的使用能为用户提供创新的服务和更Cool的用户体验。

跨媒体大数据对于媒体表达、数据关联、模型建立、统一存储和混合计算提出了新要求,原有架构不在适用,需要新的处理架构。统一的非结构化数据智能处理平台,需实现文本、图像、音频、视频、图形、时序、跨媒体等处理库,并需在应用层面进行统一封装与融合,可对上面的多类数据进行融合与关联分析,从而产出价

值创新。

深度学习是跨媒体大数据处理的热点,是跨媒体大数据处理场景中内容感知的关键支撑。它通过“多层次”的学习而得到对于原始数据的“不同抽象层次”的表示,进而提高分类和预测等任务的准确性。微软、谷歌、百度等公司已突破深度学习工程技术,并将其应用到语音识别、图像匹配、广告预测等产品中,取得初步成功。深度学习将跨媒体大数据处理时产生更大的作用,需要加强研究和实践。

同时,深度学习只是机器学习的一个分支,并非机器学习的全部,在跨媒体大数据分析时,需要根据不同的应用场景和需求,选择更合适的技术,不能全部寄托在深度学习上。

2.4 面向行业的定制化大数据解决方案成为企业的选择方向

在M-ICT时代,大数据是企业最重要的资产渐成行业共识,关键是如何应用落地。企业的大数据解决方案不仅需要收集、处理、分析和展示企业内部的各种结构化与非结构化数据,而且需要引入丰富的外部数据:互联网数据(浏览历史、新闻、评论、政策等)、微博和微信等社交数据。全面整合利用这份多维度完整的数据,更有利于发现新的商业机会,做出更合理的商业决策。

传统的商业智能方案基于封闭IOE体系架构,主要面向企业内部结构化数据,难以适应M-ICT时代的需求,正迅速被以开放高性价比的云计算和分布式数据库为代表的新型大数据企业架构体系所取代。

由于行业间的差异性,很难有一套统一的大数据商业智能分析系统,适用于所有行业,垂直行业的数据分析涉及到领域专家知识和领域建模,迫切需要进行跨学科和跨领域的大数据技术和应用研究,促进和推动大数据在典型和重大行业中的应用和落地。大数据解决方案需在统一架

构的大数据平台上,与不同的合作伙伴合作定制,形成面向不同行业的个性化智能大数据解决方案。其中电信、金融、智慧城市是大数据应用的热点。如金融大数据:从更多视角搜集经济运行数据,全方位分析客户行为与需求,做出合理的决策。通过网络交易情况进行征信,判断客户状况、潜在需求和偿还能力,让急需资金的中小企业和个人更容易获得金融支持,并有效降低金融行业风险,降低金融成本。

大数据将与物联网、移动互联网、云计算、社会计算、等热点技术领域相互交叉融合,产生很多综合性应用,前景非常广阔。

3 智能服务与轻量级的协同

3.1 自动智能服务

随着大数据与云计算的成熟,机器学习特别是深度学习获得了快速发展,并在语音识别、人脸识别、物体识别等领域取得喜人的成果,带动了人工智能的热潮。人工智能在业务领域有三大热点:面向企业服务的智能客服问答、面向个人的智能个人助理、面向专业或者行业智能咨询服务等。

(1)智能客服问答是人工智能在业务领域最有潜力的领域

传统的客服主要是靠建立在排队机基础上的坐席来提供服务的。其中,自动语音服务,流程相对固定,需用户根据提示选择,交互不友好,使用不便;人工客服的服务质量好,但做到7×24小时动态服务,成本很高。由机器自动提供自然语言的应答服务,成为客服新的主要的发展方向。

用户采用自然语言与系统进行交互,系统首先进行语音识别,将语音转成文本,系统进行自然语言的理解,从系统中找到或者推理出合理的答案,然后再转成语音与用户进行交互。也可以通过短信、微信、及时消

息等文本的方式进行。

从应用范围上可以分为两大类：

(a) 开放域的自然语言问答系统。其典型是苹果的Siri与微软的小冰,用户可以随意与系统交互各种各样的问题,以娱乐为目的。

(b) 特定域的自然语言问答系统。类似于10000、10001、10086、114等的自动智能问答面向特定领域,要求答案准确。

两类系统基本技术类似,但是重心不同,面向专有领域的智能问答系统目标更明确,更实用,但需要解决如下关键技术:

- 自然语言处理。这是基础也是核心。除传统的分词、新词发现、实体识别、语义角色标注和语义消歧外,重点是解决口语化、表达模糊、简称别称、缩略语与拼写错误等,进行自动纠错与规范,理解真正的语义。

- 对话管理。问答系统有上下文语境,需要解决语境识别、问题拆分、问题澄清等问题,在理解不清时,系统具有问题反问或者追问非常关键,是新的技术难点。

- 问题处理。除问题判定、指代消解、省略回复、文本复述等技术外,还需要在问题分类、关键词抽选、语义解析和意图识别上做关键的工作。

- 问题检索。问题归一化、语义搜索和知识推理是进一步要解决的问题。

- 知识库的构建是一个传统的问题,需要在词典知识库、文法库、业务知识库等的基础上构建针对企业专业知识库,并进行管理和加工,适应专有领域自然语言的深层理解。

基于自然语言理解的智能问答系统,将能大大降低客服中心的运营成本,方便提供7×24小时服务,提升服务质量。建好对应企业的知识库,让企业自行管理,提供特定领域的专业服务,该服务具有非常广阔的应用前景。

(2) 智能个人助理

智能个人助理根据个人的要求

提供贴身针对性的服务,具有非常大的想象空间。智能个人助理归个人私有专用,可帮助收集和整理信息,做摘要、推荐合适的信息、定制化提醒、日常事务的处理(如特定时间的例行祝福邮件等等),并根据你的历史处理习惯帮你做行为建议。

交互方式建立在自然语言的基础上,可以采用多种方式你进行交流,会成为你个人真正的专业助手。系统自身具备学习功能,能够持续的完善,原来越来越智能,越来越私有化,甚至比你更了解自己。

(3) 智能咨询服务

智能咨询服务系统会成为新一代专家系统,在专有领域内,具有非常强的自我学习和知识推理能力,随着计算能力的增长和大数据的结合,会极大地影响很多服务领域,将颠覆很多以知识和记忆为主服务的行业,带来服务方式的革命。

在医疗领域,系统通过海量知识的学习,并跟踪最新研究成果,结合实际数据(如医学化验单和病人的历史数据),给出专业的指导服务,在新病诊疗、日常医疗咨询等方面甚至比普通的医生更全面更及时。在金融领域,专业机构很早就在使用专业算法在做投资,当前主要用于限于专业机构,未来针对个人理财和投资的定制服务,会越来越的走向智能专业助理。

只要行业服务内部逻辑关系理清,或者是依据已有案例、现有知识和经验来判定的,智能咨询服务就可以根据这个规律来提供自动服务,如各种咨询服务(如法律咨询等)都可能被颠覆。IBM沃森已经在开始这方面的尝试,这是一个非常可行也非常有前景的业务领域。

3.2 轻量级的企业协同服务

WebRTC是基于浏览器原生支持多媒体通信的实现技术,开发者可以快速地开发富媒体应用,无需关注专业化的媒体处理技术(包括媒体采

集、编码、发送、接收、解码和渲染等)和通信技术。可实现一次开发统一部署,无需针对不同浏览器类型、异构终端类型、异构操作系统做适配开发;用户无需安装各种APP或者浏览器插件,通过浏览器即可享用丰富的多媒体应用。

WebRTC技术和BYOD应用的结合,将为企业协同办公、企业社交等应用提供安全、便捷、泛在的多媒体服务。基于WebRTC实现的多媒体会议、桌面共享、即时消息等协同办公功能,可在以较低成本开发、部署和运维的同时,为企业和消费者提供个性化、多样化和一致性的协同工作现实体验。

基于WebRTC实现的联络中心应用,可大幅降低客服中心的服务成本,客服人员通过浏览器即可随时随地提供音视频、即时消息、文件传输、桌面共享等丰富的媒体服务,同时结合后台的智能搜索、智能问答、数据挖掘等技术,即可为消费者提供高效、实时、准确的个性化、全方位的服务,极大地提升消费者的用户体验。

在4G LTE时代,WebRTC和VoLTE的结合也将进一步促进传统的电信服务和互联网技术的融合,从而进一步加速WebRTC技术的成熟和相关生态链的完善。

4 结束语

业务领域处于高速发展期,传统的业务实现、支撑和使用方式在正发生根本性的变革,面临巨大的挑战,也带来更大的机遇。云数据中心领域正在面临软件定义一切的变革,从软件定义计算、软件定义存储、软件定义网络、数据定义数据到软件定义数据中心,服务效率全面提升,降低了TCO。在管理和调度层面还有更多的工作要做,易用性、跨层次的资源调度,高可用性和高性能的结合等,进一步提升系统的效率,让用户更容易使用;业界火热的去IOE并非针对特定公司的行动,更多地体现在

计算模式、存储模式和服务模式的需求变革,是技术发展的必然。新型的服务器、存储革命到来,并呈现定制化和模块化趋势,正在颠覆传统的存储和计算行业,分布式数据库正在对传统的数据库模式造成重大冲击,这是一次变格局的大机遇;云计算、大数据与人工智能相互促进,将催生更多的新业务与新模式,更多地体现在与行业应用与应用场景的结合上,为业务领域的发展带来前所未有的变局和机遇,需要我们快速把握和应对这种变化,在个人通信方式上体现的融合与智能,在企业领域是专业、智能与实现的轻量级与多样化。同时,可信与安全是影响业务领域一大关键问题,在大数据与人工智能时代,迎来了新的挑战,需要深入开展专题研究与实践。业务领域发展空间巨大,前景广阔,抓住这次业务领域的大机遇,就会带来大发展。

参考文献

[1] Gartner. 2014年十大策略性技术与趋势 [EB/

- OL]. [2015-02-10]. <http://cloud.chinabyte.com/news/372/12746872.shtml>
- [2] CCF. 2014年大数据白皮书与发展趋势报告 [EB/OL]. [2015-02-10]. <http://cloud.chinabyte.com/news/372/12746872.shtml> <http://www.csdn.net/article/2014-12-13/2823106>
- [3] 李乐民. 未来网络的体系结构研究 [J]. 中兴通讯技术, 2013, 19(6): 39-42
- [4] 董振江, 李从兵, 王蔚, 吕达. 移动互联网 WebRTC 及相关技术 [J]. 中兴通讯技术, 2013, 19(6): 28-32
- [5] 赵慧玲, 解云鹏, 史凡. 网络虚拟化及网络功能虚拟化技术探讨 [J]. 中兴通讯技术, 2014, 20(3): 8-11
- [6] 薛一波. 大数据的前世、今生和未来 [J]. 中兴通讯技术, 2014, 20(3): 41-43
- [7] Mary Meeker Internet Trends 2014 - Code Conference [EB/OL]. [2015-02-10]. http://s3.amazonaws.com/kpcbweb/files/85/Internet_Trends_2014_vFINAL_-_05_28_14_-_PDF.pdf?1401286773
- [8] WebRTC 1.0: Real-time Communication Between Browsers [EB/OL]. [2015-02-10]. <http://www.w3.org/TR/WebRTC/#rtcpeerconnection>
- [9] Network Functions Virtualisation(NFV). Network Operator Perspectives on Industry Progress, ETSI [EB/OL]. [2015-02-10]. http://portal.etsi.org/NFV/NFV_White_Paper2.pdf
- [10] DONG Z J, LIU L X, WU B, LIU Y. MBGM: A Graph-Mining Tool Based on MapReduce and BSP [J]. ZTE Communications, 2014(4): 20-26
- [11] Mary MEEKER. Internet Trends [EB/OL]. [2015-02-10]. http://211.157.29.42/F10Data/HYGB_NEW/DOC/180.PDF

上接第 55 页

输到设备端进行处理,处理完成后再将数据由设备端传回至主机端,这个过程需要消耗一定的时间,当待处理图像的像素数较多时,可以降低传输时间在处理时间中的比重,使得加速比增大。同理,在 MPI 中,进程间的通信也有一定的代价,所以,当像素数较少时,仅使用 OpenCL 比使用 MPI 并行实现的效果理想,但当图像像素数较多时,使用 MPI 并行效果强于只应用 OpenCL 时的效果。

参考文献

- [1] 图像处理(卷积) [EB/OL]. (2014-05-10). http://blog.sina.com.cn/s/blog_6d0e97bb010130p2.html
- [2] 姚敏. 数字图像处理 [M]. 北京: 机械工业出版社, 2006
- [3] KARAS P, SVOBODA D. Convolution of large 3D images on GPU and its decomposition [J]. EURASIP Journal on Advances in Signal Processing, 2011, 350(5): 120-128
- [4] 刘伟峰, 蒋楠, 宋付英. 图像处理的 GPU 加速技术与评价 [C]. 福州: 第十四届全国图像图形学学术会议论文集, 2008
- [5] HARTUNG S, SHUKLA H, MILLER J P. GPU Acceleration of Image Convolution Using

- Spatially-varying Kernel [C]//Proceedings of the 2012 19th IEEE International Conference on Image Processing, 2012: 1685
- [6] 佟凤辉, 樊晓旭, 王党辉, 等. 基于 SIMD 技术的图像卷积处理器体系结构研究 [J]. 微电子学与计算机, 2003, 23(3): 14-17
- [7] 贾斯特. OpenCL 异构计算 [M]. 张云泉等译. 北京: 清华大学出版社, 2012
- [8] Khronos OpenCL Registry [EB/OL]. (2014-05-10). <https://www.khronos.org/>
- [9] LAM/MPI Parallel Computing [EB/OL]. (2014-05-10). <http://www.lam-mpi.org/>
- [10] FUJII Y, AZUMI T, NISHIO N, KATO S, EDAHIRO M. Data Transfer Matters for GPU Computing [C]//Proceedings of the International Conference on Parallel and Distributed Systems (ICPADS), 2013: 275-82
- [11] MARQUES R, PAULINO H, ALEXANDRE F, MEDEIROS P D. Algorithmic Skeleton Framework for the Orchestration of GPU Computations [C]//Proceedings of the Euro-Par 2013 Parallel Processing. Proceedings of the 19th International Conference, 26-30 Aug. 2013, Aachen, Germany: LNCS 8097, 874-85, 2013
- [12] Rocks [EB/OL]. (2014-05-10). http://www.rocksclusters.org/wordpress/?page_id=449
- [13] TORQUE [EB/OL]. (2014-05-10). http://en.wikipedia.org/wiki/TORQUE_Resource_Manager
- [14] TORQUE [EB/OL]. (2014-05-10). <http://db.apache.org/torque/torque-4.0/index.html>
- [15] DAVID J, QUINN S, MARK C. Core Algorithms of the Maui Scheduler [M].

Lecture Notes in Computer Science, 2001

作者简介



王继刚, 中兴通讯战略规划总监; 主要从事 ICT 领域基础软硬件及相关创新产品的技术规划和蓝海拓展工作; 已主持、参加基金项目 10 余项; 已发表学术论文 20 余篇, 其中 SCI、EI 收录 9 篇。



刘惠, 西安电子科技大学软件学院副教授、硕士生导师; 主要研究方向为多核与异构集群调度、并行计算、移动计算; 已主持、参加基金项目 20 余项, 发表学术论文 33 篇。



姜滨, 北京趣拿软件科技有限公司(去哪儿网)任软件开发工程师; 从事软件开发相关工作。

《中兴通讯技术》杂志(双月刊)投稿须知

一、杂志定位

《中兴通讯技术》杂志为通信技术类学术期刊,通过介绍、探讨通信热点技术,展现通信技术最新发展动态,并促进产学研合作,发掘和培养优秀人才,为振兴民族通信产业做贡献。

二、稿件基本要求

1. 投稿约定

- (1) 作者需登陆《中兴通讯技术》投稿平台: www.zte.com.cn/paper,并上传稿件。第一次投稿需完成新用户注册。
- (2) 编辑部将按照审稿流程聘请专家审稿,并根据审稿意见,公平、公正地录用稿件。审稿过程需要大约1个月左右。

2. 内容和格式要求

- (1) 稿件须具有创新性、学术性、规范性和可读性。
- (2) 稿件需采用 WORD 文档格式。
- (3) 稿件篇幅一般不超过 6 000 字(包括文、图),内容包括:题名、作者姓名、作者单位、中文摘要、关键词(4~8 个)、英文摘要、正文、参考文献、作者简介。
- (4) 中文题名一般不超过 20 个汉字,中、英文题名含义应一致。
- (5) 摘要尽量写成报道性摘要,包括研究的目的、方法、结果与结论,以 150~200 字为宜。摘要应具有独立性和自明性,采用第三人称。中英文摘要内容应一致。
- (6) 文稿中的量和单位应符合国家和国际标准。外文字母的正斜体、大小写等须写清楚,上下角的字母、数据和符号的位置皆应明显区别。
- (7) 图、表力求少而精(以 8 幅为上限),应随文出现,切忌与文字重复。图、表应保持自明性,图中缩略词和英文均要在图中加中文解释。表应采用三线表,表中缩略词和英文均要在表内加中文解释。
- (8) 参考文献以 20 条左右为宜,未公开发表的资料不宜列入。所有文献必须在正文中引用,文献序号按其在文中出现的先后次序编排。主要种类参考文献的书写格式为:
 - 期刊[序号]作者. 题名[J]. 刊名, 出版年, 卷号(期号): 起止页码
 - 书籍[序号]作者. 书名[M]. 出版地: 出版者, 出版年: 起止页码
 - 论文集中析出文献[序号]作者. 题名[C]//论文集编者. 论文集名(会议名). 出版地: 出版者, 出版年(开会年): 起止页码
 - 学位论文[序号]作者. 题名[D]. 地点: 学位授予单位, 授予年
 - 专利[序号]专利所有者. 专利题名. 国别: 专利号[P]. 公布日期
 - 国际、国家标准[序号]标准编号, 标准名称[S]
- (9) 作者原则上不超过 3 人,超过 3 人时,可以感谢形式在文中提及。作者简介包括:姓名、工作单位、职务或职称、学历、毕业于何校、现从事的工作、专业特长、科研成果、已发表的论文数量等。
- (10) 提供正面、免冠、彩色标准数码照片一张,最好采用 JPG 格式(文件大小超过 100 kB)。
- (11) 尽可能标注出研究课题的资助基金或资助项目名称。
- (12) 作者姓名中含有多音字时,应标注作者姓名的汉语拼音。
- (13) 提供联系方式,如:通信地址、电话(含手机)、Email 等。

3. 其他事项

- (1) 请勿一稿多投。凡在 2 个月(自来稿之日算起)以内未接到录用通知者,可致电编辑部询问。
- (2) 为了促进信息传播,加强学术交流,在论文发表后,本刊享有文章的版权(包括英文版、电子版、网络版和优先数字出版)。作者获得的稿费包括版权酬金。如对此持有不同意见,请在投稿时说明。

编辑部地址:安徽省合肥市金寨路 329 号国轩凯旋大厦 1201 室, 邮政编码: 230061

联系电话: 0551-65533356, 联系邮箱: magazine@zte.com.cn

本刊只接受在线投稿, 欢迎访问本刊投稿平台: www.zte.com.cn/paper

中兴通讯技术

ZHONGXING TONGXUN JISHU

双月刊 1995 年创刊 总第 121 期
2015 年 4 月 第 21 卷第 2 期

主管:安徽省科学技术厅
主办:安徽省科学技术情报研究所
中兴通讯股份有限公司
编辑:《中兴通讯技术》编辑部

总编:孙枕戈
副总编:黄新明
责任编辑:杨勤义
编辑:徐烨, 卢丹, 朱莉, Paul Sleswick
排版制作:余刚
发行:王萍萍
编务:王坤

ZHONGXING TONGXUN JISHU

《中兴通讯技术》编辑部
地址:合肥市金寨路 329 号凯旋大厦 12 楼
邮编:230061
网址: www.zte.com.cn/magazine
投稿平台: www.zte.com.cn/paper
电子信箱: magazine@zte.com.cn
电话: (0551)65533356
传真: (0551)65850139

出版、发行:中兴通讯技术杂志社
发行范围:全球发行
印刷:合肥添彩包装有限公司
出版日期:2015 年 4 月 10 日
刊号: ISSN 1009-6868
CN 34-1228/ TN
广告经营许可证:皖合工商广字 0058
定价:每册 20.00 元, 全年 120.00 元