



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10月 • 第5期

专题：虚拟运营业务和网络



《中兴通讯技术》第7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）

副主任 马建国（天津大学电子信息工程学院院长） 陈前斌（重庆邮电大学通信与信息工程学院执行院长）

编委（按姓氏拼音排序）

曹淑敏 中国信息通信研究院院长

陈建平 上海交通大学教授

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

葛建华 西安电子科技大学通信工程学院副院长

管海兵 上海交通大学电子信息与电气工程学院副院长

侯为贵 中兴通讯股份有限公司董事长

洪 波 中兴发展股份有限公司总裁

洪 伟 东南大学信息科学与工程学院院长

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

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

李红滨 北京大学教授

李建东 西安电子科技大学副校长

李 军 清华大学信息技术研究院院长

李乐民 中国工程院院士, 电子科技大学教授

李融林 华南理工大学教授

李少谦 电子科技大学通信与信息工程学院院长

李 涛 南京邮电大学计算机学院院长

李 星 清华大学教授

刘建伟 北京航空航天大学电子信息工程学院党委书记

马建国 天津大学电子信息工程学院院长

孟洛明 北京邮电大学教授

糜正琨 南京邮电大学教授

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

谈振辉 北京交通大学教授

唐雄燕 中国联通网络技术研究院首席专家

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副总经理

王 京 清华大学教授

王文东 北京邮电大学软件学院副院长

卫 国 中国科学技术大学教授

吴春明 浙江大学计算机科学与技术学院教授

邬贺铨 中国工程院院士

徐安士 北京大学教授

续合元 中国信息通信研究院技术与标准研究所总工

薛一波 清华大学教授

杨义先 北京邮电大学教授

杨 震 南京邮电大学校长

尤肖虎 东南大学教授

张宏科 北京交通大学教授

张 平 北京邮电大学网络技术研究院执行院长

张云勇 中国联通研究院副院长

赵慧玲 中国电信股份有限公司北京研究院总工程师

赵先明 中兴通讯股份有限公司执行副总裁

郑纬民 中国计算机学会理事长、清华大学教授

钟义信 北京邮电大学教授

钟章队 北京交通大学计算机与信息技术学院院长

周 亮 南京邮电大学通信与信息工程学院副院长

朱近康 中国科技大学教授



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
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124期 第21卷 第5期 2015年10月

专题:虚拟运营业务和网络

- 02 移动通信转售业务发展趋势及影响 张志峰
07 全球移动虚拟网络运营商发展情况报告 George GUO
10 全球移动虚拟网络运营商发展经验对中国的启示 王雪梅
13 移动通信转售批发价对转售产品的影响 彭荣林
17 移动通信转售业务标准化发展之路 孟然,李娜,李硕
21 移动虚拟网络运营技术研究 朱玉娜,王修中
24 基于移动终端显示移动转售企业标识的解决方案 沈彬,李硕,孟然

专家论坛

- 27 大数据环境下的可靠存储技术思考 李挥,张宇蒙,陈俊
32 半监督多视图学习在大数据分析中的应用探讨 蓝超,饶泓,浣军

企业视界

- 35 M-ICT时代下的有线网络 柯楚,王翔

技术广角

- 39 安全可信智能移动终端研究 张大伟,郭烜,韩臻
45 数据中心绿色技术及其发展研究 岳玉,曹可建,郭亮
50 基于信息熵的POMDP模型观测 钟可立,王小捷
56 机会调度模式下的嵌套码研究 胡留军,许进
60 基于大数据的业务并发度分析 易正磊,顾军,张兴

综合信息

人工智能将改写未来10年ICT产业面貌(06) 未来3年网络信息基础建设投资将超1.2万亿元(49) 光纤入户比近半提速降费按下“快进键”(55) Gartner调查报告称:信息安全管理实践正日趋完善(64) 2016年第1—6期专题征文(65)

期刊基本参数:CN 34-1228/TN*1995*b*16*64*zh*P*¥ 20.00*15000*12*2015-10

Contents 目次

ZTE TECHNOLOGY JOURNAL Vol. 21 No. 5 Oct. 2015

Special Topic: Service and Network of Virtual Operation

- 02 Trends and Impact of Resale of Mobile Communication Business ZHANG Zhifeng
07 The Global MVNO Footprint George GUO
10 Development Experience of Global MVNO for China WANG Xuemei
13 Mobile Communication Resale Wholesale Price for Resale Products PENG Ronglin
17 Standardization Development of Mobile Resale Business MENG Ran, LI Na, LI Shuo
21 MVNO Technology ZHU Yuna, WANG Xiuzhong
24 The Solution for Display of Mobile Resale Business Logo
on Mobile Terminals SHEN Bin, LI Shuo, MENG Ran

Expert Forum

- 27 Reliable Storage Technologies for Big Data LI Hui, ZHANG Yumeng, CHEN Jun
32 Semi-Supervised Multi-View Learning
in Big Data LAN Chao, RAO Hong, HUAN Jun

Enterprise View

- 35 The Wireline Network in M-ICT KE Chu, WANG Xiang

Technology Perspective

- 39 Security and Trusted Intelligent Mobile Terminal ZHANG Dawei, GUO Xuan, HAN Zhen
45 Green Technology of Data Center and Its Development YUE Yu, CAO Kejian, GUO Liang
50 Information Entropy Based Estimation for Observation
Function of POMDP Model ZHONG Keli, WANG Xiaojie
56 Nested Codes in Opportunistic Scheduling Model HU Liujun, XU Jin
60 Service Concurrency Analysis Based on Big Data YI Zhenglei, GU Jun, ZHANG Xing

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。



续合元

中国信息通信研究院技术与标准研究所总工程师，教授级高级工程师，曾任 ITU-T 物联网全球标准化行动的技术策略协调人、M2M 焦点组主席，现任 ITU-T SG13 的副主席，中国通信标准化协会 TC3、TC10 总体工作组组长。长期从事电信网络与交换方面的研究和标准制订工作，对于信令和网络协议有深入的研究；目前致力于未来网络、泛在/物联网、车联网、工业互联网等的相关研究。

专题导读

随着电信技术的发展，用户对电信业务需求不断提高，电信业务种类也变得更加丰富，传统电信运营企业将业务重点集中于其最为擅长的网络建设与维护，而对于大量的增值业务和功能化业务则转售给更加专业的企业，开展合作业务运营，虚拟运营商因此应运而生。虚拟运营商的出现，改变了以往电信运营的模式。从商业运作上看，虚拟运营商并不具有网络，但是可以通过网络的租赁和使用，为客户提供服务。这样以来，虚拟运营商就可以将更多的精力投入到新业务的开发、运营、推广和销售，为用户提供更为专业的服务。

虚拟运营商于 1999 年在欧洲开始兴起，之后包括中国香港、美国、日本和韩国等市场都相继开放，并均将其视为激活电信市场竞争、打破垄断的重要举措。在中国，关于开放虚拟运营商政策的呼吁已持续数年。2012 年 6 月 27 日，工信部颁布了《关于鼓励和引导民间资本进一步进入电信业的实施意见》（实施意见）；2013 年 1 月工信部发布了《移动通信转售业务试点方案》；2013 年 12 月 26 日，工信部向京东、迪信通、巴士在线等 11 家民营企业发放了首批虚拟运营商牌照，允许其从基础电信业务经营者购买移动通信服务，并按照相关规定自主运营。到目前为止，工信部共向 42 家民营企业发放了移动通信转售业务试点批文。

中国移动通信转售业务的开展，为电信市场带来了很多变化，如市场竞争者数量迅速增加，产品服务多元化等。纵观全球移动通信转售业务发展，目前仍保持着勃勃生机。中国虚拟运营商自 2014 年 5 月陆续放号以来，在一年的时间内已超过 600 万用户。但虚拟运营业务还只是刚刚开始，未来其该如何发展，都需要进一步探讨和研究。我们这期专题的论文凝聚了作者多年的研究成果和工作经验，希望能给读者有益的启示与参考。在此，对各位作者的积极支持和辛勤工作表示衷心的感谢。

续合元

2015 年 9 月 17 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技术与标准研究所 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 副院长

移动通信转售业务发展趋势及影响

Trends and Impact of Resale of Mobile Communication Business

张志峰/ZHANG Zhifeng

(中国信息通信研究院, 北京 100037)
(China Academy of Information and
Communications Technology, Beijing
100037, China)

1 产业发展趋势及影响

1.1 全球移动通信转售业务市场持续增长

据 Informa Telecoms & Media 统计,截至到2012年底,全球移动虚拟网络运营商(MVNO)的用户规模达1.2亿,占全球移动电话用户数的1.9%。目前,发达通信市场移动通信转售业务仍保持快速增长。Informa Telecoms & Media 预测,到2018年,全球MVNO的用户将达到2.7亿,年均增长率超过14%,占全球移动电话用户数的3%。INFORMA对2013—2018年全球MVNO用户的发展预测如图1所示。

(1)美国:移动电话普及率高,但未成MVNO发展阻力

美国MVNO市场起步非常早,正式启动于1999年,初期发展缓慢,随后进入高速发展期,到2006年MVNO用户占有率相对稳定,达到7%。此后,受智能终端补贴、LTE网络建设等多因素影响,基础运营商不断探寻新的增长点,探索合作驱动发展。Sprint、T-Mobile等弱势运营商希望借助MVNO的力量,让自己能够集中精

中图分类号:TN91 文献标志码:A 文章编号:1009-6868(2015)05-0002-005

摘要: 指出中国移动通信转售业务的开展为电信市场带来了变革:市场竞争者数量迅速增加、业务种类进一步丰富、产品服务多元化、移动用户规模及流向发生变化。认为纵观全球移动通信转售业务产业,移动通信转售业务市场目前生机勃勃;随着移动通信转售业务日趋成熟,移动通信转售业务产业发展、商业模式、产业合作将进一步演变,对通信市场必将产生深远影响。

关键词: 移动通信转售;产业发展;业务模式;产业合作

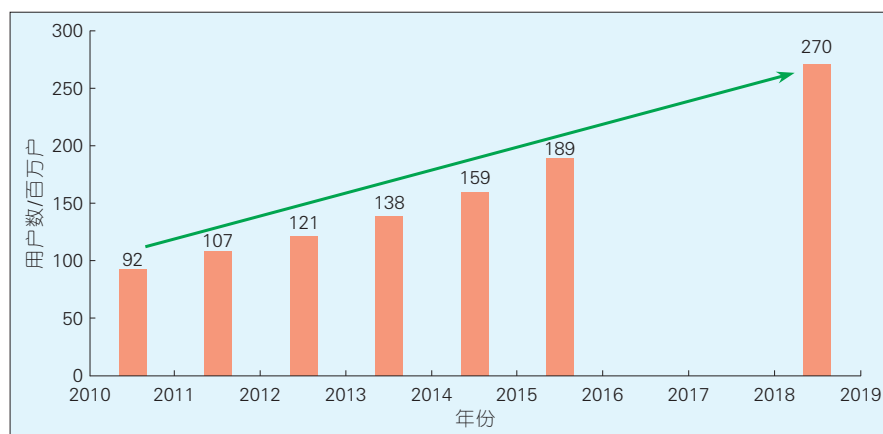
Abstract: The development of China's mobile telecommunication resale service has brought many changes, such as the rapid increase in market players, greater variety of business, and more diverse products, and a different direction for mobile users. Throughout the industry of resale of the world's mobile communications, market of resale of mobile communications services remains vitality. As the mobile communications resale business matures, the evolution of the development, business model and industry cooperation of resale of mobile communications business have a profound impact on the communications market.

Keywords: resale of mobile communication business; industrial development; business model; industrial cooperation

力于更核心的战略,多采取同MVNO合作制订资费并从中分成的方式维系中低端用户;AT&T、Verizon通过MVNO发展弥补用户流失、收入下滑的损失;基础运营商则持更开放的态

度,从而造就MVNO在美国市场再度活跃。目前美国共有100余家MVNO。MVNO用户份额得到持续提升,2013年已经达到11%。

除了移动网络运营商(MNO)态



▲图1 Informa对2013—2018年全球MVNO用户发展预测

收稿日期: 2015-07-03

网络出版时间: 2015-08-24

度的转变带来虚拟运营业务的增长外,近两年,美国涌现出一批新型MVNO也对移动通信转售市场产生了极大的促进作用。部分MVNO开始对新兴的智能终端和移动宽带市场进行产品创新和市场争夺,如Ting率先推出多设备共享流量套餐,Zact主打个性化定制资费模式,Kamma为用户提供独特的社会化宽带分享服务,Freedompop更是提出了免费赠送的商业模式。

(2)英国: MVNO数量快速增加,已达248家

MVNO在全球兴起的起点始于1999年的英国市场,在有利的监管政策推动下,英国MVNO市场表现活跃,开放3年内MVNO数量就增至50家左右。MVNO类型丰富,除了Virgin这样的多领域品牌价值类MVNO之外,还有以Tesco为代表的大型零售连锁企业、由手机终端零售商转型为固网宽带运营商的TalkTalk、专注于国际长话细分市场的Lebara和针对青少年群体提供广告媒体服务的Blyk等。由于市场定位清晰,差异化业务区分明显,多家MVNO均有不错的市场表现。

近年来新进入市场的虚拟运营商数量有所增加,截至2013年英国市场MVNO数量达248家,用户份额达15.6%,用户规模呈增长态势。英国MVNO市场发展状况如图2所示。

1.2 中国移动通信转售业务起步顺利

中国正处于移动通信转售业务

业务运行起步阶段,截至2014年底,有21家转售企业发布了移动通信转售业务品牌,3家开始预约放号,16家开始正式放号,移动转售用户超过210万。

经过2014年的摸索,大部分转售企业已经明晰了业务定位和方向,2015年随着更多转售企业在更多地域开展业务,移动通信转售业务将会进入快速发展期。一方面,蜗牛、苏宁、国美、阿里等第一阵营的企业可能会跑得更快,第二阵营的可能会迎头猛追;另一方面,随着一些其他行业的“大鳄”,例如海航、民生、平安等进入移动通信转售业务市场,搅动的动静也许会超乎预期;此外,优酷、263、小米等互联网企业也可能给我们带来颠覆性创新模式上的期待。2014年,作为移动转售试点元年,净增用户已突破百万大关;2015年,净增用户突破千万量级应该问题不大。

1.3 市场分化和优胜劣汰将导致市场高度集中

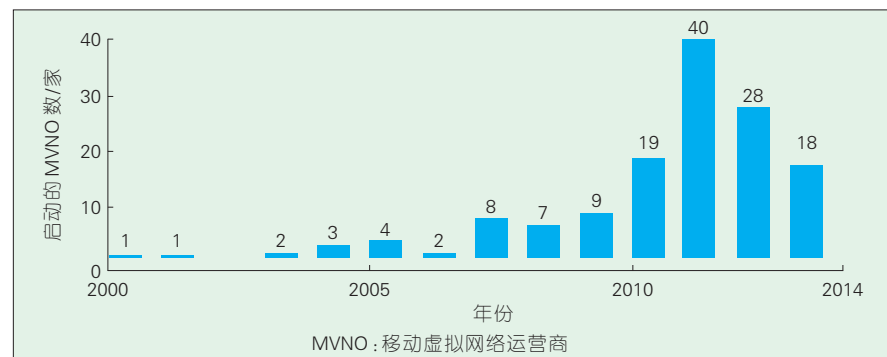
在试点政策公布以后,一共近两百家民营企业正式提交了参与试点的意向,最终一共42家企业获得试点资格。这超出了大多数人的预期,但也在情理之中。因为移动转售业务开放也给基础运营商提供了一个发展的机遇,理论上讲是一个多方共赢的好政策。

移动通信市场,由3家基础运营商之间的直线竞争,变成42家转售企业之间的网状竞争关系,是否会陷

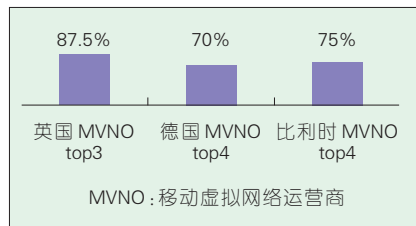
入过度竞争、恶性竞争?本文认为不会,因为中国移动通信市场容量较大且仍有很大增长空间,决定了42家转售企业并不算多。中国移动通信业务收入接近9000亿元,盈利空间较大。当前移动互联网和智能终端高速增长,移动数据流量规模每年翻倍增长,移动数据流量收入每两年翻倍增长,移动用户信息需求不断扩大,为民营企业发展提供了更多创新空间。移动转售企业将产生“群狼效应”,推动移动业务繁荣与发展。

具体到试点企业而言,不能奢望每个试点企业都能成功。从国际经验来看,在成熟的虚拟转售业务市场中马太效应凸显,即领先的MVNO更容易获得生存空间。在欧洲国家,虚拟运营市场排名前列的MVNO拥有本国整体MVNO市场份额70%以上。例如,英国排名前3的MVNO占MVNO总市场份额87.5%;德国排名前4的MVNO占MVNO总市场份额70%;比利时排名前4的MVNO占MVNO总市场份额75%。上述现象在亚洲同样显现,如在日本,354家MVNO中用户规模超过50万的企业数量也仅有4家。同时,从国际发展的角度来看,有24.5%的虚拟运营商最终停业或是被收购(引自Informa统计数据)。Valoris Viewpoint统计的领先的MVNO占MVNO总市场份额如图3中所示。

中国未来在出现更多转售商的同时,市场也将逐步整合与分化,发展较好的转售企业将占据较大份额,大多数转售商都将是用户规模较小、覆盖细分市场的小型转售商。市场竞争的残酷性,使得转售企业出现业绩分化乃至优胜劣汰,属于正常现象,但绝大部分转售企业通过创新探索,依然可以得以生存和发展。由于中国市场容量的特殊性,中国可能会诞生用户数世界领先的虚拟运营商。扣除收购的影响,国际上一般大型移动通信转售业务商也就四五百万用户规模,在中国产生这样规模的



▲图2 英国MVNO市场发展状况



▲图3 领先的MVNO占MVNO总市场份额

虚拟运营商是可期待的。

2 业务模式发展趋势及影响

2.1 MVNO将引领电信行业资费计划和流量服务创新

在国际领先的电信市场, MVNO都为原有的市场带来了资费和流量服务创新。以美国新兴MVNO为首的新型虚拟运营商们,在市场上紧密契合移动宽带时代的客户需求,在个性化定制套餐、流量充分利用方面不遗余力地进行“微创新”,创新产品虽然五花八门,如美国虚拟运营商Ting率先推出了可供用户个性化定制的资费计划,用户可根据实际需求自主选择每月订购的语音通话时长、短信数量以及数据流量。美国虚拟运营商Zact将个性化定制的资费模式推向新高度,用户可使用智能终端上的应用程序,并允许用户随时修改资费计划并实时生效。

从中国转售企业目前推出的业务服务情况看,个性化定制资费、流量服务几乎成为标配。转售企业给市场带来的第一轮冲击波主要体现在资费方案和流量服务上。最早放号的19家企业中有12家企业采用了固定套餐方式,不过转售企业正不断突破基础运营商的传统资费模式,依托IT支撑系统建设的后发优势,积极探索个性化定制套餐、多人共享套餐、话音流量互转等新模式。

2.2 移动互联网给移动通信转售市场带来颠覆性创新

随着移动互联网浪潮掀起,部分虚拟运营商也在纷纷主动在其产品

服务中植入“移动互联”基因。目前,在全球范围内,一些新型MVNO运用互联网思维开发电信产品,结合社交平台、硬件购买、Wi-Fi等互联网手段,推出了主打“免费使用”、“无限使用”和“云管端一体化”等旗帜的“颠覆性创新”产品模式,受到用户的广泛好评。

(1) FreedomPop以“免费”的概念切入

美国虚商FreedomPop推出“免费”理念套餐产品,向用户提供每户每月200分钟语音、500条短信、500 MB的免费使用量。此外,通过“邀请新用户加入”的形式,原有用户还可获得更多免费流量。

(2) Republic Wireless推出“Wi-Fi无限”产品

美国虚商Republic Wireless向用户推出混合制式定制手机。用户购买定制手机,便可通过“优先连接Wi-Fi”进行包括通话路由在内的所有通信服务,在各档套餐包中,Wi-Fi均可无限使用,但当用户每月流量超过5 GB时,Wi-Fi速度受到限制。这对数据流量需求大的用户具有极大吸引力。

(3) 无限量使用套餐

虚商Republic Wireless通过设置不同档位套餐,向其分别提供Wi-Fi、语音+短信、语音+短信+流量的无限套餐。

Voyager Mobile向用户推出“19美元无限通话+短信”、“39美元无限通话+短信+上网流量”等套餐。

FreedomPop更是面向用户推出了“10.99美元无限语音和短信”套餐档位,吸引了众多用户转网。

(4) Amazon推出“流量+硬件+内容”模式

美国电商巨头亚马逊结合自身平台和硬件优势,向用户推出Kindle捆绑套餐,售价每年49.99美元,套餐包含250 MB流量、20 GB云存储空间和10美元App消费额度,吸引新用户入网并增强其原有用户黏性。

Karma推出了“用户花费79美元购买一台热点设备,即可获得1 GB的免费网络带宽”的电信服务套餐。

目前中国转售企业从创新深度上看,绝大部分转售企业的市场动作还属于“微创新”层面,未来有望出现更多“颠覆性”创新的尝试。

2.3 MVNO创新模式将被主流基础运营商借鉴推广

从国际经验来看,虚拟运营商通信产品的创新将被主流运营商迅速推广借鉴。其中最典型的莫过于美国虚拟运营商Ting。自2011年10月开始试运营之后,Ting即推出了开创性的共享资费计划。不同于以往老牌运营商的所谓家庭计划,Ting的共享计划不但支持多个终端共享套餐中的话音和短信,还允许多个终端共享数据流量。这种共享资费模式不但有利于吸引家庭用户,而且有效地降低了新型移动宽带设备,如平板电脑、具有蜂窝连接功能的消费电子产品等进入运营商网络的门槛,有利于运营商扩大用户基础,建立新的市场增长点。同时,Ting还允许用户对资费计划进行个性化的定制,自主选择每月订购的语音通话时长、短信数量以及数据流量。当用户每月的实际使用量超出或低于其选择的资费等级时,Ting将对用户做自动的升级或降级处理,并将当月多余的话费返还至用户帐户。

Ting的这些探索成功地开启了美国移动通信市场的新模式。Verizon Wireless、AT&T等一线运营商在Ting推出共享资费计划之后的半年到一年内,也都纷纷推出了自己的共享资费计划。目前,多设备共享套餐已经基本成为美国运营商资费计划中的标准配置,而Ting正是这一趋势的开创者。

中国移动转售企业所推出的许多创新产品,也均被快速推广应用。譬如,苏州蜗牛推出流量不清零卡以及零月租卡之后,中国电信迅速推出

了 iFree 卡,提供零月租、流量免费赠送等与苏州蜗牛类似的产品内容。此外,在巴士在线推出流量银行后不久,中国联通也迅速推出了流量银行产品。总的来说,这些创新将从整体水平上提升消费者用户感知,对中国电信业产生积极影响。

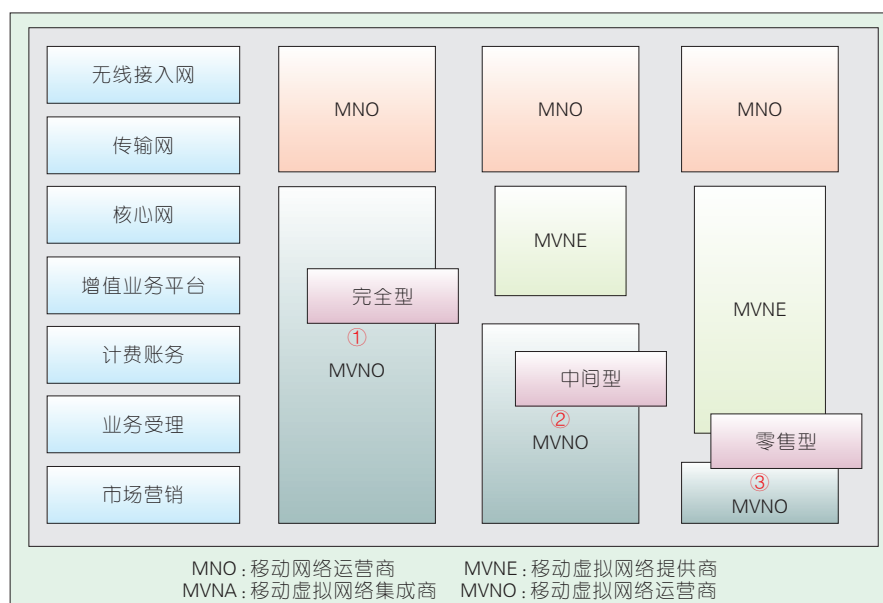
3 产业合作模式发展趋势及影响

3.1 发达市场 MVNE/MVNA 模式普遍推广

移动通信转售业务提供商与基础运营商、终端供应商、内容提供商、服务提供商、软件提供商、平台提供商、系统集成商、设备提供商等建立合作伙伴关系,通过向网络运营商租用网络资源和设备,从设备制造商获得必要的服务设备,并与自身的技术和服务加以整合,重新包装形成全新的服务产品,直接面向用户或通过专业化的销售渠道进行服务,并通过向最终用户直接收取月通信消费额及入网时可能的购机费用获取收入。从全球发展来看,成熟市场上虚拟运营商根据涉及的产业链相关内容的多少,有 MVNO、MVNE、MVNA 3 种分类。MVNO 分类如图 4 所示。

从国际经验来看,在虚拟运营商最发达的欧洲,支持 MVNO 快速灵活的开展业务已经成为重要节点,即移动虚拟网络提供商(MVNE)。MVNE 是虚拟运营商(MVNO)和移动网络运营商(MNO)之间的桥梁,提供从内容整合、服务集成、业务流程管理、IT 支撑、专业服务管理等多种功能。

MVNE 并不是从虚拟运营行业一开始就出现的。最早的 MVNO,在开展自己的业务时,和中国目前各家 MVNO 一样,从网络对接到 IT 能力建设,全部都是自己投资建设。随着虚拟运营行业参与方越来越多,很多的 MVNO 的相近需求就暴露出来了:首先,各家 MVNO 都需要和少数的 MNO 做网络对接,而 MNO 提供的对接方



▲图 4 MVNO 的分类

式往往是统一的;其次,MVNO 核心业务所需的客户关系管理(CRM)、服务开通、计费、客服等功能和业务流程非常类似,完全可以标准化实施甚至标准化外包运营;第三,MVNO 在业务发展初期,尽管用户数不多,但 IT 系统的建设前期投入仍然很大,对于部分有试水心态的 MVNO 来说,在没有看到明确的业务回报之前,他们希望有按需供应的服务商来支撑他们开展相关业务,于是,MVNE 产业应运而生。

MVNE 激发了虚拟运营商的业务灵活性,降低了 MVNO 进行电信业务运营的风险与门槛,可以使企业将更多地精力放在品牌构建、渠道策划、产品宣传和业务创新上。因此可以说,MVNE 的出现是虚拟运营产业进一步分工协作的标志,也是虚拟运营产业成熟的标志。

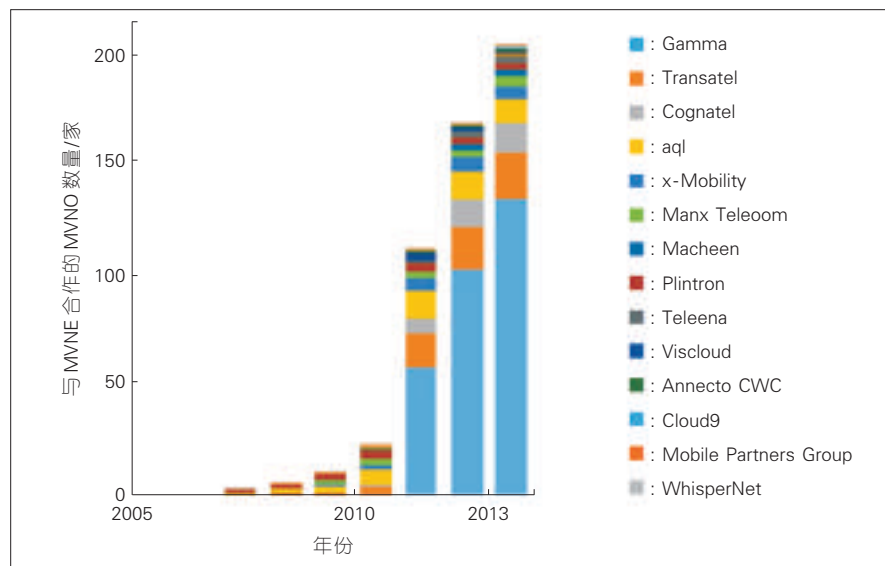
在英国市场上,通常基础运营商至少拥有两家合作 MVNE。MVNE 的出现使得很多拥有共同需求的 MVNO 聚合在一起,大大降低了 MVNO 的建设运营成本,并使 MVNO 能够专注于细分市场开发、业务创新等差异化能力。2013 年英国市场与 MVNE 合作的 MVNO 数量如图 5 所

示。Piran Partners 2013 年的统计数据显示,在英国,有 132 家 MVNO 都与本国最大的 MVNE Gamma 公司合作,开展虚拟运营业务,未利用 MVNE、移动虚拟网络集成商(MVNA)而直接与基础运营商合作开展虚拟运营业务的 MVNO 仅 33 家。

3.2 MVNE 将可能出现在中国虚拟运营市场

目前来看,在中国移动转售市场上尚未出现 MVNE,而转售企业对于 IT 系统大多采用自建或租用运营商的方式,但目前租用运营商系统的转售商,大多也有后期自建系统的计划。对于不同的基础运营商,与转售商合作的方式也不尽相同,中国移动和中国电信向 MVNO 提供自建 IT 系统和租用 IT 系统两种合作模式,而中国联通则明确要求 MVNO 自建系统。由此可见,在中国基础运营商目前正扮演着 MVNE 的角色。

那么,对于选择自建系统的中国虚拟转售商,选择自建的原因主要由以下 3 点:第一,申请企业规模较大,有能力自建 IT 系统;第二,企业普遍重视商业信息的保密性,认为自建系统是最安全的选择;第三,认为自建



▲图5 2013年英国市场与MVNE合作的MVNO数量

系统能够提升业务创新能力,将系统能力作为企业核心竞争力。

未来,随着更多的中小型企业加入虚拟运营行列,它们希望快速启动

业务的需求将愈加迫切,同时,将有越来越多的MVNO希望以轻资产、弹性化的形式地开展转售业务,在没有看到明确的业务回报之前能够尽可能减少不必要的固定资产投资。此时,MVNE将很好地解决这些企业的核心诉求,帮助他们把开展虚拟运营业务的重心放在业务创新等核心差异化能力上。

作者简介



张志峰,中国信息通信研究院产业与规划研究所工程师;主要从事移动转售业务发展和网络发展规划的研究;已发表学术论文2篇。

综合信息

人工智能将改写未来10年ICT产业面貌

人工智能技术近年来日渐兴起,全球科技巨头均在积极布局这一领域,专业研究机构认为人工智能应用前景广阔,将改写未来10年ICT产业面貌。

目前,人工智能应用于信息和通信技术之间的业务正在如火如荼地展开,苹果、谷歌、IBM、微软与雅虎等国际知名厂商争相通过并购、人才争夺战等手段来强化布局人工智能技术相关领域。

拓璞产业研究所近日发布报告预测:2015年以企业为主的人工智能系统市场价值接近2亿美元,到2020年将达20亿美元以上,5年间成长倍数高达10倍。

拓璞通讯产品分析师尤克熙认为,人工智能原本只是一项用于研究领域的专精科技,近年来随着技术不断拓展,开始进入一般人工智能时代,大企业开始利用并购获取技术,并结合大数据与云端服务器进行改进其服务。

“人工智能导入企业将于未来5年成为重心。”尤克熙表示,包含制造业、软件业、资讯工程业、生物科技、零售通路业、网络业、汽车产业等都将陆续导入人工智能技术。例如导入人工智能于机器人技术可提高制造业产能与效率;导入人工智能于医疗产业,结合穿戴式装置将可提升整体医疗水平等。

在研究机构看来,人工智能将成为IT领域一场重要的技术革命,目前市场关心的IT和互联网领域的一些主题和热点(智能硬件、O2O、机器人、无人机以及工业4.0),其发展突破的关键环节都是人工智能。人工智能有望成为未来10年乃至更长时间内IT产业发展的焦点。

2015年“两会”期间,全国人大代表、科大讯飞董事长刘庆峰在两会议案中建议制订国家人工智能战略,加快人工智能布局。刘庆峰表示,人工智能是未来IT发展的前瞻方向。未来5~10年,人类将进入智能一切的时代,机器人将像水、食物、互联网一样,成为人类每天生活中的标配,难以脱离。

在中国国内,人工智能技术正被尝试应用于各种日常生活场景中。比如2015年腾讯和阿里争相推出刷脸支付功能,从中也可以看出人脸识别等生物识别和图像识别技术的规模化商业应用已经初见端倪。

为促进中国人工智能产业发展,在国务院不久前印发的《关于积极推进“互联网+”行动的指导意见》所提出的11项具体行动计划中,其中一项便是“互联网+”人工智能。《指导意见》表示,将加快人工智能核心技术突破,提升终端产品智能化水平。

(转载自《中国信息产业网》)

全球移动虚拟网络运营商发展情况报告

The Global MVNO Footprint

George GUO

(亚太信宇科技(北京)有限公司, 北京
100007)
(Syniverse Technologies, Beijing
100007, China)

截至2014年底,全球移动运营商(MNO)共托管了992家移动虚拟网络运营商(MVNO)及运营商旗下260家子品牌公司。根据GSMA Intelligence数据报道,这代表相比2012年全球移动运营商托管的总移动服务供应商数量从1036个增长到了1252个。研究表明,全球移动虚拟网络运营商在成熟市场中继续保持旺盛的发展态势,且渗透率(基于连接数)已超过100%。全球2/3的移动虚拟网络运营商来自于欧洲(585家),随后的是亚洲(129家)和美国(107家)。对比之下,非洲市场的移动虚拟网络运营商还是属于婴儿期。该地区目前只有8家移动虚拟网络运营商。

GSMA Intelligence协会将移动虚拟网络运营商划分成8个种类,它们分别是:折扣、电信、媒体/娱乐、零售、商务、迁移、漫游和机器对机器(M2M)服务。在2012年,“折扣”和“电信”类移动虚拟网络运营商是市场上最普遍存在的类型。其占比为46%,而另外19%来自相邻的行业(如零售商、银行、电视/媒体组织),余下的34%聚焦于商务、迁移、漫游

收稿日期: 2015-06-18
网络出版时间: 2015-08-24

中图分类号: TN91 文献标志码: A 文章编号: 1009-6868 (2015) 05-0007-003

摘要: 基于全球移动虚拟网络运营商(MVNO)的最新发展态势,按照区域分类和目录类型(折扣、电信、媒体/娱乐、零售、商务、迁移、漫游和机器对机器),对移动虚拟网络运营商发展情况分别进行了阐述。认为移动虚拟网络运营商与基础运营商不存在严格的竞争关系,虚拟运营商发展的繁荣反而是驱动传统运营商业务增长的动力。

关键词: 移动虚拟网络运营商; 移动运营商; 用户接入量; 每用户平均收益值

Abstract: In this paper, the latest global mobile virtual network operator (MVNO) footprint is introduced. By dividing MVNOs into region and classification, e.g., discount, telecom, retail, media/entertainment, business/migrant, roaming and M2M, the landscape of each segment is described. MVNOs and traditional operators do not have a strictly competitive relationship, and the prosperity of virtual operators is driving the growth of traditional operators.

Keywords: MVNO; MNO; connection volume; average revenue per user (ARPU) value

和M2M。随着运营商持续与知名品牌跨越电信业、媒体/娱乐和零售的合作,移动虚拟网络运营商已经有了相当数量的用户连接规模。维珍移动,一家从固定宽带和数字电视发展起来的英国移动虚拟网络运营商; PosteMobile,一家意大利以邮递业务起家的移动虚拟网络运营商。截至2013年底,两者都有近300万的用户连接数。挪威和瑞士的移动虚拟网络运营商数量分别为17家和16家,德国有122家虚拟网络运营商。从欧洲移动虚拟网络运营商的每用户平均收益(ARPU)效果分析,移动虚拟网络运营商的数量与该国家的ARPU值存在相关性。

1 虚拟运营商的分类

虚拟运营商可以按多个维度进行分类,本文主要介绍按区域和目录

类型划分的虚拟运营商发展情况。

1.1 区域划分

2012年,当时全球57个国家有812个移动虚拟网络运营商。两年来这个数量增长了16%,共涉及69个国家。其中发展最显著的市场在巴西和土耳其,因为在这两个国家,移动虚拟网络运营商可以申请临时牌照。移动虚拟网络运营商推出的数量在2011年和2012年间下降了20%,直到2013年才恢复增长(注:分析只包括那些已经证实推出日期的移动虚拟网络运营商,占总量的2/3左右)。

虽然欧洲地区的MVNO数量占全球的份额在2013年和2014年间从65%下降至61%,但是大多数的移动虚拟网络运营商仍然位于欧洲。截至2014年5月底,美国地区的虚拟运

营商数量显著增长 49% 至 128 家, 占全球虚拟运营商的 14%。同时期, 亚洲的移动虚拟网络运营商数量也提高到 79 家 (占全球虚拟运营商的 8%), 而大洋洲仍然保持在 53 家 (占全球虚拟运营商的 6%)。非洲的虚拟运营商数量则从 4 家发展到了 8 家。额外的区域, 即那些没有地理边界且目标客户群体为移动漫游和无特定市场的虚拟运营商。这些移动虚拟网络运营商占全球 10% 份额, 并自 2012 年以来数量没有变化。区域多元化虽然略有增加, 但是移动虚拟网络运营商仍主要集中在少数国家。截至 2014 年 5 月底, 排名前 10 的移动虚拟网络运营商市场是德国 122 家、美国 96 家、英国 76 家、荷兰 56 家、法国 51 家、澳大利亚 43 家、丹麦 43 家、西班牙 32 家、波兰 28 家和比利时 24 家。在欧洲, 移动虚拟网络运营商仍然主要集中在北部和西部的一些区域。德国、英国、荷兰、法国的集合占一半以上。相比之下, 全球有 12 个市场只有一个移动虚拟网络运营商, 它们分别是阿根廷、白俄罗斯、保加利亚、克罗地亚、塞浦路斯、关岛、斐济、冰岛、印度、拉脱维亚卢森堡。

有移动虚拟网络运营商的国家仍然有限——只有不到世界三分之一的国家有移动虚拟网络运营商。主要原因在于移动虚拟网络运营商牌照的发放。有许可的移动虚拟网络运营商主要适用于饱和的移动市场。在这些市场, 运营商寻求创新的方式来吸引新顾客。在早期的 69 个国家中, 相比全球平均水平的 98%, 托管移动虚拟网络运营商平均普及率为 129%。移动虚拟网络运营商在欧洲有较大的优势。因为该地区具有最高的地区普及率 (141%) 及较早支持并建立移动虚拟网络运营商的缘故。然而, 随着中国和巴西等一些国家移动虚拟网络运营商的增加, 预计欧洲移动虚拟网络运营商市场所占的大份额比率将会在未来几年逐

步减少。

1.2 目录划分

GSMA Intelligence 已确定 8 个不同类别的移动虚拟网络运营商: 即折扣、电信、媒体/娱乐、迁移、零售、商业、漫游和 M2M。8 个类别定义分别如下:

- 折扣: 以低成本服务为主旨的移动虚拟网络运营商
- 电信: 提供一系列电信服务如固话及宽带互联网的移动虚拟网络运营商
- 媒体/娱乐: 关联媒体和娱乐行业的移动虚拟网络运营商
- 迁移: 主要提供国际语音服务的移动虚拟网络运营商
- 零售: 与消费零售业相关的移动虚拟网络运营商
- 商务: 主要为目标商务客户服务的移动虚拟网络运营商
- 漫游: 通常针对国际旅行者, 以多个国家 MNO 的漫游协议为基础的移动虚拟网络运营商
- M2M: 支持 (嵌入式) 机器对机器服务的移动虚拟网络运营商

此外还有一种跨所有上述 8 个类别的类别——仅数据 (通过数据卡、加密狗或嵌入式设备提供数据服务, 但不提供蜂窝语音服务的移动虚拟网络运营商)。

2012 年, 折扣是最突出的移动虚拟网络运营商商业模式, 占全球移动虚拟网络运营商的 1/4 采用此模式。接着是占 21% 的电信类型。而邻近行业的分类, 如零售, 银行和媒体机构则占 19%。剩下的 34% 为迁移、M2M 和漫游。自 2012 年以来, 8 种类型中商业和媒体/娱乐增长基本持平, 其他类型都出现了两位数增长。在这些类别中最高增长比例的是漫游 (32%), 其次是 M2M (26%)、零售 (24%) 和折扣 (22%)。同时, 通过加密狗、便携本和嵌入式设备等方式仅提供数据服务的移动虚拟网络运营商 (跨所有上述 8 类别) 增加了 34%

到 102 家, 占市场 11%。

2 MVNO 的用户接入数量

需要重点指出的是: 虽然媒体和零售类别分别只占全球移动虚拟网络运营商总数的 8% 和 11%, 但从用户连接数量上看两者的占比却非常大。媒体/娱乐业的维珍移动截至 2013 年底在英国本土的连接用户数为 300 万。在零售分类中, 意大利的 PosteMobile 和 COOPVoce 拥有用户连接数分别为 284 万和 48.5 万。而法国的 La Poste 则有 94.3 万的连接数。因此, 这些移动虚拟网络运营商均拥有前 10 的用户连接数。同时, 电信和折扣类移动虚拟网络运营商占了前 20 位的一半。这些类别中也包括虚拟运营商集团, 如德国的 Freenet 和 Drillisch 电信。这两家移动虚拟网络运营商由多个移动运营商托管, 报告的连接数分别为 1 329 万和 194 万。

如上所述, 商务和漫游类是瞄准准机会市场的分类, 并不列入前 20 之中。M2M 分类属于专业范畴, 其中的一个大玩家, Numerex, 具有排名第 3 的连接数。Numerex 研发了一些包括安全、政府、卫生、能源、金融服务和运输领域的 M2M 产品和服务。因此需要相应访问移动网络的能力。前 20 中迁移类的代表为德国的 Turkcell 和意大利的每日电讯。Turkcell 的目标在德国本土, 连接数据为 40 万。而每日电讯服务超过 50 万的用户, 其中一半为居住在意大利的中国人 (总计 24.9 万), 这些人使用每日电讯的语音电话打给中国和香港。

移动虚拟网络运营商的角色有助于运营商提升用户连接数量。发展移动虚拟网络运营商对移动运营商来说是一个有吸引力的战略。基于 2 个原因: (1) 移动运营商可以利用移动虚拟网络运营商的不同品牌获得更多的客户, 以维持或增加成熟市场中移动运营商的市场份额。(2) 通过出租自己的网络拓展业务细分, 并产生额外的收入。例如, 智利的

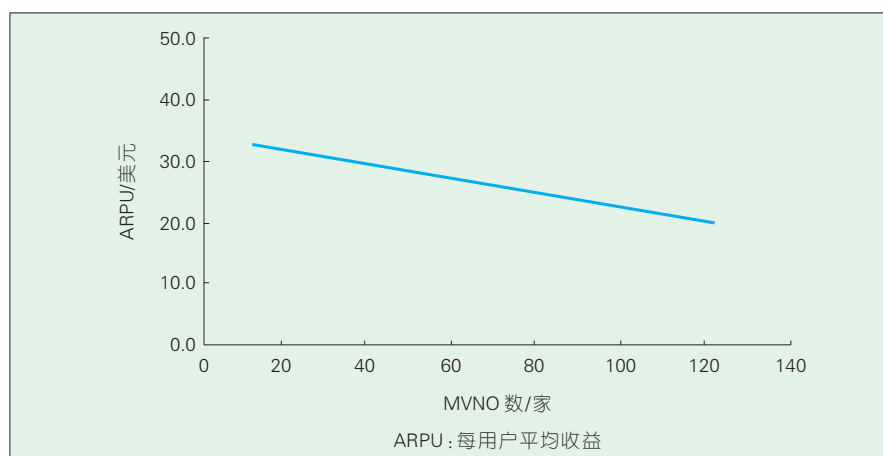
Nextel 公司最近在智利宣布打算推出一个预付移动虚拟网络运营商,此举使 Nextel 公司更具竞争力。该运营商的总裁,Estanislao 解释说,“当前商业模式的重点在于聚焦高价值客户和当前的成本结构以确保高水平的服务质量、频谱容量和售后服务。这是一种盈利微薄的预付费市场。如果移动虚拟网络运营商的监管获得批准,公司有意作为移动虚拟网络运营商提供预付费服务同时继续经营自己的网络服务合同”。GSMA Intelligence 已确定 4 个指标性的市场以展示移动虚拟网络运营商在推动运营商的用户连接数量增长方面的作用以及移动虚拟网络运营商模式受到的限制。

3 MVNO ARPU 值影响的研究

移动虚拟网络运营商对许多运营商富有吸引力。他们也越来越被国家监管机构看作是一种引入更大竞争市场而无需大规模结构调整的工具。早期的移动虚拟网络运营商牌照导致了我们在今天在欧洲看到的高水平竞争。试图复制这种情况的国家包括中国和巴西。

巴西主要是由四大控制了 98% 市场的运营商统治。尽管监管机构 Anatel 允许移动虚拟网络运营商在 2010 年开始运营,到目前为止也仅有 3 家活跃的虚拟运营商: Anaconda Travel SIM、Datora 电信和 Porto Seguro。最近,Anatel 为增加竞争已经修改了规则。其中一项新的规定将要求移动运营商明确在他们的网站上公布其向移动虚拟网络运营商收取的批发价格。

高比例的移动虚拟网络运营商集中于“折扣”的分类。它们的存在能够对市场的价格和 ARPU 值带来显著影响。图 1 显示了欧洲北部、西部和南部这些成熟市场目前移动虚拟网络运营商数量与地区 ARPU 值之间的相关性。



▲ 图 1 移动虚拟网络运营商数量与地区 ARPU 值之间的相关性 (2013 年 4 季度)

其他因素也影响收入,2013 年底挪威的 ARPU 值为 44.76 美元,瑞士为 43.50 美元。两地区移动虚拟网络运营商数量分别为 17 和 16 家。前两者的 ARPU 值是德国的两倍 (18.66 美元),但德国地区有 122 家移动虚拟网络运营商。

从移动虚拟网络运营商数量上看,德国是世界上最大的移动虚拟网络运营商市场。其领先作用已经存在于市场相当长一段时间,其平均 ARPU 值还在继续下降。2014 年 4 月份,E-PLUS 公司与 WhatsApp 公司签署了一份合同,旨在有效地在德国地区服务于 WhatsApp 应用。E-PLUS 负责销售以 WhatsApp 冠名的预付费 SIM 卡。即使客户没有足够的数据流量,他们使用该卡也可以无限制地访问 WhatsApp。SIM 卡的基本包为 10 欧元 (或 14 美元)。但是支付额外的 10 欧元可以获得 600 积分,每积分可以兑换 1 MB 的数据流量、1 min 语音或一条短信。据报道该地区 90% 的智能手机都安装了 WhatsApp。这给 E-PLUS 提供了一个良好的机会:从竞争对手——运营商那挖掘顾客,同时帮助 WhatsApp 自身建立电信部门。

4 MVNO 的演变

移动虚拟网络运营商市场的发展一直受到欧洲监管机构的鼓励。最近已经有移动虚拟网络运营商被

直接用作监管工具的例子。欧洲委员会 (EC) 希望利用移动虚拟网络运营商作为工具来维持市场竞争,如在运营商并购批准之前运用相关的移动虚拟网络运营商接入条件可作为判断依据。在爱尔兰 O2 (Telefonica) 合并之前,运营商按承诺应该出售高达 30% 合并公司的网络容量给两家移动虚拟网络运营商。为了收购德国 E-PLUS 获得欧盟 EC 的批准,Telefonica 也同意类似的条件:可能会在德国建立 3 个新的移动虚拟网络运营商。

随着市场的成熟,其他国家的监管机构对移动虚拟网络运营商也越来越受欢迎,如在中国,虚拟运营商牌照首次授予于 2014 年并且年底总数超过 40 家。然而,移动虚拟网络运营商对整体市场的影响仍然可以忽略不计。

在中国移动虚拟网络运营商市场大的进展是获得虚拟运营商牌照的企业名单,包括互联网的运营商如百度、阿里巴巴和手机厂商小米,以及联想和富士康。不过这些公司将如何利用其牌照仍有待观察。在每一种情况下,他们会看到一个虚拟运营商与他们的核心业务垂直整合的潜在利益以及整体移动客户生态系统发挥的更大作用。再比如,在欧洲已经看到第三方生态系统的运营商

➡ 下转第 20 页

全球移动虚拟网络运营商发展经验 对中国的启示

Development Experience of Global MVNO for China

王雪梅/WANG Xuemei

(中国信息通信研究院泰尔数据研究中心, 北京 100191)
(Data Research Center, China Academy of Information and Communications Technology, Beijing 100191, China)

自2013年起至今,移动通信转售业务申请审批工作已全部结束,工信部共发放了5批试点批文,42家企业获得了虚拟运营商牌照。按照工信部的规定,目前这些企业拿到的只是两年试点批文,只有完成考核之后才能拿到正式牌照。业内预估,到2015年底试点期结束,至少有60%以上的虚拟运营商将面临淘汰。业内预计2015年底中国MVNO总用户数有望达到1 800万。

1 全球MVNO用户占比稳定,欧美地区MVNO发展成熟

虚拟运营的萌芽始于欧洲。1998年,挪威的Sense公司成为全球第一家虚拟运营商,但受当时管制政策限制,致使Sense的经营一直未能取得进展,最终只能宣布破产。随后瑞典和法国相继出台了真正的开放政策,自此虚拟运营商开始蓬勃发展,走向了北美、亚太、非洲等地区。从虚拟运营发展到现在的17年间,其经营主体呈现多样化趋势,涉及零

收稿日期: 2015-06-25
网络出版时间: 2015-08-24

中图分类号: TN91 文献标志码: A 文章编号: 1009-6868 (2015) 05-0010-003

摘要: 指出统计结果显示全球移动虚拟网络运营商(MVNO)用户数始终围绕在总移动用户数1.7%的比例左右,最成熟的市场可达到10%以上的占比,而且从国际虚拟运营商发展经验来看,在这个存活率仅为三成的市场中,与基础运营商良好合作、专注市场细分才是发展关键;认为虽然中国虚拟运营商放号一年多的时间已经发展780万用户,但对基础运营商的用户市场尚构不成威胁;建议中国虚拟运营利用4G移动宽带发展带来的新商机,避开以低价争夺低端市场的竞争手段,定位于智能终端和移动宽带市场,通过新的业务模式拓展市场空间。

关键词: 移动虚拟网络运营商;移动基础网络运营商;市场细分

Abstract: In terms of the global market, mobile virtual network operator (MVNO) has a roughly 1.7% market occupancy of total mobile subscribers. This could reach up to 10% in a mature market. Given the development of MVNO in the foreign market, a collaborative relationship with MNO and focus on market segmentation are the key factors to the development of MVNO in a "30% survival rate" market. Since MVNO launched the business in May 2014, its subscriber has reached more than 7.8 million, yet the existing MNO is still dominating the Chinese market. It is suggested that MVNO in China should seize the opportunities of 4G mobile broadband development, avoid low-prices competition, positioning in the smart terminal and mobile broadband market, and expand the market space of the new business model.

Key words: mobile virtual network operator; mobile network operator; market segmentation

售、广告传媒、娱乐休闲等行业^[1]。

据不完全统计,目前全球共有1 300家左右虚拟运营商,其中八成左右集中分布在欧洲和北美地区,亚太地区占比不足10%^[2]。

自2011年下半年全球MVNO用户首次突破1亿大关以来,一直处于稳定而缓慢上升状态,但始终占总移动用户1.7%左右的比例。截至2015年1季度末,全球共有移动用户71.67亿,其中MVNO用户1.21亿,占比为1.69%。

从区域来看,对于管制政策宽

松、市场竞争激烈而发展成熟的西欧和北美地区,集中了八成以上的移动虚拟运营商(MVNO)用户,这两个地区的MVNO用户渗透率也达到10%以上,而其他地区MVNO用户占比均不足1%。

我们同时应该看到,对于市场成熟的欧美地区,其发展也花费了十余年的时间,2001—2004年,管制机构提供支持使得虚拟运营商们的生存环境得到改善,而且2000年西欧地区的移动通信渗透率为62%,北美在2003年时移动渗透率刚刚突破50%,

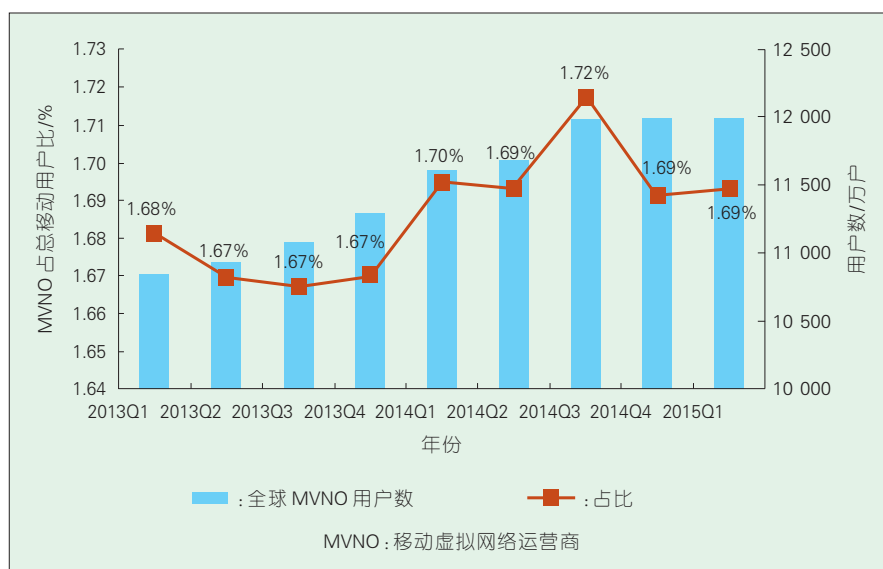
都为虚拟运营商这一新进入者提供了较大的发展空间。而中国2013年底工信部发放移动转售牌照时,中国的移动通信渗透率已达到九成以上,趋于饱和状态,更为虚拟运营商的发展增加了不小的难度。因此,对于刚起步的中国虚拟运营市场来讲,短期之内,民营资本不会对基础运营商造成实质性冲击,而随着移动通信市场的不断饱和,相信长期来看,MVNO也不会对基础运营商们构成威胁。2013Q1—2015Q1全球MVNO用户发展情况如图1所示。

从国家来看,全球MVNO用户排名前十的国家中,大多数也集中在欧美地区,其中德国的MVNO用户占比高达近三成,该国电信市场共有122家虚拟运营商,居全球首位;美国电信市场共有96家虚拟运营商和英国电信市场共有72家虚拟运营商^[3]。全球MVNO用户分区域占比如图2所示。全球MVNO分区域用户数及在移动用户中的占比如表1所示。全球MVNO用户数TOP10国家及在移动用户中的占比如表2所示。

2 MVNO市场淘汰率高

相关数据表明,全球的虚拟运营商存活率仅为三成,其中实现盈利的MVNO更是少之又少,而导致其经营不善乃至被市场淘汰,除受监管限制外,未与基础运营商达成良好合作关系、市场细分错误及业务创新不足也是重要原因。而研究那些运营成功的MVNO,无外乎都是懂得与基础运营商开展合作关系从而避开与其直接竞争,以及专注于细分市场的虚拟运营商。

英国维珍移动作为全球公认最成功的MVNO,据不完全统计,目前在全球已拥有超过2500万MVNO用户。维珍移动在拓展全球市场时,几乎每到一国都会拉上当地一基础运营商成立合资公司,以确保本地化和细分市场互补。在英国、澳大利亚、美国、印度等国家,维珍移动与本地



▲图1 2013Q1—2015Q1全球MVNO用户发展情况

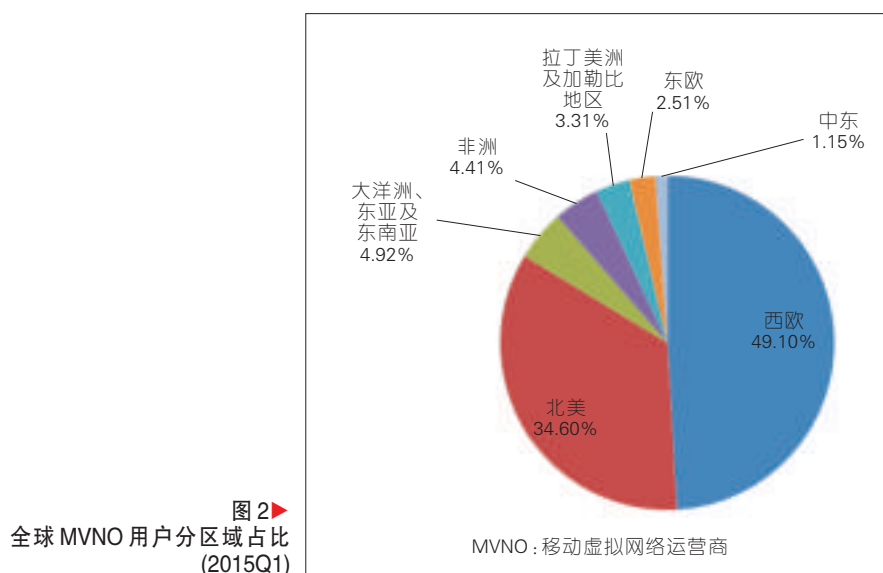


图2
全球MVNO用户分区域占比 (2015Q1)

▼表1 全球MVNO分区域用户数及在移动用户中的占比

地区	MVNO用户数	MVNO用户占比/%
西欧	5 957.14 万	11.10
北美	4 197.84 万	10.23
拉丁美洲及加勒比地区	304.90 万	0.60
东欧	535.53 万	0.59
非洲	401.73 万	0.55
中东	139.47 万	0.39
大洋洲、东亚及东南亚	597.19 万	0.25

运营商成立50:50的合资企业,将网络运营商的利益捆绑进MVNO公司,开展MVNO业务,从而获得网络运营

商的支持。在2001年,维珍移动进入美国发展时,发现了每用户平均收益(ARPU)较低的低端年轻用户市场

▼表2 全球 MVNO 用户数 TOP10 国家及在移动用户中的占比

国家	MVNO 用户数	MVNO 用户占比/%
美国	4 109.08 万	10.76
德国	3 175.31 万	29.07
英国	963.23 万	11.84
西班牙	665.49 万	12.92
意大利	402.00 万	4.69
法国	347.56 万	4.92
哥伦比亚	308.00 万	6.02
南非	257.77 万	3.09
马来西亚	206.96 万	4.48
塞内加尔	178.23 万	12.50

空缺,与美国音乐电视(MTV)网签订了多年的内容提供和市场推广独家协议,为用户提供音乐、游戏和其他 MTV 娱乐内容,打造的手机品牌 Virgin Xtras 添加了诸如“解救电话”等一系列年轻人喜闻乐见的功能,这都能够使得 Virgin Mobile 第一年就在美国市场实现了突破 100 万用户的好成绩。

Amazon 在某种意义上也在经营着移动转售业务,在 2012 年推出的具有 LTE 功能的 Kindle Fire HD 时,与 AT&T 合作推出了支持 LTE 网络的数据流量套餐:一年费用 49.99 美元,包含每月 250 MB 流量、Amazon 云存储服务中 20 GB 的存储容量,并提供给支付账户 10 美元,可在软件商店消费。与 AT&T 每月 250 MB 的数据流量相比,用户一年可节省大约 130 美元。通过这种模式,降低了用户浏览网页、选购商品的成本,增强了用户黏性。Amazon 开展的此项业务不计较移动服务的成本,但是从整体系统性利益最大化考虑这条策略,值得中国经营移动转售业务的电商们参考。

3 中国 MVNO 发展时间紧任务重,4G 是发展突破口

2013 年年底,工信部颁发首批“移动通信转售业务”运营试点资格,2014 年被认为是中国 MVNO 发展元年。自 2014 年 5 月份开始陆续放号以来,中国一些虚拟运营的先行者已

在中国市场获得了较大的客户群。截至 2015 年 6 月底,28 家虚拟运营商开展了业务,在一年的时间内发展超过 780 万的 170 号段用户,而这些用户也集中在蜗牛移动、苏宁互联、国美极信等少数前两批拿到试点批文牌照的 MVNO 中,后几批拿到牌照的虚商们预计 2015 年下半年将陆续展开业务。尽管近期中国虚商用户数以每月 100 万的速度增长,但这一乐观的数字背后也面临着用户离网率高、每用户平均收益(ARPU)低等关键性问题^[4]。

在 2015 年年底试点期结束后,国家将吊销发展不佳的企业转售业务牌照。对于较晚拿到牌照的 MVNO 来说,距离试点结束的时间所剩不多,搞清移动转售的发展方向才是关键。对于前期已经大规模放号的企业,势必要继续投入下去,如果中途退出则前期的投入成本也将石沉大海,血本无归。前期 MVNO 每发展一个用户都要企业贴钱,业内人士估算当用户规模达到千万量级时才可能实现盈利。而近期,MVNO 发展遭遇了寒流,继乐语妙 more 关闭了北京旗舰店后,连连科技也被传解散线下运营团队,尽管很快消息被辟谣,但是种种情况都加重了后面几批拿到试点牌照企业的顾虑。

尽管 2013 年时,中国的移动通信渗透率已高达 90%,对于 MVNO 们来说已错过发展黄金时期,比全球虚拟

运营发展高峰期错后了十余年,但是此时的 MVNO 也迎来了 4G 时代。移动宽带的发展,会给虚拟运营商带来新的商机,竞争焦点不再集中于以低价争夺低端市场,而是定位于新兴的智能终端和移动宽带市场,通过新的业务模式寻找新的市场空间。与此同时,基础运营商也以开放合作的姿态迎接 MVNO,中国移动已率先做出表率,公布全面向虚拟运营商优化 4G 业务资费,全力支持虚拟运营商的市场发展。近期凤凰资产在中国电信支持下,成为中国第一家开展电信 4G 移动转售业务的虚拟运营商,主打金融与互联网融合产品。中国联通也宣布启动 4G 转售,将在 7 月向虚拟运营商开放 4G 网络服务。3 家运营商全面开启 4G 转售将为虚拟运营商的 4G 业务运营打下良好的基础,为其提供更多选择与创新机会^[5]。

参考文献

- [1] 张晶. MVNO 发展之路艰难而漫长 [R]. Teleinfo 智库观察, 2015
- [2] 全球虚拟运营商发展情况概述 [R]. ARCH, 2015
- [3] 西欧市场虚拟运营商的生存之道 [R]. 德国电信国际咨询公司, 2014
- [4] 王雄伟. 前景广阔道路艰难——虚拟运营商(MVNO)中国发展前景浅析 [J]. 中国新通信, 2014, 12(8): 49-51
- [5] 于斌. 中国虚拟运营商在 4G 时代的发展策略 [J]. 中国商贾, 2014, 25(34): 171-173

作者简介



王雪梅, 中国信息通信研究院泰尔数据研究中心助理工程师; 主要研究方向为通信行业市场、宽带接入与业务、虚拟运营商等; 已发表科技论文 1 篇。

移动通信转售批发价对转售产品的影响

Mobile Communication Resale Wholesale Price for Resale Products

彭荣林/PENG Ronglin

(深圳中兴视通科技有限公司, 深圳 518057)
(Shenzhen ZTE SeeCom Tech. Co., Ltd., Shenzhen 518057, China)

1 中国移动转售业务的发展

2013年1月7日,《移动通信转售业务试点方案》公开征求意见,2013年5月17日工信部通[2013]191号文件发布《工业和信息化部关于开展移动通信转售业务试点工作的通告》,移动转售的试点工作正式开始,转售业务的试点期截止期为2015年12月31日,试点结束后将根据试点情况适当调整相关政策。

2013年12月26日,工信部正式颁发第一批牌照,至2014年12月18日,合计为42家颁发了转售业务试点牌照。42家虚拟运营商,一般可分为8类:渠道、互联网、电商、终端、金融、行业、内容/服务提供商(CP/SP)、宽带商^[1]。虚拟运营商分类及明细如表1所示。

从2013年5月17日,虚拟运营商正式推出转售业务开始,截至2015年5月底,共发展410万户^[2],中国联通转售业务发展量占比80%以上。

由媒体宣传及用户反馈来看,转售业务主要经历了以下几个阶段:基础运营商跳槽潮、首家拨测成功、首家放号、规模放号、套餐及业务融合

收稿日期: 2015-06-29
网络出版时间: 2015-08-24

中图分类号: TN91 文献标志码: A 文章编号: 1009-6868 (2015) 05-0013-004

摘要: 分析了因中国的移动转售产品依赖于转售批发价的限制,在转售价格和开放的转售功能的影响下,各虚拟运营商推出的产品受到的影响以及各虚拟运营商产品受自身转售业务的目标定位、自有业务以及成本价格的影响;针对现阶段转售业务现状,提出了批零倒挂问题需要有标准、运营商不能按自身价格体系来制订转售结算价格、虚拟运营商在商业模式不清晰情况下可以和有通信提供能力企业先合作试点等建议,以避免高投入无回收情况。

关键词: 移动转售业务;转售批发价;虚拟运营商;移动转售产品

Abstract: The resale product of China relies on the limitation of resale wholesale price. Products launched on the market by virtual network operators have been affected by the resale price and open resale function since mobile resale business went live. At the same time, the products of each virtual network operator are affected by their own resale business positioning, existing business, and operating cost. Focusing on the status of resale business, this paper proposes that certain criteria should be established for zero upside down, mobile operators cannot set resale products' selling rate according to their own pricing system, and virtual network operators are allowed to cooperate with companies which are capable of offering mobile service to launch pilot projects if the business pattern is not clear. In this way, high input and low return can be avoided.

Keywords: mobile resale service; resale wholesale price; virtual network operator; mobile resale product

创新、规模上量等,其中夹杂着互联互通、银行无法互通短信等问题以及到中期的低潮期媒体唱衰期。但也有企业开始大规模销售放号,发展用

户超过100万户。

目前,仍有一半的虚拟运营商仍处于积极准备上线阶段,预计2015年下半年开始还没有上线的企业将

▼表1 虚拟运营商分类及明细

分类	虚拟运营商
渠道	天音、爱施德、乐语、迪信通、话机世界、中邮普泰、恒大和
互联网	蜗牛、优酷、巴士在线
电商	阿里、京东、苏宁、国美
终端	小米、联想、海尔、富士康、联想、海信、红豆
金融	平安保险、民生银行、银盛、中期、连连科技
行业	海航、朗玛、用友、长江时代、中兴视通、青牛、凤凰、星美
CP/SP	北纬、远特、三五互联、远特、分享通信、世纪互联、263网络
宽带商	鹏博士

陆续上线推出业务。

2 基础运营商转售条件

工信部的转售业务试点方案规定转售业务的申请者应与基础电信业务经营者签订移动通信转售业务商业合同,签订合同后,再提交牌照申请^[1]。所以虚拟运营商与基础运营商的合同条件十分重要,转售的价格和转售条件主要由基础运营商的合作协议确定。

2.1 3家运营商的转售价格

按截至2015年6月份的转售条件,根据本文的分析主要为语音、流量的价格。转售价格的结算方式主要分为资源池模式,即按单价批发与基础运营商结算。模组结算方式是根据基础运营商设置的套餐及模组包,按用户订购量进行结算。三大运营商转售价格标准^[4]如表2所示。

除了转售的基本价格外,三大运营商开放的通信功能也是虚拟运营商开展业务的基础,初期运营商除了基础通信功能外,开放的功能并不多。三大运营商开放的功能列表如表3所示。

2.2 运营商转售条件

(1)转售结算方式

三大运营商转售价格条件中,在

▼表3 三大运营商的转售开放功能

运营商	基础通信功能	语音附加功能	国际业务	其他功能
联通	语音、数据、短信、彩信	来电显示、呼叫转移	9.5 ~ 9.8 折	无
电信	语音、数据、短信、彩信	来电显示、呼叫转移	无折扣	无
移动	语音、数据、短信、彩信	来电显示、来电提醒、呼叫转移	无折扣	中国 WLAN、网内语音

初期,中国联通采用了最为灵活的资源池模式;中国电信提套餐模组转售,转售折扣最低,中国电信在2014年中期,也提供了资源池的结算方式;中国移动采用模组套餐结算方式,虚拟运营商可根据模组组合成转售产品。如果把虚拟运营商比喻成卖面包的话,联通提供的是面粉,虚拟运营商可以做成各种各样的面包;电信提供的是现成的面包,虚拟运营商销售电信现成的面包;移动提供的是面包块,虚拟运营商可以把几个面包块组成不同大小的面包组合^[5]。

随着转售业务的开展,三大运营商根据运营情况,也进行了调整,联通在转售结算基础上,分别对各虚拟运营商进行某段时间发展用户的结算价折扣的调整,进行了短期的促销政策;电信在套餐转售的基础加上了资源池的结算方式;中国移动也在模组套餐方式上,加入了模组转售和资源池模式可自由选择的方式^[6]。在基础运营商提供的转售条件中,虚拟运营商仅可使用一种结算方式,不能在

同类的转售业务中使用两种结算方式,例如中移动的语音转售,虚拟运营商可以使用模组转售或者资源池方式,不可以一部分用户使用模组,一部分用户使用资源池的方式。

(2)转售价格定价

由于当前用户的消费行为主要为语音和流量的消费,短信和彩信由于OTT业务的冲击,在用户使用中,需求较低,本文主要分析语音和流量的转售价格。

中国联通采用资源池,全国统一价,语音价格0.15元/min,流量价格0.2元/MB,7折折扣,转售的含税价格为,语音: $0.15 \times 0.7 \times (1+11\%) = 0.12$ 元/min;流量: $0.2 \times 0.7 \times (1+6\%) = 0.1484$ 元/MB;结算价格和中国联通的接近一致,从6折至7折,联通按结算给联通的结算金额,每结算2000万元,在2000万元阶梯上降低0.1折,例如,在2000万元内结算价格为7折,2000万元~4000万元结算价格为6.9折。鉴于前期各虚拟运营商的业务发展较少,一般为7折结算。

中国电信提供的套餐折扣,按套餐的折扣价格为40%~52%;中国电信后期提供资源池结算方式,结算价格和中国联通的接近一致,从6折至7折,联通转售价格阶梯为2000万元,电信为1000万元。

中国移动提供的套餐模组和中国移动的在售的语音、流量的加油叠加包一致,6.6~7折提供。在资源池的单价批发条件上,流量基础价格为0.18元/MB,7折折扣,结算价格为联通的9折,税后成本为0.1336元/MB。语音分为长市语音,基础价格则为0.13元/min,折扣6.6折,语音的基础价格为0.0858元/min,税后结算价格

▼表2 三大运营商转售价格标准

运营商	流量		语音		短信		来显
	资源池	模组	资源池	模组	资源池	模组	
联通	0.2元/MB,折扣6~7折	无	0.15元/min,折扣6~7折	无	0.1元/条,折扣6~7折	无	5元
电信	0.2元/MB,折扣6~7折	电信3G套餐模组,按ARPU值结算比例为40%~52%	0.15元/min,折扣6~7折	电信3G套餐模组,按ARPU值结算比例为40%~52%	0.1元/条,折扣6~7折	电信3G套餐,按ARPU值结算比例为40%~52%	5元
移动	0.18元/MB,折扣7折	移动销售模组,从5元30MB至280元11GB,折扣7折	长市语音:本地主叫0.13元/min;国内漫游主叫0.6元/min,被叫0.4元/min;长市漫合一:主叫0.17元/min。折扣6.6折	长市模组:10元60min至100元820min;长市漫模组:50元230min至400元390min	0.08元/条,折扣6.6折	5元60条到50元1000min,折扣6.6折	6元
ARPU:每用户平均收益				注:转售价格为不含税价格			

为 0.095 元/min, 长市语音用户在漫游时, 拨打电话 0.6 元/min, 接听电话 0.4 元/min; 长市漫语音基础价格为 0.17 元/min, 6.6 折折扣, 税后成本为 0.1245 元/min。

三大运营商给出的转售价格体系, 资源池的单价结算方式, 流量含税成本在 0.13 元至 0.15 元, 语音含税成本在 0.12 元左右。

(3) 转售提供功能

三大运营商提供的转售功能包括基础通信功能包括: 语音、上网、短信、彩信等基础功能, 移动比其他运营商多提供了国内的无线局域网 (WLAN) 功能、网内通话功能。运营商的定向流量和闲时流量这些个性化的流量定制功能暂未开放。

3 虚拟运营商转售产品分析

据 IBM 发布的虚拟运营白皮书, 约 90% 的虚拟运营商希望通过移动互联网转型, 积极迈向“互联网+”时代^[7]。虚拟运营商根据原有业务的不同, 对移动转售业务采用不同的定位, 并根据定位和移动转售批发价格进行不同的产品资费设计。

3.1 虚拟运营商发布产品资费分析

42 家虚拟运营商按转售业务的定价不尽一样, 由于中国移动转售业务进展较慢, 截至 2015 年 6 月上线的产品多为中国联通和中国电信的转售产品^[8]。当前虚拟运营商已发布套餐卖点如表 4 所示。

42 家虚拟运营商, 已发布产品资费的企业 21 家, 除专做 Wi-Fi 转售的北纬通信外, 20 家发布的产品资费有一些的共同点, 根据联通的资源池转售模式, 大部分虚拟运营商都采用了流量不清零、可共享、可转赠等创新做法。由于电信前期套餐转售模式的限制, 电信合作的转售业务都按电信的原有模组套餐或者叠加电信的增值包后重新打包套餐进行销售。

(1) 套餐的转售价格

在套餐的价格上, 各大虚拟运营

▼表 4 虚拟运营商已发布套餐卖点

虚拟运营商	资费卖点
蜗牛	免卡系列, 按年计费
苏宁	至简产品, 越用越优惠, 购物送流量
阿里	懒人套餐, 流量计费, 最低从 7 元起
国美	语音流量均衡配置
迪信通	模组自由组合
巴士在线	可共享套餐, 单价低至 0.08 元
爱施德	C2B 套餐, 自由定制套餐
天音	多人共享套餐
京东	通信特权卡, 购物送流量
话机世界	组合套餐
分享	套餐超出部分按七折优惠
三五互联	0 月租、无套餐
朗玛信息	大众乡音卡、高贵至尊卡和畅爽游戏卡
远特通信	企业及行业融合通信平台
乐语	与健康装备融合销售
连连科技	赢+手机套餐
华翔联信	X 计划, 流量、语音、短信组合
星美生活	最低 5 元, 送来电, 每月送通话
用友	随身、随行套餐, 融合行业应用
长江时代	云套餐、充值更优惠

商在产品套餐毛利定位上, 主要为少量毛利、毛利为 0、略为亏损的 3 种定价策略, 联通转售业务各虚拟运营商转售价格为 0.12 元/MB 到 0.2 元/MB, 语音从 0.09 元/min 到 0.15 元/min 不等; 电信转售业务的套餐折扣从 30% 到 50% 区间。

(2) 套餐创新模式

按联通转售业务的资源池结算模式, 各大虚拟运营商进行了创新设计, 包括阿里按统一单位懒人计费模式^[9]、苏宁的至简产品越用越优惠模式、蜗牛的按年卡计费模式、爱施德的 C2B 自由定制套餐模式^[10]、巴士在线的资源银行模式等。这些套餐创新方式也影响了三大运营商在后期的产品资费的仿照跟进, 各虚拟运营商之间也有相互参照学习做法。

(3) 转售业务融合方式

产品资费融合了虚拟运营商自有的业务体系, 电商行业, 京东、苏宁采用购物送作法, 用户购买商品时赠送流量; 行业应用企业, 例如用友, 在

套餐里加入了赠送自有的行业应用作法; 渠道行业, 乐语将移动转售业务和健康终端融合; 其他行业也出现“通信+金融”、“通信+医疗”、“通信+系统集成”等整合业务。

3.2 虚拟运营商的产品困境

(1) 转售批发价格批零倒挂, 产品资费不具备优势

按现有的转售批发价格, 用户在流量的单价, 按前文的分析, 语音的含税价格在 0.12 元/min, 流量的含税价格为 0.15 元/MB, 在现在通信市场大部分用户都有强流量需求, 虚拟运营商的流量结算价格较高, 在批零倒挂的问题上, 没有定论。我们拿联通最新的套餐来计算, 76 元含 400 MB 和 200 min 赠送来显, 用户参与存 120 得 360 元全国统一活动, 相当于用户实际支付成本为 56 元/月, 而虚拟运营商税前的成本为 80.5 元, 虚拟运营商价格高出 44%。流量套餐在 1 GB 以上的套餐, 差距明显更大。以致有虚拟运营商高管开玩笑说, 和客户介绍转售业务的融合解决方案, 客户对其他方面都很满意, 就是转售业务的号码能否换成三大运营商号码, 价格差距过大, 无法说服用户使用。

故前期各虚拟运营商在推出联通转售产品的时候, 都集中在套餐创新上做文章, 无套餐、无月租、不清零、越打越优惠等各类灵活方式, 在低月租的语音通话市场上寻找机会。

(2) 转售产品开放功能缺乏, 创新业务较少

目前, 除了简单的有关语音、流量的创新如流量不清零、无限语音以及简单的与自身业务叠加, 如购物送流量、玩游戏送流量外^[11], 虚拟运营商并没有真正带来全新的商业模式。创新也依赖于开放的各功能及与自有业务的融合。

在各虚拟运营商的运营过程中, 各虚拟运营商为了融合自有业务, 需要一定的定向功能及定向的优惠功能。例如蜗牛作为一个手机游戏运

营商,前期已实现了三大运营商的定向后量流量,用户在蜗牛的游戏商店下载免流量费,但蜗牛发展的自有移动转售用户却没有相应的定向流量免费功能,在产品资费上缺乏自身业务融合的作法,用户也缺少粘性。

也有部分企业希望进行国际业务的销售,现有的国际业务的结算标准并无可操作空间,虽若按照接近运营商基础价进行销售,则产品将出现亏损,如果提价销售,则可能会给用户带来体验上的问题。

3.3 移动通信转售业务定价思考

(1) 批零倒挂问题需有标准

由于批发价高于零售价,虚拟运营商在发展业务的时候,很难有可操作空间,虚拟运营商都有主营业务,即使企业愿意补贴转售业务用户,在批发价和零售价差距过大的情况下,大投入补贴用户使套餐与基础运营商接近,也会背上较大的成本压力。

无论零售价按基础运营商现行的套餐资费体系以及普遍统一的促销政策进行计算,还是按用户实现使用通信资源和实际支付的价格进行计算,批发价格均应低于零售价格。

(2) 基础运营商定价方式需更加开放

各大运营商的转售价格定价部门在各公司的市场部,基础运营商的定价往往参考了来自有的价格体系来定价,且避免给现行的资费体系带来冲击,基础运营商往往将虚拟运营商当成自己价格体系里的一部分进行管理。

转售价格体系应该在整个中国的价格体系内考虑定价,一个产品套餐的竞争力,除了使用的网络制式之外,更多是公司品牌、服务等关联。

(3) 虚拟运营商应与基础运营商更多的业务互补

在各大虚拟运营商中,主营业务各种各样,在一些细化市场的能力要高于基础运营商,在这些细化市场中,虚拟运营商和基础运营商需要有

互补性,合作共赢。

(4) 虚拟运营商在批发价无优势及自身商业模式不清晰时的对策

虚拟运营商不仅碰到批发价格高,还存在商业模式不清、IT投入成本过高、运营模式不成熟、过早引发的虚拟运营商之间的价格战等问题。一方面,虚拟运营商需要优化自有用户的用户体验,明确自己的商业模式;另一方面,在与基础运营商进行批发价格申请的时候,需要与基础运营商针对批发价格、功能开放和系统实现的多方位考虑才能推动业务往前发展。

虚拟运营商并不一定需要自建系统及独立运营转售业务,企业有可能需要的是会员服务,引流入口,融合通信,串联企业各业务的纽带,更高频次地服务用户等定位需求。商业模式是业务开展的基础,但真正打动用户的是产品,只有产品形成用户的粘性,才能让商业模式成功。而在产品要素中,价格也是一个比较重要的因素。

为避免初期在商业模式未清晰情况下降低风险,IBM建议移动虚拟运营商可以通过几家企业众筹实施IT能力,这样不但可以降低风险和IT成本,还将提升移动虚拟运营商的竞争力。依靠通信能力的提供,也有企业转变商业模式为更多企业提供服务,例如中兴视通作为虚拟运营商,同时可为企业提供基础的通信能力(包括虚拟运营业务及其他运营商能力提供),这也是虚拟运营商一种新模式创新。

4 结束语

虚拟运营商在中国从2014年5月份开始有企业开始放号,至今仅有一年多的时间,在中国不管对于监管部门、基础运营商和虚拟运营商都是一个新生事物,各个部门和企业都在探索尝试各种方法让行业有序健康地发展下去。在转售价格问题上,2014年9月,移动通信转售业务批发

价格动态调整机制发布,双方商务协议中规定业务开通后每6个月协商调整一次,考虑到基础运营商间市场竞争加剧,促销频繁及4G资费主动下调等因素,双方可协商缩短批发价调整周期,甚至视情况适时调整^[12]。

转售业务作为中国的新业务,在工信部、基础运营商、虚拟运营商共同努力下,相信将会有更多创新产品及业务出现。

参考文献

- [1] 新华网. 虚商里土豪扎堆 八大派系抢滩 [EB/OL]. (2015-06-20). http://news.xinhuanet.com/fortune/2014-08/04/c_126829638.htm
- [2] 腾讯科技. 工信部:虚拟运营商总用户超410万 [EB/OL]. (2015-06-20). <http://tech.qq.com/a/20150415/022907.htm>
- [3] 通信发展司. 移动通信转售业务试点政策解读 [EB/OL]. (2015-06-20). <http://txs.miit.gov.cn/n11293472/n11295244/n11297893/15417573.html>
- [4] 宿艺. 虚拟运营商转售成本调查:价格优势尽失外,又来了6个新难题 [EB/OL]. (2015-06-20). <http://www.huxiu.com/article/35958/1.html>
- [5] 三明. 中国式虚拟运营商,就是移动转售 [EB/OL]. (2015-06-20). <http://www.tmtpost.com/74591.html>
- [6] 郭晓峰. 中移动向虚拟运营商开放4G转售业务 [EB/OL]. (2015-06-20). <http://tech.qq.com/a/20150228/011156.htm>
- [7] IBM. 《厚积薄发协同共赢—构建移动互联网时代的虚拟运营架构》白皮书 [R]. IBM, 2015
- [8] 孙卫涛. 体验差+成本高 42家试点虚拟运营商一半未发号 [EB/OL]. (2015-06-20). <http://biz.zjol.com.cn/system/2015/04/23/020618830.shtml>
- [9] 黄远. 用流量打电话 阿里通信公布170“懒人计费”方案 [EB/OL]. (2015-06-20). <http://biz.zjol.com.cn/system/2015/04/23/020618830.shtml>
- [10] 新华信息化. 爱施德发布U.友套餐 开创C2B自由定制新时代 [EB/OL]. (2015-06-20). http://news.xinhuanet.com/info/2014-04/22/c_133280136.htm
- [11] 张海龙. 虚商别把批零倒挂当成不创新的借口 [EB/OL]. (2015-06-20). <http://www.c114.net/vno/4329/a854064.html>
- [12] 中国信息产业网. 移动业务转售建立“动态价格”机制 [EB/OL]. (2015-06-20). http://www.cnii.com.cn/VirtualOperating/2014-09/25/content_1451792.htm

作者简介



彭荣林,深圳中兴视通科技有限公司产品总监;从事基础运营商产品管理、虚拟运营商产品规划工作。

移动通信转售业务标准化发展之路

Standardization Development of Mobile Resale Business

孟然/MENG Ran

李娜/LI Na

李硕/LI Shuo

(中国信息通信研究院, 北京 邮编 100191)
(China Academy of Information and
Communication Technology (CAICT), Beijing
100191, China)

为了落实国务院《关于鼓励和引导民间投资健康发展的若干意见》^[1]精神,鼓励民间资本参与电信建设、鼓励民间资本以参股方式进入基础电信运营市场、支持民间资本开展增值电信业务。工信部于2012年6月出台《关于鼓励和引导民间资本进一步进入电信业的实施意见》(工信部通[2012]293号)^[2],提出鼓励和引导民间资本进入电信行业的八大重点领域,其中移动通信转售业务位于八大领域之首。

2013年5月17日工信部公布了《移动通信转售业务试点方案》(简称《试点方案》)^[3],这标志着中国移动转售业务进入电信业的时代到来。《试点方案》发布后,工信部于2013年12月26日起,先后向42家民营企业颁发了移动通信试点企业批文。2014年5月,首批移动转售企业正式对外放号开展业务,自此移动通信转售业务正式步入了中国通信行业的舞台。

移动通信转售业务正式开展已经一年多的时间,但发展情况与前期各界的预期还存在一定的差距,其中

中图分类号:TN91 文献标志码:A 文章编号:1009-6868(2015)05-0017-004

摘要: 基于中国移动转售业务开展的特点及对标准化体系建立的迫切需求,提出中国移动转售标准化体系建设思路。认为标准化体系建设应从顶层设计入手并为未来业务发展预留充分的可扩展空间,现阶段应以移动转售业务能力开放所急需的互联互通相关标准为重点,参照成熟的通信行业标准体系,逐步形成健全完善的移动转售标准体系。

关键词: 移动通信转售; 标准化; 互联互通; 业务能力

Abstract: This paper proposes a method for constructing a mobile resale standardization system according to the characteristics of mobile resale business and the urgent needs of standard system in China. A standardization system should start from top-level design. Reserving sufficient space for future business development is also important. Currently, the interconnection and interworking standards, which are urgently demanded by developing the ability opening of mobile resale service, are the focus. With reference to the mature communication industry standard system, a sound perfect mobile resale standard system will be built in the future.

Keywords: mobile resale; standardization; interconnection; service ability

忙于解决业务运营中遇到的各类技术问题是阻碍移动转售业务快速发展的因素之一。传统电信行业经过多年的发展具有规范的体系流程,无论在网络技术以及业务规划等方面都形成了一套成熟、完善的标准化体系。移动转售企业来自于各行各业,作为新晋的移动业务运营者,要实现长期健康、有序、可持续发展的业务发展和技术演进,需要借鉴传统电信企业的技术演进管理机制,从自身领域特点和行业发展需要出发,对行业发展中面临的相关业务和技术进行规范,从而形成一套移动转售领域的标准化机制。

1 移动通信转售业务能力开放面临的问题

截止2015年6月底,已有20余家

转售企业正式面向市场开展业务,累计发展用户超过700万户,转售业务基本覆盖全国各省的主要城市。中国联通领跑移动转售业务,用户发展呈加速趋势,到2015年6月中国联通全网移动转售用户规模已经超过600万,进入2015年更是以每月超百万的速度增长。作为移动转售企业的领头羊,苏州蜗牛在2015年3月25日对外宣布,其170用户数已经突破100万大关。进入2015年后,移动转售业务进入平稳发展阶段,用户规模和市场收入均在不断提升,但我们看到,移动转售在互联互通、业务能力提供方面还面临一些发展“瓶颈”。

1.1 企业同时签约多家运营商,互联互通难度增加

《试点方案》中对移动转售业务

收稿日期: 2015-06-29
网络出版时间: 2015-08-24

定义为:移动转售企业从拥有移动网络的基础电信业务经营者购买移动通信服务,重新包装成自有品牌并销售给最终用户的移动通信服务。移动通信转售企业不自建无线网、核心网、传输网等移动通信网络基础设施,必须建立客服系统,可依据需要建立业务管理平台以及计费、营账等业务支撑系统。

根据试点方案的要求,移动转售企业需要通过与基础电信企业的业务平台、网元设施等进行互联对接,才具备提供移动通信服务的能力。

在已经批准的42家试点企业与3家基础运营商之间为多对多的合作关系,其中签约的转售企业,中国电信26家,联通签约25家,移动签约17家。同一家转售企业可以同多家基础运营商签约,转售企业在获取批文后还可以与其他的基础运营商合作,目前转售企业与基础企业签约情况统计如图1所示。

基础运营商虽然针对移动转售业务,面向转售企业提供了对接平台,但也存在一家基础企业面对几十家转售企业,一家转售企业面对3家基础企业展开互联互通的局面。与以往传统的3家基础企业间的互联互通相比,协调和推动的难度增大。

1.2 增值业务开展缓慢,推动协调工作复杂

目前移动转售企业提供的各类产品套餐,主要是基于语音、短信、数据三大基础业务进行的重新设计组

合,而对于彩铃、位置服务、定向流量等基础企业网内发展较为成熟或已有实际应用的业务,移动转售用户还未全面享受到。根据对基础企业与转售企业的调研了解到,对于增值和扩展服务未面向转售用户提供的原

因主要包括如下方面的原因:
首先,基础企业增值业务大多通过基地或者独立的公司进行运营,在集团提供的移动转售一点接入平台中并没有包括增值业务的对接互联功能。移动转售企业如需开通增值服务,需要自行与基础企业增值服务基地或运营公司进行一对一协商,并根据需求确定对接方式。

其次,基础企业网内还处于试验摸索阶段,未形成统一的标准和流程。因此,暂时也未向移动转售用户开放。

综上,目前移动转售业务能力开放中存在的主要问题是业务对接互联过程中的沟通、开发、部署、测试工作都是每家单独进行,造成工作步骤重复,增加成本消耗,并且未来的扩展性差。而造成上述问题的根源在于移动转售企业与基础企业之间的业务能力互联之间没有形成统一的标准和规范,因此要加快业务能力开放进程,降低移动转售企业和基础企业在互联互通工作中的投入,迫切需要对移动转售所设计的业务、技术领域推行标准化体系建设。

2 移动通信转售业务标准化现状

工信部业务主管部门高度关注移动通信转售业务的发展情况,针对业务能力互联互通过程中遇到的各种问题,积极组织相关企业协调解决。2014年2月18日,在移动通信转售试点工作第一次例会上,工信部通信发展司闻库司长指出“业务系统对接工作需要规范化,在此前的工作基础上,求同存异,形成一套统一的接口标准”。为落实第一次工作例会的会议精神,以标准规范保障移动通信

转售的顺利开展,促进转售企业与基础企业之间的互联互通,由中国信息通信研究院(原工信部电信研究院)首席科学家龚双瑾教授负责,3家基础电信企业和获得批准的各家移动通信转售企业参加,共同组建移动通信转售技术标准工作研究小组,启动研制移动转售业务试点需要的业务开通、计费结算、网络服务质量等互联互通技术规范工作。

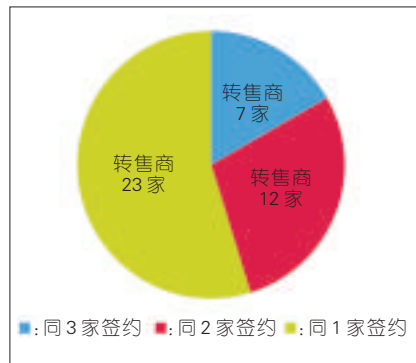
移动转售技术标准研究组于2014年3月启动工作,本着“结合实际,急用先行”的原则,从移动转售业务开通最基础的互联业务功能和互联业务接口两个方面着手进行相关标准的研制。这两项标准经过编制、征求意见和审查,于2014年12月在工信部网站进行公示。

移动通信转售标准中首先明确了双方业务支撑系统的互联架构,移动通信转售企业与基础电信企业通过双方的互联网关平台进行对接,实现业务流程的扭转及业务信息的交互。双方之间的网元、系统和业务平台不直接进行对接互联架构如图2所示。

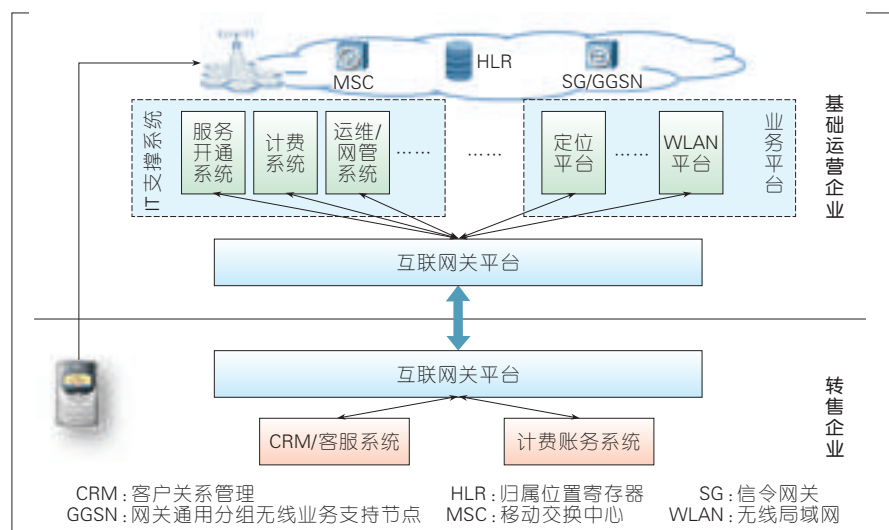
标准中仅是对转售企业与基础电信企业之间的信息交互方式提出要求,而对于各自企业互联网关平台与本企业内部系统平台之间的数据信息交互方式均由各企业自行确定。

《移动转售企业与基础电信企业互联业务功能要求》^[4]是移动通信转售系列标准中的最基础的规范,该标准覆盖移动转售业务开展过程中的各类业务功能要求,包括业务运营、计费结算、服务保障、卡号资源四大方面。

业务运营分为基础业务运营和扩展服务,其中基础服务包括开户、销户、停复机、补换卡、远程写卡、业务变更、业务状态查询、网络状态查询等确保基本移动通信服务的业务能力;扩展服务包括增值服务订购、位置信息服务开通/取消/变更、用户位置信息查询、集团虚拟专网设置、



▲图1 转售企业与基础企业签约情况统计



▲图2 转售企业与基础企业互联总体架构

Wi-Fi转售等。由于移动转售企业对扩展服务的需求以及3家基础电信企业网络支持力度的不同,在移动转售业务开展前期,扩展服务为可选服务项,双方根据实际情况协议提供。互联业务功能标准中针对上述业务运营的每个具体业务场景的互联功能、互联信息内容等进行了明确的相关规定。

计费结算主要是基础企业传递给移动转售企业的话单文件的传递,标准中对话单文件的内容以及话单文件的生成发送时间间隔进行了明确规定。

服务保障对基础企业提供给转售企业的网络及业务支撑系统层面的保障方式进行了约束,分为主动保障和被动保障。其中主动保障是在基础企业发生网络割接或故障中主动通知转售企业;被动保障指转售企业发起故障申告,基础企业进行申告处理的过程。标准中对故障通知及申告处理的方式和交互的信息内容进行了规定。

卡号资源指基础企业向转售企业分配具体的号码资源,根据试点政策工信部向3家基础企业分配移动转售专用号段,转售企业根据业务需要向基础企业申请用户号码资源。双方之间的用户号码资源以及制卡

数据也需要进行交互。标准中对于交互的卡号资源的时机以及交互的内容进行了具体规定。

《移动转售企业与基础电信企业互联业务接口技术要求》^[9]是以《移动转售企业与基础电信企业互联业务功能要求》技术标准中规定的各类互联业务功能为基础,明确了具体的互联接口方式。根据各类业务场景传送的数据内容和要求的不同,互联网关之间的接口包括实时调用接口和文件传送两类。实时调用接口主要针对时效要求高的业务处理,考虑到接口规范及安全等因素,标准规定实时接口采用简单对象访问协议(SOAP)1.2版本协议。文件接口主要针对实时性要求低的大数据量交互,标准规定文件接口采用安全文件传送协议(SFTP),通过二进制传输方式上传/下载文本。

标准中针对每个实时业务接口场景定义了具体的接口业务说明、接口消息命令码、接口提供方式、接口调用方式、接口业务流程以及接口消息字段内容及类型要求;针对每类文件接口,定义了接口的具体描述、文件类型、文件命名方式、文件内容格式以及文件传递周期等要求。

移动转售企业与基础电信企业互联系列标准的制订,既满足了移动

转售业务互联业务的发展需求,又兼顾现有产业发展现状,使基础电信企业、移动转售企业在实现移动转售业务互联互通的工作上,最大程度地减少建设成本,节约资源,对中国移动转售企业在试点初期实现快速业务开通起到了有效地推动作用。

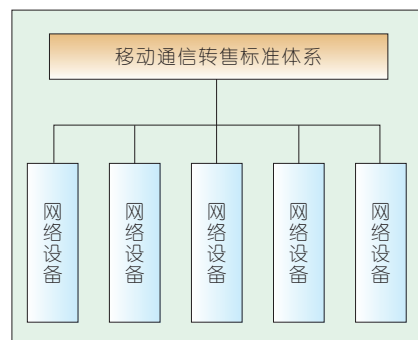
3 移动通信转售业务标准化发展规划

在国际上,欧、美等虚拟运营业务开展较早。根据各国对虚拟运营业务的定义,虚拟运营商对网络业务的涉入程度不同,导致不同开放政策下虚拟运营商与协议基础运营商的互联互通的层面也不同。而且相比于中国试点期间移动转售企业直接与协议基础运营商实现互联对接的情况不同,国际上虚拟运营商有很大一部分是采用通过移动虚拟网络提供商(MVNE)实现与基础企业的对接,互联互通的工作全部交给MVNE与基础运营商去解决,而虚拟运营商将更多的精力投入到品牌和市场的经营上。MVNE多出身于电信行业,大部分还是基础运营商的子公司,因此国际上的相关业务能力对接相对简单,因此也没有形成一套体系化的标准规范。

中国移动通信转售业务试点开展的特殊性,决定了要保障业务顺利、健康开展,要以急需的技术标准和合理的评价测试标准为抓手,从标准体系顶层设计开始,充分考虑未来业务发展需要,整合资源、构建架构层次分明的标准体系。移动通信转售标准体系的建设遵循急用先行、逐步完善的原则,最终实现与现有通信产业标准的有机融合。

中国移动转售标准体系的建立借鉴现有基础通信行业的标准体系,可分为“网络设备”、“互联互通”、“服务支撑”、“终端设备”、“辅助支撑”等五大部分。标准体系规划如图3所示。

(1)网络设备



▲ 图3 移动通信转售标准体系总体规划

网络设备类为未来扩展类标准，主要是针对未来如向移动转售或虚拟运营服务者开放部分网络基础设施，则需要对开放设施的具体技术要求进行规范。

(2) 互联互通

互联互通类标准是目前最为急需的标准系列，包括基础性的功能、接口标准以及具体的业务互联标准，如增值业务、号码携带业务等。未来，随着新业务以及新的服务提供者的出现，互联互通标准化的范围逐步扩展到转售与第三方服务商之间的互联互通。

(3) 服务支撑

服务支撑类主要是针对移动通信转售中一些服务开展进行规范，包括客服号码、充值服务、服务提供商

(SP)服务等。

(4) 终端设备

终端设备类主要是针对转售企业对于移动终端特别是智能终端，提出的一些区别于传统移动运营商的要求制订技术标准以及检测方法。

(5) 辅助支撑

辅助支撑类指针对保证转售业务顺利发展的辅助管理支撑手段进行规范，包括移动转售业务监管平台及配套测试方法等。

4 结束语

标准化工作的开展是移动通信转售业务长期可持续发展的基础，移动通信转售业务的开展离不开基础企业网络和业务系统的支撑，未来业务的发展更离不开跨行业的合作与融合。众多的转售企业都需要与基础电信企业、互联网企业以及其他行业进行互联，只有制订了统一标准，才能使各方在互联互通方面尽快达成一致，从而节省企业的资源，降低成本投入。移动转售业务互联过程中标准化的实施的程度，从某种程度上保障了移动转售业务能够健康、有序、持续发展。

参考文献

- [1] 国务院. 关于鼓励和引导民间投资健康发展的

若干意见[R]. 2010

[2] 工业和信息化部. 关于鼓励和引导民间资本进一步进入电信业的实施意见[R]. 2012

[3] 工业和信息化部. 移动通信转售业务试点方案[R]. 2013

[4] YD/T 2787-2014. 移动转售企业与基础电信企业互联互通功能要求[S]. 工业和信息化部, 2015

[5] YD/T 2788-2014. 移动转售企业与基础电信企业互联互通接口技术要求[S]. 工业和信息化部, 2015

作者简介



孟然，中国信息通信研究院通信标准研究所工程师；主要从事电信网码号资源与移动转售领域相关政策、技术的研究。



李娜，中国信息通信研究院通信标准研究所工程师；主要从事移动转售、虚拟运营、号码携带、码号资源相关领域的政策、标准、技术的研究工作。



李硕，中国信息通信研究院通信标准研究所助理工程师；主要从事移动转售、虚拟运营、码号资源相关领域的政策、标准、技术的研究工作。

← 上接第9页

进入了虚拟运营商市场。WhatsApp在德国推出预付E-PLUS(合并O2之后)网络服务。被消费者视为运营商品牌可以给WhatsApp带来更大的杠杆：推出语音业务可进一步的创收。

报告显示，谷歌也了解到虚拟运营商整合的潜在利益并计划在美国推出基于Sprint和T-Mobile的网络服务。美国移动虚拟网络运营商市场拥挤，不产生高水平的ARPU值。因此任何这样的举措可能是一个更广泛战略的一部分。谷歌把在固定宽带和Wi-Fi基础设施的补充投资作为驱动其互联网应用使用量的最终目标。整合谷歌的全球业务和客户基

础，该公司可能会对移动运营商的业务造成相当大的破坏。参考这种模式，其他大型生态系统运营商也会给移动虚拟网络运营商市场带来非对称业务模型的可能性。

5 结束语

至2014年年底，在全球范围内运作的移动虚拟网络运营商数量上升到了近千家。由于监管政策不断鼓励提升运营商竞争力和激发非传统电信领域运营商的兴致。众多的互联网服务供应商和设备制造商逐步地加入了虚拟运营商市场。

此外，移动虚拟网络运营商与基础运营商不存在严格的竞争关系。

移动虚拟网络运营商生态系统的繁荣反而往往是驱动运营商增长的動力。这对传统的基础运营商来说在当前发展缓慢的市场环境下是一个好消息。

作者简介



George Guo，美国Syniverse公司全球副总裁暨亚太区总经理；全面负责公司亚太区的企业和智能方案事业部业务；曾就职于洛克希德·马丁公司担任技术运营主管，就职于NeuStar公司担任NeuStar亚太区副总裁兼总经理。

移动虚拟网络运营技术研究

MVNO Technology

朱玉娜/ZHU Yuna

王修中/WANG Xiuzhong

(中兴通讯股份有限公司, 江苏 南京
210012)
(ZTE Corporation, Nanjing 210012, China)

移动虚拟网络运营(MVNO)是网络共享的一种模式。网络共享是多个运营商可以共享全部或者部分网络设备。根据网络设备共享的类型,又分为无线网络共享和核心网网络共享^[1-15]。MVNO的应用场景有:

(1)核心网共享无线接入网:每个运营商拥有自己的核心网,共享无线接入网,但不共享频段。

(2)无线接入网共享核心网:各无线接入网,接入到共享的核心网设备。

(3)网络分布式共享:每个运营商拥有自己的核心网和无线接入网,通过策略实现共享。

(4)公共网络共享。

(5)公共频段共享。

要实现MVNO,必须首先解决其所涉及到的技术问题。下面分网元讨论为支持MVNO各个核心网网元需要提供的关键技术。

1 HLR 关键技术

当MVNO采用单纯转售(Wholesaler)或者部分MVNO组网方

收稿日期: 2015-06-16

网络出版时间: 2015-08-24

基金项目: 国家科技重大专项(2012ZX03002-008); 江苏省未来网络创新研究院未来网络前瞻性研究项目(BY2013095-4-16)

中图分类号: TN91 文献标志码: A 文章编号: 1009-6868 (2015) 05-0021-003

摘要: 探讨了为实现移动虚拟网络运营(MVNO)核心网网元需要解决的关键技术问题,包括:归属位置寄存器(HLR)需要提供“虚拟”HLR功能;电路交换(CS)网元的移动交换中心服务器(MSC Server)/拜访位置寄存器(VLR)需要提供等效公众陆地移动网(EPLMN)功能、网络标识和时区(NITZ)功能, MSC Server/VLR 还需要解决路由选择的问题;分组交换/演进分组核心网(PS/EPC)网元服务通用分组无线业务支持节点/网关通用分组无线业务支持节点/移动性管理实体/分组数据网网关(SGSN/GGSN/MME/PGW)需要支持多公众陆地移动网(PLMN)功能以及虚拟路由转发表(VRF)等。

关键词: 移动虚拟网络运营;等效公众陆地移动网;路由选择;多公众陆地移动网

Abstract: Core network elements require solutions in order to realize mobile virtual network operator (MVNO). The home location register (HLR) needs to support virtual HLR function; the mobile switching center (MSC) server and visitor location register (VLR) of circuit switching needs to support equivalent public land mobile network (EPLMN), network identity and time zone (NITZ), and the routing-selection strategy; and serving GPRS support node (SGSN), gateway GPRS support node (GGSN), mobility management entity (MME) and PDN Gateway (PGW) all need to support multiple PLMN functions and virtual routing forwarding (VRF).

Keywords: mobile virtual network operator; equivalent public land mobile network; routing selection; multiple public land mobile network

式时, MVNO 运营商和移动网络运营(MNO)运营商会共用归属位置寄存器(HLR)网元。

为提供MVNO解决方案,归属位置寄存器(HLR)需要提供“虚拟HLR”技术。

“虚拟HLR”技术就是将MVNO运营商和MNO运营商的用户数据集中存放在一个物理HLR中,并将此物理HLR从逻辑上划分成多个虚拟的HLR,不同的虚拟HLR归属不同的运营商,不同虚拟HLR管理的国际移动用户识别码/移动用户的号码(IMS/MSISDN)号段不同。

虚拟HLR具备如下特性:

(1)独立的业务提供能力

(2)独立的业务管理能力

1.1 虚拟HLR独立的业务提供能力

每个虚拟HLR具备独立的业务提供能力。每个虚拟HLR的容量可单独设置,虚拟HLR的容量可在线动态调节。

虚拟HLR划分方式灵活,可以按照IMS/MSISDN号段划分某个用户归属的虚拟HLR,也可以按照IMS/MSISDN单个号码划分某个用户归属的虚拟HLR。

虚拟HLR还需具备独立的业务定义能力,包含如下方面:

(1)可为每个虚拟HLR定义不同的业务功能。

(2) 每个虚拟HLR可定义不同的模板数据(如用户开户模板)、不同的配置数据(如漫游限制列表)。

(3) 针对某个虚拟HLR的用户数据操作不会对其他虚拟HLR产生相关影响。

1.2 虚拟HLR独立的业务管理能力

虚拟HLR独立的业务管理能力包括:分域管理功能、分权管理功能、独立的性能统计功能、独立的告警管理功能。

分域管理功能:属于某一虚拟HLR的操作维护人员,只能对该虚拟HLR内的数据(如用户数据、配置数据)进行查询和修改,不能操作其他虚拟HLR的数据。

分权管理功能:对操作不同虚拟HLR的操作维护人员的权限进行控制,实现权限划分,为不同虚拟HLR对应的操作维护人员分配不同的IMSI/MSISDN号段的操作权限,控制各运营商维护人员只能操作各自对应“虚拟HLR”下的用户数据,且可以进一步控制某些操作维护人员拥有查询、修订的权限,而另外一些操作维护人员只有查询的权限,实现权限的细化控制。

独立的性能统计功能:针对不同虚拟HLR提供相互独立的统计功能,如虚拟HLR内的用户总数、漫游用户数、开通某业务(如开通通用分组无线服务技术(GPRS)业务)的用户数等统计项目,每一个虚拟HLR都可以单独输出各自的统计数量。

独立的告警管理功能:每个虚拟HLR可以定义相互独立的告警策略和告警内容,并单独输出各自的告警信息。

2 电路交换域关键技术

2.1 等效公众陆地移动网

等效公众陆地移动网(EPLMN)主要解决共享网络之间的用户驻留和漫游策略。EPLMN可以让用户自

己选择利用哪个通信网络的资源来提供服务。

在移动交换中心服务器(MSC Server)中配置EPLMN列表,在终端发起位置更新时,MSC Server将EPLMN列表数据通过位置更新接受(LUA)消息下发给移动用户,移动用户将MSC Server下发的EPLMN列表保存在用户识别(SIM/USIM)卡中。

2.2 网络标识和时区

利用网络标识和时区(NITZ)技术,可以在终端上自动显示为用户提供服务的运营商的名称。对于提供虚拟运营业务的运营商而言,虽然使用了MNO运营商的网络,但是还是希望在其用户的终端上显示自己的名称。

方法1:SIM/USIM存储PLMNID(PLMN标识)和PLMN名称的映射列表,终端根据接入的PLMNID显示运营商名称。

方法2:在终端进行位置更新过程中,MSC Server使用移动性管理信息(MM Information)消息中的NITZ参数发送运营商的名称给终端。在MSC Server中可以配置不同的国际移动用户识别码(IMSI)号段发送不同的运营商名称信息,就可以实现使用同一张网络的不同运营商,发送自己的名称给各自的用户。

2.3 路由选择

当采用部分移动虚拟网络运营(Simple MVNO)或者完全移动虚拟网络运营(Full MVNO)组网方式时,MVNO运营商会使用MNO运营商的MSC Server/拜访位置寄存器(VLR)网元。共用的MSC Server/VLR网元如何把MVNO运营商的用户相关的消息发送给MVNO运营商的归属位置寄存器/网关移动交换中心/短消息服务中心/业务控制点(HLR/GMSC/SMSC/SCP)等网元是需要解决的关键技术之一。

MNO的MSC Server/VLR采用以下

方式来解决路由问题:

(1) MNO的MSC Server/VLR根据用户的IMSI,将位置更新请求等消息路由到不同运营商的HLR。

(2) MNO的MSC Server/VLR根据主叫MSISDN/IMSI信息,确定是否需要将呼叫等消息路由到MVNO的网关移动交换中心(GMSC)。

(3) MNO/MVNO GMSC根据被叫用户的MSISDN确定将路由请求等消息路由到MNO HLR或MVNO HLR。

3 分组交换/演进分组核心网关键技术

完全MVNO组网方式时,MNO的GGSN/MME如何把消息发给MVNO的GGSN/PGW;部分MVNO组网方式时,MNO的GGSN/PGW如何把消息转发给MVNO的计费系统,以及GGSN/PGW如何把消息转发给MVNO的应用服务器是PS域相关网元需要解决的关键技术问题。

3.1 SGSN支持多PLMN

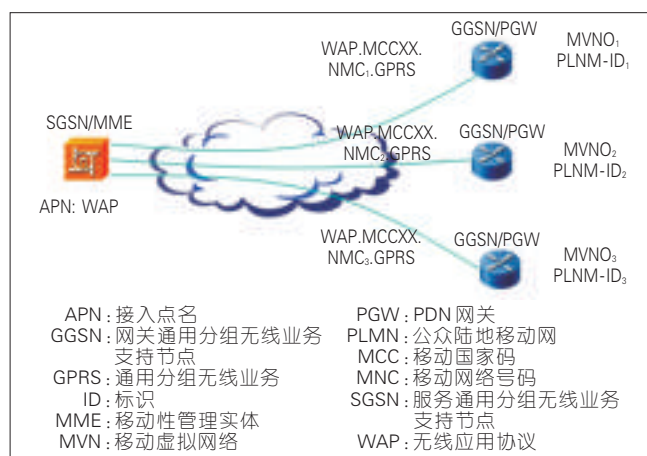
MNO的SGSN支持多PLMN,来解决SGSN/MME如何把消息发给MVNO的GGSN/PGW的问题。SGSN支持多PLMN原理如图1所示。SGSN/MME支持多PLMN,根据MNC来判断用户归属的网络。SGSN将不同的PLMN用户分别解析到不同MVNO的GGSN。

3.2 GGSN/PGW支持多PLMN

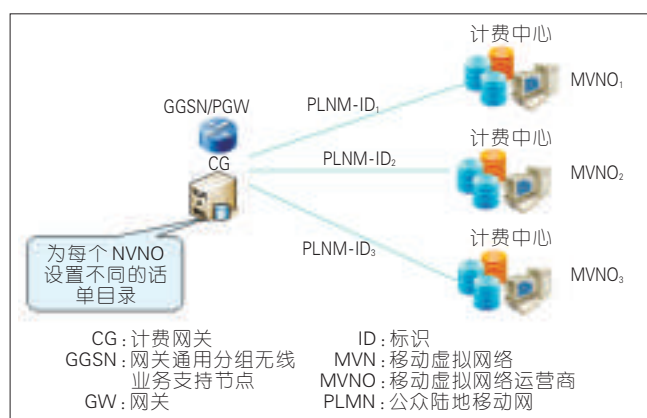
MNO的GGSN/PGW支持多PLMN,来解决GGSN/PGW针对不同MVNO的用户启用不同的业务控制策略以及分别计费等问题。GGSN支持多PLMN原理如图2所示。GGSN/PGW支持多PLMN根据MNC,基于不同的APN,不同的IMSI号段来判断用户归属的网络,并针对不同的PLMN用户分别计费。

3.3 GGSN支持VRF

MNO的GGSN支持VRF,针对不



◀ 图 1
SGSN 支持多 PLMN



◀ 图 2
GGSN 支持多 PLMN

同的 MVNO 和 MNO 启用不同的 VRF。GGSN 支持 VRF 原理如图 3 所示。

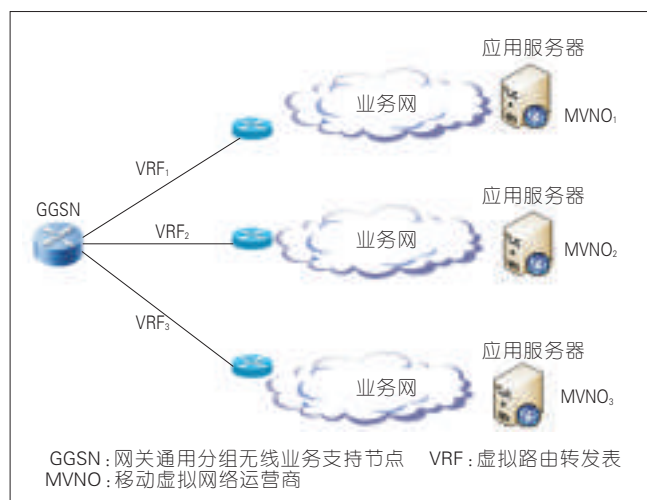
4 结束语

MVNO 是运营商在提供通信服务时候的一种新商业模式, MVNO 有利于节省运营商建网的成本。MVNO

的方案及其关键技术是 MVNO 和 MNO 都需要考虑的。MVNO 方案对用户而言是透明的, 实现了享受业务服务的无缝性。

致谢:

本论文受江苏省未来网络创新



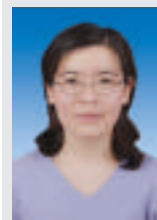
◀ 图 3
GGSN 支持 VRF

研究院未来网络前瞻性研究项目资助支持, 谨致谢意!

参考文献

- [1] 3GPP TS22.951. Service aspects and requirements for network sharing [S]. 2010
- [2] 3GPP TS24.008. Mobile radio interface Layer 3 specification; Core network protocols; Stage3 [S]. 2012
- [3] 3GPP TS25.413. UTRAN Iu interface Radio Access Network Application Part (RANAP) signaling [S]. 2012
- [4] 3GPP TS29.272. Evolved Packet System (EPS); Mobility Management Entity (MME) and Serving GPRS Support Node (SGSN) related interfaces based on Diameter protocol [S]. 2013
- [5] 3GPP TS22.042. Network Identity and TimeZone (NITZ); Service description; Stage 1 [S]. 2010
- [6] 3GPP TR23.038. Alphabets and language-specific information [S]. 2011
- [7] 3GPP TS29.018. General Packet Radio Service (GPRS); Serving GPRS Support Node (SGSN) – Visitors Location Register (VLR); Gs interface layer 3 specification [S]. 2013
- [8] 3GPP TS29.002 Mobile Application Part (MAP) specification [S]. 2013
- [9] 3GPP TS23.012. Location management procedures [S]. 2011
- [10] 3GPP TS23.018. Basic call handling; Technical realization [S]. 2011
- [11] 3GPP TS23.002. Network architecture [S]. 2011
- [12] 3GPP TS23.003. Numbering, addressing and identification [S]. 2011
- [13] 3GPP TS23.060. General Packet Radio Service (GPRS); Service description; Stage 2 [S]. 2011
- [14] 3GPP TS29.060. General Packet Radio Service (GPRS); GPRS Tunneling Protocol (GTP) across the Gn and Gp interface [S]. 2013
- [15] 3GPP TR21.905 Vocabulary for 3GPP Specifications [S]. 2010

作者简介



朱玉娜, 中兴通讯股份有限公司无线规划部高级工程师; 长期从事移动核心网的技术研究以及产品规划设计工作; 已发表学术论文 4 篇, 拥有专利技术 18 项。



王修中, 中兴通讯股份有限公司南京研究所工程师; 主要从事 IMS 核心网络方面的研发工作, 并负责相关标准和技术的跟踪与研究; 已完成授权标准专利 1 篇, 授权普通专利 9 篇。

基于移动终端显示移动转售企业标识的解决方案

The Solution for Display of Mobile Resale Business Logo on Mobile Terminals

沈彬 / SHEN Bin

李硕 / LI Shuo

孟然 / MENG Ran

(中国信息通信研究院, 北京 100191)
(China Academy of Information and
Communications Technology (CAICT),
Beijing 100191, China)

中国移动转售业务发展迅速,截至2014年12月底,中国已有42家企业从工信部获得了移动转售业务的牌照并开展转售业务。这些转售企业有的虽然只是刚刚起步,但也有一些企业已向用户提供服务。

根据中国《移动通信转售业务试点方案》中的要求,转售企业从基础电信企业购买移动通信服务,重新包装成自有品牌并销售给最终用户。当用户购买了转售企业销售的号卡装进终端后,终端上仍然显示“中国电信”、“中国移动”、“中国联通”等3家基础电信企业标识,而没有标识出提供业务的转售企业。且工信部目前对转售业务号码的分配原则是:监管部门为移动转售业务规划独立的号段,在这个号段下分配给基础运营企业子号码段,转售企业从基础运营商处获取号码。当前工信部已规划170、171号段用于转售业务,并将1700—1702分配给中国电信,1703、1705—1706分配给中国移动,1707—1709分配给中国联通,这3家基础电

收稿日期: 2015-06-29
网络出版时间: 2015-08-24

中图分类号: TN91 文献标志码: A 文章编号: 1009-6868 (2015) 05-0024-003

摘要: 提出了针对在不同场景下在移动终端上显示移动转售企业标识的解决方案: 在开机界面显示需要转售企业提出需求, 由终端厂商配合完成; 在开机问候语中显示需要用户识别 (SIM) 卡和终端的支持, 并在 SIM 卡中进行设置; 在终端状态栏显示需要 SIM 卡配置, 并建议出台相关标准; 在终端软件中显示需要有准确的数据来源, 能够实时更新和正确显示。

关键词: 移动终端; 移动转售; 标识; 用户识别卡

Abstract: In this paper, according to different scenarios, we propose the solutions for displaying mobile resale business logos on mobile terminals. If it is displayed in the startup interface, the resale business needs to cooperate with terminal manufactures. If it is displayed in the welcome greeting interface, the SIM card and terminal should be supported, and the SIM card needs to be set up. If it is displayed in the status bar of the terminal, the SIM card needs to be configured, and the relevant standards are recommended. If it is displayed in the software of the terminal, accurate data source is needed and should be updated on time.

Keywords: mobile terminals; mobile resale; LOGO; SIM card

信企业再将所获得的号码资源以1万号的颗粒度为协议分配给转售企业, 因此终端用户无法像以往一样通过移动用户号码的网号来区分出当前提供业务的转售企业。现在中国已规划并使用了35个移动通信网网号, 用于基础电信企业为终端用户提供移动通信服务, 如果再通过移动通信网网号来分辨服务提供商, 那么用户将需要记忆更多的网号与服务提供商之间的对应关系, 显然该方法不可取。所以当前通过移动终端来解决转售企业的标识问题是最好的途径, 转售企业也希望通过该方法来解决^[1-2]。

本文针对移动终端当前显示基础电信企业标识的方法进行研究, 提

出在移动终端上显示转售企业标识的解决方案, 为转售企业、移动终端厂家、制卡商和相关软件厂商等提供参考。

1 显示服务提供商的位置及现状

为了了解终端上都有哪些位置显示服务提供商标识, 我们分别对苹果、三星、华为、小米等主流品牌手机进行调查, 包括非定制机和定制机, 最终我们发现, 有如下4处位置可显示服务提供商。

1.1 开机界面

对于非定制机的开机界面显示的是各手机品牌 LOGO, 而不显示服

务提供商。对于定制机,则显示服务提供商所提供业务的品牌 LOGO 或相应的动画,如中国电信 3G 业务定制机的开机界面为天翼品牌的 LOGO 动画,中国移动 4G 业务定制机的开机界面为 4G 业务品牌 LOGO,中国联通 3G 业务定制机的开机界面为沃品牌的 LOGO 动画。该位置所显示的内容,是终端厂商根据服务提供商提出的需求在终端系统上设置完成。如果想要修改开机界面内容,需要专业人士修改终端系统内设置。

开机界面处显示服务提供商标识只能标志终端是由哪个服务提供商定制的,而不能根据用户所使用的用户识(SIM)卡的服务提供商来更换标识。例如目前大多数手机都提供多网络制式,即手机不但支持中国移动的 TD-LTE,还支持中国联通的 WCDMA,如果手机是由中国移动定制的,且中国移动要求手机厂商在手机的开机界面显示中国移动 4G 业务品牌 LOGO,而使用此部手机的用户在手机中插入的是中国联通提供的 SIM 卡,这样用户是可以正常通信,且能够正常使用手机中的全部功能,只是在开机界面处显示的是中国移动 4G 业务品牌 LOGO,而在其余地方显示中国联通等相关标识。

1.2 开机问候语

在终端开机后,部分手机会弹出服务提供商的问候语,如中国移动动感地带品牌的 SIM 卡插入手机后,在手机开机时,就会自动弹出“没错我就是 mzone 人”的开机问候语。该处所显示的内容,是通过服务提供商在 SIM 卡中写入数据时设置了相关应用,使得用户在开机时,自动向终端推送应用中所设置的带有服务提供商标识的文本信息。

开机问候所显示的服务提供商标识,主要是根据用户当前所使用的 SIM 卡的服务提供商来显示相应的标识,如果用户更换了其他服务提供商的 SIM 卡,则该位置也就会显示成其

他服务提供商的标识。

1.3 终端状态栏

在终端开机正常运行后,我们会看到终端状态栏的左上角显示网络信号的强弱以及服务提供商标识,且目前市场上大多数的终端在该位置处显示“中国电信”或“中国移动”或“中国联通”。该位置所显示的内容,是通过用户所使用的 SIM 卡中有关公众陆地移动网(PLMN)、公众陆地移动网络名(PNN)、运营商公众陆地移动网列表(OPL)以及服务提供商名(SPN)等文件内容设置,和当前网络情况来共同决定的。也有部分终端通过当前网络情况及终端中的文件配置来决定所显示的内容。

终端状态栏所显示的服务提供商标识,主要是根据用户当前所使用的 SIM 卡的服务提供商来显示相应的标识,当用户更换了其他服务提供商的 SIM 卡,则该位置也会随之显示成其他服务提供商的标识。例如用户的手机即支持中国移动的网络又支持中国联通的网络,如果用户在手机中插入的是中国移动的 SIM 卡,用户所使用的服务即为中国移动所提供,在中国移动的网络覆盖范围内,在此位置则显示“中国移动”。如果用户在手机中插入的是中国联通的 SIM 卡,用户所使用的服务即为中国联通所提供,在中国联通的网络覆盖范围内,在此位置则显示“中国联通”。

1.4 终端软件

用户终端自身提供的或者是用户自行安装的用于显示号码归属服务提供商的软件,也能够在使用软件时显示号码和服务提供商的对应关系,如来电时显示来电号码及其对应的服务提供商,或者在去电时显示本机的服务提供商等。此类软件目前由终端厂商、专用软件厂商提供,该位置所显示的内容,是终端厂商将号码与服务提供商的对应关系

文件内置在手机系统内,或者是软件厂商将此对应关系内置在软件里,只有终端预装该软件,才能在相应的位置显示号码对应的服务提供商。

终端软件中显示服务提供商标识,完全由软件中内置的号码和服务提供商的对应关系文件内容决定,与终端是否为服务提供商的定制机无关,与当前用户所使用的是哪个服务提供商提供的通信服务(即所插入的 SIM 卡)也无关。

2 显示移动转售企业标识解决方案的研究

根据前述终端能够显示服务提供商标识的位置以及当前显示的方法分析,我们对于终端显示转售企业标识问题提出 4 种解决方案。

2.1 在开机界面显示转售企业标识的解决方案

终端开机界面显示服务提供商标识均为定制机,且不随着 SIM 卡的更换而更换。转售企业想要在终端开机界面显示自有标识,则需要与目标终端厂商协商,包括想要在开机画面显示的内容和方式,由终端厂商按要求内置在终端系统中即可。

2.2 在开机问候语显示转售企业标识的解决方案

开机问候语中所显示的服务提供商标识是配置在 SIM 卡中的。因此,转售企业想要在终端开机问候语种显示自有标识,首先 SIM 卡应支持 SIM 应用工具箱(SAT),在写卡时,激活 SIM 卡业务列表中的第 29 号业务,即主动式 SIM。且终端根据硬件能力能够支持主动式 SIM,这样 SIM 就可以通过通用集成电路卡(UICC)主动要求终端执行某个操作。对于开机问候语, SIM 发送的命令为 DISPLAY TEXT。SIM 通过 DISPLAY TEXT 指令要求终端显示文本消息和图标,并可定义消息的优先级和文本串的格式。

开机问候语的过程如下:支持主

动式SIM的终端,在SIM初始化阶段通过终端发送命令TERMINAL PROFILE,支持SAT的SIM在成功执行指令后,将状态字SW1 SW2置为“91XX”来通知终端从SIM中取DISPLAY TEXT指令并执行,终端收到状态字“91XX”后,应向SIM发送FETCH指令来取要SIM要求执行的DISPLAY TEXT指令,SIM收到FETCH指令后将DISPLAY TEXT指令发送给终端,终端收到该指令后立即尝试执行,并通过TERMINAL RESPONSE将执行结果反馈给SIM,由SIM应用根据命令TERMINAL RESPONSE中包含的信息决定后续的操作。

转售企业只需将自有标识或者是需在终端显示的问候语设置在DISPLAY TEXT指令中即可。

2.3 在终端状态栏显示转售企业标识的解决方案

终端状态栏显示服务提供商标识是配置在SIM卡中或者是终端的配置文件中。因此,转售企业想要在终端状态栏显示自有标识,就需要在SIM卡的相应字段进行配置,或者是将自有标识设置在终端的配置文件中。但由于目前市场上终端型号较多,且终端厂商也较多,修改终端的配置文件中在短时间内未必能够解决,并且随着转售业务的发展,会有新的企业进入,也会有转售企业退出,转售企业列表将是随着变化而动态更新的。因此,转售企业想要在终端状态栏中显示自有标识,最有效的解决方案是在SIM卡中进行配置。

在写卡时,将业务提供商名称(EFSPN)设置为自有标识即可。

在以往2G/3G时,SIM未启用SPN,因此终端获取到的字符串为空,从而获取网络运营商的名称。导致终端状态栏不能正常显示转售企业标识。另外,当前虽然有部分转售企业已将自有标识写入到EFSPN字段中并且正确设置,但在终端上仍不能正常显示,这就是终端根据自身的

配置文件内容进行显示,而未读取SIM中的EFSPN字段。为避免出现这些问题,建议尽快出台相关标准,使得SIM卡的配置和终端厂商的读取方式都根据标准要求执行。

2.4 在终端软件中显示转售企业标识的解决方案

终端软件中显示服务提供商标识,完全由软件中内置的号码和服务提供商的对应关系文件内容决定。目前,大多数软件都是根据移动用户号码的网号,如13X、15X、17X、18X,来识别网络运营商,也就是在转售业务出现之前固有的3家服务提供商,即中国电信、中国移动、中国联通。而随着转售业务的开展,转售企业的进入,服务提供商已经远不只这3家,且根据工信部现有的转售业务用户号码的分配方式,所有转售企业均使用170号段,因此再根据网号来判断服务提供商是不可取的。

要想解决在终端软件中能够显示转售业务的用户号码显示正确的转售企业标识,重点需解决软件中内置的号码和服务提供商的对应关系文件内容的数据来源及数据内容问题。随着移动网络的发展,手机应用(APP)花样繁多,无论是企业还是有实力的开发者均可以开发相关的软件,而软件中的数据大多来自于网络或者是从网络上搜集号码信息的企业,极少数的企业是从运营商获得,数据来源问题也给开发者带来了很大的困扰。另外,用户号码的数量和状态是动态变化的,开展转售业务的转售企业也是动态变化的,号码和服务提供商的对应关系是动态更新的,所有需要内置号码和服务提供商对应关系的软件中的数据就需要动态更新。要想解决这些问题,建议建立一个准确的数据信息平台,凡是在软件中需内置号码和服务提供商对应关系的软件开发者都到统一的平台上来取数据,并且随着平台数据的更新而更新,使其正确显示在终端上。

这样不但能解决在终端软件中正确显示转售企业作为服务提供商标识的问题,还能解决软件开发企业或开发者目前存在的问题。

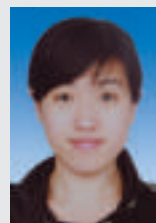
3 结束语

综上所述,要想在移动终端上正确显示转售企业标识,不但是转售企业自己需要多方沟通努力改进的过程,还需要终端厂商的配合,标准的制订以及码号数据信息平台的建设等。只有这样,才能保证在移动终端中的各个位置均能按照转售企业的需求显示其标识,彻底的解决转售企业的问题,为转售业务的顺利开展扫清障碍。

参考文献

- [1] 3GPP TS 51.011 V5.0.0 (2001-12). 3rd Generation Partnership Project; Technical Specification Group Terminals; Specification of the Subscriber Identity Module - Mobile Equipment (SIM-ME) interface (Release 5) [S]. 3GPP, 2001
- [2] YD/T 1762.3-2008. TD-SCDMA/WCDMA数字蜂窝移动通信网通用用户识别模块(USIM)与终端(ME)间CU接口技术要求第3部分:通用用户识别模块应用工具箱(USAT)特性[S]. 工信部, 2008

作者简介



沈彬, 中国信息通信研究院技术与标准研究所工程师; 主要从事电信网码号资源、号码携带、移动转售等方面的政策、技术研究。



李硕, 中国信息通信研究院技术与标准研究所助理工程师; 主要从事移动转售、虚拟运营、码号资源相关领域的政策、标准、技术研究。



孟然, 中国信息通信研究院技术与标准研究所工程师; 主要从事电信网码号资源与移动转售领域相关政策、技术研究。

大数据环境下的可靠存储技术思考

Reliable Storage Technologies for Big Data

中图分类号: TN929.1 文献标志码: A 文章编号: 1009-6868 (2015) 05-0027-005

摘要: 针对分布式容错技术的研究,提出了两点关键要求:降低冗余开销、提高节点修复效率。分析目前主流的容错策略:复制、纠删码、再生码、基于局部可修复码,并认为这些容错策略存在不同程度的缺陷,因此设计出容错能力、计算效率及存储利用率更高的容错策略,仍是未来很长一段时间内值得深入研究的问题。

关键词: 大数据; 可靠性; 分布式存储; 容错技术

Abstract: Two key requirements of fault tolerance technology are proposed in this paper: minimal storage overhead and maximum node recovery performance. Four main strategies for fault tolerance are analyzed: replication, erasure codes, regenerating codes and locally repairable codes. It is considered that these fault tolerance strategies have different defects. Designing a fault tolerance strategy with higher fault tolerance, better computational efficiency and memory utilization will still be a problem needs to be solved in the future.

Key words: big data; reliability; distributed storage; fault tolerance technology

李挥/LI Hui

张宇蒙/ZHANG Yumeng

陈俊/CHEN Jun

(北京大学深圳研究生院, 广东 深圳 518055)

(Shenzhen Graduated School, Peking University, Shenzhen 518055, China)

- 纠删码的低效修复已经成为限制其广泛应用的瓶颈所在
- 再生码和局部可修复码通过适量增加存储开销,有效降低了纠删码的修复带宽
- 再生码编码策略的未来研究方向,应结合安全问题、网络拓扑和磁盘 I/O 复杂度进行设计

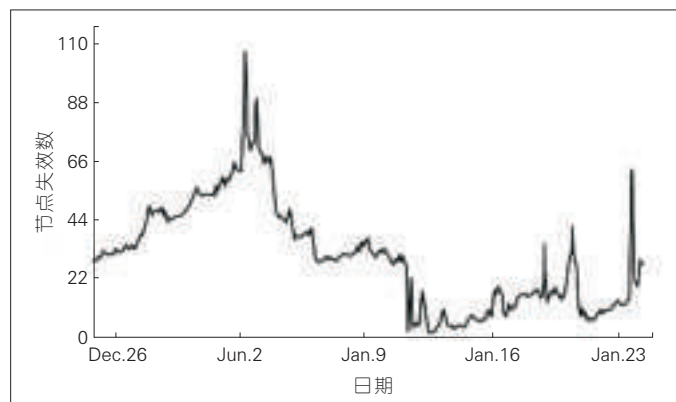
随着经济全球化的发展和科技改革的推进,网络覆盖面积不断加大,信息交互随之增强,全球数据正在以爆炸式的速度增长。国际数据公司 (IDC) 报告指出,从 2010—2020 年全球数据量将有 50 倍的增长,预测达到 40 ZB 数量级^[1]。同时海量数据对存储系统提出了巨大的挑战,根据统计,数据存储的需求每年的增速在 50% ~ 62% 之间。大规模分布式存储系统以其海量存储能力、高吞吐量、高可用性和低成本的突出优势成为存储海量数据的有效系统并被广泛使用。当前最主流的分布式系统是开源的 Hadoop 分布式文件系统 (HDFS)^[2],作为 GFS^[3] 的一个开源实现,它被应用于众多大型企业,如 Yahoo、Amazon、Facebook、eBay 等。

随着分布式存储系统的规模越来越大,为节省成本,存储节点大多采用廉价、可靠性差的设备,这直接导致节点故障越来越频繁。图 1 给出了 Facebook 部署的 Hadoop 集群的日节点失效数。集群共 3 000 个节点,涉及 45 PB 数据,平均每天有 22 个节点失效,最高的日节点失效超过 100 个^[4]。如何有效保障数据可靠性

成为了当前分布式存储系统首要关注的问题。

为了提供可靠的存储服务,分布式存储系统通过引入冗余信息来提高系统的容错能力。这种冗余存储的方式能够使系统容忍一定数量的节点故障^[5-6],同时系统还需要一个良好的节点修复机制,在发生故障时能快速有效地修复失效数据,维持系统

图 1
Facebook 日节点失效数



收稿日期: 2015-08-16
网络出版时间: 2015-09-20

冗余度。

1 基于复制的容错技术

复制策略是引入冗余最简单的方法,其基本思想是为系统中的每一个数据对象都建立若干个相同的副本,并把这些副本分散存储在不同的节点上,当遇到某个数据损坏或失效而无法正常使用时,可通过访问最近的存储节点来获取与原件完全一致的数据备份,这样只要数据对象还有一个存活副本,分布式存储系统就可以一直正常运行。修复过程也十分简单高效,只要向所有存储副本的节点中最近的节点发出请求、下载并重新存储,即可恢复系统冗余度。复制策略存储方式简单,易于实现,故障修复容易,并且便于扩展。此外,存储的多个副本也可以均摊读文件时的负载,如通过为热点文件配置更高的副本数来支持高效的并发读操作。

但是在节点数量庞大,存储结构复杂的大规模分布式系统中,要实现快速高效的容错技术,必须解决3个问题:副本数量的设置、副本的放置方式和副本的修复策略。

1.1 副本数量设置

设置副本数量一般有两种方式:一是静态设置,主流的分布式文件系统如HDFS^[2]和GFS^[3]都是采用3副本固定机制,这种方法操作简单,但灵活性差;二是动态设置副本数量,亚马逊分布式存储系统S3提供用户可以自行设定副本数的功能。另外,文献[7]提出一种动态的容错机制,系统根据数据的访问频率、出错概率、网络状况以及存储时间等动态因素决定副本数,同时动态地删除或添加副本,这种动态机制能大大增加存储空间利用率、提高数据的获取性能,但动态决策方式会加大系统的处理开销。

1.2 副本放置策略

副本的放置策略不但影响分布

式存储系统的容错性能,还关系到副本的存储效率和访问效率。HDFS采用的3副本放置策略,如图2所示^[2]。

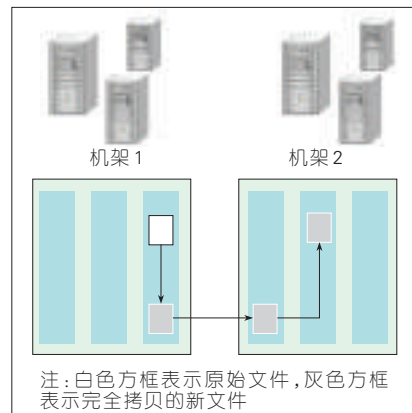
3副本放置策略为:本地放一份,同机架内其他任一节点放一份,不同机架的任一节点放一份。同机架内存放两个副本,可减少机架间的数据传输,方便本地节点对于数据需求时的读取。若本地数据损坏,节点可以从同一机架内的相邻节点获取数据,读取速率快。而数据块存放在两个不同的机架中能避免机架故障导致的数据不可用。同时,为了降低整体的带宽消耗和读取延时,HDFS会尽量让读取程序读取离它最近的副本。如果在读取程序的同一个机架上有一个副本,那么就读取该副本。如果一个HDFS集群跨越多个数据中心,那么客户端也将首先读本地数据中心的副本。

1.3 副本修复策略

容错技术的修复过程事实上就是恢复系统的冗余度,保证其在一定的可接受范围内。实际的存储系统采用的修复策略有两种:一种是“主动”修复策略^[8],一旦检测到一个副本失效立刻创建一个新副本;另一种是基于阈值的“惰性”修复策略,这种策略只有当备份数量小于某个阈值才进行修复,如Total Recall^[9]。根据资源的访问频率,可以分为热门资源和冷门资源,热门资源一般采用主动修复,而访问量小的冷门资源则可以采用惰性修复策略,减少修复临时失效等不必要的开销。

2 基于纠删码的容错技术

纠删码起源于通信传输领域,由于其数学特性,被逐渐应用于大规模存储系统中,特别是分布式存储环境,实现数据的冗余保护。相较于复制策略,纠删码技术在相同可靠性条件下可以最小化冗余存储,学术界和工业界已将纠删码广泛应用于分布式文件系统。例如卡耐基梅隆大学



▲图2 HDFS的3副本放置策略

研究的DiskReduce^[10]、Facebook的HDFS-RAID^[11]、谷歌的Colossus^[12]、微软的Azure^[13]存储系统均采用了纠删码并实现了更经济的可靠性。

2.1 纠删码基本原理

纠删码的基本原理如图3所示,存储原始文件 O ,首先将其切分成 k 个数据块,记为 O_1, O_2, \dots, O_k ,然后编码生成 n 个编码块,记为 $B_1, B_2, \dots, B_n, n > k$,最后将这 n 个编码块按照一定的放置规则分别存储在不同的节点上。编码过程中生成了冗余数据,当系统中有存储节点失效时,只要留下足够的编码块就可以利用这些剩余的编码块恢复丢失的数据,维持系统的冗余度。若 n 个编码块中任意 k 个块即可重构原始文件 O ,则这种纠删码满足最大距离可分特性(MDS)^[14],在可靠性和冗余的权衡上达到最优,最常用的编码方法是RS码^[15]。

2.2 基于纠删码的分布式存储模型

在分布式存储系统中,数据分布在多个相互关联的存储节点上,通常情况下,映射生成的编码块需要存储在不同的节点上。图4给出了一种基于纠删码的分布式存储模型^[16],假设系统含有 n 个存储节点,其中 k 个是数据节点, m 个是编码节点,即满足 $n = k + m$ 。 k 个数据节点存储原始数据块,标记为 D_0, D_1, \dots, D_{k-1} ; m 个

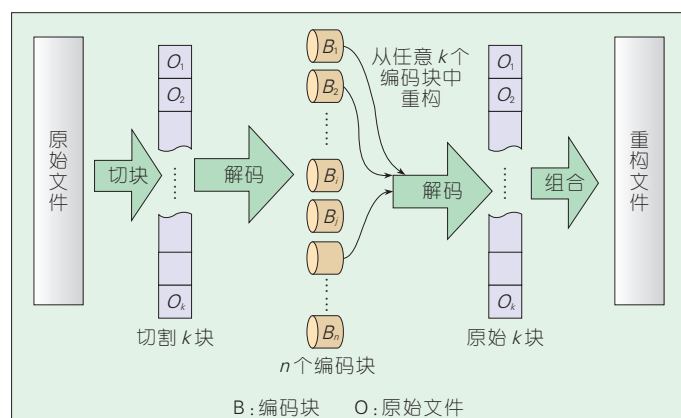


图3
纠删码原理示意

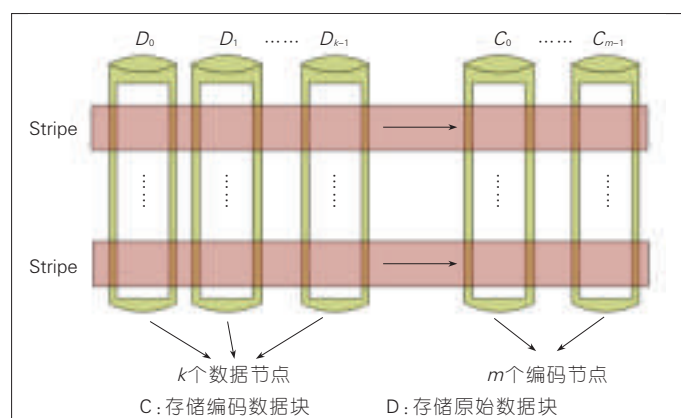


图4
基于纠删码的
分布式存储模型

编码节点存储编码数据块,标记为 C_0, C_1, \dots, C_{m-1} 。纠删码算法需要将原始文件切割成 k 等份后依次存储在 k 个数据节点中,并将编码生成的 m 份放入 m 个编码节点。当存储大文件时,需要对原始文件进行二次切割,即每次从文件中读取指定大小的数据量进行编码,我们将一次编码过程中涉及的原始数据和编码数据称为一个 stripe^[16]。一个 stripe 独立地构成一个编码的信息集合,不同 stripe 之间相互无关。但是,逻辑上的 stripe 与实际物理节点的对应关系并不是恒定不变的,可以通过 stripe 的轮转实现数据存储负载均衡。

与复制策略相比,纠删码策略可以有效地降低维持可靠性所需的存储开销,提供令人满意的存储效率^[5]。

2.3 纠删码技术的缺陷

然而,基于纠删码的容错技术未能在实际的大规模分布式存储系统

中真正应用,除了其结构较复制策略复杂外,纠删码本身在数据恢复时存在致命的缺陷。在基于纠删码的分布式存储系统中,当一个节点失效时,为维持系统冗余度,新节点需要首先从 k 个节点中下载全部数据恢复出原始文件,再重新编码生成失效的数据,这个过程中传输的数据量是失效数据的 k 倍。当节点在网络中分布较分散时,节点的修复需要消耗大量的网络带宽。这一缺陷在普通分布式系统中已有制约,在大数据环境下,数据量和存储节点在成倍甚至几何级增长时更为明显。同时,需要的下载量太大势必会导致节点修复过程变慢,对于不断发生故障的分布式存储系统来说,节点的修复速率直接影响到系统可靠性。如果修复速率过慢,甚至赶不上节点发生故障的速度,那么系统将无法维持其可靠性。据 Facebook 在 HotStorage'13 上发布的论文指出,纠删码的低效修复已

经成为限制其广泛应用的瓶颈所在^[4]。

针对纠删码的修复问题,Rodrigues 提出了一种混合策略^[5]:采用纠删码的同时维护一个副本,从而有效减少修复带宽。然而,这种混合策略节省带宽有限,但存储开销大,同时使得系统设计复杂化。Dimakis 创造性地将网络编码应用于分布式存储,提出再生码的概念^[17],显著降低了修复带宽。

3 基于再生码的容错技术

3.1 再生码的基本原理

再生码的描述如下:将原始文件编码后存储到 n 个节点中,每个节点存储大小为 α 。当一个节点失效时,新节点连接剩余 $n-1$ 个节点中的 d 个节点 ($k \leq d \leq n-1$),从每个节点下载大小为 β ($\beta \leq \alpha$) 的数据进行修复,即修复带宽为 $\gamma = d \times \beta$ 。再生码的参数集可表示为 $\{n, k, d, \alpha, \beta, B\}$,其平均修复带宽 γ 小于文件大小 B 。再生码的编码、再生及重构过程如图5所示。

随着每个节点的存储量的提高,节点修复时需要下载的数据量将降低,通过在信息流图上求最小割界的方法,给出了节点修复带宽消耗的下界曲线,而再生码正是在存储开销 α 和修复带宽 γ 的最优曲线上。如图6所示,最优曲线上存在两个极值点,分别代表最优存储效应和最小修复带宽效应,达到这两个极值点的编码称为最小存储再生码(MSR)和最小带宽再生码(MBR),已有一些明确的编码实现^[18]。理论上,当 $d=n-1$ 时,再生码的修复带宽达到最小值。

3.2 再生码技术的瓶颈及前景

虽然理论上再生码可以达到最优的存储开销和修复带宽,但由于它依赖于复杂的参数和晦涩难懂的数学理论,其实现方式非常复杂。现有的再生码大多在有限域 $GF(2^n)$ 上进行域元素的多项式运算^[18]。计算机处理中,加法较为简单,但乘法和除法

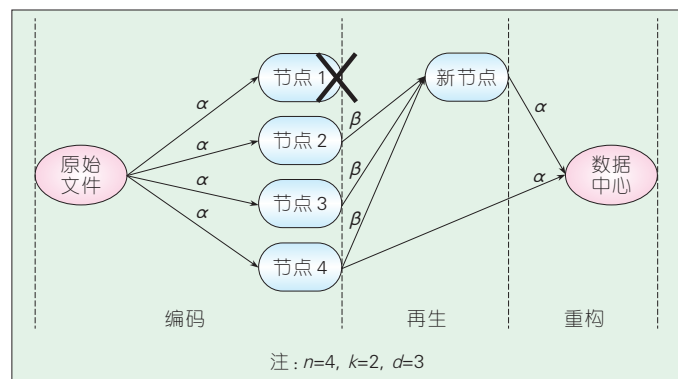


图5
再生码编码、再生
及重构过程

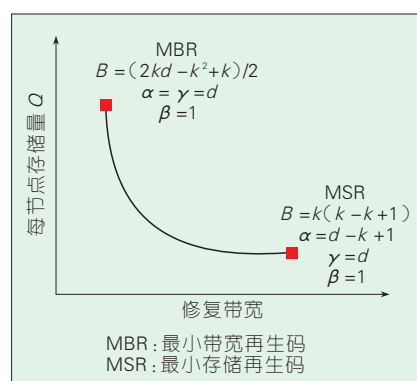


图6 “存储-带宽”的最优曲线

却非常复杂,甚至需要借助离散对数运算和查表才能实现。这使得再生码的编解码计算开销大,无法适应存储系统对计算效率的要求。很多研究都表明,设计一种结构简单、计算复杂度低的策略至关重要。文献[19]中分析了3种再生码:随机线性网络码(RL)^[20]、精确线性码(EL)^[21]和生成矩阵码(PM)^[22]。其中,PM码利用一种紧凑的表示方式和高效的编解码算法大大提高了编解码速率,然而与纠删码相比,PM码仍需要更长的计算时间。

再生码作为对纠删码的改进,具有很好的理论支撑。但目前提出的大多数再生码、编解码复杂度较高且码率较低。如何提出码率较高并且复杂度低的编码策略就很有意义。深圳市融合网络技术实验室在该领域进行了深入研究,并取得了一定的研究成果:1)提出BASIC^[23]编码框架,利用一种新颖的卷积形式来表示编码运算过程,可以将有限域运算转化

为GF(2)内简单的移位和异或操作;2)提出一种改进的Zig-Zag编码^[24],采用移位和异或的Zig-Zag解码算法,避免解码时所需要的复杂计算,达到了最低的编解码复杂度。这些编码都可以应用在再生码的构造上,以更好地实现码率较高并且复杂度低的编码。

对于再生码编码策略的未来研究方向,应结合安全问题、网络拓扑和磁盘输入输出(I/O)复杂度进行设计,从而使再生码更为实用。

4 基于局部可修复码的容错技术

除再生码外,局部可修复码技术(LRC)^[25]可以通过增加本地数据实现修复带宽的降低。文献[25]给出了修复局部性 r 、编码距离 d 、每个节点的

存储大小 α 以及存储编码长度 n 之间的权衡。Facebook在HDFS中实现了LRC技术^[4],微软也在Azure上添加了LRC技术^[26]。

文献[4]给出了LRC技术的一种实现:如图7,原始文件被等分成10个数据块,通过RS编码生成4个冗余块,图中显示为4个绿色方框。为降低修复带宽,在RS码基础上进行二次编码产生3个额外的冗余块,标记为 S_1, S_2 和 S_3 ,图中显示为3个橙色的方框。 S_1 是由前5个数据块编码产生, S_2 是由后5个数据块编码产生,这两个由局部原始数据块编码产生的冗余块称为本地校验块。而 S_3 则是由4个冗余校验块编码产生,称为隐式校验块。实际存储中,我们将10个原始数据块标记为 c_1, c_2, \dots, c_{10} ,将7个冗余块标记为 c'_1, c'_2, \dots, c'_7 ,存放在7个不同的节点。当1个数据块丢失时,只需要1个额外的冗余块和4个数据块即可修复失效数据,与传统纠删码相比,修复带宽降低了大概一半。

从图7可以看出LRC技术以额外的14%存储开销为代价,降低RS码的修复带宽。但其编码方式仍是RS码,因此编码效率没有提高。另外,LRC编码不满足MDS特性,系统还需要增加额外信息标示二次编码数

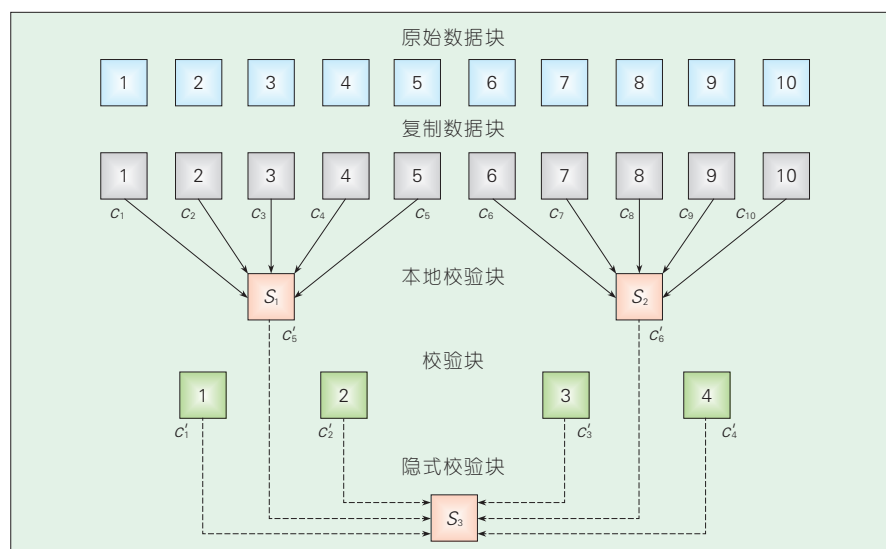


图7 基于LRC容错技术的实现过程

据。当修复一个节点故障时, LRC 具有很好的修复局部性, 但修复两个或两个以上的节点故障时就需要连接 k 个节点, 修复带宽与纠删码相同, 仍是失效数据的 k 倍。随着存储系统规模变得越来越大, 出现两个或者多个故障的几率也随之增大。

除此之外, 针对大数据存储系统中的容错修复问题, 我们不断对存储编码的构造方式进行改进^[27-28], 以获得更低的冗余开销和更高效的修复性能。

5 结束语

介绍了大数据环境下的可靠存储技术, 并针对分布式存储系统, 介绍多种容错策略及相关技术。基于复制的容错技术冗余度大, 性能提升艰难, 很多研究者将目光聚集于基于纠删码的容错技术。而再生码和局部可修复码通过适量增加存储开销, 有效降低了纠删码的修复带宽。

这些容错策略在容错能力、计算效率、存储利用率等方面都存在不同程度的缺陷, 如何平衡这些影响系统可靠性的因素, 设计出容错能力、计算效率及存储利用率更高的容错策略, 仍是未来很长一段时间内值得不断深入研究的问题。

致谢:

感谢北京大学先进网络技术实验室博士研究生侯韩旭对于研究的理论支持。

参考文献

- [1] 潘凯西. 大数据支持企业转型与创新[R]. 北京: 大数据重构企业智慧发布会, 2014
- [2] Shvachko K, Kuang H, Radia S, et al. The hadoop distributed file system[C]//Mass Storage Systems and Technologies (MSST), 2010 IEEE 26th Symposium on. IEEE, May 3-7, 2010, Lake Tahoe, Nevada, USA, 2010: 1-10
- [3] Ghemawat S, Gobioff H, and Leung S T. The Google file system [J]. ACM SIGOPS operating systems review, 2003, 37(5): 29-43
- [4] Sathiamoorthy M, Asteris M, Papailiopoulos D, et al. Xoring elephants: Novel erasure codes for big data[J]. VLDB Endowment, 2013, 6(5): 325-336
- [5] Weatherspoon H, and Kubiatowicz J D.

- Erasure coding vs. replication: A quantitative comparison [M]. Peer-to-Peer Systems. Germany: Springer Berlin Heidelberg, 2002: 328-337
- [6] Rodrigues R, Liskov B. High availability in DHTs: Erasure coding vs. replication [M]. Peer-to-Peer Systems IV. Germany: Springer Berlin Heidelberg, 2005, pp. 226-239
- [7] Wu L, et al. A Dynamic Data Fault-Tolerance Mechanism for Cloud Storage [C]//Emerging Intelligent Data and Web Technologies (EIDWT), 2013 Fourth International Conference on, Xi'an, China, 2013: 95-99
- [8] Sit E, et al. Proactive Replication for Data Durability [C]//5th International workshop on Peer-To-Peer Systems (IPTPS 2006), Santa Barbara, USA, 2006
- [9] Bhagwan R, Tati K, Cheng Y, et al. Total Recall: System Support for Automated Availability Management [C]//1st Symposium on Networked Systems Design and Implementation (NSDI 2004), San Francisco, USA, 2004: 25-25
- [10] Fan B, Tantisiriroj W, Xiao L, et al. DiskReduce: RAID for data-intensive scalable computing [C]//Proceedings of the 4th Annual Workshop on Petascale Data Storage. ACM, 2009: 6-10
- [11] Saving capacity with HDFS RAID [EB/OL]. [2014-06-05]. <https://code.facebook.com/posts/536638663113101/saving-capacity-with-hdfs-raid/>.
- [12] Colossus [EB/OL]. <http://www.quora.com/Colossus-Google-GFS2>
- [13] Calder B, Wang J, Ogus A, et al. Windows Azure Storage: a highly available cloud storage service with strong consistency [C]//Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles, Cascais, Portugal, 2011: 143-157
- [14] Blaum M, et al. On lowest density MDS codes [J]. Information Theory, IEEE Transactions on, 1999, 45(1): 46-59
- [15] Stephen B W, and Vijay K B. Reed-Solomon codes and their applications [M]. USA: John Wiley & Sons, 1999
- [16] Xianghong L, and Jiwu S. Summary of research for erasure code in storage system [J]. Journal of Computer Research and Development, 2012, 49(1): 1-11
- [17] Dimakis A G, Godfrey P B, Wu Y, et al. Network coding for distributed storage systems [J]. Information Theory, IEEE Transactions on, 2010, 56(9): 4539-4551
- [18] Dimakis A G, Ramchandran K, Wu Y, et al. A survey on network codes for distributed storage [J]. Proceedings of the IEEE, 2011, 99(3): 476-489
- [19] Jiekak S, Kermarrec A M, Le Scouarnec N, et al. Regenerating codes: A system perspective [J]. ACM SIGOPS Operating Systems Review, 2013, 47(2): 23-32
- [20] Ho T, Médard M, Koetter R, et al. A random linear network coding approach to multicast [J]. Information Theory, IEEE Transactions on, 2006, 52(10): 4413-4430
- [21] Suh C, Ramchandran K. Exact-repair MDS code construction using interference alignment [J]. Information Theory, IEEE Transactions on, 2011, 57(3): 1425-1442. doi: DOI: 10.1109/TIT.2011.2105003
- [22] Rashmi K V, Shah N B, Kumar P V. Optimal

- exact-regenerating codes for distributed storage at the MSR and MBR points via a product-matrix construction [J]. Information Theory, IEEE Transactions on, 2011, 57(8): 5227-5239
- [23] Hou H, Shum K W, Chen M, et al. BASIC regenerating code: Binary addition and shift for exact repair[C]//Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on. IEEE, Istanbul, Turkey, 2013: 1621-1625
- [24] Chen J, Li H, Hou H, et al. A new Zigzag MDS code with optimal encoding and efficient decoding[C]//Big Data (Big Data), 2014 IEEE International Conference on. IEEE, Washington DC, USA, 2014: 1-6
- [25] Papailiopoulos D S, Dimakis A G. Locally repairable codes [C]//information theory systems, proceedings (ISIT), 2012 IEEE international symposium on. IEEE, Cambridge, USA, 2012: 2771-2775
- [26] Huang C, Simitci H, Xu Y, et al. Erasure Coding in Windows Azure Storage[C]//USENIX Annual Technical Conference, BOSTON, MA, 2012: 15-26
- [27] Zhu B, Shum K, Li H. Heterogeneity-Aware Codes with Uncoded Repair for Distributed Storage Systems[J]. Communications Letters, IEEE, 2015, 19(6): 901-904. doi: 10.1109/LCOMM.2015.2415495
- [28] Zhu B, Shum K, Li H, et al. General fractional repetition codes for distributed storage systems [J]. Communications Letters, IEEE, 2015, 18(4): 660-663. doi: 10.1109/LCOMM.2014.030114.132694

作者简介



中SCI和EI收录45篇。

李挥, 北京大学教授、博士生导师, 广东省粤教云计划总体组专家, 广东省计算机学会大数据专委会副主任委员, 广东省云计算行业协会副会长等; 研究领域为未来网络大数据、分布式存储系统及网络编码理论; 近5年在国内外核心期刊和会议上发表论文 50 多篇, 其中



张宇蒙, 北京大学先进网络技术实验室在读硕士研究生; 研究领域为分布式存储系统及网络编码理论。



陈俊, 北京大学先进网络技术实验室在读硕士研究生; 研究领域为分布式存储系统及网络编码理论。

半监督多视图学习在大数据分析中的应用探讨

Semi-Supervised Multi-View Learning in Big Data

中图分类号: TN929.1 文献标志码: A 文章编号: 1009-6868 (2015) 05-0032-003

摘要: 半监督多视图学习是机器学习领域一种极具潜力的大数据处理和分析方法,该方法能有效处理异构和半监督数据,并能方便地在线化和并行化,适合处理海量数据。该方法在大数据时代的应用前景值得研究人员和业界关注。指出未来需要通过引入其他领域新的研究技术和成果,不断丰富和完善半监督多视图学习的理论体系和算法设计,并在实验和实践中不断检验和探索。

关键词: 半监督; 多视图; 大数据; 并行化

Abstract: This paper introduces a promising machine-learning paradigm called semi-supervised multi-view learning. With this paradigm, information is extracted from heterogeneous and semi-supervised data sets. Lately, multi-view learning has been scaled up online and through parallelization to deal with emerging big data challenges. Due to its successful application in many research domains and the fact that it has been explored and used by leading companies, multi-view learning may have a future in the big-data era as a major data analytic technique. New research techniques should be introduced into this area to improve the theoretical system and algorithm design of semi-supervised multi-view learning.

Key words: semi-supervised; multi-view; big data; parallelization

蓝超/LAN Chao¹
饶泓/RAO Hong²
浣军/HUAN Jun¹

(1. 堪萨斯大学工程学院, 堪萨斯 66045;
2. 南昌大学信息工程学院, 江西 南昌 330031)

(1. College of Engineering, University of Kansas, Kansas 66045, USA;
2. College of Information Engineering, Nanchang University, Nanchang 330031, China)

- 如何有效利用无监督数据, 是半监督学习研究的核心问题
- 提高无监督数据的样本复杂度的效率是值得关注的问题
- 多视图学习的样本复杂度分析主要集中于主动学习

机器学习已成为产业界大数据分析的主流工具。在2015年北京全球软件开发大会上, 基于大数据的机器学习和数据挖掘专题讨论得到了业界广泛参与, 如百度、搜狗、阿里巴巴、京东、美团、猿题库等知名企业参与了该专题的讨论。当前, 机器学习在产业界的推广正处于白热化阶段。与此同时, 如何结合大数据的特点, 选择恰当的学习方法仍是值得探讨的问题。

从数据角度而言, 大数据之“大”并不仅限于容量, 也指数据类型多和收发速度快, 三者统称为大数据的

“3V”特征^[1]。因此, 若想迅速有效地处理和分析大数据, 不仅需要从数据库和计算机体系结构等入手(如IBM的Hadoop, 微软的大数据生态系统), 更需选择合适的数据分析方法, 才能充分挖掘海量数据中潜在的信息和价值。

除了3V, 大数据普遍存在却鲜有讨论的另一特征是半监督性。在传统机器学习中, 用于训练分类系统的数据大多是有监督的, 即数据的实际分类已知。然而, 许多应用问题如自然语言处理, 大部分数据都不知道实际分类, 这类数据称为无监督数据。如何有效利用无监督数据, 使其帮助有监督数据一起训练分类系统, 是半

监督学习^[2]研究的核心问题。随着大数据时代的到来, 大量无监督数据将被采集, 导致越来越多应用问题演变成大规模的半监督学习问题。

大规模半监督学习在学术界的探讨可追溯到2005年Delalleau提出的高效无参归纳法^[3]。在该论文中, 作者提出将用于构造核矩阵的(无监督)数据集进行下采样, 减小核矩阵大小, 从而提高算法的存储和计算效率。在此思想基础上衍生出来的工作有很多, 其中以大规模化基于核或基于图的半监督学习方法为主。除了数据下采样, 也有研究员通过构造混合分布模型以减小参数估计的计算复杂度。此外, 半监督学习的大规

收稿日期: 2015-08-27
网络出版时间: 2015-09-23

模化也可通过提高优化算法效率或并行程度来实现。

尽管有着近十年的研究历史,目前大部分的大规模半监督学习方法仍具有局限性。一方面,它们大都基于传统的单视图机器学习方法,忽略了大数据的异构性,容易导致信息丢失;另一方面,规模化算法的策略较为传统,如下采样或加速优化,无法有效满足大数据带来的新的挑战,如数据的在线化和分布化等。

文章结合大数据的特点介绍机器学习领域的一种新兴的半监督学习方法——多视图学习。它不仅在许多研究领域已得到成功应用,更因具备有效利用无监督异构数据的能力和迅速发展的大规模化算法,有望成为大数据时代最具潜力的数据分析工具之一。

1 多视图学习优点多

1.1 多视图学习及其优点

多视图学习^[4]是指专门针对多视图数据而进行建模和学习。其中,多视图数据是指由多组(往往具有不同意义的)特征进行描述的数据,而每一组特征称为一个视图。多视图学习的主要思想是基于无监督数据的视图一致性,即分类器在同一无监督数据不同视图下的分类结果应基本一致。将此约束加入学习法则,多视图学习便能巧妙利用无监督数据帮助分类器的训练。

多视图学习的一个显著优点是缓解过学习问题,即由于模型过于复杂而将数据噪声也学入分类法则的现象。传统的机器学习方法大多将数据的所有视图堆砌成一个高维的单视图数据,进行建模和学习。此时,如果视图间存在冗余信息,为高维数据所建的模型将比实际需求更为复杂,容易导致过学习。而多视图学习则为每个视图分别建模,有效降低了模型复杂度。

多视图学习的另一优点是提升

数据的总体分类能力。当数据特征所蕴含的分类信息总体较弱时,可将特征集拆分成多个视图进行多视图学习^[5]。通过各视图的弱分类器协同训练,达到各视图“单独学习弱,集成学习强”的目的。

另外,多视图学习还能有效处理异构数据^[6]。大数据时代,数据的异构性越来越强。比如,客服中心为了提高工作效率,需要根据来电客户的个人信息和语音信息对其来电目的进行快速预测。这里,用户的个人信息是静态文本数据,而语音信息则是动态时序数据,两者不仅数据类型不同,也往往服从不同的后验(预测)概率分布。强行统一或堆砌两类特征不仅给建模带来困难,也容易导致信息丢失,失去大数据分析的优势。多视图学习则允许各视图分别选择合适的分类器,再通过协同训练提高视图整体的分类能力。

1.2 多视图学习的理论研究

多视图学习的理论分析主要基于其分类器的泛化误差。一个分类器的泛化误差是其在指定数据分布下,错分数据的概率。许多理论研究证明:多视图中的协同学习方法在满足条件时,分类器的泛化误差将以极大的概率收敛到极小的范围之内。

多视图学习的早期理论分析基于两个假设^[7]:视图充分性,即每个视图能分别提供充分的(但不必很强)分类信息。该假设在大数据中较容易被满足;各视图间条件独立,即给定数据分类,其各视图间统计独立。

多视图学习另一个理论研究是其样本复杂度。样本复杂度指通过多少数据的训练就能保证分类器收敛到预期的泛化误差。目前,多视图学习的样本复杂度分析主要集中于主动学习,即选取哪些无监督数据进行人工分类,才能使分类器的训练最快收敛。在此问题中,有监督数据的样本复杂度被证明与泛化误差的倒数成 \log 比^[8],而无监督数据样本复杂

度则与泛化误差的倒数成正比^[9]。在大数据时代,无监督数据的容量常常远大于有监督数据,是计算机的存储和计算的主要负担。因此,提高无监督数据的样本复杂度的效率是值得关注的问题。

1.3 多视图学习的常见算法及

大规模化算法

大部分多视图算法为每个视图分别建立分类器,基于协同训练方式的不同,有两种主要的多视图算法:一是基于迭代的协同训练算法^[7],另一个是基于协同正则化的算法^[10]。基于迭代的协同训练算法的每个回合由一个视图的分类器标注一部分无监督数据的类别,并将它们加入有监督数据集,一起重新训练其他视图的分类器,以达到视图一致,提高视图总体的分类能力。这类算法直观有效,也较容易处理异构数据,但因其迭代的训练方式,学习效率相对较低;基于协同正则化的算法为所有视图的分类器统一建立一个目标函数,通过特定约束条件达到视图一致,并一次性地训练完所有分类器。该方法避免了迭代训练,计算效率往往更高。

为各视图分别建模能最大程度保留视图的个性特征,但对视图一致性的要求也更高。当一致性不能被较好满足时,该类算法的效果会有所下降^[11]。另一类多视图学习算法则侧重于将多个视图进行有机融合成单视图,再进行单视图学习^[12]。这类算法有利于提高特征的鲁棒性和泛化能力,不同于简单的堆砌视图,但也具有丢失视图个性特征的风险。为此,有学者提出同时学习视图的共性特征和个性特征^[13]。

除了改良传统算法,多视图学习也正沿着在线化和并行化的方向发展,以迎接大数据的新挑战。在线多视图学习假设训练数据分批,依次呈现给分类器进行学习。此时,如何为分类器设计高效的更新算法,是在线

多视图学习的研究问题。目前研究的主流算法是基于协同正则化的算法,而其在线化的主要策略是通过引入随机梯度下降法,实现目标函数的在线优化^[14]。该策略的优点是计算速度快,存储空间小,但对学习参数的敏感性较高。

目前的并行多视图学习研究主要基于 MapReduce 的框架。如爱立信研究院提出^[15]将不同数据的不同视图被分配到不同计算单元进行计算,以提高计算效率。但由于视图一致性的要求,不同计算单元间往往需要频繁通信,这成为系统效率的决定性瓶颈,也是当前分布式机器学习的一个研究热点。并行或分布式多视图学习中通信量如何增长?有哪些影响因素?如何设计低通信量的高效算法?这些都是大规模多视图学习需要探索的问题。

2 多视图学习应用广

多视图学习在许多研究领域都已获得成功应用,包括网页分类、自然语言处理、计算机视觉、医疗诊断、药物分析和化学分析。在网页分类中,每个网页的内容和链表可表示为两个视图;在文本分类中,同一文本的不同语言版本可表示为不同视图;在用户决策识别系统中,用户的语音和姿势可表示成两个视图;在图像检索和标识中,图像本身和周边的文本信息可表示成两种视图。即使在一些数据并不自然展示出多个视图的问题中,也可通过从单视图中提取多个视图进行多视图学习,以充分发挥多视图学习的优势。

多视图学习在工业界也被积极探索和推广,涉及领域包括机器翻译、情感预测、图像检索和推荐系统等。在微软研究的机器翻译中^[16],测试文章在不同解码器下的译文被视作不同视图下的准参考译文加入训练集参与翻译器的特征权重的训练,从而使翻译器更全面的适应测试集领域的文章的特点,达到领域自适应

的目的。

3 结束语

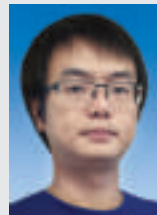
半监督多视图学习是大数据时代极具潜力的分析工具。其在许多研究领域已获得成功应用,并在工业界被积极推广。但同时也需指出,大数据的大容量、分布式和在线化等特征为半监督多视图学习带来了新的挑战。比如,如何有效降低无监督数据的样本复杂度,如何降低分布式视图间的通信量等。要解决这些问题,需要通过引入其它领域新的研究技术和成果,不断丰富和完善半监督多视图学习的理论体系和算法设计,并在实验和实践中不断检验和探索。

参考文献

- [1] LABRINIDIS A, and JAGADISH H V. Challenges and Opportunities with Big Data: A community white paper developed by leading researchers across the United States [R]. 2012
- [2] ZHU X J. Semi-supervised learning literature survey [EB/OL]. <http://pages.cs.wisc.edu/~jerryzhu/research/ssl/semireview.html>
- [3] DELALLEAU O, YOSHUA B and ROUX N. Efficient non-parametric function induction in semi-supervised learning [C]// Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, Bridgetown, Barbados, 2005: 96-103
- [4] XU C, TAO D C, and CHAO C. A Survey on Multi-view Learning [Z]. arXiv preprint: 1304.5634, 2013
- [5] CHEN M M, CHEN Y X and WEINBERGER K Q. Automatic feature decomposition for single view co-training [C]//Proceedings of the 28th International Conference on Machine Learning, Washington, USA, 2011: 953-960
- [6] LIAN W Z, RAI P, SALAZAR E and CARIN L. Integrating Features and Similarities: Flexible Models for Heterogeneous Multiview Data [C]//Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin Texas, USA, 2015
- [7] BLUN A and MITCHELL T. Combining labeled and unlabeled data with co-training [C]//Proceedings of the eleventh annual conference on Computational learning theory, Madison, USA, 1998: 92-100
- [8] WANG W and ZHOU Z H. On multi-view active learning and the combination with semi-supervised learning [C]//Proceedings of the 25th international conference on Machine learning, Helsinki, Finland, 2008: 1152-1159
- [9] BALCAN M F and BLUM A. A Discriminative Model for Semi-Supervised Learning [J]. Journal of the ACM (JACM), 2010,57(3):19. doi: 10.1145/1706591.1706599

- [10] SINDHWANI V, NIYOGI P and BELKIN M. A co-regularization approach to semi-supervised learning with multiple views [C]// Proceedings of ICML workshop on learning with multiple views, Bonn, Germany, 2005: 74-79
- [11] CHRISTOUDIAS C, URTASUN R and DARRELL T. Multiview learning in the presence of view disagreement [C]// Conference on Uncertainty in Artificial Intelligence, California, USA, 2012
- [12] CHEN N, ZHU J and XING E P. Predictive subspace learning for multi-view data: a large margin approach [C]//Advances in neural information processing systems, Hyatt Regency, Vancouver Canada, 2010
- [13] JING X Y, HU R M, ZHU Y P, WU S S, LIANG C. and YANG J Y. Intra-View and Inter-View Supervised Correlation Analysis for Multi-View Feature Learning [C]// Twenty-Eighth AAAI Conference on Artificial Intelligence, Qu é bec City, Qu é bec, Canada, 2014
- [14] Ruijter T D, Tsitsivladze E and Heskes T. Online co-regularized algorithms [M]. Germany: Springer Berlin Heidelberg, 2012
- [15] HARIHARAN C. and SUBRAMANIAN S. Large scale multiview learning on mapreduce [C]// 19th International Conference on Advanced Computing and Communications, Chennai, India, 2013
- [16] MAEIREZO B, LITMAN D and HWA R. Co-training for predicting emotions with spoken dialogue data [C]// Proceedings of the ACL 2004 on Interactive poster and demonstration sessions, Barcelona, Spain, 2004

作者简介



蓝超,美国堪萨斯大学计算机系在读博士;研究方向为机器学习和模式识别;在国际会议及期刊上发表论文10余篇。



饶泓,南昌大学信息工程学院教授;主要研究方向为机器学习、数据挖掘、智能信息处理;主持多项国家及省市科技项目;在国内外学术刊物及国际会议发表论文30余篇。



浣军,美国堪萨斯大学计算机系教授,美国国家卫生局、航空航天局以及自然科学基金评审委员会、委员,Elisver 及 Springer 大数据期刊编委等;研究方向为机器学习和大数据挖掘,以及在生物信息学中的应用;于2009年获美国国家科学基金会教师早期职业发展奖;在《自然-生物技术》等国际学术期刊及会议上发表论文100余篇。

M-ICT 时代下的有线网络

The Wireline Network in M-ICT

柯楚/KE Chu
王翔/WANG Xiang

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

在 M-ICT 时代下, 随着 IT 与 CT 的不断融合, 互联网化取得了巨大成功。以“服务无处不在”、“体验至上”为重要标签, M-ICT 时代的互联网不仅仅作为信息的网络中介, 同时还与各种传统行业相互结合, 充分发挥了其在社会资源配置中的优化和集成作用。将互联网的创新成果深度融合于经济、社会各领域之中, 可以提升全社会的创新力和生产力, 形成更广泛的以互联网为基础设施和实现工具的“互联网+”的经济发展形态^[1]。

作为信息社会的基础设施, 有线网络在 M-ICT 时代也需要适应这一变革。我们认为, M-ICT 时代下的有线网络将呈现出“极速、简单、开放”三大特征。

1 网络“极速”的需求

带宽的增长一直是通信网络发展的源动力。在 M-ICT 时代下, “无处不在的服务”使得信息总量不断快速增长, 而“极致的体验”又要求信息的数字洪流在整个有线网络高速公路中能够顺畅快捷的传输。具体来说, 当前推动有线网络升级提速的主要动力在于: 4G、4K 与云计算。

收稿日期: 2015-08-20
网络出版时间: 2015-09-20

中图分类号: TN929.1 文献标志码: A 文章编号: 1009-6868 (2015) 05-0035-07

摘要: 在 M-ICT 时代, 有线网络需要适应 IT 与 CT 融合的技术变革, 总体上呈现出“极速、简单、开放”的三大特征: “极速”表现为网络各层次、各部分的全面提速; “简单”体现于网络架构的简单化、网络虚拟化以及网络的智能化; “开放”的目的是为了促进网络业务创新, SDN 与 NFV 是实现网络开放的关键技术, 两者间互为补充, 最终实现有线网络的互联网化创新环境。

关键词: 有线网络; 接入网; 承载网; 网络虚拟化; 网络功能虚拟化; 软件定义网络

Abstract: In the M-ICT era, with the convergence of IT and CT, “fast, simple and open” are the three characteristics. “Fast” refers to the comprehensive speedup of the network layers and segments; “simple” refers to simplification, virtualization and intelligence of network; “open” refers to drive network service innovation, and software defined network (SDN) and (NFV) are the key technologies. These two technologies complement each other and eventually achieve the innovation environment of wireline network.

Key words: wireline network; access network; bearer network; network virtualization; NFV; SDN

• 相较 3G 移动通信而言, 4G 技术为移动终端提供了 10X 以上的带宽, 而后续的 pre5G、5G 网络则进一步实现移动网络的速率提升, 给后面的承载网络带来了巨大的带宽容量压力。

• 对最终用户而言, 视频目前仍是产生最大信息量的业务。围绕着“极致体验”, 视频的清晰度不断上升 (目前已经超越 1 080 P 的 FHD 画质并朝向 4 K 进发), 传送视频信息所需的带宽也急剧上升, 这些都对网络端到端的超高清视频传输能力提出更高的要求。

• 云计算的发展使得云计算中心成为网络的关键节点。随着计算能力的不断提升、云计算范围不断扩展, 云计算中心内部的信息交换容量迅速增长, 数据中心互联网络也得到

快速发展。

每一次的带宽提速都会使得一部分计算、存储被转移到云端, 南北向的流量增长率终将迎来拐点, 而东西向的流量则仍然将维持指数级的增长。

数字的洪流对有线网络提出了更高的要求, 包括更高的线路速率、更长的传输距离、更大的交换容量、更高的集成度以及容量能耗比, 而这些一并构成了所谓的网络“极速”的涵义。网络极速体现在网络的各个层次, 需要有线网络中的光网络、分组承载网络、宽带接入网络全面同步演进。

1.1 光传送网络

光传送网络作为最基础的网络传输通道, 以努力挖掘光纤信道传输

能力、降低单位 bit·km 的传送成本为永恒目标。在成本/收入难以增加甚至逐步下降的情况下,速率、容量的持续提升并超过需求是光传送网络发展唯一的出路。

在传输容量上,单通道速率由过去单纯的不归零码(NRZ)调制的 10G 系统,转向如今偏振复用矢量调制的 100 G 系统,即将商用的超 100 G 系统则以灵活调制结合多载波技术实现给定场景的最高速率,单端口速率也不再局限于固定的 400 G,而是 $N \times 100$ G 灵活可配置的系统。未来单纤总传输容量的提升则由单纯的波分复用(WDM)转向空分(少模/多芯)复用 SDM+WDM 传输,这可实现单纤传输容量阶跃式的提升。预计未来可以实现单纤 Pbit 级传输。

在交换容量上,单纯的光传送网(OTN)电交换难以匹配传输容量的提升,集群交换虽可实现交换容量大幅增加,但体积和功耗惊人。未来,我们一方面需要努力探索新型的光电组合或融合的交换技术,充分利用光交换的低功耗与电交换的灵活性;另一方面我们需要通过以硅光技术为代表的光器件创新及应用,逐步实现高集成度、低功耗、低成本的极速光传送网络。

1.2 分组承载网

在 M-ICT 时代,业务的 IP 化已经基本完成,分组承载网当前包括 3 部分:以分组传送网(PTN)/IPRAN 设备构建的城域分组网络、以核心路由器构建的骨干 IP 网络、以互联网数据中心(IDC)交换机/路由器构建的数据中心网络(含云内网络与云间网络)。

对各种分组承载网设备而言,极速体现在线路接口速率的提升、分组交换容量的增长、转发表容量与查表效率的增长。

M-ICT 时代的分组承载网络物理接口朝向以太网接口统一,随着 IC 工艺水平的不断升级,接入设备组网带宽从 1 G 快速发展至 10 G,而核心

与汇聚设备线路速率已规模商用 100 GE 并向 400 GE、1TE 演进;单卡容量从 200 G、400 G 向 1 T、2 T 不断发展,整机容量从 Tbit/s 向上百 Tbit/s 至 Pbit/s 不断延伸。

1.3 宽带接入网络

作为信息高速公路的“最后一公里”,宽带接入网连接终端与网络,其发展受终端业务带宽需求影响。

无源光网络(PON)由于具有传输容量大、可靠性高、便于维护等一系列优势,已占据宽带接入网的主流,并形成一系列光纤接入(FTTx)的宽带接入方案。PON 技术朝向更高线路速率、更大分光比、更长传输距离进行演进。当前 EPON/GPON 技术已可为用户提供几十(Mbit/s)/户的接入带宽,10G PON 技术可为每户提供 100 Mbit/s 以上接入带宽。采用单波长 NRZ 技术进一步提速遇到技术瓶颈;下一代无源光网络(NG-PON2)选择采用结合 WDM 技术进一步提速,提供单纤 40 Gbit/s 以上容量,并支撑 GE 到桌面的能力。进一步的速率提升将朝向密集波分 PON 以及采用正交频分复用(OFDM)调制技术的方向不断发展。

“光进铜退”持续深化,但受用户入户资源限制等原因,铜线仍将长期存在。铜线挖潜也是传统运营商的重要战略,并逐步向用户侧推进的方向演进,从而提供更高的带宽。第 2 代超高速数字用户线路技术(VD2)是当前主流的数字用户线路(DSL)部署技术,可以提供最高 100 M 的接入速率。通过 Vectoring 技术,可显著提升 VD2 接入速率和高速业务的普遍覆盖,传统 DSL 市场的格局基本确定,Vectoring 技术的引入,是市场重新洗牌的机会。作为当前最新的铜线接入技术 G. fast,可以为“最后一百米”提供 1 Gbit/s 的接入速率,并可在相当长的一段时间内满足接入带宽需求。在光纤到户(FTTH)部署场景下,纯 FTTH 与光铜混合 FTTH(G.

fast)将长期存在,G. fast 解决最后一段接入的难题,将有效推进 FTTH 的建设,是 FTTx 解决方案的重要组成部分。

2 网络“简单化”的主要趋势

“简单”的涵义比较广泛,体现在基础设施、网络架构、运营体系以及交互界面上。M-ICT 时代下,网络简单化的主要趋势有:网络融合、虚拟化、智能化。

2.1 网络融合

M-ICT 时代的有线网络融合体现在两种纬度:

(1)网络架构横向层面,接入、汇聚、骨干之间的界限模糊、设备融合。接入光线路终端(OLT)设备容量不断增长,L3 功能进一步增强,并逐渐与汇聚交换机结合,直连宽带远程接入服务器(BRAS);BRAS 的控制面功能被虚拟化至电信城域网之后,其转发面功能则有可能向 OLT 侧转移,形成融合的接入节点。

(2)网络架构纵向层面,光与分组不同层次网络的协同与融合。OTN 本身在适应业务分组化的趋势,其设备的内部交换也逐步采用分组交换技术,与分组化的 PTN/IPRAN 传送设备采用统一的交换架构,形成融合的 P-OTN 设备,并通过多层网络优化(MLO)技术实现对 P-OTN 融合网络的统一控制。未来光分组交换(光突发交换(OBS)、光分组交换(OPS))技术不断成熟后,光与分组网络进一步深度融合,形成统一的光分组传送网络。

2.2 网络虚拟化

网络的虚拟化趋势既包括将多个物理网络设备虚拟为一个网元的“多虚一”虚拟化技术趋势,也包含着将一张物理网络虚拟化为多个逻辑网络的“一虚多”技术趋势。这两种技术趋势分别面向不同的网络简单化诉求,在不同的网络层次采用不同

的虚拟化技术。网络虚拟化一直是网络发展的重要方向,在 M-ICT 时代下,软件定义网络(SDN)技术与网络功能虚拟化(NFV)技术得到发展,成为网络虚拟化的重要支撑,同时也给网络虚拟化带来与传统不一样的新特色。

通过网络“多虚一”的虚拟化技术,多个物理的网络设备被视为一个逻辑设备进行控制管理,大大简化网络的控制管理复杂度。M-ICT 时代一些典型的虚拟化技术方案有:

- 中兴通讯发布的 SDN-based IP-RAN 虚拟化技术方案。在该方案中,在 SDN 控制器的控制下,海量远端接入层 IP-RAN 设备被虚拟成汇聚 IP-RAN 设备的板卡进行管理,汇聚层和接入层设备组成大的虚拟集群系统,使得接入层组网成为设备内部处理,大大简化了由于接入层引入 L3 处理后大量的业务配置复杂问题,提高网络自动化部署、运维和管理能力,提高网络服务能力。

- 中兴通讯发布的接入网虚拟化方案。在该方案中,首先通过虚拟远端网关设备(vRG)功能将光网络单元(ONU)的功能上移到 OLT 之上,使得在 FTTH 场景下,通过管理 OLT 设备即可实现对下挂 ONU 的管理;其次通过虚拟光网络单元(Vonu)功能将多用户居住单元(MDU)逻辑上虚拟为多个 ONU,使其可以与 ONU 一样来进行 MDU 的管理。vRG + vONU 功能可以整体实现在 FTTH/光纤到楼(FTTB)场景下,将整个接入网络虚拟化为以 OLT 为核心的单一网络转发设备。

网络的“一虚多”虚拟化主要需求来源于将网络分割为不同区域应用,如将网络提供给不同的运营商运维,提供给运营商不同的业务部门运营,按不同的安全性要求隔离网络,以网络即服务(NaaS)的形式提供给不同的用户使用等。“多虚一”与“一虚多”的虚拟化技术互为补充,往往首先实现网络“多虚一”后,再通过

“一虚多”实现业务应用隔离。M-ICT 时代下,“一虚多”的功能主要由 SDN 控制器完成,一些典型的虚拟化技术方案包括:

- S-PTN 技术方案。在该方案中,在 PTN 网络中引入集中 SDN 控制器架构,通过 SDN 控制器将网络分割为两张不同的网络,分别承载移动 Backhaul 业务与集团客户的专线业务,再交给运营商不同的业务部门进行分别运营。

- 接入网虚拟化方案。在该方案中,先将接入网“多虚一”虚拟化为一个网元后,由 SDN 控制器实现对该接入网元的“一虚多”虚拟化,不仅可以以不同板卡、不同 PON 口为虚拟化粒度,也能实现基于 ONU 端口甚至虚拟局域网(VLAN)为粒度的虚拟化分割,由不同的运营商运维或根据不同的业务实现不同的网络控制。

- 逻辑路由器技术方案。在该方案中,通过专用的操作配置接口配置独立的硬件资源如物理接口、线卡槽位等到不同逻辑路由器, Fabric、电源、冷却系统、硬件转发引擎、Host CPU 进程、路由表数据库等资源也将限制在不同的逻辑路由器内。旨在保护用户投资并最大限度提供灵活的方法映射物理资源到不同的虚拟设备来简化管理,并按需部署专用设备,保证不同虚拟设备之间的隔离,从而提高用户数据的安全性。

- 虚拟化数据中心(VDC)虚拟化数据中心技术方案。在该方案中,以虚拟可扩展 LAN(VXLAN)作为底层转发技术,以 NFV 方式进行安全服务编排,SDN 控制器收集数据中心物理网络拓扑,统一管理,虚拟化为多个独立的逻辑网络,每个逻辑网络可以单独进行配置和管理。这使得每个用户可以对计算、存储和网络资源统一管理,就像管理一个真正的数据中心一样管理这些资源。

无论是“多虚一”还是“一虚多”的虚拟化,都是为了实现网络的“按需服务”,屏蔽了网络内部复杂的各

种细节,给网络用户提供简单、方便的使用体验。

2.3 网络智能化

网络的智能化进程类似于生物智能进化过程。传统的承载接入网络的智能体现在各个网络层次,形成独立、互不协调的网络智能,例如自动交换光网络(ASON)/基本波分系统的自动交换光网络(WASON)的智能仅局限于同步数字体系(SDH)/波分复用(WDM)光网络,多协议标签交换(MPLS)则通过一系列的分布式协议实现了 IP 网络的智能。这样的智能化水平使得网络各区域间难以保持统筹,也难以实现端到端的业务质量保障。

M-ICT 时代下网络智能进一步“进化”,从低等生物“脑神经节”智能向高等生物“脑”智能发展,最终形成网络整体的智能,构建整体电信网络的“大脑”。

M-ICT 时代下网络智能有两方面的趋势:一方面,有线网络只是作为电信网络的“躯干”,在网络智能化进程中主要是向“大脑”提供信息并接受“大脑”的控制。在网络 SDN 化的趋势下,有线网络的控制逐步集中,越来越多的受到电信网整体的智能“大脑”控制,从另一个角度看也可以说,有线网络本身的智能化水平在下降。但有线网络本身仍将会保留一部分低等级的智能,如实现快速的保护倒换,类似于生物的“条件反射”系统。另一方面,在朝向“大脑”进化的过程中,有线网络本身的“感知”系统则进一步进化发达。以往一些网络单元也纷纷被加入“智能感知”环节,如光接入网络中无源的光配线网络(ODN)朝向智能光配线网络(eODN)发展,通过内置的光时域反射仪(OTDR)、电子标签等技术,使无感知能力的 ODN 能实现智能监测。另外,除传统的操作、管理、维护(OAM)机制外,网络中也会增加各类业务探针(内置或外置)的部署,实现

端到端的监测感知网络质量^[2]。

这种类似高等生物智能的新型网络架构,再加上智能组织的编排器,共同指挥多个智能实体相互协调,打通了从上到下、从局部到全局的智能通道,使得所有业务行为都由软件自动在全网实现,实现真正意义上的简单。

3 网络“开放”的技术支持

相比 IT 业,CT 业环境一直比较封闭,为了获得高性能与电信级业务水平,采用垂直一体的“烟囱式”网络设备结构,设备间的互通依靠电信标准组织形成统一标准。新业务出现时,往往需要等待新的电信标准出现,业务创新周期以“年”为单位。在 M-ICT 时代,IT 与 CT 日益走向融合,CT 在保持高性能与电信级可靠性的同时,利用 IT 的开放思路对电信网络进行重构,形成互联网化的电信设备研发与网络运营体系。

在 M-ICT 时代下,SDN 与 NFV 技术形成网络开放的技术基础。SDN 定位在网络转发功能的集中控制和调度,而 NFV 定位在网络业务功能的虚拟化、云部署。在城域网的范畴,原企业用户终端设备(CPE)、宽带网络网关(BNG)、运营商级的网络地址转换(CGN)等设备的控制面上移,集中放置在数据中心部署,可以更好地共享资源,加快业务部署速度,也有利于新业务的快速、灵活开发。因此,未来在城域网汇聚层引入云设施是一个重要的发展趋势。对于光传输、核心路由器来说,NFV 需求尚不迫切。

转发面是否放在数据中心进行虚拟化部署还存在不同意见,一是服务器作为高速转发设备是否合适;二是对服务质量(QoS)功能的支持方面当前仍比较欠缺。因此,传统的高性能转发设备在相当一段时间有可能继续存在,但会接受 NFV 管理面的统一调度^[3-4]。

SDN 与 NFV 两种技术互为补充,

最终形成芯片→网络设备→网络操作系统→网络应用的类 IT 产业架构。各层之间接口开放,容纳产业链,同时不同开发者根据自己所擅长的方面,快速实现各种网络业务创新,从而能够实现互联网化的快速业务创新。

当前 ONF 所倡导的 Openflow 协议是网络设备控制接口开放的最重要协议标准。我们需要看到,目前的 Openflow 还存在一些不足,如其主要针对网络转发行为的抽象,对网络 OAM 等非转发行为的抽象不够;另外 Openflow 仍是网络协议相关的,随网络协议变化 Openflow 协议本身不断进行增补,限制了网络协议的创新。为了保持对大量已有传统网络转发芯片的兼容,Openflow 采用了表类型模式(TTP)的方式对已有的网络转发芯片进行抽象,能尽快对已有的有线网络进行 SDN 化集中控制,这是当前网络向 SDN 化演进的必经阶段。随着 SDN 的发展,Openflow 必将继续解决网络其他行为的抽象,并实现协议独立,使得网络能够实现真正软件定义^[5]。

在有线网络领域,还有另一种设备的开放,即类似于 Facebook 的开放计算项目(OCP)计划中的 OCP 交换机架构。OCP 开放结构实现交换设备板卡通用、交换机软件的第三方加载,实现类似于兼容机的白盒设备架构,原则上与 Openflow 的 SDN 化思想并不矛盾,但在采用 Openflow 的网络集中控制之后,Openflow 转发设备本身白盒化,业务创新通过 Openflow 接口实施,对交换机上软件创新的压力将大大减轻。两种开放技术存在一定的竞争关系。当 SDN 技术深入发展后,OCP 的白盒架构意义表现得并不突出。

4 结束语

总之,在 M-ICT 时代,“极速、简单、开放”贯穿于有线网络的技术、架构与产业链的发展全程。数字的洪

流促使网络容量不断增长,驱动网络的各层次全面提速;网络架构走向简洁扁平,同时运用软件工程技术,网络的内部复杂度被屏蔽,网络的智能化能力持续提升;SDN 与 NFV 技术深刻影响有线网络的各个层面,通过网络软件化、接口抽象,最终实现完全开放的网络生态环境,获得如同 IT 一般的互联网化网络创新能力。

参考文献

- [1] 孙枕戈. ICT 行业未来趋势:拥抱 ICT 的 M (Mobile)时代[J].中兴通讯技术, 2014,12(1): 2-4. doi: 10.3969/j.issn.1009-6868.2014.01.001
- [2] 周磊. 智能 eODN 的先进运维能力[EB/OL]. [2014-05-23]. <http://www.cww.net.cn/cwwMag/html/2014/5/23/201452314146951.htm>
- [3] NFV White Paper [EB/OL]. [2015-01-3]. <http://portal.etsi.org>
- [4] 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. doi: 10.3969/j.issn.1009-6868.2015.04.009
- [5] Open Networking Foundation. Openflow Switch Specification V1.4.0[S], 2013

作者简介



柯楚,中兴通讯股份有限公司高级工程师,有线规划部副部长,中兴通讯技术与战略专家委员会专家成员;主要研究方向为宽带通信网络及关键技术;通信设备相关专利 10 余项。



王翔,中兴通讯股份有限公司副总裁,有线规划部部长;承担多项国家及省市重大项目;主要研究方向为有线网络、业务软件、核心网及其关键技术。

安全可信智能移动终端研究

Security and Trusted Intelligent Mobile Terminal

中图分类号: TN929.1 文献标志码: A 文章编号: 1009-6868 (2015) 05-0039-006

摘要: 从软件方案、基于可信执行环境(TEE)方案和基于典型安全元件(SE)方案3个方面对智能移动终端安全技术进行了探讨。软件层面探讨了一般运行环境中的安全技术,基于TEE的方案探讨了TEE的系统架构、隔离技术和安全执行技术,基于SE的方案探讨了基于本地SE和云端SE的安全增强技术。认为只有将可信硬件平台和可信软件加以结合,才能为智能移动终端提供完整的安全保障。

关键词: 智能移动终端;可信执行环境;可信计算;安全元件

Abstract: This paper discusses the smart mobile terminal security technology from three aspects: software solutions, solution based on trusted execution environment (TEE) and solution based on secure element (SE). A software-level solution involves security technology used in the rich execution environment. A solution based on TEE involves the system architecture of TEE, isolation technology of TEE and the trusted execution technology. A solution based on SE involves security-enhancement technology based on local SE and cloud of SE. A combination of trusted software and trusted hardware platform guarantees security for smart mobile terminals.

Key words: intelligent mobile terminal; trusted execution environment; trusted computing; secure element

张大伟/ZHANG Dawei
郭烜/GUO Xuan
韩臻/HAN Zhen

(北京交通大学, 北京 100044)
(Beijing Jiaotong University, Beijing 100044, China)

随着移动互联网的发展,智能移动终端的数量急剧增加,功能也日益完善。2013年全球智能移动终端出货量近10亿部;全球计算平台(含PC和智能移动终端)中移动操作系统(Android和iOS)的占比超过50%^[1];2013年中国智能移动终端用户规模为3.2亿,2014年已达10.6亿,较2013年增长231.7%^[2]。截至2014年12月底,手机网民规模达5.57亿,较2013年底增加5 672万人。网民中使用手机上网人群占比由2013年的81.0%提升至85.8%^[3]。

收稿日期: 2015-03-02
网络出版时间: 2015-05-01

基金项目: 国家自然科学基金(61402035);新世纪优秀人才项目(NCET-11-0565);中央高校基本科研业务费专项资金资助项目(2015JBM041)

随着智能移动终端应用的普及,移动终端中存储的敏感信息越来越多,但丰富的通信和数据交换功能为信息泄露和恶意软件传播提供了通道,各种安全问题日益凸显。

1 智能移动终端的安全需求

智能移动终端已从过去的基本通信工具演变为工作、生活工具。它们已经包括多媒体播放、照相、定位、移动钱包、移动办公、移动医疗等新功能。随着用户敏感数据和关键业务在智能移动终端上的不断积累,智能移动终端也越来越需要被保护。

智能移动终端应用环境下的安全需求如下:

(1) 开放环境下的安全需求

不同于传统手机的封闭系统,新

型智能移动终端设备通常都是构建在提供开放式操作环境的操作系统之上,如Android、iOS操作系统。使用这些操作系统的一个主要优点是用户可以随时添加应用程序,同时可以几乎不必考虑对设备稳定性的影响。然而,这种开放式的环境也是设备暴露在不断增长的多种形式的攻击之下。

(2) 数据安全需求

智能移动终端设备上存储着不断增长的个人敏感信息(如联系人、邮件、照片等)甚至是敏感数据(证书、密码等)。为了防止在设备丢失、被盗或者其他不良情况,必须有足够的安全措施来保护这些隐私信息。

(3) 安全连接的需求

通过多种网络技术如3G、4G或者Wi-Fi,以及个人通信手段如蓝牙、近场通信(NFC),越来越多的用户可以使用他们的设备进行P2P通信和访问网络。如何保证连接过程的安全,尤其是终端上的安全接入问题也有待进一步深入研究。

(4) 交易安全需求

使用智能移动终端进行金融交易已经成为移动市场的主流。2014年底,移动金融整体用户规模达到

8.7 亿,较年初翻一番^[2],越来越多的用户选择使用移动端金融交易服务。移动支付的实现方式包括远程支付和近场支付。在移动支付中需要保证信息的机密性、完整性和不可抵赖性、交易的真实性以及解决交易中的身份鉴别等问题。

(5) 管理策略的安全需求

智能移动终端正不断被企业用于承载关键技术及核心应用,同时携带个人设备(BYOD)策略也被大量引入企业。为了防止企业数据泄露和个人使用环境中的恶意软件对企业数据的窃取,必须提高设备安全并引入有效的移动设备管理措施。

2 一般运行环境的终端安全技术

一般运行环境(REE)主要包括运行于通用嵌入式处理器中的一般操作系统(Rich OS)及其上的客户端应用程序。诸如 Android、iOS 等一般操作系统赋予了智能移动终端功能的可扩展性和使用的便利性。与此同时,也带来了多种安全威胁。

2.1 传统的设备访问控制

智能移动终端提供了包括密码配置、用户身份鉴别等传统的设备访问控制机制。以 Android 和 iOS 系统为例。Android 系统提供了身份鉴别、口令设置、重鉴别和鉴别失败处理机制^[4]。Android 系统提供基本密码配置选项,包括设置图案密码、数字密码、混合密码等多种密码方式。有些机型还为用户配置了基于用户生物特征,如面部识别和指纹识别的身份鉴别机制。iOS 的系统管理者可以设定密码强度,可确定用户频繁使用后需要设定新密码的周期^[5]。Android 和 iOS 的用户还可以设置用户错误登录的上限,以及超过这个上限后系统是否擦除设备信息。

2.2 设备数据加密机制

智能移动终端操作系统为设备

中的数据提供了数据加密机制。

Android 3.0 及之后的版本的系统提供了文件系统的加密机制^[4]。所有的用户数据均可使用 AES-128 算法,以密码分组链接(CBC)模式进行加密。文件系统密钥通过使用由用户口令派生出的密钥以 AES128 算法进行保护。生成加密文件系统密钥的加密密钥时,采用标准的基于口令的密钥派生 PBKDF2 算法,由用户口令派生出加密密钥。

iOS 系统中,所有用户数据强制加密^[6]。每台 iOS 设备都配备了专用的 AES-256 加密引擎,它内置于闪存与主系统内存之间的直接存储器访问(DMA)路径中,可以实现高效的文件加密。加密解密所使用的密钥主要来自设备的唯一标识(UID)以及设备组标识(GID)。设备的 UID 及 GID 全部被固化在芯片内部,除了 AES 加密引擎,没有其他方式可以直接读取。只能查看使用它们进行加密解密后的结果。每台设备的 UID 是唯一的。使用 UID 的加密方式将数据与特定的设备捆绑起来,因此,如果将内存芯片从一台设备整体移至另一台设备,文件将不可访问。除了 GID 及 UID,其他加密使用的密钥全部由系统自带的随机数生成器产生。

除了 iOS 设备内置的硬件加密功能,iOS 系统还提供了名为文件数据保护的数据保护方法,进一步保护设备闪存中的数据。每次在数据分区中创建文件时,数据分区都会创建一个新的 256 位密钥,并将其提供给 AES 引擎,以对文件进行加密。这些密钥被称作文件密钥。每个文件的文件密钥是不同的,被加密封装于文件的元数据中。

2.3 应用运行时的隔离机制

智能移动终端为在其上运行的应用程序提供了应用隔离机制。Android 系统提供了沙盒机制,为每个应用在运行过程中提供了一个沙盒^[7]。其具体的实现是,系统为每个

应用提供了一个 Dalvik 虚拟机实例,使其独立地运行于一个进程,并为每个应用创建一个 Linux 底层的用户名,设置 UID。具有相同用户签名的应用通过设置 SharedUserID 方式来共享数据和权限。iOS 沙盒的实质是一个基于 TrustBSD 策略框架的内核扩展模块访问控制系统^[9-10]。应用间不能查看或者修改数据和运行逻辑,并且应用在执行过程中也不可能查看到设备上已安装的其他应用。

2.4 基于权限的访问控制

智能移动终端为在其上运行的应用程序提供基于权限的访问控制机制。在 Android 系统中,每个应用程序都会有一个嵌入式的权限列表,只有用户授予了该项权限,应用才能使用该项功能^[8]。iOS 系统中,GPS 定位功能、接受来自互联网的通知提醒功能、拨打电话、发送短信或电子邮件这 4 项功能需要授权使用^[5]。

2.5 应用逆向工程的防止策略

智能移动终端的应用程序通常会使用各种手段来防止逆向。在 Android 系统中,通常的做法是应用程序的混淆和加壳技术。此外,还有使用动态链接(SO)库和采用 Android 类动态加载技术的方法^[11]防止逆向。iOS 系统中,通过使用统一资源定位(URL)编码加密、方法体方法名高级混淆和程序结构混排加密等方式防止逆向^[5]。

2.6 系统安全更新

类似于桌面操作系统,智能移动终端系统具有不定期系统安全更新机制。通过不断的系统安全补丁或者发布带有新的安全机制的系统升级减少攻击的发生。

3 可信执行环境技术的终端安全技术

尽管在 REE 中采取了诸多安全措施来保障应用和数据的安全,众多

的攻击案例和系统漏洞表明,这些仍然无法保证敏感数据的安全性。因此开放移动终端组织(OMTP)首先提出了可信执行环境(TEE)概念。2010年7月,全球平台组织(GP)第一个提出了TEE标准^[12]。

3.1 TEE 概述

TEE是运行于一般操作系统之外的独立运行环境。TEE向一般操作系统提供安全服务并且与Rich OS隔离。Rich OS及其上的应用程序无法访问它的硬件和软件安全资源。TEE的架构如图1所示^[13]。

图1中,TEE向被称作可信应用程序(TA)的安全软件提供安全可信执行环境。它同时加强了对这些可信应用程序中数据和资源的机密性、完整性和访问权限的保护。为了保证TEE的可信根,TEE在安全引导过程中进行认证并且与Rich OS分离。在TEE内部,每一个可信应用都是独立的。可信应用程序不能未经授权的访问另一个可信应用程序的安全资源。可信应用程序可以由不同的应用提供商提供。TEE中,通过TEE内部接口(TEE internal API)控制可信应用对安全资源和服务的访问。这些资源和服务包括密钥注入和管理、加密、安全存储、安全时钟、可信用户界面(UI)和可信键盘等。TEE将执行一个度量程序,其中包括功能性测试和安全性评估。

TEE提供了介于典型操作系统和典型安全元件(SE)之间的安全层。如果我们认为Rich OS是一个易于被攻击的环境,SE是一个能够抵抗攻击但是应用受限的环境,那么TEE就扮演着介于两者之间的角色。Rich OS、TEE、SE所处的位置和比较如图2所示^[13]。

在一般情况下,TEE提供了一个比Rich OS更高安全等级的运行环境,但它的安全等级比SE所提供的要低。TEE提供的安全性足以满足大多数的应用。此外,TEE提供比SE

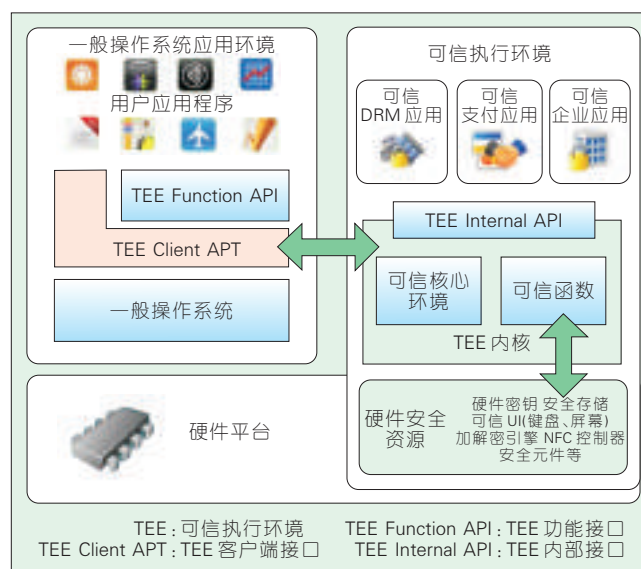


图1
TEE架构

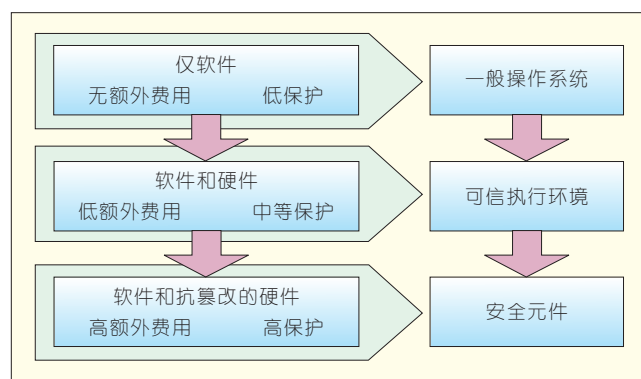


图2
Rich OS、TEE、SE所处的位置和比较

更强大的处理能力和更大的可访问的内存空间。由于TEE比SE支持更多的用户接口和外围连接,它允许在其上开发有一定用户体验的安全程序。此外,因为TEE与Rich OS是隔离的,它能够抵御在Rich OS中发生的软件攻击。

3.2 TEE 系统架构

3.2.1 TEE 硬件架构

芯片级别的TEE硬件它连接着如处理器、RAM和Flash等组件^[14]。REE和TEE都会使用一些专有硬件,如处理器、RAM、ROM和加解密引擎。处理器之外的实体被称作资源。一些能够被REE访问的资源也能够被TEE访问,反之,REE不能访问未经TEE授权的TEE资源。可信

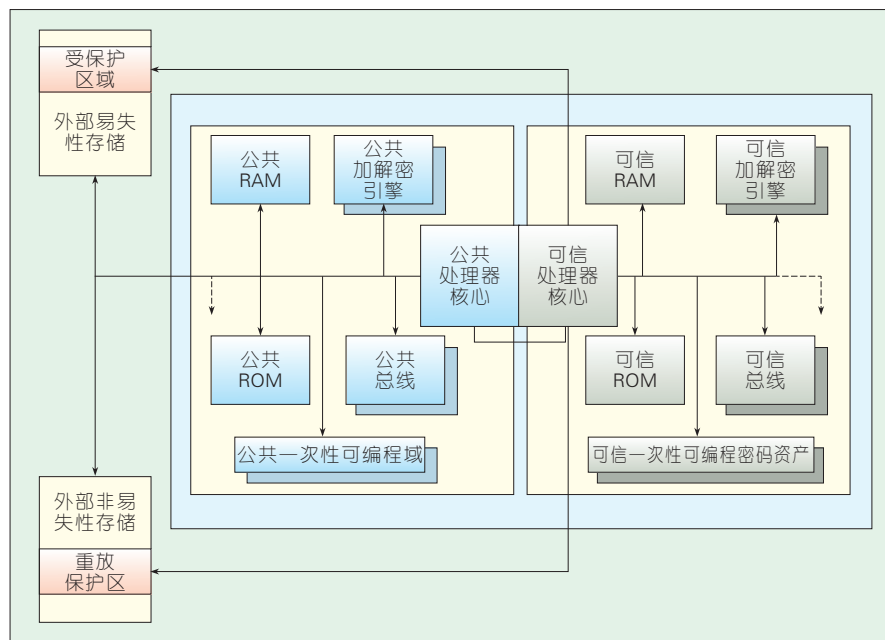
资源仅能由其他可信资源访问,从而保证了与一般操作系统隔离形成封闭系统。一个封装在片上系统(SoC)上的资源组合结构如图3所示^[14]。

3.2.2 TEE 软件架构

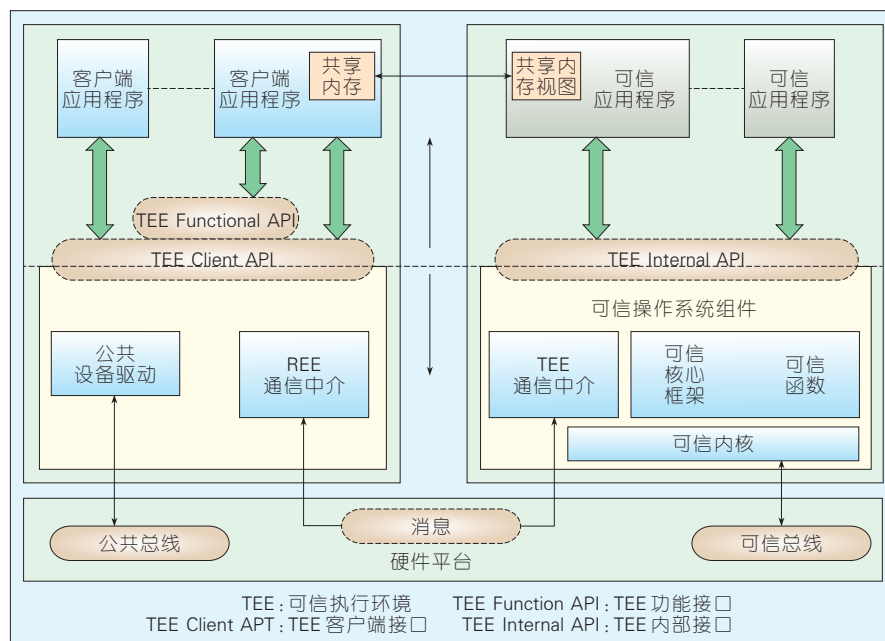
TEE系统软件^[14]架构如图4所示。TEE软件架构的目标是为可信应用程序提供隔离的和可信服务,并且这些服务可以间接的被客户端应用程序(CA)使用。

TEE软件架构包括4部分:REE调用接口、可信操作系统组件、可信应用程序和共享内存。

REE调用借口包括两类API接口,TEE功能接口(TEE Function API),TEE客户端接口(TEE Client API)和一类通信中介。TEE Function API向CA提供一套操作系统友好



▲ 图3 REE和TEE的硬件架构



▲ 图4 TEE系统软件架构

API。允许程序员以类似于编写操作系统应用的方式调用TEE服务,如进行密码运算和存储。TEE Client API是一个低级的通信接口。它被设计用于使运行于Rich OS中的应用程序访问和交换运行于TEE中的可信应用程序中的数据。REE通信中介提供了CA和TA之间的消息支持。

在TEE内部有两类不同的软件结构:可信操作系统组件和可信应用程序。可信操作系统组件由可信核心框架、可信函数和TEE通信中介组成。可信核心框架向可信应用程序提供了操作系统功能,可信函数向开发者提供功能性调用。TEE通信中介与REE通信中介一同工作,为CA

和TA之间的信息交互提供安全保障。可信应用程序是调用可信操作系统组件的API的内部应用程序。当一个客户端应用程序与一个可信应用程序开启一个会话进行交互时,客户端应用程序与可信应用程序的一个实例进行连接。每个可信应用程序的实例都与其他所有的可信应用程序的实例在物理内存空间上隔离。共享内存能够被TEE和REE共同访问,它提供了允许CA和TA之间大量数据有效快速交互的能力。

3.3 使用TrustZone技术的TEE实现

得益于ARM优秀的设计和商业模式,ARM架构非常适用于智能移动终端市场。因此,目前智能移动终端所使用的CPU内核以ARM架构居多。ARM TrustZone技术是ARM提出的系统范围的安全方法。TrustZone技术包括在ARM处理器架构和系统架构上添加的处理器安全扩展、附加总线等技术。使用ARM TrustZone技术构建TEE是绝大多数智能移动终端的实现方式。

3.3.1 TrustZone的隔离技术

TrustZone技术的关键是隔离^[15]。它将每一个物理处理器核心划分为两个虚拟核心,一个是非安全核心(Non-secure),另一个是安全核心(Secure)。同时提供了名为Monitor模式的机制来进行安全上下文切换。TrustZone技术隔离所有SoC硬件和软件资源,使它们分别位于两个区域(用于安全子系统的安全区域以及用于存储其他所有内容的普通区域)中。支持TrustZone总线构造中的硬件逻辑可确保普通区域组件无法访问安全区域资源,从而在这两个区域之间构建强大边界。将敏感资源放入安全区域的设计,以及在安全的处理器内核中可靠运行软件可确保资产能够抵御众多潜在攻击,包括通常难以防护的攻击如使用键盘或触摸屏输入密码等。TrustZone技术的硬

件和软件架构如图5所示^[15]。

3.3.2 TrustZone 技术中的安全启动

安全可信系统周期中一个重要的阶段是系统启动过程。有许多攻击尝试在设备断电后破解软件。如果系统从未经检验真实性的存储上引导镜像,这个系统就是不可信的。

TrustZone 技术为可信域内的所有软件和普通区域可能的软件生成一条可信链。这条可信链的可信根是难以被篡改的。使用 TrustZone 技术的处理器在安全区域中启动。使用 TrustZone 技术的处理器的启动过程如图6所示^[15]。

4 基于安全元件的终端安全技术

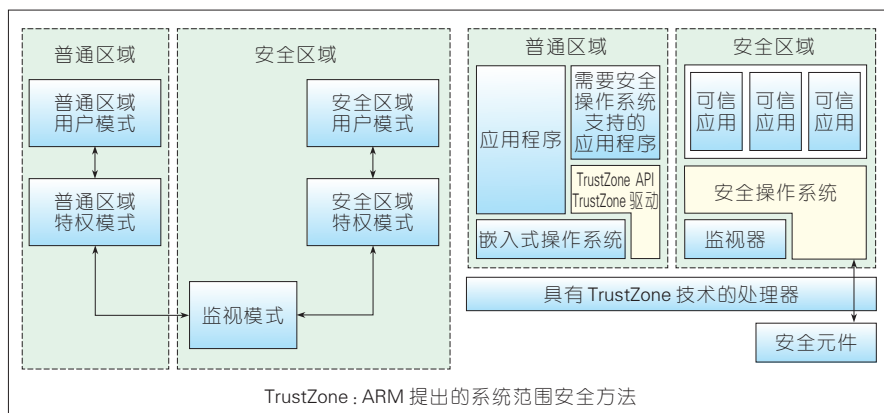
4.1 安全元件概述

GP 组织将安全元件(SE)定义为由硬件、软件和能够嵌入智能卡级应用的协议组成的防篡改组合。它可以通过一组由可信方预设的安全规则和要求来保护应用程序和其机密信息。SE 的典型实现包括 SIM/UICC 卡、嵌入式 SE 和可移动存储卡。SE 提供比 TEE 更高的安全级别,但与此同时它的花费也最高。在近场移动支付中,通常使用 SE 模拟非接触卡。SE 与终端进行通信,发送查询响应,生成动态密码等。最新的安全方案是使用基于主机的卡模拟方式(HCE),这种方式将安全存储和运行环境转移到云端,而不是存储在本地的 SE 中。

4.2 基于本地 SE 解决方案

在 SE 的实现方式中,通常把嵌入在智能设备中的 SE 和嵌入在运营商 SIM 卡中的 SE 称为本地 SE。以谷歌钱包(Google Wallet)和 ApplePay 为例,来探讨在移动支付中基于本地 SE 的解决方案。

谷歌钱包 1.0 版本和 ApplePay 都是基于本地 SE 实现的。他们的工作



▲ 图5 TrustZone 硬件(左)和软件(右)架构

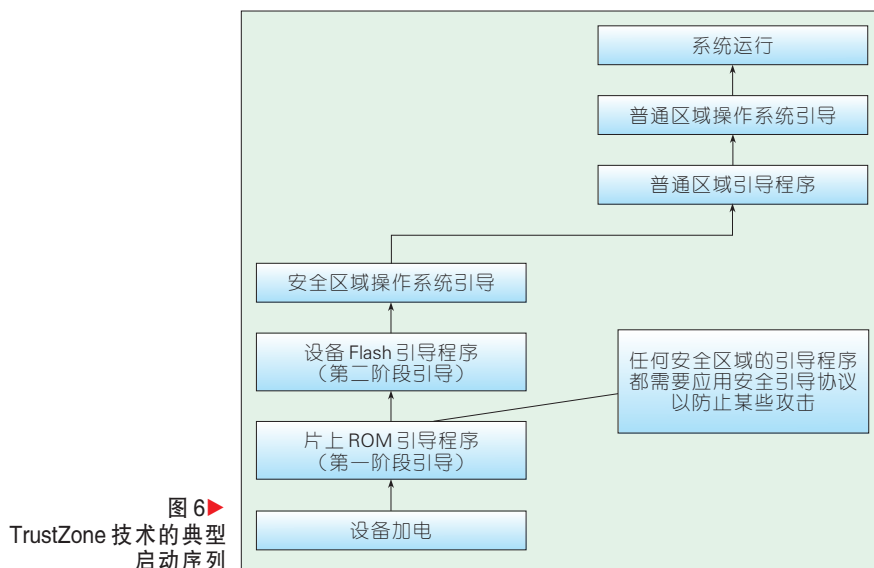
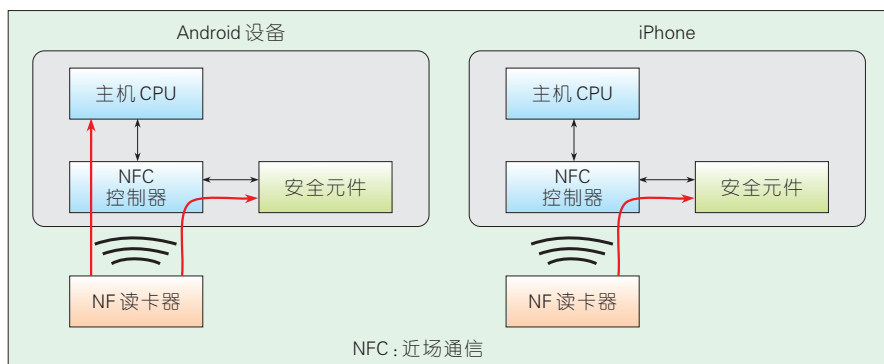


图6
TrustZone 技术的典型启动序列

方式如图7所示。

在谷歌钱包 1.0 中,Android 设备的 NFC 控制器工作在卡模拟模式。在 SE 中存储的移动支付应用程序模拟非接触卡片,使用标准的应用协议数据单元(APDU)指令与终端进行通

信。ApplePay 中使用本地设备 SE 执行卡模拟和安全存储。在许多方面它与谷歌钱包 1.0 类似,但也有重要的差异。ApplePay 在 SE 中不存储真实的主账号(PAN),这与谷歌钱包 1.0 完全相反。ApplePay 存储的是符



▲ 图7 Google 钱包 1.0 和 ApplePay 工作方式

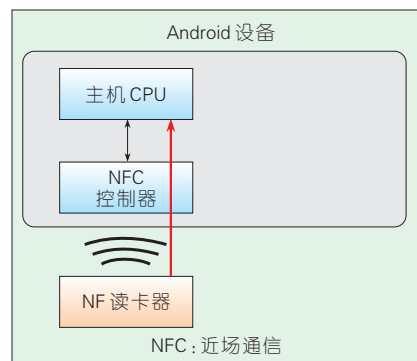
合 EMVco 令牌化标准^[16]的令牌。在交易过程中,这个令牌被发送给终端。在授权流程中,网络识别令牌,去令牌化,生成真正的 PAN,将 PAN 交给发卡行以授权。

4.3 基于云的 SE 解决方案

基于本地 SE 的解决方式本质上是安全的,这是它的一个非常大的优势。然而,它也有一个很大的缺点。SE 的拥有者决定了市场准入。其他所有人都需要通过复杂的商业模式、合作方式和依赖关系进入市场。这让整个过程变得复杂而昂贵。此外,SE 本身的存储容量和处理速度有限也是这种方式的一个缺点。

一种可行的解决方案是使用基于主机的卡模拟方式。Android 4.4 及其后版本的操作系统提供了 HCE 模式的 API^[17]。HCE 模式的运行方式如图 8 所示。

当消费者把手机放置在 NFC 终端上时,NFC 控制器将所有的数据直接发送到直接运行应用程序的主机 CPU 中。然后 Android 应用程序(移动钱包)和特定的支付程序开始进行处理,进行卡仿真,进行请求和响应。由于主机 CPU 本身是不安全的,因此任何真实的支付凭证不会存储在手机钱包中。以谷歌钱包 3.0 为例,谷歌将所有这些真实的数据托管到云服务中,在那里进行安全存储和安全处理。从本质上来说,这是一种基于云的 SE 方式。实现了从基于本地的 SE 到基于云的 SE 的转变。



▲ 图 8 HCE 模式运行方式

这种方法也有它的缺点,如安全性和交易过程中需要网络连接。同时还需要使用如支付卡令牌化的技术来保证它的安全性。然而另一方面,这种方式可以使商业模式、合作关系变得简单,而且没有对本地 SE 的接入限制。这使得应用提供商可以轻松的提供服务。

5 结束语

本文从软件方案、基于 TEE 的方案和基于 SE 的方案 3 个层面对智能移动终端安全技术进行了探讨。软件层面上,在一般运行环境中,主要使用传统的设备访问控制、数据加密、应用运行时隔离机制、基于权限的访问控制、逆向工程的防止、系统安全更新等措施保护智能移动终端的安全。在基于 TEE 的方案中,使用特殊的软硬件体系结构、安全隔离和安全启动机制等来保护智能移动终端的安全。在基于 SE 的方案中,使用了基于本地 SE 和基于云端 SE 的安全隔离和可信执行技术来保证智能移动终端的安全。本文认为智能移动终端的安全解决方案是多层面立体式的解决方案。任何一个层面都有安全性或使用方便性的不足。只有将软件和可信硬件平台加以结合,才能为智能移动终端提供完整的安全保障。

参考文献

- [1] 工业和信息化部电信研究院. 移动互联网白皮书(2014年) [EB/OL]. [2015-03-01]. <http://www.miit.gov.cn/n11293472/n11293832/n15214847/n15218338/index.html>
- [2] TalkingData. 2014 移动互联网报告 [EB/OL]. [2015-03-01]. https://www.talkingdata.com/index/#/datareport/-1/zh_cn
- [3] CNNIC. 第 35 次中国互联网络发展状况统计报告 [EB/OL]. [2015-03-01]. <http://www.cnnic.net.cn/hlwzfzyj/hlwxbzg/>
- [4] Android System and kernel security [EB/OL]. [2015-03-01]. <http://source.android.com/devices/tech/security/overview/kernel-security.html>
- [5] iOS Security Overview [EB/OL]. [2015-03-01]. https://developer.apple.com/library/ios/documentation/Security/Conceptual/Security_Overview/Introduction/Introduction.html
- [6] iOS Security February 2014 [EB/OL]. [2015-03-01]. http://www.apple.com/ipad/business/docs/iOS_Security_Feb14.pdf

- [7] JOSHUAJDRAKE P, OLIVAFORA P, LANLIER Z, et al. Android Hacker's Handbook [M]. John Wiley & Sons, Inc., 2014
- [8] ENCK W, ONGTANG M, MCDANIEL P. Understanding android security [J]. IEEE security & privacy, 2009,250(1): 50-57
- [9] 凌宁, 张文, 牛少彰, 等. 基于 iOS 系统的安全性研究 [J]. 中国电子商情·通信市场, 2013,12(4):91-95
- [10] iOS Technology Overview [EB/OL]. [2015-03-01]. <https://developer.apple.com/library/ios/documentation/Miscellaneous/Conceptual/iPhoneOSTechOverview/Introduction/Introduction.html>
- [11] 梅瑞, 武学礼, 文伟平, 等. 基于 Android 平台的代码保护技术研究 [J]. 信息安全, 2013,23(7):10-15
- [12] Trusted execution environment [EB/OL]. [2015-03-01]. http://en.wikipedia.org/wiki/Trusted_execution_environment
- [13] The Trusted Execution Environment White Paper [EB/OL]. [2015-03-01]. www.globalplatform.org
- [14] TEE System Architecture Version 1.0 [EB/OL]. [2015-03-01]. <http://www.globalplatform.org/specifications/device.asp>
- [15] ARM Security Technology [EB/OL]. [2015-03-01]. http://infocenter.arm.com/help/topic/com.arm.doc.pr29-genc-009492c/PRD29-GENC-009492C-trustzone_security_whitepaper.pdf
- [16] EMV Payment Tokenisation Specification-Technical Framework [EB/OL]. [2015-03-01]. http://www.emvco.com/download_agreement.aspx?id=945
- [17] Host-based Card Emulation [EB/OL]. [2015-03-01]. <http://developer.android.com>

作者简介



张大伟, 北京交通大学计算机与信息技术学院讲师; 主要从事可信计算、智能卡安全方向的教学、科研工作。



郭炬, 北京交通大学计算机与信息技术学院在读硕士研究生; 从事移动安全方向的研究。



韩臻, 北京交通大学计算机与信息技术学院教授、博士生导师, 中国计算机学会信息保密专委会副主任委员, 教育部高等学校信息安全专业教学指导委员会副主任委员; 从事信息安全体系结构和可信计算技术方面的研究和教学工作。

数据中心绿色技术及其发展研究

Green Technology of Data Center and Its Development

中图分类号: TN929.1 文献标志码: A 文章编号: 1009-6868 (2015) 05-0045-005

摘要: 提出了数据中心在制冷、供配电和 ICT 设备等方面的绿色技术, 并指出通过对 xUE 系列指标的分析可以了解 PUE 参数的局限性。认为能源使用效率(PUE)并不是追求的最终目标, 但不会影响对于绿色技术的创新和使用。指出未来需要不断加大技术、标准和政策等各方面的投入力度, 才能逐步实现中国数据中心的绿色健康发展。

关键词: 制冷; 供配电; ICT 设备; PUE 绿色技术

Abstract: This paper proposes some green technologies for cooling, power supply and distribution, and ICT equipment in Internet data center (IDC). By analyzing the configurations of xUE, the limitation of PUE can be clearly shown. PUE (power usage effectiveness) is not the ultimate goal that we pursue, but will not affect our innovation and use of green technology. In the future, we should increase the support for such technologies, standards, and policies to realize the high energy efficient development of IDCs.

Key words: cooling; power supply and distribution; ICT equipment; PUE green technology

岳玉/YUE Yu¹
曹可建/CAO Kejian¹
郭亮/GUO Liang²

(1. 国家计算机网络应急技术处理协调中心, 北京 100029;

2. 中国信息通信研究院, 北京 100191)

(1. National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100191, China;

2. China Academy of Information and Communication Technology, Beijing 100191, China)

一直以来, 信息、通信和技术 (ICT) 产业特别是互联网数据中心 (IDC) 的能耗是一个令业界头痛的大问题。欧美地区较早就开始关注数据中心的能耗和绿色化问题。早在 2008 年, 欧盟就发布了《能效行为准则》, 该准则是企业界对于电子产品能效的一种自愿性承诺, 目前已经推出针对数据中心的能效行为准则。在美国, 越来越多的科技企业如微软、Facebook 和 Twitter 等都积极去人口密度较低、电力能源价格较低的小镇展开数据中心建设。亚洲各国和地区也逐渐开始关注能耗问题: 韩国于 2008 制订了《低碳绿色增长的国家战略》, 提出大力发展低碳技术产业、强化应对气候变化能力、提

高能源自给率和能源福利, 以全面提升绿色竞争力; 日本政府于 2009 年公布了《绿色经济与社会变革》的政策草案, 提出通过实行削减温室气体排放等措施, 推动低碳经济发展; 新加坡于 2011 年发布《SS564》, 该标准是一个认证管理体系, 它为数据中心实现节能提供了方法^[1]。

作为一个数据中心大国, 中国很有必要对绿色技术进行研究, 并通过国际通行的能源使用效率 (PUE) 参数比较, 从根本上促进中国数据中心产业的健康绿色发展, 实现数据中心产业由大到强的转变。

1 IDC 绿色技术研究

1.1 制冷技术

据研究统计, 制冷耗电占到整个

数据中心耗电的将近一半, 所以制冷采取绿色技术进行节能的意义非常重大。

目前, 绿色制冷技术主要是利用自然冷源进行冷却以节省能源, 常见的有新风节能技术、乙二醇自然冷却技术等。

(1) 新风技术

新风技术是利用机房室外的自然环境为制冷源, 当室外空气温度比室内低一定程度时, 通过新风节能装置引进符合机房空气质量要求的室外自然新风, 与机房内热空气直接进行热量交换, 降低机房内环境温度以满足机房环境要求, 从而达到节省空调制冷量、节约能源的目的。机房新风节能通常采用的主要有通过风扇直接进入、采用湿膜通风设备进入以及利用智能换热设备进入等一系列新技术。

(2) 乙二醇自然冷却技术

空调制冷主要靠压缩机的运转将系统内的氟利昂从气体压缩成高温高压的液体, 再通过室外冷凝及室

收稿日期: 2015-07-28
网络出版时间: 2015-09-23

内蒸发,吸收室内的热量,这需要消耗大量的电能。乙二醇节能机组利用冬季室外冷源,通过水泵的运转,将室内的热量传送到室外冷凝器,再由室外冷凝器将乙二醇液体的热量散发出去,大大减少系统的耗电量,在保证机房的恒温、恒湿和洁净条件下,节能降耗。

1.2 供配电技术

在数据中心中,供配电系统的耗电占到整个数据中心耗电的四分之一左右。随着IDC中各种ICT设备板卡密度越来越大,数据中心机柜的功率密度也越来越高,这对供配电系统提出了更高的技术要求。

(1) 开关电源休眠技术

开关电源整流模块的能量消耗包括输出功耗、带载损耗、空载损耗3个部分,其中输出功耗是根据负载电流大小决定的,无法降低能耗;带载损耗取决于整流模块的工作效率;而空载损耗则是负荷未达额定容量造成的。

开关电源整流模块休眠技术就是根据负载电流大小,与系统的实配模块数量和容量相比较,通过智能“软开关”技术,来自动调整工作整流模块的数量,使部分模块处于休眠状态,把整流模块调整到最佳负载率下工作,从而降低系统的带载损耗和空载损耗,实现节能目的。

(2) 高压直流供电技术

24 V、48 V、60 V等直流供电技术在通信领域的应用已经有相当长的时间。为了区分,业内将200 V以上的直流称为高压直流。目前,欧洲的研究机构一般进行350~400 V标准的研究和实验,而中国制订的标准是240 V。

相比传统不间断电源(UPS)供电,高压直流供电系统在开关电源系统侧减少了一次直流(DC)/交流(AC)变换,在IT设备侧减少了一次AC/DC变换,从而提高了电源的使用效率,并且能够有效地提高供电系统

的可靠性。

1.3 ICT设备技术

(1) 核心网设备动态节能技术

核心网设备大部分时间处于轻载状态,此时大部分业务处理设备或者单板却仍然在消耗能源,如果能够将处于空闲状态的设备或单板进行休眠或关闭,就能够减少设备的整体功耗。

单板休眠技术是在话务量低谷时,使部分单板平滑进入省电的休眠待机状态,并在话务量提升到一定程度时自动唤醒单板,以提供更多的处理能力。

CPU动态降频降压技术是根据系统负载动态调整系统的处理能力。系统运行时,实时监控CPU占用率情况,当发现当前CPU占用率低于设定阈值时,就对CPU进行降频操作;当发现当前CPU占用率高于设定阈值时,就对CPU进行升频操作,以降低CPU功耗。

(2) 接入网设备动态节能技术

• 按端口关断技术

随着技术能力的提高,现在单个业务板的接口密度越来越大,但在实际的应用中并不是所有的端口都在使用,这样以来就会带来一些硬件上资源的浪费。

按端口关断技术就是通过对端口在位状态的定时查寻,根据其在位状态进行对应的数据通道使能或关断。目前还不是所有的硬件都能支持在位状态检测。

• 按流量/按业务低功耗设计

有统计结果显示,通常情况下端口的峰值流量不足50%,一天内平均流量不足20%。因此,提高设备的硬件资源利用率,基于业务量的动态功耗设计成为重要的节能技术。按流量的动态节能主要包括流量检测和预判、硬件资源适配调整、流量突发的快速响应等关键技术。

(3) 服务器动态节能技术

• 处理器降频和休眠技术

基于硬件支持,目前服务器已经实现了处理器内核工作频率随业务压力变化而动态调整。内置的分组控制单元(PCU)可对CPU内核进行精确的电源状态调节,提供多达15级功耗(频率)状态以满足不同应用负荷的能耗控制需求,使CPU的能耗曲线更接近应用负载曲线。当业务负载下降到某一门限值时,处理器内核工作频率也会随之下降到某一新的频点,根据新的频点,处理器核电压做出相应的优化调整,从而使实际功耗最低,服务器以更低功耗获得同样的性能。

当处理器处于空闲状态时,可以使其中的一个或几个内核进入休眠状态;根据业务需要再逐步唤醒处理器内核。如果整个系统都处于空闲状态,那么可以使系统进入深度休眠状态;当业务忙时,唤醒系统。对于刀片系统,当业务负荷低时,将业务集中到某几块线卡中,使空闲线卡进入深度休眠状态;当业务量上升到一定程度时空闲线卡唤醒^[2]。

(4) 存储设备动态节能技术

• 智能动态能耗控制技术

存储设备主要是由基于CPU的控制器和磁盘阵列组成的,在存储设备的使用过程中,在不同时段设备的“忙”“闲”可能不一样,智能动态能耗管理就让设备“忙”的时候能耗高,“闲”的时候能耗低。CPU有“忙”有“闲”,现在的CPU架构基本都支持动态能耗控制,即“忙”的时候CPU用较高的电压全速运行,“闲”时用较低电压低速运行,在无业务处理时,CPU便可以进入“休眠”状态,从而实现节能目的。

• 磁盘休眠和下电技术

磁盘休眠和磁盘降速技术可按磁盘负荷情况控制磁盘工作于4种状态:全速工作、磁盘空闲、磁盘休眠、磁盘下电。有一套自适应控制算法控制磁盘的下电时间。以SAS盘为例,磁盘正常运行功耗约为20 W,在休眠时其功耗大约为1 W,单盘可

节能 19 W。一个存储设备的盘数越多,节能效果越显著。

2 xUE 的研究

2.1 PUE 的概念

2007 年,绿色网格(TGG)首次在数据中心领域提出了 PUE 的概念。此后,PUE 逐渐成为国际上比较通行的数据中心电力使用效率衡量指标,也是考察数据中心绿色、节能、环保水平的重要指标。PUE 是一个比率,越接近 1 表明能效水平越好,其计算公式为:

$$PUE = \frac{\text{数据中心总能耗}}{\text{IT 设备能耗}} \quad (1)$$

数据中心的总能耗是维持数据中心正常运行的所有能耗,包括 IT 设备、制冷设备、供配电系统和其他设施的能耗的总和。如果数据中心所在建筑同时用于办公等其他用途,则办公等所耗电能不包括在数据中心总能耗中。在数据中心中,只有 IT 设备能耗被认为是“有意义”的电能。

2.2 其他参数概念

2.2.1 可再生能源利用率

可再生能源是指在自然界中可以循环再生的能源,主要包括太阳能、风能、水能、生物能、地热能和海洋能等。可再生能源对环境无害或危害极小,而且资源分布广泛,适宜就地开发利用。与可再生能源相对的是煤、石油等化石燃料及核能^[3]。

可再生能源利用率(RER)用于衡量数据中心利用可再生能源的情况,以促进太阳能、风能等可再生、无碳排放或极少碳排放的能源利用。

RER 的定义为:

$$RER = \frac{\text{可再生能源供电}}{\text{数据中心总耗电}} \quad (2)$$

2.2.2 CUE 和 WUE

TGG 分别在其 32 号和 35 号白皮

书对碳使用效率(CUE)和水使用效率(WUE)进行了定义和解释。一般来讲,数据中心的能源都来自于电网,不会产生本地 CO₂,那么 CUE 被定义为:

$$CUE = \frac{\text{整个数据中心能耗产生的 CO}_2\text{排放}}{\text{IT 设备能耗}} \quad (3)$$

在(3)中,整个数据中心能耗与 PUE 计算公式中的分子是一样的参数,所以 CUE 的单位是 kg/kWh。

数据中心的 WUE 被定义为:

$$WUE = \frac{\text{数据中心每年的水使用量}}{\text{IT 设备能耗}} \quad (4)$$

可见,WUE 的单位是 L/kWh。

CUE 能够帮助数据中心的管理人员更好地了解 and 衡量基础设施产生的温室气体排放量以及影响,而 WUE 则是能够让工作人员更好地了解 and 衡量对水资源的使用。它们跟 PUE 一起,形成 xUE 系列参数,用以衡量和比较数据中心的绿色化水平。

2.3 PUE 的局限性

2.3.1 绿色技术对 PUE 的影响

毫无疑问,绿色技术的使用将降低数据中心的总耗电 $P_{\text{总}}$,和 ICT 设备的总耗电 P_{IT} ,假设制冷、供配电以及 ICT 等绿色技术减少的总的耗电量为 $\Delta P_{\text{总}}$,而单纯 ICT 绿色技术减少的总耗电量为 ΔP_{IT} ,则会产生(5):

$$PUE_{\text{绿色}} = (P_{\text{总}} - \Delta P_{\text{总}}) / (P_{\text{IT}} - \Delta P_{\text{IT}}) \quad (5)$$

可以得到使用了绿色技术后的 $PUE_{\text{绿色}}$ 与原来 PUE 的差值为:

$$PUE_{\text{绿色}} - PUE = (P_{\text{总}} - \Delta P_{\text{总}}) / (P_{\text{IT}} - \Delta P_{\text{IT}}) - P_{\text{总}} / P_{\text{IT}} = \frac{(P_{\text{总}} * \Delta P_{\text{IT}} - P_{\text{IT}} * \Delta P_{\text{总}}) / [(P_{\text{IT}} - \Delta P_{\text{IT}}) * P_{\text{IT}}]}{\quad} \quad (6)$$

整理可以得:

$$\begin{aligned} (PUE_{\text{绿色}} - PUE) * (P_{\text{IT}} - \Delta P_{\text{IT}}) = \\ (P_{\text{总}} * \Delta P_{\text{IT}} - P_{\text{IT}} * \Delta P_{\text{总}}) / P_{\text{IT}} = \\ PUE * \Delta P_{\text{IT}} - \Delta P_{\text{总}} \end{aligned} \quad (7)$$

能够确定的是, $P_{\text{IT}} - \Delta P_{\text{IT}} > 0$,由此可以得到结论:

(1)如果当 $\Delta P_{\text{总}} / \Delta P_{\text{IT}} > PUE$ 时, $PUE_{\text{绿色}} - PUE < 0$,绿色技术的使用将带来 PUE 绝对值的降低。

(2)如果当 $\Delta P_{\text{总}} / \Delta P_{\text{IT}} < PUE$ 时, $PUE_{\text{绿色}} - PUE > 0$,绿色技术的广泛使用将带来 PUE 绝对值的升高。

可见,各种绿色技术的使用给 PUE 带来的影响是不确定的。很可能出现使用了较多的绿色技术,但 PUE 反而升高的情况。此时作为一个指标,PUE 显现出一定局限性,如何避免这种局限性需要进一步研究。

2.3.2 波谷用电对 PUE 的影响

电力是一种难以储存或储存成本很高的特殊商品,应该尽量做到即时生产、即时消费。早在 2010 年,发改委、电监会等六部委就联合印发《电力需求侧管理办法》,针对电力需求提出了 16 项定性或定量的管理措施,其中提到“将推动并完善峰谷电价制度,鼓励低谷蓄能”等,该《办法》于 2011 年 1 月 1 日起开始实施。

该《办法》促使 IDC 企业积极采用蓄冰池等技术来充分使用波谷电进行制冷蓄能,这样既贯彻了政府的政策,也会在一定程度上降低数据中心的制冷费用,是一个降低运营成本的很好的做法。这种方式将高峰用电转移到低谷时段,既缓解了高峰电力供需缺口,又促进了电力资源的优化配置,是一项“削峰填谷”的双赢策略。但是对于数据中心来说,是否进行波谷蓄能对总的制冷量和总的电能消耗的影响不会太大,所以这种方式在 PUE 上不一定会体现出来。但是很明显,这种方式对整个社会的用电平衡具有积极的贡献^[4]。

3 中国 IDC 绿色技术的应用现状

3.1 大型 IDC 运营商的新技术尝试

2012 年底到 2013 年初,云计算发展与政策论坛对百度、腾讯、阿里以及三大运营商的 6 个典型机房进行

了绿色技术的调研和分级评估。这6家目前中国最大的数据中心运营商已经开始在实践中采用更多的绿色技术来实现数据中心的节能。

其中,百度 M1 数据中心实践了“数据中心是一个整体,节能从源头开始”这一理念,通过基础设施、IT 设备和软件协同,打造出高效绿色数据中心,取得了 PUE 第一的佳绩。位于天津的腾讯某数据中心单元大规模应用冷热通道隔离、水侧自然冷却、自然风冷却、变频技术、分时式制冷等多项节能技术,绿色节能能力居于中国领先水平。中国电信的永丰国际数据中心从设计、建设到运营维护,均采用先进的绿色节能理念,通过冰蓄冷、水冷自然冷源、高压直供、分散式供电、太阳能等新技术以及科学运维,获得了显著的节能效果。

位于杭州的阿里巴巴某数据中心是阿里巴巴集团与杭州电信联合建设的高效绿色数据中心,积极采用高压直流、市电直供、新风自然冷却等技术,并在直流服务器等方面获得了突破,节能效果十分显著。位于无锡的中国移动某数据中心通过密闭精确送风、配电系统监测等绿色节能技术和完善的节能管理方法,实现能源效率的显著提升,为传统中小型数据中心的绿色节能改造树立了典范。位于天津的中国联通华苑国际数据港采用了谐波治理、高压水喷雾等绿色节能技术,实施了提高机房送风温度、电能精细化管理等节能措施,为探索租赁型数据中心的绿色节能积累了宝贵的经验。

3.2 IDC 产业存在的问题

据统计,近7年来,中国 IDC 市场规模增长了约6倍。相比欧美发达地区,中国数据中心数目庞大,但规模偏小。当前,为满足快速增长的市场需求,各地加快了数据中心特别是中大型数据中心建设步伐。但是,在发展的同时,我们还面临着一些挑战和持久战问题。

首先是新技术创新和引入缓慢。中国数据中心平均能效利用率普遍在 2.2~3 之间,而发达国家多为 1.5 左右,谷歌等大型互联网公司的数据中心甚至能够达到 1.1 或更低。由于缺乏足够的政策支持和引导,以及早期的资金投入比较大等因素,导致中国数据中心运营企业通常采取跟随战略,投资意愿不强,创新能力不足,技术引进缓慢。

其次是共享意识不强。就国际上来看,虽然各互联网巨头一向对其数据中心讳莫如深,但还是能够开放一些最佳实践,特别是 Facebook 的 OCP 项目,将其硬件设计开源,体现出行业领导者的责任感和使命感。在中国,企业对于新技术的共享意愿不强,只有极少数领先企业愿意向社会公开他们的技术细节。表面上看这是企业的共享意识问题,但是往深层次看,实际上这是知识产权保护的问题。

还有一点是数据中心布局不合理。据统计,目前中国可以提供 IDC 服务的数据中心有 500 个左右,主要分布在北京、广东、华东等经济发达、业务量大和带宽资源充足的地区。这些地区本来的自然条件、能源供应的情况就不太理想,数据中心的扎堆给这些地区的能源使用更是带来了严峻考验。

4 对中国数据中心发展的建议

我们发现 PUE 存在一定的局限性,所以 PUE 不是我们追求的最终目标,但这不会影响我们对于绿色技术的创新和使用。中国数据中心能耗使用效率较低是不争的事实,我们需要采取一定的措施来推动和促进数据中心产业的发展。

(1) 需要加大新技术的研发力度。推动产业发展的根本是技术,数据中心的各个相关方面,包括供配电、制冷和 ICT 设备等,中国都有领先的制造商,但是各种先进技术还主

要为其他国家的厂商掌握。面对巨大的数据中心市场,需要这些企业将使命感和责任感转化为对新技术研发和推广的动力。只有掌握了核心技术才能真正掌握主动权。

(2) 加大标准化力度。目前,已有多个通信行业标准组织已经积极开展数据中心相关标准的研究工作。2013 年 4 月,工信部发布了《电信互联网数据中心(IDC)总体技术要求》、《电信互联网数据中心(IDC)的能效测评指标》、《互联网数据中心分级分类技术要求》等 4 个相关标准。这几个标准仅仅是一个起点,对于支撑中国 IDC 行业的发展还远远不够,需要进一步加强 IDC 运营维护等其他方面的标准化进度。

(3) 加大政策支持力度。政策对数据中心的发展具有推动作用,2012 年 11 月 30 日由工业和信息化部发布的《关于进一步规范因特网数据中心(IDC)业务和因特网接入服务(ISP)业务市场准入工作的实施方案》放开了 IDC 牌照的申请;2013 年初由工业和信息化部等部门发布的《关于数据中心建设布局的指导意见》为数据中心提供了宏观的建设指导。行业还需要一些实践指南、与电力政策挂钩等等鼓励性政策,来进一步推动中国数据中心的良性发展。另外,政策对数据中心的发展具有保驾护航的作用。需要有更多的知识产权保护相关方面的政策出台,以保护和推动企业的技术创新。

(4) 加大评测体系支撑力度。测评有利于行业内的横向对比,并可以促进竞争。同时,测评也有利于政府对 IDC 行业的掌控。2013—2014 年,工信部电信研究院开展了两轮绿色数据中心评测工作,评估对象包括阿里巴巴、腾讯、百度等三大互联网公司,以及中国电信、中国联通和中国移动等三大运营商以及业内一些 IDC 运营商的数据中心。这也是中国国内依据行业标准、由第三方组织开展的权威评估活动,评估得到了 TGG 等

国际组织的大力支持。从发布的评估结果来看,部分云计算数据中心通过一系列先进技术的应用,能源有效利用率已经能够达到 74%(PUE 为 1.36),中国的数据中心行业与国际领先水平的差距正在逐渐缩小。

5 结束语

目前,中国已成为数据中心大国,只是很多数据中心是直接从机房改造而来,当时的改造主要从业务快速部署的角度出发,而关于能效方面考虑得较少,这导致中国数据中心的平均能效水平较低。但我们也应该注意到,中国新建的很多数据中心已经在积极地进行更多新技术的尝试和应用。百度 2015 年 7 月建成的阳泉数据中心太阳能光伏发电项目成功并网发电,这是太阳能光伏发电技术在国内数据中心的首例应用;9 月,阿里巴巴的千岛湖数据中心正式启用,因地制宜地采用深层湖水进行

制冷,这是中国首个利用自然水冷却的数据中心,与普通数据中心相比,它一年可节电约数千百度,减少的碳排放量相当于 1 万多吨标煤。我们通过研究 IDC 在制冷、供配电和 ICT 设备方面的绿色新技术,不断加大技术创新和引入力度,加大政策的支持力度,加快相关标准的制订力度以及测评体系的支撑力度,从而实现中国数据中心的绿色健康发展。

参考文献

- [1] TGG22 White Paper. Usage and Public Reporting Guidelines for PUE/DCiE [Z]. 2009
- [2] TGG32 White Paper. Carbon Usage Effectiveness (CUE): A Green Grid Data Center Sustainability Metric [Z]. 2010
- [3] TGG35 White Paper. Water Usage Effectiveness (WUE™): A Green Grid Data Center Sustainability Metric [Z]. 2011
- [4] 李洁,石友康.通信节能减排技术研究[J].《电网技术》,2011(1): 36-40

作者简介



岳玉,国家计算机网络应急技术处理协调中心处长、高级工程师;现从事数据中心建设规划指导、机房节能等新技术研究与应用。



曹可建,国家计算机网络应急技术处理协调中心助理工程师;现从事数据中心建设与运维管理、动力环境监控、机房节能等新技术研究与应用。



郭亮,中国信息通信研究院技术与标准研究所云计算标准与测试验证北京市重点实验室高级工程师;长期从事数据中心相关的政策研究、标准制订和测试认证等工作;主持和参加了多项数据中心相关标准的编写工作,并参与了工信部、发改委、科技部等众多云计算、数据中心国家专项和研究课题;已发表论文十余篇。

综合信息

未来 3 年网络信息基础建设投资将超 1.2 万亿元

国家发改委副主任林念修在 2015 中国“互联网+”峰会上提出,发改委将加大政策支持,在政策环境方面研究制订准入互联网清单;在全面改革创新试验区,和国家自主创新示范区等地,组织开展互联网创新发展试点;将综合运用政策先行先试、单项建设资金、投资基金等各类手段,对区域试点进行支持。

会议上,国家发改委高技术产业司联合中国信息通信研究院发布《国务院关于积极推进“互联网+”行动的指导意见》解读报告。

林念修在解读中介绍,“十三五”规划与“十二五”规划最大的变化就是突出创新驱动发展,因此实施“互联网+”计划对于推进“十三五”意义重大。他预计,未来 3 年,网络信息基础设施建设投资将超过 1.2 万亿,“互联网+”新技术制造、能源等将形成一大批新的投资热点;同时“互联网+”将催生大量新消费,2014 年中国信息消费规模达到 2.8 万亿,预计 2015 年将达到 3.2 万亿。目前互联网直接关联产业对中国 GDP 贡献率达 7% 以上。

对于指导意见中提到的“实施重大工程包”计划,

林念修在会上表示,将在 11 个重点领域,推进实施“互联网+”重大工程,抓好一批有影响能落地的重点项目;同时在全面改革创新试验区,和国家自主创新示范区等地,组织开展“互联网+”创新发展试点;将综合运用政策先行先试、单项建设资金、投资基金等各类手段,对区域试点进行支持。

他还提到,将尽快组建国家新型产业创业投资引导基金,更多投向“互联网+”领域,支持初创期创新型企业发展。另外,国家发展改革委正在联合 32 个部门和单位,加快组建“互联网+”行动机制。

对于“互联网+”如何落地,国家信息化专家咨询委员会常务副主任周宏仁则认为,中国目前“互联网+”发展,不能脱离中国制造 2025,应该聚焦企业发展,特别制造业核心企业,否则中国要彻底错失新兴工业化的发展机遇。他认为,当下“互联网+”的发展,在电子商务和服务业方面走得比较快,这是前十年的成绩,当下“互联网+”的核心内涵还是在制造业上。特别是装备制造业的“互联网+”,是“互联网+”的硬骨头,这会决定中国未来的发展。

(转载自《中国信息产业网》)

基于信息熵的 POMDP 模型观测函数估计

Information Entropy Based Estimation for Observation Function of POMDP Model

中图分类号: TN91; TN919.8 文献标志码: A 文章编号: 1009-6868 (2015) 05-0050-006

摘要: 部分可观测马尔可夫决策过程(POMDP)广泛应用于建模决策任务。模型中的观测矩阵主要用来建模环境的不确定性,通常很难从训练数据中直接获取,需要引入额外的信息进行估计。通过引入信息熵来修正模型中的观测矩阵,修正后的观测矩阵更能反映环境的不确定性。模拟环境下的实验表明,引入信息熵进行修正估计的观测矩阵有效提高了 POMDP 模型的性能,而在基于 POMDP 模型的对话系统中,修正的估计提高了系统的决策准确度。

关键词: 部分可观测马尔可夫决策过程;不确定性;意图识别;观测矩阵;信息熵

Abstract: Partially Observable Markov Decision Process (POMDP) is a decision model used extensively for decision tasks. The observation matrix of the model is a channel that reflects the uncertainty of surroundings, which is hard to do directly from the corpus. Extra information needs to be introduced for estimation of the observation matrix and better reflection of surroundings. The concept of information entropy is introduced to modify the observation matrix in the model by which the modified observation matrix can reflect the uncertainty of the situation more precisely. Simulated experiment and real situation show that introducing information entropy to modify the observation matrix improves performance of the POMDP model together with the decision-making accuracy in a dialogue system based on POMDP.

Key words: partially observable markov decision process; uncertainty; intention identification; observation matrix; information entropy

钟可立/ZHONG Keli
王小捷/WANG Xiaojie

(北京邮电大学, 北京 100876)
(Beijing University of Posts and
Telecommunications, Beijing 100876, China)

略, POMDP 将客观世界的动态特性用状态转移来描述,在状态不完全可观测的情况下,系统通过与环境交互,进行决策,管理对话过程。

虽然基于 POMDP 模型的对话管理具有更大的灵活性,可用于状态不完全可观测的情况,但是模型需要统计状态转移矩阵,在训练过程比较依赖于数据;需要定义回报函数,目前没有很好的估计方法;需要估计观测概率矩阵,一般情况下是很难从数据中直接估计。这些问题会给基于 POMDP 模型的对话管理带来困难。

基于 POMDP 模型的对话管理是根据当前的信念状态分布来选择动作,因此信念状态的估计会直接影响到系统决策性能,而观测函数的估计也会直接影响到信念状态的估计。在对话系统中, POMDP 模型的观测函数一般是建模从用户对话意图到系统状态的映射,但是,用户对话意图本身并不是直接可获得的,它是由用户的语言输入来体现的。一方面,用户对话意图十分丰富,且同一对话意图可以用几乎是无限种不同的语言输入来表现,另一方面,对一个确定的语言输入的意图识别也会存在

人机对话是语言信息处理中的一个重要应用任务,对话管理是对话系统中的核心组成部分,决定了对话系统的质量,为此出现了很多关于对话管理模型的研究。

已有的对话管理技术^[1]主要包括:基于自动机的方法^[2-3],基于框架的方法^[4-5],基于信息状态的方法^[6],基于概率模型的方法^[7-8]。基于自动

机的方法虽然易于设计,但是不灵活,不自然,难以应付复杂的任务。基于框架的方法实现的复杂度较低,但是对话比较机械,人机交互的自然度较低。基于信息状态的方法比较灵活,能丰富的表达对话的状态,把对话过程看作是一列规则的变化,但是需要人工的定义规则与策略。概率方法主要基于马尔可夫决策过程(MDP)与部分可观测马尔可夫决策过程(POMDP)模型。与前3种方法相比,基于 POMDP 的方法不需要人工定义策略,而是通过数据学习策

收稿日期: 2014-11-12

网络出版时间: 2015-02-22

基金项目: 国家自然科学基金(61273365); 国家高技术研究发展(“863”)计划(2012AA011104)

不确定性,更不用说语言输入本身的识别也可能存在不确定性。因此,很难直接从训练数据直接获得,而需要另外的技术来进行估计。对观测函数专门研究的文献不多。文献[9]简单地把语音识别错误率引入到观测函数的估计中,主要建模语音输入识别错误的影响,但并不考虑对识别结果进行用户意图识别时的错误。这种估计方法也没有把训练数据的信息考虑到里面,性能较差。文献[10]结合连续信息与离散信息并应用到贝叶斯网路里面,观测的集合大小有4种,即观测是离散的,连续信息指的是观测的概率,但需要相关阈值确定离散信息,阈值的确定给计算POMDP的回报值带来困难。文献[7]对观测函数进行扩展,在原来的基础上加入置信度(Confidence)分数,用置信度来衡量该观测包含的信息量和对策略规划的重要程度,但是此方法的参数 h 很难估计,由于是直接引入置信度,因此很难在现有的工具训练模型。

观测函数是描述内部状态与外部环境之间的联系。对话管理的一个关键性的问题是部分可观测带来的不确定性^[11-12],而信息的不确定性可以使用信息熵来衡量,因此为了使观测函数更能反映环境的不确定信息和其所包含的信息量,本文引入信息熵来修正观测概率。

1 POMDP模型及观测函数的估计

一个POMDP模型可以用一个六元组 $\langle S, A, T, R, Z, O \rangle$ 来描述,其含义如下:

- * S : 系统的状态集合
- * A : 系统的动作集合
- * $T(s', a, s)$: 系统的状态转移函数,描述的是当系统在状态 s 下执行动作 a 可能转移到 s' 的概率
- * $R(s, a)$: 系统的报酬函数,描述的是当在状态 s 下执行动作 a 时,系统获得的立即回报值

* Z : 系统的观测集合

* $O(s', a, o)$: 系统的观测函数,其中 $o \in Z$

POMDP问题结构如图1所示。每个时刻,系统会处于一个隐状态 s ,系统会根据当前的信念分布 b ,选择一个动作 a ,得到一个立即回报 r ,然后转移到下一个隐状态 s' , s' 依赖于 s, a 。此时系统会得到新的观测 o' , o' 依赖于 s', a ,然后根据动作 a ,新的观测 o' ,更新当前的信念分布,继续选择动作。其中关于 o' 的观测概率,可用于确定转移到 s' 状态的置信度。

在POMDP中,观测函数一般情况下是很难直接从数据中估计得到,需要额外的信息去加强对观测函数的估计,反映环境的不确定性。

文献[9]在研究基于语音输入的人机对话系统建模时,把语音识别错误率引入到观测矩阵的估计中,而语音识别结果到意图的映射是一对一的。虽然这里是针对语言识别错误,假设输入为文本,不存在语音识别错误时,这里的语音识别错误也可以看成是意图识别的错误,因此,相同的模型可以建模文本输入时存在意图识别错误的情形。因此,后文中我们会替换使用语音识别和意图识别,其反映的都是类似的观测中的不确定性,可以类似进行建模。设 p_{err} 为语音识别错误率,如果用户的意图与系统观测到的用户意图或行为一致的话,认为观测概率为 $1 - p_{err}$;否则观测概率 $\frac{p_{err}}{|A_u| - 1}$,其中 $|A_u|$ 为观测集合的个数。这种方法把语音识别的整

体错误率信息引入到观测函数里面,但是没有反映当前对话的信息。

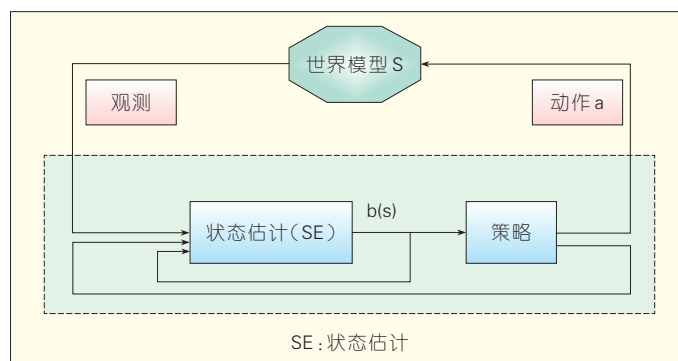
文献[7]和文献[13]对文献[9]观测函数进行扩展,在原来的基础上加入置信度分数,反映当前对话的信息。在每个观测概率前乘以相应的观测置信度分数。这种方法把观测 o 分成两部分,一部分是离散的观测值,代表语音识别或意图识别的假设;一部分是连续分数,衡量该假设的置信度。由于直接在观测函数里面加入连续性的分数,目前训练工具基本都是要求观测部分是离散的,因此无法在现有的工具上直接训练。

2 改进的观测函数估计方法

在文献[14]中,每一个语音输入只有一个识别结果,即只使用识别率最高或分数最高的结果,相当于使用1-best识别结果。

但是,语音识别通常并不完美,采用1-best显然没有采用 n -Best($n > 1$)得到的信息丰富。直接利用1-best结果作为观测已经一定程度损失了语音信息。例如:假设一个语音输入的标准识别结果是“我要购买从北京到广州的机票”,其1-best结果为“我要购买东京到广州的机票”。如果还有2rd-best的结果“我要购买从北京到广西的机票”以及3rd-best的结果“我要购买从北京到杭州的机票”,甚至 N -best的结果...其中 N -best列表中的第一个地点很有可能出现“北京”的次数多于其他的地点,采用语音识别的 N -best结果可以使观测的结果更能反映真实的状态,使信念

图1
POMDP结构



分布提供更具体的信息。

本文关注输入为文本的人机对话系统管理。如上所述,对于文本输入的对话,语言输入假设是正确的,那么需要考虑的是意图识别的错误。在采用 POMDP 建模时,需要基于文本输入估计对话意图作为真正的观测。一般,对每轮对话中用户的输入进行一次意图识别,对于一段语言输入判断其言语意图,可以用分类器(意图识别器)来进行识别,本文在实验中会选择采用两种不同的意图分类器进行考查,无论哪种分类器,均可以得到 N -best 识别结果。随后的问题就是,如何有效利用这些 N -best 信息。

本文提出利用 N -best 信息,计算其信息熵,引入到观测函数中,用于修正观测矩阵的值,使观测函数包含环境不确定性的信息,更加客观的描述环境。

为每个观测得到的是一个 N -best 列表(由意图识别器获得), N -best 列表记录的最可能的 N 个用户意图假设, N -best 列表中同时还记录了意图识别器为每个用户意图分配的概率,例如:

$$o = [\langle \tilde{a}_u^1, p_1 \rangle, \langle \tilde{a}_u^2, p_2 \rangle, \dots, \langle \tilde{a}_u^N, p_N \rangle]$$

其中向量的每一个元素都是(意图、概率)对,而对于概率 $p_n = p(a_u^n | o), n = 1 \dots N$; $\sum_i p_i = 1, i = 1 \dots N$, 我们通过下面的公式引入信息熵进行修正。

$$p(\tilde{a}_u^n | o, a_u^n) = \begin{cases} \alpha_{correct}(p_n)(1 - p_{err}) & \tilde{a}_u^n = a_u^n \\ \alpha_{incorrect}(p_n) \frac{p_{err}}{|A_u| - 1} & \tilde{a}_u^n \neq a_u^n \end{cases} \quad (1)$$

$$\alpha_{correct}(p_n) = \frac{e^{p_n * H(o)}}{e^{H(o)} - 1} \quad (2)$$

$$\alpha_{incorrect}(p_n) = \frac{e^{p_n * (1 - H(o))}}{e^{1 - H(o)} - 1} \quad (3)$$

其中, $\alpha_{correct}$, $\alpha_{incorrect}$ 大于 0, $H(o) = -\sum_i p_i \log p_i$, 为了方便计算与分析,把观测 o 的信息熵做一个转换,

使其的取值范围在 $[0, 1]$ 之间,令

$$H(o) = \frac{H(o)}{\log |N|}, N \text{ 为观测集合的大小。}$$

从 $\alpha_{correct}$ 可以看出, $H(o)$ 越接近 0, 即信息熵越小,人处理信息的可靠度越大,越相信该观测结果,该观测概率越大。反之,该观测结果越不可信,观测概率越小。

对话管理的一个关键性的问题是部分可观测带来的不确定性。根据信息熵的概念可知,描述环境的信息越精确,信息的熵越大,即信息的精确性与熵之间是正相关的。因此把信息熵引入到观测函数里面,能有效反应环境的不确定信息。

3 实验

3.1 实验环境

本节基于对话系统对修正后的观测函数对 POMDP 模型的性能进行定量分析。实验分别在模拟和实际的对话系统中进行。

模拟系统实验是在给定一组策略下,经过多轮决策,估计总的期望回报。使用基于 SARSOP 算法的 Approximate POMDP Planning 工具包,模拟决策次数是 1 000 次。

真实系统实验是在一个以机器人教学为目的的对话系统下进行的,通过接收实际的人类语言输入, POMDP 模型采用修正后的观测函数(其他参数不变)进行决策,产生相应的应对句子。

实验采用的对话系统结构如图 2 所示, UserState 为用户的状态即意图, Obs 为系统的观测部分,由于需要计算 Obs 的信息熵,因此该观测值不是 1-best, 而是 N -best, Action 是对话系统的动作, Reward 是系统得到的长期累计回报值,并不是当前得到的立即回报值。

系统根据用户的意图来选择动作,达到学习,回答,询问的结果。在该系统的设计中,使用 156 个对话作为训练语料,建模一个基于 POMDP

的对话系统,需要完成 7 部分工作。

(1) 状态

在 POMDP 模型框架之上构建的对话系统,模型中的信念状态包含当前状态的信息和历史信息,允许系统不直接知道确定的状态,根据信念状态分布做出决策,所以可以把用户的意图直接作为状态变量。用户意图总共分 4 种类型:对话开始/结束状态;教学意图,教学形状、类别、颜色、名称、部位;查询意图,查询名称、颜色、形状、部位;其他意图。

(2) 动作

动作是对话系统根据当前信念状态分布信息采取的操作,用于与用户交互。对话系统的动作可以分为 5 类:对话开始/结束;回答,回答形状、类别、颜色、名称、部位;提问,提问名称、颜色、形状、部位;学习,学习形状、类别、颜色、名称、部位;其他。

(3) 观测

这个系统的观测是用户说话意图的表现,主要是用于描述在系统的角度下用户意图的表现,所以把观测状态集合定义为和状态集合一样。

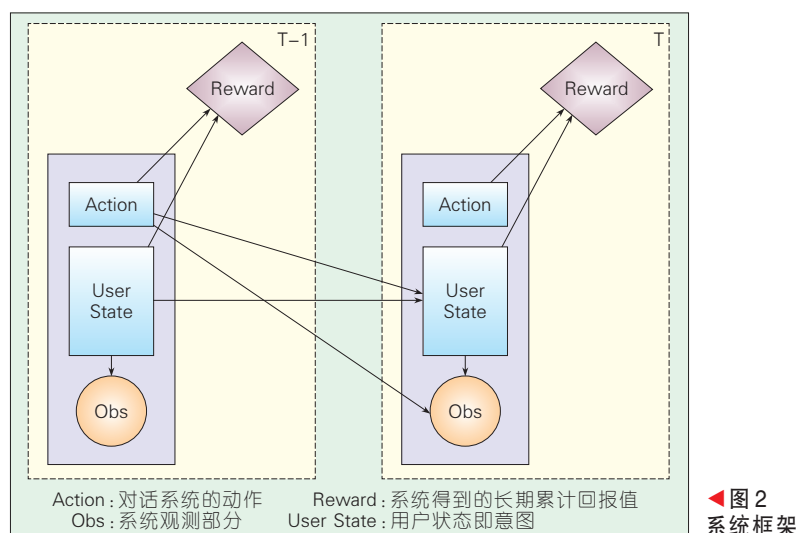
(4) 状态转移函数

系统是根据当前的信念状态得出动作,从而会影响用户的下一步的状态,由于把用户状态定义为用户意图,即会影响用户意图。不同的动作选择会对用户意图的改变有着不同的影响。状态转移函数用于记录当前意图下,执行动作后,下一个出现某意图的可能性。

(5) 观察函数

一般情况下观测函数是很难直接从训练语料直接统计出来,也是 POMDP 模型最复杂,最有挑战性的部分。不同的状态、动作会导致不同的观察出现。观测函数的性能也是直接影响信念状态的更新,从而会影响对话系统的决策。为了方便计算,该系统认为系统的观测只跟当前的意图相关,与前一个动作无关,即 $p(ols_u', a) = p(ols_u')$ 。

(6) 回报函数

图2
系统框架

回报函数是用于描述在当前意图的情况下执行某动作后得到的回报。回报函数的定义虽然比较简单,但是却不能很好的定义回报函数。目前一般都是按照系统的实际用途人工定义回报函数。系统根据用户意图采取正确的动作就可以得到正的回报,执行错误的动作就会得到负的回。比如,如果当前用户意图是 Teach 类别的话,系统采取 Rlearn 动作,就会得到+200 的回报,采取 Rquery 类别的动作就会得到+150 回报,否则会得到-200 回报;如果当前意图是 End, 用户采取的动作是 Rend, 代表对话结束,可以给出+1000 的回报。

(7) Belief 初始值

如果没有对话的开始状态,一般情况下是每个状态的 Belief 值定义为一,效果比较好。

根据这 7 部分的定义,搭建一个基于 POMDP 模型的对话系统,这个模型总共有 12 种状态,9 种动作,12 种观测,折扣因子为 0.9,因为引入开始/结束状态,初始状态的信念值只有 Start 的值为 1,其他状态的值为 0。

为了加快 POMDP 模型的速度,训练时没有采用精确求解算法,而是采用 SARSOP 近似算法。

为了更全面的考察上文提出的修正观测概率方法在基于 POMDP 的

对话系统的效果,设计了两个实验。

3.2 模拟环境实验

意图识别模块使用条件随机场 (CRF) 和支持向量机 (SVM) 两种分类器来估计观测概率,同时为了体现加入修正方法后,系统的健壮性,在估计状态转移时同时使用 4 种不同的平滑方法(不使用平滑 (Origin),最大熵,最大似然,Good-turing) 估计状态转移概率。模拟次数是 1 000 次。

表 1 的意图识别分类器都是使用 CRF,而表 1 的第二列实验的 POMDP 模型中的观测概率是没有经过修正的,第三列的观测概率是经过修正的。从表 1 可以看出,经过修正后的系统回报值比没有经过修正的系统回报值要大,最高提高了 2.5%,最低提高了 2%。

表 2 的意图识别分类器都是使用 SVM,而第二列实验的 POMDP 模型中的观测概率是没有经过修正的,第三列的观测概率是经过修正的。从表 2

可以看出,经过修正后的系统回报值比没有经过修正的系统回报值要大,最高提高了 35%,最低提高了 3.4%。

主要结论是:修正带来了性能提高,且不论是 CRF 还是 SVM,都能提高。主要原因是观测函数是反映环境的不确定性与意图识别的错误信息的一个渠道,因此加入信息熵的概率来修正观测概率在理论上也有所支持。

3.3 真实对话实验

上一组实验是直接经过 POMDP 模型仿真模拟得出的结果,下面这一组实验是放在真实对话中测试经过修正观测概率后给对话系统带来的影响。

表 3、表 4 中的每个实验都包含 3 个测试样本集,第一个样本集的对话轮数是 212,第二个样本集的对话轮数是 87,第三个样本集的对话轮数是 100。分别对比表 3 和表 4,可知对观测概率经过修正后,在真实对话中,该对话系统的决策准确度比没有使用修正的观测矩阵的系统是要高的,对系统的决策性能有所提高。

通过实验一与实验二两组实验可知,使用 CRF 分类进行意图识别得到的期望回报整体上会比使用 SVM 方法的要大,主要是因为本文使用的基于 SVM 分类器的意图识别模块的识别错误率 p_{err} 为 0.17,而基于 CRF 分类的意图识别错误率 p_{err} 为 0.05,也就是说意图识别错误率的大小会直接影响到模型的性能大小。

使用 SVM 作为意图识别模块的对话系统经过修改观测概率方法后,系统决策性能的提升幅度比使用

表 1 意图识别的分类器使用 CRF

平滑方法	修正前的期望回报	修正后的期望回报
Origin	1 798.09	1 843.83
最大似然	5 119.82	5 248.29
Good-turing	5 121.78	5 239.79
最大熵	5 148.59	5 255.27
CRF: 条件随机场		

▼表 2 意图识别的分类器使用 SVM

平滑方法	修正前的期望回报	修正后的期望回报
Origin	-786.89	-505.78
最大似然	3 283.7	3 398.11
Good-turing	3 255.95	3 398.26
最大熵	3 288.62	3 419.15
SVM: 支持向量机		

▼表 3 意图识别的分类器使用 CRF

平滑方法	样本集编号	对话数	对话轮数	修正前决策对的对话轮数个数	修正后决策对的对话轮数个数
Origin	1	53	212	91	91
最大似然	1	53	212	107	112
Good-turing	1	53	212	108	112
最大熵	1	53	212	108	112
Origin	2	16	87	27	27
最大似然	2	16	87	81	81
Good-turing	2	16	87	81	81
最大熵	2	16	87	81	81
Origin	3	16	100	32	32
最大似然	3	16	100	84	85
Good-turing	3	16	100	84	85
最大熵	3	16	100	84	86
CRF: 条件随机场					

▼表 4 意图识别的分类器使用 SVM

平滑方法	样本集编号	对话数	对话轮数	修正前决策对的对话轮数个数	修正后决策对的对话轮数个数
Origin	1	53	212	92	92
最大似然	1	53	212	171	172
Good-turing	1	53	212	171	172
最大熵	1	53	212	176	177
Origin	2	16	87	12	12
最大似然	2	16	87	29	33
Good-turing	2	16	87	29	32
最大熵	2	16	87	31	33
Origin	3	16	100	17	17
最大似然	3	16	100	29	41
Good-turing	3	16	100	29	41
最大熵	3	16	100	29	41
SVM: 支持向量机					

CRF 的要大,主要是因为本文使用的基于 SVM 分类器的意图识别模块的识别错误率比基于 CRF 分类的意图识别错误率要大,也就是说 CRF 的意图识别率已经很高,修正前与修正后

的概率变化不大,而 SVM 的意图识别错误率比 CRF 的大,修正的观测概率给对话系统带来比较大的影响。如果从信息熵的角度来看的话,经过计算,使用 CRF 的意图识别的系统得到

的观测平均信息熵是 0.3168,而是用 SVM 的意图识别的系统得到的观测平均信息熵是 1.0557。平均信息熵越高,不确定性越大,越需要修正观测矩阵,可提升的性能也越大。平均信息熵越低,代表观测比较可靠,越不需修正,甚至可以直接相信观测,同样可提升的性能也不大。

因此在意图识别错误率比较高的情况是有必要对观测概率进行合理的修正,以达到提高对话系统的决策性能的效果。

3.4 实验总结

意图识别模块的性能很明显对模型的期望回报有着很大的影响,模型的期望回报是反应该模型的性能重要指标,系统是通过意图识别的结果对具有不确定性的环境的进行估计,根据不完全可观测的结果,进行决策,执行动作,得到回报奖赏,进而与环境交互。因此意图识别的性能和对环境的估计准确程度与对话系统的性能有着密切关系。通过实验与分析可知,在意图识别错误率一定的情况下,在观测函数估计中引入对环境不确定性的信息,通过信息熵描述环境的不确定性,提高系统对环境的估计与判断,使系统在不完全可观测的情况下,提高决策的准确度与模型的期望回报。

4 结论

在本文中,把描述环境不确定性的信息熵加入到观测函数里,修正观测概率。以一个简单的机器人教学系统为例子实现具体的 POMDP 模型,并通过两个实验考察观测概率经过修正后与修正前的性能差异,验证使用修正后的观测概率的系统得到较好的效果。

作为使用 POMDP 模型来搭建对话系统这个方向的初步工作,本文的结果说明对话的决策性能受很多方面影响。可以通过很多方面提高对话决策性能,回报函数反映执行某动

作得到的立即回报,也是与环境交互的结果之一,因此今后可以把回报函数的信息加入到观测函数里面,增强观测函数对环境的描述。

参考文献

- [1] BUI T H, ZWIERS J, POEL M, NIJHOLT A. Toward affective dialogue modeling using partially observable Markov decision processes [C]//Proceedings of the Workshop Emotion and Computing, 29th Annual German Conference on Artificial Intelligence, Bremen Germany, 2006
- [2] BADLER N, STEEDMAN M, ACHORN B, BECHET T, DOUVILLE B, PREVOST S, CASSELL J, PELACHAUD C, STONE M. Animated conversation: Rule-based generation of facial expression gesture and spoken intonation for multiple conversation agents [C]//Proceedings of the SIGGRAPH, 1994:73-80
- [3] HUANG F, YANG J, WAIBEL A. Dialogue management for multimodal user registration [C]//Proceedings of the International Conference on Spoken Language Processing (ICSLP'2000), Beijing, China, 2000:37-40
- [4] GODDEAU D, MENG H, POLIFRONI J, SENEFF S. A form-based dialogue manager for spoken language application [C]//Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP'96,Pittsburgh,PA). ICSLP, 1996: 701-704
- [5] MICHAEL F M. Spoken Dialogue Technology: Enabling the Conversational User Interface [R]. ACM Computing Surveys, 2002
- [6] LARSSON S, TRAUM D. Information state and dialogue management in the TRINDI dialogue move engine toolkit [J]. Natural Language Engineering, 2000,5(3-4):323-340
- [7] WILLIAMS J D, YOUNG S J. Partially observable Markov decision processes for spoken dialog systems [J]. Computer Speech and Language, 2007,21(2):393-422
- [8] ESTHER L, ROBERTO P, WIELAND E. A stochastic model of human-machine interaction for learning dialog strategies [J]. IEEE Transactions on speech and audio processing, 2000, 8(1): 11-23
- [9] WILLIAMS J D, POUPART P, YOUNG S J. Factored partially observable Markov decision processes for dialogue management [C]//Proceedings of the Workshop on Knowledge and Reasoning in Practical Dialog Systems, Int. Joint Conf. on Artificial Intelligence (IJCAI), Edinburgh, 2005
- [10] ZHANG B, CAI Q, MAO J, CHANG E, GUO B. Spoken Dialogue Management as Planning and Acting under Uncertainty [M]. Eurospeech Denmark, 2001
- [11] 张波, 蔡庆生, 郭百宁. 口语对话系统的 POMDP 模型及求解 [J]. 计算机研究与发展, 2002,39(2):22-27
- [12] TPAEK E. Uncertainty utility and misunderstanding: A decision-theoretic perspective on grounding in conversational systems [C]//Proceedings of the Fall Symp on Psychological Models of Communication in Collaborative Systems. North Falmouth, Massachusetts, USA, 1999
- [13] WILLIAMS J D, YOUNG S J. Scaling POMDPs for spoken dialog management [J]. IEEE Audio, Speech and Language Processing 2007,15 (7):2116-2129
- [14] WILLIAMS J D, POUPART P, YOUNG S J. Partially observable Markov decision processes with continuous observations for dialogue management [C]//Proceedings of the SigDial Workshop on Discourse and Dialogue, Lisbon, 2005

作者简介



钟可立, 北京邮电大学在读硕士研究生;研究方向为自然语言处理,研究兴趣集中在汉语切分与对话管理模型;曾在国际 SIGHAN 汉语切分技术评测中获第 1 名。



王小捷, 北京邮电大学教授、博导, 中国人工智能学会常务理事, 中国人工智能学会自然语言理解专委会主任;主要研究方向为自然语言处理与认知计算。

综合信息

光纤入户占比近半 提速降费按下“快进键”

工信部最新发布的通信业经济运行情况显示,截至 2015 年 8 月,光纤接入用户近 9 800 万户,8 Mbit/s 及以上用户占比接近 60%。2015 年 1—8 月,3 家基础电信企业互联网宽带接入用户净增 867.2 万户,总数达 2.09 亿户。“宽带中国”战略的加速推进,宽带提速效果日益显著,8 Mbit/s 及以上接入速率的宽带用户总数达 1.22 亿户,占比达 58.4%,比 2014 年末增加 17.5 个百分点;20 Mbit/s 及以上宽带用户总数占比达 23.1%,比 2014 年年末增加 12.7 个百分点。光纤宽带建设进度加快,光纤接入用户比上年末净增 2 965.7 万户,总数达 9 797.2 万户,占比达 46.8%。

“宽带中国”战略加速推进、光纤入户占比近半的数据背后,表明在全球通信行业普遍陷入滞涨的背景下,中国运营商的提速降费事实上还是获得了显著的推进。

中国电信给出的数据显示,截至 2015 年 7 月底,中

国电信移动网络及有线宽带投资达 450 亿元,其中新建光纤到户家庭为 6 600 万户,已提前完成年底覆盖 6 500 万户的目标;南方超过 80% 城市区域已实现光网覆盖;83% 的铜缆小区已完成光纤化改造;已建成 4G 基站 35 万套,预计年底建成 50 万套。

中国移动方面,截至目前,中国移动 4G 用户数已经突破 2 亿户,4G 基站数也超过 70 万个,其 TD-LTE-A 下行三载波聚合的理论峰值能达到 330 M。中国移动 4G 资费还将继续改进,目前正在与 70 多家企业开展合作,发展定向流量的后向收费模式。

中国联通方面,依据其在全光网络工作推进会上公布的数据,截至 2015 年 7 月底,中国联通在全国 39 个市建成了全光纤宽带网络覆盖,光纤到户覆盖家庭达到 1.18 亿户。目前全国 20 M 单产品价格综合降幅已达 36%,50 M 单产品综合降幅达 38%;受惠用户规模 1 601.6 万用户,预计全年惠及 5 500 万用户。

(转载自《中国信息产业网》)

机会调度模式下的嵌套码研究

Nested Codes in Opportunistic Scheduling Model

中图分类号: TN929.11 文献标志码: A 文章编号: 1009-6868 (2015) 05-0056-004

摘要: 将机会调度模式引入嵌套码体系, 通过各个收发节点与中继节点的信道状况自适应调整嵌套码的码速率, 可以增加信道状况好的链路传输的数据量, 提升无线网络的整体吞吐量。仿真结果表明, 相比传统非机会调度嵌套码而言, 机会调度嵌套码能够有效提升网络吞吐量和误比特性能。

关键词: 中继系统; 嵌套码; 机会调度; 吞吐量; 误比特率

Abstract: This paper introduces the concept of opportunistic scheduling in a nested coding system. In this scheme, adaptive rate nested codes are used according to channel conditions in order to increase the transmission load of the link while maintaining a good channel condition. In this way, network capacity is no longer constrained by the worst link, and the overall throughput of wireless networks is increased. Simulation shows that compared to traditional non-opportunistic scheduling nested coding, opportunistic scheduling nested coding increases throughput and reliability of single relay multicast relay networks.

Key words: relay system; nested codes; opportunistic scheduling; throughput; bit error rate

胡留军/HU Liujun
许进/XU Jin

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

网络的容量和可靠性。

1 网络编码系统模型

文章提出的方法主要针对典型组播单中继系统。系统由一个中继站和多个收发节点构成。各个收发节点既是信源节点也是目的节点。即各个收发节点都要通过中继将自己的信息发送给所有其他收发节点, 同时也接收所有其他收发节点发来的信息。典型的4收发节点组播单中继系统如图1所示^[7]。

在该系统中采用的网络编码传输信息的过程如下: 4个收发节点(A, B, C, D)将各自的信息(i_A, i_B, i_C, i_D)发送至中继节点r; 中继节点将收到的部分或全部信息进行网络编码; 中继节点r将网络编码后的信息广播

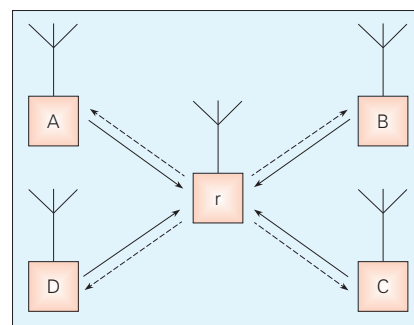
当前无线网络系统正经历对系统容量的需求提升及网络密集化的过程^[1]。网络编码因能够采用较少的链路获取较大的网络容量而被广泛关注^[2-3]。在传统网络编码中, 网络节点需要预知其他节点的边信息, 然而在实际情况下, 边信息往往很难获取^[4]得到。

针对边信息获取难的问题, L. Xiao等将卷积码与网络编码相结合, 提出采用嵌套码的补偿方法来解决没有边信息情况下的网络编解码问题^[5]。因此, 我们可以认为嵌套码可以进一步提高未来网络编码投入到实际应用的可行性。

在现有嵌套码体系下, 全网中的

所有链路采用相同的码速率。然而, 在无线通信网络场景下, 不同链路所受信道衰落各不相同。为了保证质量最差的链路具有可接受的可靠性, 全网所有链路都需要采用较低的传输速率。在这种情况下, 无线网络的容量瓶颈是通信质量最差的无线链路, 网络中信道质量好的无线链路得不到充分利用。

针对以上问题, 文章将机会调度机制^[6]引入嵌套码体系。在典型无线组播单中继网络架构下, 基于信道状态信息, 通过动态选择参与网络编码的信源节点, 提升整个网络的容量。此外, 针对无线信道的时变特性, 文章还提出了在机会调度网络编码体系中引入可变码率的删余卷积码(RCPC)^[6]来动态适配无线链路的信道质量, 从而进一步提升整个无线网



▲ 图1 网络编码系统模型

收稿日期: 2015-05-25
网络出版时间: 2015-07-21

至各收发节点;各收发节点将从中继节点收到的信号解码,获取从其他收发节点传来的信息^[8]。

2 机会调度与嵌套码的原理

2.1 嵌套码

嵌套码将卷积码与网络编码技术相结合,每个卷积码的编码矩阵相互线性独立,再将各个卷积码编码进行网络编码形成嵌套码。嵌套码的优势在于接收端仅通过嵌套码译码就可以获取组成该码的各个源节点的数据信息。

本系统中源节点通过相互线性独立的卷积码生成器产生编码数据包并发送至中继节点,中继节点进而对其中一部分接收信息进行嵌套编码^[9-10],数学表达式如下:

$$c_{\text{nested}} = i_1 G_1 \oplus i_2 G_2 \oplus \dots \oplus i_{|s|} G_{|s|} \\ = [i_1, i_2, \dots, i_{|s|}] [G_1, G_2, \dots, G_{|s|}]^T \quad (1)$$

其中, $G_1, G_2, \dots, G_{|s|}$ 表示相互线性独立的生成矩阵, $i_1, i_2, \dots, i_{|s|}$ 表示源节点数据包, \oplus 表示异或操作。

嵌套码的设计依据以下准则:各节点的编码向量相互线性独立;合并后的嵌套码的码率小于1;避免使用恶性卷积码^[11]。基于各节点的编码向量相互独立,可以降低嵌套码译码时对边信息的依赖,若码率大于1或者若源节点使用了恶性卷积码,将导致卷积码性能损失。

2.2 机会调度与RCPC码

网络模型中全部链路均为相互独立的衰落信道,在 μ 时刻,源节点到中继节点的瞬时信噪比(SNR)可写为:

$$\gamma_{sr}(\mu) = \frac{P|h_{sr}(\mu)|^2}{\sigma_{sr}^2} \quad (2)$$

其中, P 为发射功率, $h_{sr}(\mu)$ 为信道系数, σ_{sr}^2 则为高斯白噪声的噪声方差。

中继节点将源到中继节点的瞬

时SNR设置为一组升序数组。 r_k 表示升序数组中第 k 个瞬时SNR, N_d 表示目的节点的总数,共有 $N_i = N_d - k + 1$ 个目的节点参与机会调度, k 为调度等级。 r_k 为 N_i 个目的节点与中继节点的链路上的最小SNR。

参与机会调度的信源节点至中继节点的每单位带宽下的系统容量定义^[7]:

$$C_{\text{inst}}^k = N_i \log_2(1 + r_k) \quad (3)$$

式(3)中,当 N_i 逐渐增大时, r_k 的取值逐渐减小,因此可以通过选取合适的 k 值,使得式(3)表示的系统容量最大。

中继节点对调度到的节点进行异或操作, S_{os} 表示参与机会调度的源节点的集合。因此,中继节点进行嵌套编码的表达式如下:

$$c_{\text{nested}}^{OS} = \bigoplus_{j \in S_{os}} i_j G_j \\ = [i_1, i_2, \dots, i_{|j|}, \dots, i_{|s|}] [G_1, G_2, \dots, G_{|j|}, \dots, G_{|s|}]^T \quad (4)$$

为了提高嵌套码的码率,中继节点采用RCPC编码。例如可采用码率1/3,约束长度为3的最佳距离谱(OVS)码作为母码。

2.3 译码算法

目的节点获取的软信息可以表示为:

$$\tilde{\tau} = c_{\text{nested}-os} + e \quad (5)$$

其中 e 表示噪声。进一步,以在目的节点 d 处为例, $\tilde{\tau}$ 可表示为:

$$\tilde{\tau} = \bigoplus_{l \in k_d} i_l G_l \oplus \bigoplus_{l \notin k_d} i_l G_l + e \quad (6)$$

k_d 表示目的节点 d 的已知先验信息的节点的索引, $\bigoplus_{l \in k_d} i_l G_l$ 表示目的节点已获取的信息。通过以上集合可推导表达式如下:

$$\tilde{\tau}_d = \tilde{\tau} \oplus \bigoplus_{l \in k_d} i_l G_l = \bigoplus_{l \notin k_d} i_l G_l + e \quad (7)$$

其中,

$$\bigoplus_{l \notin k_d} i_l G_l + e \quad (8)$$

表示目的节点 d 尚未获知的信

息。为了从未知信息中获得目的信息,嵌套码通过各行向量相互线性独立的编码矩阵 $[G_1, G_2, \dots, G_j, \dots, G_{|s|}]^T$ 从 $\bigoplus_{l \notin k_d} i_l G_l + e$ 中获取目标源节点信息 i_l 。

2.4 编码可达速率

文献[8]分析了有、无调度机制下的系统可达速率:

$$R = \left(\frac{T}{t} \right) \cdot W \cdot I(s, r, d) \quad (9)$$

R 表示系统的可达速率, T 表示每个时隙的时间长度, t 表示所有时隙的总的时长, W 表示可用带宽, $I(s, r, d)$ 表示源节点 s 到目的节点 d 的互信息。

首先,各源节点使用相同的传输时隙 T 。如图2(a)所示,当不采用机会调度时,由于要兼顾链路条件最差的那个节点,中继节点传输网络编码包需要用较低的码率,例如在图2(a)中用了3个时隙(3T)。在图2(b),由于采用了机会调度的机制,中继节点可根据不同的机会调度等级选择不同的时隙传输网络编码包。

由图2可以推导出有或者无机会调度的情况下系统传输网络编码包所需要的时隙数:

$$t_{\text{withoutOS}} = |S| T_{\text{withoutOS}} + 3 T_{\text{withoutOS}} \quad (10)$$

$$t_{\text{withOS}} = |S| T_{\text{withOS}} + X \cdot T_{\text{withOS}} \\ (x = 3/2, 2, 5/2 \text{ or } 3) \quad (11)$$

整理可得:

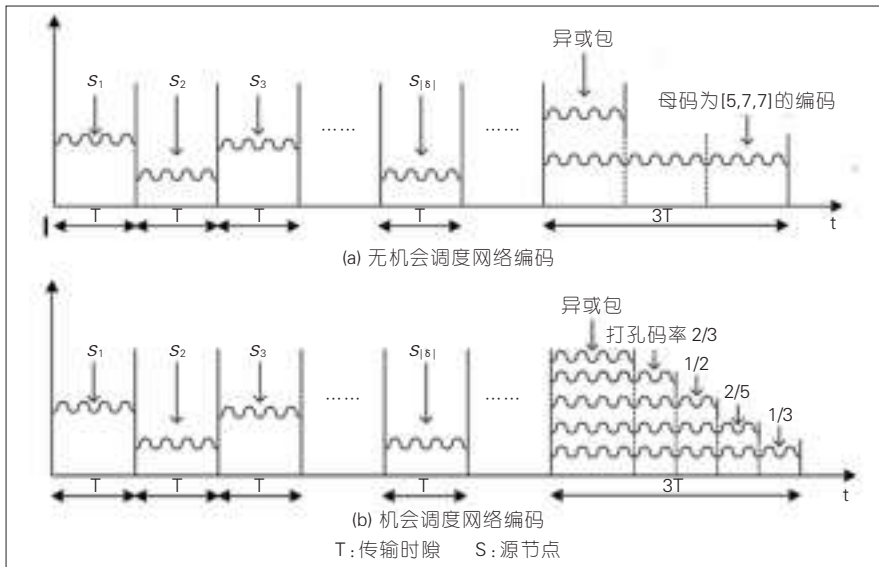
$$T_{\text{withoutOS}} = \frac{t_{\text{withoutOS}}}{|S| + 3} \quad (12)$$

$$T_{\text{withOS}} = \frac{t_{\text{withOS}}}{|S| + X} \\ (x = 3/2, 2, 5/2 \text{ or } 3) \quad (13)$$

其次,通过中继转发后源节点与目的节点之间的互信息 $I(s, r, d)$ 可以表示为^[9]:

$$I(s, r, d) = \log_2 \left(1 + r_{sd} + \frac{r_s r_{sd}}{\frac{\sigma_{s'}^2}{s_{s'}} + r_{sd} + \frac{\sigma_{s'}^2}{\sigma_d^2} \sum_{s_j \in |s|} |r_{s_j}|} \right) \quad (14)$$

其中, s_e 表示采用机会调度后源



▲图2 机会调度

节点的集合； $r_{sd} = |h_{sd}|^2 p_s / \sigma_d^2$ ， $r_{sr} = |h_{sr}|^2 p_s / \sigma_r^2$ 和 $\sigma_{z_d}^2$ 分别表示接收节点d的网络编码噪声方差。在文章所述的系统模型中，有 $|s|_{\max} = |s|$ ， $\sigma_{z_d}^2 = \sigma_d^2$ ，其中 $|s|_{\max}$ 表示机会调度机制下最大的源节点集合中包含的节点数。在有机会调度和没有机会调度的情况下，源节点与目的节点之间的互信息分别表示为：

$$I_{\text{withoutOS}} = \log_2 \left(1 + r_{sd} + \frac{r_{sr} r_{rd}^1}{|s| + r_{rd}^1 + \sum_{j=1}^{|s|} r_{sjr}} \right) \quad (15)$$

$$I_{\text{withOS}} = \log_2 \left(1 + r_{sd} + \frac{r_{sr} r_{rd}^2}{|s_e| + r_{rd}^2 + \sum_{s_j \in s_e} r_{sjr}} \right) \quad (16)$$

r_{rd}^2 与 r_{rd}^1 分别表示有或者无机会调度下目的节点d的SNR。通过式(12)~(16)和式(10)换算出系统的可达速率：

$$R_{\text{withoutOS}} = \frac{W}{|s|+3} \times \log_2 \left(1 + r_{sd} + \frac{r_{sr} r_{rd}^1}{|s| + r_{rd}^1 + \sum_{j=1}^{|s|} r_{sjr}} \right) \quad (17)$$

$$R_{\text{withOS}} = \frac{W}{|s|+x} \times \log_2 \left(1 + r_{sd} + \frac{r_{sr} r_{rd}^2}{|s| + r_{rd}^2 + \sum_{s_j \in s_e} r_{sjr}} \right) \quad (18)$$

(0 < $s_e \leq s$, $x = 3/2, 2, 5/2$ or 3)

考察在相同码率的条件下，有、无机会调度机制对系统可达速率的影响，假设 $x=3$ ， $r_{rd}^2 = r_{rd}^1$ 。因为 $|s| \geq |s_e|$ ，有 $\sum_{j=1}^{|s|} r_{sjr} \geq \sum_{s_j \in s_e} r_{sjr}$ 。因此

$$R_{\text{withoutOS}}^{P_{ap}} \leq R_{\text{withOS}}^{P_{ap}} \quad (19)$$

即采用机会调度的可达速率高于不采用机会调度时的可达速率。当可以自适应调整网络编码的码率时，自适应打孔可达速率为：

$$P_{ap} = \arg \max_p R^p \quad (20)$$

且：

$$R_{\text{withoutOS}}^{P_{ap}} \leq R_{\text{withOS}}^{P_{ap}} \leq R_{\text{withOS}}^{op} \quad (21)$$

由此可见，机会调度模式能够有效提升网络的可达速率并提高网络容量。通过在机会调度嵌套码架构下提高打孔码率，能够进一步提升网络的可达速率和网络容量^[6]。

3 仿真分析

假设所有的节点有相同的发射功率且所有链路均为瑞利独立慢衰落信道。基于文献[10]的结果，文章采用的编码矩阵为如式(22)。选取任意目的节点并假定节点获取全部边信息、部分边信息和无边信息。假设 i_k 为节点C想要获取的目标信息，那么节点C获取全部边信息就表示节点C侦听到了相邻节点(B,D)的信息(i_B, i_D)；节点C获取部分边信息则表示节点C侦听到(B,D)其中之

一的信息，无边信息则表示节点C仅有自己发射的信息 i_C 。

$$\begin{bmatrix} G_A \\ G_B \\ G_C \\ G_D \end{bmatrix} = \begin{bmatrix} 5 & 6 & 5 & 6 & 7 & 4 \\ 7 & 0 & 7 & 3 & 6 & 2 \\ 4 & 5 & 2 & 6 & 5 & 0 \\ 6 & 1 & 5 & 7 & 2 & 5 \end{bmatrix} \quad (22)$$

下面分两种情况分别分析文章中机会调度嵌套码的性能。在第1种情况下，假设中继节点已经获取所有收发节点的先验信息，并利用先验信息进行解码。在第2种情况下，中继节点仅具有收发节点的先验信息。在传输的过程中，中继节点采用维特比解码算法进行解码^[11]。在这两种情况下，机会调度与非机会调度嵌套码的误比特率分别如图3和图4中所示。

由图3可见，当中继节点已知要发送至目的节点的所有信息时，目的节点获取到全部边信息时误码率(BER)性能最优，没有获取任何边信息时BER性能最差。以上情况由节点获取先验信息的多少以及不同编码矩阵的性能差异所致，如果目的节点获取其他节点的先验信息较少，则目的节点的译码性能较差。

由图4可见，当中继节点仅已知要发送至目的节点的部分信息时，目的节点无边信息时的性能反而略优于目的节点获取全部边信息时的性能。这可以从差错传播的角度来解释，虽然更多的边信息意味着译码端可以利用译码信息更多，但是由于在中继节点不知道目的节点的全部先验信息时，过多的边信息也可能引入过多的差错。不完整的先验信息引入的差错会对最终的译码性能造成一定的影响。从仿真结果上可以看到边信息数量越多反而性能越差。

基于以上分析，有、无机会调度下嵌套码的瞬时最大容量分别为：

$$C_{\text{inst}}^{\text{withOS}} = (N_d - k + 1) \log_2(1 + r_k) \quad (23)$$

$$C_{\text{inst}}^{\text{withoutOS}} = (N_d) \log_2(1 + r_{\min}) \quad (24)$$

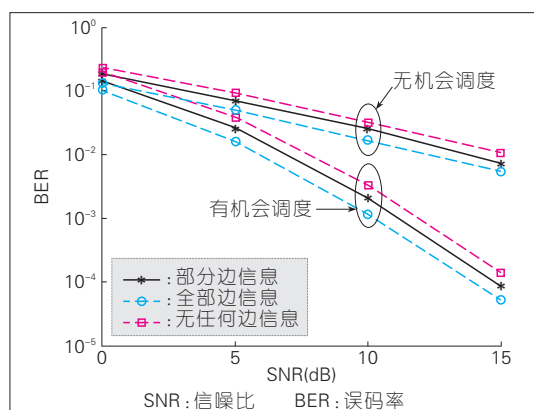
此时， r_{\min} 表示接收瞬时SNR中的最小值。

有、无机会调度下网络编码的平均最大容量的表达式:

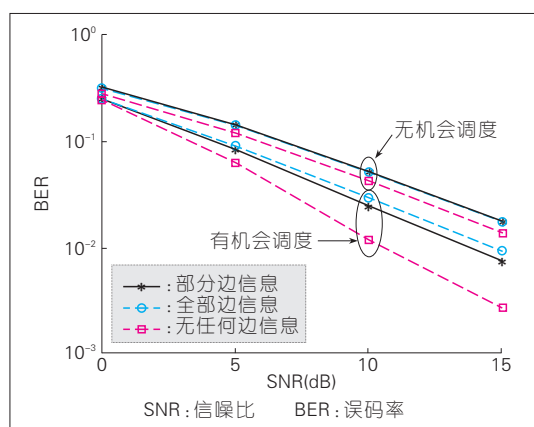
$$C_{Ave}^{withOS} = (N_d - k + 1) \int_0^\infty \log_2(1 + r_k) p(r_k) d_{rk} \quad (25)$$

$$C_{Ave}^{withoutOS} = (N_d) \int_0^\infty \log_2(1 + r_{min}) p(r_{min}) d_{rmin} \quad (26)$$

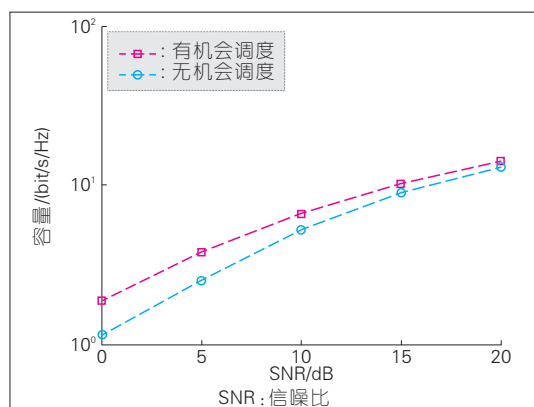
其中, $p(\cdot)$ 为概率密度函数。



▲图3 中继节点获取全部收发节点先验信息的机会调度网络编码误比特率



▲图4 中继节点获取部分收发节点先验信息的机会调度网络编码误比特率



▲图5 机会调度对嵌套码容量的影响

在有、无机会调度的情况下,嵌套码的容量分别如图5所示。由图5可见,机会调度机制对应的最大容量优于非机会调度机制,尤其在低信噪比区域。因为机会调度机制以系统容量最大为优化目标,会选择向链路质量较好的用户传输网络编码包。随着链路SNR的逐渐提高,两种传输机制的容量曲线相接近,由于链路SNR的提高使得所有链路均可传输网络编码包,两种机制的传输策略基本相同。因此性能差异也逐渐缩小。

4 结束语

文章将机会调度引入嵌套码,通过提升高信噪比无线链路的传输负荷,充分利用网络中的无线链路,提升多播单中继无线网络的网络容量。仿真结果表明,机会调度方法能够有效提升嵌套码的容量和误码性能。在中继节点获取全部收发节点先验信息时,并且在BER为 10^{-2} 时采用机会调度机制能够获得6 dB的增益,在中继节点获取部分收发节点先验信息时能够获得近4.5 dB的性能增益。当SNR为0~20 dB时,采用机会调度相比非机会调度机制的网络容量最大能有66%的提升。

综上所述,机会调度机制与嵌套码相结合的传输方案相比非机会调度机制,无论从BER性能还是网络容量方面都存在显著提升。

参考文献

- [1] PENG M., YANG C, ZHAO Z, and WANG W. Cooperative Network Coding in Relay-Based IMT-Advanced System [J]. IEEE Communication Magazine, 2012, 50(4): 76-84. doi: 10.1109/MCOM.2012.6178837
- [2] LIN H, LIN Y, and KANG H. Adaptive Network Coding for Broadband Wireless Access

Network [J]. IEEE Trans. Parallel and Distributed System, 2013, 244(1): 4-18. doi: 10.1109/TPDS.2012.101

- [3] AMERIMEHR M H and ASHTIANI F. Delay and Throughput Analysis of a Two-Way Opportunistic Network Coding-Based Relay Network [J]. IEEE Trans. Wireless Communication, 2014, 13(5): 2863-2873. doi: 10.1109/TWC.2014.040914.121461
- [4] XIAO L, FUJA T, KLIEWER J, and COSTELLO D. Nested Codes with Multiple Interpretations[C]//40th Annual Conference on Information Sciences and Systems, Waikoloa, HI, USA, 2006: 851-856
- [5] YOMO H and POPOVSKI P. Opportunistic Scheduling for Wireless Network Coding[C]// Proceeding of IEEE International Conference on Communications '07, Jun. 2007: 5610-5615
- [6] FREMGER P, ORTEN P., OTTOSSON T, and SVENSSON A. Multi-rate Convolutional Codes[M]. United Kingdom: Chalmers University of Technology Press, 1998
- [7] LIN Z and VUCETIC B. Power and Rate Adaptation for Wireless Network Coding with Opportunistic Scheduling [C]// The IEEE International Symposium on Information Theory, Toronto, Canada, 2008.
- [8] SHARMA S, SHI Y, LIU J, and HOU Y T. Is Network Coding Always Good for Cooperative Communications [C]// Proc. IEEE INFOCOM, San Diego, USA, 2010
- [9] SHARMA S, SHI Y, HOU Y T, SHERALI H. D, and KOMPELLA S. Optimizing Network-coded Cooperative Communications via Joint Session Grouping and Relay Node Selection[C]// Proc. IEEE INFOCOM, Shanghai, China, 2011
- [10] CHANG J, WANG D. and LIN M. Some Extended Results on the Search for Good Convolutional Codes [J]. IEEE Trans. Inform. Theory, 1997, 43(5):1672-1697. doi: 10.1109/18.623175
- [11] FORNEY G D. The Viterbi Algorithm [J]. Proceedings of the IEEE, 1973, 61(3):268-278

作者简介



胡留军,中兴通讯股份有限公司高级工程师、无线预研部部长;主要研究方向为移动通信网络及其关键技术;广东省专利金奖获得者,多项国家及省市重大项目的负责人;已发表论文、专利80余篇。



许进,中兴通讯股份有限公司高级工程师;主要研究方向为新型调制编码、网络编码等无线新技术;在国内外学术刊物及会议上发表论文10余篇。

基于大数据的业务并发度分析

Service Concurrency Analysis Based on Big Data

中图分类号: TN929.1 文献标志码: A 文章编号: 1009-6868 (2015) 05-0060-05

摘要: 指出不同业务之间的关系对于网络优化具有很重要的意义。使用大数据的分析方法处理蜂窝网络的实测数据,可以得到各种业务的并发关系,同时将并发关系通过业务关系网络的方式展现出来,具有很好的可视性。通过业务关系网络可以直接看出业务并发情况,为蜂窝网络的优化扩容等方面提供理论指导。

关键词: 大数据; 蜂窝网络; 业务并发度; 网络优化

Abstract: The relationship between different services is important for network optimization. In this paper, we process the real data in cellular networks with a method used in big data to attain a concurrent relationship between the various services. Then the service concurrency is presented by the way of a relationship network with good visibility. Service concurrency can be seen directly through the relationship network, providing theoretical guidance for cellular network optimization and expansion.

Key words: big data; cellular networks; service concurrency; network optimization

易正磊/YI Zhenglei¹
顾军/GU Jun²
张兴/ZHANG Xing¹

(1. 北京邮电大学无线信号处理与网络实验室, 北京 100876;
2. 中兴通讯股份有限公司, 上海 201203)
(1. Wireless Signal Processing and Network Laboratory, Beijing University of Posts and Telecommunications, Beijing 100876, China;
2. ZTE Corporation, Shanghai 201203, China)

1 大数据的意义和影响

1.1 大数据的价值和挑战

大数据是由于规模、复杂性、实时而导致的使之无法在一定时间内用常规软件工具对其进行获取、存储、搜索、分享、分析、可视化的数据集合。由于大数据的上述特点,如何将数据进行合理应用是目前大数据领域的核心目标^[1]。

大数据对于整个社会来说蕴含着巨大的潜在价值,大数据的价值并不在于数据本身,而在于如何将数据的作用反馈于社会决策。维克托·迈尔-舍恩伯格在《大数据时代》中指出,人类从依靠自身判断做决定到依

靠数据做决定的转变,是大数据做出的最大贡献之一。因此,能否正确利用大数据的内在规律,是决策成功或者失败的关键因素^[2]。

一般来讲,数据的生命周期包括数据采集、数据归纳、数据重构、数据挖掘、数据预测、数据可视化等6个方面,大数据亦是如此。但是由于大数据的体积庞大、结构复杂,常规的处理方法并不能挖掘出数据的内在价值,这也正是大数据时代人们面临的最大挑战。

1.2 大数据对移动互联网的影响

在移动互联网的环境中,大数据的特点并不仅仅体现在数据量的巨大,更体现在数据的实时性与关联性,这些特点让数据搜集变得容易,对数据的分析可以反过来快速影响基础网络,这也正是移动互联网的自

身特点。在大数据与移动互联网高度融合的背景下,最重要的目的就是发现和挖掘真正有价值的数据,使我们能够根据这些信息精确地指导每一次的网络规划和优化,而这些有价值的信息往往来源于对用户大量网络行为数据的抽象和分析。

如图1所示,大数据应用于移动互联网的基本目的包括:负载优化、用户划分、趋势预测等。

2 蜂窝网络中的大数据应用

2.1 蜂窝网络的发展现状

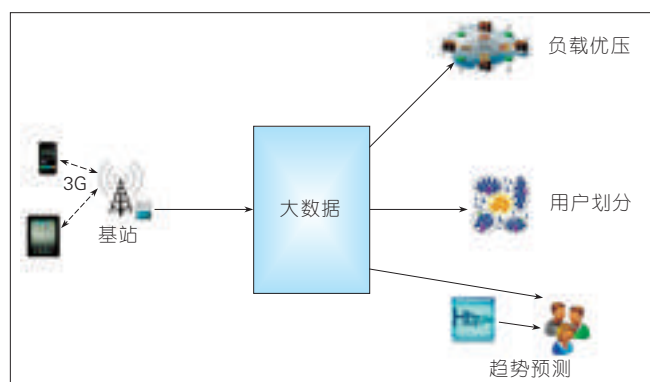
蜂窝网络是一种最常见的移动通信网络结构,通过把移动设备的服务区分为很多正六边形的子区域,并在每个子区域设置一个基站,形成了形状似“蜂窝”的结构,因而把这种移动通信网络称为蜂窝网络。

作为支撑信息传输的关键基础设施,蜂窝网络的性能优劣对于整个通信系统的效率起着至关重要的作用。近年来,为了应对移动网络流量爆炸性增长,欧洲、美洲、日韩等地区

基金项目: 国家自然科学基金(61372114); 国家重点基础研究发展(“973”)规划(2012CB316005)

收稿日期: 2015-08-17

网络出版时间: 2015-09-21



已经大规模建设长期演进(LTE)网络,中国也迎来了LTE网络建设高峰。在这一建设进程中,除了基于IP多媒体子系统(IMS)的VoLTE作为最终的语言解决方案之外,Small Cell和无线局域网(WLAN)将成为提升网络容量的重要手段,在超宽带移动网络部署^[3-4]中发挥越来越重要的作用。此外,行业应用正从窄带向宽带演进,除了语音通信外,数据、视频传输需求逐步增加。LTE的100 Mbit/s高速数据传送能力,可以更好地服务于政务网、公共安全和应急救援等行业。因此,利用新一代无线技术来实现行业应用,已成为一个发展趋势^[5]。

随着分布式计算和云平台的逐步实现,作为云服务的关键应用之一,大数据应用具备了大规模发展的条件^[6]。运营商的大数据业务正从分散孤立系统向统一的标准化平台方向发展、从数据的采集、存储向检索和挖掘的方向发展。在网络大数据的收集与预处理方面,分布式存储方案、内存数据库技术将成为主流技术;Hadoop和一体机成为主流的数据分析平台。

根据以往的研究,蜂窝网络的承载情况随着通信协议的发展也发生着巨大变化。根据某一地区的真实历史数据分析可以得知,全市所有基站的平均吞吐量在2G、3G和4G网络的时代是有很大的区别的,具体数值如图2所示。

根据图2可以看出,发展速度最快并且承载压力最大的信道是下行

数据信道,而目前国内的4G网络还处于发展初期,吞吐量的增长了已达到将近100%。由此可知,数据业务的有效传输是蜂窝网络的核心难题。

此外,随着移动互联网、物联网的发展以及4G技术的逐渐普及,无线通信网络的各种功能需求也日益扩大,用户的网络行为也随之发生改变,所产生的数据体积也达到了新的数量级。因此,如何利用有限的蜂窝网络资源,应对大数据时代的挑战,是当今通信领域的一个重要课题。

2.2 有效的网络优化

当今的蜂窝网络的发展速度明显慢于基站吞吐量的增长,而用户日益丰富的行为对蜂窝网络的承载能力提出了更高的要求。可见,优化网络,提升网络负载能力已经成为一个亟待解决的问题。

我们已经提到,大数据时代的到来使得人们的思维方式发生了巨大的变革,数据驱动了决策制定,因此对蜂窝网络的优化策略正是通过对

蜂窝网络中产生的数据分析处理来制订。同时,蜂窝网络系统本身就是一个巨大的数据仓库,我们可以从中采集到丰富的数据,通过对这些数据的分析,可以识别用户的地理位置,洞察客户接触不同信息的渠道,了解用户的各种网络行为。常用的分析方法包括:预测业务流量、探寻不同业务之间的关联、探寻不同业务模型下的资源瓶颈等等。

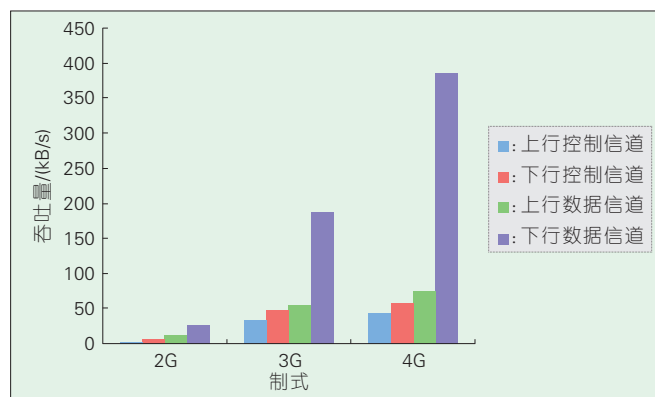
接下来以业务并发度探寻实例来阐述在蜂窝网络系统中如何应用大数据来解决问题。在该实例中,我们分析各种常见业务在不同场景不同时间的并发度,最终得出不同区域的业务并发以及对网络资源的消耗情况,并据此提出网络优化策略。

3 大数据时代的业务并发度分析

3.1 应用场景

随着通信技术的飞速发展和移动终端的功能逐渐丰富,蜂窝网络中的数据业务流量迅猛增长,移动互联网下终端用户需求更加多样化和复杂化,这也促使移动互联网由传统的单业务向着多业务平台发展。多业务的出现使得网络数据更趋向于复杂多样与结构各异,这给用户的行为分析带来了很大的困难和挑战。与此同时,由于业务种类的繁多,分析某一种业务对于整体蜂窝网络影响甚至微乎其微,因此为了提高网络的承载能力,需要科学准确地分析各数

图2
不同制式下基站平均吞吐量



据业务之间的并发性。

由图3可见,由于目前蜂窝网络用户数和终端数快速增长,以及业务场景多种多样,数据种类也趋于全面,包括用户使用各业务的时间信息、位置信息、链接次数、业务量大小等,因此,用户——业务网络资源之间的映射十分复杂,海量的数据具有极低的价值密度,如果仅仅对某一部分数据进行分析,不能反映整体网络情况,所具有的价值意义也就很小。

此外,无线侧采集到的数据格式并不统一,应用传统数据库无法建立统一的数据表结构。根据数据的以上特点,我们采用Hadoop平台进行数据的预处理与所有算法的执行,实现了高效处理非结构化数据。

3.2 核心技术

为了分析不同场景下的海量数据,首先平台应该具有较高的数据处理能力。以无线侧的呼叫细节记录(CDR)数据为例,某地区全市全天CDR数据条数达到千亿数量级,数据体积约为70 GB,连续一个月的数据量将达到2 TB,如果空间维度扩展到全国,时间维度扩展到几年,这样的数据量是传统数据库和单机环境的处理能力远远不能达到的,而以Hadoop为代表的大数据处理平台则可以轻松应对^[7-8]。

其次,平台能够处理并管理非结构化数据。非结构化数据相对于结构化数据而言,不方便用数据库二维逻辑表来表现的数据即称为非结构

化数据,包括所有格式的办公文档、文本、图片、标准通用标记语言下的子集XML、HTML、各类报表、图像和音频、视频信息等等。这些数据需要特殊的筛选方法进行预处理和归一化,之后才可以应用到各类数据挖掘算法中去。

再次,算法需要分析复杂场景并对场景进行区分。如果将全市或者全国所有的数据不加以区分而进行统一处理,那么所有的特殊区域都将被平均化,然而某些区域的特殊情况的内在原因正是数据分析者们更感兴趣的地方,因此将数据进行合理的场景划分,是数据预处理的中重要一步。

最后,平台必须能够分布式结构并行处理,提高海量数据的处理速度。由于大数据更新速度快的特点,能否快速实时的对海量数据进行处理是整个数据分析的基础。

3.3 处理流程

为了探究未来一段时间某地区的业务并发关系,我们需要对该地区采集到的海量数据做预处理。原始数据包括了许多字段,如基站信息(经纬度)、网络类型、上下行流量、数据包数、业务持续时长等。我们主要研究各业务之间的联系,因此只需提取出与业务量有关的字段。首先,从所有数据中剔除了许多不常用的或是数据不全的业务,最终得到了60种有效业务,这些业务对象包括了除语音、短信外所有数据类业务,如

即时消息、社交、流媒体、邮件等;然后再从这60种需要分析的业务中选取可能会使用到的各业务的上下行流量、用户数、数据包个数等;最终我们选取了各业务流量来探究业务并发关系。我们将相关的数据整理为表1格式。

现在各种网络业务越来越多,但是各种业务之间并不是孤立存在的,用户使用习惯、业务本身属性等都会使得各业务是息息相关的。为了衡量业务之间关系的大小,我们定义了各业务之间的距离。

对于采集到的 N 种业务,要得到第 i ($i=1,2,3,\dots,N$)种业务与第 j ($j=1,2,3,\dots,N, j \neq i$)种业务之间的距离,首先需要计算出第 i 种业务与第 j 种业务的皮尔逊相关系数,计算公式为:

$$\rho_{ij} = \frac{\text{Cov}(x^{(i)}, x^{(j)})}{\sqrt{\text{Var}(x^{(i)})} \sqrt{\text{Var}(x^{(j)})}} \quad (1)$$

其中 $x^{(i)}, x^{(j)}$ 分别为第 i, j 两种业务流量的时间序列。在得到业务相关系数的基础上,我们就可以计算第 i 种业务与第 j 种业务之间的距离 d_{ij} ^[9],计算公式为:

$$d_{ij} = \sqrt{2(1-\rho_{ij})} \quad (2)$$

为了直观地看出各业务之间的并发关系,我们使用kruskal算法构建最小生成树网络^[10-11],对于由 N 种业务之间的 C_N^2 个距离构建的集合 U ,首先找出 U 中的最小值,即距离最小的两种业务,在这两种业务之间添加一条无向边,连接这两种业务,接着在剩下的 $C_N^2 - 1$ 条边中继续寻找最小值连接业务,同时保证业务之间不连成环,直至遍历所有距离值。其流程如图4所示。

3.4 结果展示

图5所示为某使用地区,从2014年1月5日开始连续15天全网60种业务构建的最小生成树业务网络。

在该业务网络中,各节点代表了不同业务,如快播、优酷、微信、淘

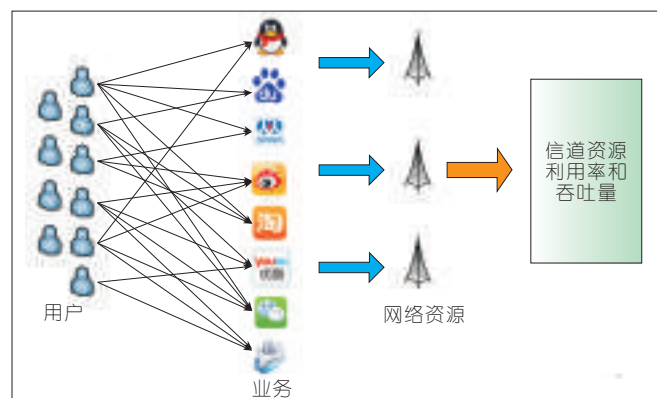
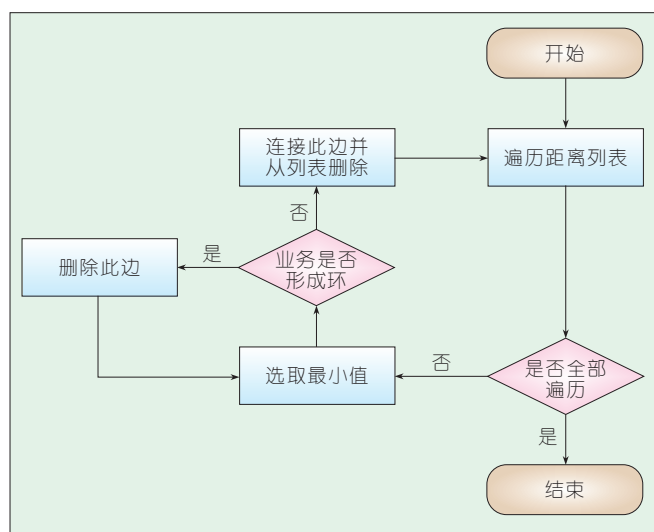


图3
用户—业务—网络资源之间的映射

▼表 1 业务流量数据示例

时间	JABBER 的流量/M	GTALK 的流量/M	56VIDEO 的流量/M	优酷的流量/M	土豆的流量/M	风行的流量/M	阿里旺旺的流量/M
2014/1/5 8:00	13 22	0	2 858	714 084	32 236	27 502	595
2014/1/5 9:00	23 740	194	577 924	163 836	615 363	70 646	2 997
2014/1/5 10:00	20 451	0	529 369	174 232	335 228	90 245	3 293
2014/1/5 11:00	19 979	0	155 743	137 776	56 886	11 620	3 085
2014/1/5 12:00	17 595	0	151 262	953 784	118 481	98 255	3 155
2014/1/5 13:00	16 151	0	210 168	829 077	189 708	36 301	2 594
2014/1/5 14:00	17 448	175	198 973	679 424	93 845	84 362	2 685
2014/1/5 15:00	22 507	0	69 295	546 109	40 653	60 004	2 778
2014/1/5 16:00	31 648	0	130 530	780 224	108 563	66 752	3 788
2014/1/5 17:00	31 085	0	151 783	716 352	75 607	58 511	4 587
2014/1/5 18:00	27 859	1 067	96 042	195 300	44 656	8 781	6 342
2014/1/5 19:00	33 841	710	32 882	139 539	15 262	13 070	6 676
2014/1/5 20:00	26 495	475	98 747	115 989	30 487	15 968	5 618
2014/1/5 21:00	23 522	965	89 497	103 930	46 340	28 453	4 424
2014/1/5 22:00	31 415	200	189 714	105 917	97 643	2 898	5 971

▲图 4
业务网络构建流程

▲图 5 最小生成树业务网络

宝、新浪微博等。节点大小是由该业务在网络中介数的大小决定的，节点越大代表了该业务介数越大。而对于这样一个由业务构成的网络来说，业务介数越大，其他的业务就越容易通过它关联在一起，例如在图 5 中，淘宝与优酷这两种业务并未直接连接在一起，这说明它们之间的相关性不是最大的，但是它们还是可以通过一定的路径连接起来，从图中具体来看就是：淘宝——IMessage——快播——优酷，而且从我们构建该网络的方法来看，这样连接起来的路径一定是相关系数权重最大的。对于其他业务也可以此类推，可以看出任意两种业务相关的路径通过 IMESSAGE 的次数是最多的。

该网络中的边选取了不同颜色来标注，不同颜色代表不同的业务距离范围。在阐述该网络的构建方法时我们就已经说明，边连接的节点是两种距离最小的业务，因此从结合点与边即可看出各业务之间的关系：距离越小就说明两种业务越容易并发。从图 5 中可以发现任意一种业务与其他业务的并发情况。

对于不同地区或不同时间段的业务数据，我们都可以构建出最小生成树业务网络，从该网络中找到各业务之间的关联关系，通过业务关联关系可以预测出任意一种业务在未来一段时间与其他业务的并发情况。例如采用某地区一定时间内的数据可以得到图 5 所示的业务网络，从中可以预测出，在出现优酷这种业务时，很可能会同时出现快播、乐视网（乐视）、56VIDEO、PPLIVE 这 4 类相关的业务。

该业务关系网络图是根据历史数据得出。为了预测未来一段时间的业务并发关系，需根据预测需要，不断使用新数据来更新业务关系网络图，从而保证预测的准确性。

得到不同业务之间的并发关系后，我们就可以结合不同业务对网络资源的消耗情况对网络的调控与优

化提供一定的理论指导。例如,如果某地区的业务呈现出图5所示并发情况,则可以知道IMessage业务会与多种业务并发,那么在做网络调控时,需要优先满足该业务消耗的信道资源。同时对于容易并发的业务,在做优化时可以当做同一类业务来处理,因为它们会同时消耗不同的网络资源。

4 结束语

提出了一种大数据背景下基于业务并发度来分析用户网络行为的方法,该方法分析所得到的结果可以对网络规划和优化进行理论指导。我们需要进一步分析不同种类的业务对于蜂窝网络资源消耗的映射关系,从而精确预测整体网络的负载情况,并据此提出更准确、更全面的网络优化指导。

参考文献

[1] Howe D, Costanzo M, Fey P, et al. Big data:

The Future of Bio Curation [J]. Nature, 2008, 455(7209): 47-50

[2] Viktor Mayer Schonberg. Big Data: A Revolution That Will Transform How We Live, Work, and Think [M]. UK: Hodder & Stoughton, 2013

[3] Hwang I, Song B, Soliman S S. A Holistic View on Hyper-Dense Heterogeneous and Small Cell Networks [J]. Communications Magazine, IEEE, 2013, 51(6), pp.20-27

[4] Hoadley J, Maveddat P. Enabling Small Cell Deployment with HetNet [J]. Wireless Communications, IEEE, 2012, 19(2): 4-5. doi: 10.1109/MWVC.2012.6189405

[5] LTE: the UMTS Long Term Evolution [M]. New York: John Wiley & Sons, 2009

[6] 张建勋, 古志民, 郑超. 云计算研究进展综述 [J]. 计算机应用研究, 2010, 27(2): 429-433

[7] 吴吉义, 平玲娣, 潘雪增. 云计算: 从概念到平台 [J]. 电信科学, 2009 (12): 23-30

[8] Zikopoulos P, Eaton C. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data [M]. McGraw-Hill Osborne Media, 2011

[9] Zheng Z, Yamasaki K, Tenenbaum J N, et al. Carbon-Dioxide Emissions Trading and Hierarchical Structure in Worldwide Finance and Commodities markets [J]. Physical Review E, 2013, 87(1): 012814

[10] Gallager R G, Humblet P A, Spira P M. A Distributed Algorithm for Minimum-Weight Spanning Trees [J]. ACM Transactions on Programming Languages and Systems (TOPLAS), 1983, 5(1): 66-77

[11] Cheriton D, Tarjan R E. Finding Minimum Spanning Trees [J]. SIAM Journal on

Computing, 1976, 5(4): 724-742

作者简介



易正磊, 北京邮电大学无线信号处理与网络实验室在读研究生; 目前主要研究领域为大数据与社会网络分析。



顾军, 中兴通讯 FDD LTE 无线网络规划总工; 长期从事 3/4G 无线组网方案研究与应用; 申请国家专利 10 项, 国际专利 2 项, 发表技术论文 5 篇。



张兴, 北京邮电大学副教授、博士生导师; 主要研究领域为异构无线网络融合、物联网与蜂窝网络融合等; 近期出版专著 3 本, 发表 SCI 检索论文 7 篇, 其他国内期刊会议论文等 30 余篇。

综合信息

Gartner 调查报告称: 信息安全管理实践正日趋完善

全球领先的信息技术研究和顾问公司 Gartner 近日表示, 越来越多除 IT 以外的其他部门也在设置安全保障, 这一趋势反映了有效安全管理的必要性。

根据 Gartner 公司有关终端用户隐私、IT 风险管理、信息安全、业务连续性和合规管理的年度调查结果显示: 信息安全管理实践正在日趋成熟。2015 年 2—4 月, Gartner 在全球 7 个国家开展调查, 964 名在大型机构工作的受访者参与了调查, 这些大型机构在 2014 财政年度的总收入不低于 5 000 万美元, 拥有不少于 100 名员工。

Gartner 的副总裁兼院士级分析师 Tom Scholtz 表示: “人们对数字业务的风险意识日益增强, 再加上有关网络安全事件的深度宣传, 促使 IT 风险列为董事会级别的讨论事宜, 71% 的受访者表示, IT 风险管理数据影响董事会层面的决策。这也反映出, 越来越多的企业开始重视把应对 IT 风险作为公司管理的一部分。”

Scholtz 指出, 在 IT 以外的部门建立汇报程序的主

要原因在于促进执行与监督相分离, 提升企业信息安全的形象, 打破员工和利益相关者认为“安全只是 IT 问题”的思维定势。企业日益认识到, 必须将安全性作为企业风险问题进行管理, 而不仅仅是一个 IT 运营问题。

支持安全项目的高管级别也在提高。63% 的受访者表示, 他们的信息安全项目获得的资助和支持来自 IT 以外部门的领导层, 这一调查数据相较于 2014 年的 54% 有了大幅上升。企业首席执行官和董事会的支持率保持不变, 为 30% (2014 年是 29%), 而指导委员会的支持率从 7% 升至 12%。地区差异引人注目, 57% 的北美受访者表示信息安全项目获得的支持来自 IT 以外部门, 明显低于西欧的 63% 和亚太地区的 67%。

Scholtz 认为, 企业高管重视安全项目至关重要, 否则安全项目将难以得到企业其他部门的必要支持。在安全策略的有效性方面, 虽然有一半的受访者表示管理层参与评估和审批这些策略, 但仅 30% 的受访者表示, 业务部门会积极参与到这些会影响其业务的策略制订中来。

(转载自《中国信息产业网》)

2016年第1—6期专题征文

第1期

网络空间安全

专题策划人

专题策划人:北京邮电大学 杨义先 教授
南京邮电大学 杨庚 教授

互联网的发展势头较为强劲,在人们日常生活中也扮演着越来越重要角色。没有网络安全就没有国家安全。2014年中央网络安全和信息化领导小组正式成立,2015年6月,国务院学位委员会、教育部正式决定在“工学”门类下增设“网络空间安全”一级学科。本专题将就上述问题展开相关讨论,希望能为关心网络空间安全发展的专家、学者们提供一个交流与沟通的平台。

专题征文范围包括:

- 网络空间安全学科的发展规划与布局
- 网络空间安全人才培养体系与实践
- 网络空间安全体系与关键技术
- 面向数据的安全体系结构研究
- 大数据安全关键技术研究
- 云计算安全模型与关键技术

第2期

大数据分析处理与应用

专题策划人

清华大学 郑纬民 教授

大数据顺应“互联网+时代”国家治理、产业经济、政府决策、社会运行等模式变革而诞生,已经成为国家经济社会发展和全面信息化的重要战略部署。大数据分析处理与应用需要在理论和实践两个方面寻求新方法、新技术和新变革。本专题将就此展开讨论,希望能为中国大数据产业技术的发展和广泛应用做出贡献。

专题征文范围包括:

- 大数据系统结构
- 大数据存储管理技术
- 大数据系统优化技术
- 大数据处理、挖掘与分析方法
- 金融与经济大数据、社会治理大数据、智慧城市大数据、媒体大数据处理等大数据处理分析应用

第3期

5G技术与业务创新

专题策划人

清华大学 王京 教授
中兴通讯无线首席科学家 向际鹰 博士

ITU已经正式将5G命名为IMT-2020,并发布了IMT-2020愿景。ITU通过雷达图的方式全面体现了对5G技术指标的需求,其中最引人关注的是发布了将应用场景与技术指标相结合的雷达图。这就意味着5G的技术创新将与业务创新紧密联系在一起,它们将相互促进、共同发展。以往的技术发展相对独立,对业务的抽象非常简单,从很大程度上限制了移动通信网络的可持续发展。本专题的目的就是期望面向未来各种应用场景研发新业务,引出对未来5G技术的需求,从而探索5G技术的创新思路。

专题征文范围包括:

- 面向各种物联网需求的新业务模式
- 面向工业互联网的新业务模式
- 面向智能终端的新业务模式
- 满足多种业务需求的开放式网络架构
- 满足多种业务需求的开放式空口传输技术
- 满足特定业务需求的网络切片技术与定制化网络技术
- 未来移动通信系统的运营模式和产业生态环境
- 新业务模式的质量评估体系与方法

2016年第1—6期专题计划

期次	专题名称	专题策划人
1	网络空间安全	北京邮电大学 杨义先 教授 南京邮电大学 杨庚 教授
2	大数据分析处理与利用	清华大学 郑纬民 教授
3	5G技术演进与业务创新	清华大学 王京 教授 中兴通讯无线首席科学家 向际鹰 博士
4	天地一体化信息网络	中国工程院 张乃通 院士 哈尔滨工业大学 顾学迈 教授
5	工业互联网与智慧工厂技术	中国工程院 邬贺铨 院士 湖南大学 王耀南 教授
6	SDN/NFV的实践与规模应用	中国信息通信研究院科技委 蒋林涛 主任

2016年第1—6期专题征文

第4期

天地一体化信息网络

专题策划人

中国工程院 张乃通 院士
哈尔滨工业大学 顾学迈 教授

天地一体化信息网络是陆地通信网的扩展,是人类活动拓展至空间、远海乃至深空的重大信息基础设施,也是当今全球科技和产业热点。天地一体化信息网络是地面网络、天基网络、天地一体化通信协议以及其共同支撑的应用集成,可以实现真正意义上的全球无缝覆盖,并可以帮助人们及时、准确和全面地了解地球空间。为及时报道广大科技工作者在天地一体化信息网络方面取得的研究成果,推动中国天地一体化信息网络理论和技术的研究,特策划本期专题。本专题主要针对天地一体化信息网络中的卫星移动通信、空间互联网、星地混合网络及深空通信等研究领域征集论文,欢迎广大学者、专家、研究生、工程技术人员积极投稿。

专题征文范围包括:

- 天地一体化信息网络发展趋势及关键技术
- 天地一体化信息网络传输体制与接口标准
- 天地一体化信息网络协议体系
- 天地一体化信息网络资源管理与安全机制
- 天地一体化信息网络的星间链路
- 空间互联网相关技术及应用
- 移动通信卫星轨道与星座方案设计及优化
- 移动通信卫星平台及有效载荷技术
- 升空平台组网与传输技术

第5期

工业互联网与智慧工厂技术

专题策划人

中国工程院 邬贺铨 院士
湖南大学 王耀南 教授

工业互联网作为计算进程和物理进程的统一体,能够集成计算、通信与控制于一体。工业互联网通过人机交互接口实现和物理进程的交互,并可以使用网络化空间以远程、可靠、实时、安全、协作的方式操控一个物理实体。智慧工厂的核心技术是连接,需要把设备、生产线、工厂、供应商、产品、客户紧密地连接在一起,使得产品与生产设备之间、不同的生产设备之间以及数字世界和物理世界之间能够互联,并使得机器、工作部件、系统间通过网络持续地保持数字信息的交流。以工业互

联网、工业大数据、智能机器人、工业软件等为核心的智慧工厂关键技术正在加速推动新一轮工业革命的到来!本专题将就上述问题展开讨论,希望能为关心工业互联网与智慧工厂的专家学者、企业家等提供一个交流与沟通的平台。

专题征文范围包括:

- (1) 工业互联网
 - 工业互联网的系统构成
 - 工业互联网的物理构成、工业互联网的实现架构
 - 工业互联网与嵌入式系统、工业互联网与工业控制系统
 - 工业互联网与无线传感网络
 - 工业互联网与控制系统、工业互联网与物联网
- (2) 智慧工厂协同关键技术
 - 基于物联网技术的智能工厂
 - 机器人化智能设备
 - 工业大数据挖掘与分析
 - 工业互联网信息安全
- (3) 智慧工厂技术
 - 智慧工厂的车间
 - 基于工业互联网的智能工厂
 - 智慧工厂的增强现实技术
 - 智慧工厂的机器人技术
 - 智慧工厂的智能装配

第6期

SDN/NFV的实践与规模应用

专题策划人

中国信息通信研究院科技委 蒋林涛 主任

SDN/NFV是目前电信界的研究热点,SDN由美国提出的,而NFV是由欧洲主导。目前,业界对SDN/NFV有着不同的定义,看法也并不一致。本专题从多个角度研究和分析SDN/NFV技术,给出3~4个SDN/NFV的实践和规模应用实例,让读者对SDN/NFV技术有正确和深入的了解。欢迎电信设备制造商、电信运营商、大专院校、科研机构中的专业技术人员能够积极投稿!

专题征文的范围包括:

- SDN/NFV技术综述
- SDN/NFV的标准化工作
- SDN/NFV关键技术的分析和实现
- SDN/NFV在云化数据中心的实践和规模应用
- SDN/NFV与移动通信网、SDN/NFV与固定通信网的实践和规模应用
- SDN/NFV的问题与未来展望

《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 124 期
2015 年 10 月 第 21 卷第 5 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
常务副总编: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 10 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