

编辑委员会

顾问：邬贺铨 林金桐

主任委员：韦乐平

副主任委员：张顺颐 张英海 王晓丹

编委：(按姓氏笔画为序)

王思伟 石文孝 卢捍华

兰巨龙 毕军 曲桦

刘云 纪越峰 李小文

李丹 李安民 肖志辉

吴巍 余少华 沈连丰

张云勇 张同须 张宏莉

陈山枝 陈德人 林春庭

易东山 周涛 赵季红

赵慧玲 段晓东 唐雄燕

卿斯汉 陶成 黄玉兰

琚春华 彭木根 蒋林涛

韩江洪 童晓渝 谢显中

蔡康 廖建新 滕勇

电信科学

DIANXIN KEXUE

(月刊,1956年创刊)

2017年(第33卷)第11期

主管单位 中国科学技术协会

主办单位 中国通信学会

人民邮电出版社

出版单位 北京信通传媒有限责任公司

总经理 梁海滨

副总经理 易东山 康荣 王建军

编辑单位 《电信科学》编辑部

北京市丰台区成寿寺路11号

邮电出版大厦8层(100078)

主编 韦乐平

编辑部主任 吴娜达

传真 010-81055494

电话 010-81055443/5459/5467(编辑部)

010-81055476(广告部)

投稿网址 www.telecomsci.com

正文排版 《电信科学》杂志排版室

印刷 北京时捷印刷有限公司

国内发行 北京报刊发行局

订购处 全国各地邮局

邮发代号 2-397

国外发行 中国国际图书贸易集团有限公司

国外代号 M841

刊号 ISSN 1000-0801
CN 11-2103/TN

广告经营许可证 京东工商广字第8032号

出版日期 2017年11月20日

定价 48.00元

读者热线:010-81055459/5476

010-81055598(发行部)

E-mail: dxkx@ptpress.com.cn



《电信科学》微信订阅号

C 目录

视点聚焦

10/40/100 Gbit/s PON 技术的发展和展望

张成良,王 庆,余景文,万 芬 [2017300](1)

研究与开发

异步卫星协同通信系统中的双采样差分空时编码方案 章坚武,屠贺嘉琦 [2017309](9)

3D MIMO-OFDMA 系统中基于垂直波束成形的能效优化算法

李 汀,仇林杰,季 薇 [2017266](17)

基于深度学习的软件定义网络应用策略冲突检测方法

李传煌,程 成,袁小雍,岑利杰,王伟明 [2017305](27)

高速铁路宽带无线通信系统越区切换技术 刘云毅,赵军辉,王传云 [2017278](37)

基于用户体验质量和系统能耗的异构网络联合接入选择和功率分配策略

季石宇,唐良瑞,李淑贤,杜施默 [2017251](47)

基于元胞蝙蝠算法的无线传感器网络节点定位研究

孟凯露,岳克强,尚俊娜 [2017311](56)

水声通信系统中基于迭代自适应的脉冲噪声抑制方法

周桂莉,李有明,余明宸,王晓丽 [2017240](66)

基于主题模型的垃圾邮件过滤系统的设计与实现 寇晓淮,程 华 [2017313](73)

基于图转换的无线接入点关联优化算法

陶 雯,沈 艳,管昕洁,万夕里 [2017310](83)

基于比例因子转移概率的 AAC 音频压缩历史检测算法

黄其娟,王让定,严迪群 [2017277](93)

电信科学

第 33 卷第 11 期 2017 年 11 月

基于 SVD 和 SAE 的医学图像多功能零水印算法

赵彦霞,孙洁丽,周增慧,邢 胜 [2017254](102)

综述

5G 关键技术与标准综述

王庆扬,谢沛荣,熊尚坤,魏 焱,刘 昱,李文苾,吴锦莲 [2017312](112)

物联网技术专栏

工业通信网络现有架构的梳理总结和未来运营商的发展策略

常 洁,王 艺,李 洁,陈正文 [2017301](123)

工业互联网中的标识解析技术

杨 震,张 东,李 洁,张建雄 [2017296](134)

物联网使能平台的体系结构与服务模式

彭 昭 [2017302](141)

面向智能制造的工业连接现状及关键技术分析

李 洁,张 东,常 洁,杨 震 [2017295](146)

对象存储系统的自动化管理技术

胡豆豆,罗 飞,张 园 [2017306](154)

运营技术广角

云网协同时代运营商 IP 承载网发展

朱永庆,黄晓莹,张文强 [2017256](162)

上行 SCMA 信道估计及性能评估

胡显安,曾祥念 [2017267](169)

暂态社区感知的 ICWN 数据转发机制

杜 剑,夏元轶,赵俊峰,王 峥,王 鹤 [2017262](179)

NFV 故障关联及故障自愈方案研究

毛斌宏,阳志明 [2017261](186)

小基站部署方案规划算法

刘 旻 [2017253](195)

Telecommunications Science

November 2017 (Vol.33 No.11)

Development and forecast of 10/40/100 Gbit/s PON technology	ZHANG Chengliang, WANG Qing, YU Jingwen, WAN Fen	[2017300](1)
Dual-sample differential space-time coding scheme in asynchronous satellite cooperative communication system	ZHANG Jianwu, TU Hejiaqi	[2017309](9)
Energy-efficient optimization algorithm based on vertical beamforming for 3D MIMO-OFDMA system	LI Ting, QIU Linjie, JI Wei	[2017266](17)
Policy conflict detection in software defined network by using deep learning	LI Chuanhuang, CHENG Cheng, YUAN Xiaoyong, CEN Lijie, WANG Weiming	[2017305](27)
Handover technology in high-speed railway broadband wireless communication system	LIU Yunyi, ZHAO Junhui, WANG Chuanyun	[2017278](37)
Joint user association and power control scheme based on energy consumption and QoE in heterogeneous network	JI Shiyu, TANG Liangrui, LI Shuxian, DU Shimo	[2017251](47)
Wireless sensor network nodes localization method based on cellular automata bat algorithm	MENG Kailu, YUE Keqiang, SHANG Junna	[2017311](56)
Impulsive noise mitigation based on iteration adaptive approach in underwater acoustic communication	ZHOU Guili, LI Youming, YU Mingchen, WANG Xiaoli	[2017240](66)
Design and implementation of spam filtering system based on topic model	KOU Xiaohuai, CHENG Hua	[2017313](73)
An association optimization algorithm for wireless access points based on graph transformation	TAO Wen, SHEN Yan, GUAN Xinjie, WAN Xili	[2017310](83)
AAC compression detection based on scaling factor transition probability	HUANG Qijuan, WANG Rangding, YAN Diqun	[2017277](93)
Multipurpose zero watermarking algorithm for medical image based on SVD and SAE	SUN Jieli, ZHAO Yanxia, ZHOU Zenghui, XING Sheng	[2017254](102)
Key technology and standardization progress for 5G ...	WANG Qingyang, XIE Peirong, XIONG Shangkun, WEI Yao, LIU Yu, LI Wenyi, WU Jinlian	[2017312](112)
Summary of existing framework in industrial communication networks and future development strategies for communication operators	CHANG Jie, WANG Yi, LI Jie, CHEN Zhengwen	[2017301](123)
Identifier technology in industrial internet	YANG Zhen, ZHANG Dong, LI Jie, ZHANG Jianxiong	[2017296](134)
Architecture and service model of IoT enabled platform	PENG Zhao	[2017302](141)
Situation description and critical technology analysis for industrial connection of intelligent manufacturing	LI Jie, ZHANG Dong, CHANG Jie, YANG Zhen	[2017295](146)
Automatic management technology of object storage systems	HU Doudou, LUO Fei, ZHANG Yuan	[2017306](154)
Development of carrier's IP bearer network in cloud-network coordinated era	ZHU Yongqing, HUANG Xiaoying, ZHANG Wenqiang	[2017256](162)
Channel estimation and performance analysis for uplink SCMA system	HU Xian'an, ZENG Xiangnian	[2017267](169)
Transient community aware data forwarding mechanism for intermittent connected wireless network	DU Jian, XIA Yuanyi, ZHAO Junfeng, WANG Zheng, WANG He	[2017262](179)
Research on NFV fault association and fault self-healing	MAO Binhong, YANG Zhiming	[2017261](186)
Planning algorithm of small cell deployment scheme	LIU Yang	[2017253](195)

Chief Editor: WEI Leping

Director: WU Nada

Sponsor: China Institute of Communications

Posts & Telecom Press

Editor: *Telecommunications Science* Magazine Office (F8, You Dian Publisher Building, No.11

Chengshousi Road, Fengtai District, Beijing 100078, China)

General Distribution Office: Beijing Newspapers and Periodicals Distribution Office

Overseas Distributor: China International Book Trading Corporation (P.O.Box 399, Beijing, China)



10/40/100 Gbit/s PON 技术的发展和展望

张成良¹, 王庆², 余景文¹, 万芬²

(1. 中国电信股份有限公司北京研究院, 北京 102209;

2. 湖北邮电规划设计有限公司, 湖北 武汉 430024)

摘要: 当前, EPON/GPON 向更高速率的升级演进已成为光接入领域颇受关注的议题。基于此, 对 XG(S)-PON、TWDM PON、WDM PON、100 Gbit/s EPON 的现状与面临的问题进行了分析, 并对各种技术的应用现状与前景进行了展望。

关键词: XG(S)-PON; TWDM PON; WDM PON; 100 Gbit/s EPON

中图分类号: TN929.1

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017300

Development and forecast of 10/40/100 Gbit/s PON technology

ZHANG Chengliang¹, WANG Qing², YU Jingwen¹, WAN Fen²

1. Beijing Research Institute of China Telecom Co., Ltd., Beijing 102209, China

2. China Comservie Hubei PTPD Co., Ltd., Wuhan 430024, China

Abstract: Currently, the evolution of EPON/GPON to higher bitrate technology has become a hot topic in optical access field. Based on this, the current situation and problems of XG(S)-PON, TWDM PON, WDM PON and 100 Gbit/s EPON were analyzed, then some forecast of above technologies was given.

Key words: XG(S)-PON, TWDM PON, WDM PON, 100 Gbit/s EPON

1 引言

众所周知, 提高单波长的承载速率和挖掘单根光纤的可用带宽是光传输领域研发与创新不变的主题, 同样地, 在光接入领域这一规律同样适用。经过近几年国内三大运营商基于 EPON/GPON 技术大力建设 FTTH, 中国已有上亿家庭享受到了光纤到户带来的 20~100 MHz 接入带宽的极速上网体验。但在政府“提速降费”的要求及运营商自身推出“千兆入户”“4K 高清视频”等超高速业务需求的驱动下, EPON/GPON

向更高速率的升级演进已成为当下光接入领域颇受关注的议题。

IEEE 在 EPON 之后制定了 10 Gbit/s EPON 标准, 且已发布多年, 2013 年成立了 IEEE 802.3ca 工作组, 标志着 IEEE 已开始关注 100 Gbit/s EPON 标准的制定; FSAN/ITU-T 在 GPON 之后制定了 10 Gbit/s PON (对称、非对称) 标准, 较 10 Gbit/s EPON 晚几年发布, 随后马上启动了 40 Gbit/s PON 标准的制定, 2016 年标准已定稿, 因其通过 4 个波长、每波长 10 Gbit/s GPON 的方式实现 40 Gbit/s GPON, 故称为 TWDM PON, 40 Gbit/s

收稿日期: 2017-10-10; 修回日期: 2017-11-10



PON 在 FSAN 的标准体系中被称为 NG-PON2(最初 10 Gbit/s GPON 被定义为 NG-PON, 不过这个名字在业界很少使用), NG-PON2 除定义了 TWDM PON 之外, 还以附录形式定义了一种波分复用的可选机制, 使得在 TWDM PON 系统中能以点到点波分的方式为有特殊需求的客户提供专属的高速连接, 目前定义的单波长速率最高为 10 Gbit/s, 这一可选方式在业界被认为是 WDM PON 的实现标准。因此, PON 技术的演进需关注 10 Gbit/s EPON、10 Gbit/s GPON、TWDM PON、WDM PON、100 Gbit/s EPON 的技术进展与产业链面临的问题。其中, 10 Gbit/s EPON 已经在现网规模部署, 不是本文讨论的重点。本文将对 10 Gbit/s GPON、TWDM PON、WDM PON 的应用前景与面临挑战进行分析阐述, 同时对 100 Gbit/s EPON 标准的进展做简单介绍。

2 10 Gbit/s GPON 技术进展及应用情况

在标准上, 10 Gbit/s GPON 细分为两种配置: 上下行对称 10 Gbit/s 的 XG(S)-PON 与上行 1.25 Gbit/s、下行 10 Gbit/s 的 XG-PON。受芯片与光模块成熟度限制, 目前只有 XG-PON 可提供商用设备。在前期多次实验室测试的基础上, 2016 年中国电信开展了 XG-PON 的现场试验, 主要验证了高速上网业务承载、设备性能和功能、网管配置、WDM1r 对 ODN 的影响等, 包括以下内容。

- XG-PON 用于新建 FTTH 场景。
- XG-PON (外置 WDM1r) 用于 FTTH 升级场景: GPON 与 XG-PON 共存, OLT 新增 XG-PON 板卡, GPON ONU 按需更换为 XG-PON ONU。
- XG-PON 用于 FTTB 升级场景: 选择已有 GPON FTTB 网络, 局端更换为 XG-PON 板卡, MDU 上联板更换为 XG-PON 板卡。试点表明, 非对称的 XG-PON 技术与设备已

满足现网规模部署的要求, 预计 2017 年底中国电信将完成数十万端口的 XG-PON 建设规模。XG-PON 的成熟商用得益于互通性和与 GPON 兼容性两个关键问题的解决。

2.1 10 Gbit/s GPON 互通性测试进展

为推动 XG(S)-PON 的成熟商用, FSAN 曾组织了多次 XG-PON 的互通性测试, 但是由于全球各运营商的部署计划尚不明确, 厂商参与积极性不高, 进展始终比较缓慢。但中国电信较早地意识到了 XG(S)-PON 对网络升级演进的重要性, 并且认为需通过推动异厂商互通促进产业链的成熟, 分别在 2012 年、2013 年、2015 年、2016 年组织了共计 4 次互通性测试。在最近一次的 2016 年 6—8 月的测试中, 共有华为、中兴、烽火和诺基亚 4 个厂商参与, 解决了前期遗留的若干问题, 同时进行了互通场景下的多业务承载和长期稳定性测试, 为现网规模部署开展了大量准备工作。

互通测试主要包括以下几部分。

- 基本传输汇聚层互通性测试: 包括 ONU 发现和激活, OAM 管理操作, DD40 差异接入操作, 带宽分配, 对不同 ONU 的相邻分配, 嵌入式 OAM 操作 (PON-ID 设置、断电报警 (dying gasp) 识别、PLOAM 队列状态识别、动态带宽报告), PLOAM 通道运行 (多 PLOAM 消息共帧、突发配置版本更改、ONU 停用和序列号失效操作), Alloc-ID 分配及 De-allocation 删除, 前向纠错 (FEC)。
- ONU 二层功能操作: 基于 LOID 的认证, ONU 的 VLAN 功能测试, ONU 的 QoS 功能测试, ONU 上行业务流分类、标记和调度功能测试, ONU 上/下行端口限速功能测试, ONU 的多播功能测试, IGMP snooping 功能测试, ONU 跨 VLAN 多播功能复制测试, 强行剥除 ONU 下行 VLAN 多播功能测试, 多播容量 (group capacity) 测试。

- OMCI 互通性测试：状态告警上报、重启，能力集查询，配置模型测试，OMCI 的通道建立互通性测试，OLT 对 ONU 能力集查询测试，ONU 的状态告警上报功能测试，重启 ONU 功能测试，ONU 支持的 TCONT 的类型和数量的互通性测试，MDU SNMP 的配置测试，MIB reset 和 MIB upload 过程中逻辑标识认证测试，ONU 环路检测功能测试，ONU 的远程重启功能测试，N:MP 配置模型验证测试，基于 VEIP 1:MP 配置模型测试。
- 单播和多播业务混合组网测试。
- 系统以太网业务 RFC2544 测试。

尽管测试中还存在一些小问题留待后续进一步解决完善，但是与前 3 次相比，XG-PON 的异厂商互通性有了明显的进步，几个主要厂商的设备已经实现互通。总体来讲，XG-PON 产业链渐趋成熟，具备了小规模商用的条件。

2.2 10 Gbit/s GPON 与 GPON 兼容性进展

10 Gbit/s GPON 由 ITU-T G.987 系列标准定义，标准定义之初考虑了与 GPON 系统的共存。G.987 标准为 10 Gbit/s GPON 定义的频谱规划如图 1 所示，其中，10 Gbit/s GPON 上行频谱范围为

1 260~1 280 nm、下行频谱范围为 1 575~1 580 nm，上述频段被称为基础波段，其他系统的频谱范围为增强波段，GPON 系统的上下行频谱正好处于 1 310 nm 窗口与 1 550 nm 窗口的增强波段内，分别为 1 290~1 320 nm 和 1 480~1 500 nm。因此，从频谱规划可知，10 Gbit/s GPON 与 GPON 的兼容共存是通过波分复用的方式实现的。

图 2 是 G.987 定义的 GPON 与 10 Gbit/s GPON 共存方式。其中，WDM1r 器件完成的功能是将 GPON、10 Gbit/s GPON 两个系统的信号进行合分波，这一器件的引入在技术上较易实现，但在实际部署实施时会对 ODN 带来较大的调整割接工作量，是 GPON 网络存量较大的运营商所不愿意看到的。为此，中国电信开展了内置 WDM1r 器件技术方案的研究，在业界被称为双模光模块技术，该方案将 GPON MAC、10 Gbit/s GPON MAC、WDM1r 器件集成在一个业务板卡甚至一个模块内实现，网络升级改造时，运营商只需对设备板卡进行升级，而不用对 ODN 进行任何改造，具有易规模实施的特点。目前该技术已经通过了实验室测试，测试结果表明，GPON/XG-PON 双模光模块方案较为成熟，稳定性较好，光路指标及业务性能均满足现网部署要求，可用于现网 GPON 的升级场景。

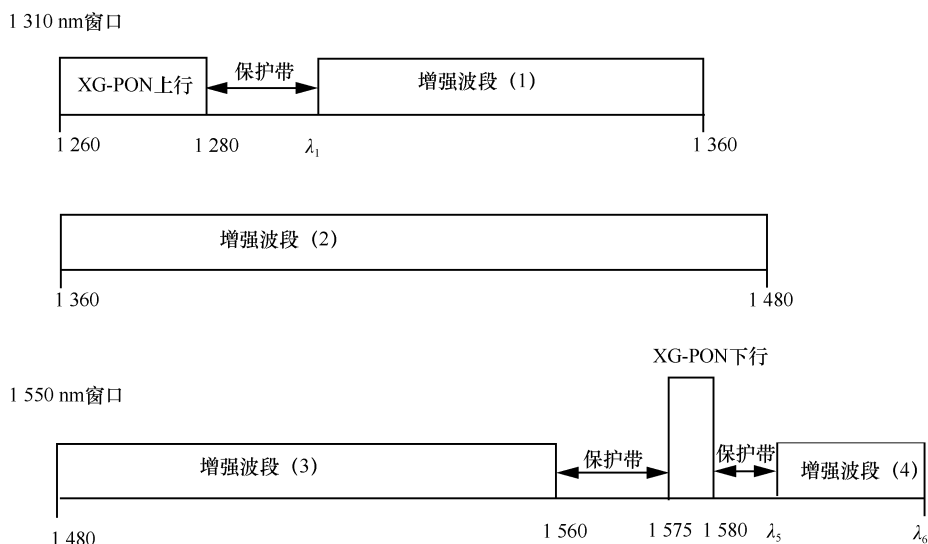


图 1 10 Gbit/s GPON 的频谱规划

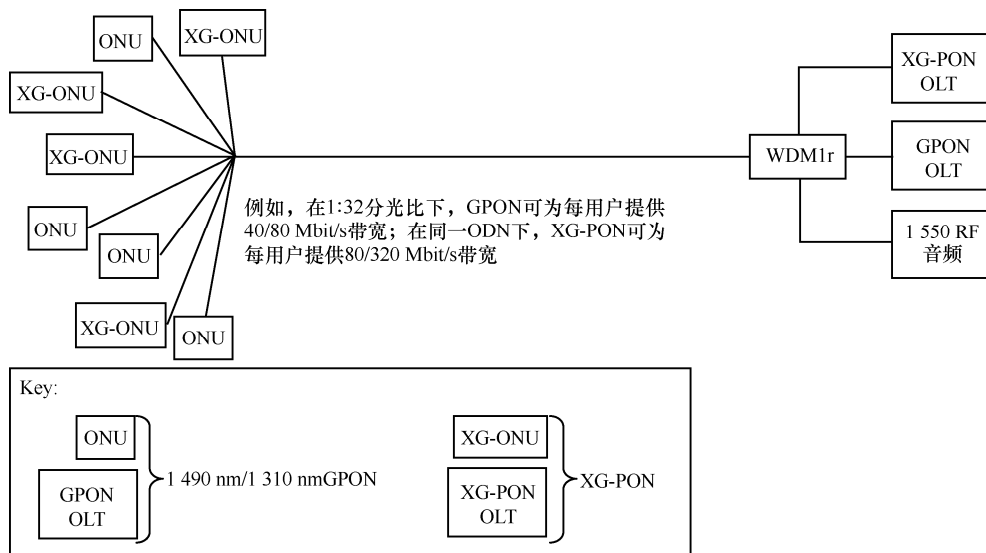


图2 G.987定义的GPON与10 Gbit/s GPON共存方式

目前双模光模块技术指标已经写入了 CCSA 相关标准，主要包括以下两点。

- class D1 级别的光指标要求，对应 GPON class B1、XG-PON class N1a。
- class D2 级别的光指标要求，对应 GPON class C++、XG-PON class N2a。

可以说，无论采用外置 WDM1r 器件，还是采用双模光模块方案，10 Gbit/s GPON 与 GPON 的兼容性都已解决，可支持现网规模部署。

3 TWDM PON 技术进展与面临的问题

3.1 TWDM PON 技术架构

TWDM PON 由 ITU-T G.989 系列标准定义，因突发模式的光模块在 10 Gbit/s 以上成本高昂，且技术实现难度太大，故标准定义的实现 40 Gbit/s 速率的方式是采用 4 个波长进行粗波分复用，每波长工作在 XG(S)-PON 模式下，一个典型的 TWDM PON 系统架构如图 3 所示。

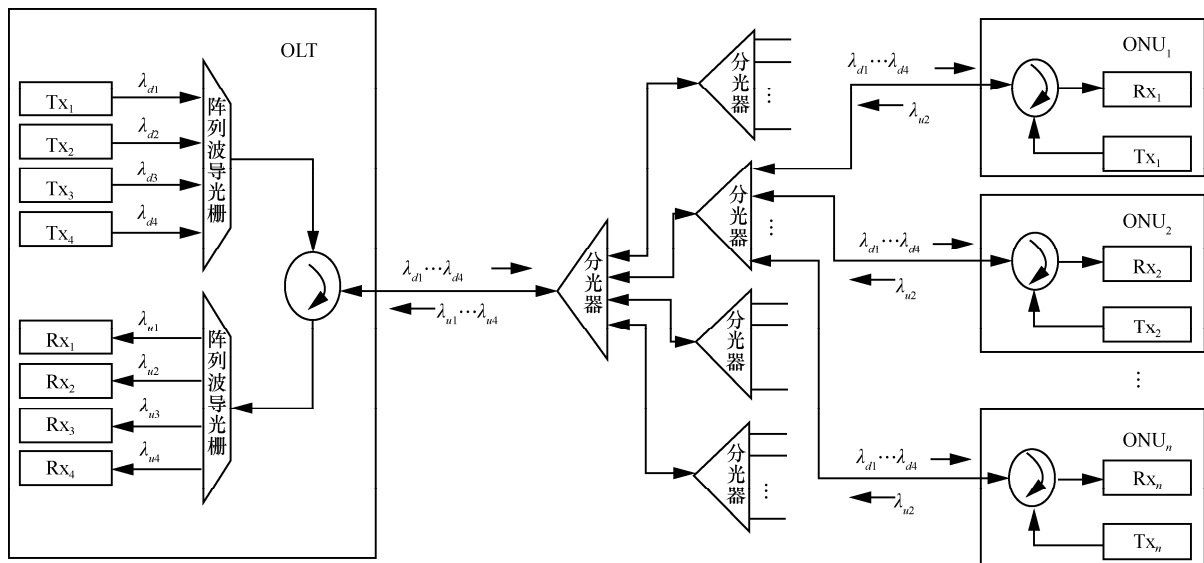


图3 典型的 TWDM PON 系统架构

3.2 TWDM PON 面临的问题

从技术架构可知, TWDM PON 重用了 XG(S)-PON 的技术方案, 在设备研发上只需增加粗波分复用及远端光模块的无色可调机制即可, 可以说相对于 XG(S)-PON 增加的实现难度并不大。但因为 XG-PON 刚刚成熟, XG(S)-PON 迟迟没有商用芯片与光模块支持, 何时可成熟商用尚不可预测, 因此, 设备商何时能推出可商用的 TWDM PON 设备仍需等待, 目前个别厂商推出的基于自研芯片或 FPGA 版本的 TWDM PON 板卡多数处于技术验证与演示阶段, 距离以商用芯片、商用光模块、可异厂商互通为标志的成熟阶段尚有一定的差距。

出现上述问题的原因在于绝大多数国际主流运营商都对 TWDM-PON 持观望态度, 不能给出明确的应用时间表, 导致主流设备厂商和光模块厂商投入有限。这一现象与 10 Gbit/s EPON、10 Gbit/s GPON 标准发布初期极其相似。回顾当初 EPON/GPON 迅速成熟商用的发展过程可知, 只有在业界出现几百万甚至上千万级别的市场需求才能促进设备的成熟商用。而目前国内运营商采用的是比较稳妥的逐步试验推广 XG-PON、XG(S)-PON 的发展战略, 在一定程度上更加让人看不到 TWDM PON 应用的明朗前景。

不过, 值得注意的是 Verizon 在 TWDM PON 应用上表现了积极的态度。2016 年, Verizon 与合作厂商在实验室测试 TWDM PON, 关注系统波长调谐能力、同平台承载家庭和政企业务、互通性等, 并宣称计划在 2017 年实现 TWDM PON 的政企场景商用部署。Verizon 接入技术总经理 Vincent O'Byrne 在 2017 年接受采访时表示, NG-PON2 可以在一根光纤上支持 4 个波长, 每个波长均可以承载 10 Gbit/s 上下行对称速率, 极大地增加了带宽; 另外, 采用 TWDM 的波长绑定技术, 可以将多个波长绑定提供给 1 个用户使用, 最大可以将所有的 40 Gbit/s 提供给 1 个用户。Vincent

O'Byrne 也提到, 2020 年北美家庭接入速率有望达到 10 Gbit/s, 届时必然需要下一代的 PON 接入技术来保证带宽。Verizon 跳过 XG-PON 选择 TWDM PON 的战略, 有可能会一定程度上加快该技术的商用与规模部署。

3.3 Google 光纤的 TWDM PON 技术实现

Google 光纤(Google Fiber)作为宽带接入领域的新进入运营商, 在技术创新上有很大的动力, 同时为快速部署新技术, 其对 ITU-T 等标准组织的依赖程度比传统运营商要小, 在这种背景下, Google 光纤推出了基于私有技术的 TWDM PON, 并已现网部署, 其项目名称为 go long。go long 系统由 12 波、每波 2.5 GB (GPON) 构成, 总系统容量为 30 GB。因其利用了现已成熟的 GPON 技术, 且通过 AWG+分光器的 ODN 架构, 用户终端不需要成本较高的可调光接收机, 因此实现难度与整体成本比 ITU-T G.989 定义的 TWDM 要低, 且支持每波长向 10 Gbit/s 速率的平滑演进。这一创新实现虽然不是经过国际标准化推广的技术, 但其思路是非常值得借鉴的。

4 WDM PON 技术的发展和應用

4.1 WDM PON 技术方案

在 G.989.2 的附录中, 针对波长选择型 WDM PON 的技术实现和相关技术指标进行了规范, 具体内容介绍如下。

- 波长选择型 WDM-PON 系统架构。
- 波长选择型 WDM-PON PMD 层要求: 4 个等级的线路速率、FEC 码、线路码型、2 个波段的工作波长以及相关参数。
- 波长选择型 WDM-PON 系统 OLT 和 ONU 的发射激光器相关内容。

WDM PON 为每个最终用户(可以是商业用户、基站、家庭客户等)提供了一个波长级的接入通道, 在带宽能力上无疑比 TDM 方式有了较大的提升及无限的想象空间。在讨论 WDM PON 的



技术特性时,业界公认的一点是 ONU 的无色化是实现 WDM PON 规模应用、最终走进家庭宽带用户的关键。无色化是指虽然每个 ONU 的工作波长不同,但在 ONU 发布到用户手中时是不需要每个 ONU 预先配置不同波长的,工作波长的分配与选择完全由系统上电后 OLT 自动化地实现远程的配置,就好像 ONU 是无波长特性的,即“无色”。无色的概念既是 WDM PON 最重要的应用特性之一,也是技术方案上最大的难点,业界在近些年进行了各种实现方案的探索。

表 1 列出了 5 种经过尝试的实现无色光模块的技术方案。5 种实现方案中,外部注入方案的要点是通过在 OLT 的种子光源注入实现无色光模块,ONU 采用 RSOA 或 FP-LD;在实际部署时由于需要种子光源,整体实现成本较高,由于受到光链路反射影响严重,传输距离受限在 20 km 以内。波长重用方案的要点是 ONU 采用 RSOA 激光器,上下行采用不同的调制格式(IRZ/RZ);此方案由于不同方向的调整格式不同,并且上行方向需要重用下行数据光作为种子光导致复杂性增加;此方案的优点是无需额外的种子光源,缺点是上行发射功率受限,光功率预算较小。可调型方案的要点是 ONU 采用可调激光器,可以灵活地调整上下行的波长;由于可调激光器内需要控制

波长的电路,结构复杂,同时可调激光器的成本也比较高,需要采用新型的无制冷器和波长锁定器的激光器;此方案的优点也比较明显,功率预算高并且可以支持很高的带宽(10 Gbit/s 以上)。超密集方案的要点是相干 PON 技术、ONU 可调激光器作为上行发射和下行接收本振;由于上下行采用相干接收技术,光功率预算大,具有更高的灵敏度,从而支持更大的分光比(1:1 000)和更远的传输距离(100 km),但是成本过高、技术尚不成熟导致近几年来并无突破性进展,此方案适合作为未来 NG PON3 的可选技术。自注入方案属于波长路由型 WDM PON,采用自注入锁定 RSOA 方案,无需种子光源,采用法拉第旋转反射镜,可以控制注入种子光偏振,结构简单,成本很低,但是目前只能支持 2.5 Gbit/s 以下的速率。

在这些 WDM PON 的技术方案中,业界普遍认为比较可行的是可调型和自注入两种,主要原因是这两种方案性能好,实际部署上也比较可行。早期在 2.5 Gbit/s 速率以下时,自注入获得更多青睐,主要由于其成本较低、实现简单。但是随着近期 WDM PON 的主要应用场景定位为无线前传,并且 CPRI 的速率逐渐升高到 10 Gbit/s 级别,目前 5G 的前传带宽需求也基本确定为 25/50 Gbit/s,自注入技术方案只能被放弃,可调型方案成为唯

表 1 WDM-PON 各种实现方案比较

技术方案	OLT	ONU	技术难点	性能	成本	方案介绍
外部注入	RSOA+L BLS	RSOA+远端 C BLS	OLT 阵列集成技术、RSOA 温度性能	较差	高	器件相对成熟,受种子光功率、反射等影响,性能较差
波长重用	激光器阵列	RSOA	OLT 阵列集成技术、RSOA 温度性能,下行信号重用技术,反射问题	差	较高	不需要独立的种子光,成本较低,但性能差,业界认同较低
可调型	激光器阵列	可调激光器	OLT 阵列集成技术、低成本可调激光器技术	好	高	可以实现灵活的 WDM PON 组网,结合可调滤波器可以重用 ODN,可用于长短路等各种应用
超密集	相干接收机+MZI	可调激光器+相干接收机	相干接收技术、可调激光器技术	很好	超高	功率预算高,支持长距大分光比应用,但技术成熟度低,成本高昂,目前仍处于实验室研究阶段
自注入	RSOA	RSOA	OLT 阵列集成技术、RSOA 温度性能	差	较高	与重调制一样不需独立种子光,成本较低,但性能差,业界认同较低,仅限实验室研究

一有可能实现 25/50 Gbit/s 速率的无色光模块技术方案。

无色光模块中的低成本可调激光器也是 TWDM PON、WDM PON 和 Q6 G.Metro 的核心器件。最近两年, 10 Gbit/s 级别的无色光模块渐渐成熟, 主要采用 QSFP28 和 SFP28 封装, 但目前成本仍然较高, 未来在降低成本的同时将向 25 Gbit/s 速率继续发展, 借助于 100 Gbit/s DWDM 客户侧光模块 (4×25 Gbit/s) 的成熟产业链, 25 Gbit/s 速率从激光器和调制器等器件角度来讲, 有望较快成熟。

4.2 WDM PON 应用进展

可以说, 在光接入领域, WDM PON 是目前最为前沿、距离落地部署最遥远的技术。因为其 2016 年才基本完成多种技术方案的归一化 (可调激光器成为主流), 所以 2016 年有关 WDM PON 的标准才有实质性的进展; 产业链的进展也比较缓慢, 目前厂商可提供的 WDM PON 样机在传输性能上都不能支持单波长 10 Gbit/s。但因为 WDM PON 可应用于 4G、5G 网络的移动前传场景, 在这方面的研究与试验还是有一定的进展。

中国电信在 2015 年完成了 WDM PON 在商用环境中承载 LTE 前传的试验, 采用了 2.5 Gbit/s 速率的无色光模块, 验证了 WDM PON 用于前传

的可行性。图 4 展示了试验环境, 将 WDM-PON 设备直接部署在无线 BBU 和 RRU 设备之间, WDM PON 的 OLT 和 ONU 之间通过无色光模块连接, 可以自动适配任何支持的波长, OLT 与 BBU 以及 ONU 与 RRU 之间通过白光光模块通信。试验测试了前传的物理性能, 包括抖动、频偏、时延, 并对业务承载能力进行了验证, 包括无线终端实时观看高清视频 (包括上网卡与手机) 与 FTP 大文件上传/下载等。测试结果表明, 采用 WDM PON 的 LTE 前传性能完全等同于采用光纤直驱的性能。

此外, 中国电信在安徽试点了无源彩光前传方案, 韩国运营商规模部署了各种点到点 WDM 方案承载 LTE 和 LTE-Advanced 前传, 有源和无源 WDM 均有应用, 法国电信也在实验室验证了 CPRI 速率压缩后采用 WDM-PON 承载前传。

总体来讲, LTE 时代运营商对于各种点到点 WDM 前传技术 (包括 WDM PON) 进行了大量的试验甚至部署, 在前传带宽、传输时延要求更高的 5G 时代, WDM PON 在某些光纤资源紧张的场景下将是一种比较可行的技术选择, 但因 eCPRI 最新发布的版本显示 25 Gbit/s 将是 5G 前传的主要接口带宽, 在标准支持、光模块与设备研发方面, 还需进一步推进才能促进其商用化进程。

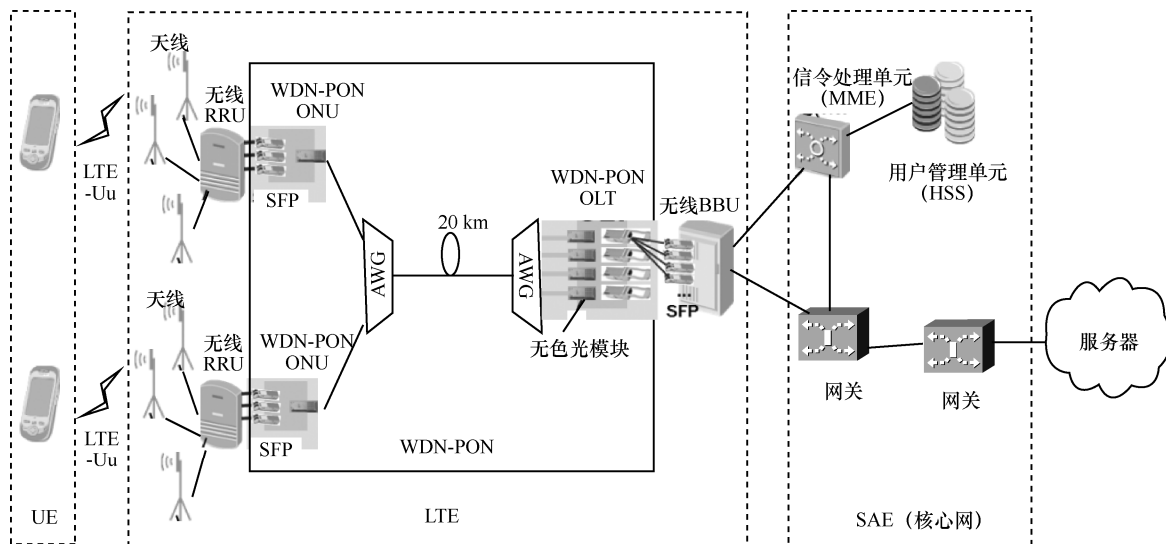


图 4 WDM PON 承载商用 LTE 前传试验拓扑



5 100 Gbit/s EPON 标准化介绍

NG-EPON 于 2013 年 7 月在 IEEE 802 全会提出了立项讨论,得到了多数厂商的响应,并由此建立了兴趣组讨论 NG-EPON 的需求以及可能的技术方案。正式立项为 IEEE 802.3ca(包含 25/50/100 Gbit/s EPON)标准后,预计 2017 年底前发布第一版草案,2019 年上半年发布标准正式版本。标准实际研发进度目前落后当初的预期,主要由于物理层和协议层影响因素多,关键技术的决策缓慢,主要包括以下两点。

- 物理层指标:通道波长、功率预算、发射功率、接收灵敏度等指标相互影响。
- 协议层机制:MPRS、MPCP(CH、OLT、ONU)机制的调整复杂。

为了提高国内运营商和厂商在 IEEE 100 Gbit/s EPON 标准制定中的话语权,体现国内未来可能的应用需求,由中国电信牵头,召集国内主要设备和光模块厂商于 2016 年 4 月成立 100 Gbit/s EPON 讨论组,讨论关键技术,明确技术建议,并共同推动 IEEE 会议上的讨论和决策。目前已经取得的工作成果如下。

- 明确国内未来可能的应用场景中,100 Gbit/s EPON 系统的技术需求。
- 共同推动 100 Gbit/s EPON 中“1+3”波长组合,在 IEEE 全会上决策通过。
- 研究讨论波长规划,对方案 A(上下行所有波长通道都在 O 波段)进行修改。
- 讨论并明确 10 Gbit/s EPON 上行波长收窄方案。

6 结束语

综上所述,1 Gbit/s 速率 PON 的演进路径并不是唯一的,10 Gbit/s GPON 与 TWDM PON 都有可能是运营商下一步的选择,甚至在 5G 前传场景下有可能直接部署 WDM PON。可以说,技术

标准提供了丰富完善的可选项,芯片、光模块、设备商、运营商也分别进行了有益的技术探索与积累,当前各种技术进展缓慢的原因还在于需求不明朗,国内运营商虽然在进行 10 Gbit/s PON 的部署,但总体来说目前国内外还是处在 EPON/GPON 的时代,在市场没有实质性需求的现阶段,10/40/100 Gbit/s 的 PON 技术都是应重点关注的方向。

参考文献:

- [1] ITU-T G.987.1. 10-Gigabit-capable passive optical networks (XG-PON): general requirements[S]. 2010.
- [2] ITU-T G.987.2. 10-Gigabit-capable passive optical networks (XG-PON): physical media dependent (PMD) layer specification[S]. 2016.
- [3] ITU-T G.989.2. 40-Gigabit-capable passive optical networks (NG-PON2): physical media dependent (PMD) layer specification[S]. 2014.

[作者简介]



张成良(1970—),男,中国电信股份有限公司北京研究院副院长、教授级高级工程师,主要研究方向为光接入及光传输技术。



王庆(1969—),男,湖北邮电规划设计有限公司总工程师、教授级高级工程师,主要研究方向为光接入及光传输技术。



余景文(1973—),男,中国电信股份有限公司北京研究院高级工程师,主要研究方向为光接入及光传输技术。

万芬(1978—),女,湖北邮电规划设计有限公司高级工程师,主要研究方向为光接入及光传输技术。



异步卫星协同通信系统中的双采样差分空时编码方案

章坚武, 屠贺嘉琦

(杭州电子科技大学, 浙江 杭州 310018)

摘要: 针对双卫星中继协作通信系统中存在非符号周期整数倍的时延差问题, 提出了一种双采样的差分空时正交频分复用(D-OFDM2)方案。所提方案在卫星信道条件未知的情况下, 基于具有时延容忍的 D-OFDM 编码, 在接收端构造了一种双采样方法。该方法在以符号速率采样的同时, 对当前主瓣大于相邻旁瓣的区间内增加一次采样, 并将两次采样结果等增益合并。仿真结果表明, 在系统时延差为非符号周期整数倍的情况下, 该方法较单采样方法能获得更高的接收信噪比, 同时改善了系统误码性能。

关键词: 卫星通信; 异步协同通信; 分布式差分空时编码; 双采样

中图分类号: TN927

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017309

Dual-sample differential space-time coding scheme in asynchronous satellite cooperative communication system

ZHANG Jianwu, TU Hejiaqi

Hangzhou Dianzi University, Hangzhou 310018, China

Abstract: Focusing on asynchronization in relay communication system over satellite fading channels, a dual-sampling differential space-time orthogonal frequency division multiplex (D-OFDM2) coding scheme was proposed. The proposed scheme constructed a dual-sampling method at the receiver based on the D-OFDM coding with delay tolerance when the satellite channel condition was unknown. The method added another sample to the current main lobe interval and equal gain merge the two sampling results. Comparing with the traditional scheme when the fractional delay is present, the D-OFDM2 scheme can obtain higher signal-to-noise ratio than single sampling method, and improves the system error performance.

Key words: satellite communication, asynchronous cooperative communication, distributed differential space-time coding, dual-sampling

1 引言

作为空地一体化系统^[1]的重要构成部分,

卫星通信指利用人造卫星作为中继站, 为地面水、陆、空域中无线电通信站提供通信^[2]。由于信号经卫星信道传输后会产生多径衰落^[3], 为

收稿日期: 2017-08-21; 修回日期: 2017-11-10

基金项目: 国家自然科学基金资助项目 (No.61471152); 浙江省自然科学基金资助项目 (No.LZ14F010003)

Foundation Items: The National Natural Science Foundation of China(No.61471152), Zhejiang Provincial Natural Science Foundation of China (No.LZ14F010003)



此可将多颗卫星作为相互协作的中继节点,组成虚拟分布式多输入多输出(multiple-input multiple-output, MIMO)系统,以达到分集的效果来对抗多径衰落。在低地球轨道(low earth orbit, LEO)卫星移动通信系统中,LEO卫星作为中继转发信号时可使用空时编码^[4,5],其中分布式空时编码(distributed space-time coding, DSTC)^[6,7]通过各中继相互协作对信号进行组合、转发处理,使信号到达接收端时能形成空时分组码(space-time block code, STBC)的形式,从而使系统在不牺牲额外带宽下得到分集增益和编码增益。然而采用DSTC编码时,接收端需要拥有瞬时信道状态信息(channel-state information, CSI)以实现信号的相干检测,难度较大。为此参考文献[8]提出了一种差分分布式空时编码(differential distributed space-time coding, D-DSTC)方法,使接收端在不明确CSI时即可进行非相干解码。另外,由于中继网络分布式特点,使转发信号在接收端产生时延差而导致符号间干扰(inter-symbol-interference, ISI)。参考文献[9]在时域中设计了一种具备较好的时延容忍度的螺旋代数空时编码,但其时间复杂度较大。参考文献[10,11]设计了一种有限时延容忍的时域交错反转码,可在保证满分集的情况下将解码过程分解为多个独立的子过程来降低解码复杂度。参考文献[12]为避免信道估计设计了一种基于扩频的差分空时编码。在频域方面,参考文献[13-15]采用一种正交频分复用(orthogonal frequency division multiplexing, OFDM)的方法来抵抗两中继网络中产生的同步误差,该方法可在保持编码正交的同时将时域的偏移转为频域的相位偏移。以上方法均针对时延差为符号周期整数倍的情况,当考虑时延差小数部分时,接收端对异步信号采样的位置会较其峰值发生偏移,从而使系统对有用信号的采样值减小并叠加进对旁瓣的采样值。针对中继网络中由

小数部分时延差带来的ISI,参考文献[16]对采用相干DSTC编码的双向中继网络采用了一种过采样方案,参考文献[17]将其进行扩展,提出了一种多中继情况下将差分DSTC与OFDM技术相结合的D-OFDM方法,但参考文献[16,17]方案仅适用于瑞利衰落信道下的小范围陆地中继传输。另外参考文献[18]给出了异步情况下采用差分DSTC编码的中继网络采用放大转发及解码转发时的系统性能比较,并优化了发射端和中继间的功率分配。根据以上情况,本文针对卫星协作通信下的异步问题设计了一种适用于在卫星衰落信道传输的双采样差分正交频分复用(dual-sampling differential orthogonal frequency division multiplexing, D-OFDM2)方案。该方案采用具有时延容忍的D-OFDM编码,并构造了一种双采样方案,使接收端对信号采样时始终能保持在当前符号主瓣大于其他符号旁瓣的区间内有两个不同采样点,将所得两采样结果进行等增益合并后,能使系统的平均接收信噪比增加,进一步减小系统异步带来的影响。

2 系统模型

考虑由一个发射端S,两颗中继卫星 R_1 、 R_2 及一个接收端D组成的分布式卫星协作通信系统,如图1所示,系统中节点均为单天线结构,传输模式选择半双工。系统传输信号分为两个阶段:第一阶段,S对信号进行D-OFDM编码并将其广播至 R_1 、 R_2 ;第二阶段, R_1 、 R_2 分别对接收到的信号进行空时编码处理并采用放大转发(amplify-and-forward, AF)协议转发信号至D。假设信号传输过程中收发两端不存在直射信号,信道为服从莱斯分布的准静态信道,各信道间互不相关,且每条信道均由 L 路独立的多径组成。两阶段中的各路多径信道系数分别由 $p_{i,l}$ 、 $q_{i,l}$ 表示,其中, $i=1,2$ 表示中继卫星编号, $l=1, \dots, L$ 。由于多径效应及各卫星相对收发两端位置的不

同,造成两路信号经传输后达接收端时存在时延差 ν ,系统因此变为异步系统。

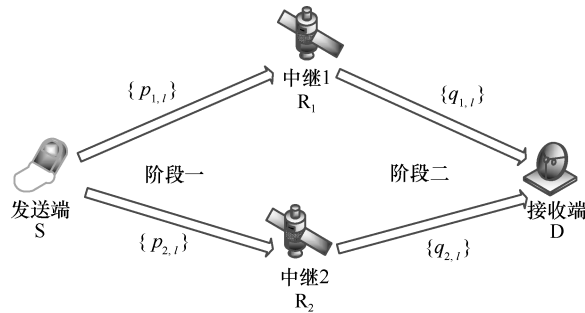


图1 两中继卫星协同通信系统模型

2.1 发射端编码

发射端对将要发送的信息进行编码,如图2所示。

将 N 个连续调制符号组成一组向量,记为 $x[n]$,然后将两组相邻的向量 $\{x_1[n], x_2[n]\}$ 构成一个酉空时矩阵 $X[n]$:

$$X[n] = \frac{1}{\sqrt{x_1[n]^2 + x_2[n]^2}} \begin{bmatrix} x_1[n] & -x_2^*[n] \\ x_2[n] & x_1^*[n] \end{bmatrix} \quad (1)$$

其中, $n=0, \dots, N-1$, $(\cdot)^*$ 表示共轭转置。

然后,对第 k 个酉空时矩阵 $X[n]^{(k)}$ 矩阵进行差分编码:

$$s[n]^{(k)} = X[n]^{(k)} s[n]^{(k-1)} \quad (2)$$

其中, $s[n] = [s_1[n], s_2[n]]^T$, 初始迭代值 $s[n]^{(0)} = [1, 0]^T$ 。

再将编码后的信号进行 OFDM 处理:

$$S[m] = \text{IDFT}\{s[n]\} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} s[n] \exp\left(j \frac{2\pi n m}{N}\right) \quad (3)$$

其中, $m=0, \dots, N-1$ 为 OFDM 中的第 m 个子载波。

随后,为信号添加循环前缀并进行串并转换,经脉冲整形后,在连续的两个 OFDM 时隙内依次将 $S_r^{(k)}[m]$ 广播发送至中继,其中, $r=1, 2$ 表示当前为第 r 个时隙, $S_r[m]$ 为 $S[m]$ 的第 r 行向量。

2.2 中继端处理

中继端接收到广播信号后,对其进行如图3所示的预处理。

处理后信号可表示为:

$$Z_{i,r}[m] = \sqrt{P_0 R} (p_i[m] \otimes S_r[m]) + \Psi_{i,r}[m] \quad (4)$$

其中, P_0 为发射端的符号发射功率, R 为中继卫星个数, $p_i[m] = \sum_{l=0}^{L-1} p_{i,l} \delta[m-l]$ 为第一阶段信道的离散冲击响应, $\Psi_{i,r}[m] \sim CN(0, N_0)$ 为发射端到中继引入的加性高斯白噪声。

对处理后的信号进行空时编码,见表1。

表1 中继对信号的处理

时隙	中继 1	中继 2
时隙 1	$V_{1,1}[m] = AZ_{1,1}[m]$	$V_{2,1} = -AZ_{2,2}^*[\langle -m \rangle_N]$
时隙 2	$V_{1,2}[m] = AZ_{1,2}[m]$	$V_{2,2} = AZ_{2,1}^*[\langle -m \rangle_N]$

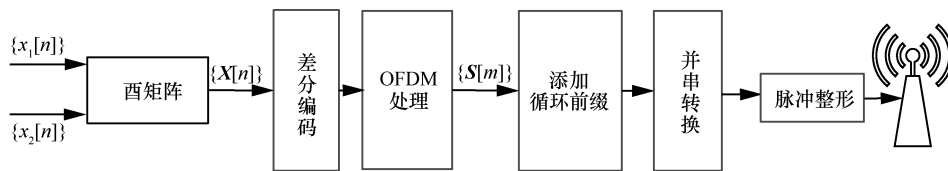


图2 发射端编码过程

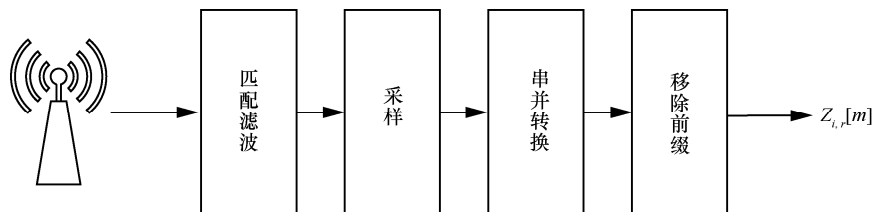


图3 中继端信号预处理过程



其中, $A = \sqrt{P_r / (P_0 + N_0)}$ 是放大系数, P_r 为中继端的符号发射功率, $Z_{i,r}[\langle -m \rangle_N]$ 为 $Z_{i,r}[m]$ 的圆周时域反转, 可表示为:

$$Z_{i,r}[\langle -m \rangle_N] = \begin{cases} Z_{i,r}[0], & m=0 \\ Z_{i,r}[N-m], & \text{其他} \end{cases} \quad (5)$$

各中继分别为信号添加循环前缀并进行并串转换, 经脉冲整形后在连续两个 OFDM 时隙内将信号 $V_{i,r}$ 发送到接收端。

2.3 接收端解调

信号到达接收端后, 进行中继端相同的预处理:

$$Y_r[m] = \sum_{i=1}^R (q_i[m] \otimes V_{i,r}[m]) + \Phi_r[m] \quad (6)$$

其中, $q_i[m] = \sum_{l=0}^{L-1} q_{i,l} \delta[m-l]$ 为第二阶段中信道的离散冲击响应, $\Phi_r[m] \sim CN(0, N_0)$ 为中继端到接收端引入的加性高斯白噪声。

然后, 对处理后信号进行 OFDM 解调:

$$y_r[n] = \text{DFT}\{Y_r[m]\} \quad (7)$$

随后在信道状态信息未知的情况下即可进行最大似然译码, 恢复出原始信号。

3 双采样改进方案

由于分布式中继网络的特点造成经不同中继转发的信号到达接收端会存在时延差。假设以 R_1 转发的信号到达接收端的时刻为准, 则经 R_2 转发的信号到达接收端产生的时延差可表示为:

$$v = (dT_s + \tau) \quad (8)$$

其中, T_s 表示一个符号周期, d 为整数, $0 \leq \tau < T_s$ 为时延差的小数部分。

当 v 的整数部分存在时, 表示两路信号到达接收端有 d 个符号的相对偏移, 对此可通过改变 OFDM 循环前缀长度, 在不影响编码正交性的前提下将时域的偏移转化为频域的相移^[12], 但该方法仅适用于 $\tau=0$ 的情况。当 $\tau \neq 0$ 时, 系统采样点位置较符号峰值位置产生错位, 导致采样后引入

相邻符号的旁瓣干扰, 为此本文提出了一种双采样方案。该方案通过在一个符号周期内设置两个采样点, 即在以符号速率 $0, \pm T_s, \pm 2T_s \dots$ 采样的同时, 在 $\pm T_s/2, \pm 3T_s/2, \pm 5T_s/2 \dots$ 处也进行一次采样, 并在采样完成后将两采样值进行等增益合并。改进后的接收端采样器设计如图 4 所示。

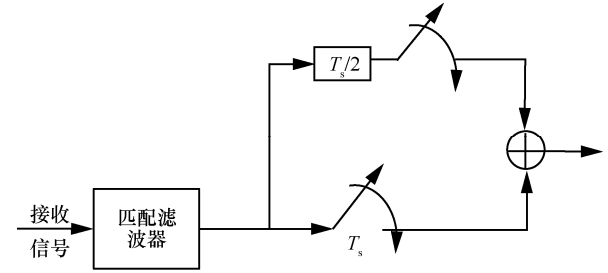


图4 改进的接收端采样器设计

采样过程中假设 $\tau \neq 0$, 旁瓣数 $L_{mf} = 1$, 如图 5 所示, 由于 τ 的存在, 原有采样点 1 对主瓣信号的采样位置较其波峰处产生偏移, 且随 τ 的增大, 对主瓣的采样值减小, 旁瓣的叠加值增大, 系统误码性能受到很大影响。因此, 本文提出了一种双采样方案, 通过在 $\pm T_s/2$ 处增加一个采样点 2, 其位置与波峰相距 $|T_s/2 - \tau|$, 这使得在一个符号周期中主瓣大于旁瓣的区间内, 始终有两个采样点对信号进行采样。 τ 的变化使两采样点对叠加信号的采样值发生变化, 当其中一个采样值变小时, 总有另一采样点位置更接近波峰, 采样更准确。

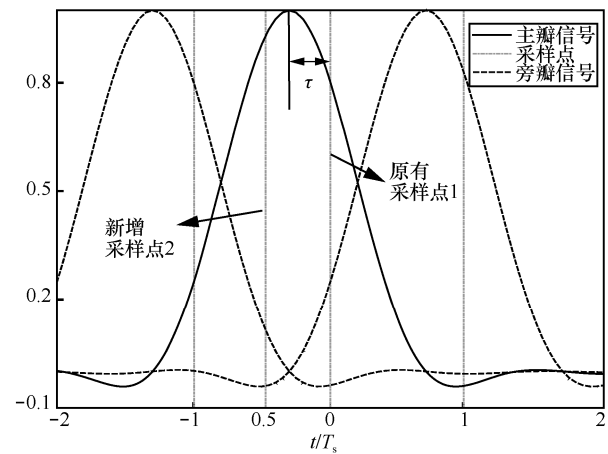


图5 改进后采样示意

本文采用升余弦滤波器:

$$p(t) = \sin c(t/T_s) \cos(\pi \beta t/T_s) / (1 - 4\beta^2 t^2/T_s^2) \quad (9)$$

则采样点 1、2 处的采样值分别为:

$$Y_r[m] = \sum_{i=1}^R \sum_{\tilde{l}=-L_{mf}}^{L_{mf}} p(\tilde{l}T_s - \tau_i) (q_i[m] \otimes V_{i,r}[m - d_i - \tilde{l}]) + \Phi_r[m] \quad (10)$$

$$\begin{aligned} \hat{Y}_r[m] &= \sum_{i=1}^R \sum_{\tilde{l}=-L_{mf}}^{L_{mf}} p((\tilde{l} + 0.5)T_s - \tau_i) \\ &\quad (q_i[m] \otimes V_{i,r}[m - d_i - \tilde{l}]) + \hat{\Phi}_r[m] \end{aligned} \quad (11)$$

其中, $\Phi_r[m] \sim CN(0, N_0)$, $\hat{\Phi}_r[m] \sim CN(0, N_0)$ 。

接着, 将两采样值进行等增益合并:

$$\begin{aligned} Y_r[m] + \hat{Y}_r[m] &= \\ &\sum_{i=1}^R \sum_{\tilde{l}=-L_{mf}}^{L_{mf}} (p(\tilde{l}T_s - \tau_i) + p((\tilde{l} + 0.5)T_s - \tau_i)) \\ &\quad (q_i[m] \otimes V_{i,r}[m - d_i - \tilde{l}]) + \Phi_r[m] + \hat{\Phi}_r[m] \end{aligned} \quad (12)$$

然后, 对合并后的信号进行 OFDM 解调:

$$\begin{aligned} y_r[n] &= \text{DFT} \{Y_r[m] + \hat{Y}_r[m]\} = \\ &\frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} (Y_r[m] + \hat{Y}_r[m]) \exp(-j \frac{2\pi mn}{N}) \end{aligned} \quad (13)$$

将式 (12) 代入式 (13) 中, 得到离散时域接收信号:

$$\begin{aligned} y_r[n] &= \sum_{i=1}^R \sum_{\tilde{l}=-L_{mf}}^{L_{mf}} (p(\tilde{l}T_s - \tau_i) + p((\tilde{l} + 0.5)T_s - \tau_i)) \\ &\quad Q_i[n] e^{-j \frac{2\pi n(d_i + \tilde{l})}{N}} v_{i,r}[n] + \phi_r[n] \end{aligned} \quad (14)$$

其中, $\phi_r[n] = \text{DFT} \{\Phi_r[m] + \hat{\Phi}_r[m]\}$, $v_{i,r} = \text{DFT} \{V_{i,r}[m]\}$,

$$Q_i[n] = \sum_{\tilde{l}=1}^{L_{mf}} q_{i,\tilde{l}} e^{-j \frac{2\pi n \tilde{l}}{N}}。$$

将 $y_r[n]$ 中的一部分用 $\tilde{Q}_i[n]$ 表示, 即:

$$\begin{aligned} \tilde{Q}_i[n] &= Q_i[n] e^{-j \frac{2\pi n d_i}{N}} \\ &\sum_{\tilde{l}=-L_{mf}}^{L_{mf}} (p(\tilde{l}T_s - \tau_i) + p((\tilde{l} + 0.5)T_s - \tau_i)) e^{-j \frac{2\pi n \tilde{l}}{N}} \end{aligned} \quad (15)$$

则接收端信号可表示为:

$$y_r[n] = \tilde{Q}_i[n] v_{i,r}[n] + \phi_r[n] \quad (16)$$

由参考文献[17]可知, 由两中继构成的分布式协作通信系统的信号传输过程的整体表达式为:

$$\begin{bmatrix} y_1[n] \\ y_2[n] \end{bmatrix} = A \sqrt{2P_0} \begin{bmatrix} s_1[n] & -s_2^*[n] \\ s_2[n] & s_1^*[n] \end{bmatrix} \begin{bmatrix} H_1[n] \\ H_2[n] \end{bmatrix} + \begin{bmatrix} w_1[n] \\ w_2[n] \end{bmatrix} \quad (17)$$

其中, $H_r[n]$ 及 $w_r[n]$ 分别为发射第 n 个符号时所对应子载波的等效信道系数及噪声。

结合系统第一阶段的离散频域信道系数:

$$P_i[n] = \sum_{l=0}^{L-1} p_{i,l} \exp(-j \frac{2\pi nl}{N}) \quad (18)$$

系统等效信道系数可表示为:

$$H_1[n] = P_1[n] \tilde{Q}_1[n] \quad (19)$$

$$H_2[n] = P_2^*[n] \tilde{Q}_2[n] \quad (20)$$

经系统传输后引入的等效加性高斯噪声为:

$$w_1[n] = A(\tilde{Q}_1[n] \psi_{1,1}[n] - \tilde{Q}_2[n] \psi_{2,2}^*[n]) + \phi_1[n] \quad (21)$$

$$w_2[n] = A(\tilde{Q}_1[n] \psi_{1,2}[n] + \tilde{Q}_2[n] \psi_{2,1}^*[n]) + \phi_2[n] \quad (22)$$

其中, $\psi_{i,r}[n] = \text{DFT} \{\Psi_{i,r}[m]\}$ 。

对于给定的 $q_{i,l}$, 系统等效噪声 $w[n] \sim CN(0, \sigma^2[n] \mathbf{I}_R)$, 其中:

$$\sigma^2[n] = N_0 \left(2 + A^2 \sum_{i=1}^R |\tilde{Q}_i[n]|^2 \right) \quad (23)$$

在实际情况下通常假设已知最大时延差整数部分 $d_{\max} = \max_{i=1, \dots, R} \{d_i\}$, 因此在循环前缀足够长且 $q_{i,l}$ 已知的情况下, 系统的平均接收符号信噪比 γ 可表示成小数时延差 τ_i 与子载波 n 的函数:

$$\gamma(n, \{\tau_i\}_{i=1}^R) = A^2 P_0 \frac{\sum_{i=1}^R |\tilde{Q}_i[n]|^2}{\sigma^2[n]} \quad (24)$$

当 τ_i 固定时, 对平均接收信噪比 γ 求导:



$$\frac{d\gamma}{d\sum_{i=1}^R|\tilde{Q}_i[n]|^2} > 0 \quad (25)$$

即平均接收信噪比 γ 是 $\sum_{i=1}^R|\tilde{Q}_i[n]|^2$ 的增函数。

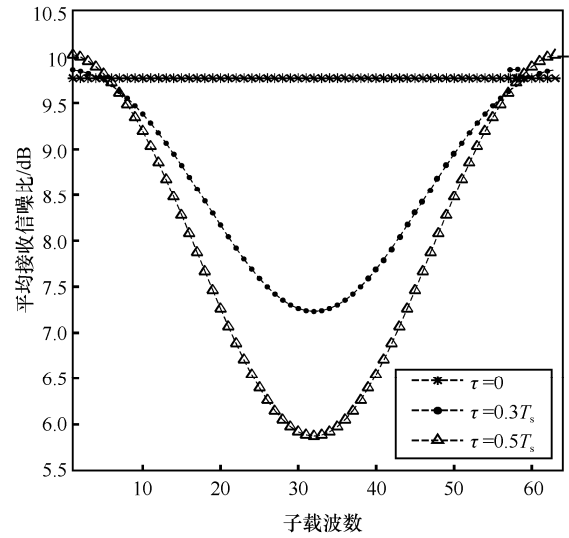
由于本文所提的双采样方案在等增益合并处理后, $|\tilde{Q}_i[n]|$ 较单采样方案增大, 因此系统的平均接收信噪比也随之增大。

4 性能仿真

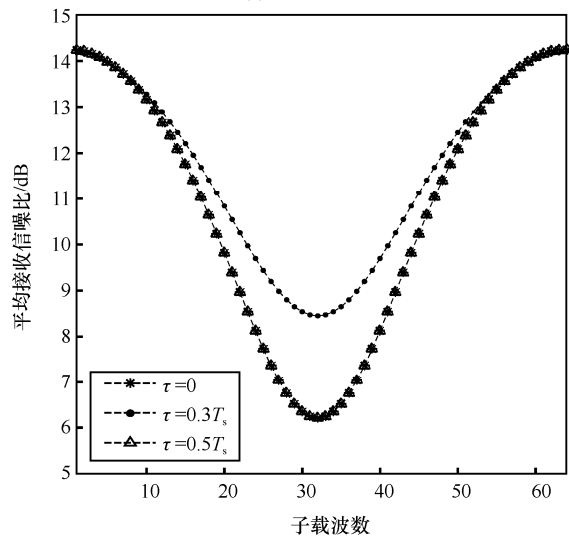
仿真中假设信道服从莱斯分布, 莱斯因子为 10, 将 L 条多径进行功率归一化处理, 即 $\sum_{l=0}^{L-1}\sigma_{q_{i,l}}^2=1$, $\sum_{l=0}^{L-1}\sigma_{s_{i,l}}^2=1$, 载频选取铱星系统用户链路频段, 信源调制方式采用 QPSK, 子载波数 $N=64$, 升余弦滚降系数 $\beta=0.9$, 旁瓣数 $L_{mf}=1$, 假设系统总功率为 P , 发射端和中继端的符号发射功率选取最优功率分配 $P_0=P/2$, $P_r=P/4$, 记单采样方案为 D-OFDM1, 所提双采样方案为 D-OFDM2。

为了简化, 假设系统多径数 $L=1$, 信道系数 $|q_{i,l}|^2=1$, $P/N_0=25$ dB。图 6 分别给出了 D-OFDM1 和所提 D-OFDM2 方案在不同 τ 下, 平均接收信噪比随子载波数的变化曲线。从图 6 中可以看出, 系统的平均接收信噪比 γ 关于子载波数中心对称, 这是因为在其他参数均给定的情况下, γ 仅为 $e^{\pm j\frac{2\pi n}{N}}$ ($\Theta=(1,2)$) 的一次函数, 而 $e^{\pm j\frac{2\pi n}{N}}$ 呈中心对称, 在 τ 相同的情况下, D-OFDM2 的值均大于 D-OFDM1。从图 6(b) 可以看出, $\tau=0$ 和 $\tau=0.5T_s$ 拥有相同的接收信噪比曲线, 这是因为当 $\tau=0$ 和 $0.5T_s$ 时, 采样点 1、2 的位置均处于相同位置, 只发生位置调换, 因此等增益合并后的采样值大小相等。

图 7 所示为不同 τ 情况下, 采用 D-OFDM1 及 D-OFDM2 方法得到的系统误码性能曲线。从图 7 中可以看出, 随着系统信噪比的增加, 两种



(a) D-OFDM1 方案



(b) D-OFDM2 方案

图 6 平均接收信噪比随子载波数的变化情况

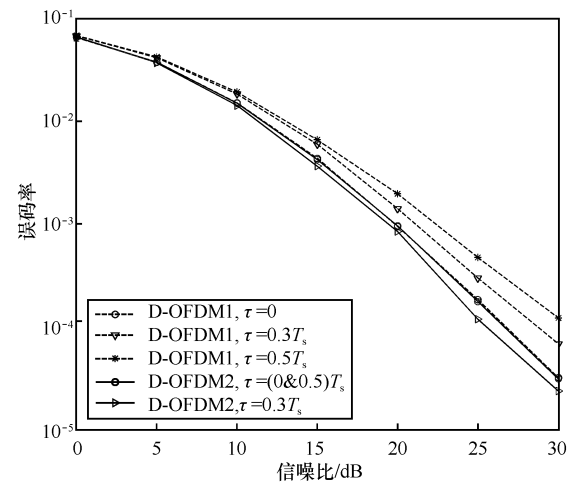


图 7 不同时延差下的系统误码性能

方案的误码率均减小,在 τ ($\tau \neq 0$) 相等的情况下, D-OFDM2 的误码率均低于 D-OFDM1, 且随着 τ 的增加, 系统的误码率差异进一步增大。当 $\tau=0.5 T_s$ 时, 二者差异达到了最大, 在误码率为 10^{-4} 情况下, 所提方法较原方法有 4 dB 的性能增益。且由于 $\tau=0.3 T_s$ 相比于 $\tau=0.5 T_s$, 系统能获得更高的平均接收信噪比, 故 $\tau=0.3 T_s$ 较 $\tau=0.5 T_s$ 能获得更低的系统误码率, 但二者差异很小, 近似相等。因此得出结论, 所提 D-OFDM2 方案能够提高系统的误码性能, 且几乎不受小数部分时延差大小的影响。

图 8 展示了当时延差 $\tau=0.3 T_s$ 的情况下, 不同旁瓣数 L_{mf} 对 D-OFDM2 系统误码性能的影响。可以看出, 当旁瓣数增加时, 系统误码性能差异并不明显。当系统误码率为 10^{-4} 时, $L_{mf}=4$ 较 $L_{mf}=1$ 只有 0.5 dB 的性能损失, 说明异步系统误码性能受第一旁瓣影响较大, 受其余旁瓣影响十分有限, 因此本文在仿真中选取 $L_{mf}=1$ 。

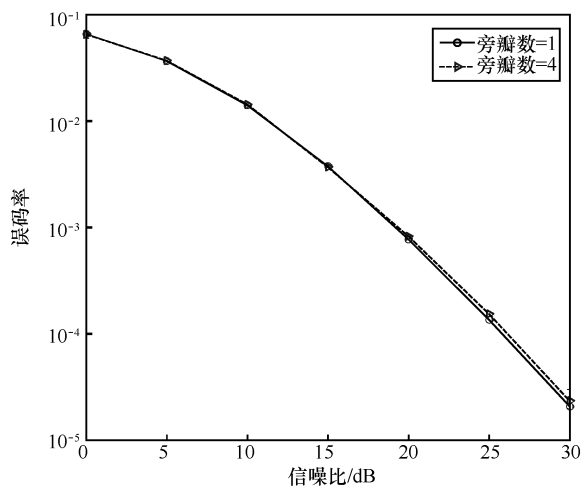


图 8 不同旁瓣数下的系统误码性能

5 结束语

由于分布式特点, 卫星协作通信系统中的接收端存在时延差不可避免。本文在频率选择性信道下设计了一种基于差分 DSTC-OFDM 编码的双采样方案。所提方案中接收端通过双采样, 成

功提高了接收端的平均接收信噪比, 消除了因小数部分时延差存在对系统误码性带来的影响。理论及仿真结果表明, 改进的 D-OFDM2 方案在系统小数部分时延差存在时, 能够较好地提高误码性能。

参考文献:

- [1] 李风华, 殷丽华, 吴巍, 等. 天地一体化信息网络安全保障技术研究进展及发展趋势[J]. 通信学报, 2016, 37(11): 156-168.
LI F H, YIN L H, WU W, et al. Research status and development trends of security assurance for space-ground integration information network[J]. Journal on Communications, 2016, 37(11): 156-168.
- [2] 易克初, 李怡, 孙晨华, 等. 卫星通信的近期发展与前景展望[J]. 通信学报, 2015, 36(6): 157-172.
YI K C, LI Y, SUN C H, et al. Recent development and its prospect of satellite communications[J]. Journal on Communications, 2015, 36(6): 157-172.
- [3] 李斗, 项海格. LEO/MEO 卫星通信系统发展展望[J]. 电信科学, 2003, 19(2): 48-51.
LI D, XIANG H G. The Development of LEO/MEO Satellite Communication Systems[J]. Telecommunications Science, 2003, 19(2): 48-51.
- [4] 蒋涉权, 王晶, 彭超. MIMO 系统中空时编码性能仿真和分析[J]. 电信科学, 2015, 31(2): 29-35.
JIANG S Q, WANG J, PENG C. Simulation and analysis on performance of space-time coding in MIMO[J]. Telecommunications Science, 2015, 31(2): 29-35.
- [5] 李娟, 邱晓红. 分层空时编码及应用研究[J]. 电信科学, 2011, 27(4): 91-95.
LI J, QIU X H. Research on layered space-time coding technology[J]. Telecommunications Science, 2011, 27(4): 91-95.
- [6] JING Y, HASSIBI B. Distributed space-time coding in wireless relay networks[J]. IEEE Transactions on Wireless Communications, 2007, 5(12): 3524-3536.
- [7] ANGHEL P A, KAVEH M. On the performance of distributed space-time coding systems with one and two non-regenerative relays[J]. IEEE Transactions on Wireless Communications, 2006, 5(3): 682-692.
- [8] JING Y, JAFARKHANI H. Distributed differential space-time coding for wireless relay networks[J]. IEEE Transactions on Communications, 2008, 56(7): 1092-1100.
- [9] DAMEN M O, HAMMONS A R. Delay-tolerant distributed-TAST codes for cooperative diversity[J]. IEEE Transactions on Information Theory, 2007, 53(10): 3755-3773.
- [10] LIU Y, ZHANG W, CHING P C. Full-diversity distributed



- space-time codes with an efficient ML decoder for asynchronous cooperative communications[C]//IEEE International Conference on Acoustics, Speech and Signal Processing, May 26-31, 2013, Vancouver, BC, Canada. New Jersey: IEEE Press, 2013: 5011-5015.
- [11] LIU Y, ZHANG W, CHING P C. Time-reversal space-time codes in asynchronous two-way relay networks[J]. IEEE Transactions on Wireless Communications, 2016, 15(3): 1729-1741.
- [12] 章坚武, 章谦骅. 卫星通信中基于扩频技术的异步差分空时协同编码研究[J]. 空间科学学报, 2015, 35(5): 634-640.
- ZHANG J W, ZHANG Q H. Asynchronous differential space-time block codes based on spreading techniques over satellite communication[J]. Chinese Journal of Space Science, 2015, 35(5): 634-640.
- [13] LI Z, XIA X G. An Alamouti coded OFDM transmission for cooperative systems robust to both timing errors and frequency offsets[J]. IEEE Transactions on Wireless Communications, 2008, 7(5): 1839-1844.
- [14] RAJAN G S, RAJAN B S. OFDM based distributed space time coding for asynchronous relay networks[C]//IEEE International Conference on Communications, May 19-23, 2008, Beijing, China. New Jersey: IEEE Press, 2008: 1118-1122.
- [15] LEE Y, KANG S G, CHONG D, et al. An alamouti coding scheme for asynchronous cooperative communication systems over frequency selective fading channels[J]. 2011, 4(4): 1-5.
- [16] BARGHI S, JAFARKHANI H. Exploiting asynchronous amplify-and-forward relays to enhance the performance of IEEE 802.11 networks[J]. IEEE/ACM Transactions on Networking, 2015, 23(2): 479-490.
- [17] AVENDI M R, JAFARKHANI H. Differential distributed space-time coding with imperfect synchronization in frequency-selective channels[J]. IEEE Transactions on Wireless Communications, 2015, 14(4): 1811-1822.
- [18] XU W, WANG Q, LIN M. Distributed space-time coding scheme with differential detection and power allocation for cooperative relay network[J]. Telecommunication Systems, 2017: 1-15.

[作者简介]



章坚武 (1961-), 男, 杭州电子科技大学通信工程学院教授、博士生导师, 主要研究方向为移动通信系统、多媒体通信技术、网络安全等。



屠贺嘉琦 (1992-), 男, 杭州电子科技大学通信工程学院硕士生, 主要研究方向为移动通信系统。



研究与开发

3D MIMO-OFDMA 系统中基于垂直波束成形的能效优化算法

李汀, 仇林杰, 季薇

(南京邮电大学通信与信息工程学院, 江苏 南京 210003)

摘要: 针对三维多输入多输出 (3D MIMO) 正交频分多址 (OFDMA) 系统, 提出了一种能效优化算法。该算法在垂直波束成形技术下, 以能量效率最大化为目标, 通过调整资源分配、功率分配、天线的波束下倾角来提高系统能量效率。根据分数优化理论, 将复杂的分数优化问题转化为较易求解的整式优化问题, 然后引入拉格朗日乘子通过不断迭代得到能量效率的最优值。仿真结果表明, 所提算法在较少迭代次数下可以获得更高的能量效率。

关键词: 三维多输入多输出; 垂直波束成形; 能量效率; 分数优化理论

中图分类号: TN929

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017266

Energy-efficient optimization algorithm based on vertical beamforming for 3D MIMO-OFDMA system

LI Ting, QIU Linjie, JI Wei

School of Communication and Information Engineering,
Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Abstract: An energy-efficient optimization algorithm was proposed for the 3D MIMO-OFDMA communication system. The proposed algorithm based on vertical beamforming technology maximized energy efficiency by adjusting the resource allocation, power allocation and the downtitles of the beam. According to the theory of fractional optimization, the complex fractional optimization problem could be converted into a more tractable integral optimization problem. The lagrangian operator was introduced to obtain the optimal value of energy efficiency by iteration. Simulation results show that the proposed algorithm can achieve higher energy efficiency with fewer iterations.

Key words: 3D MIMO, vertical beamforming, energy efficiency, fractional optimization theory

收稿日期: 2017-07-02; 修回日期: 2017-09-20

基金项目: 国家自然科学基金资助项目 (No.61771254, No.61471200), 江苏省自然科学基金资助项目 (No.BK20140881), 南京邮电大学横向项目 (No.2016 外 71)

Foundation Items: The National Natural Science Foundation of China (No.61771254, No.61471200), Jiangsu Provincial Natural Science Foundation of China (No.BK20140881), The Horizontal Project of Nanjing University of Posts and Telecommunications (No.2016 Out-of-School 71)



1 引言

随着无线通信技术的飞速发展,各种新型传输技术的使用以及传输方案的设计都可以不同程度地提高系统的吞吐量,这使得移动数据的传输速率呈现指数级增长趋势,但是传输速率飞速增长的同时,能量消耗速度急剧增加,通信设备的成本也随之不断提高。同时,无线通信系统巨大的能量消耗产生了大量的气体排放,引起环境污染。因此,在传输过程中,要充分考虑能量损耗因素,实现绿色通信,这也体现了可持续发展的思想。

在通信系统中,由于基站和用户终端的数量庞大,产生了巨大的能量损耗,其中,基站产生的能量占据了绝大部分比例,随着4G网络的不断完善以及5G网络的兴起,通信产业将会产生更多的能量损耗。大量的能量被消耗意味着巨额的电费成本开支,在整个网络通信过程中,电费开支大约占总成本的20%~30%,并且随着用户数量的增长和用户对服务质量要求的不断提高,电费开支占总成本的比例会进一步提高。对于运营商而言,为了使利润持续增长,需要尽可能地降低能耗来减少运营的成本,从而提高收入。对于终端用户而言,提高能量效率能够延长设备的使用寿命。致力于推进5G发展的欧洲研究小组METIS明确提出,能量效率应作为5G的关键指标之一^[1]。5G系统中能量效率的研究在多蜂窝系统、超蜂窝系统、中继系统、多用户系统等不同场景中都展开了广泛的研究^[2-6]。

3D MIMO (three dimensional multiple-input multiple-output, 三维多输入多输出) 技术是5G的重要技术之一,因此,在3D MIMO场景下探讨能量效率也具有重要的研究意义。

对于通信系统传输过程的研究,能量效率问题受到越来越多的关注。参考文献[7]证明了在OFDMA无线系统中,能量效率会随着信道增益

及子载波数目的增加而增加,随着电路能耗的增加而减少。在此基础上,提出了功率及子载波资源分配方案,该方案提高了系统的能量效率。参考文献[8]研究了SISO-OFDM系统下行链路的能效资源分配问题,将非凸优化问题转化为凸优化问题,给出了有效的资源分配算法。但考虑的场景是SISO (single input single output, 单输入单输出) 系统,在MIMO场景下该方案并不适用。参考文献[9]基于MIMO-OFDMA无线通信系统提出了一种服务质量约束条件下的功率、子载波资源分配方案,该方案可以实现系统能效性能的最大化。在3D MIMO下行链路传输系统中,基站天线发送给小区用户的波束下倾角也影响系统的能量效率^[10,11]。大多数能效优化的文献都是首先以系统总的吞吐量与总功耗的比值得到系统的能量效率,能效优化问题实质就是一个分数优化问题,然而直接求解此分数优化问题通常比较困难。参考文献[12]将分数优化问题转化为等效的整式优化问题,这样可以更有效地解决优化问题。

在调研上述文献之后,本文针对3D MIMO-OFDMA这一新型系统,充分利用垂直维的自由度,综合考虑功率分配、资源分配以及天线的波束下倾角,将能效最大化的分式优化问题转化为整式优化问题,提出了一种基于拉格朗日乘子法的能效优化迭代算法,用于求解能效最优时的功率和资源分配方案以及天线的波束下倾角。

2 3D MIMO 信道模型和天线阵列结构

本文采用的信道模型为三维空间信道模型 (three dimension spatial channel model, 3D SCM)^[13]。与传统2D SCM只考虑水平维度的到达角和离开角不同,3D SCM还考虑了垂直维的自由度。3D SCM架构如图1所示。

假设发送端和接收端之间有 N 条主径,每条主径有 M 条子径。在发送端,每条路径都有水平

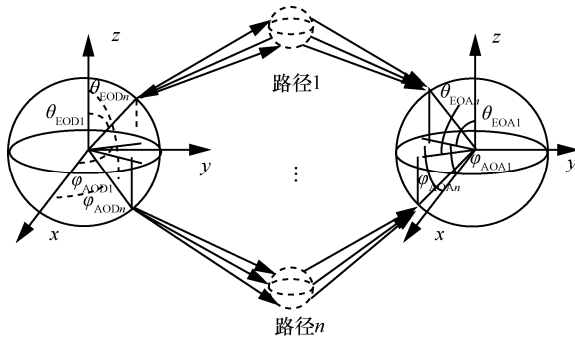


图1 3D SCM 架构

离开角 (azimuth angle of departure, AOD) 和垂直离开角 (elevation angle of departure, EOD); 在接收端, 每条路径都有水平到达角 (azimuth angular of arrival, AOA) 和垂直到达角 (elevation angle of arrival, EOA)。发送天线 s 经过第 n 条主径到达接收天线 u 的信道系数可表示为:

$$H_{u,s,n}(t) = \sqrt{P_n/M} \times \sum_{m=1}^M \begin{bmatrix} F_{\text{rx},u,\theta}(\theta_{n,m,\text{EOA}}, \varphi_{n,m,\text{AOA}}) \\ F_{\text{rx},u,\varphi}(\theta_{n,m,\text{EOA}}, \varphi_{n,m,\text{AOA}}) \end{bmatrix} \times \begin{bmatrix} \exp(j\phi_{n,m}^{\theta\theta}) & \sqrt{\kappa_{n,m}} \exp(j\phi) \\ \sqrt{\kappa_{n,m}} \exp(j\phi_{n,m}^{\theta\theta}) & \exp(j\phi_{n,m}^{\varphi\varphi}) \end{bmatrix} \times \begin{bmatrix} F_{\text{tx},s,\theta}(\theta_{n,m,\text{EOD}}, \varphi_{n,m,\text{AOD}}) \\ F_{\text{tx},s,\varphi}(\theta_{n,m,\text{EOD}}, \varphi_{n,m,\text{AOD}}) \end{bmatrix} \times \exp(j2\pi\lambda^{-1}(r_{\text{tx},n,m} \cdot d_{\text{tx},u})) \times \exp(j2\pi\lambda^{-1}(r_{\text{tx},n,m} \cdot d_{\text{tx},s})) \times \exp(j2\pi\nu_{n,m}t) \quad (1)$$

其中, $F_{\text{rx},u,\theta}$ 和 $F_{\text{rx},u,\varphi}$ 分别代表第 u 根接收端天线的垂直基底方向极化复增益和水平基底方向极化复增益; $F_{\text{tx},s,\theta}$ 和 $F_{\text{tx},s,\varphi}$ 分别代表第 s 根发送端天线的垂直基底方向极化复增益和水平基底方向极化复增益; $\theta_{n,m,\text{EOA}}$ 和 $\varphi_{n,m,\text{AOA}}$ 分别代表第 n 条主径的第 m 条子径的 EOA 和 AOA; $\theta_{n,m,\text{EOD}}$ 和 $\varphi_{n,m,\text{AOD}}$ 分别代表第 n 条主径的第 m 条子径的 EOD 和 AOD; κ 为交叉极化鉴别率 (cross polarization power ratio, XPR); $d_{\text{rx},u}$ 和 $d_{\text{tx},s}$ 分别代表接收端第 u 根天线和发送端第 s 根天线的位置矢量; $\nu_{n,m}$ 代表第 n 条主径的第 m 条子径的多普勒频率分量。

发送天线 s 的极化复增益矩阵可以进一步

写为:

$$\begin{bmatrix} F_{\text{tx},s,\theta}(\theta_{n,m,\text{EOD}}, \varphi_{n,m,\text{AOD}}) \\ F_{\text{tx},s,\varphi}(\theta_{n,m,\text{EOD}}, \varphi_{n,m,\text{AOD}}) \end{bmatrix} = \begin{bmatrix} \cos \alpha \sin \theta_{n,m,\text{EOD}} + \sin \alpha \sin \theta_{n,m,\text{EOD}} \cos \varphi_{n,m,\text{AOD}} \\ \cos \alpha \cos \varphi_{n,m,\text{AOD}} \end{bmatrix} \quad (2)$$

其中, α 表示天线极化角度。

接收天线 u 的极化复增益矩阵为:

$$\begin{bmatrix} F_{\text{rx},u,\theta}(\theta_{n,m,\text{EOA}}, \varphi_{n,m,\text{AOA}}) \\ F_{\text{rx},u,\varphi}(\theta_{n,m,\text{EOA}}, \varphi_{n,m,\text{AOA}}) \end{bmatrix} = \begin{bmatrix} \cos \alpha \sin \theta_{n,m,\text{EOA}} + \sin \alpha \sin \theta_{n,m,\text{EOA}} \cos \varphi_{n,m,\text{AOA}} \\ \cos \alpha \cos \varphi_{n,m,\text{AOA}} \end{bmatrix} \quad (3)$$

$r_{\text{tx},n,m}$ 为对应于到达方向 $(\theta_{n,m,\text{EOA}}, \varphi_{n,m,\text{AOA}})$ 所表征的单位矢量, 即:

$$r_{\text{tx},n,m} = [\cos \theta_{n,m,\text{EOA}} \cos \varphi_{n,m,\text{AOA}}, \cos \theta_{n,m,\text{EOA}} \sin \varphi_{n,m,\text{AOA}}, \sin \theta_{n,m,\text{EOA}}] \quad (4)$$

$r_{\text{tx},n,m}$ 为对应于离开方向 $(\theta_{n,m,\text{EOD}}, \varphi_{n,m,\text{AOD}})$ 所表征的单位矢量, 即:

$$r_{\text{tx},n,m} = [\cos \theta_{n,m,\text{EOD}} \cos \varphi_{n,m,\text{AOD}}, \cos \theta_{n,m,\text{EOD}} \sin \varphi_{n,m,\text{AOD}}, \sin \theta_{n,m,\text{EOD}}] \quad (5)$$

3D MIMO 系统根据三维信道状态信息, 可以调整基站端有源天线的三维波束成形权值, 使波束的主瓣在三维空间内更好地对齐所要传输的用户, 从而提高用户接收到的功率和用户端的信干噪比, 进而提高系统总的吞吐量。图2为多用户 3D MIMO 系统波束成形示意。从图2中可以看出, 若采用传统的 2D MIMO 波束成形技术, 只能区分用户 1 和用户 2, 而很难区分具有相同水平位置的用户 2 和用户 3。但是, 对于基于有源天线的 3D MIMO 波束成形技术, 虽然用户 2 和用户 3 具有相同的水平方位角, 但是他们具有不同的垂直仰角, 3D MIMO 可以利用垂直维度的信道状态信息来区分用户 2 和用户 3。因此, 基于有源天线的 3D MIMO 系统可以获得更高的系统吞吐量。

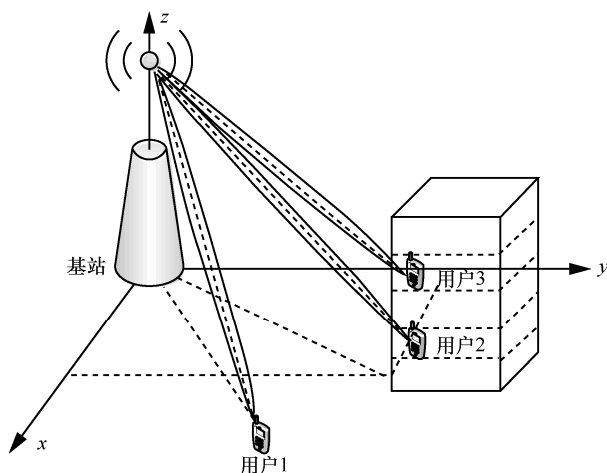


图2 多用户 3D MIMO 系统波束成形示意

3 下行链路基站能量损耗模型

本文考虑实际下行链路传输过程中的基站能量损耗模型。在下行链路传输过程中，天线发送功率在系统总的能量损耗中占据了较大部分，在传统多天线技术中系统的能量损耗也主要考虑功率放大器（以下简称功放）上消耗的能量，但是每根天线单元中的电器设备也占据着一定的能量，这些能量也是不可忽略的，尤其是当天线数目比较多时。根据 LTE 基站的模型，下行链路传输过程中各电路模块组成模型如图 3 所示。

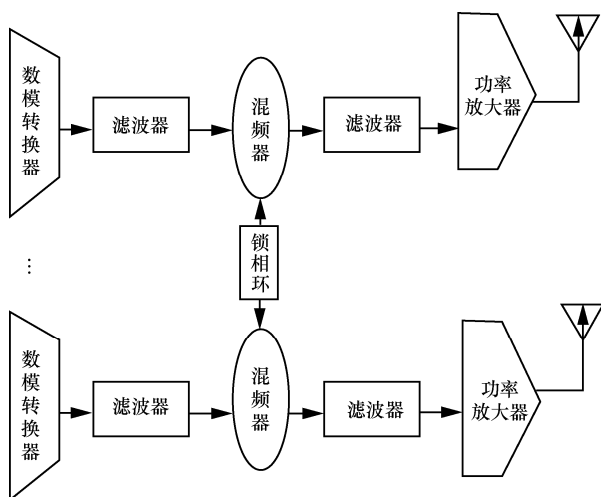


图3 下行链路传输过程中各电路模块组成模型

由参考文献[14]可知，基站每根天线在下行链路工作时，其平均功率消耗主要由两部分组成：

电路能耗和功放能耗。电路能耗 P_c 是指电器硬件设备消耗的能量，这些电路设备包括数模转换器 (DAC)、滤波器、混频器、锁相环等。故发送端电路功耗的计算式为：

$$P_c = P_{\text{DAC}} + P_{\text{filter}} + P_{\text{mixer}} + P_{\text{LO}} \quad (6)$$

其中， P_{DAC} 表示数模转换器的功率损耗， P_{filter} 表示滤波器的功率损耗， P_{mixer} 表示混频器的功率损耗， P_{LO} 为锁相环同步所需要的功率损耗。根据参考文献[15]，功放能耗模型在实际系统中的取值见表 1。

表 1 功放能耗模型在实际系统中的取值

电路模块	数模转换器	滤波器	混频器	锁相环同步器
功耗/mW	15.6	20	30.3	50

功放能耗是所有功率放大器的能量消耗 P_{PA} ，它随着天线发送功率 P_t 线性变化。它们之间的关系式近似表示为：

$$P_{\text{PA}} = (1 + \beta) P_t \quad (7)$$

其中， $\beta = \varepsilon / \eta - 1$ ， η 表示功放的漏极效率， ε 表示基带信号的峰均比 (peak to average ratio, PAR)，其值取决于调制方式和星座图点数^[16]，例如，对于 NQAM 调制， $\varepsilon = 3(\sqrt{N} - 1 / \sqrt{N} + 1)$ ；

对于 QPSK 调制， $\varepsilon = 1$ 。本文采用 QPSK 的基带调制，故下行链路总的功率损耗为：

$$\begin{aligned} P_{\text{total}} &= P_{\text{PA}} + M P_c \\ &= P_t / \eta + M (P_{\text{DAC}} + P_{\text{filter}} + P_{\text{mixer}}) + P_{\text{LO}} \end{aligned} \quad (8)$$

其中， M 为天线根数。

4 系统模型

本文所考虑的系统为下行单小区 3D MI-MO-OFDMA 系统，如图 4 所示，系统中有一个基站配置 S 根天线和 M 个接收用户。假设信道为块时变的慢衰落信道，并且基站知道信道状态信息 (channel state information, CSI)。系统总的频谱资源被划分为每组 N 个正交的资源模块，每个资源模块由 K 个频率上相邻的子载波组成。

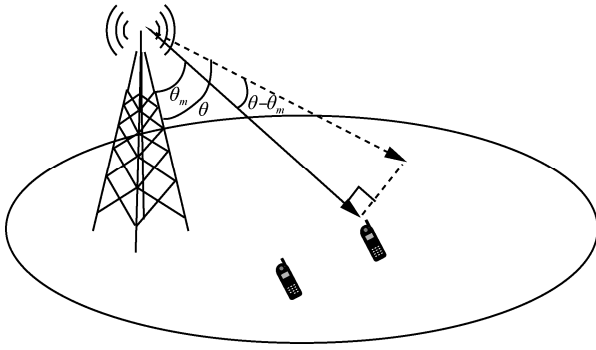


图4 3D MIMO 单小区多用户系统

设第 n 个资源模块的第 k ($k=1,2,\dots,K$) 个子载波为 $k(n)$ 。为了防止不同用户间的相互干扰,假设每个资源模块只发送给一个用户。记发送给第 m 个用户的第 n 个资源模块的第 k 个子载波的功率为 $P_{m,k(n)}$, $h_{m,k(n)}$ 代表第 n 个资源模块的第 k 个子载波发送到第 m 个用户时的信道增益。记基站天线发送给用户时垂直维度的波束下倾角为 θ , θ_m 为小区第 m 个用户基于位置的垂直到达角,则发送给第 m 个用户的波束角度与用户期望的辐射角度间的夹角为 $\theta - \theta_m$ 。本文仅考虑信道的小尺度衰落,从图4中可以看出,天线的发射功率乘以3D MIMO 信道的小尺度增益 $P_{m,k(n)} h_{m,k(n)}$ 为基站按照天线波束下倾角 θ 发送给用户的功率,即当用户的垂直到达角 $\theta_m = \theta$ 时,用户能获得的信号功率;当 $\theta_m \neq \theta$ 时,即用户的垂直到达角不等于天线的波束下倾角时,从图4中的三角关系示意可以看出,此时第 m 个用户实际接收到第 n 个资源模块的第 k 个子载波的信号功率应该为 $P_{m,k(n)} h_{m,k(n)}$ 乘以一个余弦夹角,即 $P_{m,k(n)} h_{m,k(n)} \cos(\theta - \theta_m)^{[17,18]}$ 。设用户受到的噪声为均值为0、方差为 σ^2 的加性高斯白噪声。因此第 m 个用户接收到第 n 个资源模块的第 k 个子载波后用户端的信干噪比为:

$$SINR_{m,k(n)} = \frac{P_{m,k(n)} h_{m,k(n)} \cos(\theta - \theta_m)}{\sigma^2} \quad (9)$$

由香农定理知,第 n 个资源模块的第 k 个子载波发送给第 m 个用户的传输速率为:

$$R_{m,k(n)} = B \log \left(1 + \frac{P_{m,k(n)} h_{m,k(n)} \cos(\theta - \theta_m)}{\sigma^2} \right) \quad (10)$$

其中, B 表示子载波的带宽。

故第 n 个资源模块发送给第 m 个用户的传输速率为:

$$R_{m(n)} = \sum_{k=1}^K B \log \left(1 + \frac{P_{m,k(n)} h_{m,k(n)} \cos(\theta - \theta_m)}{\sigma^2} \right) \quad (11)$$

第 n 个资源模块发送给第 m 个用户总的功率为:

$$P_{m,n} = \sum_{k=1}^K P_{m,k(n)} \quad (12)$$

则总的系统吞吐量可以表示为:

$$C_{\text{total}} = \sum_{m=1}^M \sum_{n=1}^N s_{m,n} R_{m(n)} \quad (13)$$

其中, $s_{m,n}$ 表示第 n 个资源模块是否分配给了第 m 个用户, $s_{m,n} \in \{0,1\}$ 。

由式(13)可知,系统总的吞吐量与功率分配、资源模块分配以及天线的波束下倾角有关,因此把总的系统吞吐量记为 $C(P,S,\theta)$ 。这里 P 、 S 、 θ 分别代表功率、资源模块、波束下倾角分量。此系统下,总的功率损耗计算式为:

$$P_{\text{total}} = P_c + \sum_{m=1}^M \sum_{n=1}^N \frac{P_{m,n}}{\eta} + P_{\text{LO}} \quad (14)$$

其中, P_c 为天线持续的电路损耗功率,其包括发送滤波器、混频器、频率合成器、数模转换器的电路损耗功率,与实际的发射功率无关。也就是说,式(14)的第一项表示被激活的天线所消耗的总能耗,第二项表示基站的射频放大器的总功率能耗。 η 表示功放的漏极效率, P_{LO} 表示与发送端天线根数无关的同步器的能耗。故由式(14)可知,系统总的能耗与功率分配、资源模块分配有关,把系统总的功率能耗记为 $A(P,S)$ 。因此能量效率的计算式为:



$$EE = \frac{C_{\text{total}}}{P_{\text{total}}} \quad (15)$$

可记为:

$$EE(P, S, \theta) = \frac{C(P, S, \theta)}{A(P, S)} \quad (16)$$

因此, 基于垂直波束成形的资源分配方案 (P^*, S^*, θ^*) 可通过求解下述最优化问题得到:

$$\begin{aligned} & \max_{P, S, \theta} EE(P, S, \theta) \\ \text{s.t. (C1)} & \sum_{n=1}^N s_{m,n} R_{m,n} \geq R_m \\ \text{(C2)} & \sum_{m=1}^M \sum_{n=1}^N s_{m,n} P_{m,n} \leq P_T \\ \text{(C3)} & \sum_{m=1}^M \sum_{n=1}^N s_{m,n} \leq 1 \\ \text{(C4)} & P_{m,n} \geq 0, \quad \forall m, n \\ \text{(C5)} & s_{m,n} \in \{0, 1\} \quad \forall m, n \\ \text{(C6)} & 0^\circ \leq \theta \leq 180^\circ \end{aligned} \quad (17)$$

其中, C1 是确保系统能够满足用户 m 至少需要的最小传输速率 R_m 的约束条件。C2 是下行链路基站发送功率约束条件, P_T 表示基站发出的总功率。C3 和 C5 是确保每个资源模块服务于一个用户。C4 是确保基站发送功率为正值。C6 是波束下倾角的约束条件。

5 能效优化迭代算法

式 (17) 的目标优化函数为非凸函数, 且约束条件较多, 直接求解此最优化问题基本上是不可行的, 为了得到能效优化资源分配算法, 根据分数优化理论^[19]首先对目标函数进行变换。

设约束参数的取值集合为 $\mathbf{O}$, $(P^*, S^*, \theta^*) \in \mathbf{O}$ 为式 (20) 的最优解, q^* 为最优的能量效率值。则有:

$$q^* = \frac{C(P^*, S^*, \theta^*)}{A(P^*, S^*)} \geq \frac{C(P, S, \theta)}{A(P, S)}, \quad \forall \{P, S, \theta\} \in \mathbf{O} \quad (18)$$

因此有:

$$\begin{aligned} C(P, S, \theta) - q^* A(P, S) & \leq 0 \\ \text{且 } C(P^*, S^*, \theta^*) - q^* A(P^*, S^*) & = 0 \end{aligned} \quad (19)$$

即若 $\{P^*, S^*, \theta^*\} \in \mathbf{O}$ 为式 (17) 的最优解, 则对于 $\forall \{P, S, \theta\} \in \mathbf{O}$, 有:

$$\begin{aligned} C(P, S, \theta) - q^* A(P, S) & \leq 0 \\ C(P^*, S^*, \theta^*) - q^* A(P^*, S^*) & = 0 \end{aligned} \quad (20)$$

上述计算式表明:

$$\frac{C(P, S, \theta)}{A(P, S)} \leq q^*, \quad \forall \{P, S, \theta\} \in \mathbf{O} \quad (21)$$

并且有:

$$\frac{C(P^*, S^*, \theta^*)}{A(P^*, S^*)} = q^* \quad (22)$$

故在相同约束条件下, 求 $\frac{C(P, S, \theta)}{A(P, S)}$ 的最大值可以等效为求 $C(P, S, \theta) - qA(P, S)$ 的最大值。故接下来可以关注等效的目标优化函数。

能效优化问题等效的最优化问题如下所述:

$$\begin{aligned} & \max_{P, S, \theta} C(P, S, \theta) - qA(P, S) \\ \text{s.t.} & \text{ C1} \sim \text{C6} \end{aligned} \quad (23)$$

所等效的最优化问题式 (23) 是一个联合优化问题。为了得到有效的资源分配算法, 对如何获得 (P^*, S^*, θ^*) 进行讨论。

令 $F(q) = C(P, S, \theta) - qA(P, S)$, 则当 $q = q^*$ 、 $P = P^*$ 、 $S = S^*$ 、 $\theta = \theta^*$ 时, $F(q)$ 取得最大值。对于给定的 q , 最优化问题 (23) 的拉格朗日函数为:

$$\begin{aligned} L(\mu, \lambda, \beta, P, S, \theta) & = \left[\sum_{m=1}^M \sum_{n=1}^N s_{m,n} R_{m(n)} - \right. \\ & q \left(\varepsilon \sum_{m=1}^M \sum_{n=1}^N s_{m,n} P_{m,n} + SP_c + P_{LO} \right) + \sum_{m=1}^M \mu_m \left(\sum_{n=1}^N s_{m,n} R_{m(n)} - r \right) + \\ & \lambda \left(P_T - \sum_{m=1}^M \sum_{n=1}^N s_{m,n} P_{m,n} \right) + \beta (\theta - 180) \\ & = \sum_{n=1}^N \sum_{m=1}^M (1 + \mu_m) s_{m,n} R_{m(n)} - \sum_{n=1}^N \sum_{m=1}^M (q\varepsilon + \lambda) s_{m,n} P_{m,n} + \\ & \lambda P_T - \sum_{m=1}^M \mu_m R_{m(n)} - qSP_c + \beta (\theta - 180) \end{aligned} \quad (24)$$

其中, $\mu_m \geq 0$ 、 $\lambda \geq 0$ 、 $\beta \geq 0$ 分别表示与用户的接收速率约束变量、发送功率约束变量和波束下倾角约束变量相对应的拉格朗日乘子。

首先, 对发送给小区中心用户的波束下倾角固定, 先对资源和功率分配进行优化。则:

$$\max_{\{P_n, S_n\}} L(P_n, S_n) = \sum_{m=1}^M \left[(1 + \mu_m) R_{m(n)} - (q\varepsilon + \lambda) P_{m,n} \right] \quad (25)$$

则根据 KKT (Karush-Kuhn-Tucker, 卡罗需—库恩—塔克) 条件可得:

$$P_{m,k(n)}^* = \left(\frac{(1 + \mu_m)B}{(q\varepsilon + \lambda) \ln 2} - \frac{\sigma^2}{h_{m,k(n)} \cos(\theta - \theta_m)} \right) \quad (26)$$

将式 (26) 代入式 (25), 则式 (25) 可以写为:

$$\Gamma_{m,n} = \max_{P_n^*} L(S_n, P_n^*, \theta) \quad (27)$$

则最优化资源分配方案可以通过式 (28) 得到:

$$s_{m,n}^* = \begin{cases} 1, m = \arg \max_{1 \leq m \leq M} \Gamma_{m,n} \\ 0, \text{其他} \end{cases} \quad (28)$$

在资源分配和功率分配之后, 把式 (26) 和式 (28) 代入式 (24), 然后对 θ 求偏导得:

$$\theta^* = \frac{\sum_{m=1}^M (1 + \mu_m) \theta_m - \frac{\beta \ln 2}{B}}{\sum_{m=1}^M (1 + \mu_m)} \quad (29)$$

拉格朗日乘子在第 l 次迭代的更新式为:

$$\begin{cases} \beta^{l+1} = [\beta^l - \tau_\beta^l (180 - \theta^*)]^+ \\ \lambda^{l+1} = \left[\lambda^l - \tau_\lambda^l \left(P_T - \sum_{m=1}^M \sum_{n=1}^N s_{m,n}^* P_{m,n}^* \right) \right]^+ \\ \mu_m^{l+1} = \left[\mu_m^l - \tau_{\mu_m}^l \left(\sum_{m=1}^M \sum_{n=1}^N s_{m,n}^* R_{m,n} - R_m \right) \right]^+ \end{cases} \quad (30)$$

其中, l 表示迭代因子, τ_β^l 、 τ_λ^l 、 $\tau_{\mu_m}^l$ 表示合适的迭代步长, $[x]^+ = \max\{0, x\}$ 。拉格朗日乘子根据式 (30) 不断迭代直到算法收敛。

本文提出了一种迭代算法以解决式 (24) 的最优化问题。在算法实现中, 假设信道状态信息是已知的, 并且子载波的信道带宽小于相干带宽。迭代资源分配算法流程如下。

初始化 (μ, λ, β) , 初始化最大迭代步数 $L_{\max}$ 、最大容忍值 ξ 、 τ_β^l 、 τ_λ^l 、 $\tau_{\mu_m}^l$

设迭代步数 $l=0$ 时, 最大的能量效率值 $q=0$

repeat 主循环进行迭代

计算每个资源模块 RB $n=1$ to N 送给用户 m 时的增益 $h_{m,k(n)}$

for 每个用户 $m=1$ to M

根据式 (26) ~ 式 (28) 计算 $P_{m,k(n)}^*$ 和 $s_{m,n}^*$

end for

根据式 (29) 计算 θ^*

if $C(P^*, S^*, \theta^*) - qA(P^*, S^*) < \xi$

Convergence=true

返回 $\{P^*, S^*, \theta^*\}$, $q^* = \frac{C(P^*, S^*, \theta^*)}{A(P^*, S^*)}$

else

令 $q = \frac{C(P^*, S^*, \theta^*)}{A(P^*, S^*)}$, $l = l + 1$

Convergence=false

根据式 (30) 更新 (μ, λ, β) , 转至步骤 4

end if

until Convergence=true 或者达到最大迭代次数
返回 $\{P^*, S^*, \theta^*\}$ 为所求最优解

6 仿真分析

对 3D MIMO-OFDMA 能效性能进行仿真分析。仿真中的 2D MIMO 信道模型采用参考文献[20]给出的 2D SCM 参数, 3D MIMO 信道模型采用参考文献[13]给出的 3D SCM 参数。天线阵列模型为均匀线性阵列 (uniform linear array, ULA), 天线阵元的间隔为半个波长, 发送端天线数为 2, 接收端天线数为 2。假设用户均匀分布在小区中, 且用户在小区内不移动。系统中有 3 个用户和 8 个资源



模块,每个资源模块有12个子载波(在实际系统中子载波和用户数目一般都很大),子载波带宽为 $B=10$ kHz,系统带宽 $W=5$ MHz,基带信号采用QPSK调制。小区半径为300 m,基站高度为30 m,用户受到的噪声为均值为0、方差为0.1的高斯白噪声。基站的能量损耗参数见表1,功放的漏极效率 $\eta=0.35$ 。

图5给出了在本文所提能效优化迭代算法下,作为目标函数的能量效率值随迭代步数的变化曲线。从图5中可以看出,在刚开始的几次迭代过程中,能量效率值波动比较明显,但随着迭代步数的增加,能量效率值波动渐渐平稳,最终趋于稳定值。这是因为在搜寻最优解时,最初的搜索点一般不是最优解,但随着不断迭代变化,搜寻的点将会不断向最优解靠拢。仿真结果显示,在通过较少的迭代步数之后,就能得到原最优化问题的最优解。

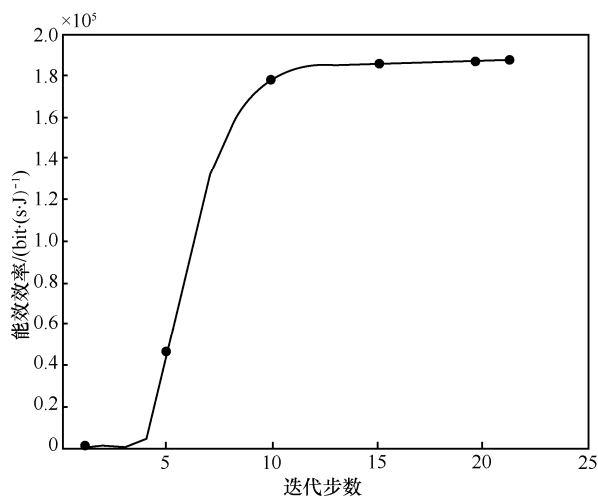


图5 目标函数的能量效率值随迭代步数的变化曲线

图6比较了3D MIMO系统和2D MIMO系统的能效性能。从图6中可以看出,随着信噪比的增加,2D MIMO系统和3D MIMO系统的能量效率也不断增大。在相同信噪比的条件下,3D MIMO系统能效性能要明显优于2D MIMO系统。这主要是因为3D MIMO系统垂直维自由度提高了系统的传输速率,因此能量效率也比2D MIMO系统高。

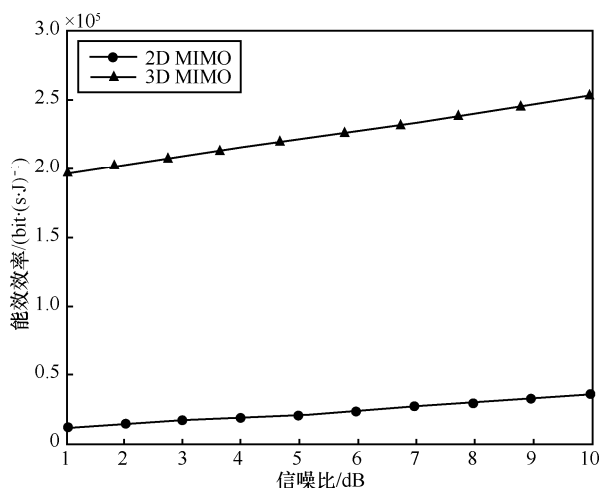


图6 3D MIMO系统和2D MIMO系统能效性能比较

图7比较了3D MIMO系统下本文所提能效优化迭代算法和参考文献[21]所提的最大化频谱效率算法的能量效率随着最大发送功率约束变化的情况。从图7中可以看出,随着基站最大发送功率约束的增加,以最大化能量效率为目标的算法和以最大化频谱效率为目标的算法,系统总的能量效率均随着基站最大发送功率的增加而增加,而且可以明显看出本文所提能效优化迭代算法的能量效率高于最大化频谱效率算法的能量效率。

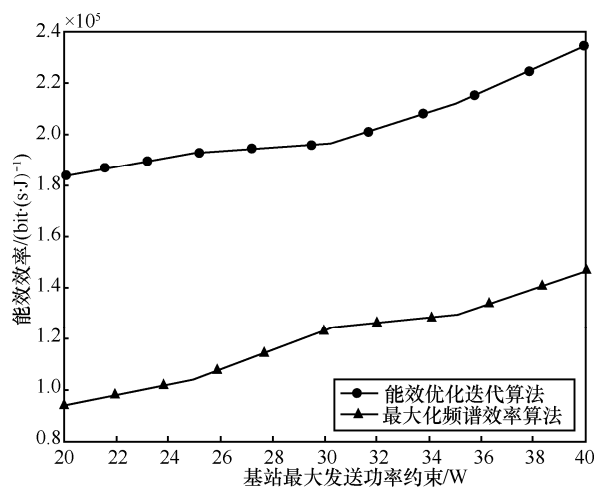


图7 不同发送功率限制下算法的能效性能比较

7 结束语

本文对3D MIMO-OFDMA下行链路传输系统的能量效率进行了研究。与传统的2D MIMO

系统相比,通过垂直波束成形动态调整发送给用户的波束下倾角以进一步提高系统的能量效率。为了求解能效优化问题,根据分数优化理论,将原始的分数优化问题转化为等价的整式优化问题,然后通过资源分配、功率分配和波束下倾角调整进行联合优化。对于求解等效的整式优化问题,本文提出了基于垂直波束成形的资源分配迭代算法进行求解。根据引入拉格朗日乘子表示的拉格朗日函数得到每次迭代的资源分配、功率分配方案以及波束下倾角的值,然后通过不断迭代直到达到最大迭代步数或者算法收敛,从而得到最大的能量效率值。仿真结果表明,本文所提算法能够有效提高系统的能量效率,系统的能效性能与 2D MIMO 系统相比有明显提高,同时也比 3D MIMO 系统最大化频谱效率方案的能量效率高。

参考文献:

- [1] FALLGREN M, TIMUS B. Scenarios, requirements and KPIs for 5G mobile and wireless system[S]. METIS deliverable D, 2013.
- [2] 牛志升, 龚杰, 周盛. 能效与资源优化的超蜂窝移动通信系统基础研究[J]. 电信科学, 2014, 30(12): 34-47.
NIU Z S, GONG J, ZHOU S. Research on energy efficiency and resource optimized hyper-cellular mobile communication system[J]. Telecommunications Science, 2014, 30(12): 34-47.
- [3] 郭杨波, 张治中, 胡昊南, 等. 关于 5G 中继技术的节能算法研究[J]. 电信科学, 2015, 31(3): 55-60.
GUO Y B, ZHANG Z Z, HU H N, et al. Research on energy saving algorithm of 5G relay cellular network[J]. Telecommunications Science, 2015, 31(3): 55-60.
- [4] LI C, ZHU W, YANG L. Optimal energy to spectral-efficiency trade-off in cooperative networks[J]. Wireless Personal Communications, 2015, 80(3): 1-20.
- [5] LI C, SUN F, JOHN M, et al. Energy efficient MIMO relay transmissions via joint power allocations[J]. IEEE Transactions on Circuits and Systems-II, 2014, 61(7): 531-535.
- [6] LI C, SONG K, LI Y, et al. Energy efficient design for multiuser downlink energy and uplink information transfer in 5G[J]. Science China Information Science, 2016, 59(2): 1-8.
- [7] MIAO G, HIMAYAT N, LI Y, et al. Energy efficient design in wireless OFDMA[C]//IEEE International Conference on Communications, May 19-23, 2008, Beijing, China. New Jersey: IEEE Press, 2008: 3307-3312.
- [8] KIM H S, DANESHRAH B. Energy-constrained link adaptation for MIMO OFDM wireless communication Systems[J]. IEEE Transactions on Wireless Communications, 2010, 9(9): 2820-2832.
- [9] ISHEDEN C, CHONG Z, JORSWIECK E, et al. Framework for link-level energy efficiency optimization with informed transmitter[J]. IEEE Transactions on Wireless Communications, 2011, 11(8): 161-176.
- [10] 王凯, 张亮, 谭国平. 一种新颖的 3D 有源天线基站组网波束下倾角选择方法[J]. 电信科学, 2015, 31(1): 90-94.
WANG K, ZHANG L, TAN G P. A novel beam downtilt selection method for 3D active antenna system based cellular network[J]. Telecommunications Science, 2015, 31(1): 90-94.
- [11] SHI Q J, XU W Q, LI D, et al. On the energy efficiency optimality of OFDMA for SISO-OFDM downlink system[J]. IEEE Communications Letters, 2013, 17(3): 541-544.
- [12] ISHEDEN C, FETTWEIS G P. Energy-efficient link adaptation with transmitter CSI[C]//IEEE Wireless Communications and Networking Conference, March 28-31, 2011, Cancun, Mexico. New Jersey: IEEE Press, 2011: 1381-1386.
- [13] 3GPP. Study on 3D channel model for LTE: TR36.873, V1.3.0[S]. 2014.
- [14] REN Z Y, ZHANG H L, GUO K. A novel method for the improvement of power efficiency in high peak-to-average-power ratio communication systems[J]. Science China Information Sciences, 2010, 53(8): 1697-1702.
- [15] ARNOLD O, RICHTER F, FETTWEIS G, et al. Power consumption modeling of different base station types in heterogeneous cellular networks[C]//Future Network and Mobile Summit, June 16-18, 2010, Florence, Italy. New Jersey: IEEE Press, 2010: 98-110.
- [16] CUI S, GOLDSMITH A J, BAHAI A. Energy-constrained modulation optimization[J]. IEEE Transactions on Wireless Communications, 2005, 4(5): 2349-2360.
- [17] ZHANG W, WANG Y, PENG F, et al. Interference coordination with vertical beamforming in 3D MIMO-OFDMA networks[J]. IEEE Communications Letters, 2014, 18(1): 34-37.
- [18] LI Z, CHEN Y, MAI Z. Energy-efficient resource allocation with 3D beamforming in 3D MIMO-OFDMA systems[C]//IEEE International Conference on Communication Software and Networks, June 4-6, 2016, Beijing, China. New Jersey: IEEE Press, 2016: 370-374.
- [19] BOYD S, VANDENBERGHE L. Convex optimization[M]. Cambridge: Cambridge University Press, 2004.
- [20] DAHLMAN E. 3G evolution: HSPA and LTE for mobile broadband[M]. Pittsburgh: Academic Press, 2008.



- [21] WU Y, LI X, HEI Y. Downtilts adjustment and power allocation algorithm based on PSO for 3D MIMO systems[C]//IET International Conference on Information and Communications Technologies, April 27-29, 2013, Beijing, China. New Jersey: IEEE Press, 2013: 557-563.

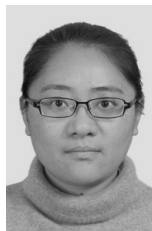


仇林杰 (1992-), 男, 南京邮电大学通信与信息工程学院硕士生, 主要研究方向为无线通信中的信号处理技术。

[作者简介]



李汀 (1979-), 男, 南京邮电大学通信与信息工程学院讲师, 主要研究方向为 MIMO 技术、3D MIMO 技术、协作通信等。



季薇 (1979-), 女, 南京邮电大学通信与信息工程学院副教授, 主要研究方向为无线通信、机器学习等。



基于深度学习的软件定义网络应用策略冲突检测方法

李传煌¹, 程成¹, 袁小雍², 岑利杰¹, 王伟明¹

(1. 浙江工商大学信息与电子工程学院, 浙江 杭州 310018;

2. 美国佛罗里达州大学大规模智能系统实验室, 美国 佛罗里达州 盖恩斯维尔 32611)

摘要: 在基于 OpenFlow 的软件定义网络 (SDN) 中, 应用被部署时, 相应的流表策略将被下发到 OpenFlow 交换机中, 不同应用的流表项之间如果产生冲突, 将会影响交换机的实际转发行为, 进而扰乱特定应用的正确部署以及 SDN 的安全。随着 SDN 规模的扩大以及需要部署应用的数量剧增, 交换机中的流表数量呈现爆炸式增长。此时若采用传统的流表冲突检测算法, 交换机将会耗费大量的系统计算时间。结合深度学习, 首次提出了一种适合 SDN 中超大规模应用部署的智能流表冲突检测方法。实验结果表明, 第一级深度学习模型的 AUC 达到 97.04%, 第二级模型的 AUC 达到 99.97%, 同时冲突检测时间与流表规模呈现线性增长关系。

关键词: 流表冲突检测; 深度学习; 异常检测; 软件定义网络; OpenFlow

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017305

Policy conflict detection in software defined network by using deep learning

LI Chuanhuang¹, CHENG Cheng¹, YUAN Xiaoyong², CEN Lijie¹, WANG Weiming¹

1. School of Information and Electrical Engineering, Zhejiang Gongshang University, Hangzhou 310018, China

2. LiLAB, University of Florida, Gainesville, Florida 32611, USA

Abstract: In OpenFlow-based SDN (software defined network), applications can be deployed through dispatching the flow policies to the switches by the application orchestrator or controller. Policy conflict between multiple applications will affect the actual forwarding behavior and the security of the SDN. With the expansion of network scale of SDN and the increase of application number, the number of flow entries will increase explosively. In this case, traditional algorithms of conflict detection will consume huge system resources in computing. An intelligent conflict detection approach based on deep learning was proposed which proved to be efficient in flow entries' conflict detection. The experimental results show that the AUC (area under the curve) of the first level deep learning model can reach 97.04%, and the AUC of the second level model can reach 99.97%. Meanwhile, the time of conflict detection and the scale of the flow table have a linear growth relationship.

Key words: policy conflict detection, deep learning, anomaly detection, SDN, OpenFlow

收稿日期: 2017-06-29; 修回日期: 2017-11-10

基金项目: 国家高技术研究发展计划 (“863”计划) 基金资助项目 (No.2015AA011901); 国家自然科学基金资助项目 (No.61402408, No.61379120); 浙江省自然科学基金资助项目 (No.LY18F010006); 浙江省重点研发计划基金资助项目 (No.2017C03058)

Foundation Items: The National High Technology Research and Development Program (863 Program) (No.2015AA011901), The National Natural Science Foundation of China (No.61402408, No.61379120), Zhejiang Provincial Natural Science Foundation of China (No.LY18F010006), Zhejiang's Key Project of Research and Development Plan (No.2017C03058)



1 引言

随着互联网业务和服务需求的不断增长,云计算网络、大型数据中心网络等已经成为承载各种大型计算和存储业务的重要网络基础设施,现有网络架构已经不能有效地满足这些新型网络平台的需求。作为下一代网络技术的重要代表,软件定义网络 (software defined network, SDN) 技术受到了学术界和产业界的极大关注,而其中基于 OpenFlow 的 SDN 技术更是得到了大家的青睐^[1,2],它已应用于如 Google 等互联网企业的大型数据中心网络中。

SDN 网络架构的重要特征是控制平面与转发平面的分离,在基于 OpenFlow 的 SDN 架构中,用户通过分离的控制器对网络中的 OpenFlow 交换机下发各种类型的转发策略,实现用户特定业务的部署,如防火墙、负载均衡、流量工程等。

SDN 应用/编排层根据某业务特征生成特定流表项时,所生成的策略可能与系统内已有的其他业务策略间产生冲突。如表 1 所示,策略 1 和 4、策略 2 和 3 以及策略 5 和 6 之间都存在重叠内容,且不同策略的行为互相抵触,这些策略间都存在冲突问题。

当对一个大型的基于 OpenFlow 的 SDN 中的所有业务进行自动部署和控制时,如果不对 OpenFlow 交换机中的流表进行冲突检测,一旦 OpenFlow 控制器对 OpenFlow 交换机添加的某些策略与之前已有的策略间存在冲突,将使得一些已存在的流表项被更改或覆盖而失效,最终对特

定业务产生致命的影响,如,当前策略冲突中有防火墙应用,则会使某些本被防火墙阻止的数据分组顺利通过,甚至给基于 OpenFlow 的 SDN 带来安全隐患,也不利于大型网络中 SDN 业务的自动部署。

图 1 展示了一个策略冲突引起防火墙功能失效的例子,该例子的拓扑图中包含了 4 台 PC 机、1 台 OpenFlow 交换机和 1 台 OpenFlow 控制器。PC 机 A 的 IP 地址为 192.168.1.10, PC 机 B 的 IP 地址为 192.168.2.2, PC 机 C 的 IP 地址为 192.168.3.7, PC 机 D 的 IP 地址为 192.168.6.5, 4 台 PC 机分别与 OpenFlow 交换机相连组成一个局域网。OpenFlow 交换机中的 Flow Table 0 中存在一条防火墙规则,该防火墙的规则是屏蔽源 IP 地址为 192.168.1.0/24、目的 IP 地址为 192.168.6.0/24 的数据分组。此时,OpenFlow 控制器即将下发一条策略,该策略是将目的 IP 地址和源 IP 地址都为 192.168.0.0/16 的所有数据分组设置一个队列 ID。如果把该策略下发到 OpenFlow 交换机中,所有源 IP 地址为 192.168.1.0/24 和目的 IP 地址为 192.168.6.0/24 的数据分组的动作行为将会冲突,从而导致原有防火墙规则的失效。

OpenFlow 相关协议本身并未提供流表冲突检测的处理机制,当前某些 OpenFlow 交换机具有简单的流表冲突检测功能,其流表冲突检测功能采用了基于传统的算法查找 OpenFlow 交换机中冲突的流表项的方法。如果 OpenFlow 交换机中的流表数目非常庞大,使用传统的流表冲突检测算法会增加 OpenFlow 交换机的运行负担、延缓流表的部署,影响数据分组正常的转发。

表 1 存在冲突的流表策略

规则	源 MAC 地址	目的 MAC 地址	源 IP 地址	目的 IP 地址	动作
1	*	*	10.20.0.1/8	10.20.5.100	drop
2	*	*	*	10.20.1.1/8	limit 8 Mbit/s
3	*	*	10.20.100.1/8	*	limit 1 Mbit/s
4	*	*	10.20.0.100	*	output port 10
5	14:10:9f:ea:7a:fd	*	*	*	*
6	*	14:10:9f:ea:50:3e	*	*	*

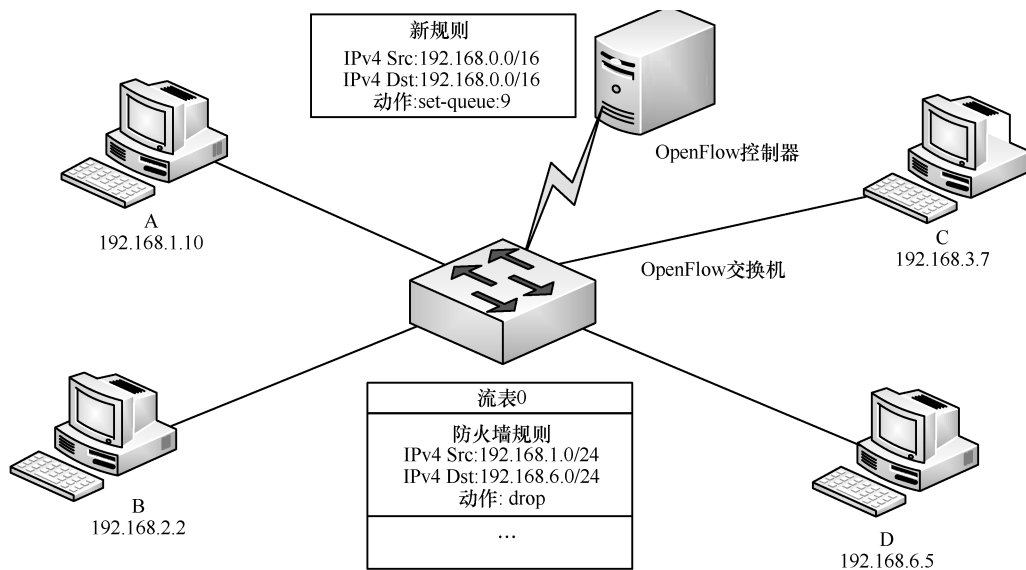


图1 策略冲突引起防火墙功能失效

当某个应用在 SDN 中进行具体部署时,若此应用需同时向多个 OpenFlow 交换机下发不同的策略,当其中一个交换机中检测出有流表冲突时,为避免对已有应用产生影响,SDN 控制器需要撤销当前应用下发的流表操作,并将对其他交换机已完成的操作进行回滚,然后将处理结果通知应用/编排层。显然因为冲突检测被放置在交换机内,给部署操作带来了冗余,增加了部署时间,也给整个部署过程带来了更多的不确定性,所以在控制器下发策略之前进行检测,预防冲突策略显得十分必要。

本文首次将 SDN 与深度学习结合,提出一种部署于控制器上的基于深度学习的 SDN 流表冲突检测方法。该方法利用深度学习抽象化高层数据、自动学习的特点,相比于传统的冲突查找算法,在大规模应用部署时能更快速地对超大规模的流表项做出是否冲突的检测。

2 相关工作

OpenFlow 协议^[3]中没有给出流表冲突的具体解决策略,仅有如下描述:如果对流表中 OFPFF_CHECK_OVERLAP 位进行了设置,

OpenFlow 交换机必须首先检测新下发的流表与交换机中已经存在的流表是否存在重复。当前对 OpenFlow 流表冲突检测的研究,都是基于传统的查找算法。

参考文献[4]提出了通过定义策略简化网络的管理,同时提出了相关算法,可用在传统网络的冲突策略和不可达策略的检测中;参考文献[5,6]提出了一种基于“first-order”逻辑来定义可能冲突的方法来检测冲突,并且在 OpenFlow 控制器下发流表前部署了“推理引擎”来检测流表冲突。

参考文献[7]介绍了一种通过使用 FlowVisor 技术来检测 OpenFlow 网络中潜在的冲突问题,FlowVisor 是一种有着特殊用处的 OpenFlow 控制器,它可以在 OpenFlow 交换机和多台 OpenFlow 控制器之间充当一个代理,他们提出的冲突检测算法是:基于散列字典树(hash-trie based conflict detection, HTCD)算法及基于冲突检测的本体论(ontology based conflict detection, OCD)。

参考文献[8]为 SDN 提出了一种减少比特向量(RBV reduced bit vector)算法对流表进行冲突检测。RBV 算法用到了比特矢量的概念并且采用



了按组分类的算法,将相同前缀长度的流表分到一个组中,从而减少了多余的比特矢量。

参考文献[9]提出了安全策略执行内核(FortNOX)作为对开放源 NOX OpenFlow 控制的一个扩展, FortNOX 采用了一种基于别名检测的方法来确定每一个应用是否都被安全授权,强制优先级最低的策略进入仲裁。

参考文献[10]提出了称为“FlowGuard”的综合框架,应用在动态的 OpenFlow 网络中检测和解决与防火墙相冲突的策略, FlowGuard 在网络状态更新时会对网络中的流表进行检测是否与防火墙策略相冲突。

参考文献[11]提出全网内所有转发规则在进行插入、更新和删除操作时,使用 VeriFlow 动态地检测违反规则的流表,他们在传统分组分类算法的基础上提出了多维前缀树,一棵多维前缀树是一种用来存储组合数组的有序的树状数据结构,这棵树与符合流表规则的数据分组相关联;参考文献[10]提出了一种形式化的模型来检测内部功能是否有异常,这里异常的定义是在同一个网络中,是否有两个或更多的相同功能被部署。

参考文献[13]提出了使用 ADRS (anomaly detecting and resolving for SDN) 解决策略和规则中存在的异常,采用了间隔树模型快速扫描流表,然后建立一个共享模型来分配网络的优先级,通过这两个模型,提出了一种自动化的算法来检测和解决

SDN 模型中存在的流表异常问题。

参考文献[14]提出了一种基于 FlowPath 的实时动态策略冲突检测及解决办法,通过获取实时的 SDN 状态检测防火墙策略的直接和间接冲突,基于 FlowPath 进行自动化和细粒度的解决冲突。

在以上所有相关研究中,未有对超大规模流表项进行冲突检测的研究及相关实验结果,也无人提出使用与机器学习相关的方法进行检测,本文首次提出一种基于深度学习的适应于超大规模流表项冲突检测的方法。

3 基于深度学习的冲突检测

3.1 冲突检测架构

当应用/编排层部署新的应用时,需检测下发到某一交换机内的策略是否与之前已存在的流表项存在冲突,使用深度学习进行检测时,以需要检测的所有流表项的域作为模型特征。由于 SDN 中的交换机内的流表项数目将随着应用的增长而不断增长,深度学习难以对非固定大小的输入进行检测,采用基于固定窗口的循环检测方法,检测架构如图 2 所示。

通常应用/编排层(或控制器)中维护着 SDN 中所有交换机的流表信息,假设当前应用产生了某条策略,需要下发到某一 OpenFlow 交换机,此时将该交换机内获取的所有流表信息按每 X 条流表项来分割,当有一条新的流表项需要下发时,将这条流表项与已分割的 X 条流表项组成一个

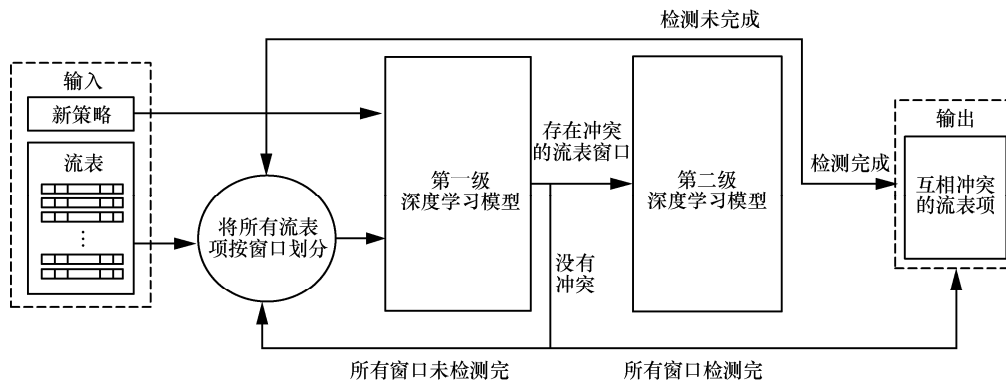


图2 流表冲突检测架构

$X+1$ 条流表项的窗口, 作为第一级深度学习模型的输入数据。其中, 第一级深度学习模型的输入数据是 $X+1$ 条流表项的所有信息特征。如果 OpenFlow 控制器中已存在的流表项数目不足 X 条, 则用无效的填充信息将窗口填充到 X 条流表项。如果分割后的片段中的流表项数目不足 X 条, 则用其他窗口中的流表项将此窗口的流表项数目填充到 X 条。分别将所有的窗口作为第一级深度学习模型的输入来判断这个窗口中的所有流表项是否存在冲突。如果存在冲突, 则将该窗口的每条流表项与即将下发的流表项组成一个窗口, 作为第二级深度学习模型的输入数据来定位具体的两条冲突的流表项。其中, 第二级深度学习模型是以两条流表项的所有特征信息作为该模型的输入数据。

3.2 深度学习模型

根据 OpenFlow1.5 协议, OpenFlow 匹配域共包括 45 个字段, 取其中的 ETH_DST、ETH_SRC、ETH_TYPE、IPv4_SRC、IPv4_DST、IPv4_SRC_MASK 和 IPv4_DST_MASK 作为深度学习模型的输入特征。其中, ETH_DST 为目的 MAC 地址, ETH_SRC 为源 MAC 地址, ETH_TYPE 为以太网协议类型, IPv4_SRC 为源 IP 地址, IPv4_DST 为目的 IP 地址, IPv4_SRC_MASK 为源 IP 地址的子网掩码, IPv4_DST_MASK 为目的 IP 地址的子网掩码。最终, 这些字段的值将会被转换成二进制值作为深度学习模型的输入特征。

第一级深度学习模型的输入层为 CNN 层, 隐藏层由 M 层 CNN 层和 N 层全连接层组合, 输出层为只有一个神经元的全连接层。输入层和隐藏层使用的激活函数为 ReLU (rectified linear unit) 函数, 该函数为非线性函数, 函数为:

$$f(x) = \max(0, x) \quad (1)$$

输出层的激活函数为非线性激活函数 Sigmoid 函数, 为:

$$f(x) = 1 / (1 + e^{-x}) \quad (2)$$

为了简化第二级深度学习模型, 并取效果最好的一个模型, 将第二级深度学习模型总共分为两类: 一类模型包括 CNN 层和全连接层; 另一类模型只有全连接层。前者的输入层为 CNN 层, 隐藏层都为全连接层, 输出层为只有一个神经元的全连接层。输入层和隐藏层的激活函数都为 ReLU 函数, 输出层的激活函数为 Sigmoid 函数。后者的输入层、隐藏层和输出层都为全连接层, 激活函数的分布情况与前者相同。

第一级深度学习模型和第二级深度学习模型的总体结构分别如图 3、图 4 所示。

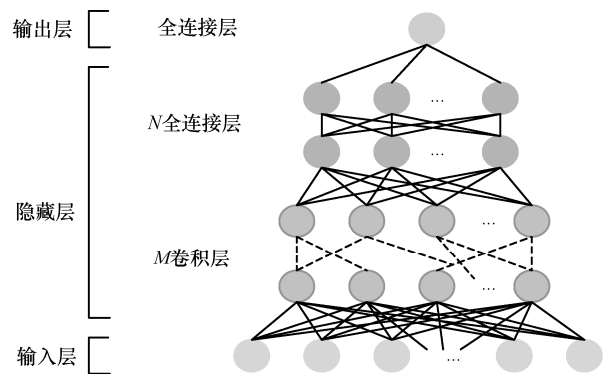


图3 第一级深度学习模型总体结构

本文将使用的深度学习模型结构用如下方式描述:

$$Model = (I, (K_0, K_1, \dots, K_i, \dots, K_n), [T, O]) \quad (3)$$

其中, $K_i = [T, L, n_0, n_1, \dots, n_i, \dots, n_L]$ 。“ I ”表示输入层神经元的个数; “ T ”表示层的类型, 卷积层用“C”表示, 全连接层用“F”表示; “ L ”表示层的数量; “ n_i ”表示隐藏层中每一层的神经元个数; “ O ”表示输出层神经元的个数。

模型训练的特征个数越多, 达到高精度模型所需的样本数量就越多。由于第一级深度学习模型具有的输入特征个数有 24 000 个, 特征数较多, 所以训练第一级深度学习模型时采用 1 200 000 个样本的数据集。为了获取最佳的 SDN 策略冲突检测的第一级深度学习模型, 本文构建了多个深度学习模型进行比较。

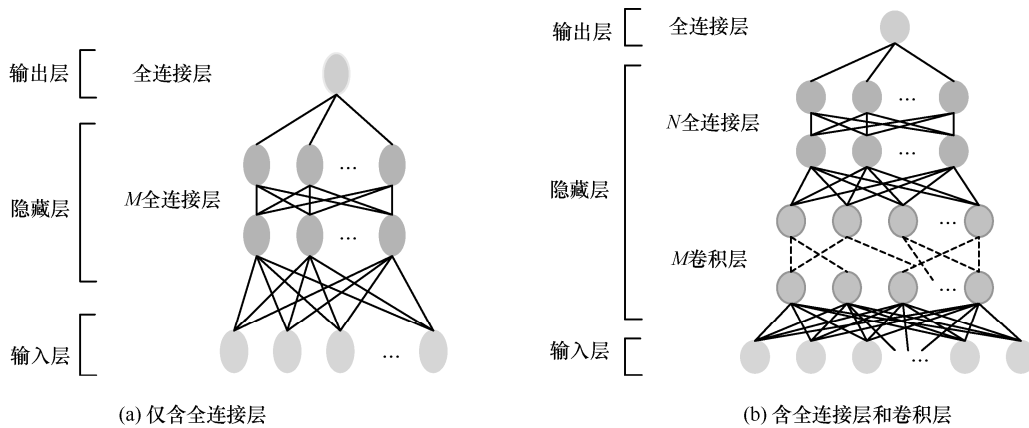


图4 第二级深度学习模型结构

本文进行训练时所使用的的第一级深度学习模型如下：

1>C3F4_1=(24 000,["C",3,512,256,128],["F",3,128,128, 128],["F",1]);

2>C3F4_2=(24 000,["C",3,128,128,64],["F",3,64,64,64],["F",1]);

3>C3F4_3=(24 000,["C",3,128,128,64],["F",3,64,32,16],["F",1]);

4>C3F3=(24 000,["C",3,128,64,32],["F",2,32,16],["F",1]);

5>C3F2=(24 000,["C",3,80,80,64],["F",1,64],["F",1]);

6>C3F5=(24 000,["C",3,128,128,128],["F",4,128,64,32,16],["F",1]);

7>C4F4=(24 000,["C",4,128,128,64,64],["F",3,64,32,16],["F",1]);

8>C5F2=(24 000,["C",5,128,64,32,16,8],["F",1,8],["F",1])。

由于第二级深度学习模型具有的输入特征个数为 480 个，特征数较少，因此训练第二级深度学习模型时采用了 400 000 个样本的数据集，且为了获取最佳的第二级深度学习模型，本文构建了两种不同类型的多个深度学习模型进行比较：一种类型是最简单的模型，模型中所有层都为全连接层；另一种类型是稍微复杂的模型，模型中包括卷积层和全连接层。

根据模型的描述方式，本文进行训练的不同类型的多个第二级深度学习模型的形式如下：

1>F3=(24 000,(["F",2,256,128]),["F",1]);

2>F4=(24 000,(["F",3,256,128,64]),["F",1]);

3>C1F4=(24 000,(["C",1,512],["F",3,512,256,128]),["F",1]);

4>C1F3=(24 000,(["C",1,256],["F",2,256,128]),["F",1])。

4 实验实现与评估

4.1 数据集产生和预处理

当前使用流表项中的 ETH_DST、ETH_SRC、IP_PROTO、IPv4_SRC、IPv4_DST、IPv4_SRC_MASK 和 IPv4_DST_MASK 字段作为每一条流表项的特征。第一级深度学习模型输入的是 100 条流表项的特征信息，所以对于第一级深度学习模型的单位模型输入特征包括了 100 条流表项的所有特征信息。第一级深度学习模型的数据集包括了两类数据，一类为具有冲突流表项的数据，另一类为没有冲突流表项的数据。第一级深度学习模型的数据集产生算法如下。

(1) 产生冲突类型的流表项数据

flow_entry_1 = GenerateRandomFlowEntry()

flow_entry_2 = GenerateConflictFlowEntry
(flow_entry1)

```

while true:
    flow_entry_i = GenerateRandomFlowEntry()
    if isConflict(flow_entry_i, conflict_
flow_entries) == False THEN
        conflict_flow_entries.append(flow_
entry_i)
        break
    end if
number of conflict_flow_entries: m
if m < 100 then
    go to 3
end if
Shuffle (conflict_flow_entries)
    conflict_sample.append(conflict_
flow_entries)
number of conflict_sample: n
if n < 600 000 then
    go to 1
end if

```

上述算法中一个冲突样本的生成方法：先随机产生一条流表项，后根据冲突规则生成一条与之冲突的流表项，再生成其余与上两条流表项均不冲突的 98 条流表项，最后将上述 100 个流表项顺序打乱。按此方法生成 600 000 个冲突样本。

(2) 产生非冲突类型的流表项

```

flow_entry_1 = GenerateRandomFlowEntry()
while true:
    flow_entry_i = GenerateRandomFlowEntry()
    if isConflict(flow_entry_i, non_ con-
flict_flow_entries) == false then
        non_conflict_flow_entries.append(flow_entry_i)
        break
    end if
number of non_conflict_flow_entries: m
if m < 100 then
    go to 2

```

```

end if
Shuffle(non_conflict_flow_entries)
    non_conflict_sample.append(flow_entries)
number of non_conflict_sample: n
if n < 600000 then
    go to 1
end if

```

一个非冲突流表项的生成的过程与冲突样本生成的过程类似，但生成的规则不同，按此方法生成 600 000 个非冲突样本。

在对使用上述算法产生的流表项进行预处理时，分别将冲突类型和非冲突类型的流表项数据转换成二进制形式，并将每个冲突类型的样本打上标签 1，将每个非冲突类型的样本打上标签 0。最后将冲突类型和非冲突类型的流表项数据合并，并且打乱顺序，最后的数据集含 1 200 000 个样本，冲突和非冲突占比为 1:1。在第 4.2 节所述的实验环境下，生成第一级模型的训练样本时间约为 5 h。

第二级深度学习模型输入的是 2 条流表项的特征信息，所以对于第二级深度学习模型的单位模型输入特征包括了 2 条流表项的特征信息。第二级深度学习模型的数据集同样包括两类数据，即具有冲突流表项的数据和没有冲突流表项的数据。第二级深度学习模型的数据集产生算法与第一级深度学习模型的数据集产生流程相似。使用了用该算法产生的 400 000 个样本对第二级模型进行训练和测试，其中冲突和非冲突占比为 1:1。第二级模型的训练样本生成时间约为 15 min。

4.2 实验结果

本文涉及的实验环境如下：两块 Nvidia K80 的显卡，128 GB 的内存，操作系统为 Ubuntu14.04，使用了基于 Keras 的深度学习框架。

针对表 2 不同的第一级深度学习模型，使用含 1 200 000 个样本的数据集进行训练，最后的测



表 2 第一级深度学习模型的测试精度

参数名	C3F4_1	C3F4_2	C3F4_3	C3F3	C3F2	C4F4	C5F2
filter_length1	3	3	3	2	5	5	2
filter_length2	2	2	2	2	3	3	2
filter_length3	2	2	2	2	3	3	2
filter_length4	/	/	/	/	/	2	2
filter_length5	/	/	/	/	/	/	2
训练精度	90.06%	90.03%	91.23%	90.82%	90.68%	88.90%	89.38%
测试精度	90.14%	91.20%	91.46%	90.84%	90.24%	89.60%	89.24%
AUC	96.45%	96.91%	97.04%	97.00%	96.61%	96.10%	96.39%

表 3 第二级深度学习模型的测试精度

参数名	F3	F4	C1F4	C1F3
filter_length	/	/	1	1
训练精度	98.53%	98.21%	98.50%	97.77%
测试精度	99.35%	98.35%	98.59%	98.60%
AUC	99.97%	99.96%	99.80%	99.93%

试精度见表 2。通过表 2 可以看出, C3F4_3 模型是测试精确度最高的模型, 因此本文采取的第一级深度学习模型是 C3F4_3 模型。

针对表 3 不同的第二级深度学习模型, 使用含 400 000 个样本的数据集进行训练, 最后的测试精度见表 3。通过表 3 可以看出 F3 模型是测试精度最高的模型, 因此本文采取的第二级深度学习模型是 F3 模型。

4.3 实验评估

本文对流表冲突检测的深度学习模型与机器学习中的 RandomForest 分类器^[13]进行了比较。在与第一级冲突检测相同的硬件环境下, 基于 Ubuntu 14.04 及 scikit-learn 机器学习框架, 使用相同的数据集, 使用 RandomForest (随机森林) 分类器进行了实验, 与深度学习的结果比较见表 4。

从该结果可以看出, 训练相同的样本数量, 深度学习方法的测试精度比其他机器学习方法的

表 4 深度学习方法和机器学习方法的比较

样本个数/个	深度学习(C3F4)	随机森林
400 000	89.56%	62.7%
1 200 000	91.46%	内存溢出

测试精度高。其中, 随机森林分类器在训练 1 200 000 个样本数据时, 由于内存溢出而无法获得具体的精度。

同时对基于深度学习的检测方法的检测时间进行了测试。在实验过程中, 在不同大小的原始流表项作为检测对象时, 第一级深度学习模型的检测时间见表 5。

第二级深度学习模型定位出具体冲突流表项 (即检测 100 个第二级深度学习模型的窗口) 的时间为 0.206 s。冲突检测总时间为第一级深度学习模型和第二级深度学习模型检测时间之和。当 OpenFlow 控制器中所有的流表项与即将下发的流表项没有冲突流表项时, 则无需使用第二级深

表 5 第一级深度学习模型的检测时间

流表项个数/个	1 000	10 000	25 000	50 000	100 000	200 000
检测时间/s	0.341	0.607	0.957	1.677	2.906	5.631

表 6 first-order logic 方法下不同数量流表项冲突检测时间

流表项个数/个	1 000	10 000	25 000	50 000	100 000
检测时间/s	0.010	0.172	0.817	9.079	257.675

深度学习模型进行检测,该级时间为 0;当 OpenFlow 交换机中的某一条流表项与即将下发的流表项有一个冲突时,第二级深度学习模型冲突检测时间为 0.206 s。当与 n 个第一级深度学习模型窗口内的流表项有冲突时,第二级检测时间为 $n \times 0.206$ s,总时间随着流表项的大小呈线性增长。

参考文献[5,6]中使用的 first-order logic 算法对流表冲突检测时的检测时间见表 6。

使用两种检测方法所耗费的总时间比较如图 5 所示。从图 5 中可以看出,在对小规模数量的流表项进行冲突检测时,参考文献[5]使用的基于 first-order logic 方法,与基于深度学习方法相比,检测时间没有太大区别;而在对大规模数量的流表项进行流表冲突检测时,参考文献[3]所使用的 first-order logic 算法,检测时间呈指数级增长,而深度学习所需时间则呈线性增长,远小于传统算法的检测时间。

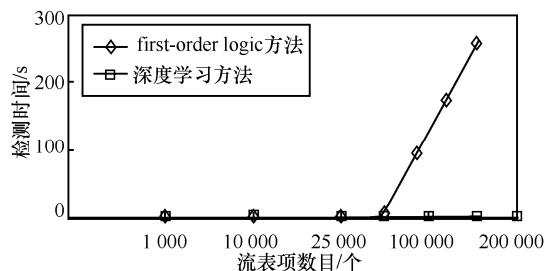


图 5 深度学习方法与 first-order logic 方法冲突检测时间比较

参考文献[6]提出使用 HTCD 算法及 OCD 算法对冲突进行检测。前者的平均算法时间复杂度为 $O(nd)$, 后者为 $O(n)$ 。其中, n 为流表项总数, d 为数据分组头检测域的个数。而本文提出的基于

深度学习的冲突检测方法,如采用窗口分割方法将所有 n 条流表项分割成每个窗口含 k ($k > 100$) 条流表项的 m 个窗口,则在任何情况下时间复杂度均为 $O(n/k)$ 。

参考文献[14]中提出的基于 FlowPath 的 SDN 安全策略动态冲突检测方法。在一个有 h 台主机、1 台交换机、平均每台交换机中的流表项数为 n 的情况下。找出全网的所有冲突的算法查找次数为:

最坏情况: $h \times n \times 1$ 次比较 (全部搜索);

平均情况: $h \times (n/2) \times (1/2)$ 次比较;

最佳情况: h 次比较 (比较时一次命中)。

本文采用两级深度学习的冲突检测算法,要完成对所有流表项的冲突检测,第一级只需完成对这 m 个窗口的基于深度学习模型的检测,检测次数为 $m = \lceil n/k \rceil + 1$,第二级检测只有在定量检测,并检测到某个窗口存在冲突时才会被启动,如果仅定性检测某流表是否与系统内流表存在冲突,则 $m = \lceil n/k \rceil + 1$ 次即最坏情况,该值远小于参考文献[14]中提到的平均检测次数及最坏次数。

5 结束语

本文通过将深度学习与基于 OpenFlow 的 SDN 相结合,提出了一种新的用来检测基于 OpenFlow 的 SDN 中的流表冲突的方法。该方法采用第一级深度学习模型定性判断在所有流表中是否存在冲突,并采用第二级深度学习模型定位出具体冲突的流表项。通过实验数据可以看出,当 OpenFlow 网络中存在大量的流表时,本文提出



的方法在检测冲突流表项所花费的时间上远小于传统的流表冲突检测方法。因此,这种新方法适合于对超大规模流表进行冲突检测。

但是,任何算法都不能保证达到 100% 的检测精确度,基于深度学习的算法也存在同样的问题,该方法的精确度与各种因素有关,如训练模型时所使用的数据集的大小、数据集的特征选择、深度学习模型的层数及每层所使用的模型等,在后续的研究过程中,将朝此方向努力,使该方法的精度能得到进一步的提高。

参考文献:

- [1] 韦乐平. SDN 的战略性思考[J]. 电信科学, 2015, 31(1): 7-12.
WEI L P. Strategic Thinking on SDN [J]. Telecommunications Science, 2015, 31(1): 7-12.
- [2] CLOUGHERTY M M, WHITE C A, VISWANATHAN H, et al. SDN 在 IP 网络演进中的作用[J]. 电信科学, 2014, 30(5): 1-13.
CLOUGHERTY M M, WHITE C A, VISWANATHAN H, et al. Role of SDN in IP network evolution[J]. Telecommunications Science, 2014, 30(5): 1-13.
- [3] SPECIFICATION O F S. Version 1.5. 1(wire protocol 0x01)[EB]. 2015.
- [4] VERMA D C. Simplify network administration using policy-based management[J]. Network IEEE, 2002, 16(2): 20-26.
- [5] LOPES A B B, LIMA DE C G A, FERNANDEZ M P. Flow-based conflict detection in OpenFlow networks using first-order logic[C]//2014 IEEE Symposium on Computers and Communication (ISCC), April 27-May 2, 2014, Toronto, Japan. New Jersey: IEEE Press, 2014: 1-6.
- [6] BATISTA B L A, DE CAMPOS G A L, FERNANDEZ M P. A proposal of policy based OpenFlow network management[C]//2013 20th International Conference on Telecommunications (ICT), Aug 14-16, 2013, Guilin, China. New Jersey: IEEE Press, 2013: 1-5.
- [7] NATARAJAN S, HUANG X, WOLF T. Efficient conflict detection in flow-based virtualized networks[C]//2012 International Conference on Computing, Networking and Communications (ICNC), Oct 8-10, 2012, Maui, Hawaii. New Jersey: IEEE Press, 2012: 690-696.
- [8] LO C C, WU P Y, KUO Y H. Flow entry conflict detection scheme for software-defined network[C]//2015 International Telecommunication Networks and Applications Conference(ITNAC), November 18, 2015, Sydney, Australia. New Jersey: IEEE Press, 2015: 220-225.
- [9] PORRAS P, SHIN S, YEGNESWARAN V, et al. A security enforcement kernel for OpenFlow networks[C]//The First Workshop on Hot Topics in Software Defined Networks, August 13, 2012, Helsinki, Finland. New York: ACM Press, 2012: 121-126.
- [10] HU H, HAN W, AHN G J, et al. FLOWGUARD: building robust firewalls for software-defined networks[C]//The Third Workshop on Hot Topics in Software Defined Networking, August 22, 2014, Chicago, Illinois, USA. New York: ACM Press, 2014: 97-102.
- [11] KHURSHID A, ZOU X, ZHOU W, et al. VerIFlow: verifying network-wide invariants in real time[C]//10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13), Apr 3-5, 2013, Lombard, IL, USA. [S.l.:s.n.], 2013: 15-27.
- [12] BASILE C, CANAVESE D, LIOY A, et al. Inter - function anomaly analysis for correct SDN/NFV deployment[J]. International Journal of Network Management, 2016, 26(1): 25-43.
- [13] WANG P, HUANG L, XU H, et al. Rule anomalies detecting and resolving for software defined networks[C]//2015 IEEE Global Communications Conference (GLOBECOM), Dec 6-10, 2015, San Diego, USA. New Jersey: IEEE Press, 2015: 1-6.
- [14] 王娟, 王江, 焦虹阳, 等. 一种基于 OpenFlow 的 SDN 访问控制策略实时冲突检测与解决办法[J]. 计算机学报, 2015, 38(4): 872-883.
WANG J, WANG J, JIAO H Y, et al. A method of OpenFlow-based real-time conflict detection and resolution for SDN access control policies[J]. Chinese Journal of Computers, 2015, 38(4): 872-883.
- [15] LIAW A, WIENER M. Classification and regression by randomForest[J]. R News, 2002, 2(3): 18-22.

[作者简介]



李传煌 (1980—), 男, 博士, 浙江工商大学信息与电气工程学院副教授, 2016 年美国佛罗里达大学访问学者, 主要研究方向为软件定义网络、深度学习、开放可编程网络、系统性能预测和分析模型, 发表 EI/SCI 检索论文 40 余篇, 申请专利 15 项。

程成 (1993—), 男, 浙江工商大学信息与电气工程学院硕士生, 主要研究方向为软件定义网络、深度学习。

袁小雍 (1990—), 男, 美国佛罗里达大学博士生, 主要研究方向为网络安全、深度学习、云计算和分布式系统。

岑利杰 (1992—), 男, 浙江工商大学信息与电气工程学院硕士生, 主要研究方向为网络安全、深度学习、软件定义网络。

王伟明 (1964—), 男, 博士, 浙江工商大学信息与电子工程学院教授, 主要研究方向为新一代网络架构、开放可编程网络, 特别是 IETF ForCES、SDN 及可重构网络等方面的协议、模型和算法。



高速铁路宽带无线通信系统越区切换技术

刘云毅¹, 赵军辉^{1,2}, 王传云²

(1. 北京交通大学电子信息工程学院, 北京 100044;

2. 华东交通大学信息工程学院, 江西 南昌 330013)

摘要: 我国现有的铁路移动通信系统已不能满足铁路信息化建设的要求, 需要开展下一代高速铁路通信系统关键技术研究。越区切换是高速铁路宽带无线通信系统的关键技术, 对保障列车的行车安全和通信可靠性具有重要意义。针对目前高速铁路通信系统中越区切换技术的研究现状和成果进行总结, 分析存在的问题, 并进一步探讨未来越区切换技术的发展方向。

关键词: 高速铁路; 越区切换; 无线通信

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017278

Handover technology in high-speed railway broadband wireless communication system

LIU Yunyi¹, ZHAO Junhui^{1,2}, WANG Chuanyun²

1. School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China

2. School of Information Engineering, East China Jiaotong University, Nanchang 330013, China

Abstract: The existing railway mobile communication system can't meet the requirements of railway information construction. Therefore, the key technology research of next generation of high-speed railway (HSR) communication system needs to be carried out. Handover is a key technology of HSR broadband wireless communication system, which has great significance to ensure the train safety and communication reliability. The research results of the handover in HSR communication system were summarized and the existing problems were investigated. Furthermore, the development directions of handover in the future were explored.

Key words: high-speed railway, handover, wireless communication

1 引言

近年来, 中国高速铁路(以下简称高铁)迅

猛发展, 高铁以其经济辐射效应极大地带动了沿线城市和地区的经济发展, 形成了我国独具特色的“高铁经济”。伴随着“一带一路”“高铁走

收稿日期: 2017-08-01; 修回日期: 2017-09-26

基金项目: 国家自然科学基金资助项目(No.61661021, No.61471031); 东南大学移动通信国家重点实验室开放研究基金资助项目(No.2017D14); 国家科技重大专项基金资助项目(No.2016ZX03001014-006)

Foundation Items: The National Natural Science Foundation of China (No.61661021, No.61471031), The Open Research Fund of National Mobile Communications Research Laboratory, Southeast University (No.2017D14), National Science and Technology Major Project of the Ministry of Science and Technology of China (No.2016ZX03001014-006)



出去”和“中国制造 2025”等国家战略的提出,高铁建设已经成为我国中长期发展的战略需要。截止到 2016 年底,我国高铁营业里程超过 2.2 万千米,稳居世界第一,占世界高铁总里程的 60%以上。2017 年 6 月 26 日,“复兴号”中国标准动车组在京沪高铁双向首发,标志着我国高铁正式进入自主知识产权时代,达到世界领先水平。根据我国《中长期铁路网规划》(2016—2030 年),预计到 2030 年,我国高速铁路网基本实现内外互联互通、区际多路畅通、省会高铁连通、地市快速通达、县域基本覆盖。图 1 为 2013—2016 年高速列车与普通列车全国日均客运发送量对比。高速列车的客运发送量及其所占比重逐年增大,在 2016 年占比达到 52.1%。随着郑徐高铁、宝兰高铁、西成高铁等新一批高铁线路开通运营,这一比重在 2017 年会继续扩大。

高速铁路的快速发展对移动通信系统也提出了极高的要求,高速铁路宽带无线通信系统需要保障高速列车在运行过程中调度信息的可靠传输以及为旅客提供更为优质的宽带通信业务。随着智能设备的普及以及互联网的高速发展,旅客在长途旅行中对各种无线服务有了更高的需求。由于列车高速运行引起的多普勒频移和信道快速时变、严重的车厢穿透损耗、大量设备同时切换引起的“信令风暴”以及用户终端处理能力有限等

问题,使得现有的高速铁路旅客无线接入系统远远不能满足旅客对服务质量 (quality of service, QoS) 的需求^[1]。此外,有限的系统容量也无法满足视频监控、多媒体应用等各种宽带服务的需求。因此,在设计高速铁路宽带无线通信系统时,如何提高 QoS 和系统容量是一个非常迫切的问题。

国际铁路联盟 (International Union of Railways, UIC) 确定将铁路宽带移动通信系统 (long term evolution for railway, LTE-R) 作为下一代铁路通信系统^[2]。LTE-R 可在 20 Mbit/s 带宽的条件下为高速列车提供 100 Mbit/s 的数据传输速率,其端到端呼叫建立时延小于 350 ms,话权抢占时延小于 200 ms,是当前铁路通信系统的研究热点^[3]。考虑智能轨道基础设施、智能移动管理、智能轨道交通服务等未来高速铁路发展需求,一些学者提出 5G-R (fifth generation for railway) 的概念,利用大规模 MIMO (multiple input multiple output)、毫米波 (mmWave)、波束成形、频谱融合等技术为未来的高铁通信系统提供超高传输容量、超高可靠性、超低时延等高质量服务^[4]。

2 越区切换

越区切换是高速铁路无线通信系统中的关键技术,在铁路通信中占有极其重要的地位。在列车高速通过小区重叠区时,性能良好的越区切换

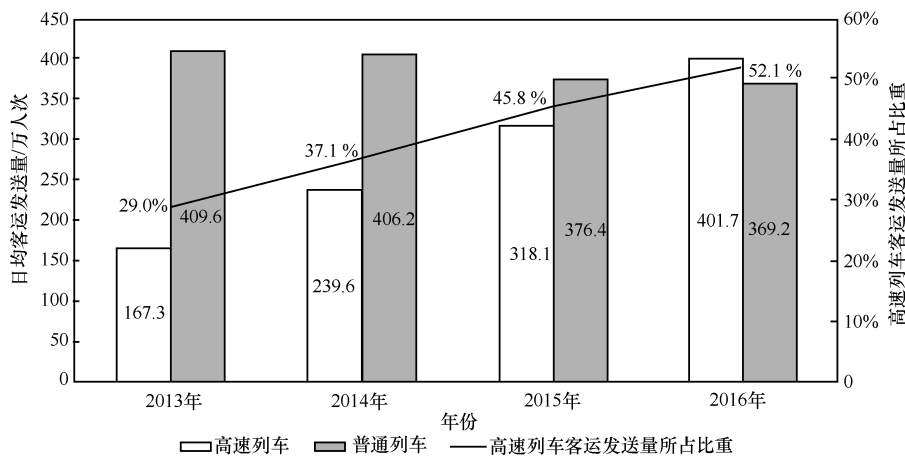


图 1 2013—2016 年全国铁路日均客运发送量

方案能够保障列车控制信息和旅客通信业务的可靠性、实时性和连续性^[5]。越区切换是指当移动终端从一个服务基站的覆盖范围移动到相邻基站覆盖范围时,改变原有的通信链路,切换到其他基站上从而保持通信连续的过程。如果切换失败,将会导致通信中断现象,严重影响用户体验,对列车安全造成不利影响。高速列车特殊的运行环境以及较强的移动性,对切换方案的性能提出了更高的要求,然而现有切换方案在高铁场景中面临严峻挑战。

(1) 切换失败率高

由于多普勒频移、信道快速时变、复杂地形环境以及车厢穿透损耗(CRH3列车的车厢穿透损耗高达24 dB^[6])等不利因素的影响,使得列车中用户设备的接收信号质量较差,降低切换成功概率,甚至引发“乒乓切换”效应,严重影响高铁乘客的用户体验。

(2) 切换频繁

2017年9月21日,“复兴号”动车组在京沪高铁正式开始以350 km/h的速度运营。随着列车速度的提高,切换将愈发频繁。假设小区半径为3 km,重叠区长度为1.2 km,那么平均每48 s列车就需要切换,在重叠区驻留的时间仅为12 s,所以要优化切换方案及流程,设计符合高速铁路环境的快速切换方案。

(3) 群切换

当列车进入新小区时,列车中的移动终端将同时执行切换。“复兴号”动车组满载时约有1 110名乘客,假设10%的乘客移动终端处于活动状态,应同时处理110个用户设备的切换请求,在基站控制面中产生大量的信令交互,从而引起“信令风暴”,带来巨大的系统开销。

(4) QoS 保证

高速列车上各种移动业务的QoS由于频繁切换而降低。例如,视频电话的分组时延应小于50 ms,而切换时间通常需要100 ms。如果通信在切换期

间中断,则无法满足视频电话的QoS要求。

3 切换方案优化设计

3.1 基于参数优化的切换方案设计

越区切换一般分为3个阶段:切换测量、切换判决以及切换执行。切换测量由基站和移动台相互协作完成,基站向移动台发送测量控制信令,移动台周期性上报测量结果,基站根据接收到的测量报告,当满足特定的条件时,切换就会被触发。针对切换过程中涉及的各种参数进行优化,提升切换性能,是一大研究方向。

(1) 基于地理位置信息的切换方案优化设计

高铁通信基站一般沿铁路呈带状分布,而高速列车沿铁轨固定行驶的特点,为切换优化提供了便利,一些学者考虑借助地理位置信息来优化切换方案。在参考文献[7]中,当列车进入目标基站的覆盖区域时,根据列车的速度自适应地决定切换触发条件。如果列车高速移动,用户设备可以提前触发切换。在参考文献[8]中使用全球定位系统(global positioning system, GPS)设备提供的方向和速度信息加快对候选基站的选择,并自适应地缩短触发时延定时器。参考文献[9]提出了一种LTE-Advanced网络的快速切换方案,当列车接近预定义的切换位置时,网络可以直接发起切换。参考文献[10]将基站的覆盖区划分为若干区域,当列车快速通过一个区域时,网络可以提前为用户设备进行切换。参考文献[11]在此基础上做出改进,根据列车行驶线路生成相邻小区列表,根据列车速度和小区间距预先对切换地理位置进行设定,在列车跨越预先设定的切换点时,源基站根据列车测量报告发送预切换信息到目标基站,实现资源预留。参考文献[12]根据列车所处位置、行驶速度以及方向确定目标小区,并根据实验数据统计出切换参考点坐标,得出此位置的时间门限值,当列车到达参考点的时间低于门限时,对目标小区进行信道分配和激活,实现快速切换。



虽然基于地理位置信息的切换方案可以精确地控制切换时间,但是会导致额外的信令开销。首先,列车必须在其测量报告中上报位置、速度等信息。其次,一些方案要求运营商预先规划切换地点,增加网络规划的开销及成本。而如何让列车知道规划的切换位置是一个需要解决的问题,一般有两种方案:一种是让运营商在列车离开车站之前将切换位置等信息发送给列车;另一种是在列车运行过程中,基站通过测量控制信息向列车发送切换位置信息。这两种方案都会导致额外的信令开销。此外,基于地理位置信息的切换方案在 GPS 信号接收不良情况下会影响切换性能。例如,当列车离开隧道时,列车上的 GPS 设备可能需要时间来搜索卫星信号,在搜索时间内, GPS 设备不能正常工作,无法及时切换。

(2) 基于判决算法的切换方案优化设计

基于参考信号接收功率 (reference signal received power, RSRP) 是常用的判决算法, RSRP 是指移动终端接收来自基站的信号功率,能够反映接收信号的强度,得到了广泛的应用。接收到某基站的 RSRP 可以表示为:

$$Pr(m, x) [\text{dBm}] = P_t [\text{dBm}] - PL(m, d_m) [\text{dB}] - \varepsilon(0, \sigma_m) [\text{dB}], m = a, b \quad (1)$$

其中, P_t 为基站的信号发射功率, PL 为路径损耗, ε 为阴影衰落, a 、 b 分别表示源基站和目标基站。基于 RSRP 判决算法的切换触发概率为:

$$\begin{aligned} & P[Pr(b, x) - Pr(a, x) \geq \tau_p] \\ &= P[\varepsilon(0, \sigma_a) \geq \tau_p + PL(b, d_b) - PL(a, d_a) + \varepsilon_0 | \varepsilon(0, \sigma_b) = \varepsilon_0] \cdot \\ & P[\varepsilon(0, \sigma_b) = \varepsilon_0] \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_b^2}} Q\left(\frac{\tau_p + PL(b, d_b) - PL(a, d_a) + \varepsilon_0}{\sigma_a}\right) \exp\left(-\frac{\varepsilon_0^2}{2\sigma_b^2}\right) d\varepsilon_0 \end{aligned} \quad (2)$$

其中, τ_p 为 RSRP 迟滞容限^[13]。但是, RSRP 需要确保移动终端使用最高的功率连接到小区,仅考虑 RSRP 而不考虑干扰、信噪比等,会导致

信道资源的不合理利用。

近年来,一些学者通过综合考虑多种因素,提出联合判决算法,保证良好的切换性能。基于 RSRP 和参考信号接收质量 (reference signal received quality, RSRQ) 的联合判决算法中, RSRQ 类似于链路信噪比, RSRQ 判决算法能够在信号接收强度不够高的情况下,选择信号接收质量最好的基站执行切换。但是 RSRQ 受系统负载和测量策略影响较为严重,不能提供服务质量下降时的顽健检测,仅考虑 RSRQ 并不能获得最好的切换性能。因此, RSRQ 通常只作为辅助判决标准。基于 RSRP 和 RSRQ 联合判决算法的切换触发概率为:

$$\begin{aligned} P(x)_{\text{handover}} &= P[Pr(b, x) - Pr(a, x) \geq \tau_p] + \\ & P[Pr(b, x) - Pr(a, x) < \tau_p] \cdot \\ & P[SIR(b, x) - SIR(a, x) \geq \tau_q] \end{aligned} \quad (3)$$

其中, SIR 为接收到某基站的 RSRQ, τ_q 为 RSRQ 迟滞容限^[14]。通过设置统一的信噪比门限值,当低于该门限值时认为发生切换中断。图 2 为两种判决算法的切换成功概率对比,可以看出,基于 RSRP 和 RSRQ 的联合判决算法拥有更好的性能。

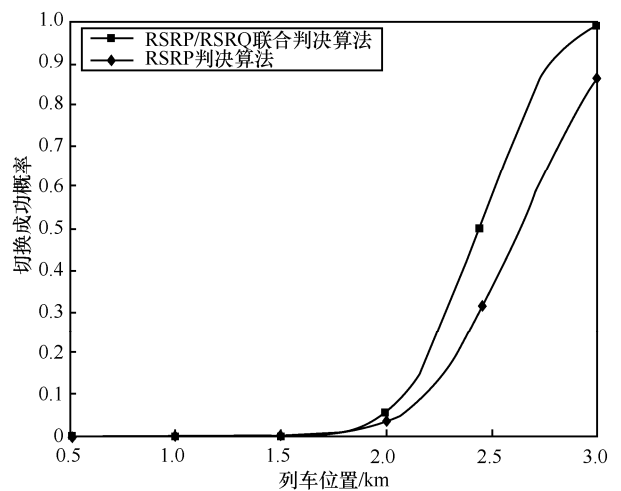


图2 切换成功概率性能对比

除此之外,也有学者综合考虑速度、距离、时间等其他因素,设计性能良好的联合判决方案,

弥补单一判决算法的缺陷^[15-17]。

(3) 基于双播机制的切换方案优化设计

高速铁路场景的切换面临的另一问题就是频繁切换、硬切换可能引发的通信中断。近几年, 基于双播 (bi-casting) 机制的切换优化方案也是越区切换领域的研究热点^[18,19]。在传统 LTE 切换中, 网络与移动台之间存在着数据传输。为了减少分组丢失率, 传统切换流程采用数据转发的机制。即在切换流程启动时, 服务基站通过 X2 接口将收到的数据副本转发给目标基站, 由目标基站将数据发送给移动台或服务网关。数据在基站之间的转发存在时延, 且服务基站和目标基站不能同时向移动台发送数据, 其通信中断时间包括切换处理时延和数据转发时延。

在双播机制中, 服务网关在切换开始后分别向源基站和目标基站发送数据副本。源基站处理收到的数据并发送给用户设备, 而目标基站则丢

弃收到的数据分组, 直到用户设备成功切换到目标基站。由于目标基站在切换开始后立即从服务网关接收数据, 所以通信中断时间约等于切换处理时延。因此, 双播机制的通信中断时间比数据转发机制短, 更适合一些实时通信服务。此外, 用于数据传输的基站和服务网关之间的 S1 接口具有比 X2 接口更高的吞吐量。基于双播机制的切换信令流程如图 3 所示, 源基站在接收到切换请求确认信息后, 便向移动管理实体进行双播请求, 随后建立服务网关到目标基站的双播数据通道, 在切换完成之后, 释放原有的服务网关到源基站的数据通道结束双播。如何减小双播带来的额外数据开销, 在双播机制的基础上进一步提高切换性能是研究人员需要考虑的重点。

3.2 基于网络优化的切换方案设计

随着实时视频监控、旅客移动互联网等各种业务需求出现, 未来高铁通信系统需要具备超高

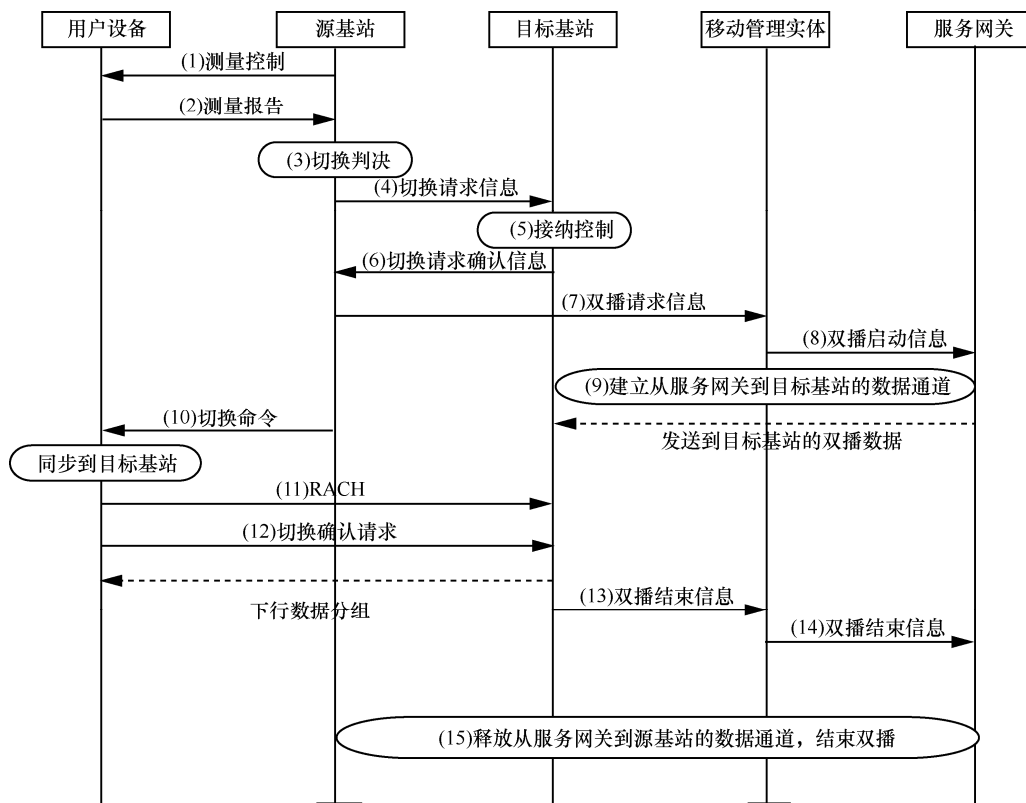


图 3 基于双播机制的切换方案信令流程



容量及超低时延等特点。目前, 5G 技术的测试与标准制定在如火如荼地展开, 未来的通信网络必定是一个多网融合、适应多种场景的超密集异构网络^[20]。同样, 对于高铁通信系统, 通过融合多种技术、优化网络架构提升切换性能, 提高用户体验将成为主流^[6]。

(1) 基于波束成形的切换方案优化设计

目前, 高速铁路场景中多天线和智能天线技术的研究和应用获得了一定进展^[21,22]。利用分布式天线是解决高铁通信系统中切换频繁问题的一种有效手段。在铁路沿线部署分布式天线单元, 不同的天线单元使用相同的工作频率等参数设置, 并通过光纤与基带单元连接。同一基带单元控制下的天线单元组成一个覆盖范围较大的逻辑小区。当高速列车在这个逻辑小区中运行时都不会发生切换, 从而减少切换次数, 降低通信中断的风险^[23]。

波束成形技术通过智能调整阵列天线各个阵元的幅度及相位, 形成定向波束, 将目标信号集中在基站与用户之间的方向, 实现能量的汇聚, 同时降低因能量扩散而对周围用户造成的干扰, 可以将其应用到高速铁路场景来集中信号能量, 提高传输可靠性^[24]。在参考文献[13]中, 波束成形技术还被用于解决高速铁路场景中因切换触发滞后而导致的切换失败问题。在高速列车驶入相邻基站的重叠区后, 基站天线由全向覆盖模式调整为波束成形模式, 改善接收信号功率。在不同的区域中, 动态调整基站的波束成形增益, 促进高速列车在短暂的切换窗口中能及时触发切换, 提高切换成功概率。图 4 为高速铁路场景中基于波束成形的切换方案和 LTE 切换方案的切换成功概率对比^[13]。其中, 横轴表示列车所在位置, 源基站和目标基站分别位于 0 km 和 4.8 km 处, 基站覆盖范围为 3 km。在 2.4 km 处由于应用波束成形技术, 波束成形切换方案的成功概率有明显的提高, 在 3 km 处的成功概率可达到 97.99%, 相比 LTE 切换方案, 其成功概率提高 7.8%。

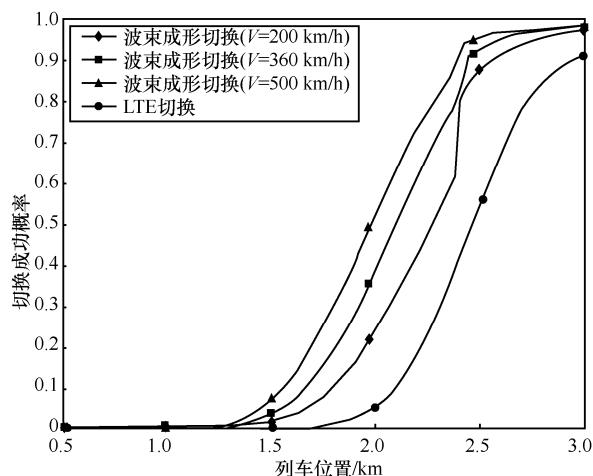


图 4 切换成功概率对比

(2) 基于车载中继的切换方案优化设计

中继目前作为一种较为成熟的技术, 是目前无线移动通信领域的研究热点, 其在高铁通信系统中也有较大的应用价值。

参考文献[18]在通过高速列车顶部安装车载中继站 (train mobile station, TRS), 用户终端通过各个车厢部署的 AP (access point, 接入点) 连接到 TRS。利用 TRS 的群移动性, 由 TRS 代替列车中的用户设备进行基站选择和越区切换, 有效避免巨大的信令开销引起的“信令风暴”^[25]。TRS 拥有比普通用户终端更为强大的计算能力, 能够对多普勒频移进行矫正, 改善信号接收质量, 节省用户终端的能量。同时, TRS 可以通过光纤与列车外部的收发天线进行连接, 从而可以有效避免车厢穿透损耗。

由于软切换需要无线网络控制器 (radio network controller, RNC) 的支持, 为了设计扁平化的网络结构以及降低端到端时延, LTE 系统采用“先断后连”的硬切换方式, 在高铁高速运行中, 容易产生较大的中断概率^[26]。为了改进硬切换带来的不利影响, 参考文献[27]提出一种基于移动中继的双天线辅助切换方案, 如图 5 所示。双天线分别安装在车首和车尾, 当列车进入重叠区并满足切换触发条件后, 先是由车首天线执行切换, 车尾天线继续保持与服务基站连接, 切换成

功后,车尾天线将工作频率同步到目标基站完成切换。若切换失败,车尾天线在进入重叠区后可再次执行切换。通过两部天线的协作通信,达到软切换的效果,大大地降低切换失败率。在此基础上,参考文献[18]通过引入双播机制进一步使信令开销减小 50%左右。

(3) 基于控制面/用户面分离的切换方案优化设计

未来 5G-R 通信系统在提高网络传输容量时,应考虑将控制面与传统用户面信息解耦设计,合理利用 GSM-R 系统中的现有带宽,均衡不同频段的差异,对系统频段弹性设计。为此,必须研究基于控制面/用户面(C/U 面)分离的高铁无线网络架构,通过控制面与用户面的分离架构,将拥有较宽连续频谱的高频频段与 GSM-R 低频频段融合,为高铁通信系统提供高容量、高可靠性的通信服务^[19]。由于控制面信息以及列车控制、列车调度等用户面信息对传输可靠性要求较高,可以由宏基站通过拥有良好传输特性的低频段来承载;对传输容量需求较高的旅客业务,可以由虚拟基站通过高频段承载。虚拟基站专门为移动用

户提供数据业务服务,而网络接入、系统信息广播等控制信息由宏基站提供,也就是将控制面与用户面放置在拥有不同特性的频段上传输,均衡不同频段的差异。由于在宏基站和虚拟基站中使用的频带不同,宏基站和虚拟基站之间不会存在同信道干扰^[28]。基于 C/U 面分离的高铁通信网络架构如图 6 所示。

基于 C/U 面分离的高铁无线网络架构已于 2016 年 6 月写入无线世界研究论坛(Wireless World Research Forum, WWRF)和未来移动通信论坛联合发布的 5G 高速移动白皮书《2016 Whitepaper V2.0H-5G Enabler: High Mobility Support》。然而,在基于 C/U 面分离的高铁通信系统中,其切换问题更加严重,原因如下^[29]。

(1) 宏基站间切换频率更高

假设一个宏基站包含 3 个虚拟基站,高速列车穿过一个宏小区就要执行 1 次宏基站间切换和 3 次虚拟基站间的切换。高速列车执行的切换越多,中断的可能性越大。

(2) 切换失败概率较高

列车首先执行“宏基站—宏基站”切换,使用

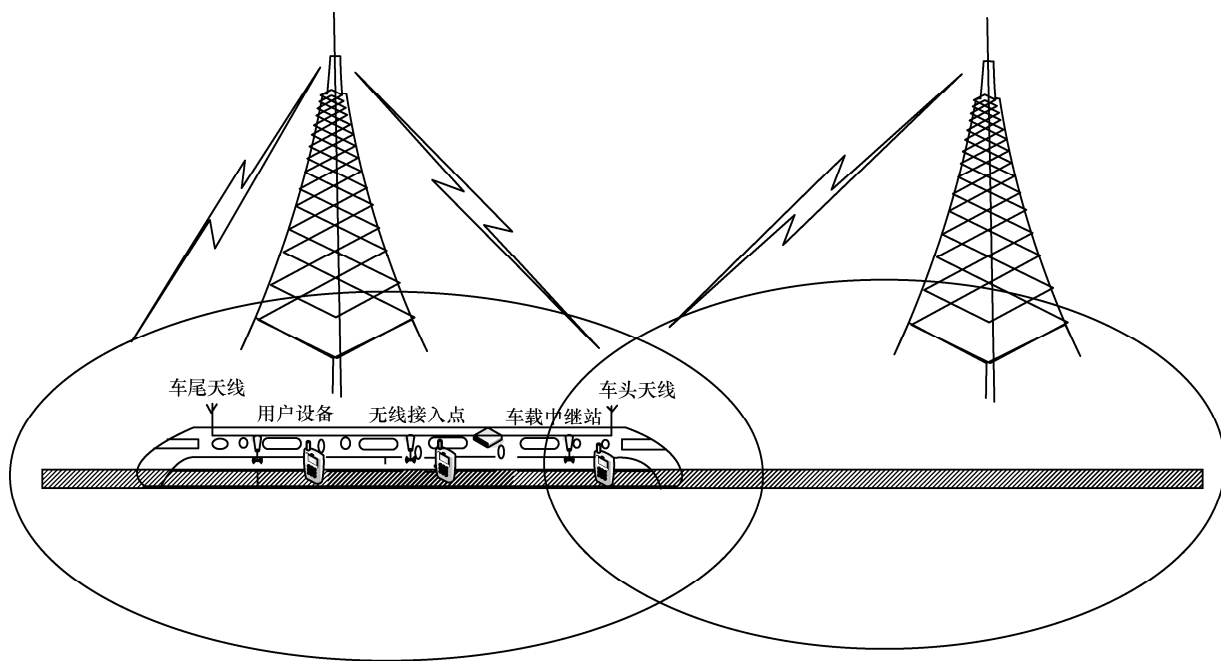


图 5 基于移动中继双天线切换方案系统模型

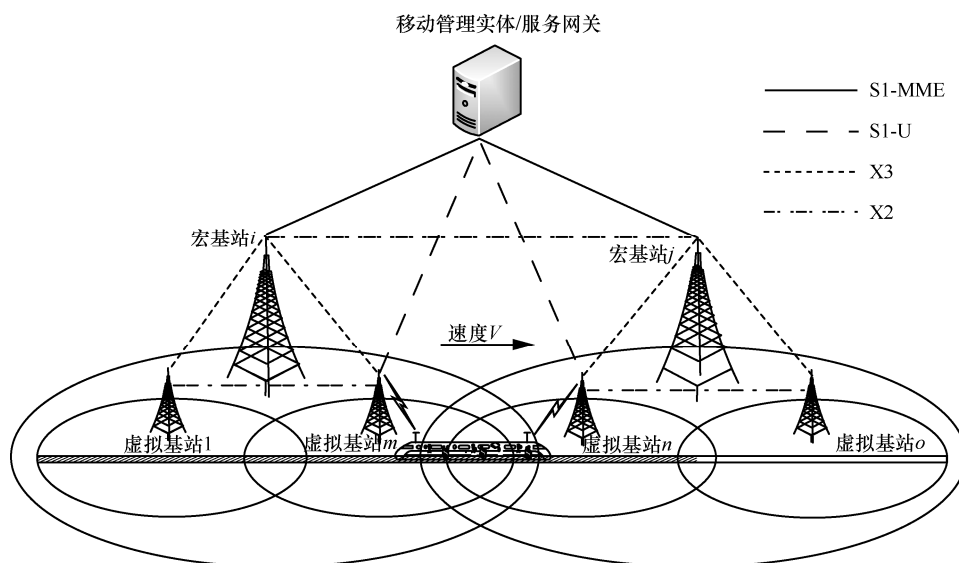


图6 基于 C/U 面分离的高铁通信网络架构

户设备连接到目标宏基站,然后在目标宏基站的控制下执行“虚拟基站—虚拟基站”切换,建立用户设备与目标虚拟基站之间的连接。当且仅当两种切换都顺利完成时,宏基站间切换才算成功。

(3) 切换重叠区较短

“虚拟基站—虚拟基站”切换作为宏基站间切换的第二步,必须在“宏基站—宏基站”切换完成之后才能执行。因此,切换必须在虚拟基站间重叠区域内全部完成。

(4) 切换触发滞后

当且仅当目标宏基站和目标虚拟基站的信号质量分别超过服务宏基站和服务虚拟基站一定阈值时,宏基站间的切换才会被触发。因此,这会造成一定的切换触发滞后。

4 结束语

为了保障高速列车在移动过程中调度信息的实时传输以及为旅客提供优质的宽带通信业务,高速铁路宽带通信系统对越区切换性能提出了极高的要求。本文从参数优化和网络优化两个方面总结了国内外的研究成果,并进一步探讨了未来越区切换技术的发展方向。基于控制面/用户面分

离的高铁通信网络架构是目前的研究热点,结合波束成形、车载中继、双播等技术对切换方案进行优化,设计适应高铁场景的快速切换或无缝切换方案,对未来高铁宽带通信系统的发展具有重要意义。

参考文献:

- [1] ZHANG J Y, TAN Z H, ZHONG Z D, et al. A multi-mode multi-band and multi-system-based access architecture for high-speed railways[C]//IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall), Sept 6-9, 2010, Ottawa, Canada. New Jersey: IEEE Press, 2010: 1-5.
- [2] SUN R, TAO C, LIU L, et al. Channel measurement and characterization for HSR U-shape groove scenarios at 2.35 GHz[C]//IEEE 78th Vehicular Technology Conference (VTC 2013-Fall), Sept 2-5, 2013, Las Vegas, NV, USA. New Jersey: IEEE Press, 2013: 1-5.
- [3] AI B, CHENG X, KÜRNER T, et al. Challenges toward wireless communications for high-speed railway[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(5): 2143-2158.
- [4] AI B, GUAN K, RUPP M, et al. Future railway services-oriented mobile communications network[J]. IEEE Communications Magazine, 2015, 53(10): 78-85.
- [5] HE R, AI B, WANG G, et al. High-speed railway communications: from GSM-R to LTE-R[J]. IEEE Vehicular Technology Magazine, 2016, 11(3): 49-58.
- [6] FAN P, ZHAO J, CHIH-LIN I. 5G high mobility wireless communications: challenges and solutions[J]. China Commu-

- cations, 2016, 13(Supplement2): 1-13.
- [7] AGUADO M, JACOB E, ASTORGA J, et al. The cross layer RMPA handover: a reliable mobility pattern aware handover strategy for broadband wireless communication in a high-speed railway domain[J]. *EURASIP Journal on Wireless Communications and Networking*, 2012(1): 298.
 - [8] CHANG F M, WANG H L, HU S Y, et al. An efficient handover mechanism by adopting direction prediction and adaptive time-to-trigger in LTE networks[C]//*International Conference on Computational Science and its Applications*, June 24-27, 2013, Ho Chi Minh City, Vietnam. Berlin: Springer, 2013: 270-280.
 - [9] HUANG Q, ZHOU J, TAO C, et al. Mobile relay based fast handover scheme in high-speed mobile environment[C]//*IEEE 76th Vehicular Technology Conference (VTC 2012-Fall)*, Sept 3-6, 2012, Québec City, Canada. New Jersey: IEEE Press, 2012: 1-6.
 - [10] HUANG Y S, LEU F Y, LIU J C, et al. A handover scheme for LTE wireless networks under the assistance of GPS[C]//*2013 8th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA)*, Oct 28-30, 2013, Compiegne, France. New Jersey: IEEE Press, 2013: 399-403.
 - [11] 原燕斌. 高速铁路环境下 LTE 切换技术的研究[D]. 北京: 北京邮电大学, 2012.
YUAN Y B. A study of handover algorithms within LTE system under high-speed railway environment[D]. Beijing: Beijing University of Posts and Telecommunications, 2012.
 - [12] 吴昊, 谷勇浩, 钟章队. 一种应用于高速铁路的 GSM-R 快速切换算法研究[J]. *铁道工程学报*, 2009(124): 92-96.
WU H, GU Y H, ZHONG Z D. Research on the fast algorithm for GSM-R switching for high-speed railway[J]. *Journal of Railway Engineering Society*, 2009(124): 92-96.
 - [13] CHENG M, FANG X M, LUO W T. Beamforming and positioning-assisted handover scheme for long-term evolution system in high-speed railway[J]. *IET Communications*, 2012, 6(15): 2335-2340.
 - [14] 贺臻桢. 基于 LTE-R 的高速移动宽带通信系统切换技术研究[D]. 长沙: 中南大学, 2013.
HE Z Z. The research on handover technology under high-speed mobile communication system based on LTE-R[D]. Changsha: Central South University, 2013.
 - [15] LUO W, FANG X, CHENG M, et al. An optimized handover trigger scheme in LTE systems for high-speed railway[C]//*Fifth International Workshop on Signal Design and its Applications in Communications (IWSDA)*, Oct 10-14, 2011, Guilin, China. New Jersey: IEEE Press, 2011: 193-196.
 - [16] YANG Y, YU P, LI W. Handover self-optimization mechanism based on velocity for cellular networks[C]//*2012 7th International ICST Conference on Communications and Networking in China (CHINACOM)*, Aug 8-10, 2012, Kunming, China. New Jersey: IEEE Press, 2012: 606-610.
 - [17] BHATTACHARYA S, DE D. An enhanced handover strategy for GSM-R technology[C]//*2012 Third International Conference on Computer and Communication Technology (ICCCCT)*, Nov 23-25, 2012, Allahabad, India. New Jersey: IEEE Press, 2012: 227-232.
 - [18] TIAN L, LI J, HUANG Y, et al. Seamless dual-link handover scheme in broadband wireless communication systems for high-speed rail[J]. *IEEE Journal on Selected Areas in Communications*, 2012, 30(4): 708-718.
 - [19] YAN L, FANG X, FANG Y. A novel network architecture for C/U-plane staggered handover in 5G decoupled heterogeneous railway wireless systems[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2017(99): 1-13.
 - [20] 倪善金, 赵军辉. 5G 无线通信网络物理层关键技术[J]. *电信科学*, 2017, 33(12): 40-45.
NI S J, ZHAO J H. The potential key technologies in physical layer of 5G wireless communications network[J]. *Telecommunications Science*, 2017, 33(12): 40-45.
 - [21] LUO W T, FANG X M, CHENG M, et al. Efficient multiple group multiple-antenna (MGMA) scheme for high-speed railway viaducts[J]. *IEEE Transactions on Vehicular Technology*, 2013, 62(6): 2558-2569.
 - [22] ZHAO J, WANG S, LIU X, et al. Geometry-based stochastic modeling for non-stationary high-speed train MIMO channels[C]//*IEEE 84th Vehicular Technology Conference (VTC 2016-Fall)*, Sept 18-21, 2016, Montréal, Canada. New Jersey: IEEE Press, 2016: 1-5.
 - [23] LANNIO B, COLLE D, PICKAVET M, et al. Radio-over-fiber-based solution to provide broadband internet access to train passengers[J]. *IEEE Communications Magazine*, 2007, 45(2): 56-62.
 - [24] NI S, ZHAO J, GONG Y. Optimal pilot design in massive MIMO systems based on channel estimation[J]. *IET Communications*, 2016, 11(7): 975-984.
 - [25] ZHANG Y. Handoff performance in wireless mobile networks with unreliable fading channel[J]. *IEEE Transactions on Mobile Computing*, 2010, 9(2): 188-200.
 - [26] LIN C C, SANDRASEGARAN K, RAMLI H A M, et al. Requirement of handover modeling in the downlink 3GPP long term evolution system[C]//*IEEE 24th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, April 20-23, 2010, Perth, WA, Australia. New Jersey: IEEE Press, 2010: 305-310.
 - [27] QIAN X Y, HAO W. Mobile relay assisted handover for LTE system in high-speed railway[C]//*International Conference on Control Engineering and Communication Technology (ICCECT)*, Dec 7-9, 2012, Shenyang, China. New Jersey: IEEE Press, 2012: 632-635.
 - [28] ISHII H, KISHIYAMA Y, TAKAHASHI H. A novel architecture for LTE-B: C-plane/U-plane split and phantom cell concept[C]//*2012 IEEE Globecom Workshops (GC Wkshps)*, Dec 3-7, 2012, Anaheim, CA, USA. New Jersey: IEEE Press, 2012:



624-630.

- [29] SONG H, FANG X, YAN L. Handover scheme for 5G C/U plane split heterogeneous network in high-speed railway[J]. IEEE Transactions on Vehicular Technology, 2014, 63(9): 4633-4646.

[作者简介]



刘云毅（1993-），男，北京交通大学电子信息工程学院硕士生，主要研究方向为LTE-R 越区切换。



赵军辉（1973-），男，博士，北京交通大学电子信息工程学院教授、博士生导师，华东交通大学信息工程学院院长，主要研究方向为 5G 通信技术、车联网、轨道交通无线通信等。



王传云（1977-），男，华东交通大学信息工程学院副教授，主要研究方向为无线通信、无线传感网。



基于用户体验质量和系统能耗的异构网络 联合接入选择和功率分配策略

季石宇¹, 唐良瑞¹, 李淑贤¹, 杜施默²

(1. 华北电力大学电气与电子工程学院, 北京 102206;

2. 中国移动通信集团浙江有限公司杭州分公司, 浙江 杭州 310006)

摘要: 为了解决无线网络能耗和用户体验质量的问题, 提出了一种应用于异构无线网络环境的基于多目标优化的联合接入选择和功率分配策略。该策略以最小化系统整体能耗和最大化用户平均体验质量为目标, 建立了多目标优化联合分配模型, 通过差分进化算法得到联合分配模型的最优解集, 并结合 TOPSIS 的方法从最优解集中得到折中的联合分配方案。仿真结果表明, 本文所提的策略能够有效地降低系统能耗, 并提升用户整体的服务体验。

关键词: 异构网络; 用户体验质量; 能耗; 接入选择; 功率分配

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017251

Joint user association and power control scheme based on energy consumption and QoE in heterogeneous network

Ji Shiyu¹, TANG Liangrui¹, LI Shuxian¹, DU Shimo²

1. School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China

2. Hangzhou Branch of China Mobile Group Zhejiang Co., Ltd., Hangzhou 310006, China

Abstract: Based on multi-objective optimization, a joint user association and power control scheme was proposed in heterogeneous network environment, aiming to reduce the system energy consumption and improve user quality of experience. The scheme took the system energy consumption minimum and the average user quality of experience maximum as objective functions and translated the joint user association and power control to multi-objective problem. Then, the differential evolution optimization algorithm and TOPSIS algorithm was used to solve the multi-objective problem. Simulation results show that the scheme can efficiently reduce the system energy consumption and improve the user quality of experience.

Key words: heterogeneous network, user quality of experience, energy consumption, user association, power control

收稿日期: 2017-07-05; 修回日期: 2017-08-21

基金项目: 北京市自然科学基金资助项目 (No.4142049); 国家高技术研究发展计划 (“863”计划) 基金资助项目 (No.2014AA01A701); 中央高校基本科研业务费专项资金资助项目 (No.2015XS07)

Foundation Items: Beijing Natural Science Foundation of China (No.4142049), The National High Technology Research and Development Program of China (863 Program) (No.2014AA01A701), The Central University Basic Business Expenses Special Funding for Scientific Research Projects (No.2015XS07)



1 引言

未来无线网络呈现出泛在化、扁平化的特性,将不再是由单一的接入网络构成,而是包含宏蜂窝网络、微蜂窝网络等多种形式、多种技术和能耗开销的异构蜂窝网络^[1,2]。异构蜂窝网络的出现能为用户提供更为充裕的资源,从而有效地保证用户的服务质量^[3,4]。但是由于无线终端设备的增多、终端类型的日益丰富,用户已经不满足于保证自身的服务质量,而更加追求提升自身的体验质量(quality of experience, QoE)^[5]。另一方面,由于温室气体的排放、煤炭等不可再生资源使用的加剧,系统能耗也是人们关注的另一个重点^[6]。异构网络的资源管理能够很好地解决用户体验提升和系统能耗降低的问题。

目前,对于异构无线网络中的资源管理已有广泛研究,参考文献[7]提出了信干噪比最大化(signal to interference plus noise ratio maximization, SINRmax)的接入方案,即用户接入信干噪比最强的网络中。参考文献[8]提出了基于负载均衡的接入选择(radio access selection, RAS),用户根据网络的负载状况和自身的服务质量(quality of service, QoS)需求,从而决定接入最优的网络之中。参考文献[9]为了降低用户干扰,提出了集中式的功率控制策略。但是上述所提的这些策略都是对单一的无线资源(如功率)进行优化和管理,并且由于现有网络资源维度的多样性,单一资源的优化难以实现网络能耗和用户体验质量的最优化。因此,越来越多的学者进一步地研究异构网络联合资源优化^[10-13]。参考文献[10]提出了在保证用户 QoS 的条件下对网络能耗的降低,但是所提的策略是基于频谱分段的情况下,即用户传输过程中不会受到同频信号的干扰,但会大大降低无线网络频谱的利用率。参考文献[11]提出了在频谱共享下联合功率和接入的方法最小化上行传输功率总和,但是系统能耗并不仅仅包

括上行传输能耗,还包括基站处理等能耗。参考文献[12]提出了基于能效最大化的联合用户接入和功率控制(EEUA)的策略,通过将功率分配和接入选择问题转变为非线性混合整数优化问题,并通过迭代算法得到问题的最优解。但是并未考虑到由于未来无线用户将会使用各类应用和业务,用户的体验质量也是十分重要的^[14-17]。参考文献[14]提出了一种用户服务体验质量模型,但是并未考虑到在实时无线系统中,模型参数难以实时获取。为此,参考文献[15]提出了在实时无线系统中的用户体验质量模型,并设计了以提升用户体验质量为目标资源管理方案。但是这些文献单一地从能耗或者用户体验质量方面进行优化,并未综合考虑用户体验质量、系统能耗以及用户接入和功率分配等问题,这样往往会造成另一个目标的劣化,从而难以得到一个较为满意的折中解。

基于以上分析,本文提出了基于多目标的接入选择和功率分配联合优化策略(multi-objective based joint user association and power control scheme, MUPS)。通过分析无线网络系统能耗成因,从用户侧和网络侧共同建立了系统能耗模型,并结合用户体验质量提升的需求,提出了系统能耗最小、用户体验质量最大化的多目标优化模型,通过多目标差分算法进行求解,并结合逼近理想解的排序方法(technique for order preference by similarity to ideal solution, TOPSIS)得到最优的接入选择和功率分配方案。仿真结果表明,该策略能够有效地降低系统能耗,并提升用户整体的体验质量。

2 系统模型

在异构网络环境下,当用户发出接入请求时,将会触发网络接入选择机制,网络控制中心将给各个用户分配相应的网络保证用户的接入,假设无线用户在异构无线网络重叠覆盖范围内

随机分布。系统模型如图 1 所示，重叠覆盖范围内共有 M 个网络，每个网络的总带宽为 B 。当用户 n 接入网络 m 中，则 $a_{mn}=1$ ，否则 $a_{mn}=0$ 。

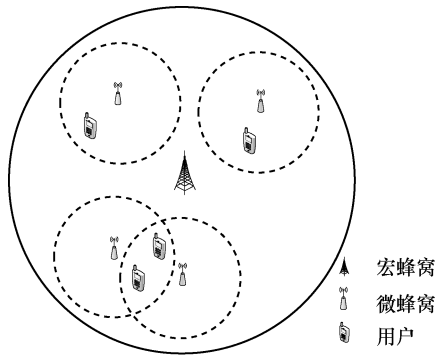


图 1 系统模型

当无线用户接入网络后，网络将会给用户分配业务传输时隙，从而保证无线用户传输业务，而当用户传输业务时，无线用户对于网络所提供的业务传输等服务存在着主观上的体验差异性，即具有不同的服务满意程度。相较于传统的服务质量，用户的体验质量进一步地引入了用户的主观感受所带来的影响。对于用户的体验质量，国际标准组织采用了平均评价分数（mean opinion score, MOS）模型作为评价模型。MOS 模型是一种主观评价模型，能够最直观地体现用户感受。它将用户感受分为 4 段，各段的统计意义见表 1^[18]。

表 1 MOS 数值与用户感受对应关系

MOS 区间	用户感受
[1,2)	十分不满意
[2,3)	不满意
[3,4)	满意
[4,5]	十分满意

通常对于无线用户，用户的体验质量是与用户的主观感受相一致的，而无线网络的传输速率与用户的主观感受息息相关，当用户的传输速率较高时，其体验质量也会提升，但是用户速率达到一定值后，用户的体验质量并不会增加；同样地，当用户的传输速率低于门限值时，由于用户

难以达到自身服务的需求，其体验质量将会是最低的，即十分不满意网络所提供的服务。本文通过参考文献[18]所提的效用函数来表示用户的体验质量，即对于无线用户 n ，其体验质量为：

$$MOS_n = \begin{cases} 1, & R < R_{\min} \\ a \ln(R_n / b), & R_{\min} \leq R \leq R_{\max} \\ 4.5, & R_{\max} < R \end{cases} \quad (1)$$

其中， R_n 为用户 n 的传输速率， $R_{\min}$ 和 $R_{\max}$ 分别是最小传输速率和最大传输速率。 a 和 b 为固定变量，保证所提的用户体验函数为连续函数。

3 多目标接入选择及功率分配联合优化策略

在无线网络中，用户十分关心自身的体验质量，但是整个无线系统更关心整个网络的能耗和终端能耗，从而能够达到节能和降低成本的目的。因此本文从网络侧和用户侧分别建立了能耗目标，同时考虑了用户体验质量的最优化，从而将用户的接入选择和功率分配问题转变为多目标优化问题，并采用多目标优化的算法对优化问题进行求解得到最优的接入选择和功率分配方案，并通过采用 TOPSIS 的方法对系统能耗和用户体验质量做出很好的折中。

3.1 多维联合优化模型

3.1.1 最小化能耗

整个无线通信系统中，不仅基站会消耗大量的能量，并且随着终端数目的增多，终端能耗也日益成为系统能耗的重要组成部分。无线网络基站能耗主要包括无线传输能耗、信号处理能耗和固定的电路能耗，计算式为：

$$E_B = \int_0^t \left(\sum_{m=1}^M (\varepsilon_m p_m(\tau) + p_m^c(\tau)) + \sum_{m=1}^M p_m^b(\tau) \right) d\tau \quad (2)$$

其中， ε_m 为网络 m 的天线增益， $p_m(\tau)$ 为网络 m 在第 τ 时刻无线网络的传输功率， $p_m^c(\tau)$ 和 $p_m^b(\tau)$ 分别为网络 m 的导频功耗和固定功耗。

用户侧的能耗主要包括两个部分，一部分是由用户终端信号处理所带来的能耗，另一部分是



终端固定能耗。终端处理能耗不仅和处理时间有关，还与信号处理功率有关^[19]，而由于终端始终保持开启状态，其固定能耗则是在整个时隙中所消耗的能量，计算式为：

$$E_U = \sum_{m=1}^M \sum_{n=1}^N \int_0^{t_n'} (a_{mn} p_n^S(\tau)) d\tau + \int_0^{t_n'} \left(\sum_{n=1}^N p_n^I(\tau) \right) d\tau \quad (3)$$

其中， $p_n^S(\tau)$ 为用户在第 τ 时刻的信号处理功率， $p_n^I(\tau)$ 为用户 n 的固定功耗。 t_n' 为用户 n 的信号处理时间，对于用户 n ，其接入网络后与网络中所有用户共用时隙，其信号处理的时间为：

$$t_n' = \sum_{m=1}^M \left(a_{mn} t / \sum_{j=1}^N a_{mj} \right) \quad (4)$$

整个系统的能量消耗应当包括网络侧的能量消耗和终端的能量消耗：

$$E = E_B + E_U \quad (5)$$

考虑到在所建立的网络场景下，用户并不存在高速移动，这就使得在整个时隙内无线网络和无线终端的信号处理功率、发射功率等几乎不发生变化，因此整个系统的能耗最小化的目标可以转变为整个系统功耗的最小：

$$f_1(a, p) = \min(P_B + P_U) \quad (6)$$

其中， P_B 为网络功耗，包括固定功耗和传输功耗。 P_U 为用户功耗，包括接收信息的处理功耗和固定功耗。具体的计算方式分别为：

$$P_B = \sum_{m=1}^M (\varepsilon_m p_m + p_m^c) + \sum_{m=1}^M p_m^b \quad (7)$$

$$P_U = \sum_{m=1}^M \sum_{n=1}^N a_{mn} p_n^S \left(\sum_{m=1}^M \left(a_{mn} / \sum_{j=1}^N a_{mj} \right) \right) + \sum_{n=1}^N p_n^I \quad (8)$$

其中， p_m 为在传输时隙内基站发射功率的均值， p_m^b 为基站固定功耗， p_n^S 和 p_n^I 分别为用户接收信息的处理功耗和用户设备的固定功耗。

3.1.2 最大化用户平均体验质量

无线网络联合优化策略的另一个目标是实现无线用户的平均体验质量最大化：

$$f_2(a, p) = \max \frac{1}{N} \sum_{n=1}^N MOS_n \quad (9)$$

其中， MOS_n 为用户 n 的体验质量，可以通过式 (1) 进行计算，通过式 (1) 可以看出，用户 n 的体验质量仅仅与用户 n 的传输速率有关，对于用户 n 而言，其传输速率为：

$$R_n = \sum_{m=1}^M \frac{B}{\sum_{n=1}^N a_{mn}} \log(1 + \gamma_{mn}) \quad (10)$$

其中， γ_{mn} 为用户 n 接入网络 m 后的信噪比，计算式为：

$$\gamma_{mn} = \frac{p_m g_{mn}}{\sum_{j \in S} p_j g_{jn} + N_0 B} \quad (11)$$

其中， g_{mn} 为网络 m 与用户 n 之间的信道增益， N_0 为高斯白噪声密度。

3.1.3 约束条件

无线系统中由于用户需求和基站性能等会对联合优化造成影响，主要包括系统发射功率的约束，用户接入和 QoS 约束。

(1) 发射功率约束

在实际无线网络中，第 m 个无线网络可用的传输功率是有限的，记为 $p_{m\max}$ ，最大传输功率的限制是为了保证基站设备能够处于线性工作区间，从而保证基站设备的良好运行，因此发射功率的约束为：

$$p_m \leq p_{m\max}, \forall m \in \mathcal{M} \quad (12)$$

(2) 用户接入约束

在无线网络中，每个用户在同一时刻最多只能接入一个无线接入网络中，即：

$$a_{mn}(a_{mn} - 1) = 0, \forall m \in \mathcal{M}, n \in \mathcal{N} \quad (13)$$

(3) QoS 约束

用户接收到的信噪比应该满足自身约束的需求，从而满足自身最低的服务质量需求，即：

$$\sum_{m=1}^M a_{mn} \gamma_{mn} \geq \gamma_{\text{req}}, \forall n \in \mathcal{N} \quad (14)$$

3.2 多目标差分进化算法

综上所述, 本文从网络能耗和用户体验两个方面对异构网络下的接入选择进行研究, 其目的是能够降低网络能耗和提升用户体验质量, 因此对于异构网络接入选择问题可以转变为多目标优化的问题:

$$\begin{aligned} & \min f_1(a, p), \max f_2(a, p) \\ & \text{s.t.} \begin{cases} p_m \leq p_{\max}, \forall m \in \mathcal{M} \\ a_{mn}(a_{mn} - 1) = 0, \forall m \in \mathcal{M}, n \in \mathcal{N} \\ \sum_{m=1}^M a_{mn} \gamma_{mn} \geq \gamma_{\text{req}}, \forall n \in \mathcal{N} \end{cases} \end{aligned} \quad (15)$$

其中, $f_1(a, p)$ 和 $f_2(a, p)$ 为联合分配算法的优化目标。考虑到所提的多目标优化问题的解空间集为无穷解集, 并且所提的优化目标的一阶导函数非连续, 因此优化问题难以通过将其转化为凸优化问题进行求解。所以采用多目标差分进化算法实现对多目标问题的优化求解^[20]。差分进化算法是一种随机的并行直接全局搜索算法, 具有简单易用和全局寻优能力强等优点, 已广泛用于单目标优化问题和多目标优化问题的求解。

(1) 种群初始化

假设初始化种群为 $X^t = (x_1^t, x_2^t, \dots, x_N^t)$, 其中, N 为种群规模, 种群中每个个体都是一组优化变量, 即包括接入方式和功率分配方式的优化变量。初始化迭代次数 $t = 0$, 其中接入方式为离散型变量而功率分配方式为连续型变量。因此本文采用混合编码的方式对每个个体进行编码。

$$x_i = \{\overbrace{b_1, b_2, \dots, b_n, \dots, b_N}^{\text{整数}}, \overbrace{p_1, p_2, \dots, p_m, \dots, p_M}^{\text{实数}}\} \quad (16)$$

其中, P_m 为网络 m 的传输功率, b_n 为用户 n 所接入的网络, 可以通过式 (17) 进行计算:

$$b_n = \sum_{m=1}^M (ma_{mn}) \quad (17)$$

(2) 变异

变异操作是通过对种群个体的接入方式和功

率分配方式的变异, 从而扩大搜索空间, 避免陷入局部陷阱的方式, 具体可表示为:

$$x_i^{t+1} = x_i^t + R \cdot (x_{r_1}^t - x_{r_2}^t) \quad (18)$$

其中, R 为变异常量, 通常表示的是前次迭代个体对下次迭代个体的影响程度。 x_i^t 和 x_i^{t+1} 分别为变异前和变异后的种群个体, $x_{r_1}^t - x_{r_2}^t$ 为当代个体的差分向量, 其中 $r_1 \neq r_2$ 。对于整个变异操作, 如果超出上下限范围则取边界值。由于本文的编码方式为混合编码方式, 当种群个体的优化变量为整数时, 需要对变异产生的值进行取整:

$$x_i^{t+1} = \lceil x_i^t + R \cdot (x_{r_1}^t - x_{r_2}^t) \rceil \quad (19)$$

其中, 符号 $\lceil \cdot \rceil$ 表示取整函数。

(3) 交叉

为了保证种群的多样性, 采用二项式交叉算子对前后两代的种群个体进行交叉, 从而保证前代个体的分量能够被继承。具体方式为:

$$v_{ij}^{t+1} = \begin{cases} x_{ij}^t, \text{rand}(\cdot) > C_r \\ x_{ij}^{t+1}, \text{rand}(\cdot) \leq C_r \end{cases} \quad (20)$$

其中, $\text{rand}(\cdot)$ 为在 $[0, 1]$ 中产生的随机数, C_r 为交叉概率。所得的 v_{ij}^{t+1} 为 $(t+1)$ 代种群中第 i 个个体 v_i^{t+1} 的第 j 个优化变量。

(4) 选择

本文采用参考文献[20]所提出的选择策略来求解含有约束条件的多目标优化问题。通过计算所有种群个体的目标函数值, 比较变异交叉后产生的 $(t+1)$ 代个体与 t 代个体的目标值大小, 若 $(t+1)$ 代个体 v_i^{t+1} 支配 t 代个体 x_i^t , 则 v_i^{t+1} 进入种群 X^{t+1} , 否则 x_i^t 进入种群 X^{t+1} 。个体 v_i^{t+1} 支配个体 x_i^t 指的是满足式 (21):

$$f_i(x_i^t) \prec f_i(v_i^{t+1}), i = 1, 2, \dots, r \quad (21)$$

其中, $f_i(\cdot)$ 为优化目标函数, $\prec$ 表示左边劣于右边, 当目标函数为最大化目标时, 即表示小于, 否则表示大于。 r 是优化目标函数的个数。



(5) 最优折中解的选取

通过上述步骤可以得到多目标问题下的最优解集, 即 Pareto 解集。在实际的网络中, 用户最终的接入方案和功率分配方案需要能从 Pareto 解集中选择出最优的折中解, 即需要均衡系统能耗和用户体验质量的最优联合分配方案^[21]。本文采用 TOPSIS 的方法实现对系统能耗和用户体验质量的有效折中, 即通过计算式 (22) 得到 Pareto 解集中各个联合用户接入和功率分配方案的最优解:

$$D_k = \frac{\sqrt{\sum_{i=1}^r (Y_{ki} - Y_i^-)^2}}{\sqrt{\sum_{i=1}^r (Y_{ki} - Y_i^-)^2 + \sum_{i=1}^r (Y_{ki} - Y_i^+)^2}} \quad (22)$$

其中, Y_{ki} 为 Pareto 解集中第 k 个解的第 i 个目标函数归一化值, Y_i^+ 和 Y_i^- 分别为 Pareto 前沿中所有个体的第 i 个目标函数归一化值的最大值和最小值。目标函数的归一化值通过式 (23) 计算:

$$Y_{ki} = \frac{H_{ki}}{\sqrt{\sum_{k=1}^K H_{ki}^2}} \quad (23)$$

其中, H_{ki} 为 Pareto 解集中第 k 个解的第 i 个目标函数值, K 为 Pareto 前沿中的个体总数。根据 TOPSIS 方式所得到的 D_k 值在 0~1 范围内, 根据 D_k 值大小对所得的 Pareto 前沿解集进行排序, D_k 值越大, 说明所得的联合分配方案越接近于最优方案, 反之, 则远离最优方案。最终, 根据排序结果得到最终的联合分配方案。

4 仿真分析

为了验证 MUPS 算法的性能, 本文对 MUPS 算法与 EEUA 和 SINRmax 进行仿真对比分析。采用异构多接入网络组网的方式, 仿真模型如图 1 所示, 其中, 宏蜂窝的信道衰落为 $10\lg\theta = 128.1 + 37.6\lg d$, 微蜂窝网络的信道衰落为 $10\lg\theta = 140.7 + 36.7\lg d$ ^[22], 具体的相关参数见表 2^[12,19]。

表 2 具体的相关参数

仿真参数	宏蜂窝	微蜂窝
总带宽/MHz	10	10
天线增益	2	2
最大发射功率/dBm	43	33
覆盖半径/m	500	50
用户接收信号处理功率/mW	5~30	
高斯白噪声密度/(dBm·Hz ⁻¹)	-174	

4.1 收敛性分析

算法收敛性分析如图 2 所示。由图 2 可知, 随着迭代次数的增多, 系统能耗不断降低, 而用户平均体验质量将不断提升, 这是由于算法将交叉变异后的个体与原来的个体进行比较, 选择出优秀个体进入下次迭代, 从而保证了后代个体的优质性。而当迭代次数为 45 次左右时, 算法达到收敛, 系统能耗和用户平均体验质量不再变换。

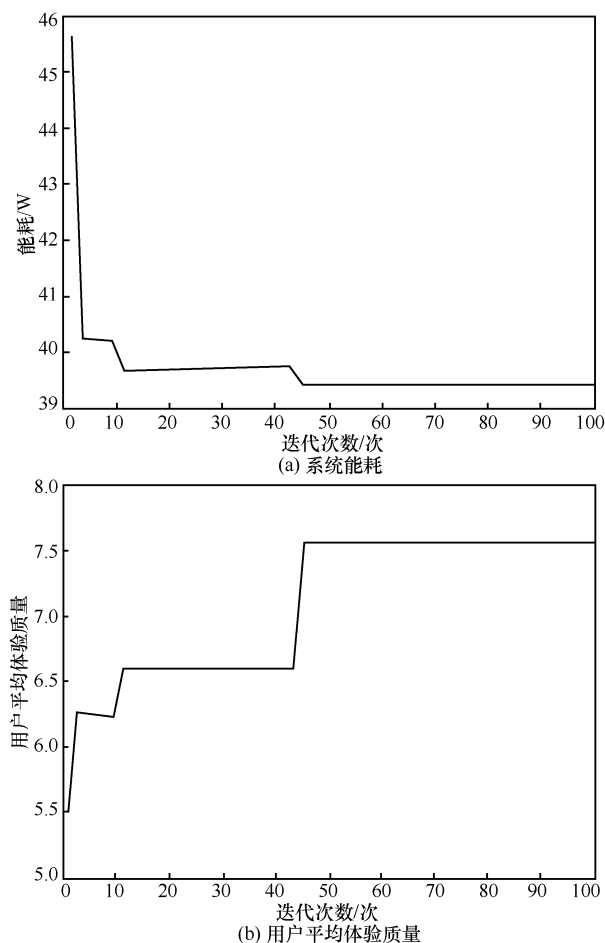


图 2 算法收敛性分析

4.2 系统能耗分析

3类不同算法的系统整体功耗如图3所示。由图3可知,随着用户数的增加,3类算法的功耗几乎并没有产生变化,尤其是SINRmax。这是由于SINRmax算法中并没有对各个网络的发射功率进行优化,即都采用最大功率进行传输,因此在网络侧的功耗并不会随着用户的增多而发生变化。此外,对于用户侧的功耗而言,从式(8)可以看出,由于可接入网络数目的限制,尽管用户数目有明显增多,但是仍然难以造成用户侧功耗的提升。而其他两类算法都可以对发射功率进行分配,但是由于用户位置的随机性和不确定性,网络的发射功率仍然难以随着用户的变化而变化。本文所提出的MUPS算法的功耗在所有算法中是最低的,这是由于MUPS算法直接以能耗为目标,相较于EEUA算法是以能效为目标,能够有效实现能耗的降低。

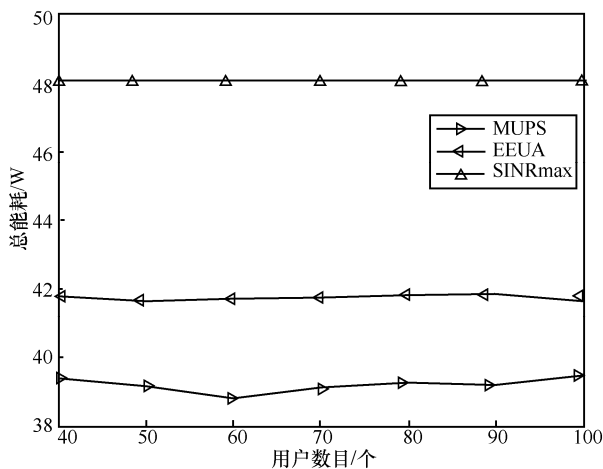


图3 3类不同算法的系统整体功耗

4.3 用户体验质量分析

3种算法的用户平均体验质量情况如图4所示。由于无线网络资源有限,随着用户数目的增多,用户难以获得充足的网络资源,因此用户平均体验质量将会降低。由图4可以看出,当用户数为40时,SINRmax的平均用户体验要高于其他两类算法,但是随着用户数的增多,SINRmax算

法中的用户体验将快速下降,当用户数分别为50和60时将开始低于MUPS算法和EEUA算法。这是由于在网络中用户数较少时,网络中干扰信道数较少,网络传输功率的最大化能够提升用户体验,但是随着用户数增多,网络之间的相互干扰将会更加严重,传输功率的最大化将会严重地劣化用户体验,从而造成用户体验质量的急速下降。相比于SINRmax算法,MUPS算法和EEUA算法的下降速率相对缓慢,但是由于EEUA算法是考虑网络传输容量,并未直接考虑用户体验质量,因此其用户平均体验质量始终低于MUPS算法。

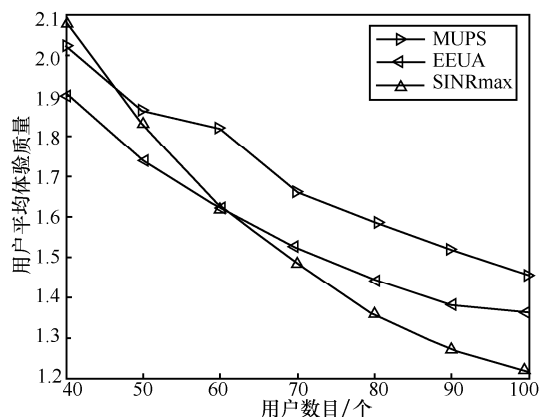
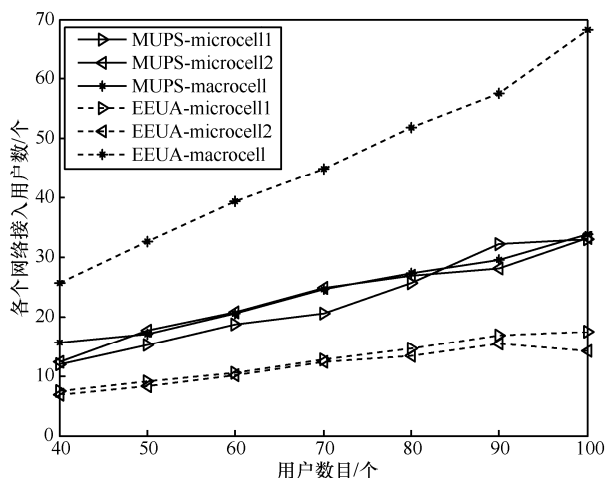


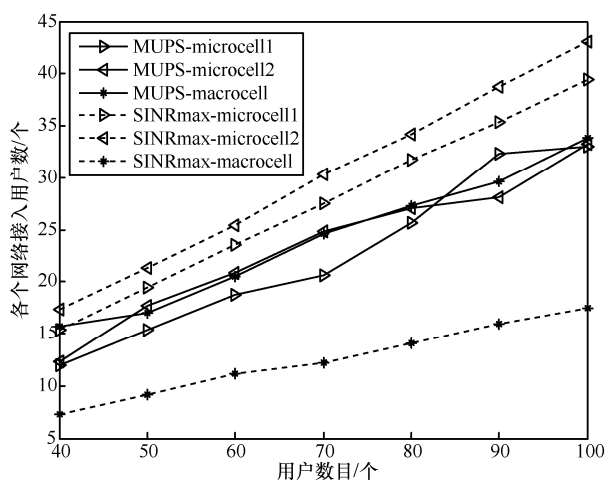
图4 用户平均体验质量情况

4.4 负载情况分析

MUPS算法的各个网络负载情况与EEUA算法和SINRmax算法的对比情况如图5所示。由图5可以看出,随着用户数增多,MUPS算法的宏蜂窝和两个微蜂窝的负载水平十分接近,呈现出了更好的均衡性,这是由于MUPS算法能够有效地折中平均用户体验和系统能耗,从而能够使得用户均衡地接入各个网络之中。在SINRmax算法中,由于微蜂窝覆盖范围较小,用户大多能够从微蜂窝处接收到更强的信号,因此,微蜂窝中的接入用户数明显高于宏蜂窝的用户数。而在EEUA算法中由于传输功率和信道状态的影响,接入宏蜂窝网络的用户数将会大于微蜂窝网络的用户数。



(a) MUPS算法与EEUA算法的对比



(b) MUPS算法与SINRmax算法的对比

图5 网络负载情况对比

5 结束语

本文针对无线网络能耗和用户体验问题提出了一种基于多目标优化的联合接入选择和功率分配算法。从用户侧和网络侧建立系统能耗模型，同时也建立以用户传输速率为变量的实时传输系统的用户体验质量模型。并在此基础上，设计了以系统能耗最大和平均用户体验最小的多目标优化问题，通过采用差分进化方法对多目标优化问题进行求解，得到最优的接入选择和功率分配 Pareto 解集，通过 TOPSIS 方法，从解集中得到最优的折中解。仿真结果表明，本文所提的算法能够有效地降低系统能耗，并提升平均用户体验质量，同时也能够保证网络的负载均衡性，具有较

好的应用价值。

参考文献:

- [1] XU C, JIN W, ZHAO G, et al. A Novel multipath-transmission supported software defined wireless network architecture[J]. IEEE Access, 2017, 5(99): 2111-2125.
- [2] NIU C, LI Y, HU R Q, et al. Fast and efficient radio resource allocation in dynamic ultra-dense heterogeneous networks[J]. IEEE Access, 2017, 5(99): 1911-1924.
- [3] 田峰, 刘翰焘, 周亮. 分层异构网络无线资源管理技术探讨[J]. 电信科学, 2013, 29(6): 32-38.
TIAN F, LIU H T, ZHOU L. Research on radio resource management for hierarchical heterogeneous wireless network[J]. Telecommunications Science, 2013, 29(6): 32-38.
- [4] 梁潇, 钱志鸿, 田洪亮, 等. 基于马尔可夫决策模型的异构无线网络切换选择算法[J]. 物理学报, 2016, 65(23): 206-214.
LIANG X, QIAN Z H, TIAN H L, et al. Markov decision model based handoff selection algorithm for heterogeneous wireless networks[J]. Acta Mathematica Scientia, 2016, 65(23): 206-214.
- [5] CASAS P, SEUFERT M, WAMSER F, et al. Next to you: monitoring quality of experience in cellular networks from the end-devices[J]. IEEE Transactions on Network & Service Management, 2016, 13(2): 181-196.
- [6] 张英杰, 章兢, 刘文强, 等. 通信行业能耗现状分析与节能战略思考[J]. 电信科学, 2011, 27(10): 103-109.
ZHANG Y J, ZHANG J, LIU W Q, et al. Overview of current state of energy consumption and energy-saving strategies exploration in communication industry[J]. Telecommunications Science, 2011, 27(10): 103-109.
- [7] SEKULOVIC N, STEFANOVIC M, DENIC D, et al. Performance analysis of signal-to-interferenceplus-noise ratio-based selection diversity over correlated rayleigh fading channels[J]. IET Communications, 2011, 5(2): 127-134.
- [8] 孙卓, 刘铁军, 张倩倩, 等. 异构无线分组网络中的接入选择机制[J]. 系统仿真学报, 2008, 20(10): 2762-2765.
SUN Z, LIU T J, ZHANG Q Q, et al. Radio access selection algorithm in heterogeneous wireless packet networks[J]. Journal of System Simulation, 2008, 20(10): 2762-2761.
- [9] TAI M H, TRAN N H, DO C T, et al. Power control for interference management and QoS guarantee in heterogeneous networks[J]. IEEE Communications Letters, 2015, 19(8): 1402-1405.
- [10] HUANG G, LIN Z, TANG D, et al. QoS-driven jointly optimal power and bandwidth allocation for heterogeneous wireless networks[J]. Electronics Letters, 2014, 51(1): 122-124.
- [11] CHITTI K, QUAN K, SPEIDEL J. Joint base station association and power allocation for uplink sum-rate maximization[C]//Signal Processing Advances in Wireless Communications, June 16-19, 2013, Darmstadt, Germany. New Jersey: IEEE Press, 2013:

- 6-10.
- [12] ZHOU T, HUANG Y, YANG L. Energy-efficient user association in downlink heterogeneous cellular networks[J]. IET Communications, 2016, 10(13): 1553-1561.
- [13] 于佳, 王野, 吴绍华, 等. 基于 CoMP 传输的分布式异构网络资源分配[J]. 电信科学, 2015, 31(9): 11-19.
YU J, WANG Y, WU S H, et al. Resource allocation of distributed heterogeneous network based on CoMP transmission[J]. Telecommunications Science, 2015, 31(9): 11-19.
- [14] YUAN D, SONG M, TENG Y, et al. QoE-oriented resource allocation for multiuser-multiservice femtocell networks[J]. China Communications, 2015, 12(10): 27-41.
- [15] CHEN F, QIN X, WEI G. QoE optimized resource allocation in multiuser OFDM systems[J]. Przegląd Elektrotechniczny, 2012, 88(7): 328-331.
- [16] WANG N, GONG S, FEI Z, et al. A QoE-based jointly subcarrier and power allocation for multiuser multiservice networks[J]. Science China, 2016, 59(12): 1-13.
- [17] 杨燕. 浅析移动通信网络中的 QoE[J]. 电信科学, 2007, 23(8): 34-38.
YANG Y. Simple analysis of QoE in mobile communication network[J]. Telecommunications Science, 2007, 23(8): 34-38.
- [18] 王大鸣, 陈松, 崔维嘉, 等. 多用户 MIMO-OFDM 系统基于 QoE 效用函数的跨层资源分配[J]. 通信学报, 2014, 35(9): 175-183.
WANG D M, CHEN S, CUI W J, et al. QoE utility function-based cross-layer resource allocation in multi-user MIMO-OFDM systems[J]. Journal on Communications, 2014, 35(9): 175-183.
- [19] WU Q, CHEN W, TAO M, et al. Resource allocation for joint transmitter and receiver energy efficiency maximization in downlink OFDMA systems[J]. IEEE Transactions on Communications, 2015, 63(2): 416-430.
- [20] 刘文颖, 文晶, 谢昶, 等. 考虑风电消纳的电力系统源荷协调多目标优化方法[J]. 中国电机工程学报, 2015, 35(5): 1079-1088.
LIU W Y, WEN J, XIE C, et al. Multi-objective optimal method considering wind power accommodation based on source-load coordination[J]. Proceedings of the CSEE, 2015, 35(5): 1079-1088.

- [21] 盛四清, 范林涛, 李兴, 等. 基于帕累托最优的配电网多目标规划[J]. 电力系统自动化, 2014, 38(15): 51-57
SHENG S Q, FAN L T, LI X, et al. Multi-objective planning of distribution network based on pareto optimality[J]. Automation of Electric Power Systems, 2014, 38(15): 51-57
- [22] LIU Y, LU L, LI G Y, et al. Joint user association and spectrum allocation for small cell networks with wireless backhauls[J]. IEEE Wireless Communications Letters, 2016, 5(5): 496-499.

[作者简介]



季石宇 (1992-), 男, 华北电力大学电气与电子工程学院博士生, 主要研究方向为无线资源管理技术。



唐良瑞 (1966-), 男, 华北电力大学电气与电子工程学院教授、博士生导师, 主要研究方向为无线通信、电力系统通信与信息处理。



李淑贤 (1992-), 女, 华北电力大学电气与电子工程学院硕士生, 主要研究方向为移动通信技术。



杜施默 (1992-), 女, 中国移动通信集团浙江有限公司杭州分公司网络工程师, 主要研究方向为通信网络流量建模。



基于元胞蝙蝠算法的无线传感器网络节点定位研究

孟凯露¹, 岳克强², 尚俊娜¹

(1. 杭州电子科技大学通信工程学院, 浙江 杭州 310018;

2. 杭州电子科技大学电子信息学院, 浙江 杭州 310018)

摘要: 为了提高节点定位精度, 解决定位误差较大的问题, 提出了基于元胞蝙蝠算法的无线传感器网络节点定位算法, 以此来获得更高的定位精度。首先将元胞自动机的思想融入蝙蝠算法, 采用了改进的元胞限制竞争选择小生境技术和灾变机制, 使得该算法在寻优过程中能够跳出局部极值, 避免早熟现象, 更快地收敛到全局最优解。通过标准测试函数的验证, 表明了该改进算法在收敛深度和广度上的优势。之后将元胞蝙蝠算法应用到无线传感器网络节点定位上来提高定位精度。实测实验中, 该算法在测试环境下平均定位误差在 0.4 m 以内, 相比于改进 PSO 算法, 获得更好的定位效果。

关键词: 无线传感器网络; 节点定位; 元胞自动机; 蝙蝠算法; 定位精度

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017311

Wireless sensor network nodes localization method based on cellular automata bat algorithm

MENG Kailu¹, YUE Keqiang², SHANG Junna¹

1. College of Telecommunication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China

2. College of Electronic Information, Hangzhou Dianzi University, Hangzhou 310018, China

Abstract: To further enhance the location precision of unknown nodes and solve the node location error in wireless sensor network, a location method based on cellular automata bat algorithm was presented. Mixed the idea of cellular automata and the bat algorithm and drew into the cellular RCS niche technique and disaster mechanism, the algorithm could jump out of local optimum and increase the convergence speed. In order to verify the feasibility and efficiency, the proposed algorithm was verified through simulation of several benchmark functions. Then the algorithm implemented the CA-BA to node location prediction to increase the precision of the unknown node location. In the measured experiment, the results show that the proposed algorithm has higher accuracy compared to the improved PSO algorithm, which average localization error is less than 0.5m.

Key words: WSN, node localization, cellular automata, bat algorithm, accuracy

收稿日期: 2017-09-02; 修回日期: 2017-09-26

通信作者: 岳克强, 290035613@qq.com

基金项目: 国家自然科学基金资助项目 (No.11603041); 广西精密导航技术与应用重点实验室开放基金资助项目 (No.DH201714); 浙江省“电子科学与技术”重中之重学科开放基金资助项目 (No.GK13020320003/004); 杭州电子科技大学研究生科研创新基金资助项目 (No.ZX170603308034)

Foundation Items: The National Natural Science Foundation of China (No.11603041), Open Funds by Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin University of Electronic Technology (No.DH201714), “Electronic Science and Technology” of the Most Important Subject Open Funds in Zhejiang Province (No.GK13020320003/004), Research and Innovation Fund for Graduate Students of Hangzhou Dianzi University (No.ZX170603308034)

1 引言

无线传感器网络^[1]由能够进行通信和计算的无线传感器节点组成, 由于其成本低、功能多、融合多门技术, 被誉为 21 世纪最具有影响力的技术之一。根据是否需要测量距离, 目前无线传感器节点定位算法可以分为两种定位机制: 基于测距和无需测距。基于测距的定位主要是通过距离来确定节点的位置, 无需测距则不需要知道节点的距离和方向, 只能通过估计定位节点的位置, 而基于测距的定位精度远远高于无需测距的方法, 将会有更好的发展空间。测距的基本方法包括接收信号强度指示 (received signal strength indication, RSSI)、到达时间差 (time difference of arrival, TDOA)、到达角度 (angle of arrival, AOA)。RSSI^[2]因其使用方便和精度高而被广泛应用。无线传感器的节点定位问题可以转换成一个多约束优化问题, 当前有学者采用智能算法来提高节点定位的精度^[3], 但常用的粒子群算法、遗传算法等智能优化算法易陷入局部最优, 定位精度、效果的提高不明显, 因此需要研究一种性能好、效率高的智能算法, 从而有效地提高定位精度。

蝙蝠算法 (bat algorithm, BA) 模拟自然界中蝙蝠利用一种声纳来寻找猎物 and 避免障碍物的生物学特性, 具有并行性、分布式和收敛速度快等特点^[4]。目前, 已经广泛应用于工程设计、分类、模糊聚类、预测和神经网络等领域。但是基础 BA 算法的搜索区域单一, 缺乏有效方法保持种群的多样性, 搜索性能较差。基于此, 参考文献[5]将改进的蝙蝠算法应用于 Dv-Hop 定位算法平均跳距的计算过程中, 使得该算法在不同锚节点密度、不同通信半径、不同节点数量以及定位精确度等方面的性能明显增强。参考文献[6]采用蝙蝠算法的拓扑结构来构建多智能体的体系结构, 通过多智能体蝙蝠算法来求解室内定位, 大大提高了算法的收敛速度和效率。多智能体之间虽然可以发生

决策交互, 也可以与环境相互作用, 但是在深度搜索和保持个体多样性方面仍有待提高。

元胞自动机 (cellular automata) 是时间、空间上均表现为离散的动力系统模型, 多个元胞在同一个环境里共同生存, 且每个元胞能够主动地、自治地活动, 基于此本文提出一种元胞蝙蝠算法 (cellular automata bat algorithm, CA-BA), 将 CA 的思想融入 BA 算法^[7,8]中, 把每个蝙蝠个体看作一个元胞, 同时采用了改进的元胞限制竞争选择 (restricted competition selection, RCS) 小生境技术, 将元胞与其邻域看作一个小生境环境, 通过小生境环境内的最优解进行速度和位置的更新, 并加入了灾变机制, 使元胞个体能够较快地跳出局部极值从而避免早熟现象, 提高了算法的效率。通过对 6 个标准测试函数的仿真, 改进的 CA-BA 算法在寻优精度和收敛速度上都比 BA 算法、CBA 算法以及 MA-BA 算法更胜一筹, 同时将其应用于无线传感器网络节点定位中, 节点的平均定位误差相对较小, 并且可以使用相对较少的锚节点达到理想的定位精度。

2 CA-BA 算法

2.1 基本 BA 算法

由 Yang 教授于 2010 年提出的蝙蝠算法^[9]是一种启发式的群体智能算法, 基于蝙蝠的回声定位能力指导它们的捕食行为。在实际的优化算法中, 每只蝙蝠为搜索解空间的一个点, 由适应度函数来决定蝙蝠位置的优劣。蝙蝠个体的每一次有效飞行就是 BA 算法的一次迭代更新。蝙蝠的位置和速度更新如下:

$$\begin{cases} f_i = f_{\min} + (f_{\max} - f_{\min})\beta \\ v_i^t = v_i^{t-1} + (x_i^t - x^*)f_i \\ x_i^t = x_i^{t-1} + v_i^t \end{cases} \quad (1)$$

其中, f_i 是第 i 只蝙蝠个体所发出的脉冲频率, 且 $f_i \in [f_{\min}, f_{\max}]$, $\beta \in [0, 1]$, 是均匀分布的随机向量。对于当前的局部搜索区域来说, 随机



选取个体按式 (2) 进行最优解扰动:

$$X_{\text{new}} = x^* + \varepsilon A' \quad (2)$$

其中, ε 为 $[-1, 1]$ 上的随机数值, A' 为在时刻 t 所有蝙蝠个体的平均响度。当发现猎物的时候, 蝙蝠个体则降低脉冲响度和增加脉冲频度。在 BA 算法中, 响度 A_i 和脉冲频度 r_i 按式 (3) 进行更新:

$$\begin{cases} A_i^{t+1} = \omega A_i^t \\ r_i^{t+1} = r_i^0 [1 - \exp(-\gamma t)] \end{cases} \quad (3)$$

其中, r_i^0 表示初始脉冲频度, γ 是发射脉冲频度增加的系数, ω 是脉冲响度衰减系数, 两者都为常数, 其中, $0 < \omega < 1$, $\gamma > 0$ 。

2.2 CA-BA 算法

元胞自动机^[10]是一种时间和空间都离散的动力系统, 散布在规格网格中的每一元胞取有限的离散状态, 依据确定的局部规则作同步更新。本文将 CA 的思想应用到 BA 算法中, 把 BA 算法中的个体看作元胞自动机系统 (cellular automata system, CAS) 中的元胞, 以此来研究种群的交流结构以及信息传递与继承机制。根据元胞自动机的原理和规则, 蝙蝠个体只与通过邻居函数确定的邻域个体进行交流^[11], 并且能够由自身的局部搜索逐步延伸到全局搜索, 这样使得寻优过程能够摆脱局部极值从而避免早熟, 有助于保持种群的多样性和探索能力, 更准确迅速地收敛到全局最优解。

2.2.1 元胞自动机与蝙蝠算法的混合机制

在所提出的 CA-BA 算法中, 把 BA 算法中的一个蝙蝠个体看作一个元胞, 且这个元胞只能与通过邻居函数确定的邻域元胞进行交流、竞争与合作, 根据邻域环境更新自身的位置和速度。本文选取的是结构简单稳定的四方网格排列, 如图 1 所示。每一个网格就表示 CA-BA 算法中的一个元胞, 即一个蝙蝠个体。元胞在网格中的位置固定不变, 每一个网格包含的信息是元胞的固定位置

以及其在解空间的位置和速度。并且元胞的边界条件是周期型 (periodic boundary) 的, 即对于二维空间而言, 上下相连, 左右相连, 元胞蝙蝠个体以相邻的 8 个元胞为邻居, 即 Moore 邻居形式, 假设图 1 中相邻两个方格的距离为 1, 它的邻域元胞可以由式 (4) 被定义:

$$N_{\text{Moore}} = \{v_i = (i' j') \mid |i - i'| \leq 1, |j - j'| \leq 1, (i, j) \in Z^2\} \quad (4)$$

其中, i' 、 j' 表示邻居元胞的行列坐标值, i 、 j 表示中心元胞的行列坐标值。此种 CA-BA 算法中蝙蝠元胞的重叠方式提供了一种隐性的迁移机制, 有利于最优个体在种群中缓慢地扩散, 能更持久地保持种群多样性。

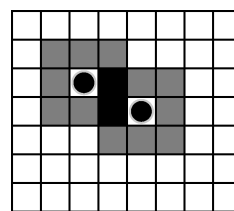


图 1 四方网格

2.2.2 改进的元胞 RCS 小生境技术

在基础 BA 算法中, 算法前期收敛速度很快, 但随着迭代次数的增加, 蝙蝠种群聚集现象比较严重, 蝙蝠个体的速度逐渐减小并趋于 0, 这导致算法后期易陷入局部极值而停滞不前, 因此如何保持种群个体的多样性对蝙蝠算法进行深度搜索具有十分重要的意义。基于上述分析本文引入小生境策略来维持种群的多样性。

将元胞 $X_{i,j}$ 与其邻域个体 $M_{i,j}(\text{num}) = (m_{1,\text{num}}, m_{2,\text{num}}, \dots, m_{N,\text{num}})$ 其中, num 为 1~8 的整数, 看作一个小生境环境, 记录其邻域最优元胞为 $M_{i,j}$, 首先比较 $X_{i,j}$ 与 $M_{i,j}$ 的适应度值。按照以下步骤对元胞个体进行进化:

若 $f(X_{i,j}) < f(M_{i,j})$, 则元胞 $X_{i,j}$ 在小生境环境内是最优的, 保留 $X_{i,j}$ 在解空间中的位置。同时按照式 (5) 分别比较 $X_{i,j}$ 与 8 个元胞邻居的海

明距离, 并将距离值按照从大到小排序。

$$d(num) = \|X_{i,j} - M_{i,j}\| = \sqrt{\sum_{i=1}^N (x_i - m_{i,num})^2} \quad (5)$$

元胞邻居个体中距离最小的邻域元胞记为 $C_{i,j}$, 对于相似性较强的元胞蝙蝠个体采用 $L'evy$ 扰动减少了陷入局部极值, 更新计算式为:

$$c'_i = (1 + L'evy(0,1))c_i \quad (6)$$

其中, $L'evy(0,1)$ 为 $L'evy$ 分布函数, 通过扰动不仅加快了算法向最优解区域靠近的速度, 而且提高了算法跳出局部极值的能力和寻优精度, 提高了算法的寻优性能。

若 $f(X_{i,j}) > f(M_{i,j})$, 仍按照式 (5) 计算距离, 并将与 $X_{i,j}$ 距离最近的元胞 $C_{i,j}$ 按照式 (6) 进行赋值。为了保证元胞个体能在邻居范围内加快搜索速度, 在有效信息的基础上吸收了最优邻居 $M_{i,j}$ 的信息, 本文中在解空间中按照式 (7) 对 $X_{i,j}$ 进行重新赋值:

$$x'_k = m_k + rand(-1,1) \times (m_k - x_k), k = 1, 2, \dots, N \quad (7)$$

通过 RCS 小生境技术, 保证了元胞个体向最优个体的学习, 元胞与邻域元胞互相传递有效信息, 并且通过不同邻域中重叠的元胞将有效信息传递到整个空间网格中, 从而与全局最优元胞进行信息的交流。并且在信息交流的过程中, 元胞的行动策略得以修正, 从而增强了局部搜索能力和加快了算法收敛到最优解。同时通过 $L'evy$ 扰动, 元胞邻居个体位置的随机更新, 提高了种群的多样性, 避免了算法陷入局部最优的情况。

2.2.3 元胞蝙蝠个体灾变机制

为避免 CA-BA 算法出现早熟现象, 本文采用了一种简单有效的方法, 在模拟自然界的状况下进行灾变操作^[12,13]。基本思想为: 以某一确定的频率对固定面积的元胞进行扰动。在该扰动半径内的个体会发生变化, 使其跳出局部最优解, 开始进行新的搜索。具体操作是: 选取

$rem(generation) = 10$, 即每隔 10 代对群体进行一次扰动, 选取扰动面积为 1×1 , 即扰动半径为 1。将灾难发生区域的个体进行清零操作, 并随机植入优良个体。引入的灾变机制, 可以使 CA-BA 算法跳出局部极值, 保持持续进化能力。

2.3 CA-BA 算法步骤

步骤 1 初始化种群。按照图 1 的形式构造元胞的四方网格排列方式, 即给定种群的生存环境和规模, 在限制范围内随机初始化元胞的速度 v 、位置 x 和发射脉冲频率 f , 并给定初始脉冲频度 r_0 和脉冲响度 A_0 。

步骤 2 计算所有元胞的适应度值 $Fitness(i)$ 。用计算出的适应度值更新元胞的极值, 并比较各个元胞的极值, 求出种群的全局最优解 $Cell(x^*)$ 和其对应的适应度值 $f(x^*)$ 。

步骤 3 构造小生境环境, 调整元胞的位置。利用式 (4) 计算出元胞 $X_{i,j}$ 的 8 个邻居, 并将其看作一个小生境环境, 找出邻域的最优元胞为 $M_{i,j}$ 。若 $X_{i,j}$ 的适应度值优于 $M_{i,j}$, 则保留 $X_{i,j}$ 在解空间中的位置, 并记为 $neigh(x^*)$, 且按式 (6) 改变与 $X_{i,j}$ 距离最近的元胞 $C_{i,j}$ 的位置; 若 $M_{i,j}$ 的适应度值优于 $X_{i,j}$, 则记 $M_{i,j}$ 的位置为 $neigh(x^*)$, 并按式 (6) 和式 (7) 改变 $C_{i,j}$ 和 $X_{i,j}$ 在解空间中的位置。

步骤 4 通过式 (1) 的 3 个计算式完成所有元胞的一次迭代过程, 即更新每个元胞在解空间中的位置 x 和速度 v , 注意式 (2) 中的全局最优值 x^* 要用邻域最优解 $neigh(x^*)$ 替代, 并记录更新后元胞的适应度值为 F_{new} 。

步骤 5 生成均匀分布随机数 $rand1$, 若 $rand1 > r(i)$, 则根据式 (2) 对当前最优解进行随机扰动, 并对新解进行越界处理和计算其适应度值。如果不满足条件则跳过该步。

步骤 6 生成均匀分布随机数 $rand2$, 若满足条件 $F_{new} < Fitness(i) \& rand2 < A(i)$, 则接受步骤 5 产生的新解, 并根据式 (3) 更新脉冲响度 A 和发射脉冲的频度 r 。如果不满足条件表示该元胞的



位置更新失败, 跳过该步。

步骤 7 比较新解和原来全局最优解的适应度值, 若满足 $F_{\text{new}} < f(x^*)$, 则进行全局最优解 $Cell(x^*)$ 和其对应的适应度值 $f(x^*)$ 的更新。否则跳过该步。

步骤 8 若 $generation = rem(generation) = 10$ (即每隔 10 代发生一次灾变), 将灾难区域个体清空, 并植入优良个体。

重复步骤 2~步骤 7, 直至满足最优解要求的精度或达到最大迭代次数, 当满足要求后输出全局最优解以及其对应的最小适应度值。

2.4 CA-BA 算法性能测试

智能算法的性能一般由标准测试函数进行验证, 因此本文选取了 6 个标准测试函数对 CA-BA 算法的性能进行仿真测试, 同时与基本 BA 算法、具有混沌搜索策略的蝙蝠算法 (chaos bat algorithm, CBA)^[4]以及参考文献[6]所改进的多智能体蝙蝠算法 (multi-agent bat algorithm, MA-BA) 进行对比。

2.4.1 参数设置

蝙蝠算法中各种参数设置目前没有明确的理论依据, 因此本文所设置的参数值根据反复实验获得的经验值来确定。CA-BA 算法中: 元胞总数 $N=64$, 四方网格的大小为 8×8 , 维数 $d=5$, 搜索

脉冲频率范围 $[0, 100]$, 最大脉冲频度 $r=0.5$, 最大脉冲音强 $A=0.25$, 脉冲音强衰减系数 $\gamma=0.95$, 脉冲频度增加系数 $\alpha=0.05$ 。BA 算法、CBA 算法、MA-BA 算法中的各个参数与 CA-BA 算法保持一致。这 3 个算法的最大迭代次数都是 200 次, 每种算法独立运行 50 次。其中, CBA 表示具有混沌搜索策略的蝙蝠算法, MA-BA 表示多智能体蝙蝠算法, CA-BA 表示元胞蝙蝠算法。

2.4.2 标准测试函数

仿真分析选取的测试函数的解析式和解空间搜索区域见表 1。对于仿真所用的标准函数可以分为 2 类: Quadric、Rosenbrock、Schwefel 是单模函数, 是测试算法全局收敛性能的经典函数; Ackley、Griewank、Rastrigin 是具有较多局部极值的多模函数, 主要用来评价算法的探索、开发能力。

2.4.3 仿真效果分析

图 2 显示了以上 6 个标准函数在 BA、CBA、MA-BA 以及 CA-BA 算法中的寻优曲线。CA-BA 算法采用了改进的元胞 RCS 小生境技术, 将元胞和其邻域个体看作一个小生境环境, 通过小生境环境中的最优解来更新自身的速度与位置, 具有较强的适应性和搜索能力, 并且采用了灾变机制,

表 1 标准测试函数

测试函数	函数表达式	搜索空间	理论最优解
Quadric	$f_1(x) = \sum_{i=1}^D \sum_{j=1}^I (x_j)^2$	$[-30, 30]$	0
Rosenbrock	$f_2(x) = \sum_{i=1}^{D-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	$[-2.048, 2.048]$	0
Schwefel	$f_3(x) = -\sum_{i=1}^n (x_i \sin \sqrt{ x_i })$	$[-100, 100]$	-418.982 9n
Ackley	$f_4(x) = 20 + e - 20e^{-0.2\sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}} - e^{\sum_{i=1}^n \cos 2\pi x_i}$	$[-30, 30]$	0
Griewank	$f_5(x) = \frac{1}{4000} \sum_{i=1}^D x_i^2 - \prod_{i=1}^D \cos(\frac{x_i}{\sqrt{i}}) + 1$	$[-600, 600]$	0
Rastrigin	$f_6(x) = \sum_{i=1}^D [x_i^2 - 10 \cos(2\pi x_i) + 10]$	$[-5.12, 5.12]$	0

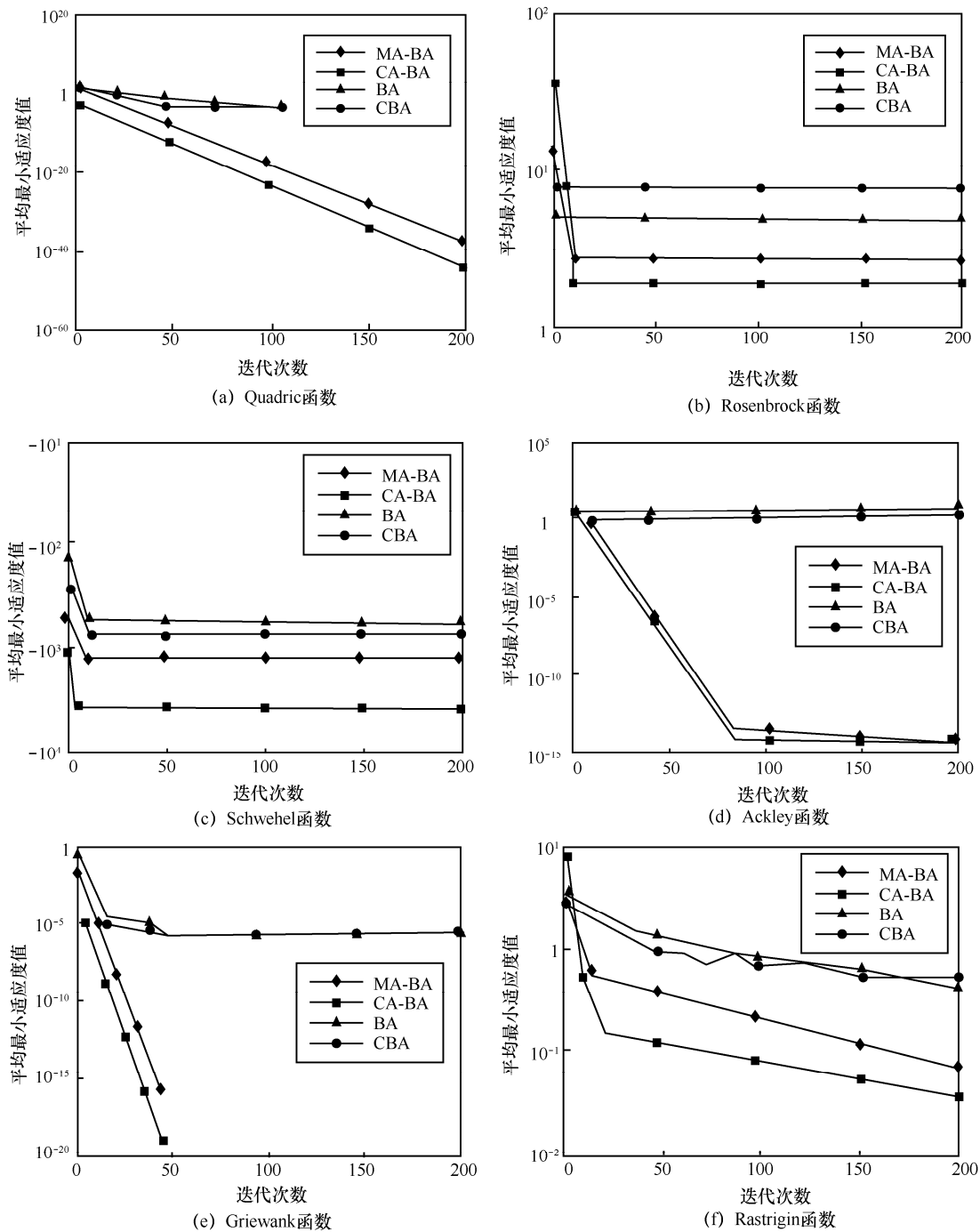


图2 标准测试函数寻优曲线

避免了元胞因陷入局部最优解而难以跳出的情况,该算法能够更高效地收敛到全局最优解。从图2的仿真结果可以看出,相比其他3个算法来说,CA-BA算法在寻优精度和收敛速度上更胜一筹,在较少的迭代次数内达到所要求的最优解。

这些都体现出了将CA的思想融入BA算法之后,大大加强了算法的搜索能力,使其能够避免早熟,更快地收敛到全局最优解。

在仿真参数如上文中所设置的情况下,将上述标准函数测试的寻优结果取100次进行平均,见表2,



表 2 标准函数测试结果

函数 $f_1 \sim f_6$	BA		CBA		MA-BA		CA-BA	
	平均最小适应度值	成功率	平均最小适应度值	成功率	平均最小适应度值	成功率	平均最小适应度值	成功率
Quadric	0.339 1	0	0.232 0	0	$1.100\ 4 \times 10^{-39}$	40%	$7.120\ 4 \times 10^{-43}$	80%
Rosenbrock	24.789 7	0	6.521 0	21%	0.099 84	47%	0.078 24	70%
Schwehel	-9.77×10^2	50%	-1.16×10^3	63%	$-1.573\ 9 \times 10^3$	100%	$-2.315\ 9 \times 10^3$	100%
Ackley	1.879 6	0	1.245 5	13%	$5.968\ 1 \times 10^{-12}$	77%	$5.267\ 5 \times 10^{-13}$	97%
Griewank	0.003 9	43%	$5.540\ 6 \times 10^{-5}$	46%	$8.278\ 5 \times 10^{-17}$	98%	0	100%
Rastrigin	32.328 3	0	13.929 4	10%	2.023 1	67%	1.691 4	87%

同时用成功率来表示算法的稳定性,对于单模函数,是以前理论最优值的 1‰为标准,多模函数以最优理论值的 5‰为标准,如果最小适应度值小于以上标准则是取得了正确解。从表 2 可以看出,CA-BA 算法的寻优成功率相比 BA 算法、CBA 算法和 MA-BA 算法来说较高,平均最小适应度值较接近于理论最优值。

3 CA-BA 算法在无线传感网络节点定位中的应用

通过上述分析,验证了 CA-BA 算法在测试函数中的性能,接下来将 CA-BA 算法应用在无线传感网络节点定位问题上,通过该算法在最优解寻优精度上的结果来验证其性能^[14]。对于无线传感网络节点网络中共有 M 个已知节点,其中, $(x_1, y_1), (x_2, y_2), \dots, (x_M, y_M)$ 为 M 个已知节点的坐标, (x, y) 为任一未知节点的坐标, $d_1, d_2, \dots, d_M$ 分别为各个锚节点到该未知节点坐标的距离。在实际的测距中存在误差,因此可以将定位问题转化为式(8)求测距误差最小值的优化问题:

$$f(x, y) = \sum_{i=1}^M |\sqrt{(x-x_i)^2 + (y-y_i)^2} - d_i| \quad (8)$$

由上述仿真分析可知,CA-BA 算法能够避免元胞陷入局部最优值以及具有较好的收敛性能,因此将该算法应用到实际测距中,即求式(8)的最小值,能够减小实际测距误差带来的影响,从而提高未知节点的定位精度。

3.1 仿真分析

仿真实验以 MATLAB 2012 为平台,节点的通信半径设置为 10 m, 30 个传感器随机分布在 $30\text{ m} \times 30\text{ m}$ 的四方网格区域内,测距误差在 20% 以下,锚节点个数为 5,仿真参数与上文设置相同,最大迭代次数为 200 次,每一个节点定位预测取 20 次平均值。将 CA-BA 算法与参考文献[6]中的 MA-BA 算法进行节点定位结果的对比,见表 3。本算法在 RSSI 的基础上对定位精度进行优化。在仿真中,以平均定位误差的绝对值 AVE 作为评判的标准,其计算式如式(9)所示:

$$AVE = \frac{1}{M} \sum_{i=1}^M \sqrt{(x-x_i)^2 + (y-y_i)^2} \quad (9)$$

其中, M 为已知节点的总个数, (x, y) 为预测位置, (x_i, y_i) 为实际位置。

表 3 选取了 10 个未知节点预测的具体数值作比较,由表 3 可知,本文所提出的基于 CA-BA 算法的节点定位精度比参考文献[6]中的节点定位精度平均提高了 0.08 m,其在定位精度上具有优势。

测距误差和不同锚节点个数对基于 RSSI 测距定位算法的节点定位精度影响很大,下面分别分析两者的影响。当锚节点个数为 10 个,测距误差分别由 0 到 30% 的平均定位误差如图 3 所示。由图 3 可以看出,在较小的测距误差下,BA、CBA、MA-BA、CA-BA 的平均定位误差几乎接近于 0,能达到较高的定位精度。但随着测距误差的不断增大,不同算法间的平均定位误差的差异开始明显变大,CA-BA 算法由于引入了元胞自动机的思

表 3 未知节点预测的具体数值比较

未知节点标号	实际位置/m	MA-BA		CA-BA	
		预测值/m	AVE/m	预测值/m	AVE/m
1	(6.104 4, 4.334 5)	(6.296 8, 4.450 7)	0.224 8	(6.185 3, 4.343 9)	0.081 4
2	(1.049 1, 0.870 4)	(1.187 3, 0.952 3)	0.160 6	(1.037 8, 0.855 5)	0.018 7
3	(9.499 7, 1.784 5)	(9.345 6, 1.968 5)	0.240 0	(9.507 8, 1.803 9)	0.021 0
4	(3.765 0, 2.711 1)	(3.891 3, 2.931 4)	0.253 9	(3.805 2, 2.739 1)	0.049 0
5	(8.176 7, 4.902 0)	(8.313 5, 5.172 9)	0.303 5	(8.146 1, 4.833 2)	0.075 3
6	(23.042 9, 23.027 5)	(23.179 5, 23.322 5)	0.325 1	(23.029 9, 23.315 9)	0.288 7
7	(26.662 8, 3.286 0)	(26.815 7, 3.495 7)	0.259 5	(26.682 8, 3.296 0)	0.022 4
8	(26.0103, 25.2551)	(26.318 5, 25.354 4)	0.323 8	(26.257 9, 25.287 9)	0.249 8
9	(23.480 0, 15.824 7)	(23.290 1, 15.687 1)	0.234 5	(23.416 0, 15.907 3)	0.104 5
10	(28.758 0, 4.574 6)	(28.587 3, 4.407 9)	0.2386	(28.752 5, 4.591 5)	0.017 8

想, 并且采用了改进的元胞 RCS 小生境技术和灾变机制, 避免了算法陷入局部最优值, 增强了算法的全局搜索能力, 使得其性能优于 BA 算法、CBA 算法以及 CA-BA 算法。当测距误差大于 10% 以后, BA 和 CBA 算法的平均定位误差大幅增长, MA-BA 的算法虽然变化平缓但误差值也大于 CA-BA 算法, 由此可见, CA-BA 算法的容错能力和稳定性都较强。

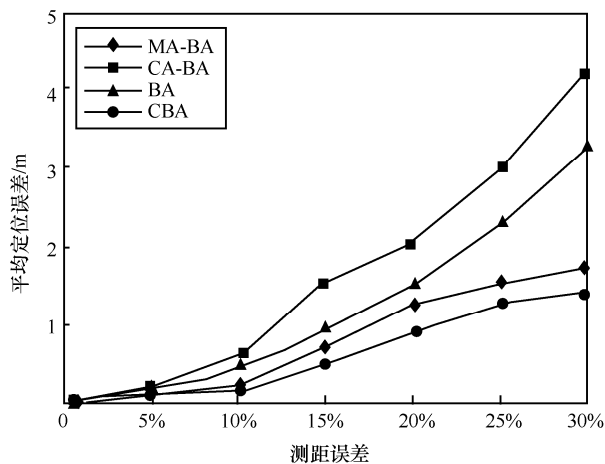


图 3 测距误差对定位性能的影响

在测距误差固定为 10% 的情况下, 锚节点个数由 4 个增加到 10 个, 不同算法的平均定位误差的情况如图 4 所示。由图 4 可以看出, 在相同锚节点个数的情况下, CA-BA 算法的平均定位误差明显小于 BA 算法、CBA 算法以及 MA-BA 算法, 精度

较高。从整体趋势看, 随着锚节点个数的增加, CA-BA 算法的整条曲线变化缓慢, 可以利用较少的锚节点来实现高精度的节点定位预测。由此可见, 在相同的定位精度要求下, CA-BA 算法可以使用较少的硬件设备, 节约成本, 应用前景广泛。

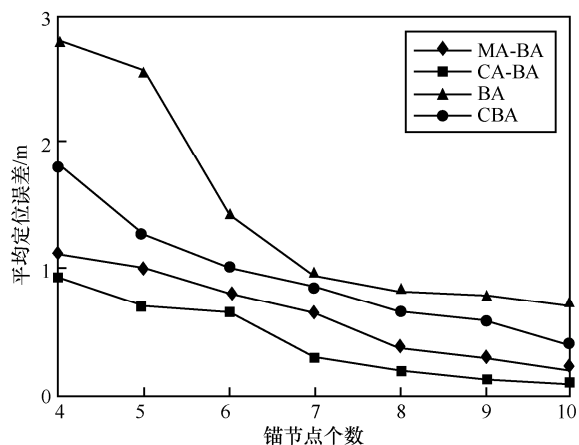


图 4 锚节点个数对定位性能的影响

3.2 实际测试分析

为了进一步验证 CA-BA 算法在实际使用中的效果, 本文通过在实际环境的定位实验来验证算法的性能, 测试中采用 CC2530 模块搭建 WSN 定位实验平台, RSSI 的读取如参考文献[15]方法所示, 通过将 RSSI 转换为距离值。实验中使用 8 个锚节点 ($T_1 \sim T_8$) 分布在边长为 6 m 的正方形边界上, 将该正方形区域均分成九宫格, 对每个



格子的中间点进行定位测试,如图5所示,实测如图6所示。

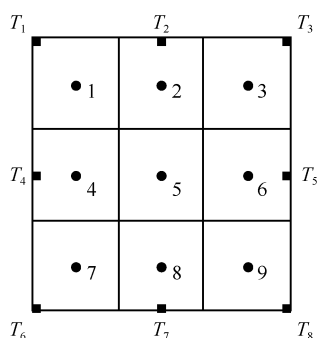


图5 定位实验点布置



图6 定位实验点实测

将实验结果与参考文献[15]中的改进 PSO 算法进行对比,定位误差如图7所示。由图7可见,本文所提出的 CA-BA 算法的平均误差在 0.4 m 左右,改进 PSO 算法的平均定位误差为 0.5 m 左右,CA-BA 定位效果要优于改进 PSO 算法的性能,实验结果表明本文所提出的 CA-BA 算法在后期的寻优精度上有较大的提高。

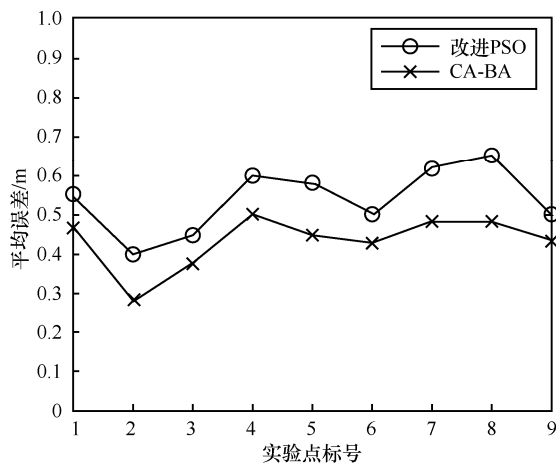


图7 锚节点为8时平均误差对比

表4给出了本文 CA-BA 算法和改进 PSO 算法的定位结果平均误差的均方根值。由表4可知,在8个锚节点的情况下,改进的 PSO 的均方根值为0.013 0,本文CA-BA算法的均方根值为0.001 1,由结果可知本文所提 CA-BA 算法性能有较大提高。

表4 定位实验误差均方根值

算法	8个锚节点
改进 PSO	0.013 0
CA-BA	0.001 1

4 结束语

在本文中,首先介绍了基本 BA 算法在开发和探索能力上性能较差,较易陷入局部极值而失去种群多样性的探索,据此提出了 CA-BA 算法,将元胞自动机的思想应用到 BA 算法中,并采用了改进的元胞 RCS 小生境技术和灾变机制,使元胞个体能够较快地跳出局部极值从而避免早熟现象,提高了算法的进化效率,能够更快地收敛到全局最优解。通过对6个标准函数的仿真测试,并与 BA 算法、CBA 算法以及 MA-BA 算法进行对比,验证了 CA-BA 算法在寻找全局最优解和收敛速度方面有很大的提高,并相对其他算法有所改善。将元胞蝙蝠算法应用到无线传感器网络节点定位上,由仿真分析可以看出,在其他因素相同的情况下,测距误差对 CA-BA 算法的影响较小,并且其能在较少的硬件设备的情况下达到所要求的定位精度;由测试分析可以看出,CA-BA 算法定位的平均误差较小,在定位精度上具有优势。

参考文献:

- [1] PENG L J, LI W W. The Improvement of 3D wireless sensor network nodes localization[C]//26th IEEE Chinese Control and Decision Conference, May 31-June 2, 2014, Changsha, China. New Jersey: IEEE Press, 2016: 4873-4878.
- [2] 方震, 赵湛, 郭鹏, 等. 基于 RSSI 测距分析[J]. 传感技术学报, 2007(11): 2526-2530.

- FANG Z, ZHAO Z, GUO P, et al. Analysis of distance measurement based on RSSI[J]. Chinese Journal of Sensors and Actuators, 2007, 20(11): 2526-2530.
- [3] 焦磊, 邢建平, 张军, 等. 一种非视距环境下具有鲁棒特性 TOA 无线传感网络定位算法[J]. 传感技术学报, 2007, 20(7): 1625-1629.
- JIAO L, XING J P, ZHANG J, et al. A new NLOS TOA-based wireless sensor network localization algorithm with robust character[J]. Chinese Journal of sensors actuators, 2007, 20(7): 1625-1629.
- [4] 刘长平, 叶春明. 具有混沌搜索策略的蝙蝠优化算法及性能仿真[J]. 系统仿真学报, 2013, 25(6): 1183-1188, 1195.
- LIU C P, YE C M. Bat algorithm with chaotic search strategy and analysis of its property[J]. Journal of System Simulation, 2013, 25(6): 1183-1188, 1195.
- [5] 赖锦辉. 基于蝙蝠优化算法的无线传感器网络节点定位研究[J]. 计算机测量与控制, 2014, 22(8): 2709-2712.
- LAI J H. Research on nodes localization method for wireless sensor networks based on bat optimization algorithm[J]. Computer Measurement & Control, 2014, 22(8): 2709-2712.
- [6] 尚俊娜, 刘春菊, 岳克强, 等. 多智能体蝙蝠算法在无线传感器中的应用[J]. 传感技术学报, 2015, 28(9): 1418-1424.
- SHANG J N, LIU C J, YUE K Q, et al. The multi-agent bat algorithm applied to wireless sensor network[J]. Chinese Journal of Sensors and Actuators, 2015, 25(9): 1418-1424.
- [7] 朱大林, 詹腾, 张屹, 等. 多策略差分进化的元胞多目标粒子群算法[J]. 电子学报, 2014, 42(9): 1831-1838.
- ZHU D L, ZHAN T, ZHANG Y, et al. Cellular multi-objective particle swarm algorithm based on multi-strategy differential evolution[J]. Acta Electronica Sinica, 2014, 42(9): 1831-1838.
- [8] 张屹, 万兴余, 郑小东, 等. 基于正交设计的元胞多目标遗传算法[J]. 电子学报, 2016, 44(1): 87-94.
- ZHANG Y, WAN X Y, ZHENG X D, et al. Cellular genetic algorithm for multiobjective optimization based on orthogonal design[J]. Acta Electronica Sinica, 2016, 44(1): 87-94.
- [9] YANG X S. A new metaheuristic bat-inspired algorithm[J]. Nature Inspired Cooperative Strategies for Optimization, 2010(284): 65-74.
- [10] 石杨. 元胞粒子群优化算法及其在柔性作业车间调度中的应用[D]. 武汉: 华中科技大学, 2010: 8-10.
- SHI Y. A thesis submitted in partial fulfillment of the requirements for the degree of master of engineering[D]. Wuhan: Huazhong University of Science and Technology, 2010: 8-10.
- [11] 张俞. 元胞遗传算法的研究[D]. 江西: 南昌航空大学, 2009: 37-41.
- ZHANG Y. Research for cellular genetic algorithm[D]. Jiangxi: Nanchang Hangkong University, 2009: 37-41.
- [12] 李新鹏, 张超勇, 高亮, 等. 基于元胞粒子群算法的数控切削参数优化[J]. 计算机工程与应用, 2014, 50(2): 252-257.
- LI X P, ZHANG C Y, GAO L, et al. NC cutting parameter optimization based on cellular particle swarm optimization algorithm[J]. Computer Engineering and Applications, 2014, 50(2): 252-257.
- [13] 鲁宇明, 陈殊, 黎明, 等. 自适应调整选择压力的灾变元胞遗传算法[J]. 系统仿真学报, 2013, 25(3): 436-444.
- LU Y M, CHEN S, LI M, et al. Self-adaptive cellular genetic algorithms with disaster based on selection pressure[J]. Journal of System Simulation, 2013, 25(3): 436-444.
- [14] 詹杰, 刘宏立, 刘述钢, 等. 基于 RSSI 的动态权重定位算法研究[J]. 电子学报, 2011, 39(1): 82-88.
- ZHAN J, LIU H L, LIU S G, et al. The study of dynamic degree weighted centroid localization algorithm based on RSSI[J]. Acta Electronic Sinica, 2011, 39(1): 82-88.
- [15] 尚俊娜, 盛林, 程涛, 等. 基于 LQI 权重和改进粒子群算法的室内定位方法[J]. 传感技术学报, 2017, 30(2): 284-290.
- SHANG J N, SHENG L, CHENG T, et al. The indoor localization based on LQI weight and improved particle swarm optimization algorithm[J]. Chinese Journal of Sensors and Actuators, 2017, 30(2): 284-290.

[作者简介]



孟凯露 (1993-), 女, 杭州电子科技大学通信工程学院硕士生, 主要研究方向为智能算法。



岳克强 (1984-), 男, 博士, 杭州电子科技大学电子信息学院讲师, 主要研究方向为进化计算、通信信号处理。



尚俊娜 (1979-), 女, 博士, 杭州电子科技大学通信工程学院副教授, 主要研究方向为通信信号处理、智能算法。



研究与开发

水声通信系统中基于迭代自适应的脉冲噪声抑制方法

周桂莉, 李有明, 余明宸, 王晓丽

(宁波大学信息科学与工程学院, 浙江 宁波 315211)

摘要: 针对水声通信系统中脉冲噪声抑制问题, 提出了一种迭代自适应的脉冲噪声抑制方法。基于 OFDM 子载波之间的正交性, 该方法首先利用空子载波矩阵从接收信号中提取出背景噪声和脉冲噪声。然后, 利用空子载波矩阵构造导频矩阵, 得到脉冲噪声的干扰协方差矩阵, 并在加权最小二乘准则下通过对代价函数的求解得到脉冲噪声的闭式解。最后, 在接收信号中减去脉冲噪声的估计值, 完成对脉冲噪声的抑制。仿真结果表明, 本文方法有效降低了水声通信系统的误码率, 且在高信干噪比下性能提升更加明显。

关键词: 水声通信; 脉冲噪声; 迭代自适应方法

中图分类号: TN929.3

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017240

Impulsive noise mitigation based on iteration adaptive approach in underwater acoustic communication

ZHOU Guili, LI Youming, YU Mingchen, WANG Xiaoli

College of Information Science and Engineering, Ningbo University, Ningbo 315211, China

Abstract: Focusing on the impulsive noise mitigation in underwater acoustic communication, an impulsive noise mitigation method based on iteration adaptive approach was proposed. Firstly, by utilizing the orthogonality of the subcarriers in OFDM system, both the impulsive noise and background noise were separated from the signals. Then null tones matrix was used to construct a steering matrix and get the impulsive noise interference covariance matrix. Furthermore, a cost function based on weighted least squares approach was proposed and a close form expression for the impulsive noise was derived. Finally, the impulsive noise was mitigated by subtracting the estimation of the impulse noise in the received signal. Simulation results show that the proposed algorithm can reduce the bit error rate (BER) efficiently and provide better performance, especially in high signal to impulsive noise ratio (SINR) case.

Key words: underwater acoustic communication, impulsive noise, iteration adaptive approach

收稿日期: 2017-06-14; 修回日期: 2017-07-31

基金项目: 国家自然科学基金资助项目 (No.61571250); 宁波市自然科学基金资助项目 (No.2015A610121); 宁波大学研究生科研创新基金资助项目

Foundation Items: The National Natural Science Foundation of China (No.61571250), Ningbo Natural Science Foundation of China (No.2015A610121), The Scientific Research Foundation of Graduate School of Ningbo University

1 引言

在水声通信中, 海洋环境噪声是影响水声通信系统性能的主要因素之一。海洋环境噪声主要由潮浪、地震活动、生物群体、交通运输、钻探等引起^[1]。在远海中, 水声噪声一般认为是高斯分布的, 但是由于在近海区域海洋工程的施工、海洋交通工具的运行及石油、天然气开采等人类活动, 水声信道不仅受到高斯白噪声的影响, 还受到脉冲噪声的干扰^[2]。脉冲噪声虽然持续时间短, 但却具有远远高于背景噪声的能量, 因此会对通信系统产生严重的影响。另外, 水声信道的可用带宽非常有限, 通常在几十至几百 kHz, 受界面反射、声线弯曲及随机散射等影响, 多径效应非常严重。由于 OFDM (orthogonal frequency division multiplexing) 采用多载波调制, 在码元中加入保护间隔来对抗多径效应, 并且能将脉冲噪声的能量分散到各个子载波上, 减小脉冲噪声对系统的影响, 因此在水声通信系统中被广泛使用^[3]。但是当脉冲噪声的能量过大时, 采用传统 OFDM 技术的接收机性能会急剧下降, 因此水声通信系统中脉冲噪声的抑制方法研究引起了学者的高度重视^[4,5]。

传统的脉冲噪声抑制方法主要有消隐法和限幅法, 其基本理论是通过设定门限值对接收端的时域信号进行判决, 如果接收信号的幅度值超过门限值, 则将其判决为受到脉冲噪声干扰的信号, 并将这些被干扰信号的幅度值置零或者设定为某一固定值, 完成对脉冲噪声的抑制。其中, 限幅法在无线通信的脉冲噪声抑制中得到了广泛使用^[6-8], 并扩展到水声通信中^[9]。参考文献[9]采用限幅法对水声通信中的脉冲噪声进行抑制, 实验结果表明该方法能在一定程度上抑制较高幅度的脉冲噪声。此类方法的优点是实现简单, 缺点是门限值的确定通常需要准确的脉冲噪声统计模型。考虑到脉冲噪声具有随机性特征, 仅存在

于少量的采样时间点上, 可以被视为时域稀疏信号。因此可以采用压缩感知理论中的信号重构算法对脉冲噪声信号进行估计, 并在接收信号中减去脉冲噪声信号的估计值, 完成对脉冲噪声的抑制。参考文献[10]提出了正交匹配追踪 (orthogonal matching pursuit, OMP) 算法, 该方法通过导频子载波的观测量对脉冲噪声进行估计和抑制。但是, 采用 OMP 算法对脉冲噪声进行估计需要已知脉冲噪声信号的稀疏度, 在稀疏度未知或者稀疏度估计不准确的情况下, 估计性能会有所下降, 从而影响脉冲噪声的抑制性能。

本文提出具有高分辨率特性的迭代自适应方法估计脉冲噪声。首先, 在基于 OFDM 调制技术的水声通信系统中, 利用构造的空子载波矩阵从接收信号中提取出背景噪声和脉冲噪声。然后, 根据空子载波矩阵构造导频矩阵, 得到脉冲噪声的干扰协方差矩阵, 并在加权最小二乘准则下利用代价函数的求解估计出脉冲噪声。最后, 在接收信号中减去脉冲噪声的估计值, 完成对脉冲噪声的抑制。与限幅法和 OMP 算法相比, 该方法不需要已知脉冲噪声的统计模型和稀疏度, 因此具有更好的稳健性。仿真结果表明, 相比于其他脉冲噪声抑制算法, 本文方法能够有效降低系统误码率。

2 系统模型

针对水声信道的频域选择特性, 采用了基于 OFDM 的水声系统进行数据的传输。假设 OFDM 信号持续时间为 T , 子载波间隔为 $\Delta f = 1/T$, 子载波总数 N , 空子载波数为 U , 数据子载波集合为 C_D , 空子载波集合为 C_N 。在基带上传输的信号可以表示为:

$$x(t) = \sum_{k \in C_D} s(k) e^{j2\pi\Delta f t} \quad (1)$$

其中, $s(k)$ 为加载在第 k 个子载波上的数据。

第 k 个子载波上的水声信道的频率响应可以表示为:



$$H(k) = \sum_{i=0}^{P-1} a_i e^{-j2\pi k \Delta f \tau_i} \quad (2)$$

其中, P 是信道多径数, a_i 和 τ_i 分别表示第 i 条路径信道响应的幅度和时延。

发送的数据信号经过水声信道和接收机的预处理^[10]之后到达接收端, 得到的信号为:

$$y(t) = e^{j2\pi \epsilon t} \sum_{k \in C_D} H(k) s(k) e^{j2\pi k \Delta f t} + e(t) + v(t) \quad (3)$$

其中, ϵ 为冗余多普勒频移, $e(t)$ 表示脉冲噪声, $v(t)$ 是均值为 0、方差为 σ^2 的高斯白噪声。将式 (3) 离散化得到:

$$y(n) = e^{j2\pi \epsilon n} \sum_{k \in C_D} H(k) s(k) e^{j2\pi k n / N} + e(n) + v(n) \quad (4)$$

因此, 信道输入输出关系用矩阵形式可以表示为:

$$\begin{aligned} y &= \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(N-1) \end{bmatrix} \\ &= DF^H A_H \begin{bmatrix} s(0) \\ s(1) \\ \vdots \\ s(N-1) \end{bmatrix} + \begin{bmatrix} e(0) \\ e(1) \\ \vdots \\ e(N-1) \end{bmatrix} + \begin{bmatrix} v(0) \\ v(1) \\ \vdots \\ v(N-1) \end{bmatrix} \\ &= DF^H A_H s + e + v \end{aligned} \quad (5)$$

其中, D 为冗余多普勒频移矩阵, $A_H = \text{diag}(H(0), H(1), \dots, H(N-1))$, $D = \text{diag}(1, e^{j2\pi \epsilon}, \dots, e^{j2\pi \epsilon(N-1)})$, F 为维数为 $N \times N$ 的离散傅里叶变换 (discrete Fourier transform, DFT) 范德蒙德矩阵, F^H 为 F 的厄米特变换。

3 基于迭代自适应的脉冲噪声估计和抑制

首先, 利用空子载波矩阵对噪声进行提取; 然后, 采用迭代自适应方法完成脉冲噪声的估计; 最后, 在接收端减去时域噪声的估计值, 得到最终的接收信号。

在 OFDM 水声通信系统的接收端接收到信号, 并构造一个维数为 $U \times N$ 的空子载波矩阵,

记为 Φ , Φ 由 F 中的第 $(N-U)$ 行至第 N 行构成, 然后在式 (5) 两边同乘空载波矩阵 Φ , 由于 OFDM 符号中的各个子载波之间的正交性, 提取出噪声:

$$r = \Phi y = \Phi e + n \quad (6)$$

其中, $n = \Phi v$ 表示频域的背景噪声, 且依旧是高斯白噪声。

然后利用迭代自适应方法进行脉冲噪声的估计, 根据参考文献[11], $U \times N$ 的空子载波矩阵 Φ 可以作为导频矩阵:

$$\Phi = [a(\omega_1), a(\omega_2), \dots, a(\omega_N)] \quad (7)$$

$$\text{其中, } a(\omega_s) = \frac{1}{\sqrt{N}} [1, e^{-j\omega_s}, \dots, e^{-(U-1)j\omega_s}]^T,$$

$s = 1, 2, \dots, N$ 。

令脉冲噪声向量为 $e = [\beta(1), \beta(2), \dots, \beta(N)]^H$, 则在采样点 s 处的脉冲噪声功率可以表示为 $p_s = |\beta(s)|^2$, $s = 1, 2, \dots, N$ 。由于背景噪声和脉冲噪声是相互独立的, 定义脉冲噪声的干扰协方差为:

$$C(\omega_s) = R - p_s a(\omega_s) a(\omega_s)^H \quad (8)$$

其中, R 为脉冲噪声的协方差:

$$R = \sum_{s=1}^N p_s a(\omega_s) a(\omega_s)^H$$

在加权最小二乘准则下, 通过求解如下代价函数^[12]可以得到脉冲噪声的估计值, 其代价函数可以表示为:

$$\min_{\beta(s)} [r - \beta(s) a(\omega_s)]^H C^{-1}(\omega_s) [r - \beta(s) a(\omega_s)] \quad (9)$$

求解得到对应于采样点 s 处的脉冲估计值为:

$$\hat{\beta}(s) = \frac{a(\omega_s)^H C^{-1}(\omega_s) r}{a(\omega_s)^H C^{-1}(\omega_s) a(\omega_s)} \quad (10)$$

为了避免在每个频率上计算 C^{-1} , 根据参考文献[13], 可以表示为:

$$\begin{aligned}
C^{-1} &= \left(R - a(\omega_s) p_s a(\omega_s)^H \right)^{-1} \\
&= R^{-1} + R^{-1} a(\omega_s) \left(1 - p_s a(\omega_s)^H R^{-1} a(\omega_s) \right)^{-1} \\
&\quad p_s a(\omega_s)^H R^{-1} \\
&= \left[I + \frac{p_s R^{-1} a(\omega_s) a(\omega_s)^H}{1 - p_s a(\omega_s)^H R^{-1} a(\omega_s)} \right] R^{-1}
\end{aligned} \quad (11)$$

由于 R 为非奇异矩阵, C^{-1} 又可以表示为:

$$C^{-1} = \left(I - p_s R^{-1} a(\omega_s) a(\omega_s)^H \right)^{-1} R^{-1} \quad (12)$$

联立式 (11) 和式 (12), 可以证得:

$$C^{-1} = \frac{R^{-1}}{I - p_s a(\omega_s)^H R^{-1} a(\omega_s)} \quad (13)$$

将式 (13) 代入式 (10), 得:

$$\hat{\beta}(\omega_s) = \frac{a(\omega_s)^H R^{-1} r}{a(\omega_s)^H R^{-1} a(\omega_s)} \quad (14)$$

因此脉冲噪声在 N 个采样点上的估计值为 $e = [\hat{\beta}(1), \hat{\beta}(2), \dots, \hat{\beta}(N)]^T$, 如图 1 所示为 128 个采样点时的脉冲噪声估计值。从图 1 中可以看出, 本文的方法能准确地估计出脉冲噪声。

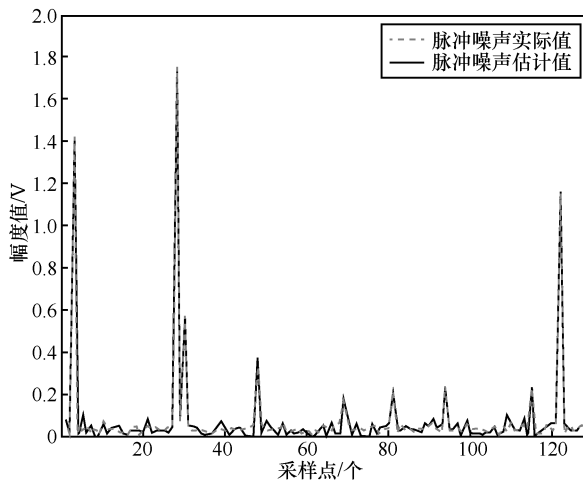


图 1 128 个采样点时的脉冲噪声估计值

4 仿真分析

为了验证算法的性能, 本文利用 MATLAB 进

行仿真。在水声通信中, 噪声 $u(n) = e(n) + v(n)$ 通常使用混合高斯模型来产生^[14], 该模型服从混合高斯分布, 由多个不同高斯分量之间的叠加, 能准确地模拟水声信道上的脉冲噪声, 其概率密度函数为:

$$f_{\text{GM}} = \sum_{i=0}^M p_i \mathcal{N}(0, \gamma_i^2) \quad (15)$$

其中, p_i 表示第 i 个高斯分量发生的概率,

$\sum_{i=0}^M p_i = 1$, γ^2 为高斯分量的方差。仿真实验中

OFDM 信号的子载波数为 128, 多径数为 5, 且每条路径的时延服从指数分布, 路径增益服从瑞利分布, 冗余多普勒频移随机产生, 取值范围为 $\left[-\frac{\Delta f}{2}, \frac{\Delta f}{2}\right]$ 。脉冲噪声模型的高斯分量个数为 3,

$p = [0.7, 0.21, 0.09]$, $\gamma^2 = [1, 10, 100]$ 。重点比较了未抑制脉冲噪声、限幅法^[9]、正交匹配追踪算法^[10]以及本文算法。为了获得比较稳定可靠的仿真结果, 该仿真结果为 5 000 次蒙特卡罗仿真取平均得到。

图 2 所示为在数据子载波数为 72 个、空子载波数为 56 个时, 分别采用 4QAM 和 16QAM 调制下, 4 种算法的误码率随 SINR 的变化情况。从图 2 中可以看出, 随着 SINR 的增加, 4 种算法的误码率均降低。在图 2 (a) 的 4QAM 调制系统中, 当 $\text{SINR} < -16$ dB 时, 限幅法的性能最好, 这是由于在低 SINR 下脉冲噪声具有较大的幅度值, 因此采用限幅法能够更好地抑制脉冲噪声分量, 保留信号分量。而本文所提算法和正交匹配追踪算法需要首先估计脉冲噪声, 而在低 SINR 条件下脉冲噪声估计的准确度下降, 从而导致误码率性能较差。而当 $\text{SINR} > -16$ dB 时, 本文所提算法能够准确估计脉冲噪声, 具有较好的误码率性能。在图 2 (b) 的 16QAM 调制系统中, 本文所提算法在 $\text{SINR} > -16$ dB 时具有最佳误码率性能。与图 2 (a) 相比, 图 2 (b) 的误码率性能有所降低, 这是因为在相同的发射能量下, 4QAM 和 16QAM 调制



的平均信号能量 E 相同。而 MQAM 调制下星座图中临近点的最小欧氏距离 $d_{\min} = \sqrt{\frac{6E_{\text{avg}}}{M-1}}$, 其中 M 为调制阶数。因此 16QAM 调制具有较小的 $d_{\min}$ 。由于 $d_{\min}$ 越小, 星座图中两点越接近, 越容易发生误判, 因此会造成误码率上升。

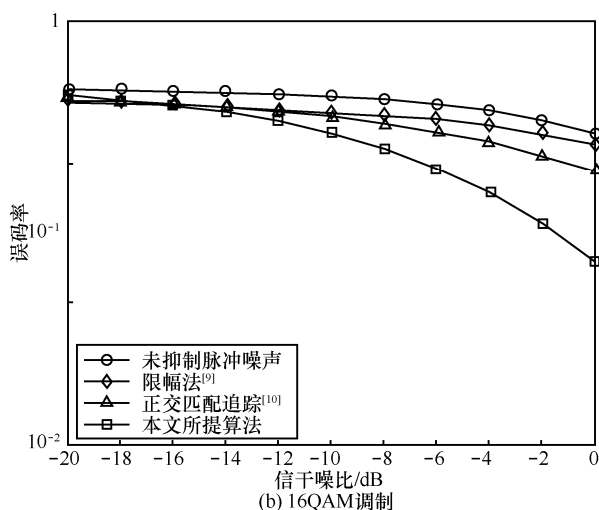
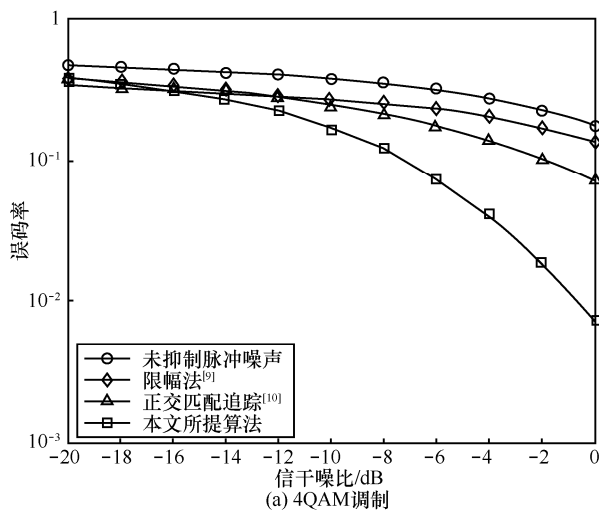


图2 不同调制模式下误码率随 SINR 的变化曲线

图3为在图2基础上进一步采用卷积码编码后4种算法误码率随 SINR 的变化情况。与图2相比, 由于采用了卷积码编码, 4种算法的误码率性能均得到了较大程度的提升。对于 4QAM 调制系统中, 当 $\text{SINR} < -14$ dB 时, 4种算法的误码率性能较为接近, 而当 $\text{SINR} > -14$ dB 时, 本文所提算法具有最佳的误码率性能, 且误码率随 SINR

的增加快速减小。从图3(b)可以看出, 受 16QAM 调制的影响, 卷积码编码对于误码率性能提升的效果不够明显。当 $\text{SINR} > -8$ dB 时, 本文所提算法的误码率性能开始优于其他算法, 并且误码率随 SINR 的增加快速减小。该组结果表明, 通过有效的信道编码技术, 可以进一步改善水声通信系统的性能。

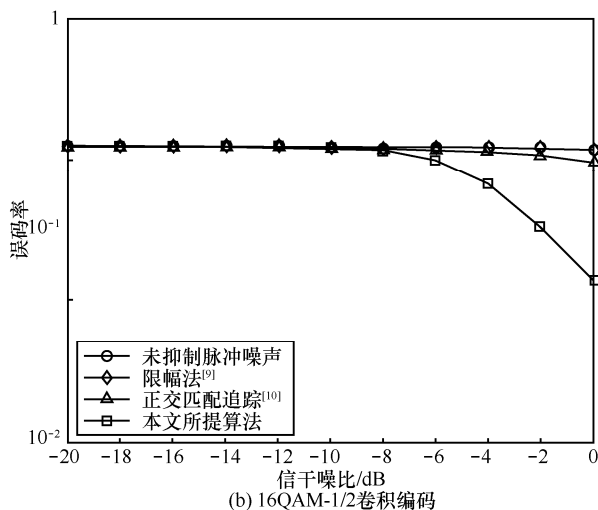
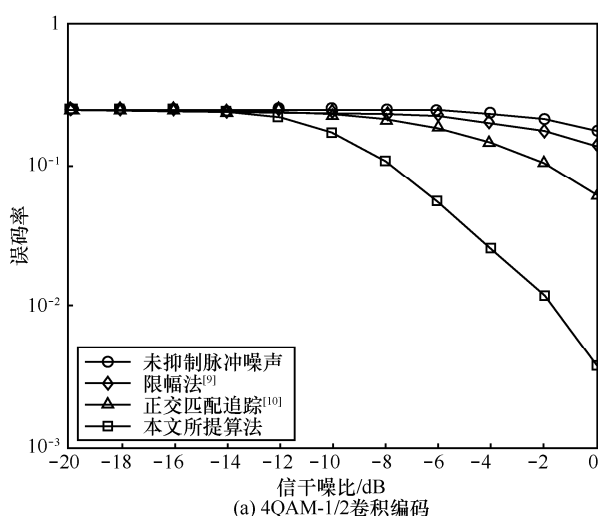


图3 1/2 卷积编码下不同调制模式下误码率随 SINR 的变化曲线

图4为 $\text{SINR} = -12$ dB 时不同调制模式下的非编码系统中误码率随空子载波数的变化曲线。从图4中可以看出, 4种算法的误码率随空子载波数的增加而下降, 这是由于空子载波个数的增加使系统获得了更多的观测值, 可以更好地估计脉冲噪声。本文算法在两种调制模式均具有较小的误码率, 性能效果明显优于其他算法。

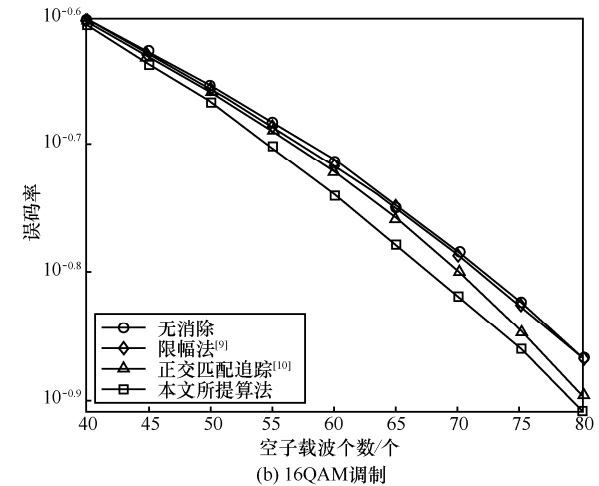
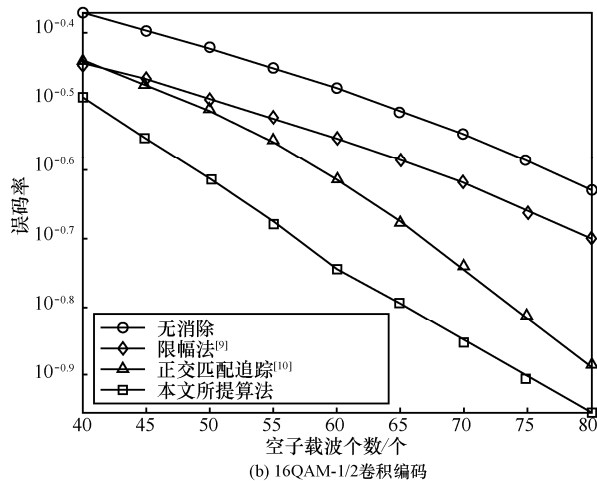
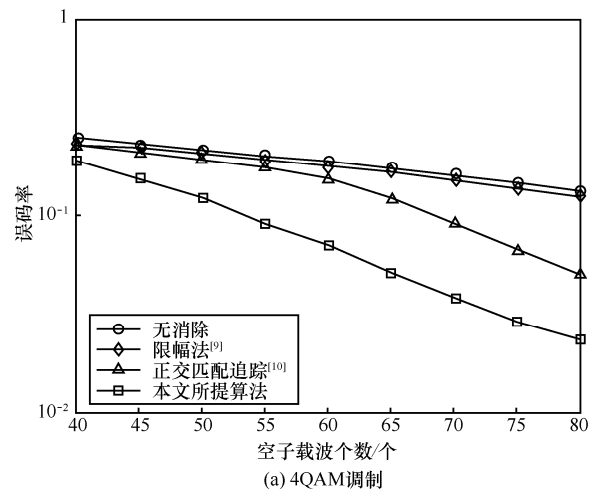
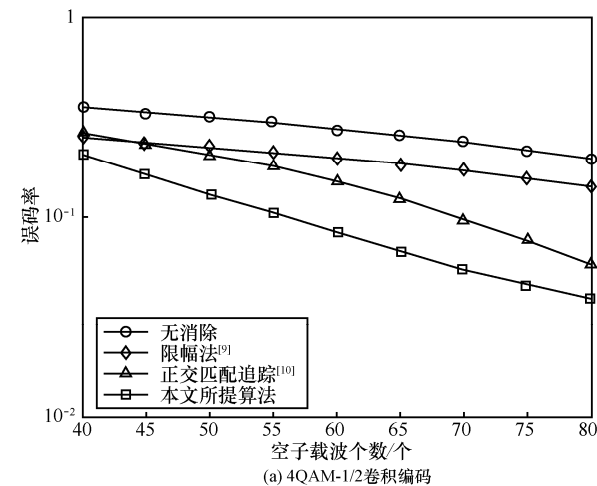


图4 $SINR=-12$ dB时不同调制模式下的非编码系统中误码率随空子载波数的变化曲线

图5 1/2卷积编码下不同调制模式下误码率随空子载波数的变化曲线

图5为在图4基础上采用卷积编码后误码率随空子载波数的变化曲线。从图5中可以看出,4种算法的误码率随空子载波数的增加而下降。当调制模式为4QAM时,本文算法在不同空子载波数下均具有较小的误码率,且空子载波数越大,性能提升更加明显。对于16QAM调制系统,当空子载波数较小时,本文算法利用导频矩阵提取的信息有限,不能准确提取接收信号信息,并且受16QAM的影响,误码率性能提升不明显。随着空子载波数的增加,导频矩阵可以更加准确地提取接收信号的信息,从而使得误码率性能提升。

复杂度分析^[15]: 本文所提算法的运算过程主要包括矩阵相乘和转置,其复杂度为 $O(U^3 + NU^2)$;

而限幅法是直接对全部采样点依次进行限幅处理,其计算复杂度为 $O(N)$; 另外,正交匹配追踪算法的计算主要包括复杂度为 $O(NU)$ 的传感矩阵与残差矩阵的乘法和复杂度为 $O(N^2U)$ 的矩阵乘法。其中, N 是子载波总数, U 是空子载波数,且 $U < N$ 。

因此,限幅法具有最低的计算复杂度,但是其误码率性能最差。而本文算法和正交匹配追踪算法的计算复杂度则主要由空子载波数决定,虽然当空子载波数较大时,本文算法的计算复杂度上升,但是其系统的性能效果优于正交匹配追踪算法;而且当空子载波数较少时,本文算法不仅计算复杂度低于正交匹配追踪算法,在系统的性能提升上也优于正交匹配追踪算法。



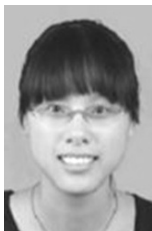
5 结束语

本文提出了一种基于迭代自适应的脉冲噪声抑制方法。该方法首先利用 OFDM 子载波之间的正交性, 从接收信号中提取出背景噪声和脉冲噪声, 然后采用具有高分辨率特性的 IAA 估计出脉冲噪声, 最后在接收端进行脉冲噪声的抑制。仿真结果表明, 本文算法有效降低了系统误码率, 在非编码和编码系统下均具有较好的性能。

参考文献:

- [1] 张歆, 张小蓟. 水声通信理论与应用[M]. 陕西: 西北工业大学出版社, 2012.
ZHANG X, ZHANG X J. Underwater acoustic communication theory and application[M]. Shanxi: Northwestern Polytechnical University Press, 2012.
- [2] 张安清. 浅海水声信道的脉冲噪声特性分析[J]. 声学技术, 2007, 26(5): 988-989.
ZHANG A Q. Analysis of impulsive noise characteristic for shallow sea channel[J]. Acoustic Techniques, 2007, 26(5): 988-989.
- [3] CHEN P, RONG Y, NORDHOLM S. Pilot-subcarrier based impulsive noise mitigation for underwater acoustic OFDM systems[C]//ACM International Conference, October 24-26, 2016, Shanghai, China. New York: ACM Press, 2016: 1-5.
- [4] 郑君杰, 阮赬, 李延宾, 等. 水声通信网络问题研究[J]. 电信科学, 2009, 25(10): 63-66.
ZHENG J J, RUAN K, LI Y B, et al. Research of underwater acoustic communication network[J]. Telecommunications Science, 2009, 25(10): 63-66.
- [5] XU X K, SUN H X, ZHOU S L, et al. Impulsive noise mitigation in underwater acoustic OFDM systems[J]. IEEE Transactions on Vehicular Technology, 2016, 65(10): 8190-8202.
- [6] 张高境, 熊兴中. 电力载波通信中的干扰特性及抑制技术分析[J]. 电信科学, 2016, 32(2): 182-188.
ZHANG G J, XIONG X Z. Interference characteristics and suppression techniques in power line communication[J]. Telecommunications Science, 2016, 32(2): 182-188.
- [7] ABDELKEFI F, DUHAMEL P, ALBERGE F. Impulsive noise cancellation in multicarrier transmission[J]. IEEE Transactions on Communications, 2005, 53(1): 94-106.
- [8] SAEEDI H, SHARIF M, MARVASTI F. Clipping noise cancellation in OFDM systems using oversampled signal reconstruction[J]. IEEE Communications Letters, 2002, 6(2): 73-75.
- [9] GUO Y, SUN Y, CHENG E, et al. Impulsive noise suppression and signal distortion effect mitigation in underwater acoustic OFDM system[J]. Journal of Convergence Information Technology, 2012(7): 181-189.
- [10] SUN H X, XU X K, MA L. Carrier frequency offset and impulsive noise estimation for underwater acoustic orthogonal frequency division multiplexing[J]. Chinese Journal of Acoustics, 2014, 33(3): 289-298.
- [11] YARDIBI T, LI J, STOICA P, et al. Source localization and sensing: a nonparametric iterative adaptive approach based on weighted least squares[J]. IEEE Transactions on Aerospace & Electronic Systems, 2010, 46(1): 425-443.
- [12] STOICA P, LI J, LING J. Missing data recovery via a nonparametric iterative adaptive approach[J]. IEEE Signal Processing Letters, 2009, 16(4): 3369-3372.
- [13] TYLAVSKY D J, SOHIE G R L. Generalization of the matrix inversion lemma[J]. Proceedings of the IEEE, 2005, 74(7): 1050-1052.
- [14] BANERJEE S, AGRAWAL M. On the performance of underwater communication system in noise with Gaussian Mixture statistics[C]// Twentieth National Conference on Communications, February 28-March 2, 2014, Kanpur, India. New Jersey: IEEE Press, 2014: 1-6.
- [15] LI C G, SONG K, YANG L X. Low computational complexity design over sparse channel estimator in underwater acoustic OFDM communication system[J]. IET Communications, 2017, 11(7): 1143-1151.

[作者简介]



周桂莉 (1992-), 女, 宁波大学信息科学与工程学院硕士生, 主要研究方向为水声通信系统中资源分配和干扰抑制。

李有明 (1963-), 男, 宁波大学信息科学与工程学院教授、博士生导师, 主要研究方向为宽带通信、电力线通信、协作中继、认知无线电等。

余明宸 (1991-), 男, 宁波大学信息科学与工程学院硕士生, 主要研究方向为电力线通信系统脉冲干扰。

王晓丽 (1975-), 女, 宁波大学信息科学与工程学院讲师、在职博士生, 主要研究方向为多载波通信及应用。



研究与开发

基于主题模型的垃圾邮件过滤系统的设计与实现

寇晓淮, 程华

(华东理工大学信息科学与工程学院, 上海 200237)

摘要: 垃圾邮件过滤技术在保证信息安全、提高资源利用、分拣信息数据等方面都发挥着重要作用。然而, 垃圾邮件的出现影响了用户的体验, 并且会造成不必要的经济与时间损失。针对现有的垃圾邮件过滤技术的不足, 基于多个主题词理论, 构建了基于朴素贝叶斯的垃圾邮件分类方法。在邮件主题获取中, 采用主题模型 LDA 得到邮件的相关主题及主题词; 并进一步采用 Word2Vec 寻找主题词的同义词和关联词, 扩展主题词集合。在邮件分类中, 对训练数据集进行统计学习得到词语的先验概率; 基于扩展的主题词集合及其概率, 通过贝叶斯公式推导得到某个主题和某封邮件的联合概率, 以此作为垃圾邮件判定的依据。同时, 基于主题模型的垃圾邮件过滤系统具有简洁易应用的特点。通过与其他典型垃圾邮件过滤方法的对比实验, 证明基于主题模型的垃圾邮件分类方法及基于 Word2Vec 的改进方法均能有效提高垃圾邮件过滤的准确度。

关键词: 文本分类; 垃圾邮件; 主题模型; 贝叶斯原理

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017313

Design and implementation of spam filtering system based on topic model

KOU Xiaohuai, CHENG Hua

College of Information Science and Engineering, East China University of
Science and Technology, Shanghai 200237, China

Abstract: Spam filtering technology plays a key role in many areas including information security, transmission efficiency, and automatic information classification. However, the emergence of spam affects the user's sense of experience, and can cause unnecessary economic and time loss. The deficiency of spam filtering technology was researched, and a method of spam classification based on naive Bayesian was put forward based on multiple keywords. In the subject of mail, the theme model was used by LDA to get the related subject and keyword of the message, and Word2Vec was further used to search keyword synonyms and related words, extending the keyword collection. In the classification of mails, the transcendental probability of the words in the training dataset was obtained by statistical learning. Based on the extended keyword collection and its probability, the joint probability of a subject and a message was deduced by the Bayesian formula as a basis for the spam judgment. At the same time, the spam filtering system based on topic model was simple and easy to apply. By comparing experiments with other typical spam filtering method, it is proved that the method of spam classification based on theme model and the improved method based on Word2Vec can effectively improve the accuracy of spam filtering.

Key words: text classification, spam, topic model, Bayesian theory

收稿日期: 2017-05-12; 修回日期: 2017-09-16



1 引言

伴随着互联网的发展和普及,电子邮件已经成为人们日常工作、生活中通信、交流的重要手段。但由于早期的 SMTP 缺乏发件人认证、大量开放式邮件中转服务器以及互联网分布式管理性质等原因,垃圾邮件已经成为亟待解决的问题。从电子邮件出现以来,研究者就在垃圾邮件拦截方面做出了大量的研究工作。然而,垃圾邮件制造者总会找到更加隐蔽且混淆的手段来躲避相关算法的检测。对于此类研究工作,目前仍然存在两个重要的问题:邮件是一种快速且便捷的通信方式,而大面积的广告推广动机促成了大量为非正当利益而开发的反过滤技术;中文词语的丰富性和特殊性导致垃圾邮件与正常邮件区分难度较大,很多国外的优秀算法在移植过程中将遭遇新的挑战。

针对以上问题,本文深入分析和比较传统垃圾邮件处理方法,指出了现有垃圾邮件过滤方法的不足,对主题模型算法及其在自然语言处理中的应用进行了研究,指出了主题模型算法应用于垃圾邮件过滤的可行性与能够解决的问题;提出了基于主题模型的垃圾邮件过滤算法;设计并实现了一种基于主题模型的垃圾邮件过滤模型,通过与其他方法的对比实验,证明本文基于主题模型的垃圾邮件过滤方法及基于 Word2Vec^[1]的改进方法均明显提升了过滤准确度,具有较高的应用价值。

2 基于邮件过滤的相关技术

2.1 面向内容的电子邮件过滤技术

常见的邮箱对于垃圾邮件的过滤策略中,基于内容对邮件过滤的方法有黑白名单、手工建立过滤规则等。手工建立规则的方法通过用户建立一系列规则来判定垃圾邮件。显然,这些方法的主观性会造成大量合法邮件的误判和垃圾邮件的

漏判,并且很难做到实时的手工维护,对邮件服务商的人力及经济造成很大压力。因此,垃圾邮件工具逐渐倾向于引入基于内容的机器学习判别方法^[2,3]。

基于内容垃圾邮件判别的机器学习方法,一般步骤如下。

步骤 1 获取训练数据集合,通过多种手段渠道获取各类电子邮件,并备注该电子邮件是否是垃圾邮件。

步骤 2 建立模型,使用训练集合训练模型,更新模型中的参数。

步骤 3 使用训练好的模型,对新的电子邮件进行过滤。

总结起来就是通过已有的训练集合(正例、反例)训练出相应的垃圾邮件规则(包括显式规则或隐式规则),然后将规则应用到新的邮件判别中。

最近几年,国内外研究者在此领域已经取得了大量的研究成果。Sheu 等人^[4]利用决策树模型构建了三步法垃圾邮件过滤模式。Feng 等人^[5]提出了基于朴素贝叶斯分类器的训练集分类方法,提升了数据处理的顽健性,提出的 SVM-NB 方法能够达到较高的垃圾邮件检测精度。而 Bansal 等人^[6]构建了基于穿梭判定算法的垃圾词语检测方法,并且在谷歌邮件系统中做了初步的应用。另外,广告产业的发展为垃圾邮件拦截与过滤提出了新的要求,Chan 等人^[7]在此方向上做了针对性研究,推出了广告环境下的垃圾邮件过滤方法。除此之外,一些其他的研究成果也引起了学术界和 IT 产业界的广泛关注^[8-10]。曹玉东等人^[11]基于改进的局部敏感散列算法实现了图像型垃圾邮件过滤,将垃圾邮件过滤方法的应用范围扩大。

2.2 垃圾邮件常用文本分类方法

(1) Decision Tree 方法

决策树利用熵的概念对每次决策产生的结

果进行分类^[4]。决策树使用树状结构对目标分类, 树中每个节点表示某个对象, 每个分叉路径代表某个可能的属性值, 而每个叶节点则对应从根节点到该叶节点所经历的路径所表示的对象的值。

决策树也可以被称为分类树, 它是非常常用的分类方法。从另一个角度来说, 决策树是一种监督学习方法。在给定样本机器类别属性后, 决策树通过学习能够得到一个固定的分类器, 从而给出新进数据的具体类别。

(2) AdaBoost 方法

自适应增强 (adaptive boosting, AdaBoost) 是加权组合多个弱分类器分类结果, 进而得到更好的分类器的方法。Carreras 和 Nicholas^[12,13]将 AdaBoost 引入垃圾邮件过滤, 获得了很高的性能。AdaBoost 方法的自适应在于: 后面的分类器会在那些被之前分类器分错的样本上训练。AdaBoost 方法对于噪声数据和异常数据很敏感。但在一些问题中, 相比于大多数学习算法, AdaBoost 方法对于过拟合问题不够敏感。AdaBoost 方法中使用的分类器可能很弱 (比如出现很大错误率), 但其分类效果只要比随机好一点 (比如它的二分类错误率略小于 0.5), 就能够改善最终模型。

(3) Rough Sets 方法

Rough Sets 算法是一种比较新颖的算法, 粗糙集理论对于数据的挖掘提供了一个新的概念和研究方法。将 Rough Sets 引入垃圾邮件过滤, 采用 11 种非文本属性 (包括收信人数、中继个数等) 来进行邮件分类 (正常、广告和反动)。

具体来说, 所有属性分为 2 种属性: 1 类为条件属性, 1 类为决策属性。本文姑且把决策属性设置在数据列的最后一列, 算法的步骤依次判断条件属性是否能被约简, 如果能被约简, 此输出约简属性后的规则, 规则的形式大体类似于 IF-THEN 的规则。

(4) kNN 方法

k-近邻方法 (*k*-nearest neighbour, *k*NN) 在线性模型中是最常见的方法, 通过选择特征与数据集中所有特征对比最近的几个样本的标签平均值表示。

2.3 用于垃圾邮件过滤的贝叶斯方法

对于邮件的垃圾分类, 一方面邮件就是文本, 属于文本分类领域。另一方面, 由邮件中的某些关键词来推断是否是垃圾邮件, 就是一种贝叶斯条件概率方法的应用。数据挖掘领域主要使用两种贝叶斯方法, 即朴素贝叶斯方法和贝叶斯网络方法。贝叶斯方法的一个显著特点, 就是在知道结果的情况下了解假设的情况, 也就是说, 当对某些知识知之甚少, 或者毫不知情的时候, 贝叶斯方法具有独特优势。

在垃圾邮件检测过程中, 其主要依据正常邮件与垃圾邮件的先验概率。而贝叶斯分类模型能够通过适当的独立性假设来简化分布, 也就是朴素贝叶斯假设。在这样的假设条件下, 能够形成朴素贝叶斯网络。

贝叶斯分类算法是基于概率统计原理的一种分类方法, 它具有运算速度快、方法简单、分类精度高等优点, 因而被广泛应用在文本分类领域, 并表现出非常好的效果。

目前, 贝叶斯过滤算法被广泛使用于智能和概率系统中, 它具有单词学习的模式和频率, 而不需要提前预设任何规则。具体来说, 贝叶斯过滤技术能够根据垃圾邮件与正常邮件的联系与特点进行判断。与传统的关键词检测过滤技术相比, 贝叶斯过滤算法更加复杂且智能, 而反过滤方法不能破解过滤器内部的配置, 从而提升了安全性与顽健性。

2.4 基于朴素贝叶斯的文本分类及流程

朴素贝叶斯分类器是垃圾邮件内容过滤中智能应用的分类方法。利用这种方法, 可以根据训练集自动训练, 训练的结果反映了训练集的性质。



因此训练者可以利用一定数量的垃圾邮件和非垃圾邮件,训练邮件过滤器,从而达到高效、准确过滤垃圾邮件的目的。

朴素贝叶斯分类的流程如图 1 表示。

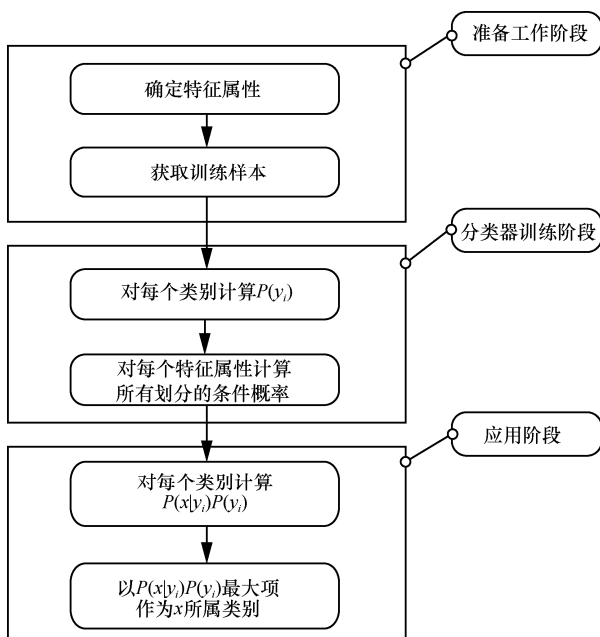


图 1 朴素贝叶斯分类流程

然而,朴素贝叶斯分类也有缺陷,它的假设是基于“各特征项相互条件独立”。在很多的实际问题中,如果此下设表现不够明显,甚至出现不成立时,错误的分类将会出现,从而影响算法的最终表现。在本文中,贝叶斯模型的使用将会被改善,而具体的内容将会在第 3 节中被介绍。

3 主题模型在垃圾邮件过滤中的研究

3.1 基于关键词的垃圾邮件过滤

3.1.1 算法思想

主要算法思想是基于关键词技术,采用朴素贝叶斯分类方法得到关键词,分析邮件内容分类到垃圾邮件的置信概率,进而产生分类结果。这种方法的优势在于复杂度低,且应用范围较广。

3.1.2 基于关键词的邮件过滤算法流程

从内容上看,邮件过滤可以看成是一个二值分

类问题,即把邮件分为垃圾邮件类和合法邮件类。基于关键词的邮件过滤算法流程简单来讲是朴素贝叶斯方法,贝叶斯过滤算法大致由以下基本步骤组成。

步骤 1 收集大量的垃圾邮件和合法邮件,建立垃圾邮件集和合法邮件集。

步骤 2 提取邮件主题和邮件体中的独立字符串,例如 sale、cash 等作为 token 串并统计提取出的 token 串出现的次数即字频。按照上述的方法分别处理垃圾邮件集和合法邮件集中的所有邮件。采用贝叶斯文本分类法对训练样本学习,得到 $P(S|W)$ 。

步骤 3 每一个邮件集对应一个散列表,合法邮件集对应表 hashtable_good,垃圾邮件集对应表 hashtable_bad,表中存储 token 串到字频的映射关系。

步骤 4 计算每个散列表中 token 串出现的概率,可以得到 $P_1(t_i)$ 和 $P_2(t_i)$, $P_1(t_i)$ 表示 t_i 在 hashtable_good 中的值(也就是 token 串 t_i 在合法邮件中的概率); $P_2(t_i)$ 表示 t_i 在 hashtable_bad 中的值(也就是 token 串 t_i 在垃圾邮件中的概率):

$$P = \frac{\text{某token串的字频}}{\text{对应散列表的长度}} \quad (1)$$

步骤 5 由步骤 2 中贝叶斯文本分类法得到的 $P(S|W)$,综合考虑散列表 hashtable_good 和 hashtable_bad,推断出当新来的邮件中出现某个 token 串时,该新邮件为垃圾邮件的概率。计算式为:

$$P(A|t_i) = \frac{P_2(t_i)}{P_1(t_i) + P_2(t_i)} \quad (2)$$

其中, A 事件表示邮件为垃圾邮件; $t_1, t_2, \dots, t_n$ 代表 token 串; $P(A|t_i)$ 表示当 token 串 t_i 出现在所收到的邮件中时,该邮件为垃圾邮件的概率。

假设该邮件共得到 N 个 token 串 $t_1, t_2, \dots, t_n$, hashtable_probability 中对应的值为 $P_1, P_2, \dots, P_n$, $P(A|t_1, t_2, \dots, t_n)$ 表示在邮件中同时出现多个 token 串 $t_1, t_2, \dots, t_n$ 时,该邮件为垃圾邮件的概率。

由联合概率公式可得:

$$P(A|t_1, t_2, \dots, t_n) = \frac{P_1 \times P_2 \times \dots \times P_n}{P_1 \times P_2 \times \dots \times P_n + (1 - P_1) \times (1 - P_2) \times \dots \times (1 - P_n)} \quad (3)$$

当 $P(A|t_1, t_2, \dots, t_n)$ 超过预定阈值 (例如 0.95) 时, 就可以判断邮件为垃圾邮件。

3.2 LDA 主题模型

LDA (latent Dirichlet allocation) 的产生和发展历经 TF-IDF、LSA、pLSA 等多种主题模型方法, 由于 LDA 模型的良好数学基础和灵活的扩展性, 一经提出即得到了来自各个领域研究者的关注, 被广泛应用在文本挖掘及信息处理的研究中^[14]。

LDA 模型最初是作为一种文本分类和主题聚类方法被提出, 它将文档集中每篇文档的主题以概率分布形式给出, 从而通过分析便能够得到聚类结果。与此同时, 它是一种典型的词袋模型。也就是说, 每篇文档将会被分解为一组词, 而不用考虑先后顺序。

LDA 是一个三层的贝叶斯概率生成模型, 由“主题—词语”和“文档—主题”构成。在 LDA 模型中需要求解“词语—主题”和“主题—文档”两个模型参数。LDA 假设文本集 D 中各文本 w 有如下生成过程, 如图 2 所示, T 表示主题的个数, D 表示文档的个数, N_d 表示第 d 篇文档中词语的个数。

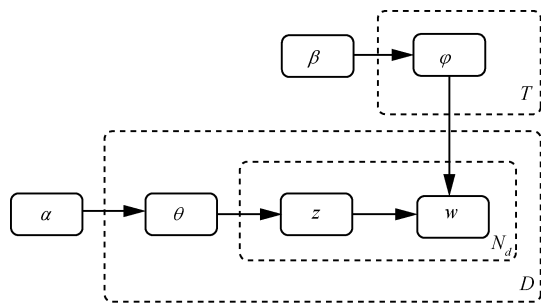


图 2 LDA 模型

步骤 1 确定文档中的词语数 N , 使之服从参数为 ξ 的泊松分布。

步骤 2 确定 θ , 使之服从参数为 α 的狄利克雷分布。

步骤 3 对于文本中 N 个词中的每一个 w_n :

确定一个主题 z_n , 使之服从参数为 θ 的多项式分布; 依照概率 $p(w_n | z_n, \beta)$ 选择每一个词语 w_n 。

3.3 基于主题模型的垃圾邮件过滤方法

基于主题模型抽取垃圾邮件的主题, 对已知的垃圾邮件样本进行训练, 提取垃圾邮件的特征, 采用贝叶斯估计分类算法, 构造垃圾邮件的过滤器。利用得到的垃圾邮件过滤器, 对新的邮件进行分析、判断, 区分垃圾邮件和合法邮件, 实现垃圾邮件的过滤。

具体实现步骤如下。

步骤 1 采集一定数量的垃圾邮件与合法邮件, 建立相应的垃圾邮件集和合法邮件集, 计算词频得到每个词语出现的情况下该邮件是垃圾邮件的概率 $P(S|W)$ 。

步骤 2 利用 LDA 主题模型对邮件进行主题抽取, 分类算法对已知的垃圾邮件样本进行训练, 对垃圾邮件集和合法邮件集中的邮件进行解析, 并提取邮件的特征, 统计相应数据。LDA 是一种文档主体生成模型, 也成为一个三层贝叶斯概率模型, 包含词、主体、文档这三层结构。生成模型, 即一篇文章的每个词都是通过以一定的概率选择了一个主题, 并从这个主题中以一定的概率选择这个词语的过程得到的。

步骤 3 由联合概率公式计算每个主题中所有词语的联合概率; 得到每个主题出现的情况下该邮件是垃圾邮件的概率; 构造邮件分类器。

步骤 4 采用贝叶斯分类器。选取一个判断垃圾邮件适当的阈值, 利用所建立的邮件分类器实现对邮件的分类。

3.4 模型的改进

传统判断两个文档相似性的办法是查看两个文档共同出现的单词的多少, 如 TF-IDF 等, 但这种办法没有考虑到文字背后的语义关联, 有可能两个文档说的是相似的内容但并没有词语上的交集。LDA 提取出来的邮件主题关键词能够表达邮



件较高级别的主题内容,能够消除主题关键词之间的歧义。但是此时每个主题关键词并不是使用向量表达,此时本文使用 Word2Vec 方法,将词语转化为向量空间,有利于计算词语之间的相似程度。同时使用主题词向量距离计算方式计算距离主题最近的词语,即用 Word2Vec 生成每个主题中词语的关联词,作为主题词语的扩容,在此基础上再进行垃圾邮件判断。

4 垃圾邮件过滤器设计及实验分析

4.1 邮件样本集的选取

4.1.1 垃圾邮件过滤的语料库

本文采用的垃圾邮件语料库从网上采集,包含正常邮件和垃圾邮件各 8 000 封。图 3 为比较典型的用于广告的垃圾邮件案例。

您好!
尊敬的客户:本公司长期代理进出口报关业务。有些发票可以为广大客户优惠代开(税率1.5%左右)以解广大客户财务票据的不足。具体有(增值税专用发票、国税商品销售专用发票、地税运输专用发票、建筑安装专用发票;广告专用发票;还有其他服务发票)等,希望有意者来电详谈,愿合作愉快,成功!可验证后付款!!
联系人:.....
手 机:.....
电 话:.....
邮 箱:.....
地 址:.....

图3 垃圾邮件案例

用这两类邮件建立垃圾邮件过滤器中词的先验概率。过程如下。

首先,解析所有邮件,提取每一个词。然后,计算每个词语在正常邮件和垃圾邮件中的出现频率。例如,假定“发票”这个词,在 8 000 封垃圾邮件中,有 200 封包含这个词,那么它的出现频率就是 2.5%;而在 8 000 封正常邮件中,只有 2 封包含这个词,那么出现频率就是 0.025%。有可能某个词在已有的某一类邮件语料中未出现,为了避免该词的先验概率出现为 0 的情况,设定该词

的出现频次为 1。假设某个词只出现在垃圾邮件中,正常邮件中没有,就设定它在正常邮件的出现频率是 0.012 5%(1/8 000),反之亦然。随着邮件数量的增加,词的先验概率计算结果会更接近于真实情况。

4.1.2 垃圾邮件评价指标

为了对垃圾邮件过滤系统的效果做分析,需要一个评价体系来进行评估,即一个系统可以判定未知文档是否属于某类。假定有 N 个邮件文档通过分类器分别分类,可以用表 1 来表示人工与系统对邮件的评判情况。 A 为人工与系统都评判为垃圾的邮件数; B 为人工评判为正常,而系统评判为垃圾的邮件数; C 为系统评判为正常,而人工评判为垃圾的邮件数; D 为人工与系统都评判为正常的邮件数。

表1 垃圾邮件测评

	实际为垃圾邮件	实际为正常邮件
判定为垃圾邮件	A	B
判定为正常邮件	C	D

定义如下几个指标来检测算法对垃圾邮件的过滤效果。

(1) 召回率 (recall)

描述收到一封垃圾邮件时,分类器判定为垃圾邮件的概率,召回率越高,表示分类器对邮件分类效果越显著,计算式为:

$$R=A/(A+C) \times 100\% \quad (4)$$

(2) 正确率 (precision)

描述分类器对正常邮件和垃圾邮件都能正确分辨的概率,将垃圾邮件判为垃圾邮件和将非垃圾邮件判为合法邮件的概率,正确率越高表示分类器的效果越理想,计算式为:

$$P=A/(A+B) \times 100\% \quad (5)$$

(3) 误判率 (misjudge)

描述正常邮件的误判率,将非垃圾邮件判为垃圾邮件的概率,这是描述一个分类器是否有效

的关键指标, 如果误判率很高, 则说明分类器没有起到很好的分类效果, 误判率越低表示正常邮件被判为垃圾邮件的概率越小, 计算式为:

$$M=B/(B+D) \times 100\% \quad (6)$$

(4) 精确率 (accuracy)

分类器对正常邮件分类的正确性, 精确率越高表示邮件对正常邮件的判别越正确, 计算式为:

$$A=(A+D)/(A+B+C+D) \times 100\% \quad (7)$$

在对实验结果的评估中将会比较以上数值。

准确率 P 是邮件被正确分类的概率, 召回率是指实验方法将邮件正确分类的概率, $F1$ 值则是指 $\beta=1$ 时的 F 值, 是最常用的 F 值之一, 可以看作模型准确率和召回率的一种加权平均。这 3 个值都是数值越高所代表的分类效果越优秀。

4.2 实验结果与分析

4.2.1 LDA 主题抽取

步骤 1 首先用 jieba 分词算法分词后得到 300 个分词文件, 名称如 1-seg.txt、2-seg.txt 等。例如, “合金” “批发” “朋友” “爸妈” 等词语。

步骤 2 再用 LDA 主题模型算法解析 300 封邮件, 得到 20 个主题词组, 如: “0.090* ‘交涉’ + 0.090* ‘小白脸’ + 0.090* ‘力阻’ + 0.090* ‘撕破脸’ + 0.090* ‘私事’”。

步骤 3 最后得到 300×20 维的权值矩阵, 300 表示 300 封邮件, 20 表示 20 个主题, 即每封邮件和 20 个主题之间的相关度。

采用 LDA 主题模型算法, 从测试集中选取 300 封邮件进行主题抽取 20 个主题, 主题词确定为 10 个, 选取两个具有代表性的主题, 结果见表 2。

4.2.2 LDA 反垃圾邮件过滤实验结果与分析

为了有效地验证该方法的可行性, 选用正常邮件和垃圾邮件各 8 000 封, 共 16 000 封作为训练集; 另取正常邮件和垃圾邮件各 150 封, 共 300 封作为测试集。用本文基于 LDA 的垃圾邮件过滤方法进行实验, 其中主题数确定为 20 个, 主题词为 10 个, 共完成 5 组实验, 结果见表 3。

表 2 主题模型结果

主题	主题词	词与主题的相关度
主题 1	贵公司	0.090
	运输	0.089
	负责人	0.089
	发票	0.086
	财务	0.086
	代	0.085
	有限公司	0.084
	实业	0.084
	开	0.083
	广告	0.083
主题 2	帐	0.088
	出任	0.081
	总裁	0.081
	恪守	0.081
	微软公司	0.081
	相违	0.081
	北京时间	0.081
	全文	0.081
主题 3,4,...,20	新浪	0.081
	微软	0.081
...		...

表 3 基于 LDA 的垃圾邮件过滤方法测试结果

	阈值设定	召回率	正确率	F1 值
1	0.40	86%	85%	85%
2	0.44	79%	94%	85%
3	0.45	76%	95%	84%
4	0.42	88%	92%	89%
5	0.44	86%	88%	86%
平均	0.43	83%	90%	86%

在实验中, 把垃圾邮件的概率跟合法邮件的概率做比较, 需要选择判定垃圾邮件概率的阈值。阈值的控制比较重要, 如果太大则会漏掉大量垃圾邮件, 通过实验确定最佳阈值为 0.43 左右。

(1) 与其他方法比较

为了更好地说明本文设计算法的有效性, 本文选取了积累典型的垃圾邮件过滤方法进行比较, 包括基于 Naïve Bayes 的邮件过滤方法^[15,16]、基于 SVM 的邮件过滤方法^[17]、基于 kNN 的邮件过滤方法^[18]、基于 MTM (message topic model) 的邮件过滤方法^[3]、基于决策树的三步邮件过滤方法^[4]、基于 SVM-NB 的邮件过滤方法^[5]。其中,



前 3 种方法是以简单机器学习为基础的邮件过滤方法, MTM 方法建立了一种有效的方式, 用来检测邮件主题词, 三步邮件过滤法是以决策树为基础的, 而 SVM-NB 方法是基于朴素贝叶斯分类分类的方式, 对比见表 4。

表 4 各种不同方法邮件测试结果

方法	召回率	正确率	F1 值
Naïve Bayes ^[15]	66%	90%	76%
SVM ^[17]	81%	88%	84%
kNN ^[18]	73%	80%	76%
MTM ^[3]	80%	82%	77%
决策树 ^[4]	79%	81%	79%
SVM-NB ^[5]	85%	87%	82%
本文提出的 LDA 方法	83%	90%	86%

由表 4 可知, 本文基于 LDA 的垃圾邮件过滤方法使垃圾邮件的召回率相比 Naïve Bayes 方法、SVM 方法、kNN 方法、MTM 方法、决策树方法有很大提升, 分别上升了 17%、2%、10%、3%、4%; 识别正确率和 Naïve Bayes 相同, 而相比 SVM 方法、kNN 方法则分别提高了 2%、10%; F1 值相比 Naïve Bayes 方法、SVM 方法、kNN 方法、MTM 方法、决策树方法以及 SVM-NB 方法分别提高了 10%、2%、10%、3%、4%、2%。

在基于决策树三步邮件过滤方法中, 它利用决策树模型构建了三步法垃圾邮件过滤模式。对于 SVM-NB 算法, 它提出了基于朴素贝叶斯分类器的训练集分类方法, 提升了数据处理的顽健性, 此方法能够达到较高的垃圾邮件检测精度。相比于这两种方法, 本文推出的 LDA 算法能够更好地提取文本特征, 从而达到更高的分类精度。基于 LDA 的垃圾邮件过滤方法在垃圾邮件正确率方面和 Naïve Bayes 方法相同, 在垃圾邮件的召回率方面高于这 3 种方法, 并且具有较高的 F1 测试值。这说明该方法在性能上要优于 Naïve Bayes、SVM、kNN 方法。

(2) 采用不同主题数的结果比较

在实验 (1) 中, 选择主题数为 20 个, 主题词为 10 个, 共完成了 5 组实验, 并且与另外 3 种

邮件过滤方法进行了比较。在本实验中选择主题词仍为 10 个, 分别选择主题数为 10、15、20、25 个进行实验, 结果见表 5。

表 5 不同主题数下的测试结果

主题个数	阈值设定	召回率	正确率	F1 值
10	0.32%	80%	80%	80%
15	0.47%	85%	84%	84%
20	0.40%	86%	85%	85%
25	0.47%	90%	88%	88%

分析实验结果如图 4、图 5 所示, 在选取合适阈值的条件下, 系统的召回率和正确率随着主题数的增加而提高。其原因是, 随着主题数的增加, 对测试集邮件的语义划分更明确, 进而使得系统的召回率和正确率明显提升。基于这样的原理, 本方法可以取得较好的垃圾邮件过滤结果。

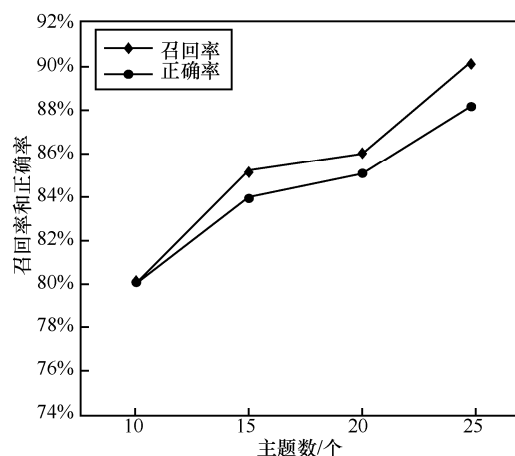


图 4 不同主题数下的召回率和正确率

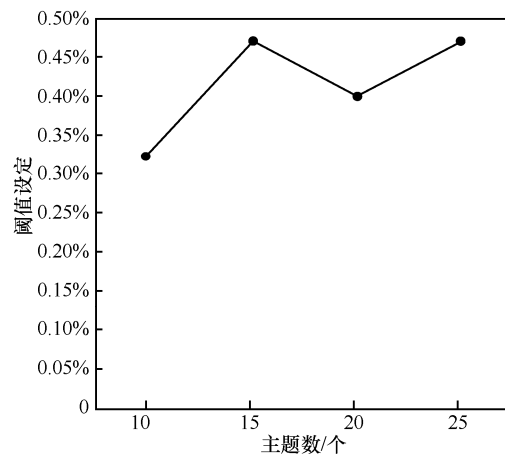


图 5 不同主题数下的阈值

4.2.3 基于改进的主题垃圾邮件过滤方法实验结果与分析

改进的方法主要将获得的主题进行扩展, 用 Word2Vec 方法得到每个主题中词的几个相关的词, 将获得的词重新构建主题组。

下面列举几个主题词经过 Word2Vec 计算后得到的结果, 如图 6 所示。

个性 0.753 134 727 478	致歉 0.737 105 131 149
性格 0.744 346 022 606	回应 0.723 619 818 687
开朗 0.742 993 593 216	此事 0.656 879 663 467
坦率 0.740 987 539 291	否认 0.652 116 060 257
直爽 0.736 198 425 293	对此 0.651 973 962 784
爽朗 0.730 711 936 951	表示遗憾 0.624 698 638 916
大而化之 0.721 120 536 327	表示歉意 0.624 465 227 127
活泼开朗 0.719 886 183 739	声明 0.622 716 546 059
天然呆 0.719 811 201 096	毫无诚意 0.608 891 963 959
轻浮 0.715 863 049 03	质疑 0.608 388 423 92

(a) 和“随和”最相关的词 (b) 和“道歉”最相关的词

交通 0.527 298 688 889	收据 0.638 935 208 321
陆上运输 0.521 845 817 566	现金支票 0.583 482 027 054
铁路 0.487 869 590 521	银行账号 0.571 007 490 158
货运 0.486 806 541 681	持票人 0.570 166 230 202
运输经济 0.486 096 262 932	收款 0.566 872 000 694
运输网络 0.478 215 456 009	退款 0.562 694 907 188
联合运输 0.476 194 471 121	开具 0.558 761 715 889
公共 0.474 913 895 13	挂失 0.556 449 472 904
运输业 0.473 794 817 924	转账 0.547 636 032 104
交通系统 0.472 240 090 37	持卡人 0.545 845 985 413

(c) 和“运输”最相关的词 (d) 和“发票”最相关的词

图 6 主题词经过 Word2Vec 计算后得到的结果

以其中一个主题为例, 通过 Word2Vec 扩展原主题词为 12 个词, 见表 6。

表 6 原主题与扩展主题

原主题词	词与主题 相关度	对应的扩展 主题词	词与主题 相关度
帐	0.088		
出任	0.081		
总裁	0.081	职位	0.081
恪守	0.081		
微软公司	0.081		
相违	0.081		
北京时间	0.081		
全文	0.081		
新浪	0.081		
微软	0.081	科技	0.081

这里实验过程分两个步骤。

步骤 1 将重建的主题组再次经 LDA 算法获得权值矩阵。

步骤 2 再用测试集进行测试, 得到最终的实验结果。

在本实验中选择主题数为 20 个, 主题词由 10 个扩展为 12 个, 共完成了 5 组实验。

测试结果见表 7。

表 7 基于 Word2Vec 改进的垃圾邮件过滤方法测试结果

	阈值设定	召回率	正确率	F1 值	表 3 中 F1 值
1	0.48	88%	88%	88%	85%
2	0.49	86%	92%	90%	85%
3	0.47	83%	84%	83%	84%
4	0.49	81%	83%	81%	89%
5	0.48	90%	88%	88%	86%
平均	0.48	85%	87%	85%	86%

分析表 7, 改进方法在增加 2 个关联主题词的情况下, 在 F1 值上比原方法改进明显, 在 5 次实验中有 3 次获得了较大的提高, 证明了改进方法的有效性。

5 结束语

本文对基于主题模型的垃圾邮件过滤系统的设计与实现进行了分析和验证, 与传统的关键词检测过滤技术相比, 贝叶斯过滤算法更加有效且智能, 从而提升了系统的安全性与顽健性。通过与其他典型垃圾邮件过滤方法的对比及验证, 证明基于主题模型的垃圾邮件分类方法及基于 Word2Vec 的改进方法均能有效提高垃圾邮件过滤的准确度。

在未来的研究中, 基于语义的文本分类具有非常大的潜力。针对自然语言的具体层次结构, 机器学习与深度学习的方式已经在其他领域表现出非常强大的处理能力。在这种背景下, 邮件拦截方法的设计可以参考相关研究成果进行深入探索。总之, 未来的邮件拦截系统将会



具有非常大的改进空间, 因此相关的研究需要被重点关注。

参考文献:

- [1] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.
- [2] 祝毅鸣, 张波. 实时黑名单在垃圾邮件过滤系统中的应用[J]. 科技资讯, 2012(12):33.
ZHU Y M, ZHANG B. Application of real time blacklist in spam filtering system[J]. Science & Technology Information, 2012(12):33.
- [3] MA J, ZHANG Y, WANG Z, et al. A message topic model for multi-grain SMS spam filtering[J]. International Journal of Technology & Human Interaction, 2016, 12(2):83-95.
- [4] SHEN J J, CHEN Y K, CHU K T, et al. An intelligent three-phase spam filtering method based on decision tree data mining[J]. Security & Communication Networks, 2016, 9(17): 4013-4026.
- [5] FENG W, SUN J, ZHANG L, et al. A support vector machine based naive Bayes algorithm for spam filtering[C]// 2016 Performance Computing and Communications Conference, Dec 9-11, 2016, Las Vegas, NV, USA. New Jersey: IEEE Press, 2017:1-8.
- [6] BANSAL R P, HAMILTON I R A, O'CONNELL B M, et al. System and method to control email whitelists: US, US 8676903 B2[P]. 2014.
- [7] CHAN P P K, YANG C, YEUNG D S, et al. Spam filtering for short messages in adversarial environment[J]. Neurocomputing, 2015, 155(C):167-176.
- [8] DEVI K S, RAVI R. A new feature selection algorithm for Efficient Spam Filtering using Adaboost and Hashing techniques[J]. Indian Journal of Science & Technology, 2015, 8(13).
- [9] AFZAL H, MEHMOOD K. Spam filtering of bi-lingual tweets using machine learning[C]// International Conference on Advanced Communication Technology, Jan 31-Feb 3, 2016, Pyeongchang, South Korea. New Jersey: IEEE Press, 2016.
- [10] DAS M, BHOMICK A, SINGH Y J, et al. A modular approach towards image spam filtering using multiple classifiers[C]//2014 IEEE International Conference on Computational Intelligence and Computing Research. Dec 20, 2014, Coimbatore, India. New Jersey: IEEE Press, 2015:1-8.
- [11] 曹玉东, 刘艳洋, 贾旭, 等. 基于改进的局部敏感散列算法实现图像型垃圾邮件过滤[J]. 计算机应用研究, 2016, 33(6):1693-1696.
CAO Y D, LIU Y Y, JIA X, et al. Image spam filtering with improved LSH algorithm[J]. Application Research of Computers,

2016, 33(6):1693-1696.

- [12] 徐凯, 陈平华, 刘双印. 基于 Adaboost-Bayes 算法的中文文本分类系统[J]. 微电子学与计算机, 2016, 33(6):63-67.
XU K, CHEN P H, LIU S Y. A Chinese text classification system based on Adaboost-Bayes algorithm[J]. Microelectronics & Computer, 2016, 33(6):63-67.
- [13] 周庆良. 一种基于 Adaboost 和分类回归树的垃圾邮件过滤算法[D]. 武汉: 华中科技大学, 2016.
ZHOU Q L. A spam filtering algorithm based on Adaboost and classification regression tree[D]. Wuhan: Huazhong University of Science and Technology, 2016.
- [14] SMITH D A, MCMANIS C. Classification of text to subject using LDA[C]//2015 IEEE International Conference on Semantic Computing (ICSC), Feb 7- Feb 9, 2015, Anaheim, CA, USA. New Jersey: IEEE Press, 2015: 131-135.
- [15] 赵治国, 谭敏生, 李志敏. 基于改进贝叶斯的垃圾邮件过滤算法综述[J]. 南华大学学报: 自然科学版, 2006, 20(1): 33-38.
ZHAO Z G, TAN M S, LI Z M. Review of spam filter algorithms based on improved Bayes[J]. Journal of Nanhua University(Science and Technology), 2006, 20(1): 33-38.
- [16] 林巧民, 许建真, 许棣华, 等. 基于贝叶斯算法的垃圾邮件过滤技术[J]. 南京师范大学学报: 工程技术版, 2005, 5(4): 61-64.
LIN Q M, XU J Z, XU D H, et al. Research on Bayes-based spam filtering[J]. Journal of Nanjing Normal University(Engineering and Technology), 2005, 5(4): 61-64.
- [17] LI L, MAO T, HUANG D. Extracting location names from Chinese texts based on SVM and KNN[C]// 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering(IEEE NLP-KE'05), Oct 30-Nov 1, Wuhan, China. New Jersey: IEEE Press, 2005: 371-375.
- [18] 林文香. 改进的KNN算法在过滤垃圾邮件中的应用研究[D]. 长沙: 湖南大学, 2010.
LIN W X. Application of improved KNN algorithm in spam e-mail filtering[D]. Changsha: Hunan University, 2010.

作者简介



寇晓淮 (1989-), 男, 华东理工大学信息科学与工程学院硕士生, 主要研究方向为信息分析与处理、智能信号处理和网络与信息安全。

程华 (1975-), 男, 博士, 华东理工大学信息科学与工程学院副教授, 主要研究方向为信息安全、信号处理、网络行为学和流量工程。



研究与开发

基于图转换的无线接入点关联优化算法

陶雯¹, 沈艳², 管昕洁², 万夕里²

(1. 江苏第二师范学院数学与信息技术学院, 江苏 南京 210013;

2. 南京工业大学计算机科学与技术学院, 江苏 南京 211816)

摘要: 无线接入点关联是 Wi-Fi 系统中的一个基础问题。已有的研究主要集中在考虑带宽容量约束条件下的各种 AP 关联问题。这些研究没有从用户的角度考虑带宽需求, 从用户的带宽需求出发, 以最多带宽分配的 AP 关联为优化目标, 考虑混合型网络架构下的 Wi-Fi 系统的无线接入点关联优化问题。与已有研究方法不同, 通过图转换技术, 并将问题转化为流图上的优化问题, 提出基于网络流的优化求解算法, 并从理论上证明算法的最优性。最后, 通过对比实验进一步证明了算法的优越性。

关键词: 无线接入; 图转换; 接入点关联

中图分类号: TP301

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017310

An association optimization algorithm for wireless access points based on graph transformation

TAO Wen¹, SHEN Yan², GUAN Xinjie², WAN Xili²

1. College of Mathematics and Information Technology, Jiangsu Second Normal University, Nanjing 210013, China

2. College of Computer Science and Technology, Nanjing Tech University, Nanjing 211816, China

Abstract: Wireless access point association problem is one of the fundamental problems for Wi-Fi systems. Existing studies focus on the AP association problems under the bandwidth capacity constrain, without considering bandwidth demand for users. Considering the user bandwidth demands, the access point association problem for a Wi-Fi system under hybrid network architecture was studied, with the objective of maximum bandwidth allocation. Different from existing studies, by utilizing graph transformation techniques, this optimization problem was transformed to a network flow optimization problem on a flow graph. Then, an algorithm was proposed based on the maximum network flow problem. Theoretic proof for the optimality of the algorithm was presented and simulations results further validated the superiority of the proposed algorithm.

Key words: wireless access, graph transformation, AP association

收稿日期: 2017-08-05; 修回日期: 2017-09-19

基金项目: 国家自然科学基金资助项目 (No.61602235); 江苏省自然科学基金资助项目 (No.BK20161007)

Foundation Items: The National Natural Science Foundation of China (No.61602235), The National Natural Science Foundation of Jiangsu Province of China (No.BK20161007)



1 引言

IEEE 802.11 技术（通常称为 Wi-Fi）因其具有高带宽、低成本等优势，随着社会的发展，基于 IEEE 802.11 协议系列的无线局域网已在校园、工厂等各种场景下不断得到部署。智能设备的普及更进一步推动了 Wi-Fi 的普及，并深入人们的生活和工作的各个方面。在企业层面，面向不同应用的 Wi-Fi 应用系统不断被推出。尽管 Wi-Fi 技术有着很多优点，随着其大规模的应用，其本身也面临很多技术上的挑战，如带宽优化、接入点的最优放置等问题。在这些挑战中，如何在给定 Wi-Fi 系统资源的前提下，使得系统能够服务更多用户是实际应用中经常考虑的一个重要问题。对该问题的深入研究，可以进一步提高 Wi-Fi 系统在实际应用的利用效率。

在实际的 Wi-Fi 的应用中，默认情况下，设备用户通过接收信号强度指示（received signal strength indication, RSSI）判断 Wi-Fi 接入点（access point, AP）的信号强度，并选择信号强度最强的 AP 与之关联。这种方法不能合理分配用户到不同的 AP 上，容易造成部分 AP 过载，而其他 AP 关联的用户偏少，从而造成系统资源不能被充分利用，降低了系统整体关联的用户数目。此外，这种方法只是以信号强度来关联用户和 AP，而没有考虑到每个 AP 的容量是有限的，即所能容纳的用户数是一定的。因此，如何在 AP 容量有限的前提下，合理调度分配系统资源，从而提升系统整体可接入用户数量是 AP 关联领域的一个重要问题^[1]。

近年来，随着 Wi-Fi 的不断普及，Wi-Fi 在企业级别也有了广泛的应用。如在机场、校园校区、商场等规模较大的公共区域，Wi-Fi 基本实现全覆盖。在这类型的 Wi-Fi 系统中，多个 AP 通过高速有线网络连接到中央控制节点，这种混合型结构一般被称为半集中式结构^[1]。在这种混合型结构

下，多个 AP 之间可以通过骨干有线网络来分享接入设备相关信息，这使得系统可以及时集中收集当前在各个 AP 信号范围内的用户信息，使得中央控制节点拥有全局视图，为设计 Wi-Fi 资源分配提供有利条件。

本文将研究在这一混合型架构下的 Wi-Fi 中最大带宽分配的 AP 关联问题。已有一些研究^[1-3]探讨了该 AP 关联问题。这些研究主要研究在带宽容量约束条件下，如何寻求带宽分配最多的一种 AP 关联方案。但这些研究并没有考虑到用户的带宽需求，随着移动互联网的发展，为了获得更好的用户网络体验，如上网、打游戏以及看视频等，用户对所分配到的带宽有着一定的要求。基于此，本文考虑从用户带宽需求出发，研究该需求下的最大带宽分配的 AP 关联优化问题。参考文献[1]首先考虑了该问题，研究了单元带宽需求下的最大带宽分配的 AP 关联优化，给出一个贪婪启发式算法来求解该问题，但该算法并不能求得该问题的最优解。已有的关于带宽分配的 AP 关联问题基本都是 NP 难问题，而且主要是应用线性、非线性规划等相关方法来设计启发式算法或者近似算法。与已有研究的方法不同，本文提出用图的优化理论来解决此类问题，并探讨在某些条件下存在最优解的可能性。本文给出了在每个设备需求是单元带宽的假设下，在基于图的最优化网络流理论基础上，发现最大带宽分配的 AP 关联优化存在最优解，理论证明和实验结果也进一步验证了该算法的正确性。

2 相关工作

随着当前接入 Wi-Fi 用户数目大量增长以及多层次结构的 Wi-Fi 系统的部署，提升系统利用效率和接纳更多用户请求的问题越发具有挑战性。为解决此问题，学术界提出了各种场景下，面向不同优化目标的资源调度和分配关联算法。参考文献[1]提出了面向带宽分配的 AP 关联分配

算法。参考文献[2]建模 AP 关联问题, 并考虑负载均衡因素, 通过松弛和取整的方法, 设计了一个具有常数倍逼近度的分配算法。参考文献[4]提出了基于最少负载优先设备关联的启发式算法, 也就是设备优先选择负载最小的 AP 接入。参考文献[5]考虑了 AP 可以再关联的优化问题, 提出了在迁徙约束条件下的面向最大吞吐量的优化算法。参考文献[6]提出了基于可用容量的负载均衡的 AP 关联算法。

其他学者从均衡用户分配的资源思路出发, 提出了多种关联算法。参考文献[3]研究了多种异构客户场景 (IEEE 802.11a/b/g/n) 下, 通过二维马尔可夫模型预测设备的上下行吞吐量, 在此基础上, 以提高 MAC 效率为目标, 考虑平衡带宽和速率, 建模在线 AP 关联问题并提出一种在线 AP 关联方法。参考文献[7]考虑了单个设备可关联多个 AP 的情况, 并提出最优的面向多个 AP 传输速率策略, 并从理论上证明该算法的逼近度为 $4+\epsilon$ 。参考文献[8]提出了度量驱动的 WLAN 设计原则, 并以此为指导优化信道分配、AP 关联和 AP 功率控制等。参考文献[9, 10]研究了在软件定义的 WLAN 框架下的 AP 关联问题, 考虑了帧延迟、数据分组丢失等多种因素, 提出面向服务质量保证的 AP 关联方法。参考文献[11]提出集中式的 AP 关联策略模型, 并应用蚁群算法求解 AP 关联解。

3 网络模型与问题描述

3.1 网络模型与定义

假设在混合型的 Wi-Fi 网络中, 有 n 个 AP 节点 $V=\{AP_1, AP_2, \dots, AP_n\}$, m 个移动设备 $U=\{d_1, d_2, \dots, d_m\}$, 网络中所有的 AP 通过一个有线的无线骨干网连接。所有 AP 可以通过该有线网络进行通信。每个 $U=\{d_1, d_2, \dots, d_m\}$ 的信号范围是有限的, 用 U_i 表示该信号覆盖范围。所有在 U_i 范围内的移动设备表示为 A_i 。网络中每个 AP_i

能接收到在自身信号范围之内内的所有移动设备的关联请求。假设每个请求关联的移动设备 j 对带宽的需求为 r_{ij} 。此外, 假设每个移动设备至少有一个可选的 AP。注意, 在任何时刻, 每个移动设备只能关联一个 AP。

图 1 的场景中有 3 个 AP $\{AP_1, AP_2, AP_3\}$ 和 4 个移动设备 $\{d_1, d_2, d_3, d_4\}$, 虚线圆圈为 AP 的信号覆盖范围。由图 1 可见, d_2 既被 AP_1 的信号覆盖, 也被 AP_2 的信号覆盖, 因此, d_2 的候选 AP 集是 $\{AP_1, AP_2\}$ 。类似的, d_4 的候选 AP 集是 $\{AP_2, AP_3\}$ 。

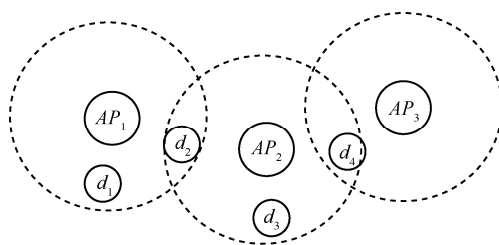


图 1 3 个 AP 和 4 个移动设备

假设 AP_i 的带宽容量为 C_i , 并且任何移动设备只需要 AP 能满足其最小的带宽即可关联该 AP。定义 L_i 为第 i 个 AP 的负载。为了方便, 与已有研究一样^[1], 本文假设每个移动设备只需要候选 AP 满足 1 个单元的带宽要求, 即可关联该 AP。因此, 第 AP_i 的负载 L_i 为该 AP 关联的移动设备的数目, 而每个 AP_i 的容量 C_i 则为该 AP_i 能关联的移动设备最大数目。因此, 在此假设下, 最终所有关联上的设备所获得的总带宽与设备总的连接数等价。

如果设备 d_j 关联到 AP_i , 定义 $a_{ij}=1$, 否则 $a_{ij}=0$ 。此外, 本文也用 $\langle u_j, v_i \rangle$ 配对来表示设备 j 关联到 AP_i 。

3.2 问题描述

定义 1 (最大 AP 关联问题) 给定一个混合型 Wi-Fi 网络, 该网络有 n 个 AP 节点和 m 个移动设备节点。假设每个移动设备候选的 AP 节点是已知的, 最大 AP 关联问题就是求解一种 AP 关联算法使得最终网络中 AP 所关联的设备数目最多。



该问题可以形式化为如下数学式:

$$\max \sum_{i \in [1, n]} L_i = \max \sum_{j \in [1, m]} a_{ij} \quad (1)$$

$$\text{s.t.} \quad \sum_{j \in [1, m]} a_{ij} \leq C_i, i = 1, 2, \dots, n \quad (2)$$

$$\begin{aligned} \sum_{i \in [1, n]} a_{ij} &= 1, j = 1, 2, \dots, m \\ a_{ij} &\in \{0, 1\} \end{aligned} \quad (3)$$

目标函数中的 $L_i = \sum_{j \in [1, N]} a_{ij}$ 表示第 i 个 AP 的负

载。式 (2) 是容量限制约束条件, 也就是每个 AP 最终的负载必须要小于或等于该 AP 的容量。式 (3) 是每个移动设备连接数目限制, 也就是每个移动设备最多只能关联一个 AP。注意, 一个 AP 可以同时被一个或多个移动设备关联。

定义 2 (二分图 $G(U \cup V, E)$) $G(U \cup V, E)$ 是一个二分图 (二分图实例如图 2 所示), 其中 U 和 V 分别是二分图上左部分和右部分点的集合, E 是 U 和 V 之间的边的集合。对于每一个 AP 节点 AP_i , 在 V 中创建一个节点 i ; 对于每一个移动设备 j , 在 U 中创建一个节点 j ; 若 AP 节点 AP_i 是移动设备 j 的关联候选节点, 则在 E 中添加一条从 i 到 j 的有向边 (i, j) 。

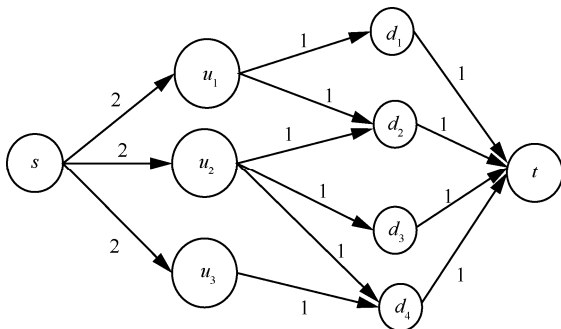


图2 图1对应的流图 G'

定义 2 实际上是将 AP 关联问题映射到对应的二分图 $G(U \cup V, E)$ 上, 是下述图转换和相应算法设计的基础。

定义 3 如果一个 AP 关联方案满足式 (2)、式 (3), 则称该关联方案为最大 AP 关联问题的可行解。

4 算法设计与分析

本节给出求解最大关联问题的最优算法以及相关最优性的证明和分析。该算法的思路是将该 AP 关联问题转化为图上的优化问题, 从而应用图的优化理论来求解最优解。首先将最大 AP 关联问题转化为二分图上的优化问题, 然后给出优化算法来求解该优化问题。

值得注意的是, 本文中的最大 AP 关联问题不能采用经典的二分匹配算法来求解, 因为对于每一个 AP 节点 (即二分图中 U 中的节点) 是可以关联多于一个的移动设备 (即二分图中 V 中的一个节点), 而二分匹配问题中 U 有且仅能与 V 中一个节点匹配。为解决此问题, 本文提出一种图转换方法, 将最大 AP 关联优化问题转化为对应的图优化问题, 从而可以应用网络流优化算法取得最优解。

4.1 基于网络流的 AP 关联算法

首先本文给出如何构造最大 AP 关联问题二分图: 创建一个二分图 $G(U \cup D, E, W)$, 对每个 AP_i 节点和移动设备 d_j , 分别在 U 和 D 中创建一个节点 i 和 j 。如果移动设备 j 在 AP_i 的覆盖范围内, 也就是 AP_i 是移动设备 d_j 的候选节点, 则在 E 中创建一条从 i 到 j 的有向边 (i, j) , 并设置 (i, j) 的权重为 $W(i, j)=1$ 。

下面给出基于网络流的最大 AP 关联算法。

算法 1 基于网络流的 AP 关联算法

输入: AP 集合 U , 移动设备集合 U , 每个移动设备的候选 AP 集合。

输出: 一组移动设备关联 AP 的解 $\langle a_{ij} \rangle$ 。

步骤 1 根据上述描述, 构造最大 AP 关联问题二分图 $G(U \cup D, E, W)$ 。

步骤 2 在 G 的基础上建立流图 G' (如图 2 所示), 在 $G(U \cup D, E, W)$ 上额外添加一个虚拟源点 S 和终点 T , 对 U 中的每个顶点 i , 连接源点 S 与顶点 i 形成有向边 (S, i) , 设置该边的权重为 C_i ; 对 V 中顶点 j , 连接 j 与终点 T 形成有向边 (j, T) ,

并设置该边的权重为 $W(j,T)=1$ 。

步骤 3 应用经典最大网络流 Push-Relabel 算法^[14,15]在新构造的流图 G' 上, 求解出 G' 上的最大流 f' 。用 $f'(i,j), i \in U \& j \in V$ 表示最大流解 f' 在边 (i,j) 最大流。

步骤 4 如果 $f'(i,j)=1, i \in U \& j \in V$ 则赋值 $a_{ij}=1$, 即将移动设备 d_j 关联到 AP_i 。

4.2 算法分析

引理 1 在图 2 中, 若所求得从源点 S 到终点 T 的最大流经过边 (u_i, d_j) , 则该边对应的最大流必定为 1, 即 $f(u_i, d_j)=1$ 。

证明: 根据网络流整数流性质^[12], 如果网络中所有边的权重 (或者容量) 为整数, 则该网络的最大流一定为整数流。算法 1 中所构造的流图 G' 的边的权重均为整数, 因此最大流经过 (u_i, d_j) 一定是整数。又因为 (u_i, d_j) 边的权重均为 1, 所以 (u_i, d_j) 若有最大流经过, 其最大流必定为 1, 即 $f(u_i, d_j)=1$ 。

引理 2 对于任何 $d_j \in D$ 节点, 若一个从源点 S 到终点 T 的流经过该节点, 则 $\max_{u_i \in U} \sum f(u_i, d_j)=1$ 。

证明: 对于任何 $d_j \in D$ 节点, 因为 (d_j, T) 的权重为 1, 根据网络流平衡性质^[13], 任何流入 d_j 的流的总和一定为 1, 因此引理得证。

定理 1 如果有一个可行性解 Γ 中, 有 k 个移动设备与 AP_i 节点关联, 则对应流图 G' 上存在一个从源点 S 到终点 t 的流, 且该流值为 k 。反之亦成立。

证明: 一个可行解 Γ 中有 k 个移动设备与 AP_i 关联, 则在流图 G' 上可以存在 k 条 (u_i, d_j) 边的流值为 1。因为源点 S 与每个 u_i 相连, 并且每个 d_j 与终点 T 相连, 根据网络流平衡性质^[13], 必定存在从源点 S 到终点 T 的流 f , 且 f 通过这 k 条 (u_i, d_j) 边, 且该流值为 $f(s, t)$ 。又因为可行解 Γ 满足式 (2), 也就是可行解满足每个 AP 的容量约束条件, 而边 (s, u_i) 上的权重则对应 AP 的容量, 因此, 该流

f 一定满足流图 G' 上的权重限制条件。

反之, 对于一个流值为 k 的从源点 S 到终点 T 的流值 $f(s, t)$, 结合引理 1 和引理 2, 可知对于任何 $d_j \in D$ 节点, 只有一条从 $u_i \in D$ 出发的边的流经过 d_j , 且 $f(u_i, d_j)=1$ 。因为 $f(s, t)=k$, 根据流平衡性质, 则必定有 k 条 (u_i, d_j) 边的流值为 1。也就是有 k 条 $f(u_i, d_j)=1$, 对应于 k 个移动设备与 AP 节点关联的方案, 并且该方案满足式 (3)。又因为 (s, u_i) 边的权重对应于 AP 的容量约束, 所以该关联方案满足式 (2), 因此该关联方案是一个有 k 个移动设备被关联的可行解。

定理 2 最大 AP 关联问题可以转化为流图 G' 上的最大网络流问题。

证明: 定理 1 揭示了单个 AP 关联问题的可行解与流图 G' 上的单源单汇流一一对应关系, 并且可行解的值对应于图 G' 的流值。值得注意的是, 最大 AP 关联问题的解是由一系列可行解构成, 且最大带宽 AP 关联问题的设备连接数值是这些可行解相应的设备连接数值之和。从而, 最大 AP 关联问题的解值 (设备连接数目) 对应于流图 G' 上的所有的单源单汇流值之和, 也就是对应于流图 G' 上的从源点 S 出发到终点 T 的网络流。因此, 最大 AP 关联问题解的最大值对应流图 G' 上的最大网络流。

定理 3 最大 AP 关联问题可以在 $O((m+n)^3)$ 时间内求的最优解。

证明: 算法 1 的复杂度是由最大网络流算法主导, 因此网络流算法的复杂度决定了算法 1 的复杂度。如果采用经典的 Push-Relabel 算法^[14,15], 则算法 1 复杂度为 $O(r^3)$, 其中, r 为流图 G' 中节点的个数。因此算法 1 的复杂度为 $O((m+n)^3)$ 。

4.3 示例

本节给出一个示例来演示算法 1。针对图 1 中的部署场景, 假定每个 AP 的容量为 2, 也就是 $C_i=2, (i=1,2,3)$ 。根据算法步骤 1 和步骤 2, 可以得到图 2 中的流图 G' 。在图 2 中, $\{u_1, u_2, u_3\}$ 分别



对应于 3 个 AP, $\{d_1, d_2, d_3, d_4\}$ 则对应于 4 个移动设备。 G' 中依附在边上的数值对应于边的权重, 其中对每个 $AP_i (i=1,2,3)$, $w(S, i) = C_i = 2$ 。其他边的权重均为 1。

应用最大网络流算法, 可求得该流图 G' 的最大流 $f=4$, 如图 3 所示, 加粗的边为网络最大流经过的边, 依附在边上加粗的数字则表示最大流经过该边的流值, 如边 (s, u_1) 的流的值为 2, 也就是 $f(s, u_1)=1$ 。根据算法 1 的步骤 4, 任何 (u_i, d_j) 边上流的值为 1, 则关联 AP_i 和移动设备 d_j 。因此, 从最大流流经 (u_i, d_j) 边的结果可以得到: AP_1 与移动设备 d_1 和 d_2 关联, AP_2 和移动设备 d_3 关联, AP_3 和移动设备 d_4 关联。

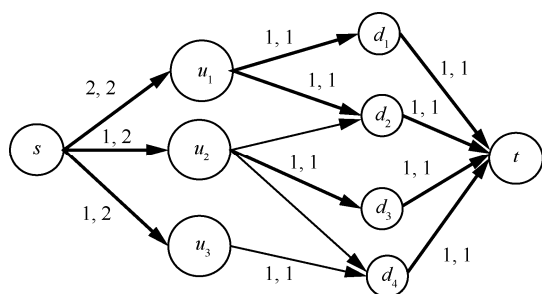


图3 图2对应的流图 G' 的最大网络流

5 实验与分析

5.1 实验参数设置

为进一步验证本文所提算法的性能, 通过实验与参考文献[1]中提出的贪婪算法对比, 通过对 AP 的关联设备数目、AP 关联数目的分布以及 AP 的利用率对比, 对算法性能进行实验分析和评估。

实验中有 n 个 AP 节点, m 个移动设备。每个 AP 节点的容量是 C , 实验中假设每个 AP 节点的带宽容量相同。本实验通过 AP 关联移动设备数目来判断本文所提出的算法的有效性和合理性。

为了更全面地突出本文算法的有效性, 从以下几个方面进行了仿真实验。

(1) 固定 AP 节点的数量为 5, 且每个 AP 节点的带宽容量为 5, 设置不同的移动设备数量, 分

别为 10、15、20, 随机运行 10 次, 比较本文算法与贪婪算法的实验结果。

(2) 固定移动设备数为 10, 每个 AP 节点的带宽容量为 5, 设置不同的 AP 节点数, 分别为 5、6、7, 随机运行 10 次, 将本文提出的算法与贪婪算法进行比较。

(3) 固定 AP 节点数量为 5, 且每个 AP 节点的带宽容量为 5, 改变移动设备的数量, 分别为 5、10、15、20、25、30、35、40、45、50, 比较这两个算法的结果。

(4) 固定 AP 节点数为 6, 设置不同的 AP 节点的带宽容量, 分别为 5、6、7, 随着移动设备逐次增加的情况下 (移动设备数量分别为 5、10、15、20、25、30、35、40、45、50), 比较本文提出的算法与贪婪算法的实验结果。

(5) 比较两算法在 3 种不同的 AP 节点数和移动设备数的条件下, AP 节点与移动设备连接分布情况。

(6) 在 AP 节点数为 5, 容量为 5 的条件下, 随着移动设备数量逐次增加, 对两个算法中 AP 节点的利用率进行了比较。

5.2 实验结果

根据要求设置参数, 进行仿真实验。在实验中将本文的优化算法与已有的贪婪算法进行比较。针对本文求最大 AP 联结数的问题, 使用贪婪算法将其分成每个 AP 节点尽可能多地连接移动设备的子问题。在给定的一个部署场景中, 将每个 AP 节点与移动设备连接的数量按降序排列, 在确保每个移动设备只连接一个 AP 节点的前提下, 按之前排列的顺序依次留下原部署场景中的 AP 节点与移动设备的连接状态, 最后计算得到 AP 联结数。

本文首先设置 AP 节点的数量为 5, 每个 AP 节点的容量为 5。然后, 设置不同的移动设备数量, 分别运行基于网络流的 AP 关联算法和贪婪算法。运行 10 次, 并且将每次这两个算法生成的 AP 关联设备的数目进行比较。图 4 给出了当移动设备数量为 10、

15、20 时的比较结果,可以看出,使用基于网络流 AP 关联算法得到的 AP 关联数普遍高于贪婪算法。

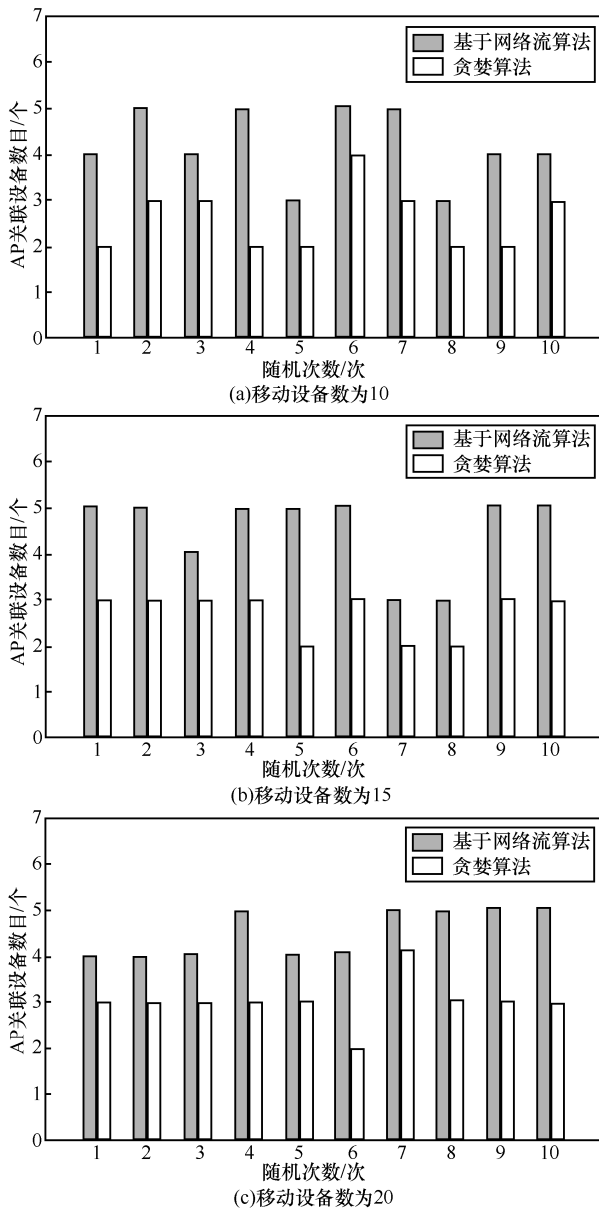


图4 AP 节点数为 5, 不同移动设备数量下运行 10 次的 AP 关联数

接下来,设置移动设备数量为 10, 每一个 AP 节点的容量为 5, 对不同数量的 AP 节点分别运行基于网络流 AP 关联算法和贪婪算法, 每一组实验运行 10 次, 对两个算法生成的 AP 关联设备的数目进行比较。图 5 给出了 AP 节点数为 5、6、7 时的比较结果。从图 5 可以看出, 基于网络流 AP

关联算法生成的 AP 关联设备的数目总是大于贪婪算法的, 且大多数时候是远大于贪婪算法的。这个结果与本文算法结论一致, 有效验证了本文算法的性能高于贪婪算法性能。

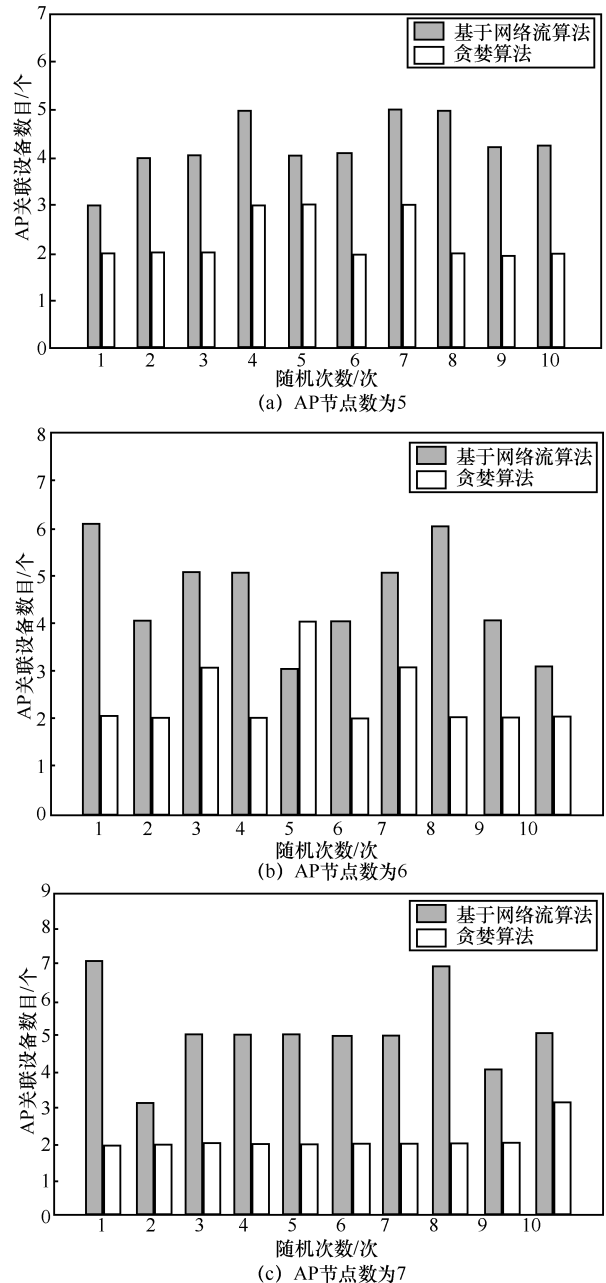


图5 移动设备数为 10、AP 节点数不同时运行 10 次的 AP 关联数

为了更进一步验证本文算法的有效性, 还进行了如下实验, 设置 AP 节点数为 5, 且每个 AP 节点的容量为 5, 移动设备数量依次为 5、



10、15、20、25、30、35、40、45、50，分别观察两种算法运行下生成的 AP 关联设备的数目，结果如图 6 所示。可以看到，在相同的 AP 数量和移动设备数量下，基于网络流 AP 关联算法得到的 AP 关联数始终大于贪婪算法得到的 AP 关联数，如此进一步验证了所提算法更为有效性。

另外，还改变 AP 节点的容量值进行两个算法的对比实验。实验中设置 AP 节点的数量为 6。然后设置不同的 AP 节点容量，分别运行基于网络流 AP 关联算法和贪婪算法，在移动设备数量逐次增加的情况下比较两种算法生成的 AP 关联设备的数目。图 7 给出了当 AP 节点容量为 5、6、7 时的比较结果。可以看到，在 AP 节点数为 6，且移动设备数量相同的时候，基于网络流 AP 关联算法得到的 AP 联结数大多数时候大于贪婪算法得到的 AP 联结数。另外还可以发现，在同一场景部署下，当该环境中的 AP 节点容量增大时，本文所提出的算法的性能明显优于贪婪算法。

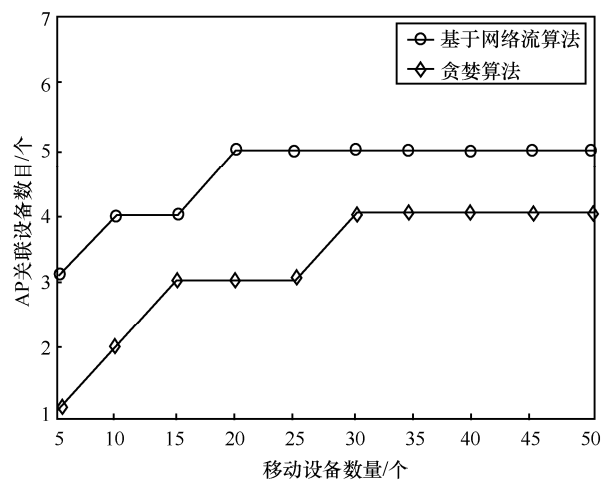


图 6 AP 节点数为 5，随着移动设备数量增加生成的 AP 关联数

接下来，分别运行基于网络流 AP 关联算法和贪婪算法，比较 3 种不同的 AP 节点数和移动设备数的条件下 AP 节点与移动设备连接分布情况。图 8 给出了 AP 节点数为 5，且每个 AP 节点的容量为 5，移动设备数为 10 的情况下，

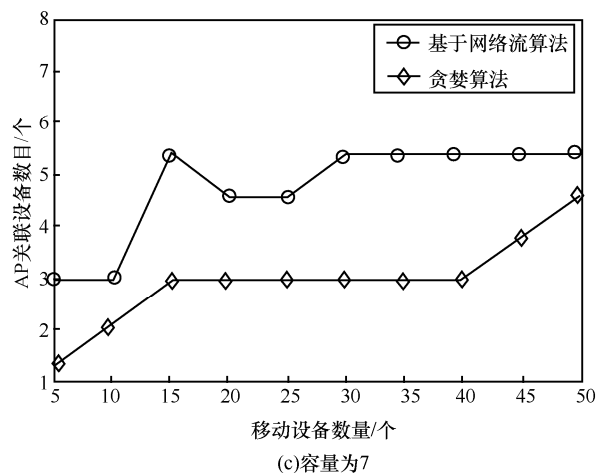
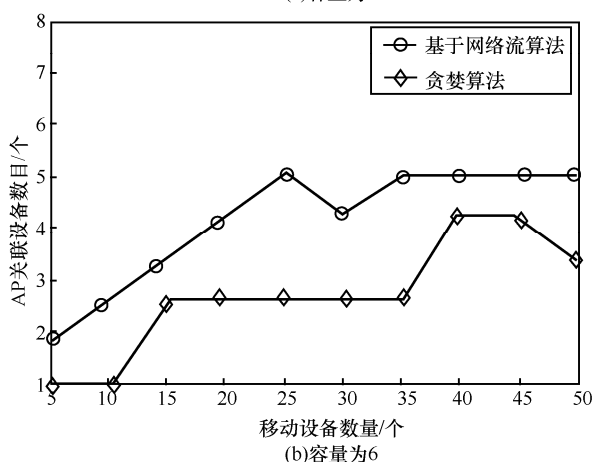
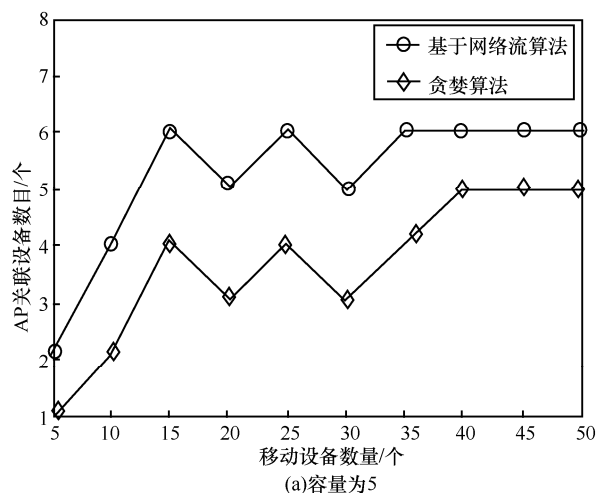


图 7 AP 节点数量为 6 时，不同容量下随着移动设备数量增加生成的 AP 关联数

AP 连接数的分布情况。从图 8 中可以看出，在基于网络流 AP 关联算法中每个 AP 节点都连接着相应数量的移动设备，而贪婪算法下第 3 个 AP 节点、第 4 个 AP 节点和第 5 个 AP 节点都

没有与移动设备相连接。所以很明显地得出本文提出的算法更为有效。另外,本文还分别进行了 AP 节点数为 6,且每个 AP 节点的容量为 5,移动设备数为 15 和 AP 节点数为 7,且每个 AP 节点的容量为 5,移动设备数为 20 的实验。结果分别显示在图 9 和图 10 中。结果与上述结论保持一致。

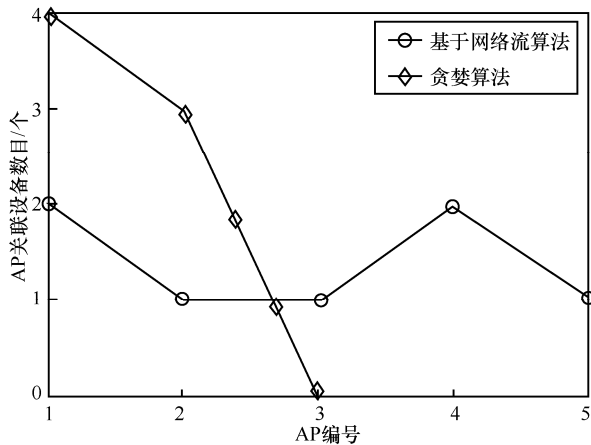


图8 AP节点数为5,移动设备数为10的AP连接数分布

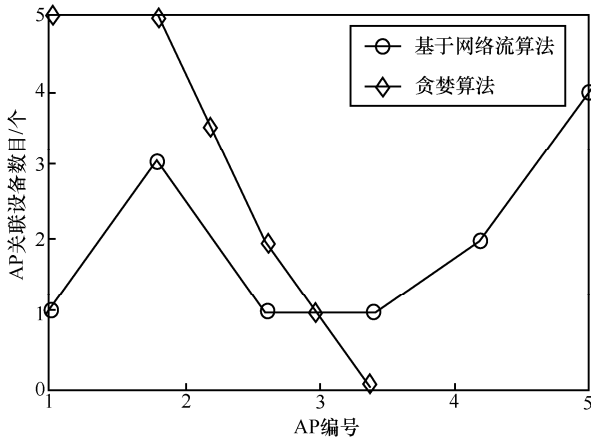


图9 AP节点数为6,移动设备数为15的AP连接数分布

此外,本文还对两个算法中 AP 节点的利用率进行了比较。图 11 给出了 AP 节点为 5,且每个 AP 节点的容量为 5,移动设备数量逐量递增的条件下两算法中 AP 节点的利用率情况。可以看出,基于网络流的 AP 关联算法下所得到的 AP 利用率一直高于贪婪算法,而且大部分情况下,AP 利用率为 100%。由此可以看出,本文提出的算法在 AP 利用情况上比贪婪算法更为高效。

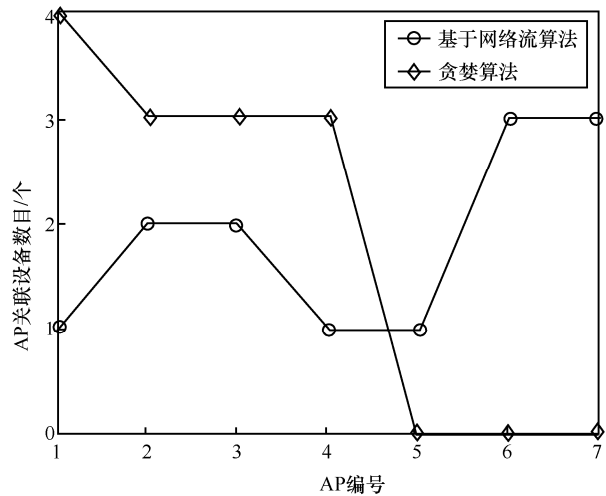


图10 AP节点数为7,移动设备数为20的AP连接数分布

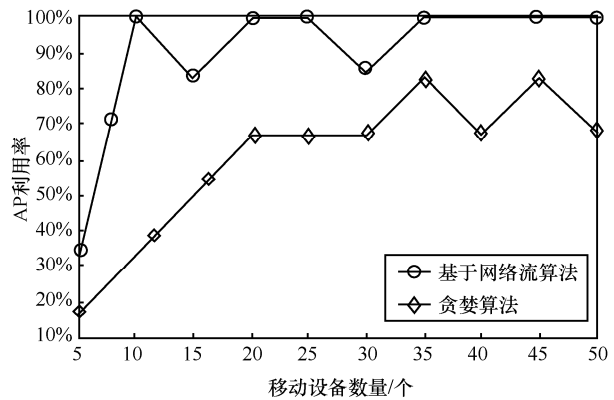


图11 AP节点数量为5,容量为5时随着移动设备数量增加生成的AP利用率

6 结束语

本文提出了在混合型架构下的一种面向最大连接数的 AP 关联算法。与已有 AP 关联方法不同,本文应用图优化方法来设计 AP 关联算法,提出图转换技术,设计流图,将该 AP 关联优化问题转化为流图上网络流优化问题,并从理论上证明了两个问题的等价性。在此基础上,应用网络流算法,提出基于网络流算法的最优 AP 关联算法。通过仿真实验,验证了该算法的有效性。

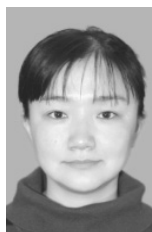
参考文献:

- [1] DANDAPAT S K, MITRA B, CHOUDHURY R R, et al. Smart association control in wireless mobile environment using max-flow[J]. IEEE Transactions on Network and Service Man-



- agement, 2012, 9(1): 73-86.
- [2] BEJERANO S J Y, HAN S J, LI L E. Fairness and load balancing in wireless LAN using association control[M]. New Jersey: IEEE Press, 2007: 560-573.
- [3] GONG D, YANG Y. On-line AP association algorithms for 802.11n WLANs with heterogeneous clients [J]. IEEE Transactions on Computers, 2014, 63(11): 2772-2786.
- [4] GONG H, KIM J. Dynamic load balancing through association control of mobile users in Wi-Fi networks[J]. IEEE Transactions on Consumer Electronics, 2008, 54(2): 342-348.
- [5] WONG W, THAKUR A, CHAN S H G. An approximation algorithm for AP association under user migration cost constraint[C]//The 35th Annual IEEE International Conference on Computer Communications, April 10-15, 2016, San Francisco, USA. New Jersey: IEEE Press, 2016: 56-62.
- [6] MURTY R, PADHYE J, CHANDRA R, et al. Designing high performance enterprise Wi-Fi networks[C]//The 5th USENIX Symposium on Networked Systems Design and Implementation, Alaska, USA, 2008: 342-349.
- [7] ABUSUBAIH M, WOLISZ A. An optimal station association policy for multi-rate IEEE 802.11 wireless Lans [C]//The 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems, October 22-26, 2007, Chania, Crete Island, Greece. New York: ACM Press, 2007: 34-40.
- [8] BROUSTIS I, PAPAGIANNAKI K, KRISHNAMURTHY S, et al. Measurement-driven guidelines for 802.11 WLAN design[J]. IEEE/ACM Transactions on Networking, 2010, 18(3): 722-735.
- [9] SOOD K, LIU S, YU S, et al. Dynamic access point association using software defined networking[C]//2015 International Telecommunication Networks and Applications Conference, Nov 18-20, 2015, Sydney, Australia. New Jersey: IEEE Press, 2015: 230-238.
- [10] CHEN J, LIU B, ZHOU H, et al. QoS-driven efficient client association in high-density software defined WLAN [J]. IEEE Transactions on Vehicular Technology, 2017, PP(99): 1-2.
- [11] 李克, 王换招, 张鹏, 等. 一种用户需求感知的无线接入点关联策略[J]. 软件学报, 2015, 26(2): 100-110.
- LI K, WANG H Z, ZHANG P, et al. User-demand-aware wireless access point association strategy[J]. Journal of Software, 2015, 26(2): 100-110.
- [12] PAPADIMITRIOU C H, STEIGLITZ K. Combinatorial optimization: algorithms and complexity[M]. New Jersey: Prentice-Hall, 1998.
- [13] CORMEN T H, LEISERSON C E, RIVEST R L, et al. Introduction to algorithms: 3rd edition[M]. Cambridge: MIT Press, 2009: 60-73.
- [14] WAN X, WU J, SHEN X. Maximal lifetime scheduling for roadside sensor networks with survivability k[J]. IEEE Transactions on Vehicular Technology, 2015, 64(11): 5300-5313.
- [15] CHERKASSKY B V, GOLDBERG A V. On implementing push-relabel method for the maximum flow problem[J]. Integer Programming and Combinatorial Optimization, 1995, 35(10): 157-163.

[作者简介]



陶雯 (1979-), 女, 江苏第二师范学院数学与信息技术学院讲师、CCF 会员, 主要研究方向为无线网络协议。

沈艳 (1984-), 女, 南京工业大学计算机科学与技术学院硕士生, 主要研究方向为网络优化。

管昕洁 (1984-), 女, 博士, 南京工业大学计算机科学与技术学院讲师, 主要研究方向为云计算和软件定义网络。

万夕里 (1982-), 男, 博士, 南京工业大学计算机科学与技术学院讲师, 主要研究方向为网络优化和算法设计。



基于比例因子转移概率的 AAC 音频压缩历史检测算法

黄其娟, 王让定, 严迪群

(宁波大学信息科学与工程学院, 浙江 宁波 315211)

摘要: 音频压缩历史的检测是音频取证的重要组成部分, 对判断音频是否经过篡改和伪造有着十分重要的意义。通过研究发现, AAC 音频比例因子的值会随着压缩次数的增加而逐渐减小。基于此, 提出了一种基于比例因子转移概率差值统计特性的 AAC 音频压缩历史检测算法。实验结果表明, 该算法能对多次压缩的 AAC 音频进行准确分类, 其低码率转高码率间 AAC 音频平均分类准确率达到了 99.75%, 同码率间准确率达 97.28%。另外, 对比实验也证明了本文算法的性能优于现有算法。

关键词: AAC 音频; 压缩历史; 比例因子; 转移概率

中图分类号: TP391

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017277

AAC compression detection based on scaling factor transition probability

HUANG Qijuan, WANG Rangding, YAN Diqun

College of Information Science and Engineering, Ningbo University, Ningbo 315211, China

Abstract: Audio compression history detection is an important part of audio forensics, which is important to detect whether audio has been tampered or forged. An algorithm of AAC audio compression history detection was presented by using the transition probability differences of scale factors as the discriminative feature. Experimental results demonstrate that the proposed method can distinguish the single, double and triple compressed AAC audios effectively, and from the low-bit-rate to high-bit-rate, the average classification accuracy achieves 99.75%, the same-bit-rate detection accuracy achieves 97.28%. In addition, the results of comparison experiments show that the proposed algorithm outperforms the state-of-the-art algorithm.

Key words: AAC audio, compression history, scale factor, transfer probability

收稿日期: 2017-07-13; 修回日期: 2017-09-26

通信作者: 王让定, wangrangding@nbu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61672302, No.61300055); 浙江省自然科学基金资助项目 (No.LZ15F020010, No.Y17F020051); 宁波大学科研基金资助项目 (No.XKXL1405, No.XKXL1420, No.XKXL1509, No.XKXL1503); 宁波大学王宽诚幸福基金资助项目

Foundation Items: The National Natural Science Foundation of China (No. 61672302, No.61300055), Natural Science Foundation of Zhejiang Province of China (No.LZ15F020010, No.Y17F020051), The Scientific Research Foundation of Ningbo University (No.XKXL1405, No.XKXL1420, No.XKXL1509, No. XKXL1503), K.C. Wong Magna Fund in Ningbo University



1 引言

随着多媒体和互联网的不断发展,人们可便捷地借助音频、图像、视频设备来分享、获取各种各样的重要信息。这些多媒体信息无时无刻不萦绕在人们的生活中,在给人们生活带来便利的同时,也存在着很多信息安全隐患。信息化的快速发展使得信息安全形势愈加严峻^[1,2],由此,多媒体取证技术应运而生,并得到了快速发展。作为多媒体取证技术的重要分支之一,音频取证也得到迅速发展。

音频压缩历史检测是音频取证技术中的热点问题之一,在对音频进行篡改、插入等操作时,都会对音频进行压缩,导致音频压缩历史的变化。音频压缩历史的检测是判断音频是否被篡改或伪造的一个必要条件。

压缩域音频历史的检测主要集中于双压缩音频的检测。近年来,国内外研究者对 MP3 压缩域音频的压缩历史的检测取得了一些研究成果。D'alessandro 等人^[3]通过分析不同比特率下音频的功率谱,发现其在 16~20 kHz 频带内呈现不同分布,从而提出通过频谱分析检测 MP3 音频真实比特率的方法。Liu 等人^[4]通过修正离散余弦变换(modified discrete cosine transform, MDCT)系数绝对值高于设定阈值的比例,实现单压缩音频和双压缩音频分类,但该阈值的设定基于作者经验值,不具备普适性。Qiao 等人^[5]使用小值 MDCT 系数数量、相邻帧间过渡连续性及系数离散数值分布等特征进行双压缩音频检测。在非压缩域音频压缩历史检测方面,Luo 等人^[6]通过提取 MDCT 系数作为其压缩历史检测的有效特征,在检测是否经过有损编码器重压缩的基础上检测其压缩采用的比特率。Luo 等人^[7]通过增加梅尔倒谱系数(Mel frequency cepstrum coefficient, MFCC)作为检测特征,提高了算法检测准确率,但特征维数较高,增加了计算量。Ren 等人^[8]去除待测音频

的前 200 个采样点,重新压缩得到校准音频,并利用待测音频与校准音频量化 MDCT 系数的不同,进行 MP3 双压缩音频检测和压缩音频原始码率估计。

在 AAC 音频压缩历史检测方面,Seichter 等人^[9]利用卷积神经网络对 AAC 压缩音频进行分类并估计压缩音频的原始码率。Jin 等人^[10]使用 Huffman 码表索引出现的概率以及其 Markov 单步转移概率作为特征,利用 LIBSVM 对单压缩和双压缩 AAC 音频进行分类。虽然该方法在低码率转高码率的压缩音频取得了较好的检测效果,但在同码率的压缩音频方面,识别率较低。李晗等人^[11]利用 AAC 在重压缩过程中频率成分的变化、能量值的不同,使用堆栈自动编码器(SAE)进行双压缩历史的检测,但其并没有对具体码率的单双压缩的音频进行检测,且检测准确率较低。

本文的研究对象仅为经过三次及三次以下的 AAC 压缩音频。可以发现,在音频压缩过程中,比例因子随压缩次数而发生变化,随着压缩次数的增加,比例因子会逐渐减小。根据音频压缩前后比例因子统计特性的差异,提出了一种基于比例因子统计特性的 AAC 音频压缩历史检测算法。该算法将比例因子转移概率差值作为检测特征,利用 LIBSVM 对多压缩的 AAC 音频进行分类。实验结果表明,本文算法检测准确率较高,能有效检测 AAC 音频的压缩历史。

2 AAC 音频比例因子特性变化分析

2.1 AAC 中的比例因子

AAC 是高级音频编码(advanced audio coding)的简称,是 ISO/IEC MPEG-2 和 MPEG-4 标准的重要组成部分^[12]。MPEG-2 AAC 于 1997 年发布,因其在相同码率,尤其是低码率的情况下,有着比 MP3 音频更好的音质^[13],因此被看好能取代 MP3 成为新一代的音频编码标准。AAC 编码框架如图 1 所示,主要由心理声学模型、增益控

制、滤波器组、时域噪声整形、强度立体声耦合、预测、M/S 编码、量化和无噪声编码等部分组成。

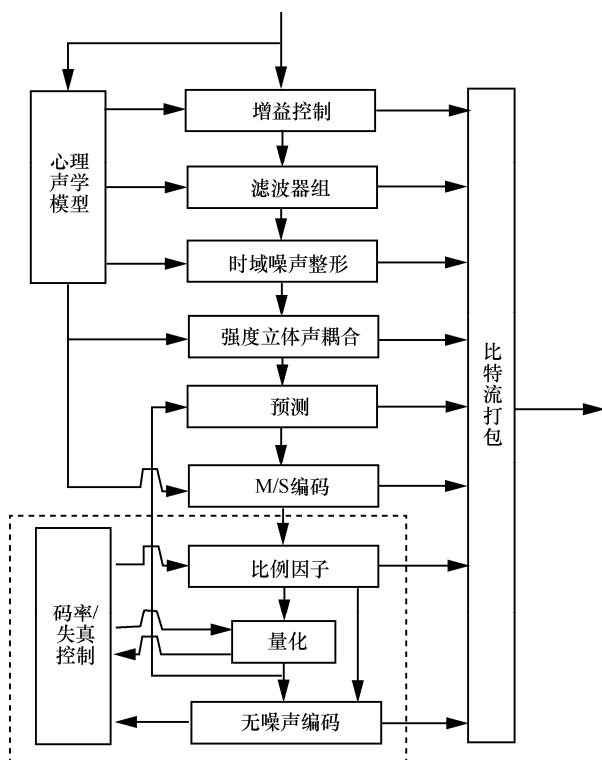


图 1 AAC 编码框架

AAC 进行 MDCT 时使用长窗块、短窗块、开始块和结束块这 4 种变换块类型：瞬时特性变化较为平稳的块使用长窗，变化较为剧烈的块使用短窗，另外两种块则用于长窗块和短窗块的过渡。使用长窗块时，AAC 具有较高的频域分辨率和较高的编码效率。而当音频信号变化剧烈时，AAC 通过使用短窗块来提高信号的时域分辨率。本文则主要关注长窗块情况下的编码。

AAC 标准将 1 024 个频域系数分成若干个连续的组分别进行量化，这样的组就称为比例因子带。比例因子带依据人耳听觉系统的临界带划分，其个数和长度取决于当前块类型和原始音频的采样率。如对原始采样率为 44.1 kHz 的音频，AAC 将其长块和短块分别划分为 49 个和 14 个比例因子带。每个比例因子对应一个比例因子带，AAC 中比例因子采用差分编码和 Huffman 编码的方式，比例因子的最大值可取到 255。

AAC 量化模块利用心理声学模型计算所得掩蔽阈值，求出各比例因子带的最佳比特分配方案。量化由帧循环、外循环和内循环 3 个迭代循环完成，其中帧循环负责初始化变量、计算可用比特数，外循环将量化环节的失真限制在可控范围内，内循环则实施对频谱系数的具体量化。AAC 内循环中的量化函数式为：

$$x_{\text{quant}} = \left\lfloor \left((|mdct_{\text{line}}| \times 2^{\frac{1}{4}(sf_{\text{decoder}} - 100)})^{\frac{3}{4}} + 0.4054 \right) \right\rfloor \quad (1)$$

其中， $\lfloor \cdot \rfloor$ 是向下取整函数， sf_{decoder} 表示量化步长， $mdct_{\text{line}}$ 、 x_{quant} 分别是量化前后的频谱系数。通过调整量化步长调节比例因子表示范围，将每个比例因子带的量化失真控制在最大可允许失真范围内，增加量化步长等价于间接减小比例因子^[14]。由于信号中高频分量的能量值较小，使得需要用较小的量化步长来保留高频信号的精度。在二次压缩之后，很多的高频信息会被量化成零值^[3]，相比一次音频压缩其量化步长增大，比例因子值减小。根据一次压缩和二次压缩时的统计信息，量化步长和比例因子是 AAC 二次压缩检测的重要特征。

2.2 比例因子特性变化分析

AAC 压缩音频（多次压缩）分为以下 3 类：低码率转高码率的 AAC 音频，这类压缩音频通常被称为假音质音频，由较差音质（低码率）的音频转码形成的虚假高品质（高码率）音频；相同码率的 AAC 音频，该类音频的产生多是由于原始压缩音频被篡改后，为了不让格式的改变引起他人怀疑而重新以相同码率重新压缩回篡改前的音频格式；高码率转低码率的 AAC 音频，这种情况虽也有可能发生，但目前在取证场景下的意义仍未被提出。因此，本文后续的压缩音频默认为上述第一类和第二类音频。

为验证 AAC 多次压缩对比例因子的影响，随机选取 100 首 10 s 的 WAV 音频，包括乡村、蓝调、流行、爵士等风格。将这 100 首 WAV 音频压



缩成 60 kbit/s、120 kbit/s 的 AAC 音频, 然后继续进行压缩操作, 分别得到 60 kbit/s→60 kbit/s、60 kbit/s→120 kbit/s 的二次压缩音频, 60 kbit/s→60 kbit/s→60 kbit/s、60 kbit/s→120 kbit/s→120 kbit/s 的 3 次压缩音频。提取每种码率 AAC 音频的比例因子, 并对其出现概率进行统计。AAC 比例因子的范围为[0,255], 如图 2 所示, 其比例因子值的分布近似服从拉普拉斯分布, 且主要值集中于[140,200], 为保证研究的普适性, 本算法仅对这个范围内的比例因子出现概率进行统计。

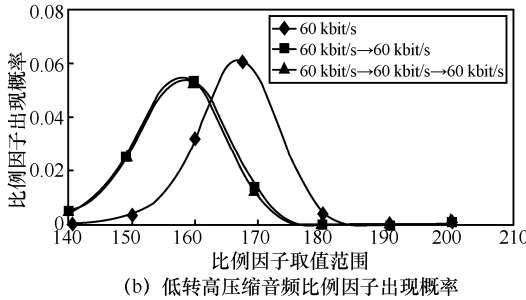
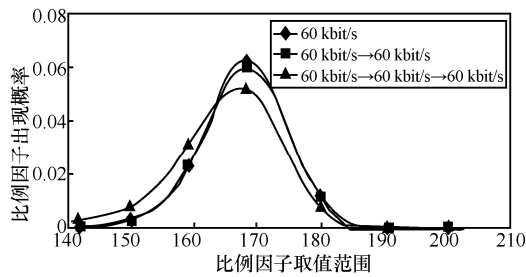


图 2 100 首 AAC 音频比例因子出现概率

由图 2 可以看出, 随着压缩次数的增加, 比例因子值相对减小。其中, 同码率音频压缩比例因子变化较小, 由低码率转高码率的音频比例因子变化相对较大。由此认为可通过分析比例因子的统计特性来判断 AAC 的压缩历史。由图 2 还可以看出, AAC 音频压缩使比例因子发生相应变化, 但压缩音频与原始音频比例因子较为接近, 难以区分音频的压缩历史。且随着压缩次数的增加, 比例因子最大值即曲线峰值的出现概率逐渐减小, 而其他比例因子的出现概率也会发生变换, 这说明压缩会使得部分比

例因子发生转化, 即可能会发生比例因子的相应转移。

3 特征构造与提取

假设一次压缩 AAC 音频为 $M1$, 对其进行解压再压缩, 得到新的 AAC 音频 $M2$, 然后继续进行三次压缩, 得到 AAC 音频 $M3$, 如图 3 所示。在本文的研究中, 假设二次压缩和三次压缩得到的 $M2$ 、 $M3$ 音频码率相等。

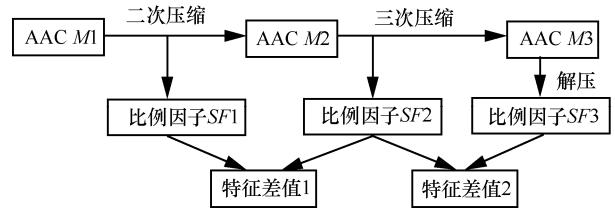


图 3 特征提取流程

图 3 中 $SF1$ 、 $SF2$ 、 $SF3$ 表示将一次压缩 AAC 音频 $M1$ 、二次压缩 AAC 音频 $M2$ 和三次压缩 AAC 音频 $M3$ 解压得到的比例因子矩阵, 可将其定义为:

$$SF1 = \begin{pmatrix} sf_{1,1} & \cdots & sf_{1,n} \\ \vdots & \ddots & \vdots \\ sf_{m,1} & \cdots & sf_{m,n} \end{pmatrix} \quad (2)$$

$$SF2 = \begin{pmatrix} sf'_{1,1} & \cdots & sf'_{1,n} \\ \vdots & \ddots & \vdots \\ sf'_{m,1} & \cdots & sf'_{m,n} \end{pmatrix} \quad (3)$$

$$SF3 = \begin{pmatrix} sf''_{1,1} & \cdots & sf''_{1,n} \\ \vdots & \ddots & \vdots \\ sf''_{m,1} & \cdots & sf''_{m,n} \end{pmatrix} \quad (4)$$

其中, n 代表比例因子带的个数, 若为单声道, 则 $m = 2I$; 若为双声道, 则 $m = 4I$ 。 I 代表待测音频长窗帧的总数。

本文统计了 140~200 内的比例因子, 但由于 140~150、180~200 内的比例因子出现概率过小, 在统计转移概率时, 仅选取在 150~180 内的比例因子进行统计。文中定义了两种转移概率的情况, 分别是块内转移概率 $p\{sf_{i,j+1}=n | sf_{i,j}=m\}$

和块间转移概率 $p\{sf_{i+1,j} = n | sf_{i,j} = m\}$, 即:

$$p\{sf_{i,j+1} = n | sf_{i,j} = m\} = \frac{\sum_{i=1}^{31} \sum_{j=1}^{30} \delta(sf_{i,j} = m, sf_{i,j+1} = n)}{\sum_{i=1}^{31} \sum_{j=1}^{30} \delta(sf_{i,j} = m)} \quad (5)$$

$$p\{sf_{i+1,j} = n | sf_{i,j} = m\} = \frac{\sum_{i=1}^{30} \sum_{j=1}^{31} \delta(sf_{i,j} = m, sf_{i+1,j} = n)}{\sum_{i=1}^{30} \sum_{j=1}^{31} \delta(sf_{i,j} = m)} \quad (6)$$

其中, $m, n \in \{150, 151, \dots, 180\}$, $\delta(X=i, Y=j) = \begin{cases} 1, X=i, Y=j \\ 0, \text{其他} \end{cases}$, $sf_{i,j}$ 是比例因子值。式(7)~式(9)

所示为得到的 $M1$ 、 $M2$ 、 $M3$ 的转移概率矩阵 $P1$ 、 $P2$ 和 $P3$, 将 31×31 组元素重组得到 961 维特征:

$$P1 = \begin{pmatrix} p_{1,1} & \cdots & p_{1,31} \\ \vdots & \ddots & \vdots \\ p_{31,1} & \cdots & p_{31,31} \end{pmatrix} \quad (7)$$

$$P2 = \begin{pmatrix} p'_{1,1} & \cdots & p'_{1,31} \\ \vdots & \ddots & \vdots \\ p'_{31,1} & \cdots & p'_{31,31} \end{pmatrix} \quad (8)$$

$$P3 = \begin{pmatrix} p''_{1,1} & \cdots & p''_{1,31} \\ \vdots & \ddots & \vdots \\ p''_{31,1} & \cdots & p''_{31,31} \end{pmatrix} \quad (9)$$

选取 100 首 $M1$ 、 $M2$ 和 $M3$ 音频, 提取比例因子块内和块间转移概率, 取其均值。图 4 所示为 $M1$ 、 $M2$ 和 $M3$ 块内和块间转移概率。图 4 中 (1)、(2)、(3) 分别代表 $M1$ 、 $M2$ 和 $M3$, 横坐标共 961 维, 代表比例因子由 M 转变为 N , M 、 N 如式(10)和式(11)所示, 纵坐标代表比例因子的转移概率:

$$M = x / 31 + 150 \quad (10)$$

$$N = x \% 31 + 149 \quad (11)$$

其中, $x = 1, 2, \dots, 961$ 。

从图 4(a) 和图 4(b) 中可以看出, $M1$ 、 $M2$ 和 $M3$ 三次压缩的比例因子转移概率差别较小, 无法利用其进行压缩音频的分类。

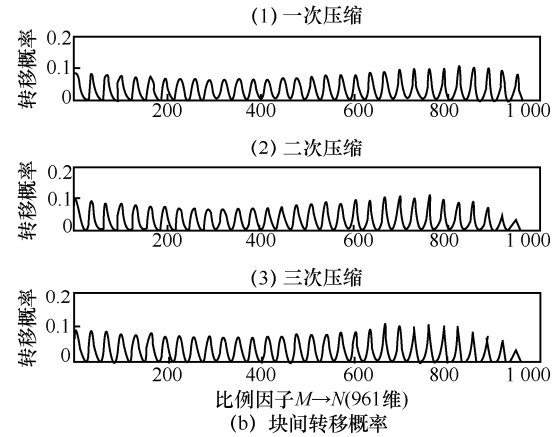
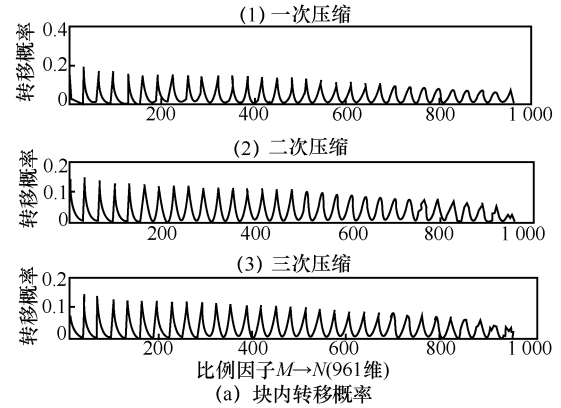


图 4 AAC 压缩音频的转移概率

利用得到的转移概率矩阵 $P1$ 、 $P2$ 和 $P3$, 计算转移概率的差值矩阵 $\Delta P1$ 、 $\Delta P2$, 如式(12)和式(13)所示:

$$\Delta P1 = P2 - P1 = \begin{pmatrix} \Delta p_{1,1} & \cdots & \Delta p_{1,31} \\ \vdots & \ddots & \vdots \\ \Delta p_{31,1} & \cdots & \Delta p_{31,31} \end{pmatrix} \quad (12)$$

$$\Delta P2 = P3 - P2 = \begin{pmatrix} \Delta p'_{1,1} & \cdots & \Delta p'_{1,31} \\ \vdots & \ddots & \vdots \\ \Delta p'_{31,1} & \cdots & \Delta p'_{31,31} \end{pmatrix} \quad (13)$$

由上述 100 首 $M1$ 、 $M2$ 和 $M3$ 音频计算得到的比例因子转移概率差值散点图如图 5 所示, 图 5(a) 为比例因子块内转移概率差值, 图 5(b) 为比例因子块间转移概率差值, 图中“+”代表 $M1$ 、 $M2$ 差值, “o”代表 $M2$ 、 $M3$ 差值。从图 5 中可以看出, $M1$ 、 $M2$ 转移概率差值大于 $M2$ 、 $M3$ 的差值, 且区别较为明显。

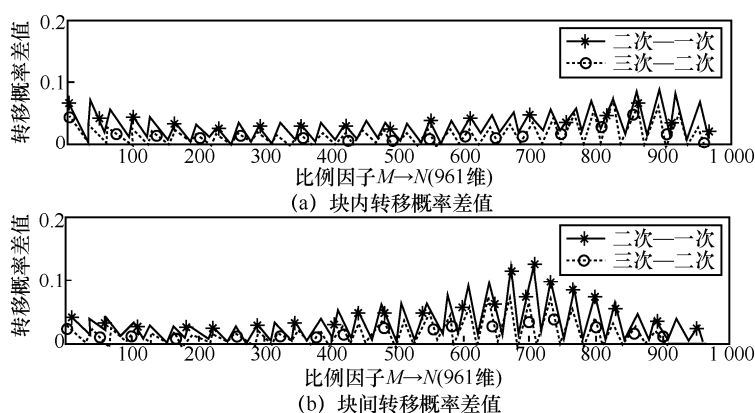


图5 压缩 AAC 音频的转移概率差值

4 实验结果与分析

4.1 实验设置

本文采用的数据库是随机选取的 800 首 10 s 长的 WAV 音频。这 800 首 WAV 音频包含各种不同的风格,如蓝调、流行、古典、乡村和民谣等。实验中使用的 AAC 编解码器为使用最为广泛的开源软件 FAAC-1.28 和 FAAD2-2.7。对上述数据库的 800 首 WAV 音频使用 FAAC-1.28 进行压缩获得单压缩的 AAC 音频,共 7 种比特率,60 kbit/s、75 kbit/s、90 kbit/s、120 kbit/s、135 kbit/s 和 150 kbit/s 各 800 首。而双压缩音频是单压音频在经 FAAD2-2.7 解码后再次使用 FAAC-1.28 编码得到的音频,共得到 28 类二次压缩音频。再对该二次压缩 AAC 音频样本解压重压缩,得到 28 类与二次压缩码率相同的三次压缩 AAC 音频样本。

本文选择 LIBSVM 作为分类器,并从上述一次、二次和三次压缩样本中随机选取 70% 用于训练模型,而剩下的 30% 作为测试模型。为了使测试的结果更加准确,表 1~表 4 中的检测准确率均是训练和测试重复 10 次结果的平均值。

为表述方便,将比例因子块内转移概率差值记为特征集 $SFTPD_{intra}$,块间转移概率差值记为特征集 $SFTPD_{inter}$ 。

4.2 检测结果

在本实验中,计算比例因子转移概率时,共

有 961 维特征,特征维数过高使得在用 SVM 进行分类时,其效果并不是很理想。因此,使用 SVM-RFE (recursive feature elimination based on SVM, 基于 SVM 的递归特征消除)^[15] 算法对选出的特征按照从优到劣进行排序。SVM-RFE 算法于 2002 年由 Guyon 等人提出。SVM-RFE 是在 RFE 的过程中使用 SVM 作为分类器,它从全集出发,根据所训练分类器的各项参数来制定特征排序的标准,逐个剔除最不相关的特征,并通过迭代直到剩下最后一个特征,从而实现特征的排序。

(1) 块内转移概率检测结果

表 1 是特征集 $SFTPD_{intra}$ 对 AAC 音频压缩历史的检测结果,其中 BR1、BR2 和 BR3 表示第一次、第二次和第三次压缩时所设置的码率,本文中仅考虑 BR2=BR3 的情况,而表 1 中的值则代表具体的检测准确率。以表中第 6 行第 8 列的值 (99.93%) 为例,该值表示本方法对先后使用 120 kbit/s 和 150 kbit/s 码率压缩 (FAAC/FAAD2) 的双压缩 AAC 音频的综合检测率达到了 99.93%。

由表 1 可知,低码率转高码率音频检测的平均正确率达到 99.88%,同码率间检测的平均正确率为 97%。该特征集在低码率转高码率的情况下取得了较好的检测效果,均超过了 99%。但针对相同码率的压缩音频,检测率相对偏低约 2.88 个百分点,这是由于在音频压缩码率相同时,其比例因子变化较小,转移概率差值变化也相对较小。

表 1 特征集 $SFTPD_{intra}$ 的检测准确率

BR1	BR2=BR3						
	60	75	90	105	120	135	150
60	100%	100%	100%	100%	100%	100%	100%
75		94.5%	100%	100%	100%	100%	100%
90			97.5%	99.5%	100%	100%	100%
105				96.5%	100%	100%	100%
120					95.83%	100%	99.93%
135						98%	99.17%
150							96.67%

(2) 块间转移概率检测结果

表 2 是特征集 $SFTPD_{inter}$ 对 AAC 音频压缩历史的检测结果。其低码率转高码率音频检测的平均正确率达 99.67%，同码率间检测的平均正确率为 96.17%。由检测结果可知，该方法对先后经过 120 kbit/s 和 150 kbit/s 的双压缩 AAC 音频检测准确率仅为 96.83%，比特征集 $SFTPD_{intra}$ 低约 3.1 个百分点，且与之前规律类似，利用特征集 $SFTPD_{intra}$ 的检测准确率在低码率转高码率上较高，在同码率间检测结果上表现一般。

表 2 特征集 $SFTPD_{inter}$ 的检测准确率

BR1	BR2=BR3						
	60	75	90	105	120	135	150
60	100%	100%	99.67%	100%	100%	100%	99.67%
75		96%	100%	99.83%	100%	99.83%	100%
90			96%	98.5%	99.83%	100%	100%
105				94.17%	100%	100%	100%
120					96.33%	99.83%	96.83%
135						95.5%	99%
150							95.17%

以上结果表明，特征集 $SFTPD_{intra}$ 与特征集 $SFTPD_{inter}$ 的检测准确率几乎相同，两个特征集对 AAC 压缩音频的检测准确率都较为理想。但是，由于在统计比例因子转移概率时，仅统计了在 150~180 内比例因子的转移情况，其特征集 $SFTPD$ 统计的范围并不是最为全面的，这些因素会使 AAC 压缩音频检测的准确率降低。

(3) 融合特征检测结果

本文选取块内、块间转移概率差值作为检测 AAC 压缩音频的特征，而块内、块间转移概率的差值特征均属于单一特征。为证明实验的可靠性，现将块内、块间的转移概率差值特征融合，即将块内、块间转移概率差值相加后，使用 SVM-RFE 对特征进行排序并选择，最终得到融合特征。

表 3 是融合特征对 AAC 音频压缩历史的检测结果，其低码率转高码率音频检测的平均正确率达 99.69%，同码率间检测的平均正确率为 97.67%。与特征集 $SFTPD_{intra}$ 、 $SFTPD_{inter}$ 对比发现，其检测准确率在低码率转高码率上相差不大，但在同码率间检测上有所提升，融合特征检测准确率比 $SFTPD_{intra}$ 高约 0.67 个百分点，比 $SFTPD_{inter}$ 高约 1.5 个百分点。总体而言，融合特征在 AAC 压缩音频分类准确率上表现较好。

表 3 融合特征的检测准确率

BR1	BR2=BR3						
	60	75	90	105	120	135	150
60	100%	99.83%	100%	100%	100%	100%	99.5%
75		96.83%	100%	100%	99.67%	99.83%	100%
90			98.67%	99.5%	100%	100%	97%
105				97.17%	100%	99.67%	100%
120					96.67%	99.83%	98.67%
135						96.67%	100%
150							97.67%

4.3 对比试验

为了更加全面地评估本文提出的 AAC 压缩音频的检测方法，重构了 Jin 等人^[10]方法的特征，并与其进行了对比。在参考文献[10]中，具体分类特征构建的方法是：Huffman 码表在 AAC 音频一次压缩和二次压缩的使用情况存在差异，将 Huffman 码表索引出现的概率作为第一特征，将

表 4 参考文献[10]与文中特征集 $SFTPD_{\text{intra}}$ 特征的双压缩检测结果比较

BR1	BR2=BR3						
	60	75	90	105	120	135	150
60	+22.83%	0	0	0	0	0	0
75		+15.92%	0	0	0	0	0
90			+21.75%	-0.33%	0	0	0
105				+22.87%	+1.5%	0	0
120					+16.66%	+2.33%	+0.03%
135						+15.92%	-0.26%
150							+11.17%

其 Markov 单步转移概率作为第二特征,将第一特征与第二特征融合得到第三特征,并用 LIBSVM 对单压缩和双压缩 AAC 音频进行分类,其低码率转高码率检测的平均正确率达 99.72%,同码率间检测的平均正确率为 77.38%。表 4 是参考文献[8]实验检测率与本文特征集 $SFTPD_{\text{intra}}$ 检测率对比,“+”代表本文算法结果高于参考文献[10]检测率的百分点,“-”代表本文算法结果低于参考文献[8]检测率的百分点。

上述结果表明, Jin 等人^[10]方法的特征对低码率转高码率的 AAC 音频其检测率较高,与本文算法中的比例因子转移概率差值的检测结果相差不大,但是本文算法解决了参考文献[10]中同码率间的检测效果较差的问题,将检测准确率提高了约 18 个百分点。

5 结束语

本文针对多次压缩的 AAC 音频,提出了一种基于比例因子转移概率差值统计特性的 AAC 音频压缩历史检测算法。通过研究压缩过程中比例因子的变化,利用其转移概率差值特征对多压缩 AAC 音频进行分类。实验结果表明,低码率转高码率间 AAC 音频检测准确率达 99.75%,而同码率间检测准确率也达 97.28%。虽然该算法在检测 AAC 音频的压缩历史时准确率较高,但也存在着一定的局限性。如未考虑噪声等其他因素攻击情况下算法的顽健性;仅针对一种编解码算法,未

涉及相同编码标准的不同编码算法的压缩历史检测算法。因此在今后的工作中将对上述问题进行更深入的研究。

参考文献:

- [1] 郑志彬. 信息网络安全威胁及技术发展趋势[J]. 电信科学, 2009, 25(2): 28-34.
ZHENG Z B. Overview of mobile communication services security[J]. Telecommunications Science, 2009, 25(2): 28-34.
- [2] 王帅, 汪来富, 金华敏, 等. 网络安全分析中的大数据技术应用[J]. 电信科学, 2015, 31(7): 145-150.
WANG S, WANG L F, JIN H M, et al. Big data application in network security analysis[J]. Telecommunications Science, 2015, 31(7): 145-150.
- [3] D'ALESSANDRO B, SHI Y Q. MP3 bit rate quality detection through frequency spectrum analysis[C]//ACM Workshop on Multimedia and Security, September 7-8, 2009, Princeton, USA. New York: ACM Press, 2009: 57-62.
- [4] LIU Q, SUNG A H, QIAO M. Detection of double MP3 compression[J]. Cognitive Computation, 2010, 2(4): 291-296.
- [5] QIAO M, SUNG A H, LIU Q. Improved detection of MP3 double compression using content-independent features[C]//IEEE International Conference on Signal Processing, Communication and Computing, Aug 5-8, 2013, Kunming, China. New Jersey: IEEE Press, 2013: 1-4.
- [6] LUO D, LUO W Q, YANG R, et al. Compression history identification for digital audio signal[C]//2012 IEEE International Conference on Acoustics, Speech and Signal Processing, March 25-30, 2012, Kyoto, Japan. New Jersey: IEEE Press, 2012: 1733-1736.
- [7] LUO D, LUO W Q, YANG R, et al. Identifying compression history of wave audio and its applications[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2014, 10(3): 1-19.
- [8] REN Y, FAN M, YE D, et al. Detection of double MP3 com-

- pression based on difference of calibration histogram[J]. Multimedia Tools & Applications, 2016, 75(21): 13855-13870.
- [9] SEICHTER D, CUCCOVILLO L, AICHROTH P. AAC encoding detection and bitrate estimation using a convolutional neural network[C]//IEEE International Conference on Acoustics, Speech and Signal Processing, March 20-25, 2016, Shanghai, China. New Jersey: IEEE Press, 2016: 2069-2073.
- [10] JIN C, WANG R, YAN D, et al. An efficient algorithm for double compressed AAC audio detection[J]. Multimedia Tools & Applications, 2016, 75(8): 4815-4832.
- [11] 李晗. AMR 和 AAC 音频双压缩检测研究[D]. 广州: 华南理工大学, 2015.
- LI H. Research on AMR and AAC audio dual compression detection[D]. Guangzhou: South China University of Technology, 2015.
- [12] INSTITUTION B S. ISO/IEC 13838-7/FPDAM 1. Information technology. Generic coding of moving pictures and associated audio information. Part 7: advanced audio coding (AAC). Amendment 1: signalling of bandwidth extension[S].
- [13] Wikipedia. Advanced audio coding[EB/OL]. (2011-06-15) [2017-07-13]. https://de.wikipedia.org/wiki/Advanced_Audio_Coding.
- [14] GAO Z H, WEI G. The core technology of the broadband MP3 audio compression[J]. Electro Acoustic Technology, 2000, 9(5): 9-13.
- [15] WESTON J, GUYON I. Support vector machine-recursive feature elimination (SVM-RFE): US, US 8095483 B2[P]. 2012.

[作者简介]



黄其娟 (1993-), 女, 宁波大学信息科学与工程学院硕士生, 主要研究方向为多媒体通信与信息安全等。



王让定 (1962-), 男, 博士, 宁波大学高等技术研究院教授、博士生导师, 主要研究方向为多媒体通信与取证、信息隐藏与隐写分析、智能抄表及传感网络技术。



严迪群 (1979-), 男, 博士, 宁波大学信息科学与工程学院副教授、硕士生导师, 主要研究方向为多媒体通信、信息安全、基于深度学习的数字语音取证等。



基于 SVD 和 SAE 的医学图像多功能零水印算法

孙洁丽¹, 赵彦霞^{1,2}, 周增慧³, 邢胜⁴

(1. 河北经贸大学信息技术学院, 河北 石家庄 050061;

2. 河北大学管理学院, 河北 保定 071002;

3. 河北经贸大学教务处, 河北 石家庄 050061;

4. 沧州师范学院计算机科学与工程学院, 河北 沧州 061001)

摘要: 针对医学图像的版权保护和篡改定位问题, 提出了一种多功能零水印算法。获取构造零顽健水印和半脆弱水印的信息矩阵, 并分别与原始顽健和半脆弱水印进行异或运算, 构造零顽健和零半脆弱水印图像。实验结果证明, 提取的水印效果较好, 算法能很好地抵抗常见攻击, 特别是抵抗强度较大的攻击的效果较好。

关键词: 医学图像; 多功能零水印; 离散小波变换; 奇异值分解; 堆叠自编码器

中图分类号: TN911.7

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017254

Multipurpose zero watermarking algorithm for medical image based on SVD and SAE

SUN Jieli¹, ZHAO Yanxia^{1,2}, ZHOU Zenghui³, XING Sheng⁴

1. College of Information and Technology, Hebei University of Economics and Business, Shijiazhuang 050061, China

2. College of Management, Hebei University, Baoding 071002, China

3. Educational Administration Office, Hebei University of Economics and Business, Shijiazhuang 050061, China

4. College of Computer Science and Engineering, Cangzhou Normal University, Cangzhou 061001, China

Abstract: A multipurpose zero watermarking algorithm was proposed according to the copyright protection and tamper location of medical image. The two information matrices were executed the XOR operation with the original robust watermarking image and the original semi-fragile watermarking image separately to generate zero robust watermarking image and zero semi-fragile watermarking image. The experimental results show that effects of extracted zero watermarking image are terrific, and the algorithm can resist the common attacks, particularly its resistance is effective for strong attacks.

Key words: medical image, multipurpose zero watermarking, discrete wavelet transform, singular value decomposition, stacked auto-encoder

收稿日期: 2017-06-27; 修回日期: 2017-08-05

基金项目: 国家自然科学基金资助项目 (No.71371063, No.61672205); 河北省应用基础研究计划重点基础研究项目 (No.16960314D); 河北省人力资源社会保障科研合作课题 (No.JRSHZ-2016-07038)

Foundation Items: The National Natural Science Foundation of China (No.71371063, No.61672205), The Key Basic Research Project of Applied Basic Research Program of Hebei Province (No.16960314D), The Science and Technology Cooperation Project of Hebei Provincial Department of Human Resources and Social Security (No.JRSHZ-2016-07038)

1 引言

医院每天都产生大量的医学图像等数字产品,对这些数字产品的保护也是对医院、医生和患者三方权益的保护。数字水印^[1]能够对数字产品进行版权保护和篡改定位,较好地保护这些产品。常用的数字水印是向原始数字产品嵌入数字水印来达到保护产品的目的,但提出的算法都会对数字产品进行修改,零水印技术^[2]被提出后,不再需要向数字产品中嵌入水印,而是提取原始数字产品的特征与数字水印一起构造零水印。目前,研究者已经提出了大量文本^[3,4]、音频^[5]、视频^[6,7]、图像^[8,9]等数字产品的零水印算法。

医学图像是医生给病人诊病的主要依据之一,当出现医患纠纷时,也可以作为医生治疗是否造成医疗事故的原始证据之一,构造顽健水印可以对医学图像进行版权保护,从而避免出现医学图像真假无法判定的问题。当有人将病灶部分修改或将病灶添加进医学图像时,希望能够通过技术手段检测到,因此,有必要构造半脆弱水印对被修改的医学图像进行篡改定位。另外,现代信息技术的快速发展,也为医疗机构之间通过医学资源共享提高诊病效率提供了可能,这也需要对医学图像进行版权保护和篡改定位。

对医学图像进行保护的水印算法研究已经成为一个研究热点,研究者提出了多种零数字水印算法。Seenivasagam 等人^[10]提出的零水印算法,利用混合 CT-SVD (contourlet transform-singular value decomposition) 域提取特征,将 QR 作为顽健水印,并利用胡氏不变矩构造零水印保证云端医学图像的安全性。Vellaisamy 等人^[11]提出了基于混合 CT-SVD 域的零水印算法,利用三角数生成函数和胡氏图像不变量抵抗反转攻击。Han 等人^[12]将三维医学图像进行三维离散小波和离散余弦变换,利用感知散列技术获取其本身特征向量,并利用 Legendre 混沌神经网络生成的混沌序列置乱

此特征向量。Dong 等人^[13]提取 DWT-DFT (discrete wavelet transform-discrete Fourier transform) 低频系数作为加密医学图像的特征向量,构造了顽健水印图像。但是,参考文献[10-13]都只构造了顽健水印对医学图像进行版权保护,没有对医学图像进行篡改定位。肖振久等人^[14]提出了利用超混沌技术的医学图像篡改定位零水印算法,但算法抵抗噪声攻击的能力还需要加强。

要构造高质量水印图像,需要获取能够代表原始图像的重要特征,并且构造的水印图像应该能够抵抗对原始图像的各种攻击。由于一般情况下,基于变换域构造的顽健水印比基于空域构造的水印的顽健性更强。奇异值分解 (singular value decomposition, SVD) 的稳定性好。属于深度神经网络的堆叠自编码器 (stacked auto-encoder, SAE) 能够获取图像的关键特征,并且还能够降低图像的维度,因此,提出的算法将 DWT、SVD 和 SAE 结合起来,构造了顽健水印和半脆弱水印,对医学图像进行版权保护和篡改定位。

2 相关技术

2.1 奇异值分解

$B \in C_r^{m \times n}$ ($r > 0$) 的大小为 $m \times n$, $m \geq n$, 则 B 的奇异值分解^[15]为式 (1)^[16]:

$$B = USV^T \quad (1)$$

其中, U 、 V 为正交矩阵, S 矩阵除对角线元素外,其他元素都为 0。

2.2 堆叠自编码器

堆叠自编码器^[17]是一个前馈神经网络,是由自动编码器 (auto-encoder, AE)^[18]堆叠而成的深度网络^[19]。堆叠自编码网络的训练分两步:自下而上的无监督学习,采用无标签样本,利用逐层贪婪学习,逐层对 AE 进行预训练;自上而下的有监督学习,对预训练得到的权重、偏置进行微调。

自动编码器是一个试图使输出与输入相同的简单神经网络。堆叠多个自动编码器可以构成深度网络。自动编码器的工作原理如图 1 所示。

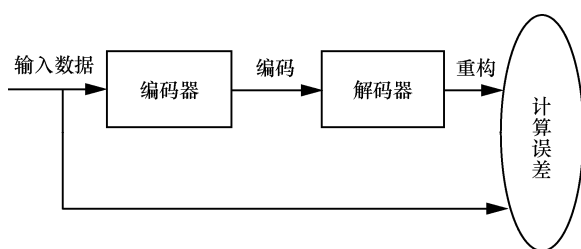


图1 自动编码器的工作原理

图1中，在网络训练期间，对每个训练样本进行训练后，会在输出层产生一个新的信号，网络学习的目的就是使输出信号与输入信号尽量相同。一个自动编码器的工作过程可分为3步：输入数据 x ，学习得到一个函数 f ，将输入数据编码成 $h=f(x)$ ；学习得到一个函数 g ，重构 h 为 $r=g(h)=g(f(x))$ ；利用损失函数 $L(r,x)$ 计算重构的 r 与输入数据 x 的误差，然后通过反向微调，最小化误差。一般情况下，中间层神经元数量远远少于输入层，在训练过程中得到的 h 的维度一般远远小于 x 的维度。如果将自动编码器与分类器结合，将 h 作为分类器的输入，将大大减少处理数据的信息量，从而大大提高计算效率。

3 基于SVD和SAE的医学图像多功能零水印算法

3.1 零水印算法构造流程

构造零水印图像的过程如图2所示，具体介绍如下。

步骤1 读出、转换和分解原始彩色图像。

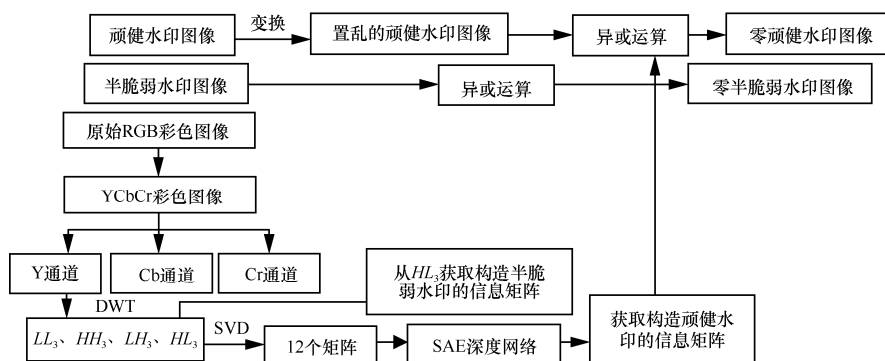


图2 零水印图像算法构造流程

读出原始图像，转换为YCbCr图像，分解出Y、Cb、Cr通道。

步骤2 读出和置乱顽健水印图像。读出顽健水印图像，利用Arnold变换对其进行置乱。设顽健水印图像尺寸大小为 $n_1 \times n_1$ 。

步骤3 读出半脆弱水印图像。设半脆弱水印图像尺寸大小为 $m_2 \times n_2$ 。

步骤4 对原始图像的Y通道进行小波分解和奇异值分解。利用三尺度的“db1”离散小波变换对原始图像的Y通道进行分解，然后再进行奇异值分解。Y通道小波变换后得到的第三层低频子带小波系数矩阵、对角方向、垂直方向和水平方向高频子带小波系数矩阵分别为 LL_3 、 HH_3 、 LH_3 、 HL_3 ，对这4个小波系数矩阵进行奇异值分解，得到12个矩阵。

步骤5 构造SAE深度网络的输入矩阵和输出标签矩阵。将步骤4中得到的12个矩阵的尺寸转换为偶数，将其最大值赋值给 m ，创建 $m \times m \times 12$ 的特征矩阵 l ，设其每个面为 $X_n(n=1,2,\dots,12)$ ，分别将步骤4中得到的12个矩阵，赋值给 $(l-1)$ ，就得到了输入特征矩阵。建立一个尺寸为 $8 \times 8 \times 12$ 的矩阵 Y_0 和一个尺寸为 $1 \times 8 \times 12$ 的矩阵 Y ，将 l 每个面的前8行×前8列矩阵赋值给矩阵 Y_0 每个面，将矩阵 Y_0 每个面的每列的和除以每个面元素总个数，所得的值赋值给 Y 的每个面。分别将 l 和 Y 转换成 $12 \times m \times m$ 和 $12 \times 1 \times 8$ 的二维矩阵，则得到的 l 和 Y 分别是SAE的输入矩阵和输出标签矩阵。

步骤 6 建立如图 3 所示的 SAE 深度网络。

图 3 所示为算法中用到的 SAE 深度网络结构, 由输入层 (V_0)、隐含层 1 (H_0 或 V_1)、隐含层 2 (H_1 或 V_2)、隐含层 3 (H_2) 和输出层构成, 各层之间神经元采用全连接, 本层神经元之间没有连接。输入层神经元个数为 $(m \times m)$, 设隐含层 1、隐含层 2、隐含层 3 的神经元个数分别为 N_1 、 N_2 、 N_3 , 隐含层 1、隐含层 2、隐含层 3 的神经元个数可根据情况确定, 但是需要满足 $(m \times m) > N_1 > N_2 > N_3$ 且 $(12 \times (N_3 + 1)) > (n_1 \times n_1)$, 输出层的神经元个数为 8。

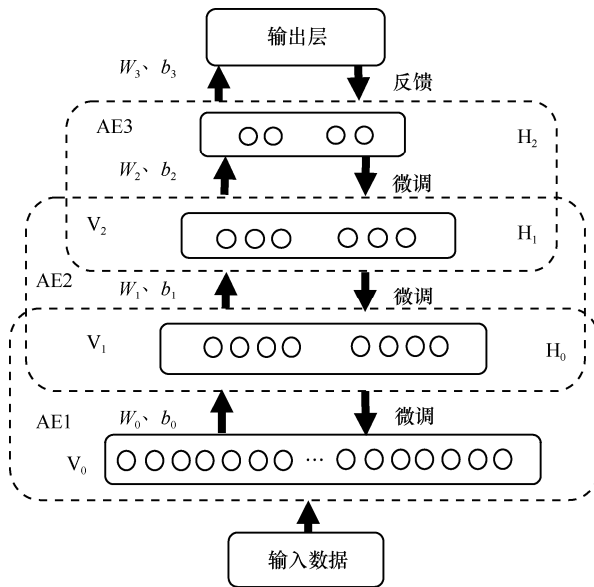


图 3 堆叠自编码器深度网络结构示意图

步骤 7 对 SAE 深度网络进行训练。将步骤 5 中产生的 SAE 的输入矩阵 x 作为堆叠自编码器隐含层 1 的输入值, 然后对 AE1、AE2 和 AE3 这 3 个自编码器进行预训练。

SAE 网络的训练分两步。

首先对 AE1、AE2 和 AE3 进行预训练。采用逐层贪婪学习方法训练每个自编码器, AE1 训练完成后, 其解码器的输出值 H_0 为 AE2 的输入值, AE2 解码器的输出值 H_1 为 AE3 的输入值。自编码器的工作示意如图 1 所示, 单个自编码器训练时, 需要经过编码和解码两个阶段。

堆叠自编码器 l 层第 k 个样本编码后的输出为 $y_k^{(l)}$ [17]:

$$y_k^{(l)} = \sigma(W_k^{(l)} x_k^{(l)} + b_k^{(l)}) \quad (2)$$

其中, $\sigma(\cdot)$ 为 Sigmoid 函数, $x_k^{(l)}$ 、 $y_k^{(l)}$ 、 $W_k^{(l)}$ 、 $b_k^{(l)}$ 分别为第 l 层第 k 个样本对应的编码器的输入、输出、连接权重和偏置。

堆叠自编码器 $(l+1)$ 层第 k 个样本解码重构后的输出为 $\hat{y}_k^{(l+1)}$ [17]:

$$\hat{y}_k^{(l+1)} = \sigma((W_k^{(l)})' y_k^{(l)} + b_k^{(l+1)}) \quad (3)$$

其中, $y_k^{(l)}$ 、 $\hat{y}_k^{(l+1)}$ 、 $(W_k^{(l)})'$ 、 $b_k^{(l+1)}$ 分别为第 $(l+1)$ 层第 k 个样本对应的解码器的输入、输出、连接权重和偏置。

再对预训练得到的权重、偏置进行微调。

将训练好的 3 个自编码器 AE1、AE2 和 AE3 按照堆叠的方式连接起来, 在 AE3 的 H2 之后加一层 softmax 分类器, 构成如图 3 所示的 SAE 深度网络, 根据输出误差, 利用反向传播算法对整个网络进行微调。反向传播算法的步骤为: 首先, 进行前向传导计算; 其次, 从输出层开始, 逐层向前传递误差, 对连接权重和偏置进行微调。

假设网络的层数为 N , 第 l 层的单元数量 (不含偏置) 用 s_l 表示。输入层第 i 个输入值 x_i 用 $d_i^{(l)}$ 表示, l 层的第 i 个单元的输用 $d_i^{(l+1)}$ 表示, $W_{ij}^{(l)}$ 表示 l 层第 j 个单元与 $(l+1)$ 层第 i 个单元的连接权重, $b_i^{(l)}$ 表示第 $(l+1)$ 层第 i 个单元的偏置值, $z_i^{(l)}$ 表示 l 层第 i 个单元输入加权和, $d_i^{(l)} = \sigma(z_i^{(l)})$, 其中, $\sigma(\cdot)$ 为 Sigmoid 函数。 $h_{w,b}(x)$ 为输出层的输出。

在前向传导阶段, SAE 第 $(l+1)$ 层的 $z_i^{(l+1)}$ 和 $d_i^{(l+1)}$ [20] 为:

$$z_i^{(l+1)} = \sum_{j=1}^{s_l} (W_{ij}^{(l)} d_j^{(l)} + b_i^{(l)}) \quad (4)$$

$$d_i^{(l+1)} = \sigma(z_i^{(l+1)}) \quad (5)$$

用向量表示各层数据后, SAE 第 $(l+1)$ 层的 $z^{(l+1)}$ 和 $d^{(l+1)}$ [20] 为:



$$z^{(l+1)} = W^{(l)} d^{(l)} + b^{(l)} \quad (6)$$

$$d^{(l+1)} = \sigma(z^{(l+1)}) \quad (7)$$

m 个训练样本的损失函数^[20]为:

$$J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \| h_{W, b}(x^{(i)}) - y^{(i)} \|^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^{N-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \quad (8)$$

其中, $x^{(i)}$ 、 $y^{(i)}$ 分别为第 i 个训练样本的输入和输出。

在反向传播阶段, m 个样本损失函数的偏导为:

$$\frac{\partial}{\partial W_{ij}^{(l)}} J(W, b) = \left[\frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b; x^{(i)}, y^{(i)}) \right] + \lambda W_{ij}^{(l)} \quad (9)$$

$$\frac{\partial}{\partial b_i^{(l)}} J(W, b) = \frac{1}{m} \sum_{i=1}^m \frac{\partial}{\partial b_i^{(l)}} J(W, b; x^{(i)}, y^{(i)}) \quad (10)$$

用向量表示的输出层 (N) 的残差 $\delta^{(N)}$ 的计算式为:

$$\delta^{(N)} = -(y - d^{(N)}) \cdot \sigma'(z^{(N)}) \quad (11)$$

$l = N-1, N-2, \dots, 2$ 层残差的计算式为:

$$\delta^{(l)} = ((W^{(l)})^T \delta^{(l+1)}) \cdot \sigma'(z^{(l)}) \quad (12)$$

偏导的计算式为:

$$\nabla_{W^{(l)}} J(W, b; x, y) = \delta^{(l+1)} (d^{(l)})^T \quad (13)$$

$$\nabla_{b^{(l)}} J(W, b; x, y) = \delta^{(l+1)} \quad (14)$$

l 层权值改变量 $\Delta W^{(l)}$ 和偏置改变量 $\Delta b^{(l)}$ 的计算式为:

$$\Delta W^{(l)} := \Delta W^{(l)} + \nabla_{W^{(l)}} J(W, b; x, y) \quad (15)$$

$$\Delta b^{(l)} := \Delta b^{(l)} + \nabla_{b^{(l)}} J(W, b; x, y) \quad (16)$$

权值更新计算式为:

$$W^{(l)} := W^{(l)} - \alpha \left[\frac{1}{m} \Delta W^{(l)} + \lambda W^{(l)} \right] \quad (17)$$

$$b^{(l)} := b^{(l)} - \alpha \left[\frac{1}{m} \Delta b^{(l)} \right] \quad (18)$$

其中, $\alpha = 1$, $\lambda = 0.5$ 。

步骤 8 获取用于构造顽健水印的原始图像

信息矩阵。取出输出层的输入矩阵 A , 创建一个尺寸为 $n_1 \times n_1$ 的元素值为 0 的矩阵 B , 从矩阵 A 隔 1 列取一个元素作为矩阵 B 的元素, 每取 n_1 个元素后, 矩阵 B 的行号加 1。创建尺寸为 $n_1 \times n_1$ 的元素值为 0 的矩阵 C 。当矩阵 B 的元素满足 $B(i, j) \leq B((i+1), j) \leq B((i+2), j)$ 时, 矩阵 C 的元素为 1, 否则为 0, 得到的矩阵 C 为用于构造零顽健水印的信息矩阵。

步骤 9 构造零顽健水印图像。将步骤 8 获得的矩阵 C 和置乱的原始顽健水印进行异或运算, 将产生的矩阵转换成像素值只有 0 和 255 的零顽健水印图像。

步骤 10 获取用于构造半脆弱水印的原始图像信息矩阵。利用 Y 通道离散小波变换后得到的水平方向系数矩阵 HL_3 得到信息矩阵, 当 $HL_3((i+2), j) \leq HL_3((i+1), j) \leq HL_3(i, j)$ 时, 信息矩阵的元素为 1, 否则为 0, 得到与半脆弱水印尺寸相等的信息矩阵。

步骤 11 构造零半脆弱水印图像。将步骤 10 得到的信息矩阵和原始半脆弱水印进行异或运算, 将产生的矩阵元素转换成像素灰度值只有 0 和 255 的零半脆弱水印图像。

3.2 零水印算法提取流程

提取零水印图像的过程如图 4 所示, 具体介绍如下。

步骤 1 读出零顽健水印图像。

步骤 2 读出零半脆弱水印图像。

步骤 3 读出待检测图像, 转换成 YCbCr 图像, 并分解出 Y*、Cb*、Cr* 通道。

步骤 4 对待检测图像的 Y* 通道进行离散小波变换, 并进行奇异值分解。对待检测图像的 Y* 通道进行三尺度的“db1”离散小波变换, Y* 通道小波变换后得到的第三层的低频子带小波系数矩阵、对角方向、垂直方向和水平方向高频子带小波系数矩阵分别为 LL_3^* 、 HH_3^* 、 LH_3^* 、 HL_3^* , 对这 4 个小波系数矩阵进行奇异值分解, 得到 12 个矩阵。

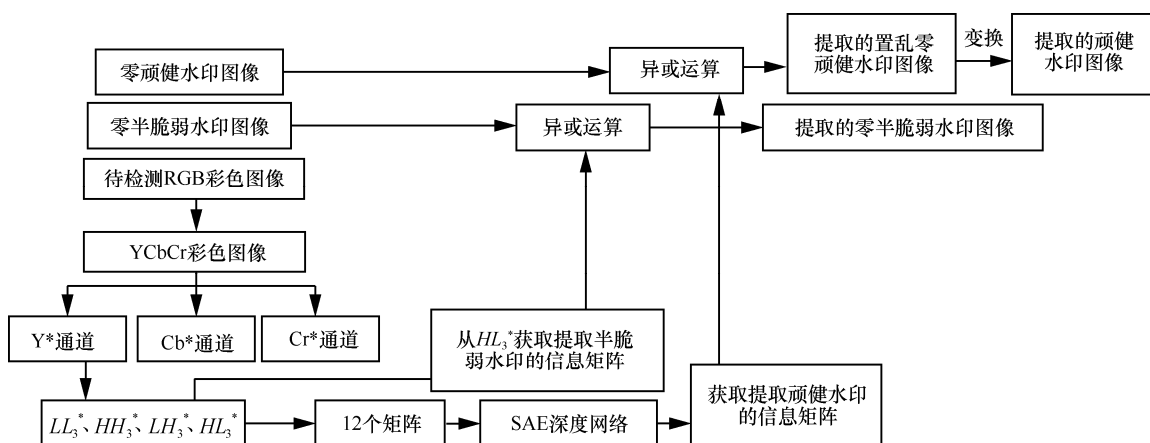


图4 零水印图像算法提取流程

步骤5 构造SAE深度网络的输入矩阵和输出标签矩阵。过程同第3.1节的步骤5。

步骤6 建立如图4所示的SAE深度网络。

步骤7 对SAE深度网络进行训练。训练过程同第3.1节的步骤7。

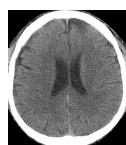
步骤8 构造用于提取顽健水印图像的信息矩阵。过程同第3.1节的步骤8。

步骤9 提取顽健水印图像。将零顽健水印和第3.2节的步骤8得到的信息矩阵进行异或运算，提取出像素值为0和1的置乱顽健水印图像，转换成像素值为0和255的图像，对其进行Arnold变换，则获得提取的顽健水印图像。

步骤10 获取用来提取半脆弱水印图像的信息矩阵。过程同第3.1节的步骤10。

步骤11 提取半脆弱水印图像。将零半脆弱水印和第3.2节的步骤10得到的信息矩阵进行异或运算，则获得提取的半脆弱水印图像。

Arnold变换的次数为60次。第3.1节的步骤6的 $m \times m = 52 \times 52 = 2704$, $N_1 = 1000$, $N_2 = 500$, $N_3 = 205$, 输出层神经元个数为8。所用的原始图像和水印图像如图5所示。



(a) 原始图像

M

(b) 顽健水印图像

WA

(c) 半脆弱水印图像

图5 原始图像和水印图像

4.1.1 零水印图像构造和提取实验

构造的零顽健和零半脆弱水印图像如图6所示。



(a) 零顽健水印图像



(b) 零半脆弱水印图像

图6 零顽健水印图像和零半脆弱水印图像

4 仿真实验和结果分析

4.1 仿真实验

算法采用MATLAB仿真实现。原始图像为 $408 \times 354 \times 3$ （单位：像素）的CT图像，顽健水印和半脆弱水印图像均采用自制图像，图像的像素灰度值只有0和255，尺寸分别为 36×36 （单位：像素）和 51×45 （单位：像素），顽健水印进行

分别用归一化互相关（normalised cross-correlation, NC）和篡改评估函数（tamper assessment function, TAF）表示提取的水印图像与原始水印图像的相似度和不相似度。在原始图像未受攻击时，提取的水印图像和半脆弱水印图像篡改定位图像如图7所示。提取的顽健水印图像和半脆弱水印图像的NC都为1，提取的半脆弱水印图像的TAF=0。

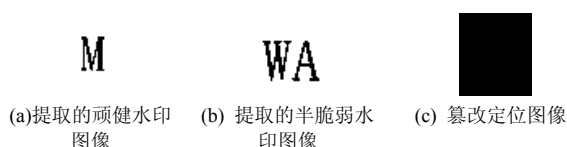


图7 提取的水印图像和篡改定位图像

4.1.2 原始图像受攻击实验

实验中,对待检测图像进行了多种攻击。对待检测图像进行剪切比例为 27.972 3%的剪切攻击后的图像、提取出的水印图像和半脆弱水印图像的篡改定位图像如图 8 所示。图 8 中,提取的顽健水印的 $NC=0.9890$, 半脆弱水印的 $NC=0.8726$, 半脆弱水印的 $TAF=0.2092$ 。

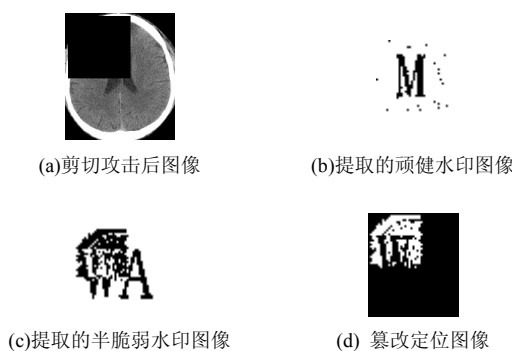


图8 剪切攻击后的图像、提取出的水印图像及篡改定位图像

如图 9~图 16 所示,分别为原始图像受到剪切、拼贴、中值滤波、旋转、JPEG 压缩、高斯噪声、椒盐噪声和缩放等常见攻击实验后,提取的顽健水印图像 NC、半脆弱水印图像 NC 和半脆弱水印图像 TAF 值的测量值。

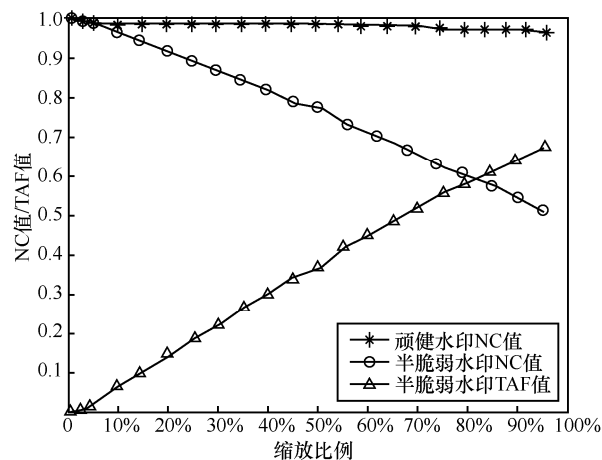


图9 剪切攻击实验结果

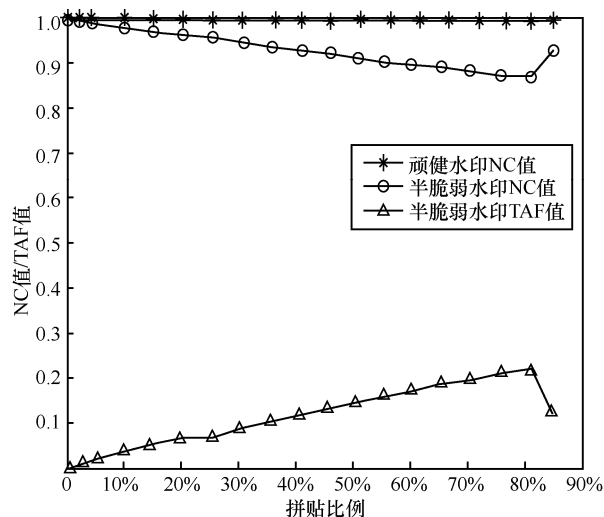


图10 拼贴攻击实验结果

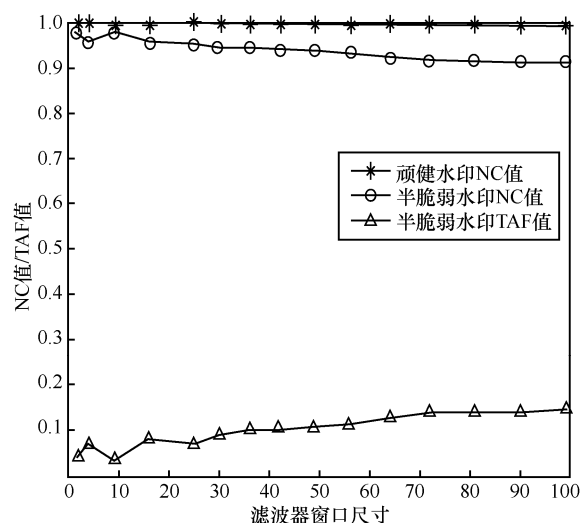


图11 中值滤波攻击实验结果

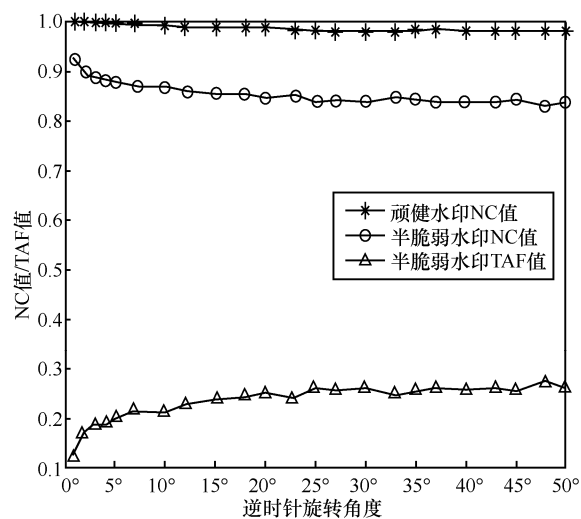


图12 旋转攻击实验结果

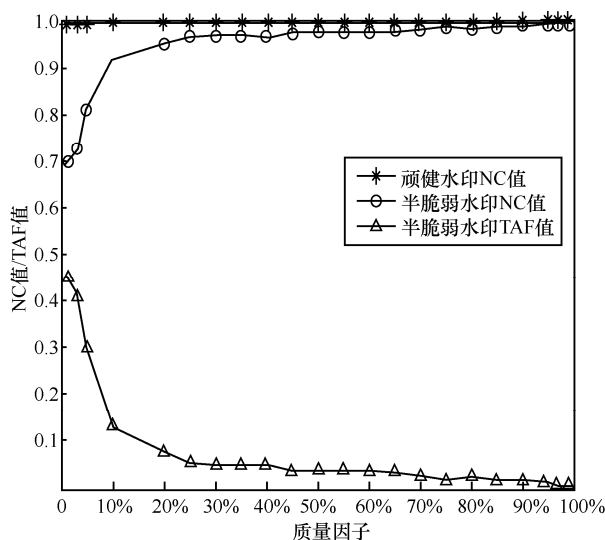


图13 JPEG压缩攻击实验结果

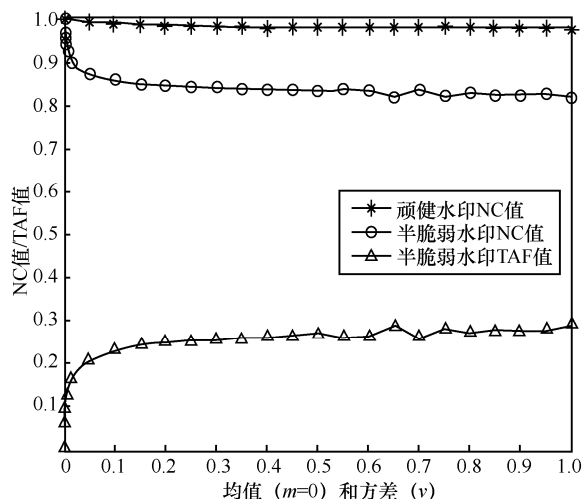


图14 高斯噪声攻击实验结果

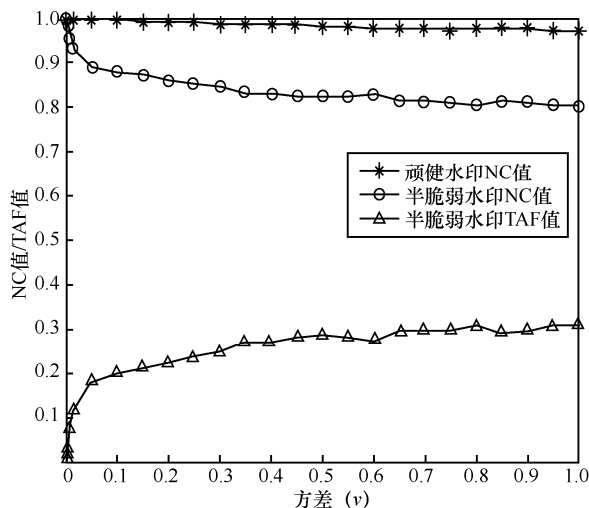


图15 椒盐噪声攻击实验结果

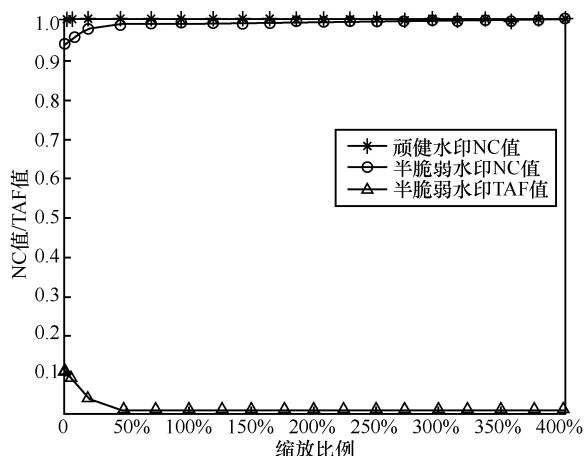


图16 缩放攻击实验结果

4.2 实验结果及分析

由实验结果可看出,本算法在原始图像未受攻击时,提取的顽健水印和半脆弱水印图像的NC值都为1。在原始图像受到多种攻击时,提取的顽健水印图像NC值在0.9649以上,说明提取的水印有较好的顽健性。另外,半脆弱水印也能够很好地对多种攻击进行篡改定位。

对于RGB彩色图像,亮度信息代表图像的主要信息,由于彩色图像的Y通道代表原始图像的亮度信息,Cb、Cr通道代表原始图像的色度信息。为了减少计算的数据量,提出的算法选择在Y通道中提取原始图像关键特征信息。在构造零顽健水印时,考虑到SVD能代表图像的代数特征,在算法中采用了SVD,又进一步利用SAE能够提取图像关键特征的优点,获取原始图像的关键特征。另外,DWT的应用也可以增加顽健水印的顽健性。在构造半脆弱水印时,选择了DWT得到的水平方向高频子带小波系数矩阵 HL_3 构造半脆弱水印,也能很好地对受到攻击后的图像进行篡改定位。

许多水印算法中仿真实验使用的图像,对应的矩阵都是方阵,这里实验所用的CT图像不是方阵,是医院真实的图像。实验结果表明,所提出的算法对于真实的医学图像具有很好的效果,具有很好的实用价值。



5 结束语

提出了一种基于 SVD 和 SAE 的医学图像多功能零水印算法。实验结果表明, 顽健水印的顽健性很好, 半脆弱水印能够很好地对篡改进行定位, 算法能够抵抗多种常见攻击。提出的算法, 创新性地将 SAE 能够较好提取图像特征的优势应用到构造顽健水印的过程中, 使构造的顽健水印对多种攻击, 特别是强度大的攻击都有很好的顽健性, 研究成果可以应用于证书、电子签字等多种数字产品的保护中, 有很好的实用价值。进一步对算法进行研究, 结合领域的实际情况, 将算法推广应用于更多的领域, 可以保证这些领域数字产品的安全, 对这些领域的发展做出贡献。

参考文献:

- [1] TIRKEL A Z, RANKIN G A, SCHYNDEL R G, et al. Electronic watermark[C]//Digital Image Computing, Technology and Applications(DICTA 1993), Dec 8-10,1993, Sidney, Australia. Sydney: Australian Pattern Recognition Society, 1993: 666-673.
- [2] 温泉, 孙铤锋, 王树勋. 零水印的概念与应用[J]. 电子学报, 2003, 31(2): 214-216.
WEN Q, SUN T F, WANG S X. Concept and application of zero-watermark[J]. Acta Electronica Sinica, 2003, 31(2): 214-216.
- [3] ZHU P, XIANG G L, SONG W N, et al. A text zero-watermarking algorithm based on Chinese phonetic alphabets[J]. Wuhan University Journal of Natural Sciences. 2016(4): 277-282.
- [4] 周新民, 鲁芳. 基于汉字结构知识的文本认证零水印算法[J]. 吉林大学学报(理学版), 2016, 54(5): 1071-1076.
ZHOU X M, LU F. Zero-watermarking algorithm for text authentication based on structure knowledge of Chinese characters[J]. Journal of Jilin University (Science Edition), 2016, 54(5): 1071-1076.
- [5] LEI M, YANG Y, LIU X M, et al. Audio zero-watermark scheme based on discrete cosine transform-discrete wavelet transform-singular value decomposition[J]. China Communications, 2016, 13(7): 117-121.
- [6] 王静, 郁梅, 李文峰, 等. 抗量化转码的 HEVC 视频流零水印算法[J]. 光电工程, 2016, 43(10): 77-83.
- [7] WANG J, YU M, LI W F, et al. HEVC-based video stream zero-watermarking algorithm with robustness to re-quantization transcoding[J]. Opto-Electronic Engineering, 2016, 43(10): 77-83.
- [7] XU Q S, LU J F, PENG X X, et al. A video zero-watermarking algorithm based on text detection[C]//2015 16th IEEE International Conference on Communication Technology(ICCT2015), Oct 18-20, 2015, Hangzhou, China. New Jersey: IEEE Press, 2015: 328-333.
- [8] 徐莲, 李京兵. 基于 DWT-PCA 和感知哈希的鲁棒零水印算法[J]. 电视技术, 2016, 40(1): 25-30.
XU L, LI J B. Robust zero-watermarking algorithm based on DWT-PCA and perceptual hashing[J]. Video Engineering, 2016, 40(1): 25-30.
- [9] LI D, LIU Z, CUI L H. A zero-watermark scheme for identification photos based on QR code and visual cryptography[J]. International Journal of Security and its Applications, 2016, 10(1): 203-214.
- [10] SEENIVASAGAM V, VELUMANI R. A QR code based zero-watermarking scheme for authentication of medical images in teleradiology cloud[J]. Computational & Mathematical Methods in Medicine, 2013(4): 1-16.
- [11] VELLASAMY S, RAMESH V. Inversion attack resilient zero-watermarking scheme for medical image authentication[J]. IET Image Processing, 2014, 8(12): 718-727.
- [12] HAN B R, CAI L S, LI W F. Zero-watermarking algorithm for medical volume data based on legendre chaotic neural network and perceptual hashing[J]. International Journal of Grid Distribution Computing, 2015, 8(1): 201-212.
- [13] DONG J T, LI J B. A robust zero-watermarking algorithm for encrypted medical images in the DWT-DFT encrypted domain[J]. Innovation in Medicine and Healthcare, 2016, 60(4): 197-208.
- [14] 肖振久, 李南, 王永滨, 等. 基于超混沌的医学图像篡改定位零水印算法[J]. 计算机工程与应用, 2015, 53(7): 115-120.
XIAO Z J, LI N, WANG Y B, et al. Zero watermarking scheme for medical image temper location based on hyper-chaos encryption[J]. Computer Engineering and Applications, 2015, 53(7): 115-120.
- [15] 薛栋, 周亚训, 金炜. 版权保护和内容认证的全盲双功能数字水印算法[J]. 电信科学, 2017, 33(2): 79-89.
XUE D, ZHOU Y X, JIN W. Bifunctional watermarking for copyright protection and content authentication with perfectly blind detection[J]. Telecommunications Science, 2017, 33(2): 79-89.
- [16] GHAZY R A, ABBAS A M, AI-ZUBI N, et al. Block-based

- SVD image watermarking in spatial and transform domains[J]. International Journal of Electronics, 2015, 102(7): 1091-1113.
- [17] BENGIO Y, LAMBLIN P, POPOVICI D, et al. Greedy layer-wise training of deep networks[C]//Advances in Neural Information Processing Systems 19(NIPS2006), Dec 4-7, 2006, Vancouver, Canada. Cambridge: MIT Press, 2007: 153-160.
- [18] BENGIO Y. Learning deep architectures for AI[J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-127.
- [19] YANN L C, YOSHUA B, GEOFFREY H. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [20] NG A. Sparse autoencoder[EB/OL]. [2017-06-18]. <https://www.mendeley.com/research-papers/sparse-autoencoder/>.

[作者简介]



孙洁丽（1969-），女，博士，河北经贸大学教授，主要研究方向为信息处理、数据挖掘。



赵彦霞（1970-），女，河北大学管理学院博士生，河北经贸大学讲师，主要研究方向为数字水印、不确定知识管理与机器学习、信息处理。



周增慧（1975-），男，河北经贸大学工程师，主要研究方向为信息管理、数据挖掘。



邢胜（1982-），男，博士，沧州师范学院讲师，主要研究方向为不确定知识管理与机器学习。



5G 关键技术与标准综述

王庆扬, 谢沛荣, 熊尚坤, 魏垚, 刘昱, 李文苾, 吴锦莲
(中国电信股份有限公司广州研究院, 广东 广州 510630)

摘要: 在全球业界的大力推动下, 5G 技术快速发展, 当前已进入标准制定与技术验证的关键阶段。基于此, 主要介绍了 5G 关键技术及其最新标准化进展, 包括大规模天线、新型无线网络架构、超密集组网、边缘计算、网络切片、网络按需定制、4G 和 5G 互操作等以及行业组织和运营商动态, 然后提出了一些思考和建议。

关键词: 5G; 大规模天线; 无线网络架构; 超密集组网; 边缘计算; 网络切片; 服务化架构

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017312

Key technology and standardization progress for 5G

WANG Qingyang, XIE Peirong, XIONG Shangkun, WEI Yao, LIU Yu, LI Wenyi, WU Jinlian
Guangzhou Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China

Abstract: Under the impetus of the global industry, the 5G technology develops rapidly and has entered the critical stage of standard setting and technical verification. Based on this, the key technologies of 5G and its latest standardization progress was introduced mainly, including massive MIMO, new radio network architecture, ultra dense network, edge computing, network slicing, networking on-demand, 4G and 5G interworking, as well as industry organization and operator dynamics, and ended with some thoughts and suggestions.

Key words: 5G, massive MIMO, radio network architecture, ultra dense network, edge computation, network slice, service based architecture

1 引言

2017 年 3 月, 3GPP RAN#75 次会议通过了将部分 5G 标准提前完成的决定^[1], 即在 2017 年 12 月首先完成面向增强移动宽带场景(包括对低时延的支持)的 NSA(non-standalone, 非独立组网) Option 3 系列选项的 Stage 3 相关工作, 而对于 NSA 的 Option 7 系列选项以及 SA(standalone,

独立组网)的 Option 2 和 Option 4 等选项仍维持在 2018 年 6 月完成的计划。提前完成 Option 3 系列选项的动力主要来自于日本、韩国等对 5G 最为激进的国家, 但是该系列选项是一种短期过渡性的网络部署方案, 商用价值不高, 真正具有商用价值的 Option 2、Option 4、Option 5 和 Option 7 等方案仍按原计划制定标准, 因此不会影响 5G 的总体进展和我国对 5G 的战略引领。本文主要介

收稿日期: 2017-08-07; 修回日期: 2017-11-08

基金项目: 国家科技重大专项基金资助项目 (No.2016ZX03001009, No.2017ZX03001001)

Foundation Items: The National Science and Technology Major Project of China (No.2016ZX03001009, No.2017ZX03001001)

绍了 5G 关键技术及其最新标准化进展、行业组织和运营商动态，并提出了一些思考和建议。

2 关键技术研发进展

2.1 大规模天线

随着 LTE 系统对物理层关键技术的深度挖掘，系统容量已经逼近香农极限，5G 频谱效率和系统容量的突破需另辟蹊径。大规模多入多出 (massive MIMO) 概念是在 2010 年由贝尔实验室的 Marzetta 首次提出的^[2]，理论假设基站侧通过部署大规模阵列天线，当天线数远远大于用户终端数时，利用波束成形技术使天线能量集中在一个较窄的方向上传播，多用户传输信道趋于正交，从空间域的维度实现频谱资源复用，能够数倍提高小区容量和频谱效率。大规模天线标准化进展如图 1 所示。

MIMO 技术在 3GPP LTE 早期版本即出现，随着容量需求和技术的发展，天线数随版本演进不断增多，16 天线可以认为是进入多天线“大规模”的门槛。为了实现全维发射，R12 中首先完成了针对 6 GHz 以下频段的 3D 化的信道及应用场景建模工作，通过球面体传播模型替代传统的平面传播模型，垂直维度的波束能够实现高楼覆盖，扩展了多天线的应用场景；紧接着 R13 中，3GPP 定义了能够支持最多 16 个端口的 FD-MIMO 方案；R14 对 6~100 GHz 频段的信道和应用场景进行了建模，同时提出了支持 32 个端口的 eFD-MIMO，支持非周期的 CSI-RS 和上行 DM-RS 增强。2017 年 3 月 3GPP RAN #75 次全会

正式开启了 R15 版本的 5G 标准制定，并在 RAN1 的物理层新空口立项中针对大规模天线技术细节开展工作，包括多天线方案、波束管理、CSI 获取、参考信号设计和 QCI 等。

大规模天线因具备提升系统容量、频谱效率、用户体验速率、增强全维覆盖和节约能耗等诸多优点，而被认为是 5G 最具潜力的无线网关键技术，甚至有商业案例将其引入 4G 系统，增强 LTE 系统能力和生命力。然而，大规模天线技术的发展和应用还需要解决诸多问题，如对于不具有上下行互易性的 FDD 系统，如何有效地实现信道估计是业界一大挑战；在现网环境中大多突发业务是小分组业务，持续并发的数据流有限，多流配对和算法执行效果不理想，小区容量增益提升将大打折扣，因此科学评估和客观对待引入大规模天线对网络带来的实际效益，如何在网络中发挥其独特的技术优势尤为重要。

2.2 新型无线网络架构

5G 新型无线网络架构重新划分了 BBU 和 RRU 功能，实现基带资源的集中化处理，并设计新前传接口实现网络架构的灵活部署，从而进一步提高资源利用率、降低基站能耗以及网络建设和运营成本。

5G 新型无线网络架构由分布单元 (distributed unit, DU)、集中单元 (central unit, CU) 以及连接两者的新一代前传接口组成。DU 包含了原本 RRU 的射频部分和原本 BBU 一部分的基带处理功能；CU 则包含了其余的基带处理功能，实现基带资源的部分集中或全部集中处理。这种新型架



图 1 大规模天线标准化进展



构一方面降低了原有架构对前传接口的带宽和时延要求；另一方面利用基带资源的集中化，实现资源共享和多小区的灵活协作调度，便于平台虚拟化，提高资源利用率，降低网络建设及运营成本。当前 3GPP 中关于新型无线网络架构的讨论范围主要集中在 CU/DU 的功能划分上，共给出了 8 种潜在的功能划分选项，并在 3GPP 中进行讨论^[3]，如图 2 所示。目前，业界倾向在无线架构中同时包含两种功能划分方案：网络兼容性更好、传输要求更低的高层功能划分方案和具有更好协助调度的低层功能划分方案。其中，3GPP 在 2017 年 4 月已确定将 Option 2（在 PDCP 与 RLC 之间进行划分）作为高层功能划分方案并开展相关标准化工作；低层功能划分由于业界无法达成共识，当前在 3GPP 只以研究项目（SI）做讨论，业界倾向于将 Option 7（物理层内划分）作为实际的低层功能划分方案。

在 5G 新型无线网络架构中，前传接口除了需要更高带宽和更低时延，以满足大规模 MIMO 等 5G 新技术带来的高带宽及低时延需求外，还需支持数据分组化以实现针对不同业务场景的 CU/DU 灵活部署。早期对下一代的前传接口研究工作以中国移动推动的 NGFI 为主，并在 IEEE 成立了相关工作组^[4,5]；而原有的前传接口规范制定者 CPRI 联盟（包括爱立信、华为、NEC 和诺基亚）则已在 2017 年 8 月发布其新一代的前传接口——eCPRI。

通过 CU 部分的虚拟化并配合高性能和开放

性的前传接口，5G 无线网络可以支持更为灵活的部署。运营商希望在此基础上实现无线网络的开放化和软硬件分离，从而进一步降低网络建设和运营成本，但是在现实中运营商将面对前传接口异厂商互通难度大、特定场景对传输网络带宽时延要求高以及如何结合应用场景优化资源部署 CU/DU 等挑战。因此，运营商有必要更为深入地研究新型无线网络架构对网络建设和运营的影响，主动投入新架构和前传接口的标准工作中，联合业界推动创新产品的研发工作。

2.3 超密集组网

5G 无线网主要从 3 个维度考虑更高数据速率的实现：更宽的频谱，空口增强，网络密集化^[6]。超密集组网（ultra dense network, UDN）就是通过更加“密集化”的无线网络部署，将站间距离缩短为几十米甚至十几米，使得站点密度大大增加，从而提高频谱复用率、单位面积的网络容量和用户体验速率。

目前 IMT-2020、3GPP 等组织对超密集组网的需求和组网场景进行了研究。3GPP 虽暂无 UDN 技术专题，但在 TR38.913 提出的室内场景、室外密集场景就是典型的 UDN 场景^[7]，UDN 理念实际上已糅合在整个 5G 无线网系统设计的过程之中。

超密集组网在提升容量的同时，也面临同频干扰、移动性管理、多层网络协同、网络回程等一系列影响用户体验或网络部署的技术问题。目前 IMT-2020 成立了专门的 UDN 工作组，针对超密集组网可能面临的问题提出了一些解决方案，

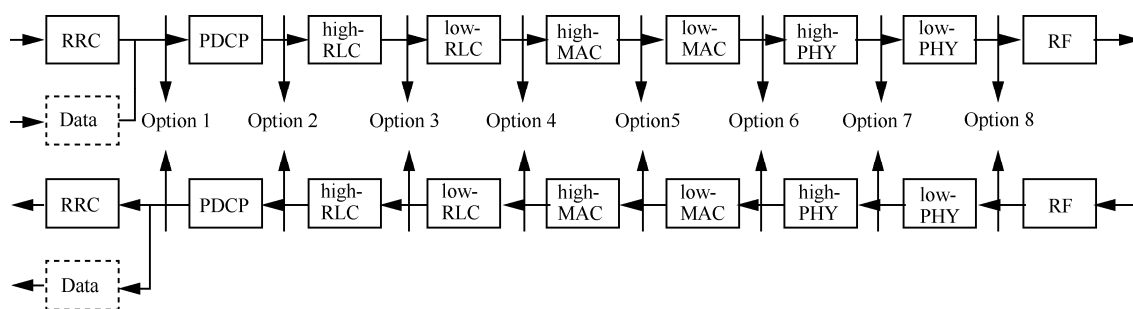


图 2 3GPP 中 CU 与 DU 的潜在功能划分方案选项^[3]

典型的包括干扰管理、小区虚拟化、接入和回程设计。

干扰管理是通过基于网络侧或终端侧的手段降低小区间同频干扰,提升网络性能。网络侧可以通过频域、时域、码域、功率域和空域等角度进行干扰规避,或者通过多小区协同将干扰信号变为有用信号,利用多个小区为同一个用户提供服务。在终端侧,目前研究较多的是干扰对齐(interference alignment)技术,利用干扰信道信息设计编码与译码矩阵,在接收机侧把多个干扰信号抑制到较低干扰空间^[8]。

小区虚拟化是指以用户为中心,将多个实体小区虚拟为一个逻辑小区,通过传输节点间协作为用户提供一致、连续的服务^[8],并通过控制层与数据层分离,避免用户频繁切换。虚拟化可以在一定程度上改善移动性能,降低小区间控制信道干扰,平滑用户体验。

接入与回程联合设计主要是为了解决超密集网络部署过程中可能面临有线回程资源不足的问题,一个建议的方案是使用自回程技术,即回程链路和接入链路使用相同的无线传输技术,通过时分/频分复用同一频带资源。无线回程可大大提高节点部署的灵活性,但需要在链路容量提升、灵活的资源分配、路径优化等方面进行增强,具

体方案还在讨论中。

2.4 边缘计算

ETSI 最早研究移动边缘计算(mobile edge computing, MEC)技术,随后 3GPP 扩大了该技术的适用范围不局限于移动网络,为边缘计算(edge computing, EC),ETSI 也在规范制定过程中将其改为多接入边缘计算(multi-access edge computing)。边缘计算技术就是指在网络边缘位置部署通用服务器,提供 IT 业务环境和云计算能力,其目的是降低业务时延、节省网络带宽、提高业务传输效率,从而为用户带来高质量的业务体验^[9]。

为此,国际标准组织 ETSI 于 2014 年 9 月成立了 MEC 工作组,于 2015 年 9 月发布了第一版 MEC 白皮书——《Mobile edge computing a key technology towards 5G》^[9],完成了需求和框架的标准工作,正式发布需求、架构等三册规范^[10-12]。ETSI 定义的 MEC 系统框架基于虚拟化平台,具备很好的开放性以及自动生命周期管理能力,同时 MEC 平台支持本地应用导入、本地分流策略执行等 MEC 的关键能力,具备 MEC 近距离部署、低时延、位置感知以及无线网络上下文信息感知等特点。多接入 EC 系统框架如图 3 所示。

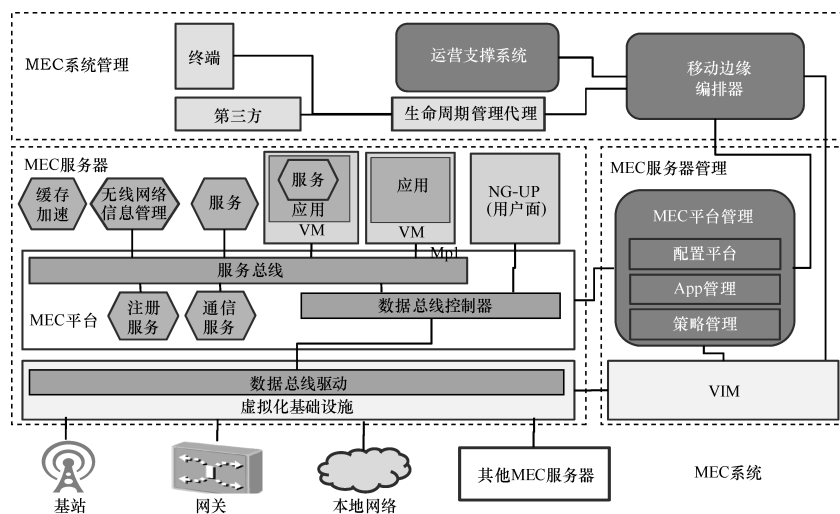


图3 多接入 EC 系统框架^[12]



MEC 系统独立部署,可以部署在无线接入侧,或者传输汇聚点,或者移动网络的核心网边缘(如分布式 DC 的网关侧)。MEC 的应用场景可以是增强现实、智能视频加速、车联网、位置相关业务等。

ETSI 的主要工作集中在 MEC 系统本身,对于 MEC 系统在移动网络中的具体应用则由 3GPP 来完成。3GPP 在技术规范中规定,边缘计算作为 5G 的高层特征之一,遵循 5G 的整体架构,并满足如下功能:本地路由、话务分流、会话和业务连续性、用户面选择和重选、网络能力开放、QoS 和计费^[13]。5G 核心网基于用户签约数据、位置、业务规则、运营商策略等为边缘计算业务选择靠近 UE 的、合适的用户面功能(user plane function, UPF)模块,然后 UPF 将用户数据转发到本地数据网络中的服务器。多接入 EC 系统在 3GPP 网络架构中的位置如图 4 所示。

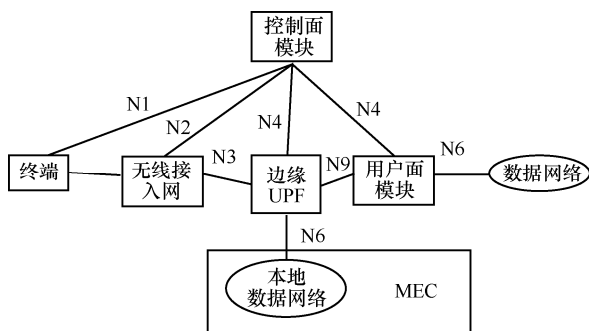


图 4 多接入 EC 系统在 3GPP 网络架构中的位置

通过 EC 可以实现一种新的生态系统和价值链。运营商除了为移动用户就近提供业务计算和数据缓存能力,还可以对可信的第三方开放自己的无线接入网络边缘,灵活快速地部署业务,实现网络从接入管道向信息化使能平台的关键跨越,是运营商未来有效部署业务的新技术之一。随着技术的逐渐成熟和业务需求的逐渐明确,EC 应用会越来越广泛,但是目前 EC 才刚刚起步,标准还不完善,与第三方及行业客户合作的商业模式也需时间和经验来摸索,因此推动 EC 的标

准化工作,研究对 EC 设备的监管、安全及计费问题,考虑 EC 与无线网、核心网整体端到端协调等,是今后 EC 方面的工作重点。

2.5 网络切片

业界对于 5G 提出了需求各异的应用场景,如高带宽、广覆盖的 eMMB;低功耗、大连接的 mMTC;低时延、高可靠的 uRLLC。这些场景的需求差异极大,已经很难用一张统一的网络来满足所有的业务需求,因此引入了网络切片(network slice)技术。

网络切片就是一个按需求灵活构建的、提供一种或多种网络服务的端到端独立逻辑网络。用户使用何种业务,就接入提供相应业务的网络切片。网络切片还是 NFV(网络功能虚拟化)应用于 5G 的关键特征,虽然网络切片在理论上并非必然要使用虚拟化技术,但只有基于虚拟化,网络切片技术才具有商用可行性与商业效益。网络切片示意如图 5 所示。

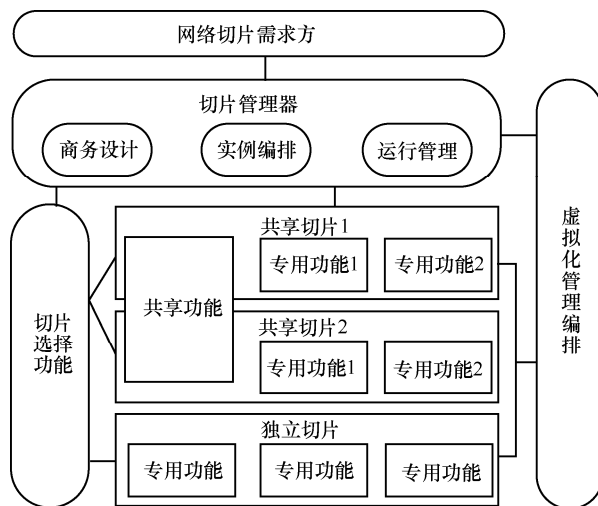


图 5 网络切片示意^[14]

为了实现网络切片,在传统网络基础上新增了切片管理器与切片选择功能两个网络功能,介绍如下。

- 切片管理器包含商务设计、实例编排和运行管理这 3 个依次工作的环节。首先在商

务设计环节, 由网络切片需求方输入切片的相关参数; 然后在实例编排环节, 切片管理器将切片描述文件输出到 MANO, 实现网络切片的实例化; 最后在运行管理环节, 切片管理器监控并动态管理各网络切片。切片管理器主要在 3GPP SA5 组进行研究, 目前切片编排研究报告 (TR28.801) 已经完成, 即将开始相关标准的制定。

- 切片选择功能: 在网络切片运行过程中, 根据用户需求与用户签约信息, 为用户选择接入的网络切片。

根据网络切片控制面功能的共享情况, 网络切片可以有 3 种不同的组网架构^[15], 且这 3 种架构在实际组网时可以混合使用。

- 完全不共享: 每个切片完全独立, 分别拥有各自完整的控制面与用户面功能实体。此架构的切片隔离性最好, 但用户在同一时间只能接入一个网络切片。
- 控制面功能部分共享: 部分控制面功能 (如移动性管理、鉴权功能) 在切片间共享, 其余的控制面功能 (如会话管理) 与用户面功能则是各切片专用的。此架构支持用户在同一时间接入控制面功能部分共享的多个网络切片。
- 控制面功能完全共享: 各切片的控制面功能完全共享, 只有用户面功能是各切片专用。此架构的隔离性最差, 只在用户面实现了隔离, 此架构也支持用户在同一时间接入控制面功能完全共享的多个网络切片。

网络切片技术目前正处于标准制定阶段, 其中 3GPP SA2 负责切片的总体架构与技术方案, 技术点主要包括切片选择、切片的漫游支持、切片的共存等方面。而 SA3 则负责研究切片的安全、隔离方案, 此外 SA5 负责切片的编排与管理的标准化网络切片除了带来技术上的重大突破, 使得用户可以按需接入最合适的网络外,

同时也将为运营商的商用模式、运维模式带来革命性变革, 移动网络由原来的用户/业务适配网络转变为网络适配用户/业务, 原来的单一网络运营方式也逐渐转变为多重网络的动态管理。因此运营商还需要从部署策略、运维模式等方面着力, 加强网络切片的划分、切片与用户/业务的对应策略、切片的上下线流程等关键问题的研究。

2.6 网络按需定制

网络按需定制, 是指网络针对不同应用场景的服务需求引入不同的功能设计, 终端接入时按需选择合适的网络功能实体。

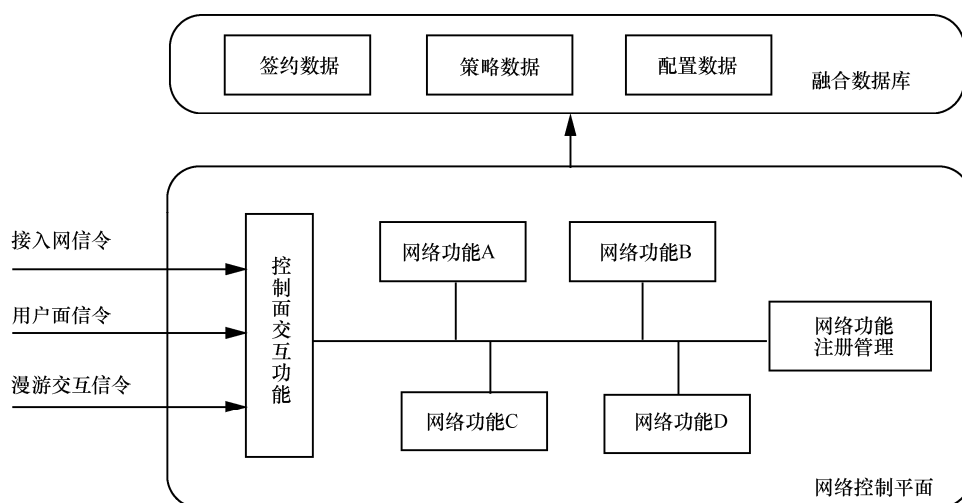
网络按需定制在具体实现上, 会把整个移动网络分为很多功能相对独立的功能实体, 如移动性管理、会话管理、鉴权授权等, 每种功能实体依据业务需求的不同又可以分为不同的小组件, 如移动性管理功能按照对移动性要求的不同, 可以分为低移动性组件与高移动性组件, 其中, 低移动性组件可应用于对位置不敏感的海量物联网终端, 而高移动性组件则用于普通 5G 智能终端。

网络按需定制已经成为 3GPP 制定 5G 核心网标准的核心理念之一, 虽然 3GPP 的 5G 网络架构标准并没有把网络按需定制作为一个专题进行独立研究, 但其思想已经充分体现在 5G 总体架构、移动性管理、会话管理等多个 5G 技术专题研究中。

按需定制的移动网络架构示意如图 6 所示。

在该架构中, 要实现网络按需定制, 需要具备以下 4 个关键特性^[14]。

- 网络功能注册管理: 实现网络功能实体的发现、注册、维护、分配等管理功能。在 3GPP 的 5G 系统架构标准^[13]中定义了 NRF (NF repository function, NF 注册功能) 来负责该项功能。
- 融合数据库: 数据库需为各功能实体提供

图6 按需定制的移动网络架构示意^[14]

用户签约数据、用户上下文等实时数据共享服务。为实现网络功能的灵活定制，5G 功能实体一般采用无状态的设计，因此各功能实体间需要通过数据库进行用户数据的实时交互。为此，3GPP 定义了 UDM (unified data management, 统一数据管理) 用以存储用户签约数据，同时还定义了 UDSF (unstructured data storage function, 非结构化数据存储功能) 用于存储用户上下文等实时数据。

- 控制面交互功能：主要负责核心网控制信令的对外收发。目前 3GPP 主要使用 AMF 与 SMF (session management function, 会话管理功能) 来完成该功能。
- 服务化的架构 (service based architecture, SBA) 与中立接口：功能实体通过统一的、服务化的中立接口对外提供服务。在传统的移动网络间，任意两个需要通信的网元间的接口都是定制的。而在 5G 架构中，由于划分了较多的功能实体，如果仍沿用原来的接口方式，将引入很多定制接口，大大增加了网络的复杂度。为此根据 IT 中微服务的理念，提出了服务化的架构与中立接口的概念，即每个功能实体对外只提供

一个统一的接口，通过该统一接口对外提供服务。

网络按需定制与前述的网络切片均充分体现了网络适配用户的思想，其中，网络切片是为用户选择合适的网络，而网络按需定制则是为用户选择合适的功能实体。网络按需定制中的功能实体划分、网络功能注册管理、服务化架构与接口等关键技术都是 3GPP 标准制定的研究热点，3GPP 5G 核心网第一版的 Stage 2 标准 (即架构及总体流程部分) 根据计划会在 2017 年内制定完成。

2.7 4G 与 5G 互操作

4G 与 5G 网络的互操作研究主要在 3GPP 的 5G 项目中进行，在 RAN 组和 SA 组的前期研究中汇总了 8 种 5G 网络的部署选项，根据大部分主流运营商和设备商的探讨，现在主要聚焦 4 种选项的规范制定^[16]，分别介绍如下。

2.7.1 Option 2 选项

Option 2 是 NR 独立组网的典型部署选项，即端到端纯 5G 网络，如图 7 所示。这种部署选项可用于两种具体场景：4G 网络和 5G 网络并存，各自独立，可以视业务需求通过 5GC (5G core, 5G 核心网) 与 EPC 互连实现两网的互操作；4G 已经退网，只有 5G 网络单独存在并承载所有业务。

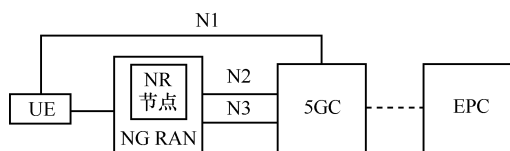


图7 Option 2选项

2.7.2 Option 3系列选项

Option 3系列选项是一系列的NR非独立组网部署选项，核心网采用EPC、UE和网络之间的控制信令锚定在LTE空口和S1-MME接口上传输，用户面具有NR和LTE双连接能力。Option 3的用户面有3种方式接入EPC，所以Option 3系列有3种具体选项：第1种选项称为Option 3，eNode B和NR节点的用户面在eNode B汇聚后再接入EPC，NR节点不需支持S1-U接口；第2种选项称为Option 3a，eNode B和NR节点的用户面各自接入EPC；第3种选项称为Option 3x，eNode B和NR节点的用户面在NR节点汇聚后再接入EPC。Option 3系列选项如图8所示。

Option 3系列选项可用于5G网络部署早期，即在5G标准和5GC设备尚未成熟前，为了快速支持eMBB业务和对4G的容量补充，只引入5G新空口、暂不引入5GC，而且NR节点只起到分

担用户面流量的作用，所有的消息和控制权都在EPC。

2.7.3 Option 4系列选项

Option 4系列选项是一系列的NR独立组网特殊部署选项，是在5G网络架构下将eNode B升级成为演进型eNode B，借助其用户面实现NR和LTE双连接能力。根据用户面的不同接入方式，Option 4系列有2种具体选项：第1种选项称为Option 4，即演进型eNode B和NR节点的用户面在NR节点汇聚后再接入5GC；第2种选项称为Option 4a，即演进型eNode B和NR节点的用户面各自接入5GC。Option 4系列选项如图9所示。

Option 4系列选项主要用于5G网络部署后期，此时5G网络已建设完成并在特定区域内提供连续覆盖，4G网络仍然具有广覆盖和一定的容量能力，因此可利用4G网络分担一部分业务流量，所有的消息和控制权都在5GC。

2.7.4 Option 7系列选项

Option 7系列选项是将Option 3系列选项的核心网由EPC改为5GC，5G业务由5GC控制，eNode B也必须升级为支持5G网络接口的演进型

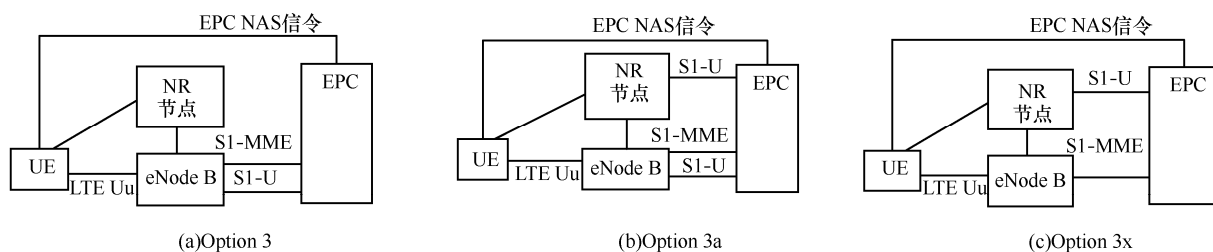


图8 Option 3系列选项

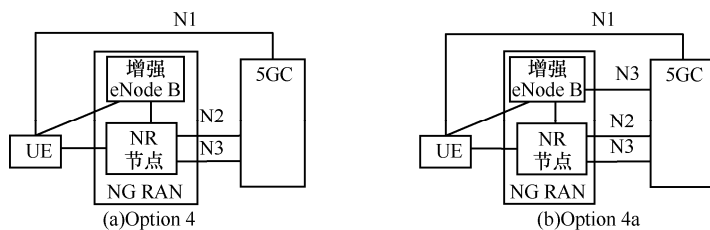


图9 Option 4系列选项



eNode B。Option 7 系列选项如图 10 所示。

这种部署方式可用于 5G 网络部署初期、引入 5G 核心网之后。由于部署初期 5G 网络的覆盖有限,为发挥 4G 网络的覆盖优势,该系列选项升级 UE 和 eNode B 接入 5GC, UE 和网络之间的控制信令锚定在 LTE 空口和 N2 接口上传输,所有的消息和控制权在 5GC, NR 节点的目的仍然是快速支持 eMBB 业务和对 4G 的容量补充,具有分担用户面流量的作用。

探讨 4G 与 5G 的互操作,首先需要明确的是,5G 网络对 4G 网络不是颠覆和替代的关系,而是 4G 网络如何向 5G 网络演进以及两网如何融合发展的过程。根据 3GPP 的路标, LTE 还有很长的生命周期, LTE 的网络能力还能满足相当多的应用场景和基本的容量需求。同时, 5G 在发展过程中出现了“5G 技术 4G 化”的现象,如大规模 MIMO 在 4G 网络中的应用等。因此,应结合业务需求和网络演进的需求,摆脱单纯的 4G 网络或者 5G 网络的框架束缚,从技术本身的特性和现实网络的客观存在出发,兼收并蓄,兼顾 4G 的持续发展和 5G 新技术的引入,从全网发展的角度考虑移动网络的演进路径和互操作关键技术的研究。

3 行业组织和运营商动态

每一代移动通信技术的诞生和推广,都意味着一个重新分配产业利益的重大机遇。虽然 5G 标准化工作尚未完成, 4G 商用取得成功的美国、日本、韩国等领先运营商出于国家战略、市场竞

争或自身技术创新的需求,纷纷发布激进的 5G 商用计划,希望在新一轮全球竞赛中继续取得先机,呈现出“你追我赶”的态势,总体聚焦在 eMBB 场景。

为在 5G 产业上取得先机,美国联邦通信委员会(FCC)在全球率先颁布 5G 频率,均为 24 GHz 以上,共 11 GHz 带宽,包括授权频段 3.85 GHz 带宽(包括 28 GHz、37 GHz、39 GHz)和非授权频段 7 GHz 带宽(包括 64 GHz 和 71 GHz)。美国第一大移动运营商 Verizon 更是“虚张声势”,2016 年 7 月发布了其企业的 5G 规范,更豪言 2017 年开始商用部署。由于 Verizon 的 5G 计划远早于最早可用的 3GPP 5G 版本发布时间(2018 年 6 月),引起业界担心 Verizon 的 5G 将与 3GPP 的 5G 在标准和产业上出现分裂,使得 5G 技术碎片化。然而, Verizon 的 5G 只是一个增强的固定无线接入系统,在用户侧依靠 Wi-Fi 解决室内覆盖,无法与 3GPP 的 5G 标准相提并论。出于美国国内市场竞争需要, AT&T 和 T-Mobile 也相继宣布他们的 5G 计划,不过两者更为务实,将基于 3GPP 标准来与业界同步推进。AT&T 虽然宣布 2016 年开始测试 5G 用于家用宽带服务,但表示 2018 年才开始部署。T-Mobile 展示了丰富的 5G 业务愿景,表示如此丰富的移动应用需要等到 2020 年。

日本、韩国作为历代移动通信技术发展的急先锋,在 5G 发展上也不例外。韩国两大运营商 SK Telecom (SK 电讯)和 KT 急于在 2018 年 2 月冬奥会期间展示其 5G 移动通信服务,已在 28 GHz 高频段开展试验,近期 SK 电讯还发布了 5G

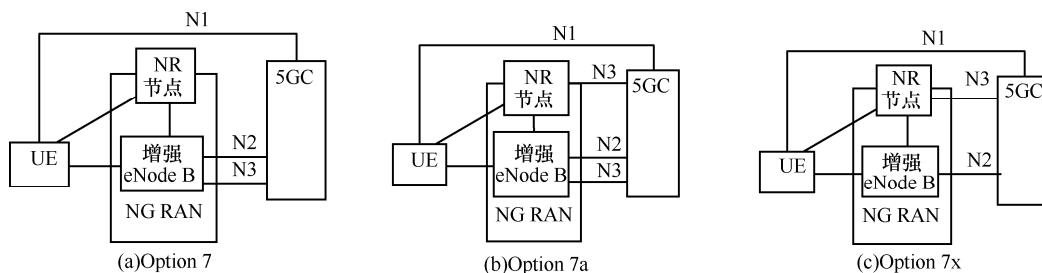


图 10 Option 7 系列选项

商用设备的招标。日本则计划在 2020 年东京奥运会前实现 5G 商用,目前 DoCoMo 正在组织验证 5G 关键技术以推动实现初期的 5G 商用目标,之后计划推动“5G+”来支持更为丰富的业务场景。

欧洲是 3G/4G 标准化的领导者,但是 4G 的发展偏慢,在 5G 的发展上也趋于稳健,目前更多的是希望能继续在 5G 时代主导推进标准。受美国 5G 激进计划的影响,欧盟在 2017 年 9 月公布了 5G 行动计划,给出了比较完整的 5G 工作时间表:在 2017 年 3 月公布具体的测试计划并开始测试,2017 年底制定完整的 5G 部署路线图,2018 年开始预商用测试,2020 年各个成员国至少选择一个城市提供 5G 服务,2025 年各个成员国在城区和主要公路、铁路沿线提供 5G 服务。同时,欧盟也在推动在 2017 年底之前确定 6 GHz 以下和 6 GHz 以上的频率划分,以支持高低频融合的 5G 网络部署。

中国政府也已明确积极推进 5G 于 2020 年商用,工业和信息化部从 2015 年 9 月至 2018 年底主导 5G 关键技术试验,三阶段试验包含关键技术验证、技术方案验证和系统验证。3 个运营商在近期都公布了自己的实验室、外场和部署计划,逐步推动产业成熟,实现 2020 年商用或试商用。不过受国内高频器件产业弱势的限制,中国更重视 6 GHz 以下频率的 5G 应用,首发的 5G 应用频段很可能为 3.5 GHz 和 4.8 GHz 频段。

全球而言,国际电信联盟(ITU)早在 2015 年就发布了 5G 愿景和需求定义,并于 2016 年正式将下一代移动蜂窝通信系统命名为“IMT-2020”,2018 年底启动 5G 技术评估及标准化,旨在 2020 年第一季度完成技术评估并最终决定技术方案。另一方面,5G 的商用将与 3GPP 的 5G 标准发展对应,在 2020 年左右首先启动 eMBB 业务场景的商用,而后再逐步支持更为丰富的业务场景^[17-19]。

4 结束语

移动通信系统不断向统一的制式演进,在 4G

时代已经统一到 3GPP 制定的 FDD 和 TDD 两种 LTE 制式,4G 终端普遍同时支持尽可能多的主流频段,4G 核心网更是遵从统一的一套 3GPP 标准,而 5G 的技术标准由 3GPP 统一制定,5G 将延续 4G 的这种模式得到发展。随着业务不断从 2G、3G 向 4G 迁移,将来承载移动通信业务的主流制式将是 4G 和 5G,产业链将比 3G 和 4G 时代更加聚焦和同质化。但是因为 4G 网络的巨大成功,在发展 5G 时除了需要考虑拓展垂直应用的需求,还要充分考虑现有移动网络的演进需求。4G 现网工作频率低、覆盖能力强、用户体验速率较低,而 5G 工作频率高、覆盖能力弱、用户体验速率较高,因此如何将两者融合起来以实现优势互补、为用户提供更好的体验,是无线组网方面需要重点研究的课题,具体涉及空口采用单连接还是双连接、各种业务的承载策略以及终端实现上的技术和成本问题。5G 核心网在理论上具备网络切片、按需定制等能力,目标是实现网络对业务的适配,但是目前业界对于网络切片的内涵、切分及应用场景存在不同的理解和需求,服务化架构的具体实现存在挑战,这些都是需要业界共同努力去研究解决的课题。

在讨论 5G 时还需要考虑 5G 和 Wi-Fi 的关系,两者相近的方面包括工作频率都包括 6 GHz 以下和 6 GHz 以上相关频段,都支持 MU-MIMO 等关键技术,在室内 5G 和 Wi-Fi 的能力相近,应用场景将发生较多重叠,两者差异化的方面包括 Wi-Fi 具有成本低、终端支持度高、后向经营商业模式较为成熟等优势,而 5G 具有服务保证、网络安全等优势,将来 5G 应该采取何种商业模式、5G 如何与 Wi-Fi 形成优势互补,都是需要尽早开展研究的课题。

5G 首次在移动宽带以外对其他场景提出了相应的技术方案,功能更为丰富,性能大幅提升,拓展了对各种垂直应用的支持,因此 5G 网络不再像前几代移动网络那样单纯和僵硬,同时随着

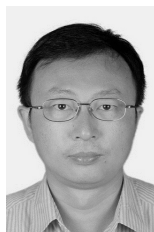


NFV 成为网络部署的主流方式以及移动网络的商业模式面临变革, 5G 网络的运营体系也需要及时做出相应的调整。在这样的大背景下, 电信运营商的理念必须及时由“业务适配网络”向“网络适配业务”转变, 面向以 eMBB、URLLC 以及 mMTC 3 种典型场景为顶点构成的综合场景, 考虑移动宽带和各种垂直应用对网络的需求, 充分利用 4G 现有网络和各级云计算基础资源, 综合运用 5G 和 NFV 的各种关键技术, 实事求是、因地制宜地制定 5G 网络部署方案。

参考文献:

- [1] 3GPP. RP-170741 way forward on the overall 5G-NR eMBB workplan[S]. 2017.
- [2] MARZETTA T L. Noncooperative cellular wireless with unlimited numbers of base station antennas[J]. IEEE Transactions on Wireless Communications, 2010, 9(11): 3590-3600.
- [3] 3GPP. Study on new radio access technology; radio access architecture and interfaces (release 14): TR38.801, V14.0.0[S]. 2017.
- [4] China Mobile Research Institute. C-RAN the road towards green ran. white paper, version 3.0[R]. 2013.
- [5] China Mobile Research Institute. White paper of NGFI (next generation fronthaul interface) version 1.0[R]. 2015.
- [6] ITU-R. Recommendation ITU-R M.2083: IMT vision-framework and overall objectives of the future development of IMT for 2020 and beyond[S]. 2015.
- [7] 3GPP. Study on scenarios and requirements for next generation access technologies (release 14): TR38.913[S].
- [8] IMT-2020. 超密集网络 (UDN) 技术研究报告[R]. 2015. IMT-2020. Research report on ultra dense network technology[R]. 2015.
- [9] ETSI. White paper No.11 mobile edge computing a key technology towards 5G[S]. 2015.
- [10] ETSI. GS MEC 001 V1.1.1 mobile edge computing (MEC) terminology[S]. 2016.
- [11] ETSI. GS MEC 002 V1.1.1 mobile edge computing (MEC); technical requirements[S]. 2016.
- [12] ETSI. GS MEC 003 V1.1.1 mobile edge computing (MEC); framework and reference architecture[S]. 2016.
- [13] 3GPP. System architecture for the 5G system (release 15): TS23.501, V1.5.0[S]. 2017.
- [14] IMT-2020(5G)推进组. 5G 网络架构设计白皮书[R]. 2016. IMT-2020 (5G) Propulsion Group. White paper of 5G network architecture design[R]. 2016.
- [15] 3GPP. Study on architecture for next generation system (Release 14): TR23.799, V1.0.1[S]. 2016.
- [16] 3GPP. Evolved universal terrestrial radio access (E-UTRA) and NR; multi-connectivity; stage 2 (release 15): TS37.340, V1.0.0[S]. 2017.
- [17] 朱浩, 项菲. 5G 网络架构设计与标准化进展[J]. 电信科学, 2016, 32(4): 126-132. ZHU H, XIANG F. Architecture design and standardization progress of 5G network[J]. Telecommunications Science, 2016, 32(4): 126-132.
- [18] 陈山枝. 发展 5G 的分析与建议[J]. 电信科学, 2016, 32(7): 1-10. CHEN S Z. Analysis and suggestion of future 5G directions[J]. Telecommunications Science, 2016, 32(7): 1-10.
- [19] 张建敏, 谢伟良, 杨峰义, 等. 移动边缘计算技术及其本地分流方案[J]. 电信科学, 2016, 32(7): 132-139. ZHANG J M, XIE W L, YANG F Y, et al. Mobile edge computing and application in traffic offloading[J]. Telecommunications Science, 2016, 32(7): 132-139.

[作者简介]



王庆扬 (1973-), 男, 博士, 中国电信股份有限公司广州研究院移动通信研究所所长、高级通信工程师, 主要研究方向为移动通信系统的关键技术及应用、移动网络规划和优化。

谢沛荣 (1977-), 男, 中国电信股份有限公司广州研究院高级工程师, 长期从事移动通信核心网络标准与技术研究工作。

熊尚坤 (1977-), 男, 博士, 中国电信股份有限公司广州研究院高级工程师, 长期从事移动通信新技术标准研究和试验工作, 现专注于 5G 演进的标准、技术方案研究和试验工作。

魏焱 (1984-), 男, 博士, 中国电信股份有限公司广州研究院工程师, 主要研究方向为 LTE、5G 无线通信系统关键技术。

刘昱 (1984-), 男, 博士, 中国电信股份有限公司广州研究院工程师, 主要研究方向为无线网络架构及新技术。

李文苒 (1972-), 女, 中国电信股份有限公司广州研究院高级工程师, 主要研究方向为各种制式的移动核心网技术和组网。

吴锦莲 (1977-), 女, 中国电信股份有限公司广州研究院高级工程师, 主要研究方向为移动通信系统新技术。



工业通信网络现有架构的梳理总结和 未来运营商的发展策略

常洁, 王艺, 李洁, 陈正文

(中国电信股份有限公司上海研究院, 上海 200122)

摘要: 工业通信网络向网络化、数字化、智能化方向发展, 已经不能被视为简单的通信网络, 而是组成自动化控制系统的关键子系统。从现有的工业通信网络系统架构出发, 对工业模型进行了系统梳理, 将其划分为企业外部和企业内部, 现有的架构主要集中于制造企业内部工厂内, 主要参考 ISA-95 模型, 包含现场级、控制机、车间级和工厂级 4 层, 每层之间的网络配置和管理策略相互独立。然后详细分析了现有的工业网络技术体系和面向制造业未来发展的网络升级需求, 最后提出通信运营商的发展策略。

关键词: 工业通信网络; 发展策略; 架构梳理

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017301

Summary of existing framework in industrial communication networks and future development strategies for communication operators

CHANG Jie, WANG Yi, LI Jie, CHEN Zhengwen

Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China

Abstract: Industrial communication networks are developing in the direction of networking, digitization and intelligence, which could not be considered as just a single communication network but as important sub-systems of the automatic control systems. Based on framework of the existing industrial communication network, the industrial model was systematically studied and divided into within and outside of the enterprise. The existing framework mainly concentrated in interior of factory within manufacturing enterprise and mainly referenced ISA-95 model, included field-level, control-level, workshop-level and factory-level. The network configuration and management strategies between each layer were independent. Then, the existing technology systems in industrial network and requirements of network upgrading facing future development of manufacturing were analyzed in detail. Lastly, the future development strategies of communication operators were put forward.

Key words: industrial communication network, development strategy, summary of framework



1 引言

全球新一轮的工业革命正在孕育兴起,未来制造业将建立在以互联网和信息技术为基础的网络平台之上,将更多的生产要素进行更为科学的整合,使其变得更加自动化、网络化、智能化。美国 GE 公司倡导的工业互联网指利用“智能设备”采集“智能数据”,利用“智能系统”进行数据挖掘和可视化展现,形成“智能决策”,为生产管理提供实时判断参考,反过来指导生产,优化制造工艺^[1];德国的工业 4.0 是制造业互联网化的一个体现,主要指在“智能工厂”以“智能生产”方式制造“智能产品”,整个过程贯穿以“网络协同”,从而整合生产资源、提升生产效率、缩短生产周期、降低生产成本^[2];我国政府相继出台的《中国制造 2025》^[3]和《关于积极推进“互联网+”行动的指导意见》^[4],以推进信息化和工业化深度融合为主线,大力发展智能制造和“互联网+”协同制造,构建工业互联网条件下的产业生态体系和新型制造模式。

可以说,工业互联网、工业 4.0、中国制造 2025 一脉相承,都面向智能制造力图实现工业制造的纵向集成、横向集成和端到端集成。在生产制造过程中,纵向集成是指基于智能工厂中的网络化的制造体系,实现分散式生产,以替代传统的集中式中央控制的生产流程,主要体现在工厂内的科学管理上,从侧重于产品的设计和制造过程,走到了产品全生命周期的集成过程,建立有效的纵向生产体系,实现从企业管理、生产执行、生产监控到现场设备的集成,纵向集成的主要效果是提升效率;横向集成是指传统的供应链、工厂及销售网络的直线型价值组织方式会逐渐联网,形成网络制造生态,原来集中的工厂组织模式将发生改变,工厂不仅仅是智能的,其生产单元也将变得更为细小,这样才便于实现更为个性化的生产,从而给产品生产制造的传统工艺带来革命性的影响,其目的是重构价值生态和创新商业模式;端到端集成包含研发、原材料采购、物流、仓储、生产、销售、交付和

服务的产品全生命周期流程的灵活集成,尤其是在一个单一的产业链上完成端到端集成,可以实现产品的更好体验和利益的最大化,从而提升产业链控制力,端到端集成是从工艺流程角度审视智能制造,主要体现在并行制造商,将由单元技术产品通过集成平台,形成企业的集成平台系统,并朝着工厂综合能力平台发展。

2 现有工业通信网络系统架构

现有的工业网络系统架构包含 ISA-95 提出的参考模型^[5]、Cisco (思科)的信息系统整体框架及其 CPwE (converged plantwide ethernet architecture, 全厂融合以太网)联合架构^[6]、美国 NIST (National Institute of Standards and Technology, 国家标准与技术研究院)给出的智能制造生态体系模型^[7]、德国电工电子与信息技术标准化委员会的 RAMI 4.0 (reference architecture model industry 4.0, 工业 4.0 参考架构模型)^[8]、我国 MIIT (Ministry of Industry and Information Technology, 工业和信息化部)和 SAC (Standardization Administration of the People's Republic of China, 国家标准化管理委员会)两部委联合发布的《国家智能制造标准体系建设指南(2015 年版)》(以下简称“建设指南”)智能制造系统架构^[9]以及 AII (Alliance of Industrial Internet, 工业互联网产业联盟)给出的工业互连网络体系架构^[10]等。

(1) ISA-95 标准参考模型

ISA-95 标准参考模型如图 1 所示,共分为 5 层, MES (manufacturing execution system, 制造执行系统)仍然在企业业务计划层与底层控制层之间,处于中间层。其中,第 0 层表示过程,通常指制造或生产过程;第 1 层表示用来监控和处理这些过程的人工或传感器以及相应的执行机构;第 2 层表示手动或自动的控制动作,使过程保持稳定或处于控制之下;第 3 层表示生产期望产品的工作流活动,生产过程的协调与优化、生产记录的维

护等,这是制造运作管理的范围,也就是MES关注的范围;第4层表示制造组织管理所需的各种业务相关活动,包括建立基础车间调度、确定库存水平以及确保物料适时适量适地进行生产。第3层和第4层之间的接口通常是工厂生产计划和运行管理及车间协调之间的接口。MES关注的是企业的制造执行,其主要的功能范围和制造执行层相对应;同时,MES也考虑与业务计划层及过程控制层所选用系统之间的信息交互。

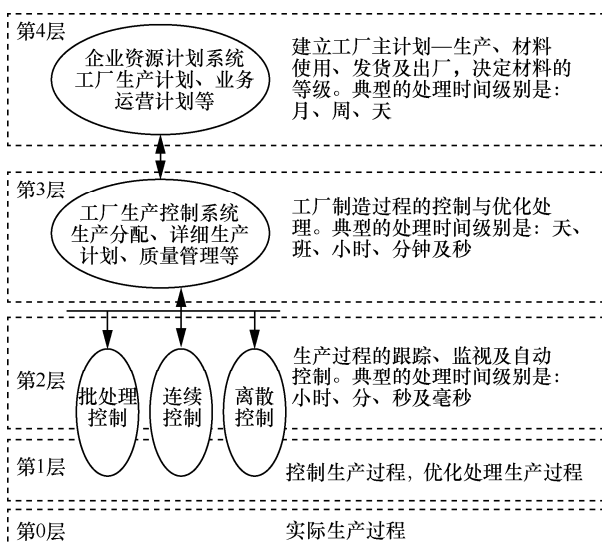


图1 ISA-95 标准参考模型

(2) 思科的 CPwE 联合架构

思科的 CPwE 联合架构如图2所示,是一个从厂房设备到MES和业务系统,也包括用户和合作伙伴的具有可预测性能和系统弹性的端到端的架构,与主流的工业标准(包括ISA-95和ISA-99)兼容。CPwE联合架构主要包含三大块:控制区、隔离区和企业区,共6个层级(0~5级),企业区分为企业网络(第5级)和业务规划网络和物流网络(第4级),隔离区中未包含任何层级,主要提供缓存区域用于实现应用和数据的共享,控制区分为生产运营和控制(第3级)、区域控制(第2级)、基本控制(第1级)和流程(第0级)。CPwE可在0~2层级之间实现基于Ethernet/IP(工业协议)的网络实时控制及流量隔离和管理,在第3级实现站点操作和控制及多服务网络的安全管理,在DMZ(demilitarized zone, 隔离区)实现应用和数据共享、访问控制和威胁防范,并在4~5级实现企业IT集成协作及应用程序优化。随着无线技术的普及,思科已经将最先进的无线技术添加到CPwE架构中,借助无线技术,可以采用工业标准Wi-Fi,将它延伸到控制型应用,并在工厂恶劣环境中实现无线操作,也可通过移动平板电脑查看

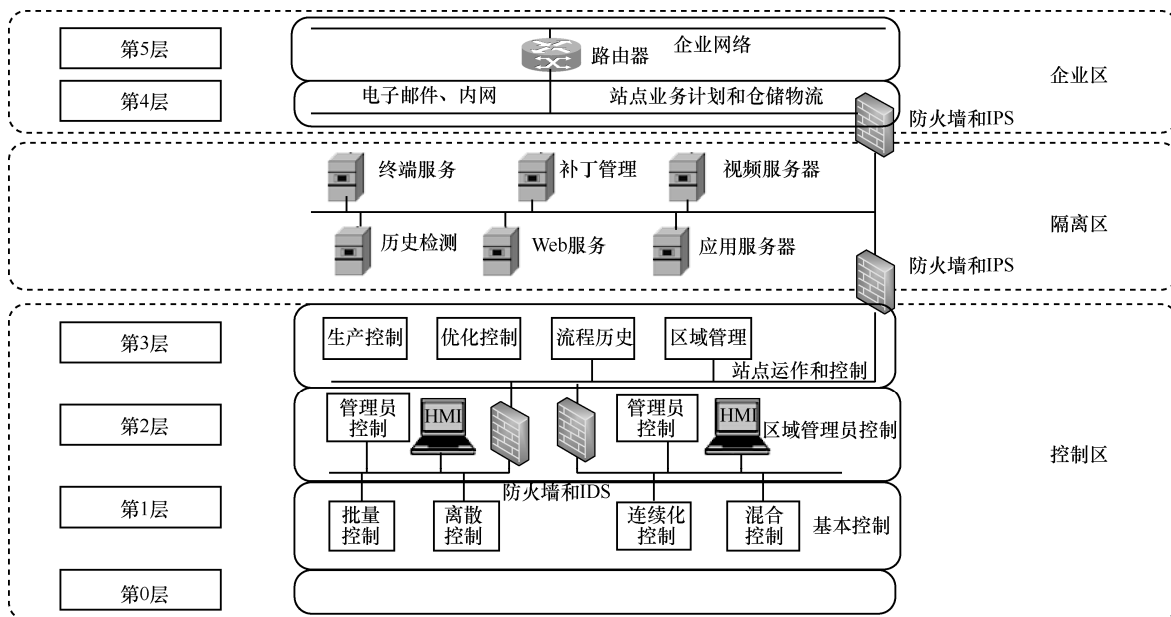


图2 思科的 CPwE 联合架构



生产设备。这对于提高生产率、加快响应速度非常有利,同时由于无需使用电缆,还可以节省资金。

(3) NIST 智能制造生态体系模型

NIST 的智能制造生态体系模型如图 3 所示,标准横跨产品、生产系统和商业这 3 项主要制造生命周期维度,展示了整个制造系统的三维空间。NIST 参考的是常用的 ISA-95 参考模型,共分为 4 层:设备层、SCADA 层、MOM (manufacturing operations management, 制造运营管理) 层和企业层。其中,设备层包含了常用的现场总线及传感器和执行器,对应于 ISA-95 参考模型中的第 1 层。SCADA 层遵循了 IEC 61512 《批控制》等一系列标准,用于现场设备的控制,等同于 ISA-95 参考模型中的第 2 层。MOM 层等同于 ISA-95 参考模型中的第 3 层 MES 层。企业层侧重于企业级的制造标准。右侧有一些跨层的标准用于定义制造系统的安全 (ISA-99)、质量管理过程 (ISO 9000)、能源管理 (ISO 50001) 和环境管理 (ISO 14000)。

(4) RAMI 4.0 参考架构

RAMI 4.0 以一个三维模型展示了工业 4.0

涉及的所有关键要素,如图 4 所示,工业 4.0 集中于产品开发和生产全过程。RAMI 4.0 从 3 个层面对模型进行阐述,左侧垂直轴从 IT 视角的架构出发,借用了信息和通信技术常用的分层概念,类似于著名的 ISO OSI 七层模型,各层实现相对独立的功能,同时下层为上层提供接口,上层使用下层的服务。RAMI 4.0 模型的第二个维度 (左侧水平轴) 描述全生命周期及其相关价值流,这一维度从业务流程视角的架构出发,此处的过程是指生产过程,完整的生命周期从规划开始,到设计、仿真、制造,直至销售和服务。RAMI 4.0 模型的第三个维度 (右侧水平轴) 从应用视角的架构出发,依据 ISA-95 和 ISA-88,主要关注的是工业生产环境下产品的制造过程控制和管理功能。更进一步,由于工业 4.0 不仅关注生产产品的工厂、车间和机器,还关注产品本身以及工厂外部的跨企业协同关系,因此在底层增加了“产品”层,在工厂顶层增加了“互联世界”层。

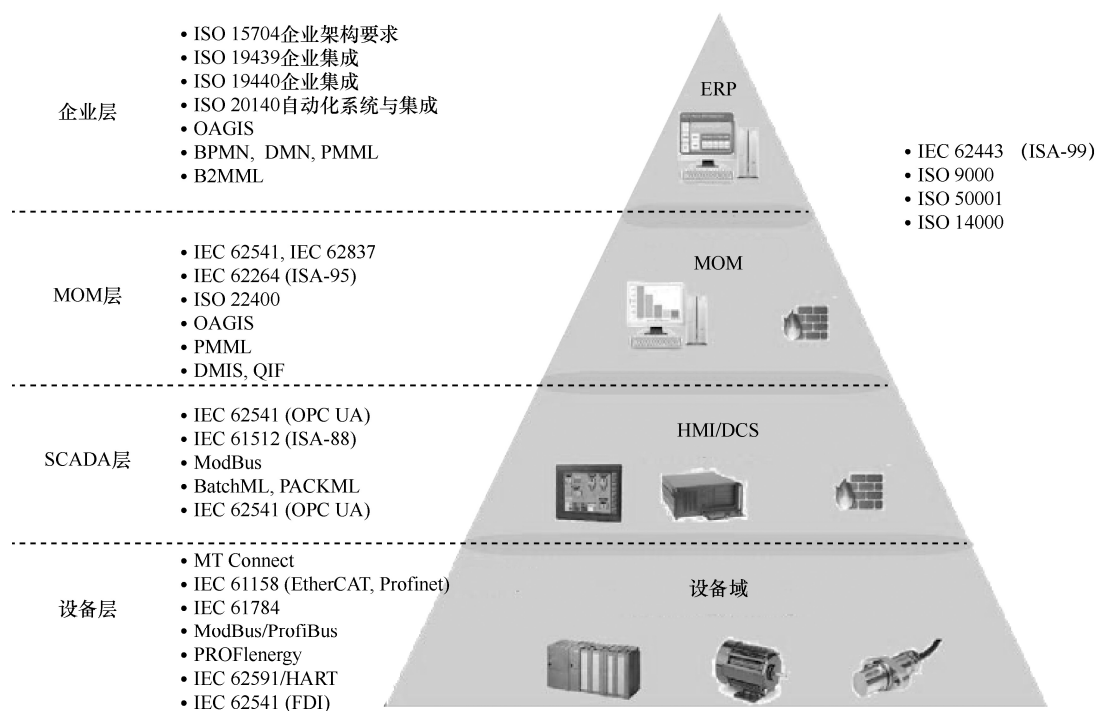


图 3 NIST 智能制造生态体系模型

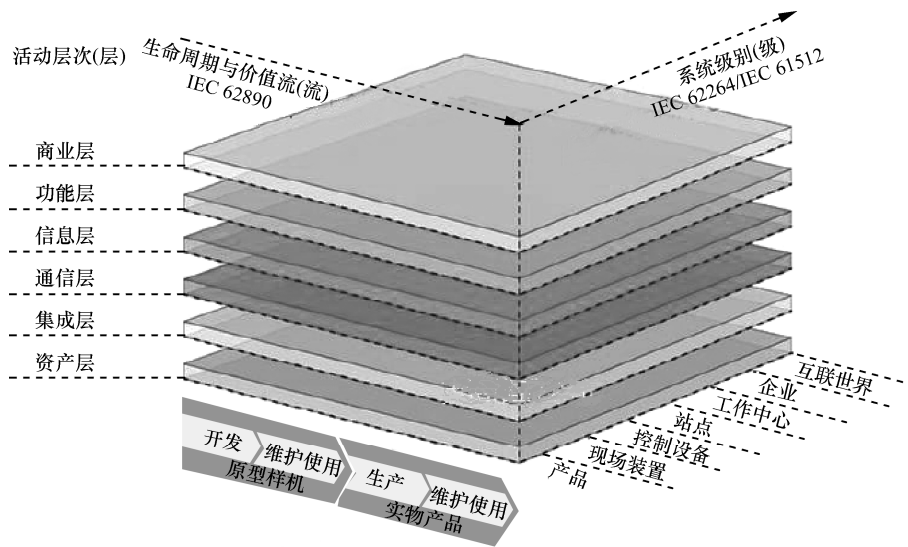


图4 RAMI 4.0 参考架构

(5) 智能制造系统架构

《建设指南》从生命周期、系统层级和智能功能3个维度，如图5所示，建立了智能制造标准体系模型，并由此提出了智能制造标准体系框架，主要完成智能制造标准体系结构和框架的建模研究工作。生命周期坐标轴是由设计、生产、物流、销售、服务等一系列相互联系的价值创造活动组成的链式集合。不同行业的生命周期构成不尽相同。在智能制造的大趋势下，企业从主要提供产品向提供产品和服务转变，价值链得以延伸。系统层级坐标轴自下而上共5层，分别为设备层、控制层、车间层、企业层和协同层，智能制造的系统层级与主流的工业标准（ISA-95）相兼容，体现了装备的智能化和IP化以及网络的扁平化趋势。智能功能坐标轴自上而下包括资源要素、系统集成、互联互通、信息融合和新兴业态5层。

(6) 工业互联网体系架构

工业互联网体系架构包括工厂内部网络和工厂外部网络，如图6所示，工厂内部网络呈现两层三级的结构。两层是指存在工厂IT网络和工厂OT网络两层技术异构的网络；三级是指根据

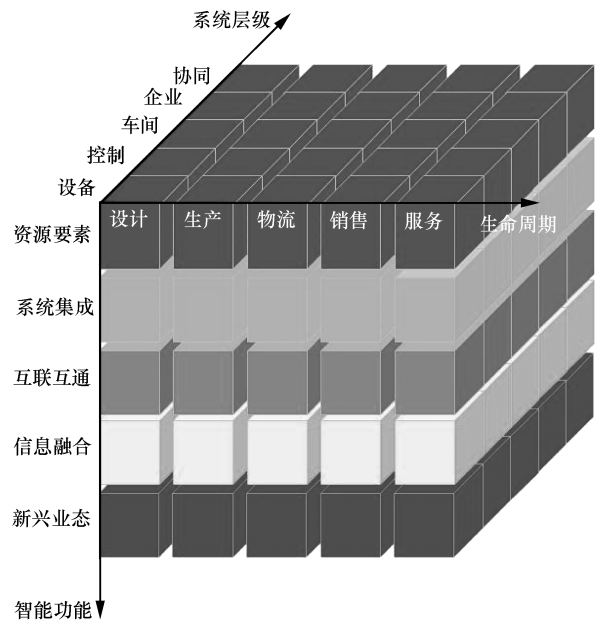


图5 智能制造系统架构

目前工厂管理层级的划分，网络也被分为现场级、车间级、工厂级/企业级3个层次，每层之间的网络配置和管理策略相互独立。工厂OT网络主要用于连接生产现场的控制器（PLC、DCS、FCS等）、传感器、伺服器、监控设备等部件。工厂OT网络的实现技术主要为现场总线和工业以太网等。工厂IT网络主要由IP网络构成，并通过

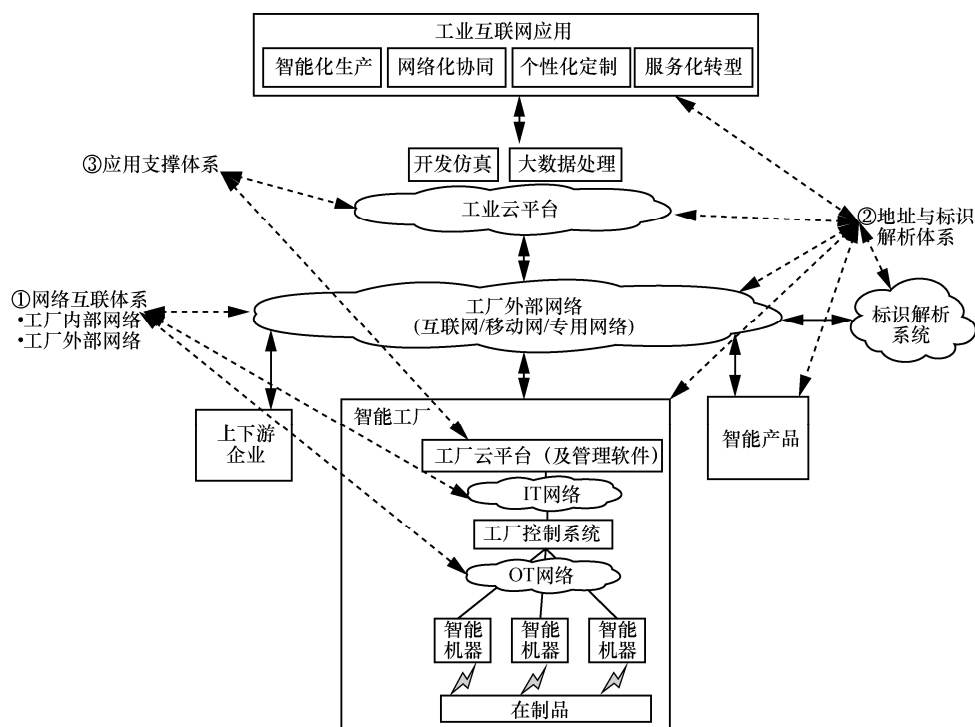


图6 工业互联网络体系架构

网关设备实现与互联网和工厂 OT 网络的互联和安全隔离。工厂外部网络主要强调工业生产信息系统与互联网正在走向深度协同与融合,包括 IT 系统与互联网的融合,即企业将其 IT 系统(如 ERP、CRM 等)托管在互联网的云服务平台中,或利用 SaaS 服务商提供的企业 IT 软件服务。

通过现有架构的对比分析和总结,给出了工业网络的系统架构模型的全视图,为之后的工业网络技术体系奠定了基调,如图 7 所示。通过现有工业网络的系统架构分析可以看出,现有的架构主要集

中于制造企业内部工厂内,以网络为依托,采取纵向分层、横向分散的策略,纵向分层主要参考 ISA-95 参考模型,包含现场级、控制机、车间级和工厂级 4 层,每层之间的网络配置和管理策略相互独立。其中,由于现场作业的特点使得业务处理位置分散、软硬件条件参差不齐,而且需要维护的设备种类数量繁多,这就要求系统用简便、高效、智能的技术手段简化系统的维护工作。同时需要通过标准化选项完成现场信息录入,在简化现场操作的同时,消除因为人员差异而对同一作业信息的不同主观描

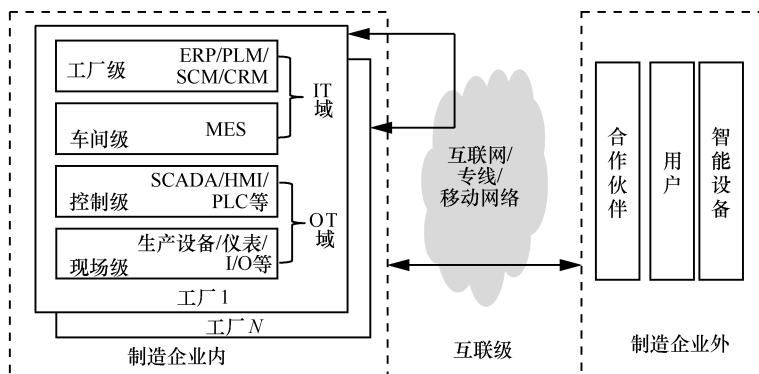


图7 工业系统模型梳理

述。在企业外部,采用互联网、专线和移动网络等与合作伙伴、用户和智能设备互联。

由工业系统的模型可以梳理出工业网络的分层架构,主要分为现场级、控制级、车间级、工厂级和互联级,其中,互联级主要处于场外,由互联网、公共移动网和专线等组成,特点是以运营商公用网络承载为主,通过调研发现,专线缺乏“随选”能力,智能设备/产品远程接入需求不能完全满足,现有的解决方式是部分企业使用非授权频段将“连接能力”扩展到公共区域。在厂内 IT 域(包含车间级和工厂级),由商业以太网、Wi-Fi、非 3GPP LPWA、RFID、蓝牙、Zigbee 以及 3GPP 蜂窝网络等承载,在工厂园区内不好布线的地方以公用承载的专用网络为主,在工厂厂区内以有线承载为主,无线应用逐步扩大,通过客户的调研反馈发现,车间有线网络质量不高,Wi-Fi 网络可靠性不高,因此工业 PON 等带动有线网络开始升级,少量大型工厂自建专用蜂窝移动网络。在厂内 OT 域(包含现场级和控制级),由现场总线、工业以太网、Wi-Fi、非 3GPP LPWA、工业 WSN 和 5G 等组成,对于传输时延、同步精准度要求较高,本质安全,总线供电,通信环境恶劣,目前呈现以有线为主、无线为辅的特点,面对远程维护和纵向集成需求,面对现场总线带宽不足、在短期内标准过多无法形成统一、部署复杂等问题,未来需要实现具有时延敏感性的工业以太网的统一网络。IT-OT 融合体系架构如图 8 所示。

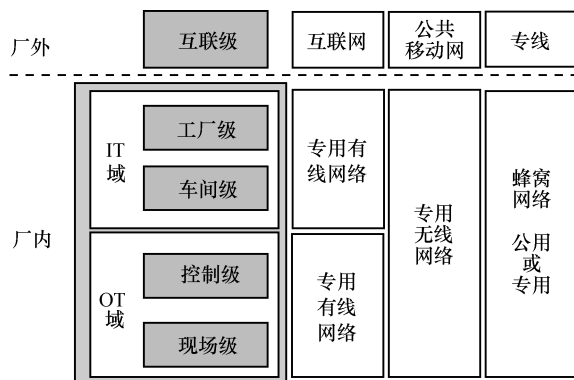


图 8 IT-OT 融合体系架构

3 工业网络技术体系

工业通信网络技术已逐步应用于工业控制领域,并呈现高速发展趋势。M2M 通信技术服务于生产线的制造协同、智能原材料的识别和定位跟踪、设备交互之后的在线数据采集和在线服务,常用工业通信技术为有线和无线通信技术。有线通信技术可满足未来智能设备之间双向交互、高宽带、低时延等需求,已广泛应用,现阶段有线通信技术由现场总线、工业以太网、实时以太网、TSN (time-sensitive networking, 时间敏感网络) 和工业 PON 等主流技术组成,基于无线技术的工业通信网络按照距离和速度可分为以 2G/3G/LTE/5G 网络和 LPWA (low power wide area, 低功耗广域) 网络为代表的广域网、以 Wi-Fi 专网为代表的无线局域网和 RFID (radio-frequency identification, 射频识别)、基于 IEEE 802.15.4 标准的无线个域网。

现场总线主要应用在长距离、短时延、信号稳定的工业控制系统里,主要用于分布式控制之间的控制信息传输,比如电机、电流、电压和位置等跟设备状态和控制相关的信息,加上配置和报警的信息。目前应用较多的有 Profinet、Ethernet/IP、Ethernet PowerLink、EtherCAT、CAN 总线 (ISO 11898)、LonWorks、HART。CAN 总线由德国 BOSCH 公司推出,广泛应用于汽车制造业以及航空工业;LonWorks 总线标准由美国 Echelon 公司推出,被广泛应用在楼宇自动化、家庭自动化、工业过程控制等行业。其余主要的现场总线标准在 IEC 61158 中规定,长期以来,关于现场总线的问题争论不休,互连、互通和互操作问题很难解决,传统的现场总线最高速率只有 12 Mbit/s (如西门子 ProfiBus-DP),于是现场总线开始转向工业以太网。

工业以太网与普通商业以太网兼容,但在产品设计、材质选用、产品强度、适用性和实时性



方面能够满足工业现场的需要。根据实现方式的不同,可将工业以太网分为 3 种类型:通用硬件、标准 TCP/IP 型(以 Ethernet/IP、ModBus/TCP 等为代表);通用硬件、自定义实时数据传输协议型(以 Ethernet PowerLink、EPA、Profinet RT 等为代表);专用硬件、自定义实时数据传输协议型(以 EtherCAT、Profinet IRT 等为代表)。类型 1 完全兼容通用以太网,成本低廉,实现方便,但只适用于实时性要求不高的工业过程自动化应用;类型 2 采用通用以太网控制器,但却引入了专门的实时数据传输协议,使用特定以太网帧进行传输,实现较高的实时性,这种结构对以太网的读取受到实时层的限制;类型 3 在类型 2 的基础上底层使用专有以太网控制器(至少在从站侧),在实时通道内由实时 MAC 接管通信控制,彻底避免报文冲突,简化通信数据处理,其缺点是成本较高,需使用专有协议芯片、交换机等,但优点是实时性强。IEC 61158 协议新旧版本对比如图 9 所示。

类型	技术名称	类型	技术名称
类型1	TS 61158现场总线	类型1	Foundation FieldBus
类型2	CIP现场总线	类型2	CIP
类型3	ProfiBus现场总线	类型3	ProfiBus & Profinet
类型4	P-NET现场总线	类型4	P-NET
类型5	FF HSE高速以太网	类型5	WorldFIP
类型6	SwiftNet被撤销	类型6	InterBus
类型7	WorldFIP现场总线	类型7	由于市场原因删除
类型8	InterBus现场总线	类型8	CC-Link
类型9	FF H1现场总线	类型9	HART, Wireless
类型10	Profinet实时以太网	类型10	HART
类型11	TC-net实时以太网	类型11	VNET/IPTC-net
类型12	EtherCAT实时以太网	类型12	EtherCAT
类型13	Ethernet PowerLink实时以太网	类型13	Ethernet PowerLink
类型14	EPA实时以太网	类型14	EPA
类型15	ModBus-RTPS实时以太网	类型15	ModBus-RTPS
类型16	SERCOS I, II现场总线	类型16	SERCOS
类型17	VNET/IP实时以太网	类型17	RAPIEnet
类型18	CC-Link现场总线	类型18	SafetyNET p
类型19	SERCOS III实时以太网	类型19	MECHATROLINK
类型20	HART现场总线		

(a) IEC 61158 Ed.4 2007

(b) IEC 61158-1 Ed.1 2014

图9 IEC 61158 协议新旧版本对比

在一些条件苛刻、无法布线的现场区域,工业无线技术已成为工业以太网强有力的延伸手段,主要包括基于 IEEE 802.15.4 标准的用于工业过程测量、监视与控制的 WirelessHART (IEC 62591:2016)、ISA100.11a (IEC 62734:2014)、

WIA-PA (wireless networks for industrial automation-process automation,用于工业过程自动化的无线网络)(IEC 62601:2015)等工业级无线传感器网络标准。

在工业网络中,RFID 技术是利用射频信号通过空间耦合(交变磁场或电磁场)实现非接触式的无线射频自动识别技术。与条形码相比,RFID 标签在智能制造环境下有着不可比拟的优势,如防水、防磁、耐高温、使用寿命长、读取距离大、标签上数据可以加密、存储数据容量更大、存储信息可在信号激活后进行更改、可以透过外部材料读取数据、能够同时处理多个标签、可以对所附着的物体进行追踪定位等。以世界著名的米其林轮胎为例,北美公司推出的带有 RFID 标签的轮胎被写入轮胎的序列号、生产日期、生产厂商代号等信息,一旦轮胎出现质量问题,可以缩小召回的范围。除此之外,RFID 技术还可以贯穿服务使用的过程,2016 年 8 月阿里巴巴 YunOS 在 ITU-T SG20 物联网研究组新立项的 Open IoT Identity correlation service (开放 IoT 身份标识协作服务),旨在打通设备、数据与人之间的协作关联,目标是可以满足 IoT 应用领域对设备的唯一安全身份标识,相应地,其连接的业务平台可以提供给信任设备相应的服务。最后,RFID 技术可以提供生产制造控制系统、生产制造执行系统和管理信息系统的服务信息,以汽车生产过程管理应用为例,戴姆勒—克莱斯勒将 RFID 标签附在汽车的底盘上,能够自动传输数据,比如要喷涂的颜色。

为满足工业应用需求并顺应技术发展,工业通信网络历经现场总线、工业以太网和工业无线网络 3 个发展阶段,实现控制级/现场级专用网络和管理级通用网络的融合以及有线到无线的补充。并且,由于数字化通信的特点,工业通信网络除基本数据传输功能外,还提供信息获取(包含位置)、标识和维护、报警诊断等智能化功能。然而,

由于传输介质不同,或者不同族的现场总线/工业以太网技术,工业通信网络具有几个共性特征:通信协议的定义都是基于 ISO/OSI 模型的 7 层通信协议;将网络技术实际应用于工业过程控制中,从通信角度来说,必须考虑数据传输的实时性、确定性、可靠性等要求;从功能完备性角度来说,在满足基本功能基础上还需考虑互操作性、本质安全性、功能安全性、通信安全性、高可用性、集成性等要求。

4 面向制造业未来发展的网络升级需求

工业通信网络需要一张无所不在的宽带网络,能把人、物、数据、流程全部连接起来,应用场景主要包含 3 类:广域应用场景、工厂级应用场景和现场级应用场景。广域应用场景主要指跨域的多工厂之间的网络通信和应用,包括多厂之间的广域网络访问和通信、协同设计、供应链协作、多厂间物流等。现有多工厂之间的广域网络通信一般采用互联网、专线网络或 VPN 虚拟专网方式,通过整机厂和周边零配件分厂之间的供应链协作系统,整机厂供应链管理人员可以通过电脑或手机远程访问周边零配件分厂的供应计划,周边零配件厂供应链管理人员也可以通过电脑或手机远程访问整机厂的需求计划。通过广域通信将生产线和辅助配套设备与供应商互联起来,各方能洞察相互之间的依赖关系、物料流动和制造周期。支持位置跟踪、远程监控资产运行状况、报告流经供应链的零件和产品的情况等诸多功能。工厂级应用场景主要包含移动办公应用、移动 MES 应用、安全管理(无线视频监控和无线巡检)、节能管理、集群通话和厂区内智能物料运输和配送等。传统的工厂级通信网络一般采用以太网+Wi-Fi 的覆盖方式,在工厂内存在大量线缆,部署复杂、施工周期长、维护成本高、能耗和空间占用大,且厂区 Wi-Fi 存在网络覆盖不全面、多 Wi-Fi 部署、网络信号不稳定和安全性不

能得到保障等问题。智能工厂现场级应用场景的生产过程包括数据采集及分析,可以实现生产过程、设备、资源监控的可操作和可视化,要能支持采集不同现场设备数据的要求,支持将生产数据及设备故障信息显示在监控站的屏幕上,实现生产过程的动态监控与管理。现有现场级通信网络一般采用工业控制总线,与厂区以太网安全隔离,具有可靠性高、安全性好和实时性好等特点。但是存在如下局限性:布线成本比较高,有些地方不好布线,工业控制总线数据采集不够全面,组网方式比较简单,新的应用场景需要更多样的拓扑结构。

以 LTE 网络为代表的高速连接网络可基本解决工厂级应用场景下广域网和厂域网的需求。现阶段部分工业企业开始尝试在厂区内自行搭建一张 LTE 专用网络来满足工业制造的需求,通过新建一张 LTE EPC 核心网,分别与企业内部各类应用系统进行互通,或者新建一张 LTE 无线接入网,采用特殊申请无线 LTE 频段覆盖整个工业厂区。但是这些方式都存在一些局限,如建设成本大、后期维护成本持续投入、LTE 无线频段资源难以获取、特殊无线频段的定制化终端成本较高等问题,难以在工业领域广泛推广,需要采用已规模部署的公众 LTE 网络,利用现有网络架构与工厂园区的基站将数据流量接入工厂网络,用户面的业务数据不经过网络侧(如核心网)而直接在本地传输。

LPWA 技术主要适用于 M2M 小流量数据传输需求,3GPP 已成立专门项目组开展了 LPWA 技术研究,同时,一些在非授权频段上定义的接入技术已经在欧洲投入商业运营,用户数正在逐步扩大。目前可以纳入 LPWA 范畴的无线接入技术主要分成几块:一是 3GPP RAN 中基于 LTE/EPC 网络优化的 eMTC;二是 3GPP GERAN 提出的基于 GSM 演进的 EC-GSM (extended coverage-GSM,扩展覆盖 GSM)技术;三是 3GPP RAN 融合由华为、高通、Neul 联合提出 NB-ClOT 及由爱立信、诺基亚等提出的



NB-LTE 的 NB-IoT 技术；四是在非授权频段上定义 Sigfox、LoRa 等接入技术。LTE-M (LTE-machine-to-machine) 是基于 LTE 演进的物联网技术，在 R12 中被称为 Low-Cost MTC，在 R13 中被称为 LTE enhanced MTC (eMTC)，旨在基于现有的 LTE 载波满足物联网设备需求。在工业应用中，低功耗的 LPWA 网络可以快速建立工厂运行设备的参数采集系统，无需用户进行大规模的信息化改造就可以实现设备实时报警、状态分析、故障诊断等多类功能。

在工业领域中，5G 可以提供高效率、高质量、低成本的网络连接。IHS 2017 年的 5G 经济报告显示，2035 年，5G 使能的工业制造空间高达 3.4 万亿美元。在智能工厂里，5G 网络能够灵活地支撑智能制造的多业务场景，包括机械臂协同控制所需毫秒级的极低时延 (uRLLC)、先进的生产辅助系统应用增强现实所需的高带宽与毫秒级的低时延 (MBB+uRLLC)。在 2017 年 3 月巴塞罗那世界移动大会上，德国电信与华为联合开发的机械臂接球用于展示超低时延切片在工业制造中的应用，该切片基于对接入、传输及核心网的域内及域间编排控制技术以及边缘计算部署来达到 ms 级的时延保障，最高可达 1 ms。总之，通信运营商已经逐步开展 5G 在工业领域的网络应用场景及相关技术研究。

针对不同应用场景，对网络的 QoS、可靠性、安全性要求不同。对于广域网络，通信实时性要求相对低、数据交互一般、带宽要求一般、可靠性要求一般、安全性要求高；工厂级通信实时性要求相对低、数据交互频繁、带宽要求高、可靠性要求高、安全性要求高；现场级通信对于实时性要求相对高（毫秒级的时延）、可靠性要求高、抗干扰性要求好、安全性要求高。

5 通信运营商的工业网络发展方向

工业通信网络的目标总体架构主要呈现 4 个

关键特征：一是在体系架构方面，实现层级打通、内外融合，传统工业系统多层结构逐渐演变为应用层、平台层和边缘层 3 层，整体架构呈现扁平化发展趋势；二是在网络互联方面，各种智能装备充分实现网络化，无线成为有线的重要补充，新型网关推动异构互联和协议转换，工厂与产品、外部信息系统和用户充分互联；三是在数据智能方面，工业云平台成为关键核心，实现工厂内外数据的充分汇聚，支撑数据的存储、挖掘和分析，有效支撑工业信息控制系统和各种创新应用；四是在安全保障方面，各种安全机制与工业互联网深度融合，实现纵深防御、立体防护，通过多种安全措施保障网络互联和数据集成安全。工业互联网目标架构的实现将是一个长期过程，需要架构、网络、数据、安全等方面逐步协同推进。

- 从提出的 IT 和 OT 融合架构出发，由于 IT 和 OT 部分功能融合，OT 部分建议运营商逐步切入，并考虑在发展过程中与现有现场总线和工业以太网共存发展，同时加强工厂外部的广域互连能力。
- 大网延伸，虽然具备一网到底能力，构建虚拟工厂专网等。
- 通信运营商除了服务智能产品的移动 M2M 可管理通道、互联企业信息系统的广域有线通道、4G/5G 网络服务下沉到生产现场之外，也可以考虑管道深度嵌入工业应用。
- 网络使用体现“专网”效果，确保网络质量，并不断满足客户日益增长的需求，做到弹性网络、实时配置等。
- 面向不同接入需求，提供综合承载能力：高带宽宽覆盖、低时延高可靠、低功耗低成本等。

总体发展策略应以完善广域互联能力为主，服务于横向集成，同时切入工厂厂域网络，服务于纵向集成。对于网络发展建议面向不同接入需求，同时管道深度嵌入工业应用，联网到底。技

术研发则以 LPWA + 5G、基于公网的工厂虚拟专网技术和随选网络为主要发展方向。

6 结束语

总之,工业通信网络是对于可靠性和时延要求很高的网络,随着 5G 网络的不断发展,通信运营商的身份和信息化的专业经验决定了其现在及未来会积极主动地参与到工业网络中。

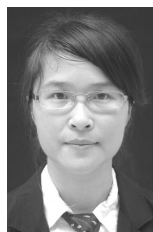
参考文献:

- [1] 卢云. 电信运营商物联网“端、管、云、应用”发展策略[J]. 电信科学, 2016, 32(Z1): 129-136.
LU Y. Development strategy of telecom operator's internet of things “terminal, network, cloud and application”[J]. Telecommunications Science, 2016, 32(Z1): 129-136.
- [2] 杨思维. 升级版德国“工业 4.0 平台”经验对我国制造业的影响[J]. 电信科学, 2016, 32(1): 108-111.
YANG S W. Impact of the upgraded version of German “industrial 4 platform” on China's manufacturing industry[J]. Telecommunications Science, 2016, 32(1): 108-111.
- [3] 崔云峰, 钟卫东, 刘东. 智能工业装备操作系统技术及创新[J]. 中兴通讯技术, 2017, 23(6): 1-9.
CUI Y F, ZHONG W D, LIU D. Intelligent industrial equipment operating system technology and innovation[J]. ZTE Technology Journal, 2017, 23(6): 1-9.
- [4] 闫建, 李瑞, 郑迦元. “互联网+”行动计划比较研究——以 22 个省(自治区、直辖市)为例[J]. 重庆理工大学学报(社会科学), 2017, 31(4): 65-72.
YAN J, LI R, ZHENG J Y. A comparative study of “Internet plus” action plan: taking 22 provinces (autonomous regions and municipalities directly under the central government) as an example[J]. Journal of Chongqing Institute of Technology, 2017, 31(4): 65-72.
- [5] 贾洪哲, 沈英儒. 浅析 ISA-95 和 MES 系统[J]. 中国信息界, 2011(1).
JIA H Z, SHEN Y R. Analysis of ISA-95 and MES system[J]. China Information Times, 2011(1).
- [6] Converged plantwide ethernet (CPwE) design and implementation guide[EB]. 2011.
- [7] LU Y, MORRIS K C, FRECHETTE S P. Current standards landscape for smart manufacturing systems[EB]. 2016.
- [8] 欧阳劲松, 刘丹, 汪烁, 等. 德国工业 4.0 参考架构模型与我国智能制造技术体系的思考[J]. 自动化博览, 2016(3): 62-65.
OUYANG J S, LIU D, WANG S, et al. Reflections on German industry 4.0 reference architecture model and China's intelligent manufacturing technology system[J]. Automation Panorama,

2016(3): 62-65.

- [9] 工业和信息化部. 国家智能制造标准体系建设指南(2015 年版)[EB/OL]. (2015-12-30)[2017-10-30]. <http://www.miit.gov.cn/n1146285/n1146352/n3054355/n3057585/n3057589/c4570069/content.html>.
Ministry of Industry and Information Technology. National smart manufacturing standards system construction guide (2015 edition)[EB/OL]. (2015-12-30)[2017-10-30]. <http://www.miit.gov.cn/n1146285/n1146352/n3054355/n3057585/n3057589/c4570069/content.html>.
- [10] 工业互联网产业联盟. 工业互联网体系架构(版本 1.0)[EB/OL]. (2016-09-09)[2017-10-30]. <https://wenku.baidu.com/view/43894ef9f12d2af90342e66a.html>.
Alliance of Industrial Internet. Industrial internet architecture (version 1.0)[EB/OL]. (2016-09-09)[2017-10-30]. <https://wenku.baidu.com/view/43894ef9f12d2af90342e66a.html>.

[作者简介]



常洁 (1984-), 女, 博士, 现就职于中国电信股份有限公司上海研究院, 主要研究方向为工业网络、5G 等。



王艺 (1973-), 男, 中国电信股份有限公司上海研究院首席物联网专家, 主要研究方向为物联网。



李洁 (1980-), 男, 中国电信股份有限公司上海研究院工程师, 主要负责车联网及工业互联网领域的产品开发、技术研究工作。



陈正文 (1973-), 女, 中国电信股份有限公司上海研究院高级工程师, 主要研究方向为物联网蜂窝通信技术。



工业互联网中的标识解析技术

杨震¹, 张东², 李洁¹, 张建雄¹

(1. 中国电信股份有限公司上海研究院, 上海 200122;

2. 中国电信集团公司政企客户事业部, 北京 100010)

摘要: 工业互联网的核心是数据的价值发现问题, 但由于历史原因, “信息孤岛”现象在企业内部、企业之间大量存在。标识解析技术是目前可见解决“信息孤岛”、完成工业大数据汇聚以及在此基础上形成信息融合理解的关键技术。分析了标识解析在工业互联网领域应用要解决的几个关键环节, 并且给出了进行工业互联网数据理解的研究思路。

关键词: 工业互联网; 标识解析; 自然语言理解; 工业大数据

中图分类号: F49

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017296

Identifier technology in industrial internet

YANG Zhen¹, ZHANG Dong², LI Jie¹, ZHANG Jianxiong¹

1. Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China

2. Government and Enterprise Customer Department of China Telecom Corporation, Beijing 100010, China

Abstract: The value discovery of data is the core of industrial internet, but for historical reasons, there are many information islands exist in enterprises. The identifier is the key technology for solving information isolated island and industrial big data gathering. Also, the identifier is the basis technology of understanding industrial big data. Several key elements in the application of identifier in industrial internet were analyzed, and some research ideas for the future study of industrial big data's understanding were given.

Key words: industrial internet, identifier, natural language processing, industrial big data

1 引言

物联网的发展需要解决两个基本问题, 一个是万物互联的通信技术问题; 另一个是联网物体的价值发现问题。而第二个问题是物联网发展的核心, 即为什么万物互联, 或万物互联能产生什么价值。智能制造、工业互联网是物联网的最重要组成部分, 随着世界经济格局的变化, 其越来越

受到各国的重视, 以中国、美国、德国为代表的世界重要经济体都在该领域加大投入, 期望利用先进技术改善传统制造业, 提升生产效率和本国的基础制造业竞争力。美国有先进制造、工业互联网; 德国有工业 4.0/数字化战略 2025; 而我国将“中国制造 2025”定位于国家战略高度, 旨在推动互联网与工业融合创新, 通过网络融合发展, 促进新技术在传统制造行业进行应用。未来

工业互联网关键技术及应用方向之一,就是智能制造所涉及的各类割裂数据的理解及应用,而要理解数据,就要对数据的来源、流动过程、用途等进行了解并掌握,目前可见,唯一有希望满足这个任务的是标识解析技术。因此,在工业互联网的各参与方中,标识解析技术越来越受到重视。

本文首先介绍了标识解析的基础概念、主流技术,然后针对物联网、工业互联网数据内容理解的需求,从实现设备互联互通互操作角度,提出了标识解析技术在工业互联网领域的标准化思路,最后根据面向数据融合理解的目标,给出了标识解析技术可能的应用架构,并进行了相关分析和介绍。

2 标识解析技术介绍^[1-5]

2.1 标识解析技术简介

标识可以理解为用于识别不同物品、实体、物联网对象的名称标记,可以是由数字、字母、符号、文字等以一定的规则组成的字符串。目前主流的标识技术有 Handle、OID (object identifier, 对象标识符)、Ecode (entity code for IoT, 物联网统一标识体系)、Epc、UCode 等,分别由不同的组织机构提出,其出发点都是面向物品对象、数字对象等进行唯一标记及提供信息查询的功能,进而发展成一种底层的信息架构,某些方面类似于互联网中的 DNS。现简述如下。

- **Handle:** 由 TCP/IP 的联合发明人、有“互联网之父”之称的 Robert Kahn 博士发明,核心系统由美国国家创新研究所 (CNRI) 开发。Handle 在全球设立若干的根节点,根节点之间平等互通,这也是其目前受到各国重视的原因之一;其另外一个特点是提供部分用户自定义的编码能力,用户可以根据实际需求在编码体系的部分字段自定义编码规则。目前 Handle 技术在国内已经成功

应用在产品溯源、数字图书馆等领域。

- **Ecode:** 我国自主研发的标识编码技术,其由 Ecode 编码、数据标识、中间件、解析系统、信息查询和发现服务系统、安全保障系统等部分组成。Ecode 目前在国内已经成功应用于农产品质量溯源等领域。
- **OID:** 由 ISO/IEC、ITU 国际标准组织共同提出的标识机制,用于对实体及数字等对象、概念或者“事物”进行全球无歧义、唯一命名。一旦命名,该名称终生有效。OID 在医疗卫生领域、信息安全等领域获得较多的应用。

此外,还有其他的标识技术,限于篇幅这里不做详细叙述。随着工业互联网的发展,人们发现标识技术是适用于工业互联网各个领域的基础共性支撑技术,可以解决“信息孤岛”问题,突破不同领域间的信息壁垒,作为工业大数据、设备智能化等目标的基本支撑技术。

2.2 对标识技术在工业互联网应用的理解

标识的本质是用于识别对象的技术(包含实体对象、虚拟数字对象等),以便各类信息处理系统、资源管理系统、网络管理系统对目标对象进行相关管理和控制。工业互联网中的标识解析,除了识别网络内各个实体的作用外,并在此基础上增加了查询实体关联信息的功能。工业互联网标识解析技术的变化在于,一是颗粒度从主机细化到了物品、信息、服务等资源;二是功能可支持对异主、异地、异构的信息实现智能化关联。

在实际中,标识从应用中一般分为两类:一类是有直接可理解含义,如道路标识等;另一类是用来唯一识别一个“对象”,而不关注这个对象的用途如何,用户的业务实现由各类解析系统及基于解析系统的业务应用系统完成。业务系统在完成定义的功能中,通常把标识的解析信息与其他信息相关联,实现业务需求所需的功能,如



关联对象的时间、空间信息后进行产品追溯等。

第二类标识技术是本文主要研究的内容。物体标识技术构成有 3 个基本元素：标识编码技术、标识载体技术和标识解析技术，介绍如下。

- 标识编码技术是将事物或概念赋予一定规律性的、易被人或机器识别和处理的数字、符号、文字或任意可以被机器识别的混合信息符号等。编码是人们统一认识、统一观点、交换信息的一种基本技术手段。编码的目的在于提高信息处理的效率，但目前不同标识体系的编码规则不统一，造成“信息孤岛”及信息融合、理解及基于此的新应用的开发困难。
- 标识载体技术规定了标识数据在“载体”上以何种形式存储和读取。离线载体有条形码、二维码、RFID、NFC 等，在线载体则以各种文本或二进制数字信息存在。需要说明的是，在数字空间中通常不需要物理世界的标识载体。
- 标识解析技术是指将对象标识映射至实际信息服务所需的信息的过程，如地址、物品、空间位置等。例如，通过对某物品的标识进行解析，可获得存储其关联信息的服务器地址。标识解析是在复杂网络环境中，能够准确而高效地获取对象标识对应信息的“信息转变”的技术过程。

3 面向工业互联网需求标识解析技术的标准化研究思路^[6-8]

标识解析技术本身并不是一个新技术，其最初的应用主要是解决生产者对所生产产品的管理，近些年国家层面的主要推动力量为监管部门，应用在药品、食品的安全追溯领域。但针对标识技术应用的市场自发推动力量不足，这其中核心的原因是标识解析技术所对应的信息价值没有被挖掘出来。随着物联网、工业互联网的发展，人

们逐渐认识到标识技术是物联网、工业互联网基础信息融合理解的支撑技术，是未来物联网、工业互联网大数据应用的最基本支撑技术。

但要发挥标识技术的应用价值有两个关键问题需要解决：标准化问题，各类标识解析技术面向应用的“互联互通”问题，虽然从基础层面业界提出了初步的解决方案，但核心是标识技术与实际需求结合的标准化问题；应用自然语言处理技术的标识技术演进问题，目的是实现在数据“互联互通”基础上的融合理解。

3.1 面向标识技术的标准化研究思路

智能制造的本质是以机器设备、原材料、控制系统、信息系统、产品以及人之间的网络互联为基础，通过对工业数据的全面深度感知、实时传输交换、快速计算处理和高级建模分析，实现智能控制、运营优化和生产组织方式变革。随着智能制造相关基础共性标准和关键技术的研发，面向具体行业应用的标准研制及实验验证显得越来越重要。一方面，其可以解决国内智能制造企业设备升级换代问题，提高企业智能制造水平，推动互联网与制造业融合，提升制造业数字化、网络化、智能化水平，加强产业链协作，发展基于互联网的协同制造新模式。另一方面，这是标识技术面向未来工业大数据最实际的应用需求，通过解决行业内部各类环境、应用需求的标识标准化技术，从而通过建立不同行业间数据“中间件”及相关标准化体系，解决行业间的数据汇聚、融合理解问题。而“智能装备之间、智能装备与物之间、智能装备与制造系统之间互联互通及互操作要求”的标准研制是解决以上问题的第一步。

制造企业在数据互联互通、互操作技术方面，需要针对同一个或一类物体在制造企业内部、行业内部、跨企业、跨行业信息应用过程中进行统一表达以及标识寻址。因此，有需要建设一套完整的数据解析及理解体系。但是，当前阶段，还

面临许多问题，特别是没有统一的数据表示及理解基础。首先，一个实体对象在不同行业、垂直领域的用途不同，造成这个实体对象的表达方式不同；其次，不同企业在不同时期建设信息系统的目的及用途不同，也造成了信息表达的差异性；第三，在使用信息时，存在不同的语境、上下文，这也对信息的理解有影响。总之，不同企业信息的存储、表示、应用差别巨大，无法供第三方调用。因此，应建立和完善制造行业标识解析与数据理解标准，并对关键技术进行试验验证，满足我国高端制造业升级演进的迫切需求。

我国工业和信息化部、国家标准化管理委员会于 2015 年底公开颁布了《国家智能制造标准体系建设指南（2015 年版）》（以下简称《制造标准体系指南》）^[1]，其对智能制造整体的标注化工业做出了顶层设计，结合该规划，本文提出了面向数据融合的标准化思路，如图 1 所示。

需要说明的是，本研究思路是在国家《制造标准体系指南》基础上，针对基础共性、行业应用需求思考提出的；此外，如果在通信技术层次上，同时考虑标识技术的应用需求和设备在网络层面的互联互通需求，使网络架构面向数据融合理解需求的方向演进，如进行工业 PON 技术的研发及相关标准化工作，将会从宏观层面更容易满足未来智能制造、工业互联网对信息应用的需求。在国家 2015 版《制造标准体系指南》基础上，首先需在 A 基础共性部分，进行基础资源面向数据融合理解、标识应用的共性研究，如进行标识、工业互联网设备语义、资源描述等面向具体落地及行业应用需求，进行符合企业及行业实际应用需求的标准化研究；其次，需要进行标准的行业应用研究，如面向行业个性化需求（智能设备、物、控制系统互联互通及互操作）标准需求，在基础共性标准的工作上进行面向行业、企业实际应用需求的标准化研制。

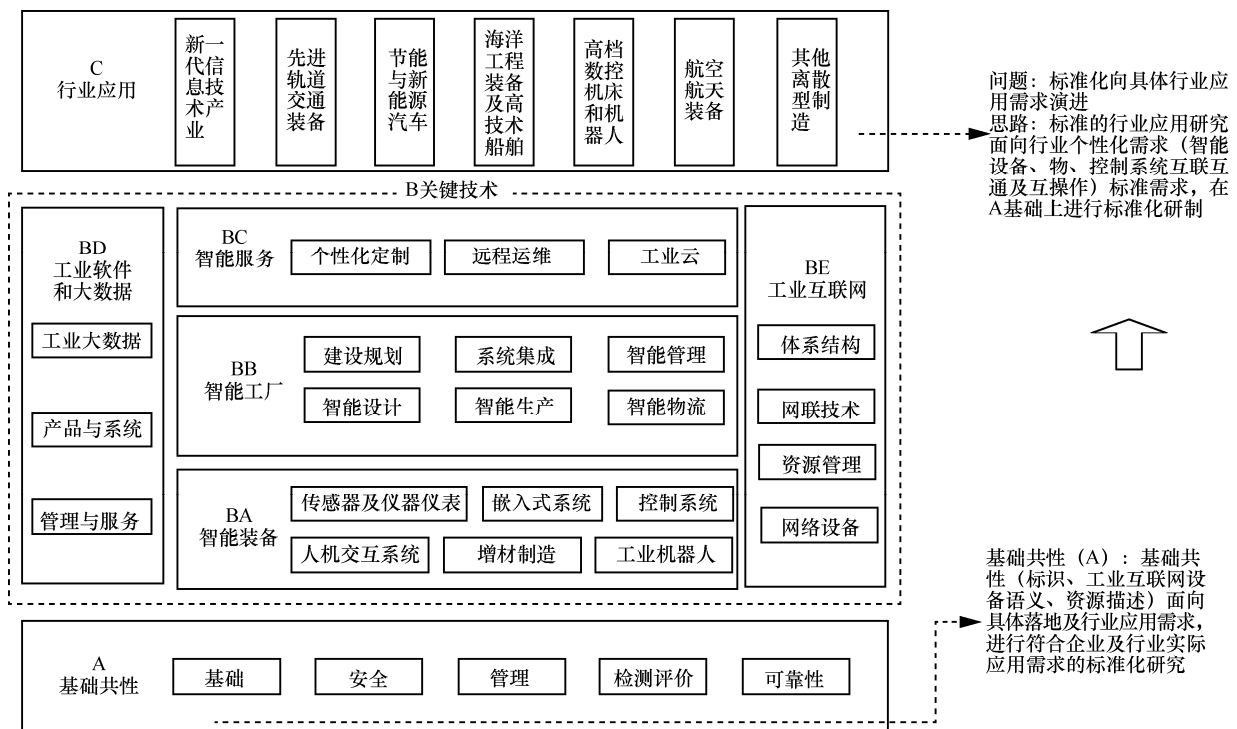


图 1 基于标识的面向信息融合理解的标准化研究思路示例



3.2 对于各类连接所传递数据的理解需求

工业大数据资源互通、共享和协作生产是工业互联网的主要诉求之一，而实现这个目标的核心是数据的理解，尤其是各类数据在语义层次上的理解。这些理解在具体行业应用中，需要具体的数据字典、数据标识、表示、理解的标准方法。数据理解是开发新应用、产生新业态的基础。目前不但在企业间，即便在同一企业内部，统一的数据表示及理解基础也存在较大欠缺。首先，一个实体对象在不同行业、垂直领域的用途不同，造成这个实体对象的表达方式不同；此外，企业内不同时期建设平台的目的及用途不同，也造成了信息表达的差异性；第三，在开发具体的应用时，存在不同的语境、上下文，这也对信息的理解有更高的需求。

工业制造企业内部数据的表达方式基本有两类，一类与互联网上的信息类似，是语言文字类型，但其特点是行业专业性强，经常存在同行业内部不同企业针对同一事物的描述表达差异的情况，更极端的情况是同一企业内部对于同一事物，分公司、分工厂或不同的需求背景下描述也有差异；第二类就是更加海量的机器数据，即企业生产经营过程中，由机器产生的各类数据。由于这些机器来源于不同制造商，遵从的标准不同，制造时期不同，所以机器所产生的数据格式、表达方式也各不相同。因此要实现互联网与工业企业的融合创新，需要强大的信息处理及支撑技术。为了实现智能装备、物、控制系统之间互联互通及互操作需求，需要一套统一的信息理解及面向这个需求的信息传递、表达的标准及处理架构。该架构主要解决的问题如下。

- 实现行业内数据表达的统一。行业、企业采用适用本行业、企业的标识技术独立存在，在体系内，实现本领域的标识数据融合，这部分不改变现在标识的运行机制及系统。
- 实现不同行业、不同标识的数据汇聚。通

过云平台汇聚各类标识技术的数据，进行各类标识数据解析后，根据需求标识之间的转换，实现基于标识的数据互通，支撑初步的数据价值发现。

- 实现工业互联网领域的人工智能技术应用。根据各类标识对应的实体、应用汇聚相关的文本信息，采用人工智能、机器学习、自然语言处理等技术对信息的内在价值进行深层次挖掘。此部分从宏观上已经可以开始与互联网信息进行融合，互联网上有大量的关于实体性质、特征、用户描述的成熟语言环境，还具有面向应用的开发及商业环境。此部分的核心技术是自然语言处理技术。基于这些基础工作，可以进行具体的工业大数据应用，如质量管理、趋势分析等。

这个体系的核心工作有两个：第一是对接各类标识独立服务系统、接口、数据交换协议、格式的设计，即标准化工作；第二是在汇聚端的自然语言处理技术的研究及应用，学术界和产业界要解决的是如何开发一套适合工业大数据理解的自然语言处理技术。面向多领域标识数据融合的逻辑设想如图2所示。

4 面向工业互联网数据融合、人工智能技术应用的关键工作

工业互联网数据融合的思想及目标可以用（cyber-physical system, CPS）的概念描述，其最早出现于2006年美国国家科学基金会（NSF）组织召开关于信息物理系统的研讨会。随着技术的发展，学术界及产业界越来越发现信息物理系统的描述很可能是未来工业互联网、互联网相互融合的直接体现，其内涵是使传统工业企业封闭的物理系统与互联网在信息层面上进行了融合，在信息融合的基础上造就新的业态，而标识技术正是实现这个目标必不可少的技术。但要实现这一

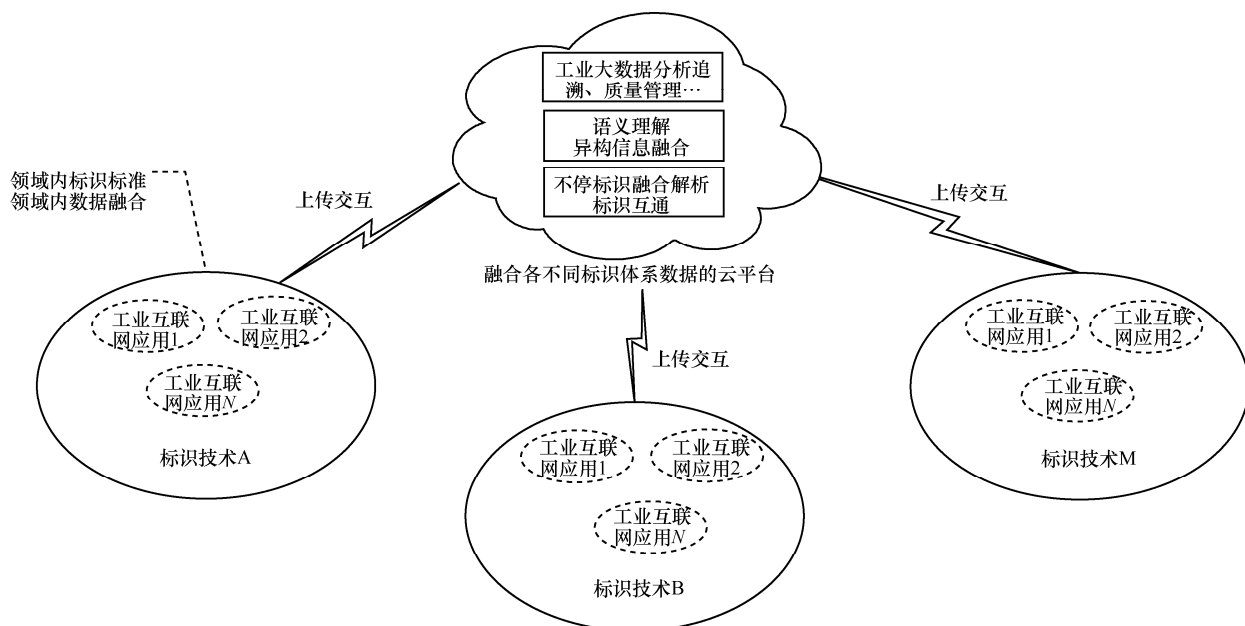


图2 面向多领域标识数据融合的逻辑设想

目标，从标识技术本身看，还要有很多工作要做。在技术研究层面还需有以下工作。

- 基于不同标识解析体系数据融合的技术研究。
- 面向物联网语义理解及语义互操作的技术研究。
- 基于标识进行行业语义定义（标准）及相应语义扩展标准化及技术研究。
- 基于标识技术支持工业互联网与互联网数据相互融合的数据开放技术研究。
- 支持基于标识解析技术智能化应用的技术研究，尤其是在信息融合基础上采用自然语言处理技术的应用研究。
- 公共云服务平台：进行标识技术开展公共服务的基础支撑平台技术研究，研究公共云服务平台技术架构及关键技术。
- 支撑标识技术的网络技术研究：结合电信网络演进，对未来网络支撑物联网标识技术的网络演进关键技术进行研究，包括传递标识关键信息的网络基础设施改造及升级、技术研究路线、增值服务和应用等。

5 结束语

互联网与工业融合创新，不仅仅是网络的互通，更重要的是网络所传递信息的融合理解及开放应用。工业互联网空间的信息采集、理解是工业互联网、智能制造发展的重要核心推动力。标识解析体系是未来工业互联网、物联网、智慧城市基础设施的重要组成部分，也是支撑网络互联互通和信息共享共用的重要基础设施。部分标识技术已经具备跨平台，初步具备了面向业务可灵活定义等特性。但在面向物联网信息空间的语义理解方面还有许多工作做，尤其是标准化方面，还需面向制造企业的实际需求，制定开放共享的标准。

参考文献：

- [1] 工业和信息化部，国家标准化管理委员会. 国家智能制造标准体系建设指南（2015年版）[S]. 2015.
Ministry of Industry and Information Technology, Standardization Administration. Guidelines for the intelligent manufacturing construction of national standard system (2015 Edition)[S]. 2015.
- [2] OID 注册中心（中国电子技术标准化研究院）. 中国对象标识符（OID）白皮书[R]. 2014.



- National OID Registration Center (China Electronics Standardization Institute). Chinese object identifier (OID) white paper[R]. 2014.
- [3] 吴双力, 闫旸. 基于物联网标识解析的药品电子监管码创新应用[J]. 首都医药, 2012(10): 11-12.
WU S L, YAN Y. Innovative application of drug electronic monitoring code based on IoT identification[J]. Capital Food Medicine, 2012(10): 11-12.
- [4] 杜娟, 王峰. 互联网的内涵、服务体系及对制造业的作用路径[J]. 电信科学, 2016, 32(1): 98-104.
DU J, WANG F. Connotation, service system of internet and its integrating path to manufacture[J]. Telecommunications Science, 2016, 32(1): 98-104.
- [5] 杨思维. 升级版德国“工业4.0平台”经验对我国制造业的影响[J]. 电信科学, 2016, 32(1): 108-111.
YANG S W. Impact of the upgraded version of German “industrial 4 platform” on China’s manufacturing industry[J]. Telecommunications Science, 2016, 32(1): 108-111.
- [6] 王志宏, 杨震. 人工智能技术研究及未来智能化信息服务体系的思考[J]. 电信科学, 2017, 33(5): 1-11.
WANG Z H, YANG Z. Research on artificial intelligence technology and the future intelligent information service architecture[J]. Telecommunications Science, 2017, 33(5): 1-11.
- [7] 杨震, 杨宁, 徐敏捷. 面向物联网应用的人工智能相关技术研究[J]. 电信技术, 2016, 8(5): 16-19.
YANG Z, YANG N, XU M J. The study on the application of artificial intelligence in the internet of things[J]. Telecommunications Technology, 2016, 8(5): 16-19.
- [8] 吴军. 大数据和机器智能对未来社会的影响[J]. 电信科学, 2015, 31(2): 7-16.
WU J. Big data, machine intelligence and their impacts to the future world[J]. Telecommunications Science, 2015, 31(2): 7-16.

[作者简介]



杨震 (1972-), 男, 博士, 中国电信股份有限公司上海研究院物联网部教授级高级工程师, 主要研究方向为人工智能、自然语言处理、搜索引擎技术。



张东 (1972-), 男, 中国电信集团公司政企客户事业部产业互联网创新发展中心主任、高级工程师, 主要从事产业互联网 (互联网+)、工业互联网和智慧城市方向的研究工作。



李洁 (1980-), 男, 中国电信股份有限公司上海研究院工程师, 主要从事车联网及工业互联网领域的产品开发、技术研究工作。



张建雄 (1969-), 男, 中国电信股份有限公司上海研究院副研究员, 主要研究方向为物联网技术与应用、工业互联网平台、工业云和工业大数据技术与应用、智能制造解决方案等。



物联网使能平台的体系结构与服务模式

彭昭

(中国电信股份有限公司上海研究院, 上海 200122)

摘 要: 现有成功的物联网应用案例多集中于局部规模的垂直领域, 且以闭环应用为主, 导致众多的物联网“应用孤岛”。互联网公司、传统设备商、云服务提供商等产业巨头纷纷构建以自身能力为核心的物联网生态圈。从物联网生态打造的关键环节平台入手, 研究物联网使能平台的技术架构和服务模式, 探讨平台的核心功能和关键技术, 并给出平台发展趋势。

关键词: 使能平台; 物联网平台; 微服务

中图分类号: TP319

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017302

Architecture and service model of IoT enabled platform

PENG Zhao

Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China

Abstract: Existing successful IoT (internet of things) application cases concentrate on the vertical field of local scale and usually focused on closed loop applications, which leading to numerous application islet. Internet companies, traditional equipment manufacturers, cloud service providers, as well as other industrial giants have built an IoT ecosystem with the core of its own capabilities. Starting with key point of building IoT ecosystem, technology architecture and service model of IoT enabled platform was studied, the core function and key technology of platform was discussed, and the platform development trend was presented.

Key words: enabled platform, IoT platform, microservice

1 引言

随着网络接入技术、云计算技术、人工智能技术的发展, 物联时代开启, 也逐步成为传统行业创新的新趋势。技术的成熟和商业模式的迭代创新推动物联网在不同行业的落地与普及, 可穿戴设备、白色家电及智能交通等应用随处可见。根据应用场景和环境的不同, 物联网终端形式各异, 导致物联网应用呈现碎片化趋势。现有成功

的物联网应用案例多集中于局部规模的垂直领域, 且以闭环的应用为主, 导致众多的物联网“应用孤岛”。物联网技术的价值不仅是将物接入云端提高自动化水平, 更是对“物”所涉及数据的深度挖掘, 实现新的业态和服务的增值。

互联网公司、传统设备商、云服务提供商等产业巨头纷纷构建以自身能力为核心的物联网生态圈, 打造自身产业链的枢纽地位, 实现自身核心价值的增值与变现。芯片、模组、网络及标准

收稿日期: 2017-09-30; 修回日期: 2017-11-10



的相应成熟推动产业的大规模发展,预计未来2~3年将成为物联网产业生态发展的关键时期。电信运营商也开始布局物联网,利用其在物联网价值链中的管道优势,快速发展物联网通道类产品,但如何在复杂的竞争态势中保证运营商在物联网领域中获取最大化价值成为全球运营商共同面临的挑战。

2 物联网生态与物联网使能

2.1 物联网生态的关键要素

物联网生态的目的是打破垂直行业的“应用孤岛”,营造物联网终端、数据和应用开放共享的环境,通过平台挖掘物联网数据深度价值,促进应用类型和应用模式的创新。构造物联网生态的关键要素包括云、管、端3个部分。

(1) 云:物联网的“大脑”

是物联网业务逻辑处理的核心单元,以云端平台的形式为物联网终端提供数据存储、远程控制甚至决策分析。

(2) 管:物联网的“神经”

是联系物联网终端与云端的重要渠道,当前主流的物联网终端接入方式包括Wi-Fi、蓝牙、Node B、eMTC、Cat1等,运营商因运营移动网络具备天然的网络优势。

(3) 端:物联网的“感知触手”

是物联网数据的重要来源,根据应用场景的不同,物联网终端形态各异,目前并没有统一的标准。

解决终端“入云”成为构建物联网生态的先决条件,多变的应用场景特性将该问题复杂化。使能平台的提出屏蔽了底层物联网终端的多样性,降低了终端入云门槛,成为物联网生态的重要研究课题。

2.2 物联网使能平台概述

业界对物联网使能平台并没有明确的定义,一般可认为是面向物联网的PaaS服务平台,以降

低物联网硬件与软件开发门槛为主要目的,通过开放API和开发工具包,提供在线的开发环境、测试环境帮助开发者简化物联网硬件与软件的开发环节与周期,同时利用开发模板、组件,实现更多创新业务逻辑的开发。物联网使能平台是水平平台,能够实现对多种不同物联网应用场景的支撑,实现对多种跨行业终端的接入。通过规模化的应用和部署,使行业、企业以及物联网应用开发者以较低的成本和更高效的方式实现物联网场景。

物联网使能平台是整个物联网价值链中数据聚合、能力聚合、资源聚合的核心,包括终端接入、终端管理、数据处理、能力开放和应用辅助开发等功能。掌握物联网使能平台成为掌握物联网生态系统话语权的关键,对于物联网使能平台的运营商而言,可以将业务范围从单纯的垂直应用提供扩展到全物联网平台服务和应用运营,实现物联网业务运营的利益最大化。

3 物联网使能平台体系框架

3.1 系统架构

依据IoT参考模型(ITU-T Y.2060),物联网架构分为应用层、业务层、网络层和感知层4层,其中,业务层对应物联网使能平台,是终端到应用的接入网关,应用通过该网关调用接口查看和控制远端的物联网终端,终端也通过该网关接入云端共享数据。因此,物联网使能平台面向应用开放能力,面向终端开放接入接口。详细的物联网系统架构如图1所示。

3.2 关键功能

(1) 终端接入

终端入云是物联网数据共享和应用使用的前提和基础,通常的解决方案是提供标准的IoT终端中间件,快速适配OS和各类厂商的智能终端。IoT终端中间件集成各层通信协议,为终端设备提供网络接入能力,屏蔽上层应用与底层终端硬件

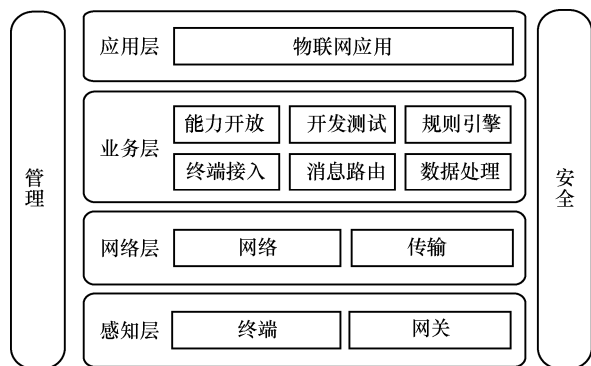


图1 物联网系统架构

的异构性。终端接入需要考虑两层的协议适配：一层是应用层承载协议，如MQTT、COAP、HTTPS等；另一层是应用层协议，如远程抄表、资产监控等涉及具体业务数据的公共或私有协议。

终端接入需要“端云”的配合，存在开放终端接入SDK和主动为终端适配两种典型的实现方式。前者需要平台提供通用的协议和终端模型，以便于应用在开发终端时能够快速适配接入，这种场景适用于通用终端接入场景；后者是面向特定的私有协议或者已经具备接入云端功能的终端，为了让其能够顺利接入物联网使能平台，一般可通过插件的形式实现对更多终端或协议的适配支持。

（2）消息路由

面对数量庞大的各种终端和物联网应用，如何保证终端的数据正确存储到对应的存储空间，发送给正确的订阅应用，如何保证应用侧和平台侧的控制指令正确地转发到对应的终端侧，是消息路由解决的主要问题。消息路由一般分为数据共享的消息路由和指令控制的消息路由两种类型。消息路由不仅需要满足海量设备的接入和处理能力，还需要考虑差异化的服务优先级和服务质量，满足不同权限等级消息的及时响应。

（3）数据处理

物联网的最大价值在于其产生的数据，根据数据形成的分析和统计结果，可作为物联网应用规划和创新的依据。物联网使能平台的数据处理

一般与云服务的数据处理相结合，通过数据的实时处理和监控为应用提供监控与告警服务，通过数据的统计和分析可总结历史规律，形成数据报告，甚至通过对脱敏数据的开放共享实现跨应用跨平台的应用模式和商业模式创新。数据处理模块包括数据存储、数据缓存、数据分析、数据开放等。

（4）规则引擎

规则引擎则是条件触发时对终端和数据执行相应的动作。数据实时处理场景中，规则引擎将某类终端数据在特定的条件下导入相关的数据处理单元，如实时分析处理模块、统计分析模块、数据呈现模块等。在告警通知场景中，终端数据达到或超过预警阈值时触发通知功能或执行模块。规则引擎包括规则的创建、修改、验证、执行等，这些能力能开放给应用直接调用。平台一般会预置通用的规则引擎，并提供自定义规则或组合规则服务供应用灵活使用。

3.3 平台核心技术

（1）微服务框架

不同的应用对应不同的终端，每个终端都有自己的应用场景和芯片模组，通过统一的框架、协议和App解决所有的连接和控制问题是不合适的，也是现阶段无法实现的。在某种程度上，服务的解耦才是物联网使能平台框架需要考虑的问题。平台应随着物联网终端的多样性发展而迭代更新，因此微服务框架成为物联网平台构建的核心架构。微服务架构将平台彻底组件化和服务化，每个组件均可独立开发、设计和运行。对于不同物联网终端接入，提供不同的终端接入微服务，对于不同的应用调用和数据共享提供相应的能力开放微服务。此外，行业知识库、行业数据分析、行业数据模板等第三方能力和服务也能够以微服务的方式引入。微服务框架为不同行业的技术与服务整合提供技术实现手段，为使能平台的构建提供灵活性和可扩展性。目前主流的微服务开发



框架有 Spring Boot、Lagom、Akka、Vert.x 和 ReactiveX。

（2）数据建模技术

面向物联网终端，解决海量终端和数据的结构问题是数据建模技术的核心。不同的终端在数据定义和数据格式上可能存在差异，这种差异为数据共享带来挑战。终端数据建模就是对物联网终端的对象、属性和操作等进行信息建模，利用标准化的方式描述实现物联网终端与数据的共享和互操作。使能平台需使用数据建模工具实现标准化的终端建模，保证终端接入时可选择平台预置模型，也可依照规则自定义实现终端建模。数据建模一般采用设备与数据流的绑定关系定义数据格式，支持终端数据点上报，在数据采集时按照数据点读取数据内容。目前，国际标准 OneM2M 已经开展了相关数据建模的研究工作。华为的 OceanConnect 平台采用 Profile 定义物联网终端的能力和服务，能力包括终端厂商、协议类型等，服务包括终端具备的某项服务或者命令、属性等。中国移动的 OneNet 通过对数据流的配置管理实现对终端元数据的管理。为了能够让平台更好地理解终端的含义，平台服务提供商均构建自己的终端建模服务。

（3）终端安全接入技术

海量的终端接入云端，终端与外部通信愈加频繁，这将对物联网系统造成安全威胁。物联网终端安全接入技术不仅需要考虑保证访问网络资源的所有设备有效的安全控制，同时也需考虑快速简便的配置接入方案。物联网是一个多网并存的异构融合网络，包括互联网、传感网和移动网等，这些网络环境各异给物联网终端接入带来安全挑战。因终端能力有限，传统基于密码的验证体系并不一定能够很好地满足安全接入的要求。终端安全接入技术需要根据不同的终端性能和应用场景安全等级进行划分。对于极小资源的设备，如传感器，通常采用预分配密钥的加密方式；对

于受限资源的设备，可采用单向认证方式；对于能力强大的设备，可采用双向认证方式。基于终端标识的应用层身份认证机制可采用 SM9、RSA 等算法来实现。

（4）数据实时处理

海量的物联网终端必定产生更加海量的数据，数据实时处理一般分为数据的实时入库和数据的实时计算。应用场景包括数据的实时变化、每天各小时的数据统计变化情况等，这些数据量大且可能并没有统计的实际价值。这类场景多用于公共服务类、资产监控类等，具备较大的规模。数据实施处理技术需要考虑流式大数据实时处理框架、分布式数据存储以及对数据密集的人工智能工作负载自适应等关键技术。在低时延和吞吐量方面，大数据、秒级时延可考虑 storm 和 flink。在数据的准确性方面，可考虑采用计算框架支持 exactly-once。

4 物联网使能平台的发展趋势

4.1 与云计算的结合

物联网的核心是实现物与物、物与人的连接，而云计算的核心是海量数据的计算。随着物联网的规模发展，未来物联网势必产生海量数据，传统的硬件架构服务器难以满足数据管理和处理要求。运用云计算模式，可对物联网中海量终端进行动态管理，数据智能分析变得可能。随着闭环物联网应用的打破，越来越多的物联网终端与数据跨行业共享，云计算将成为物联网的重要环节。通过能力的资源共享、业务的快速部署、终端的交互扩展、数据的深度挖掘等多方面带动物联网生态链的升级。

4.2 垂直细分领域

物联网使能平台是构建物联网生态的关键环节，但多从面向公共领域的通用使能入手，以解决通用物联网终端的接入和数据共享问题为主。通用使能主要面向长尾应用，能够快速聚集创新

的终端和应用,但短期并不能形成规模化效应,并不能形成一个自我运转的完善生态链。物联网具备成熟运转能力的领域多为垂直细分领域,以政府和企业应用为主,如公共设施监控、智能物流交通等。这类领域具备一定的信息化基础且规模较大,通过需求驱动能够快速落地,并形成一定的规模示范效应。因此物联网使能平台未来建议从垂直细分领域入手构建,并逐步渗透和演变成通用的使能平台。

4.3 与边缘计算的协同发展

IDC 预测到 2025 年,物联网因受限于网络带宽,50%的数据需要在网络边缘侧分析、处理与存储。物联网数据由边缘终端设备采集生成,通过“云端”平台进行数据处理和分析,这个过程可能会因为网络而不能实时响应。特别是媒体流数据,数据量大,如果数据分析和控制逻辑均在云端实现,可能并不能满足业务及时性需求。因此边缘计算将会成为云计算的补充,在工业自动化、安防监控领域尤其重要。目前,思科、亚马逊、微软都推出了相应的边缘计算及与云端协同的解决方案。随着业务的发展,减少终端与云端的交互也将成为物联网使能平台发展的一个重要趋势。

5 结束语

物联网使能平台是物联网生态构建的关键要素,是聚合物联网终端和应用的手段,也是互联

网企业、云服务提供商和电信运营商切入物联网的关键点,掌握终端接入就掌握了物联网入口。本文介绍了使能平台的体系框架,阐述了终端接入、消息路由、数据处理和规则引擎 4 个关键功能和微服务等关键技术,给出了使能平台的发展趋势,以期为研究和搭建使能平台提供参考和建议。

参考文献:

- [1] 吴思齐. 物联网资源建模工具的设计与实现[D]. 北京: 北京邮电大学, 2014.
WU S Q. Design and implementation of internet of things resource modeling tools[D]. Beijing: Beijing University of Posts and Telecommunications, 2014.
- [2] 雷吉成. 物联网安全技术[M]. 北京: 电子工业出版社, 2012.
LEI J C. Technology of security of internet of things[M]. Beijing: Publishing House of Electronics Industry, 2012.
- [3] 荣浩, 江凌云. 物联网开放服务平台[J]. 电信科学, 2016, 32(3): 120-125.
RONG H, JIANG L Y. Open service platform for IoT[J]. Telecommunications Science, 2016, 32(3): 120-125.
- [4] 陆钢, 梁柏青, 杨新章, 等. 基于云计算 PaaS 模式的产业互联网应用孵化平台[J]. 电信科学, 2015, 31(7): 119-124.
LU G, LIANG B Q, YANG X Z, et al. Industrial application incubation platform based on PaaS model of cloud computing[J]. Telecommunications Science, 2015, 31(7): 119-124.

[作者简介]



彭昭(1983-),男,博士,中国电信股份有限公司上海研究院高级工程师,主要研究方向为物联网平台。



面向智能制造的工业连接现状及关键技术分析

李洁¹, 张东², 常洁¹, 杨震¹

(1. 中国电信股份有限公司上海研究院, 上海 200122;

2. 中国电信集团公司政企客户事业部, 北京 100010)

摘要: 随着美国工业互联网、德国工业 4.0 以及中国制造 2025 的提出, 互联网、信息化与工业融合已成为实现智能制造的大趋势, 而工业连接是实现智能制造的基础能力。基于此, 首先介绍了工业领域 OT 与 IT 网络连接技术的现状, 然后结合中国制造 2025, 对工业互联网的八大场景需求进行分析总结, 并对工业连接的数据采集、IP 网络、无线网络、标识解析以及网络安全等方面的关键技术进行分析, 最后结合国内现状, 对工业连接的未来趋势进行展望。

关键词: 智能制造; 工业连接; 工业互联网; 中国制造 2025

中图分类号: F49

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017295

Situation description and critical technology analysis for industrial connection of intelligent manufacturing

LI Jie¹, ZHANG Dong², CHANG Jie¹, YANG Zhen¹

1. Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China

2. Government and Enterprise Customer Department of China Telecom Corporation, Beijing 100010, China

Abstract: With the formally issue of the industrial internet in USA, the industry 4.0 in German and made in China 2025, the integration of the IT, Internet and industrialization has become the trend of intelligent manufacturing. Industrial connection is the basic ability for intelligent manufacturing. Firstly, the technology situation of OT and IT industrial connection was described. And then combined with made in China 2025, the eight demands for industrial internet scenarios and industrial connectivity key technology were analyzed, including data acquisition, IP network, wireless network, identifier resolution and network security. Finally, according to the domestic situation, the future trend of industrial connection was prospected.

Key words: intelligent manufacturing, industrial connection, industrial internet, made in China 2025

1 引言

随着美国工业互联网与德国工业 4.0 计划的提出, 我国也于 2015 年提出了“中国制造 2025”, 并将其定位于国家战略高度。2015 年, 国务院先

后出台了《中国制造 2025》《国务院关于积极推进“互联网+”行动的指导意见》《工业和信息化部关于贯彻落实<国务院关于积极推进“互联网+”行动的指导意见>的行动计划(2015-2018 年)》等一系列文件。目的在于互联网与制造业融合进一

收稿日期: 2017-09-30; 修回日期: 2017-11-10

步深化,制造业数字化、网络化、智能化水平显著提高。

工业化与信息化及互联网的融合是实现智能制造的基础,随之也带来了对工业连接需求的进一步提升。通过工业连接可以实现工业领域全系统、全设备、全产业链的互联互通,实现工业数据的无缝集成,从而为工业智能应用与新型商业模式创新建立基础。到 2017 年,全球工业网络新增节点数将达到 1.4 亿个,年增长率超过 7%(根据埃士信咨询公司(HIS)估计)。近年来,我国工业网络节点数增长迅速,预计将从 2012 年的每年 1 800 万个节点增长至 2017 年的近 4 000 万个节点,累计的工业网络节点数总量超过 2 亿个,年增速超过 25%。以工业连接为基础的各种应用也发展迅速,根据预测,到 2020 年,我国预测性维护市场规模将达到 300 亿元,前景非常广阔。工业产品的智能化、网络化需求也在反向驱动整体工业连接的演进。

2 工业领域连接技术现状

工业领域的连接技术分为两个层次:操作技

术(OT)网络与信息技术(IT)网络。工业领域连接架构如图 1 所示。

OT 网络主要位于工厂内的现场级及车间级,对应 ISA-95 架构中的 0~2 层。现场级网络主要指底层 I/O、可编程逻辑控制器(PLC)等控制单元,工业级计算机(IPC)等上位机的互联,此类连接需要在保障可靠性、时延/抖动性能的同时大幅提升带宽和规模,确保不同协议间数据互通和物理互通。其主流技术为现场总线、工业以太网及工业无线技术。车间网络主要指车间级别内的实时运行监控与控制,包括人机界面接口(HMI)、数据采集与监视控制系统(SCADA)等。此类连接需要确保数据传输的可靠性,同时随着采集数据的剧增,对带宽要求较高。此外,由于采集接口与协议的多样性,需要适配多种采集接口协议。

OT 网络技术由于其历史原因,形成了各个制造业巨头间相互独立的多个技术标准,根据 2014 年 5 月发布的最新版的 IEC 61158,共采纳了 19 种主要类型的现场总线、工业以太网/实时以太网技术。工业总线技术从 20 世纪 80 年代开始,经过 30 多年的发展,在工业控制领域占有绝

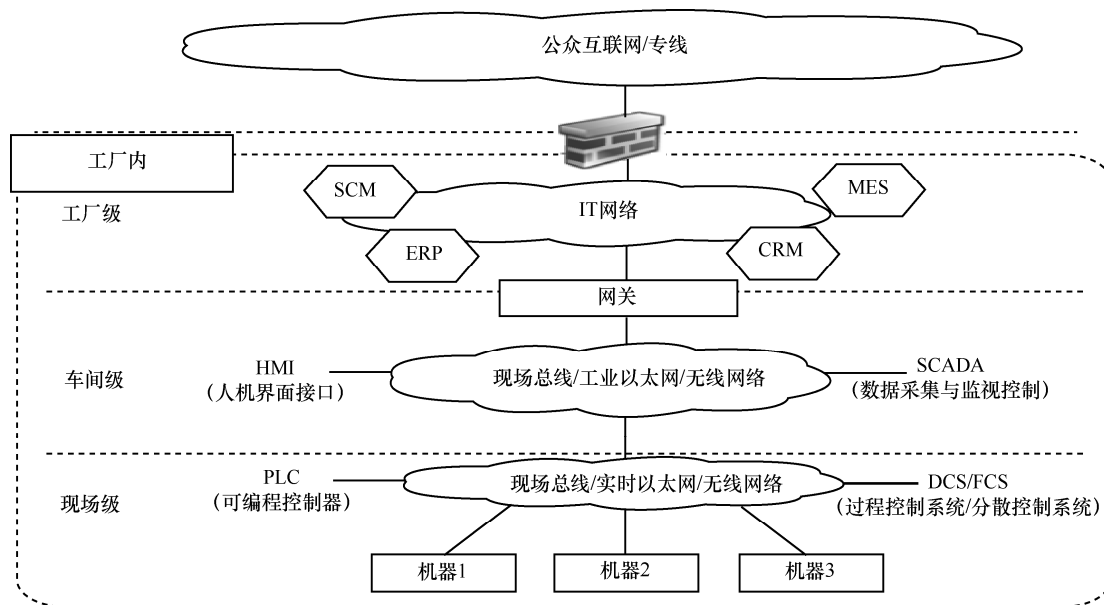


图 1 工业领域连接架构



大部分市场份额，主要技术包括西门子 Profibus、博世 CAN、FF 基金会的 FF、DeviceNet 等。随着工业信息技术的不断发展，现场总线正在向着兼容以太网通信技术的方向发展，工业以太网技术应运而生。工业以太网使得 IP 技术得以延伸至工业控制层面，使得 OT 网络与 IT 网络的融合成为可能。但由于工业控制对时延及抖动要求较高，导致各厂商对工业以太网技术中的数据链路层进行了私有化的重新定义，各标准间差异依然较大，形成了以西门子 Profinet、罗克韦尔/施耐德 Ethernet/IP、PowerLink、EtherCAT 等多种标准体系。工业无线网络作为有线网络的补充，可以有效延伸工业网络的控制范围，并提供了极高的灵活性，主要技术包括 WirelessHart、ISA100.11a 以及由沈阳自动化研究所提出的、拥有自主知识产权的 WIA-PA/FA 等。

IT 网络主要位于工厂级及工厂间网络互联，对应 ISA-95 架构中的 3~4 层。工厂级网络主要指 OT 网络与信息系统（ERP、MES、SCM 和 CRM 等系统）的互联。随着底层数据量的指数型增长、视频等高带宽应用向工业领域的渗透，对工厂内部的 IT 网络的带宽、时延及抖动的性能提出了更高的要求。工厂间网络主要指企业与工业云端应用的连接网络，是现实工业互联网的关键，此类连接主要采用专线等技术。企业与云端应用的连接则需要在可靠性、带宽、时延性能、业务形态等方面进行切合工业领域特点的创新尝试。

由于传统 OT 网络采用的通信技术相对封闭，且标准多样，使得现有的 OT 网络和 IT 网络相对隔离，难以实现数据采集与交互，因此工业数据采集技术成为打通 OT 网络与 IT 网络的关键。

3 工业连接场景及需求

工业连接涵盖了工业生产流程中机器终端及控制系统、原料、信息系统、在制品、产品以及人之间的互联，涵盖各类连接形态。随着信息通信技

术向工业领域的加速渗透，工业连接的性能不断提升，类型不断丰富，极大地拓展了传统工业网络的内涵和外延，为智能制造的发展奠定了良好基础。通过对工业领域各类连接的梳理，工业连接可以包含两类连接、四大场景、八项需求，具体如下。

两类连接包括企业内部连接和企业外部连接。其中，企业内部连接包括机器设备、控制设备、企业云平台、企业内人员间的互联关系；而企业外部连接包括不同企业间（如上下游企业间）、企业与外部公有云平台间、企业和售出产品、企业与客户间的互联关系。

四大场景分别为促进智能化生产的智能生产场景，促进网络化协同的网络协同场景，促进个性化定制的用户定制场景，促进服务化转型的服务化场景。

结合以上描述的两类连接与四大场景，本文将引申出工业连接的八项需求，具体描述如下。

需求一：现场设备与工厂控制系统的连接。此类连接主要指底层 I/O、PLC 等控制单元、IPC 等上位机的互联。需要在保障可靠性、时延性能的同时大幅提升带宽和规模，确保不同协议间的数据互通和物理互通。

需求二：现场设备与私有云平台的连接需求。此类连接主要指现场设备与运行在私有云上的 MES 等 IT 系统的直接连接。现场设备可通过以太网或者光网络，跨越上位机“e 网到底”，实现与运行在私有云上的 IT 系统的直接连接，需要通过网络和各类现场工业通信协议的高效互通。

需求三：工业控制系统与私有云平台的连接需求。此类连接主要指控制系统与信息系统的互联。特别是随着视频等高带宽应用在工业领域的使用不断普及，此类连接需要大幅提升工业以太网和通用网络技术的互通性能。

需求四：私有云平台与人的连接需求。此类连接主要指人通过 HMI、移动设备等方式与工业 IT 系统的交互，需要提升 HMI、SCADA、远程

操控、移动设备操控等不同的软硬件人机界面的性能,提升管理的敏捷性和效率。

需求五:企业与工业公有云的连接需求。企业与公有云的连接需要在可靠性、带宽、时延性能、业务形态等方面进行切合工业领域特点的创新尝试。随着新的网络技术、信息技术以及互联网商业模式的出现,工业生产过程已不仅仅局限在工厂内,开始逐步通过外部网络,使得工业生产与互联网业务模式实现深度融合,主要体现在IT系统与互联网的融合、OT系统与IT系统的融合、企业专网与互联网的融合等。

需求六:企业和企业的连接需求。此类连接可以具化为不同企业信息系统互通需求,分为两个层次,一是CRM、ERP等非核心系统的互通,二是MES等核心系统的互通。这两类互通需求都对远程互联的可靠性、时延性能、安全性提出了新的要求。

需求七:企业和用户的连接需求。用户需求与工业系统的实时互通是实现个性化定制的基础,这需要工业云平台对用户定制应用的良好支撑以及对制造、物料、物流的高效协同。

需求八:企业和智能产品的连接需求。产品和工厂的泛在连接将是预测性维护、远程维护的网络基础,也是企业实现服务化转型的基础。此

类连接需要重点解决广域大连接的问题,需要物联网技术的深度应用。

4 工业连接关键技术

为了实现工业连接中不同场景下的各种需求,其基础就是要实现工业网络技术与互联网技术的融合、海量异构数据的集成与交换以及网络与数据在工业领域的安全保障。因此,工业数据采集技术、工业网络IP化、工业网络无线化、标识解析技术以及网络安全技术成为实现工业连接的关键。

4.1 工业数据采集技术

数据是智能制造的核心,通过工业数据全周期的应用,实现机器弹性生产、运营管理优化、生产协同组织与商业模式创新,推动工业智能化发展。而如何实现数据的采集上传则成为实现数据应用的关键。工业数据采集技术结构示意图如图2所示。

现有的工业数据采集主要有以下两种方式。

方式一:由数据采集平台集成各种工业协议驱动,直接与底层设备通过网络连接并进行解析。该方式可连接设备种类多,但对于非以太网接口支持较为困难,因此需要将现存的RS232、RS485、控制器局域网络(CAN)等物理接口转换为以太

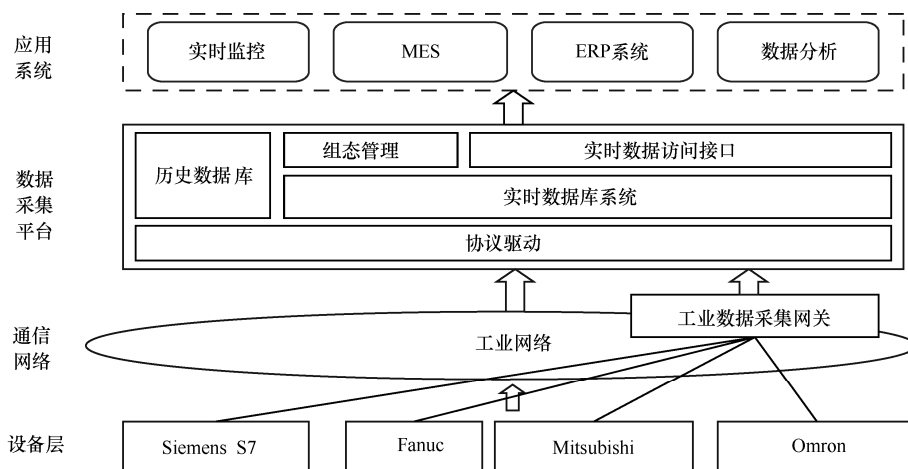


图2 工业数据采集技术结构示意图



网接口。

方式二：由工业通信网关集成各种工业协议驱动。工业现场设备种类繁多，通信接口多样化，工业通信网关作为不同接口及通信协议的设备之间通信的桥梁，可以在不同的工业接口与通信协议之间做转换，是对不同设备进行数据采集的有效手段。工业通信网关通过工业 PC 或专业芯片，集成各种主流的工业协议，适配不同厂商设备的接口，并将各种工业协议转换为标准的 OPC UA、Modbus TCP、MQTT、HTTP RESTful 等通用协议，连接至数据采集平台。该方式可以减少数据采集平台支持的协议种类，同时工业通信网关可以支持边缘计算能力，实现数据的分布式计算分析，形成本地的实时优化决策。

以上两种方式各有优缺点，实施过程中应根据实际情况结合使用：通过工业通信网关适配各种工业协议，同时对于工业通信网关尚未支持的工业协议，通过方式一的数据采集平台进行支持，实现各种工业数据采集的无缝集成。

4.2 工业网络 IP 化

为实现生产过程全流程的数据采集、流动及

智能化生产，IP 化与无线化已成为工厂内部传输网络发展共识，IP 化基于以太网技术实现从机器到 IT 系统的端到端 IP 互联，无线化是各种无线技术(在第 4.3 节中描述)在工厂内部的广泛应用。

工业以太网由于以太网应用的广泛性和技术的先进性，已逐渐成为企业商用计算机通信领域和过程控制领域中上层信息管理与通信的主流技术，并且将逐步替代现场总线，完成“e 网到底”的目标，如图 3 所示，从而实现 OT 网络与 IT 网络的无缝互通和融合。

工业以太网经历了多年的快速发展，其产品已非常成熟，EMC、高低温等工业特性已能够满足工厂对于不同环境的需求，并已经在工业领域得到了非常广泛的应用，其协议的标准化已完成，不同厂商产品互通性好。思科提出的全厂融合以太网(CPwE)架构将工业网络分为 2 个区域(6 个层级)，包括企业区(第 4~5 级)、制造区(第 0~3 级)，实现以太网在工厂 IT 与 OT 域的全覆盖打通，同时在 CPwE 架构中集成了先进的无线技术，实现无线网络在工厂复杂恶劣环境中应用。

由于现场级的工业控制网络对于时延、抖动的

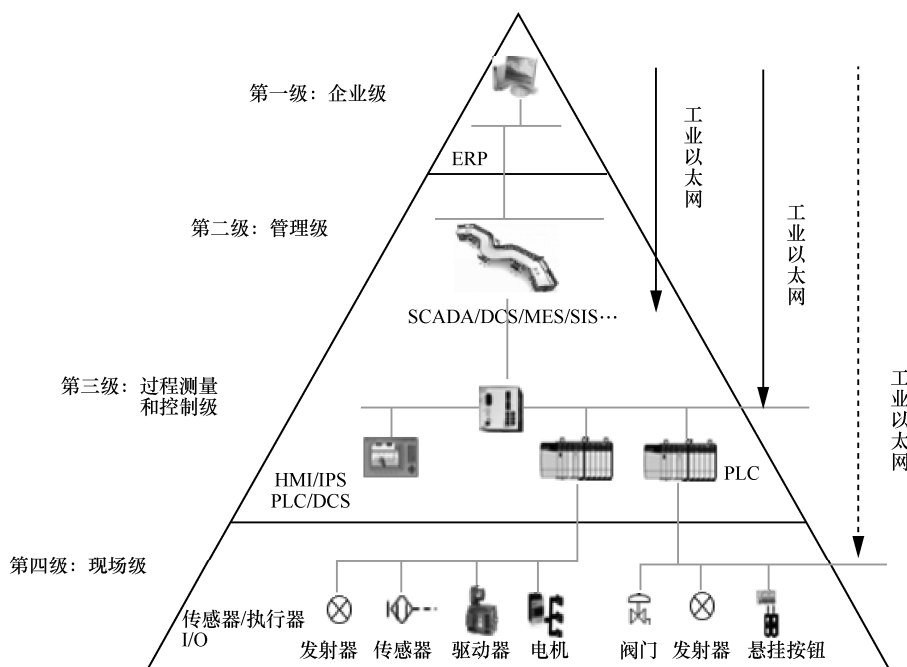


图 3 工业以太网“e 网到底”目标

要求较高,因此时间敏感网络(TSN)应运而生。相较于传统以太网技术,TSN具有时间敏感数据处理、时钟同步、帧抢占/穿插传输等功能,保证数据传输的确定实时性,可以满足智能制造实时监控或实时反馈的需求。同时,TSN可以解决现有各工业以太网协议自成体系、无法互通的问题,是实现“e网到底”目标的重要实现技术。

工业以太网虽然技术成熟,但也面临着以下问题:仅支持点到点直连组网,组网存在局限性;中间传输需多级汇聚交换机,需要预留机房建设用地以提供电力、空调等;铜缆内弱电信号易受电磁干扰和雷电影响;业务支撑能力不足,网络维护相对复杂。

因此,随着无源光网络(PON)技术在电信、电力行业的广泛应用,工业PON技术已成为除了以太网技术外车间通信网络组网技术的一种全新的方案。相对于工业以太网技术,工业PON技术具有以下优点:PON通过无源器件组网,不受电磁干扰和雷电影响;采用自愈环型网络支持并联型,切换时间短、抵抗失效能力强;点到多点传输架构,终端并行接入,部署灵活;仅需单根光纤传输,最远覆盖20 km范围;多业务承载,支持数据、视频、语音、时间同步等多种业务;安全性高,PON设置ONU安全注册机制,下行数据传送天然加密,上行数据传送时分机制隔离。

4.3 工业网络无线化

无线技术在工业领域已经得到越来越广泛的应用,成为有线网络技术的重要补充。随着工业无线技术的不断发展与完善,其应用已不单单局限在数据采集领域以及作为工业有线网络的补充,正逐渐向生产控制及全网方向发展。

工业领域使用的无线技术主要包括短距离通信技术RFID、Zigbee、Wi-Fi等,用于车间内的传感数据读取、物品及资产管理、自动导引运输车(AGV)等无线设备的网络连接;专用工业无线通信技术WIA-PA/FA、ISA100.11a、

WirelessHART等,应用于对工厂应用有一定实时性要求、高可靠和高安全的无线通信场景。

蜂窝无线通信技术2G/3G/LTE、NB-IoT等也开始应用于工业领域,用于智能产品、大型远距离移动设备或车辆、手持终端等的网络连接。随着3GPP 5G标准的推进以及商用化的临近,5G将在工业领域得到大量应用。3GPP的5G技术面向物联网定义了3类应用场景:大容量物联网(massive IoT)、低时延高可靠通信(critical communication)和增强型移动宽带(eMBB)。其中,massive IoT主要针对海量机器类通信,如各类表的远程激活、消耗流量的实时跟踪、无线视频监控的远程激活,远程监测并上报环境参数和控制机械的数据,一个用户具有多个终端和一个终端被多个用户使用等多种场景,工业互联网应用场景下有相同的需求;低时延高可靠通信要求极低的时延和更高的可用性,在工业互联网中主要应用在工厂内部的现场控制等应用场景。需要重点关注对于各种现场总线协议的承载支持,并综合利用网络切片和移动边缘计算等技术,实现工业自动化场景的支持;而eMBB场景则分为广域覆盖和热点覆盖,对于工业互联网逐渐兴起的大流量业务能够更好地支撑,如虚拟工厂和高清视频远程维护等。

4.4 标识解析技术

互联网与制造业的融合带来工业连接需求的快速提升,为实现各个层面的互联互通,工业互联网打通需要从生产到企业运营的全流程环节,实现包括智能化生产、网络化协同、个性化定制、服务化转型的四大场景。这就需要实现企业与企业、企业与用户、企业与产品的连接及异构数据互联。

随着工业设备的接入数量越来越多,使得数据呈几何式递增,面对如此庞大的数据,首先需要解决的是如何对工业领域对象进行标识的问题。通过建设统一的标识体系,可以为数量庞大的终端赋予唯一的标识,从而实现在任何时间、



任何地点对终端的定位与寻址,满足海量异构数据在不同对象间、不同领域间以及不同区域间互联互通的应用需求。

通过标识解析体系可以贯穿工业互联网整个流程,给予目标对象在整个生命周期中一个统一不变的“身份标识”。基于统一的“身份标识”实现目标对象所有数据的关联与统一,实现异构、异主、异地数据安全可控的开放及访问。

对象标识解析技术标准在国际上主要有四大体系,包括 DONA 定义的 Handle 解析服务体系、EPCglobal 定义的对象名称服务(ONS)体系、ITU 和 ISO 联合制定的面向对象标识符(OID)的对象标识符解析系统(ORS)体系、日本泛在识别中心的 uCode 解析服务体系。

通过构建统一的工业数据解析系统,可以对工业数据源进行可靠的追溯,从而避免了在浩瀚的数据海洋中查找源头的问题。同时,由于数据来源于上下游产业链的不同参与者中,所产生的数据具有数量巨大且异构的特点。因此需要把所有可用信息转变为统一格式,从而能够有效节约资源,实现对海量工业数据的分析及智能化应用。

4.5 网络安全技术

随着工业连接的创新发展,现有相对封闭的工业网络体系将更加开放,随之也将面临更多的安全问题和挑战,因此需要通过综合性的安全防护措施,保证网络与数据的安全。

工业企业各互联设备及系统之间应该进行有效可靠的安全隔离和控制。包括:OT 系统与 IT 系统之间,应部署防火墙;工厂外部对工厂内部云平台的访问应经过防火墙,并提供 DDoS 防御等功能,同时部署网络入侵防护系统,可对主流的应用层协议及内容进行识别,自动检测和定位各种业务层的攻击和威胁;所有接入工厂内部云平台、工厂信息系统、工业控制系统的设备,都必须进行接入认证和访问授权;通过外部网络传输的数据,应采用 IPSec VPN、SSL VPN 等隧道

传输机制或 MPLS VPN 等专网技术,防止数据泄露或被篡改。

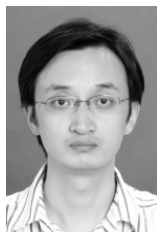
5 结束语

随着“中国制造 2025”计划的逐步推进,工业连接作为支撑智能制造的基础能力,在未来应逐步完善网络互联技术、标识解析技术以及应用支撑技术,实现数据在企业内部的各生产系统间、生产系统与商业系统间、商业系统与制成品间以及企业外部各企业间、企业与云平台间的无缝传输,构建新型的基于有线与无线通信技术安全可靠的机器连接方式,从而支撑实时感知、协同交互的智能生产模式。

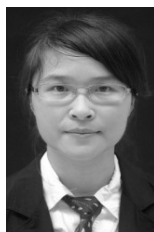
参考文献:

- [1] 高巍. 工业互联网推动工厂网络与互联网融合发展[J]. 中兴通讯技术, 2016, 22(5): 21-25.
GAO W. Industrial internet promotes the integration of plant network and internet[J]. ZTE Technology Journal, 2016, 22(5): 21-25.
- [2] 中国工业互联网产业联盟. 工业互联网体系架构[S]. 2016.
Industrial Internet Industry Alliance. Industrial internet architecture[S]. 2016.
- [3] 中国电信集团公司, 中国通信通信研究院. 工业连接计划白皮书[R]. 2016.
China Telecom Corporation, China Academy of Information and Communication Technology. Industrial connectivity planning white paper[R]. 2016.
- [4] 杨思维. 升级版德国“工业 4.0 平台”经验对我国制造业的影响[J]. 电信科学, 2016, 32(1): 108-111.
YANG S W. Impact of the upgraded version of German “industrial 4 platform” on China’s manufacturing industry[J]. Telecommunications Science, 2016, 32(1): 108-111.
- [5] 杜娟, 王峰. 互联网的内涵、服务体系及对制造业的作用路径[J]. 电信科学, 2016, 32(1): 98-104.
DU J, WANG F. Connotation, service system of internet and its integrating path to manufacture[J]. Telecommunications Science, 2016, 32(1): 98-104.
- [6] 董振江, 董昊, 韦薇, 等. 5G 环境下的新业务应用及发展趋势[J]. 电信科学, 2016, 32(6): 58-64.
DONG Z J, DONG H, WEI W, et al. New services and development trend in 5G environment[J]. Telecommunications Science, 2016, 32(6): 58-64.

[作者简介]



李洁（1980-），男，中国电信股份有限公司上海研究院工程师，主要从事车联网及工业互联网领域的产品开发、技术研究工作。



常洁（1984-），女，博士，现就职于中国电信股份有限公司上海研究院，主要研究方向为工业网络、5G 等。



张东（1972-），男，中国电信集团公司政企客户事业部产业互联网创新发展中心主任、高级工程师，主要从事产业互联网（互联网+）、工业互联网和智慧城市方向的研究工作。



杨震（1972-），男，博士，中国电信股份有限公司上海研究院教授级高级工程师、物联网部智能产品室主任，主要研究方向为人工智能、自然语言处理、搜索引擎、物联网技术。



对象存储系统的自动化管理技术

胡豆豆¹, 罗飞², 张园¹

(1. 中国电信股份有限公司上海研究院, 上海 200122;

2. 华东理工大学, 上海 200237)

摘要: 物联网视频监控领域中的对象存储系统不仅轻松突破了传统存储系统 SAN 的性能瓶颈, 而且可以实现性能与容量的线性扩展; 然而, 对象存储系统资源的远程管理成为困扰管理者的一个问题。针对该问题, 研究能够实现对象存储集群的自动化管理的方法和相关技术, 提高存储集群的使用效率。该自动化管理系统功能主要包括: 集群组件的远程启动、停止、重启; 运行状态监测; 集群扩张和日志查看及存储等功能。基于 Ceph 对象存储集群, 实现了面向 Web 应用服务平台的存储集群自动化管理系统, 最终有效地对对象存储系统进行监视和管理, 并能够自动化扩展集群的节点和存储容量。

关键词: 对象存储; 自动化; 云计算

中图分类号: TP315

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017306

Automatic management technology of object storage systems

HU Doudou¹, LUO Fei², ZHANG Yuan¹

1. Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China

2. East China University of Science and Technology, Shanghai 200237, China

Abstract: The object storage system not only breaks the bottleneck of the traditional SAN storage system, but also achieves the expansion of performance and capacity linearly in video monitoring field. However, the remote management for the resources of the object storage system has become a problem for managers. Aiming at this problem, the method and related technology of automatic management of object storage cluster was studied, so as to improve the efficiency of storage cluster. The functions of the automatic management system included remote launching, stopping and resetting of cluster components, monitoring of running status, cluster expansion and log viewing and storage. Based on Ceph object storage cluster, the automated management system for Web application service platform was achieved, monitoring and management was realized effectively ultimately, and the nodes of the cluster and the storage capacity were expanded automatically.

Key words: object storage, automation, cloud computing

1 引言

随着互联网、Web 2.0 的快速发展, Web 应用创建出数百亿的小文件, 上传海量的照片、视频、音乐, Facebook 每天新增数十亿条内容, 每天发送数千亿封邮件。据 IDC 统计, 未来 10 年间数据将增长 44 倍, 2020 年全球数据将增长到 35 ZB, 其中, 80% 为非结构化的数据, 且大部分是非活跃数据^[1-3]。

在物联网视频监控领域中, 随着高清化、智能化的推进以及平安城市等城市公共安防的发展, 面对如此庞大的数据量, 仅具备 PB 级扩展能力的块存储和文件存储显得无能为力; 人们需要一种全新的架构存储系统, 这样的存储系统需要具备极高的可扩展性, 能够满足人们存储视频数据时对存储容量 TB 到 EB 规模的扩展需求^[4,5]。

对象存储是一种基于对象的存储设备, 具备智能、自我管理的能力, 通过 Web 服务协议 (如 REST、SOAP) 实现对象的读写和对对象资源的访问。对象存储以对象 ID 为基础, 扁平化管理所有存储和桶, 根据对象 ID 便可直接访问数据, 解决 NAS 复杂的目录树结构在海量数据情况下数据查找耗时长的的问题, 这使得对象存储具备极强的扩展性, 能够轻松实现单一名字空间内支持数百万级文件的存储。

对象存储系统的出现满足了数据归档和云服务两大需求。通过面向对象的概念构建的存储系统, 能够有效地利用集群各个节点的性能, 并且有效地将多个客户端请求均匀地分布到集群中的各个节点, 充分地利用了对象存储集群的特性。集群管理员可以通过集群管理系统实时地监视集群各个部分的状态, 也可以根据需要进行调整单个节点上组件的运行状态。当存储集群需要扩张时, 通过自动化的管理系统也能够很快地对集群进行调整以及在各个存储节点之间进行动态的数据迁移, 从而保持服务的持续可用^[6]。

但随着存储资源的堆积和存储集群节点数量的不断上升, 如何通过集群自动化管理系统进行高效的管理和控制成为了一个困扰企业的问题^[7]。本文针对该问题, 研究基于对象存储文件系统 Ceph 的自动化管理系统。在 Ceph 存储集群上, 通过构建一个 Web 项目, 实现对 Ceph 集群的远程监视和控制 (包括集群的常用操作、存储节点的增加和删除等); 以此来模拟如何对 Ceph 存储集群进行自动化管理, 从而推进它的商用价值^[8-10]。

2 系统方案

自动化管理的目的是帮助用户解决存储集群在运行时的远程监视、管理和控制的问题, 即管理员通过浏览器随时随地以访问 Web 应用程序的形式来访问对象存储系统, 实现对远程对象存储集群各组件的监视、控制、节点的横向扩展及收缩, 进而改善集群资源的使用性能、提高利用率和降低运营成本。

对象存储系统自动化管理的需求主要可归纳为如下几点。

(1) 监视功能

包括集群当前运行状态监视、存储节点分布和状态监视、单个节点上运行的集群组件监视、集群中已添加节点的分布和运行状态监视。

(2) 组件的常规操作

包括 MON 的重启、MDS 的停止/启动、OSD 的停止/启动等。

(3) 存储节点动态调整

包括 OSD 的添加/删除。

(4) 群日志的查看

包括管理员操作记录查看、集群状态日志查看、存储节点监视日志查看、单节点组件运行状态日志查看、集群节点分布和运行状态的监视日志查看。

系统采用基于 MVC 框架开发的 B/S (浏览器/服务器) 模式的 Web 应用程序。用户可通过访



问 Web 浏览器的方式对存储集群进行简单的控制和管理。本文系统的模型拥有较多的处理任务, 将集群的监视、操作调用与实现和与数据库中日志信息的交互都放在该层来实现。只要网络连接正常, 互联网用户就可以通过打开浏览器随时随地访问应用系统, 从而实现对局域网内集群的所有机器的访问并进行远程管理和控制。模块划分为: 集群监视模块、存储节点添加/删除模块、组件控制模块、日志存储/调用模块 4 个主要模块, 系统功能结构如图 1 所示。

为了提高管理系统的用户体验, 采用名为 Bootstrap 的 Web 前端开发框架。Bootstrap 是 Twitter 推出的一个开源的、用于前端开发的工具包。当管理员需要监视、操作或者调整集群时, 需要在 Web 页面填写相应的表格, 并提交到 Web 服务器进行相关操作。但是和集群相关的操作并不能在很短的时间内完成, 例如: 集群中各个节点是否都已启动、添加存储节点 OSD 等。因此当提交此类操作时, 即使忽视数据在广域网中的传输和建立 SSH 连接的时间, 都很难将操作的反馈瞬间呈现在页面上。所以, 鉴于上述原因, 本文使用了 AJAX (asynchronous JavaScript XML) 技术, 实现了页面和 Web 服务器之间的异步通信。

当用户在 Web 页面点击提交事件处理的按钮

时, 视图会将该请求的相关信息传递到控制器, 控制器对用户请求进行查看并作出判断, 执行相应的模型操作, 即对应的功能模块会创建一个 SSH 协议, 打开 Java 项目与 SSH 服务器的会话, 通过该会话可以远程执行命令和 Shell 访问。本文采用了一个名为 Ganymed SSH-2 for Java 的包, 专门用于实现 SSH 协议, 建立 SSH 服务器连接, 并支持 SSH 会话。Ganymed SSH2 for Java 是一个以纯 Java 的形式实现 SSH-2 协议的库 (支持 J2SE1.4.2 和 5.0 以上的版本)。它使得在 Java 项目中与 SSH 服务器建立连接, 同时支持 SSH 的会话 (远程命令执行和 Shell 访问), 本地和远程端口交互, 本地流转发、X11 转发和 SCP 功能等。

自动化管理系统的后台除了关于 MySQL 数据库的日志操作外, 其余的集群控制都是远程调用集群上的脚本实现的。这些脚本均存放在存储集群的主管理节点上, 所以当 Ganymed SSH2 for Java 与集群主管理节点建立 SSH 会话后, 调用对应的 Shell 脚本即可执行所需要的管理操作。

3 系统实现

系统中 Web 控制页面显示的数据都是通过执行后台脚本, 截取到所需的特征数据, 并经过整理后呈现出来的。

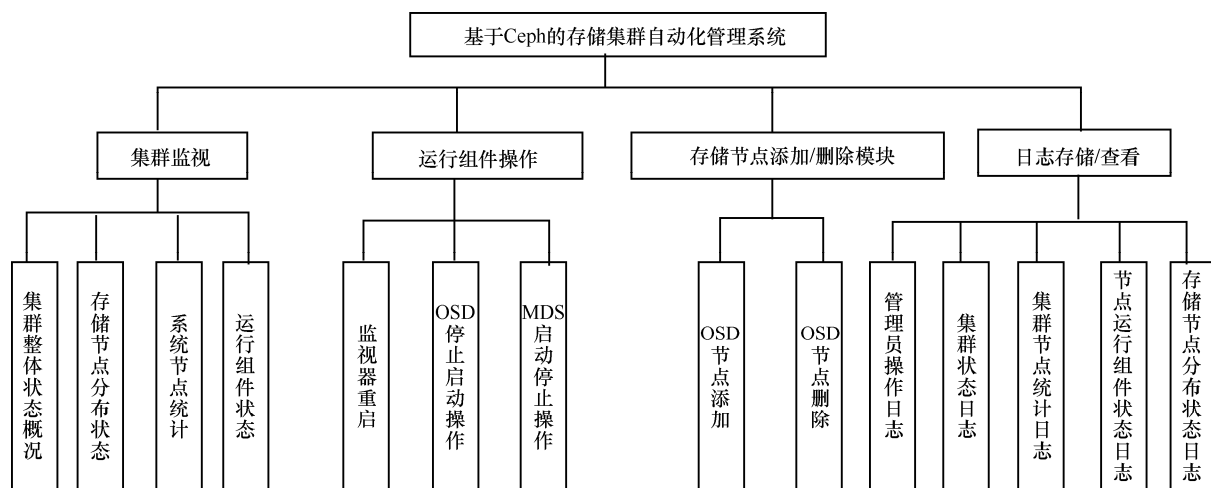


图 1 系统功能结构

3.1 全局状态监控

全局状态监测在终端输入 `ceph -s` 指令时可以查看集群整体的当前状态，快速地查看需要的关键信息，见表 1。首先从指令 `ceph -s` 输出集群当前的通用状态信息，进而通过 `grep` 和 `sed` 的筛选和替换功能实现对集群特征数据的采集。

表 1 集群整体状态

信息类别	总体信息	当前信息	状态信息
MON (监控节点)	MON 个数	主 MON 节点的 IP 地址	MON 状态列表
MDS (元数据服务器)	MDS 个数	MDS 所在节点 IP 地址	MDS 状态列表
OSD(对象存储节点)	OSD 个数	可访问 OSD 个数	OSD 状态列表
存储容量/GB	存储容量总和	可使用容量	已使用容量

3.2 存储节点分布状态监测

Ceph 是一个高可用性和高可靠性的管理硬件和软件。它不具备单点故障，并且可以为数据提供还原服务。假如系统出现故障，需要到放置组和相关的守护进程下查找问题。当 Ceph 集群遇到一些故障时，还可能阻止用户访问特定的对象，但可能访问其他对象。当运行时遇到故障，只需要按照步骤监测守护进程和放置组就能知道故障原因，然后排除。

由图 2 可知，在集群中一个 OSD 的状态或者在这个集群中 (in) 或者在这个集群外 (out) 的；并且，它或者是运行中 (up) 的，或者是不运行 (down) 的。如果一个 OSD 的状态是 up，它或者在集群 (可以读取和写入数据) 或者在 out 状态。如果它曾是 in 状态但最近被移出集群了，Ceph 将会迁移配置至其他 OSD 中。如果 OSD 不在集群中，基于 Ceph 的 CRUSH 算法，该 OSD 将不会配置在一个放置组中。如果 OSD 的状态是 down，它也会是 out 状态。

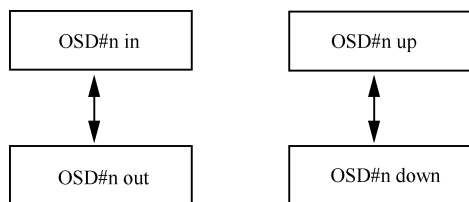


图 2 OSD 状态转换

执行 `Ceph Osd tree` 指令能够确定 `ceph-osd` 后台是否正常运行，根据控制台的输出，可以很容易地从中获取所需要的特征数据，并且有效地组织在一起，作为日志存储到日志并刷新页面中的对应部分视图。

3.3 集群节点分布监测

由于集群中的节点会逐渐增多，所以需要在 Web 页面上直观表示节点名称和 IP 地址，形式见表 2。

表 2 集群节点配置

IP 地址	节点 Hostname
192.168.232.158	server01
192.168.232.160	server02
192.168.232.170	server04

为了获取集群中加入的节点 IP 地址和相关的 Hostname，前提必须是将 Hostname 和 IP 地址写入系统文件 `/etc/hosts` 中。

同时，管理系统能够自动化管理存储节点 OSD 的添加/删除操作。集群上放置的每一个对象 (object) 都可以根据 Ceph 的 CRUSH 算法计算出具体的存放位置。所以，当 OSD 发生变化时，集群也需要做相应的调整。CRUSH 算法是 Ceph 对象存储的核心，决定了每一个对象的存放位置。当需要存放一个对象时，只需要知道对象的 ID、集群架构图和存放策略，就可以计算出应该存放在哪个 OSD 上。因此 Ceph 的元数据不需要包含位置信息，每个客户端都可以自己计算出对应的 OSD，大大减少了对元数据的请求。

RADOS 是使用 CRUSH 算法的分布式对象存储，由 OSD 和 MON 组成，OSD 承担数据的存储、备份以及恢复工作，MON 保证集群架构图与实际情况相符。RADOS 的特点之一就是数据的备份、同步以及恢复等都由 OSD 自发地完成。

RADOS 中 OSD 独自管理数据的能力是建立在 CRUSH 算法上的，CRUSH 算法又是建立在 CRUSH map (包含 OSD 的数量、组织结构以及存储策略)



上的。因此拥有一份最新的 CRUSH map，对 OSD 是至关重要的，否则 OSD 无法确定隶属同一个 PG 的 OSD，也就无法完成数据的备份、同步等操作。

RADOS 使用一个由 MON 组成的小集群生成最新的 CRUSH map。当增/减 OSD、修改 OSD 组织结构、修改存放策略或者发现 OSD 失效时，更新请求被发送到 MON 集群，MON 生成一份新的 CRUSH map。同时 MON 集群也对外提供 CRUSH map 的访问服务，其他节点可以从 MON 集群中获取一份最新的 CRUSH map。MON 集群中采用 Lease 机制确保每个 MON 对外提供的 CRUSH map 的一致性，采用 Paxos 协议选举出一个 leader，由 leader 负责生成新的 CRUSH map。

另外，OSD 节点之间互相通信时会比较各自拥有的 CRUSH map 的 epoch，用最新的 CRUSH map 替换旧的 CRUSH map。当客户端要从 OSD 中读取对象时，如果客户端中还没有 CRUSH map：客户端首先从 MON 集群中读取一份最新的 CRUSH map，然后计算出存放该对象的 PG 对应的 OSD，然后客户端与目标 OSD 比较各自的 CRUSH map 的 epoch，同步到最新的 CRUSH map。如果客户端中已经有 CRUSH map，在和 OSD 通信时，如果 OSD 的 CRUSH map 版本更新，那么更新客户端的 CRUSH map。如果客户端和 OSD 的 CRUSH map 版本一致了，但都不是最新的，使用旧的 CRUSH map 计算出的 OSD 上可能没有存放要访问的对象，导致操作失败，这时客户端从 mon 集群读取最新的 CRUSH map。

3.3.1 OSD 添加

新增一个 OSD 需要分为两个部分：准备（prepare）和激活（activate）。准备步骤的指令为：`ceph-deploy osd prepare ceph1: /dev/vdb`，步骤如下所示：

```
push 配置文件到 ceph1:/etc/ceph/ceph.conf
push ceph.bootstrap-osd.keyring 到 ceph1:/var/
lib/ceph/bootstrap-osd/ceph.keyring
```

```
udevadm trigger --subsystem-match=block
--action=add
```

```
ceph-disk-prepare --fs-type xfs --cluster ceph --
/dev/vdb //创建分区、格式化等操作
```

激活步骤的指令为：`ceph-deploy osd active ceph1:/dev/vdb`，步骤如下所示：

```
ceph-disk-activate --mark-init upstart --mount
/dev/vdb //挂载 Ceph 分区
```

```
initctl emit --no-wait -- ceph-osd cluster=ceph
id=0 //添加 ID 为 0 的 OSD 守护进程，并使其处于
非等待状态，即添加完该 osd.0 后直接启动。
```

3.3.2 OSD 删除

删除 OSD 守护进程时需要根据该 OSD 的不同状态执行一些列的指令，从而达到有效的集群 CRUSH 调整，并且完成该 OSD 上存储对象的完整迁移，最后完全将该 OSD 守护进程从集群中移除。步骤如下所示：

```
service ceph stop osd.0 //停止该设备
```

```
ceph osd crush remove osd.0 //从 CRUSH 中移
除与该 OSD 相关的数据，此时运行 ceph osd tree
指令可以看到该存储进程 osd.0 处于 down 状态
```

```
ceph auth del osd.0 //从认证中删除
```

```
ceph osd rm 0 //从集群中完全移除该 OSD 的
所有数据
```

4 系统测试

为了完成系统测试，本文在 Ubuntu12.04 的虚拟机上搭建基于 Ceph 的初始集群。该集群初始化规模为 3 个节点，并通过 `ceph-deploy` 完成 Ceph 的安装。首先将 Ceph 资料库添加到 `ceph-deploy` 的管理节点，然后安装 `ceph-deploy`。注意：以不同的用户登录上节点安装 `ceph-deploy` 时不需要在命令前面添加 `sudo` 或者以 `root` 身份执行，因为远程的主机将不会执行 `sudo` 命令。测试过程包括监视功能测试、存储节点添加/删除功能测试等。

4.1 监视功能测试

监视功能所显示当前集群的状态如图 3 所示, 包括: mon 节点的个数、所在节点、运行状态; mds 节点的个数、所在节点、运行状态; OSD 节点的个数、运行状态、存储容量等信息; 其中, OSD 存储容量已用部分 13 837 MB, 剩余可用部分为 4 210 MB, 总共的存储容量为 19 021 MB。

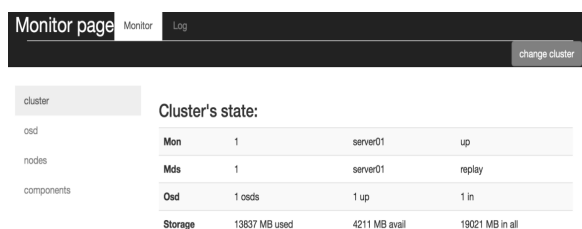


图 3 集群状态监视

集群中各存储节点状态如图 4 所示, 显示了集群中存储节点 OSD 的分布情况: ID 为 0 的 OSD 节点的权重为 0.019 99 (1 TB 大小的容量权重为 1, 所以 0.019 9 权重说明 ID 为 0 的 OSD 存储容量大约为 20 GB 左右) 以及状态为 up。



图 4 集群存储节点分布

集群内节点状态统计如图 5 所示, 这里显示的节点是已经被添加到集群的, 包括节点 IP 地址、节点名称以及该节点是否在运行中。从图 5 可以看到, 目前只有节点 IP 地址为 192.168.232.158、节点名为 server01 的节点 power 栏显示的是 up, 即在运行中。

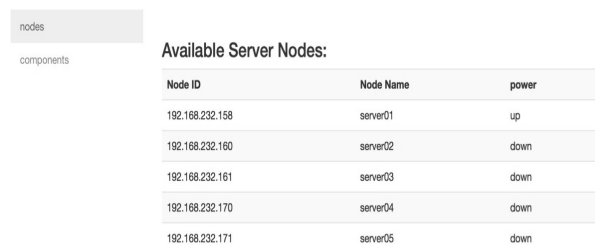


图 5 集群内节点状态统计

4.2 存储节点添加/删除功能测试

在 server02 上添加一个 OSD 守护进程, 操作和结果分别如图 6 和图 7 所示。

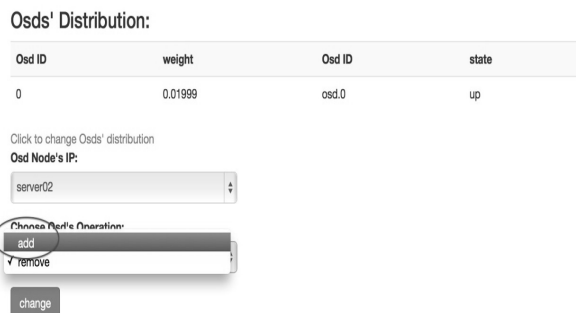


图 6 添加 OSD.1 操作

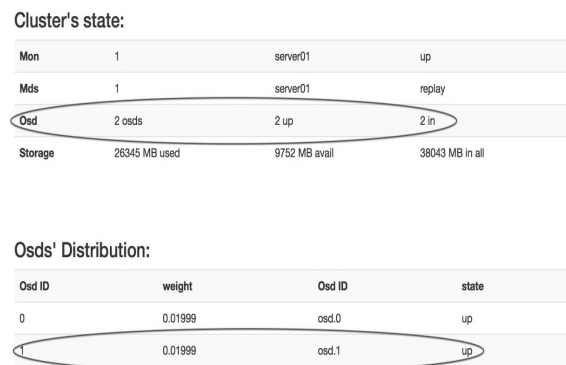


图 7 OSD.1 添加结果

将主管理节点 server01 上面的 OSD 守护进程 osd.0 删除, 操作和结果分别如图 8 和图 9 所示。

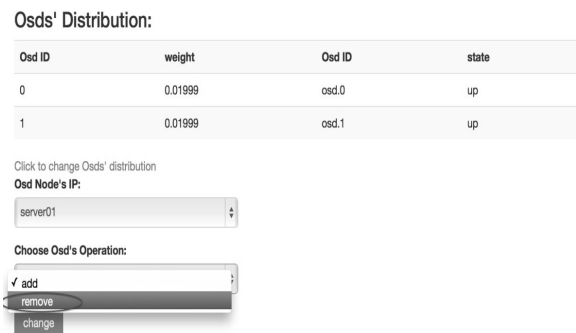


图 8 删除 OSD.0 操作示意

4.3 日志查看功能测试

查看某个时间段的存储节点分布情况, 结果如图 10 所示。



Cluster's state:

Mon	1	server01	up
Mds	1	server01	replay
Osd	1 osds	1 up	1 in
Storage	12511 MB used	5537 MB avail	19021 MB in all

Osds' Distribution:

Osd ID	weight	Osd ID	state
1	0.01999	osd.1	up

图9 删除 OSD.0 的结果

Osd Tree Distribution:

Osd ID	weight	Osd Name	Osd State	Time
0	0.01999	osd.0	up	2014-05-28 07:35:25
0	0.01999	osd.0	up	2014-05-28 07:35:43
0	0.01999	osd.0	up	2014-05-28 07:36:02
1	0	osd.1	down	2014-05-28 07:36:02
0	0.01999	osd.0	up	2014-05-28 07:36:21
1	0.01999	osd.1	up	2014-05-28 07:36:21
1	0.01999	osd.1	up	2014-05-28 07:36:33
0	0.01999	osd.0	down	2014-05-28 07:36:33
1	0.01999	osd.1	up	2014-05-28 07:36:51

Time From: 2014-05-28 07:35 To: 2014-05-28 23:55 Get Logs!

Cluster Nodes:

Node ID	Name	power
---------	------	-------

图10 某时间日志查看结果

5 结束语

基于对象存储系统 Ceph, 研究了一个对象存储集群的自动化管理系统。管理系统的功能主要集中在 3 个部分——集群状态监视、OSD 存储节点添加与删除和集群状态日志查看。以平台化方式设计, 采用基于 MVC 框架进行设计和实现。用户可通过访问 Web 浏览器的方式对存储集群进行简单的控制和管理。

作为底层存储集群的管理系统, 这些功能都是必备的, 但还是远远不够的。要真正实现自动化的存储集群管理系统, 故障排除与恢复和负载均衡优化是不可缺失的主要功能。而本系统中仅有的一些功能做得也不是非常完善, 这些工作都有待在未来进行优化。

参考文献:

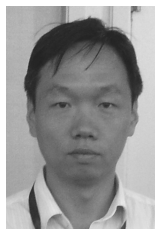
- [1] 郭立, 金海, 韩宗芬. 集群系统的单一系统映像研究[D]. 武汉: 华中科技大学, 2003.
GUO L, JIN H, HAN Z F. Research on single system image for cluster[D]. Wuhan: Huazhong University of Science and Technology, 2003.
- [2] 孙乔, 邓卜侨, 王志强, 等. 一种基于分布式服务器集群的可扩展负载均衡策略技术[J]. 电信科学, 2017, 33(9): 190-196.
SUN Q, DENG B Q, WANG Z Q, et al. A scalable load balancing strategy based on distributed server cluster[J]. Telecommunications Science, 2017, 33(9): 190-196.
- [3] 郭延辉, 陈征, 龚松春. 一个集群文件系统的虚拟化支持平台研究与实现[J]. 电信科学, 2012, 28(9): 34-37.
WU Y H, CHEN Z, GONG S C. Research and implementation of virtual platform on cluster file system[J]. Telecommunications Science, 2012, 28(9): 34-37.
- [4] 兰超, 张勇, 邢春晓. 面向复杂数据的对象存储系统[J]. 计算机科学与探索, 2014, 8(2): 128.
LAN C, ZHANG Y, XING C X. Elastic object store system for complex data[J]. Journal of Frontiers of Computer Science & Technology, 2014, 8(2): 128.
- [5] 蒋多元, 陈海雄. 云数据中心集群间网络性能优化的探讨[J]. 电信科学, 2015, 31(5): 138-142.
JIANG D Y, CHEN H X. Discussion of inter-cluster network performance optimization in cloud data center[J]. Telecommunications Science, 2015, 31(5): 138-142.
- [6] WEIL S A, POLLACK K T, BRANDT S A, et al. Dynamic metadata management for petabyte-scale file systems[C]//2004 ACM/IEEE Conference on Supercomputing (SC'04), November 6-12, 2004, Tucson, AZ, USA. New York: ACM Press, 2004.
- [7] XIN Q, MILLER E L, SCHWARZ T J E. Evaluation of distributed recovery in large-scale storage systems[C]//IEEE International Symposium on High Performance Distributed Computing(HPDC), June 4-6, 2004, Honolulu, Hawaii, USA. New Jersey: IEEE Press, 2004.
- [8] 藤本和久, 细谷睦, 岛田健太郎, 等. 集群型存储系统及其管理方法: CN200410045333.2[P]. 2004-05-20.
FUJIMOTO W, SAGO M, SHIMADA K, et al. Cluster storage system and its management method: CN200410045333.2[P]. 2004-05-20.
- [9] 胡学庆. 集群环境下海量数据存储管理技术的研究[D]. 大庆: 东北石油大学, 2010.
HU X Q. Research on massive data storage management technology in cluster environment[D]. Daqing: Northeast Petroleum University, 2010.
- [10] 黄昌华, 康慕宁, 韩朴杰. Hyper_v 虚拟机集群管理系统的设计与实现[J]. 计算机与现代化, 2013(4): 166-170.

HUANG C H, KANG M N, HAN P J. Design and implementation of integrated management system for cluster of Hyper_v virtualMachine[J]. Computer and Modernization, 2013(4): 166-170.

[作者简介]



胡豆豆（1985-），男，现就职于中国电信股份有限公司上海研究院，主要从事视频监控系统智能、云存储方向产品开发及研究工作。参与多项工业和信息化部物联网专项、“863”计划等科研项目工作，累计发表专利4项，发布 ITU-T 视频监控类国际标准2项，2013年获得上海市科学技术进步奖二等奖。



罗飞（1978-），男，博士，华东理工大学副教授、硕士生导师，主要从事嵌入式系统、云计算、大数据等方面的开发与研究工作。曾获得2012年上海市科学技术进步奖一等奖、2013年上海产学研合作优秀项目奖。



张园（1986-），女，现就职于中国电信股份有限公司上海研究院，主要从事视频监控系统智能、云存储方向产品开发及研究工作。参与多项工业与信息化部物联网专项、“863”计划等科研项目工作，累计发表专利3项，发布 ITU-T 视频监控类国际标准2项。



云网协同时代运营商 IP 承载网发展

朱永庆¹, 黄晓莹¹, 张文强²

(1. 中国电信股份有限公司广州研究院, 广东 广州 510630;

2. 中国电信集团公司, 北京 100032)

摘要: IP 承载网作为 Internet 的重要基础设施, 在云网协同发展背景下其作用越发重要。万物互联时代, IP 承载网的演进逐步从以网络扩展为中心向以用户为中心转变。在对运营商 IP 承载网络现状分析的基础上, 结合业务与技术研究, 提出了基于“体验圈”的网络发展思路。“体验圈”从流量的区域汇聚特性切入, 以用户体验为核心对架构进行优化, 同时, 对传输层以及应用层的协同提出建议。

关键词: IP 承载网; 软件定义网络; 体验圈; IP—传输协同

中图分类号: TN919

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017256

Development of carrier's IP bearer network in cloud-network coordinated era

ZHU Yongqing¹, HUANG Xiaoying¹, ZHANG Wenqiang²

1. Guangzhou Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China

2. China Telecom Corporation, Beijing 100032, China

Abstract: IP bearer network is one of the basic internet infrastructures. It plays a more and more important role in the cloud-network coordinated era. Entering the era of everything interconnected, the IP bearer network evolution has changed from network-centric to user-centric. Combined the current situation of the IP bearer network with the research of related technologies and service, some suggestions on IP bearer network development based on the concept of “experience circle” were given. The “experience circle” concept came from the research on traffic regional aggregation, which offered a way to optimize the IP bearer network based on user's experience, as well as suggestions on IP-transport synergy and application coordination.

Key words: IP bearer network, software defined networking, experience circle, IP-transport synergy

1 引言

IP 承载网由 IP 骨干网、城域网与 IDC (internet data center, 互联网数据中心) 等构成, 与用户网络共同构成了 Internet 的主体。作为用户网络的连接纽带与承载平台, IP 承载网在 Internet 的发展

过程中发挥了至关重要的作用。云计算、视频、物联网等业务的发展以及 SDN (software defined networking, 软件定义网络)、NFV (network functions virtualization, 网络功能虚拟化) 等技术的出现催生了“云网协同”时代。长期以来, IP 承载网一直在业务需求与技术发展等因素的推动

收稿日期: 2017-03-31; 修回日期: 2017-08-28

下持续平稳演进。在云网协同发展的今天,业务与技术都出现了新的变化,并为IP承载网的发展带来新的机遇与挑战,因此需要深入探讨。

2 互联网业务发展趋势

互联网诞生于20世纪70年代,随着网络空间与物理空间的逐步融合,互联网正在从最初的单纯数据通信网络逐步向“万物互联(internet of everything, IoE)”网络演进。这个过程伴随着IDC设施的云化以及终端类型的多样化。现阶段,3类互联网业务推动着互联网发展。

- 云计算相关业务:云计算是一种技术,更是一种理念。在云计算出现之前,尽管IP网络具有传送、计算与存储等功能,但长期以传送功能为主。云计算技术的出现,极大加速了IP网络功能向计算、存储等领域的拓展,使得“云”“网”逐步融合进而形成“XaaS(一切皆服务)”的态势。云计算相关应用的出现与发展进一步推动了网络东西向流量的增长,使网络流量流向呈现变化快、突发性强以及更加离散等特征,对网络的资源动态配置能力提出了极高要求。
- 视频业务:视频是当今第一大互联网应用。思科公司2016年发布的VNI(visual networking index)报告显示,2020年互联网视频流量将占全球互联网流量的79%。用户对视频体验的不断追求推动着视频业务向超高清和沉浸感体验发展,4K、8K、VR(virtual reality, 虚拟现实)以及AR(augmented reality, 增强现实)等视频应用将不断涌现,带来更加极致的视频体验。相对于其他业务,互联网视频具有大流量、高感知、长连接和高并发的特点,且在热点效应下极易出现突发的“大象流”,对IP承载网络产生非常大的冲击,需要网络具有较强的弹性承载能力。

- 物联网相关应用:物联网扩大了互联网的内涵,颠覆了传统的网络应用思路,丰富了终端类型,将催生出更加丰富、更深层次的网络应用和业务。从现有需求与技术发展来看,未来一段时间,物联网仍以M2M(machine-to-machine)通信为主(数据量小,本地连接较多),尽管连接数量巨大,甚至出现部分可靠性、即时性需求,但还不足以对现有IP网络形成较大冲击。未来,5G场景下物联网应用是否出现“井喷式”发展并对IP承载网产生较大影响,需要持续关注。

由此可见,云计算、视频等业务所导致的高通量、高突发、高并发、长连接以及流量流向离散等流量行为,将对IP网络产生深远影响。尽管可以通过CDN部署、DC分布化等方式缓解流量对网络的冲击,但在业务云化的趋势下,仍难以通过局部优化实现云网协同发展,需要从IP网络整体架构寻求解决之道。

3 IP承载网络现状及挑战

IP承载网涵盖骨干网、城域网、IDC等网络,是信息传输的高速通道。云网协同需要IP骨干网、城域网以及IDC的共同推进,其协同演进及发展对实现云网协同至关重要。

3.1 IP承载网现状

IP承载网中,城域网及IDC网络除流量转发外,主要承担用户、应用及网络资源的聚合角色,通常属于区域性网络,需要IP骨干网作为连接纽带。IP骨干网属于跨地域网络,各运营商经营理念以及发展思路的不同导致了IP骨干网形态的差异。从业务运营的角度,IP骨干网可分为单一网络融合承载与多业务分离承载两类。单一网络融合承载将所有业务承载在同一张IP骨干网上(必要时可根据业务质量实施差异化承载),美国AT&T的IP骨干网即属此类网络。多业务分离承



载则是根据业务定位,将不同服务质量要求的业务承载在不同 IP 骨干网上,中国电信利用 ChinaNet 和 CN2 分别承载不同业务,就是在 IP 骨干网层面实施多业务分离承载的范例。另外,根据拓扑结构,IP 骨干网又可分为汇聚型网络和格状网络。汇聚型网络设计的前提在于网络中存在流量规模较大的核心节点,因而汇聚型网络的普通节点通常通过核心节点实现流量交互,而较少关注节点间距离等因素。格状网络则将节点间距离作为主要设计参数,不设置核心节点,相邻节点间保证绝对可达(但不强调所有节点全连接)。随着 IP 流量持续增长以及网络架构优化等因素,上述两种拓扑结构有逐步融合的趋势。

3.2 IP 承载网面临挑战

尽管运营方式与网络拓扑不同,但随着业务发展与 IP 流量的持续增长,运营商 IP 承载网络的发展进入一个瓶颈期,主要体现在以下几方面。

(1) 扩展性不足

网络应用日益丰富与 IP 流量的不断增长需要网络持续扩展。IP 承载网的扩展性主要取决于网络容量的可扩展性,而容量的扩展则体现为节点规模的扩展与节点自身的扩展。IP 网络属于分布式网络,每个节点都同时担负着路由决策与转发功能,当节点数量较大时,业务布放速度与网络收敛速度会受到较大影响。因此,在 IP 承载网中,节点数量相对稳定,通常不会出现数量大幅增加的情况,网络的扩展性更多依赖于节点自身的扩展。IP 承载网的节点由路由器组成,受限于散热、芯片工艺等技术的发展,路由器一般 2~3 年容量提升一倍,其容量的提升速度低于全球 IP 流量的增长速率。集群路由器(通过中央框、线卡框的方式扩展路由器容量,而对外呈现为单台路由器)的出现可在一定程度上缓解此矛盾,尽管集群路由器容量理论上可达到 64 个线卡框(线卡框在非集群状态下可作为独立路由器使用,因此 64 个线卡框的集群路由器容量相当于 64 台独立路由器)

甚至更高,但由于集群路由器对机房的空间、电源和承重都有极高的要求,因此,在工程实现上,单机房集群路由器通常不超过 9 个线卡框,这对网络扩展性提出了较大挑战。

(2) 业务适配性差

传统 IP 承载网强调网络的高速转发功能,较少考虑对上层业务的适配,主要表现在:IP 网络遵循 TCP/IP 协议栈的分层原则,上层应用与 IP 层相对独立,难以根据 IP 网络资源等信息优化配置。这样不仅容易导致局部网络性能劣化,也不利于业务体验的提升;路由器等传统网元强调互联接口与协议的标准化,采用封闭的软硬件设计方式,网络能力与硬件能力强相关,网络—应用之间缺乏有效的协同/沟通机制。当新业务对网络提出新的能力要求时,通常需要升级甚至替换 IP 网元才能满足要求。

(3) 流量调度手段有限

有效的流量调度可实现网络资源利用率以及业务体验的提升,是网络运营能力的重要体现。随着业务云化,大规模分布式计算使得流量的突发性及离散性增强,网络需具备更智能的流量调度能力。目前,IP 网络采用 PHB(per hop behavior,逐跳转发)方式,各节点独立决策,因此,无法实现流量的全局调度。另外,IP 路由协议在路由决策时几乎不考虑链路质量、带宽以及时延等参数(OSPF、ISIS 等链路状态协议在设计之初考虑了相关因素,但并未真正付诸应用。MPLS 转发依赖于路由协议所确定的可达性,也不能从根本上解决该问题),对网络资源缺乏全局动态感知,从而无法对流量实施有效动态调度。

4 新技术的影响

IP 网络的发展是一个技术与业务双轮驱动的过程。近年来,随着软硬件技术特别是 IT 技术的发展,催生了虚拟化等一系列新技术,为 IP 网络的发展注入了新的活力。以虚拟化为代表的一系

列新技术开始改变信息化基础设施的构建理念，同时影响着 IP 网络的发展方向。

4.1 虚拟化技术

虚拟化本质上是在对软硬件资源逻辑抽象的基础上，通过对逻辑单元、流程等以更灵活颗粒度进行优化组合，从而实现功能重构和性能增强的一种方式。IP 网络目前诸多问题的根源在于其基于传统网元的分布式组网方式以及“面向问题”的发展模式（目前 IETF 有效的标准文档超过 8 000 个，基本都是针对特定问题而提出的协议标准），而虚拟化技术则为 IP 网络后续的发展提供了技术基础。虚拟化技术目前可分为计算虚拟化、存储虚拟化和网络虚拟化 3 类。网络虚拟化基于物理拓扑对网络连通能力进行抽象，可实现网络资源的多层次灵活隔离、共享与快速部署。网络虚拟化与服务器虚拟化、存储虚拟化等技术结合，可充分发挥 IP 网络的计算与存储能力，从而大大提升网络弹性与可扩展性。虚拟化技术也将推动云网协同进程，在对 IP 承载网架构产生深远影响的同时，进一步加速了网络流量流向的离散与突变趋势，对网络的敏捷度与响应能力提出更高要求。

4.2 SDN 技术

SDN 定义了一种在转控分离模式下信令、资源等信息的交互机制。这种机制保证了上层应用通过可编程接口与下层网络资源的交互效率，从而使网络更高效、敏捷，同时提升集约能力。

SDN 在实践过程中逐渐发展出三大流派。以 ONF（Open Networking Foundation，开放网络基金会）为代表的革命派，主张对网络进行根本性变革，网络控制层被完全抽离，网络节点只做纯粹的数据交换。从产业链的支持力度看，该模式短时间内难以在 IP 承载网推广。由设备厂商主导的 overlay 派，主张通过叠加的逻辑平面屏蔽底层网络细节，目前主要应用于数据中心互联，解决云环境下数据中心的扩展性问题，尚未在运营商 IP 网络有成熟商用。由设备厂商主导的改良派，

提出通过开放网络设备接口的方式向上层应用提供适配的资源与能力（网络节点仍可保留一定程度的控制平面与自主控制权）。这种方式对现有网络改动较少，较适合在过渡期间引入部署（初期聚焦于局部流量调度，通过引入集中化的控制技术，提升流量调度的灵活性）。

4.3 NFV 技术

NFV 是由运营商主导提出的网络功能虚拟化技术，旨在通过软硬件解耦的方式，降低设备成本并提升资源调度能力和业务部署灵活性。NFV 进一步加速了 IT 与 CT 技术的融合，为 IP 网络的发展提供新的思路。

目前，基于 x86 架构的 NFV 解决方案，其处理能力远低于基于 NP/ASIC 的专用网络设备，主要适用于计算密集型和业务功能丰富的应用场景（如 vRR、vCPE、vBNG 等）。核心路由器等以高速转发功能为主的 IP 承载网网元，在通用硬件或开源硬件技术取得突破之前，不会 NFV 化。

5 IP 承载网络演进思路

在可预见的未来，“Everything over IP, IP over everything”仍是互联网发展的基调。IP 网络在过去的演进过程中始终强调分层解耦以及自身容量和效率的提升，在与上层应用、底层资源的协同方面较为欠缺，整体运行效率难以最优化。云网协同背景下，网络发展逐渐转向以用户体验为中心，多业务融合承载将成为主流。为满足互联网的发展需求，IP 承载网的发展需要聚焦在网络架构优化、IP—传输协同、网络—应用协同等方面。

5.1 架构优化

云网协同条件下，网络服务的即时性成为用户体验的第一要素。长期以来，IP 承载网的核心诉求是对数据分组的快速和可靠转发，流量整体优化是其主要目标。长期以来，IP 承载网的架构优化以网络规划周期内流量的宏观效率提升为基础，IP 骨干网、城域网、IDC 优化缺乏有效协同，



架构保持相对稳定但难以应对短期内流量与应用大幅度的变化。为提升包括即时性在内的用户体验，同时应对流量的快速变化，需要以“体验圈”为核心推动 IP 承载网的组网优化。

IP 网络中，体验圈即具有相同/相似流量与用户特征、以特定地域范围为核心形成的网络区域。体验圈所涵盖的区域，其经济、文化或政治联系较为紧密，用户之间沟通频繁，内容源相对集中，用户行为与流量特征也存在一定相似性。一般来讲，流量汇聚比例超过 50%即可从 IP 网络角度划分为同一区域。以此类区域为基础进行网络架构调整，将使网络流量流向更为有序和集中，从而可保证区域内的用户体验。在现有网络架构下，IP 骨干网通常与 IDC、城域网相连。组建体验圈，即将具有相同/相似流量特征的 IP 骨干网、IDC 和城域网通过一定的连接法则聚合，形成特定地域范围内、具有较高网络体验的网络体系。由于流量、用户等特征的趋同性，通过把内容源、骨干网、城域网有效地组织起来，可以更好地适应流量的变化。IP 骨干网作为体验圈的重要组成部分，是连接体验圈内各网络区域（如城域网与 IDC 间、IDC 之间以及城域网之间等）的重要通道。因此，基于体验圈的 IP 承载网络优化应以骨干网为切入点。

具体而言，在骨干网内首先根据体验圈所确定的地域范围设置区域中心节点。区域中心节点是 IP 骨干网的枢纽节点，负责圈内外流量的疏导与转接。为保证区域内主要流量（如大型城域网访问大型 IDC、大型城域网间互访等）不发生路径绕转，区域中心节点应设置在大型 IDC 或城域网所在地市。考虑到组网顽健性等因素，在同一体验圈内需至少部署一对骨干网区域中心节点。为扩大体验圈的覆盖范围及提升连接效率，除区域中心节点外，某些区域还存在区域接入节点。区域接入节点可基于行政区域、地域范围和节点重要性等因素设置，该类节点主要负责本地流量的接入与转发，通常不与区域外节点连接。为保证体验圈内的流量效率，

同一体验圈内的节点连接原则为：中心节点之间全互联；接入节点采用双归方式与一对或多对中心节点连接；接入节点间可根据流量规模按需互联，最终形成星型或半网状连接。

体验圈内，区域中心节点、接入节点构成了 IP 骨干网的区域性网络。体验圈内的城域网、IDC 等则根据流量规模、重要性等确定与区域中心节点或接入节点的连接方式。通常情况下，为保证不同区域之间流量路径最短，区域中心节点之间可采用全网状（full mesh）方式互联。基于体验圈的 IP 承载网模型如图 1 所示。

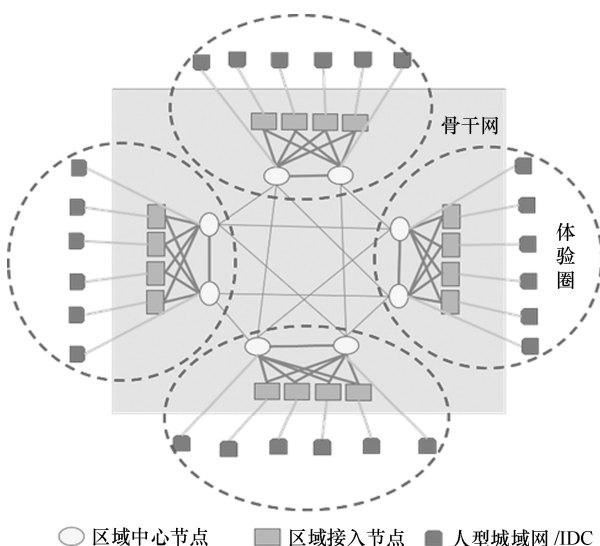


图 1 基于体验圈的 IP 承载网模型

以体验圈为核心的架构优化已在实际网络建设中实施。以西南某区域为例，该区域内所有省份流量汇聚比例达 62%，属典型的体验圈区域。将该区域看作一个整体进行各网络架构协同设计，将使带宽布放更有效率。具体来讲，则是以区域为整体进行连接优化，提升区域内连通比例，适当减少跨区域长途连接。区域内设置 IP 骨干网双核心，所有省份的接入节点双挂核心，城域网、IDC 网络通过骨干接入节点实现区域内互联；区域外，采用核心双节点一体化设计，以区域为对象实现跨区域连接与流量疏导。优化后区域呈现明显统一的组织特性，区域内连接度提升，区域间连接以区域为单位

通过核心节点实施。架构优化后,区域内流量本地化比例得到较大提升,用户体验更佳。

可以看出,体验圈的构建过程是将 IP 骨干网、城域网以及 IDC 网络作为整体实施优化的过程。它以资源的有效聚合为优化依据,最终会促进资源的进一步优化部署。而以体验圈为中心的 IP 承载网架构相对传统的格状网络或汇聚型网络,在适应流量变化与提升用户体验方面更胜一筹。同时,以体验圈为基础的组网架构,对传输网络也提出了新的需求。

5.2 IP—传输协同

为减少 IP 流量的物理路径绕转,需要 IP—传输协同组网,以保证 IP—传输之间的网络拓扑一致。组网协同属于静态协同方式。在组网协同的基础上,为实现更为精细与动态的资源调度,使传输资源更好地服务于 IP 流量承载,需要通过技术手段实现资源协同。资源协同主要解决 3 方面的问题:首先,传输技术为了适配 IP 业务统计复用的特性,开始逐步引入分组化内核,打破既有的波分、时分复用方式,通过复用方式的多样性提升数据链路层及以下层次资源的整体利用率;其次,通过更灵活的容器和封装/映射技术提高数据分组在传输单元的承载效率,从而实现任意颗粒度 IP 宽带业务在传输网络中的有效承载;第三,通过 IP 层与传输层协同调度的方式,提升从物理层到网络层资源的整体利用率。

IP 层与传输层协同调度的目的是,通过 IP 层与传输层间有效的信息互通实现资源的动态按需分配,同时提升流量传输效率和网络生存性。为实现 IP 层与传输层之间的协同调度,业界提出 GMPLS (generalized multiprotocol label switching, 通用多协议标签交换) 协议,试图通过统一的控制平面实现对 IP 层、传输层各层次用户平面(或称为转发平面)的协同调度。但是,GMPLS 主要以 UNI (user network interface, 用户网络接口) 作为 IP—传输之间信息互通的机制,并未定义能力抽象的架

构,加大了控制层的复杂性,从而导致扩展性不足,因而未能实现规模部署。随着 SDN 概念的提出,基于 SDN 的 IP—传输协同方案成为业界热点。这种方案需要在 IP 层与传输层分别部署 SDN 控制器(controller)实现能力抽象;同时,为避免 GMPLS 实现的复杂性,通过协同编排器(orchestrator)或更高层面的控制器实现 IP 层与传输层控制器的信息交互与统筹决策。与 IP—传输协同相似的解决方案,还有 IP—传输融合。该技术通常将 IP 与传输集成在同一设备中,通过 IP 提供彩光接口或 IP over OTN 等方式实现。这种方式可在一定程度上(尤其在单厂商设备组网时)降低投资成本,但对网络运维要求较高(缺乏传统的 IP 与传输专业运维界面),且标准化程度较低,难以广泛应用。

目前,IP—传输已实现组网协同,而 IP 与传输正分别进行 SDN 场景与相关技术的验证,未来将引入统一的控制面,通过南北接口与两专业控制器分别对接,从而实现更深层次的调度协同。

5.3 网络—应用协同

网络—应用协同可实现网络承载效率与应用部署的整体优化,进一步提升用户体验。云网协同背景下,网络—应用协同主要体现在组网协同与能力开放两方面。

为实现网络—应用的协同组网,通常利用 CDN 实现“存储换带宽”与“内容就近推送”,为视频业务的规模开展奠定基础。在网络应用日益丰富的形势下,还需要以网络能力开放的方式支持更多业务与应用的部署。网络能力开放就是在网络分层解耦的基础上,将网络能力抽象、封装以供用户和上层应用调用的一种网络—应用协同方式。以往,单个网元的用户管理、路径管理、业务管理、网元管理以及转发能力等被封闭在“盒子(box)”中,以此为基础的网络资源的调配、业务部署等工作通常工作量大且难以全局最优。为了实现网络—应用协同,需要逐步将网元的相关能力抽象出来,并通过封装方式提供给上



层应用。网络能力开放使得上层应用可对底层资源进行快速调用, 加快业务上线速度, 高效率的业务运转将进一步优化运营商的资源配置。

现阶段, 网络能力开放还处于初级阶段, 主要在 BGP、PCEP 等协议扩展的基础上将路径决策能力抽象、封装, 供上层应用或客户使用, 是对现有流量调度能力的提升, 属于增强型能力开放。目前, 现网主要基于增强型能力开放的方式实现局部网络的协同调度 (如在区域核心节点间部署集中控制器, 利用 BGP flowspec、PCEP 等较成熟的南向接口协议对出口设备进行策略下发, 实现特定场景下的局部流量调度)。未来, 需要将网元的业务管理、用户管理等能力向上层开放, 实现网络可编程, 从而实现网络资源多维度以及多颗粒度的调度。

随着网络能力的逐步开放, 网元形态逐渐向通用化、软件化演进。但是, 大带宽、高可用以及低功耗等特性仍是 IP 承载网的基本要求, 目前虚拟化技术网元还不能满足需求。因此, 在可预见的未来, IP 承载网网元仍然会维持目前实体的形态, 技术发展仍聚焦于大容量低功耗的关键技术。

6 结束语

满足用户与业务需求始终是运营商 IP 承载网发展的核心诉求。尽管云网协同对 IP 承载网提出新的挑战, 基于 TCP/IP 架构的 IP 承载网的演进仍将在现有基础上渐进式发展。本文提出的以“体验圈”为理念的 IP 承载网架构优化以及应用—网络—传输协同的解决方案, 从用户需求的角度切入, 推动 IP 骨干网、城域网及 IDC 架构的统筹优化, 并已在现网中得到应用且取得良好成效。未来, 将进一步通过网络—应用协同能力的提升推动云网融合发展, 为 Internet 的规模、高效发展奠定基础。

参考文献:

[1] SALTZER J H, REED D P, CLARK D D. End-to-end argu-

ments in system design[J]. ACM Transactions on Computer Systems, 1984, 2(4): 277-288.

[2] CLARK D. The design philosophy of the DARPA internet protocols[C]//Symposium Proceedings on Communications Architectures and Protocols, August 16-18, 1988, Stanford, California, USA. New York: ACM Press, 1988: 106-114.

[3] GHODSI A, KOPONEN T, RAJAHALME J. Naming in content-oriented architectures[C]//ACM SIGCOMM Workshop on Information-Centric Networking, August 19, 2011, Toronto, Ontario, Canada. New York: ACM Press, 2011: 1-6.

[4] MARK M C, CHRISTOPHER A W, HARISH V, 等. SDN 在 IP 网络演进中的作用[J]. 电信科学, 2014, 30(5): 1-13.
MARK M C, CHRISTOPHER A W, HARISH V, et al. Role of SDN in IP network evolution[J]. Telecommunications Science, 2014, 30(5): 1-13.

[5] 刘汉江, 欧亮, 陈文华, 等. 基于 SDN 的跨数据中心承载技术[J]. 电信科学, 2016, 32(3): 28-34.
LIU H J, OU L, CHEN W H, et al. Bearing technology across the data center based on SDN[J]. Telecommunications Science, 2016, 32(3): 28-34.

[6] 张敏, 梁建君, 孙丽娟. 能力开放演进与电信运营商的能力开放策略[J]. 通信企业管理, 2011(10): 72-75.
ZHANG M, LIANG J J, SUN L J. Ability open evolution and telecom carrier ability open strategy[J]. C-Enterprise Management, 2011(10): 72-75.

[7] 盖晶, 刘晶淳. 下一代分组传送网的新技术发展走向[J]. 电子制作, 2016(10x): 52.
GAI J, LIU J C. New technology trend of next generation packet transport network[J]. Practical Electronics, 2016(10x): 52.

[作者简介]



朱永庆 (1974-), 男, 中国电信股份有限公司广州研究院高级工程师, 长期从事 IP 网络研究工作, 主要研究方向为网络重构、SDN/NFV 技术的发展应用等。



黄晓莹 (1980-), 女, 中国电信股份有限公司广州研究院工程师, 主要从事中国电信骨干网络规划工作。

张文强 (1979-), 男, 中国电信集团公司网络发展部高级工程师, 长期从事 IP 网络规划、研究工作。



上行 SCMA 信道估计及性能评估

胡显安, 曾祥念

(上海诺基亚贝尔股份有限公司, 浙江 杭州 310053)

摘要: 针对两种 SCMA 资源映射方式 (RE-Sparse SCMA 和 RB-Sparse SCMA) 和已有的基于 ZC 序列循环移位正交性和正交覆盖码的导频图样方案, 提出了基于频分复用的稀疏导频图样, 并且在具有不同频率选择性的信道下进行了仿真评估, 同时对 SCMA 的 SIC-MPA 接收算法提出了目标用户选择策略。从仿真结果可以看出, 在慢衰落信道前提下, 正交覆盖码和频分复用都可以明显提升信道估计性能, RB-Sparse SCMA 与 RE-Sparse SCMA 相比, 导频图样本身有一定的频域正交性, 且性能更好, 因此能更灵活地适用于实际多用户传输场景。

关键词: 5G; 非正交多址; SCMA 上行链路; 导频图样; 信道估计; SIC-MPA

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017267

Channel estimation and performance analysis for uplink SCMA system

HU Xian'an, ZENG Xiangnian

Nokia Shanghai Bell, Hangzhou 310053, China

Abstract: There are two types of physical mapping for SCMA named RE-Sparse SCMA and RB-Sparse SCMA. A frequency division multiplexing (FDM) based pilot pattern was proposed for RE-Sparse SCMA, and existing code division multiplexing pilot patterns based on cyclic shift and orthogonal cover code (OCC) were considered. Under different frequency selective channels, simulation and evaluation work for the two mapping schemes of SCMA had done. Besides, a target user selection strategy was proposed for existing SIC-MPA. Analytical simulation results show that both OCC and FDM can improve system performance with slot fading channel. Compared with RE-Sparse SCMA, RB-Sparse SCMA with partially frequency orthogonal pilot pattern enjoys better performance gain, and it is more applicable for realistic multi-user SCMA scenario.

Key words: 5G, non-orthogonal multiple access, uplink SCMA, pilot pattern, channel estimation, SIC-MPA

1 引言

与目前 LTE 相比, 未来 5G 无线网络势必要

具备支持多样化数据业务和终端设备的需求, 如大规模用户接入、更高的频谱效率和更低的时延等^[1], 特别是在上行链路中。SCMA (sparse code

收稿日期: 2017-07-05; 修回日期: 2017-09-21

基金项目: 国家科技重大专项基金资助项目 (No.2016ZX03001018)

Foundation Item: The National Science and Technology Major Project of China (No. 2016ZX03001018)



multiple access, 稀疏码分多址) 在参考文献[2]中被提出, 并且作为 5G 候选非正交多址技术之一目前受到了业界的广泛关注^[3]。SCMA 主要是通过稀疏码本的设计和次优的 MPA (message passing algorithm, 消息传递算法) 来实现多用户过载传输。

目前 SCMA 的研究主要集中在两个方面: 一是码本和 MPA 接收算法的优化设计, 另一个是随机竞争接入场景下的用户识别和数据检测。其中, 基于软消息传递迭代检测的 MPA 复杂度很高。参考文献[4,5]提出了基于 PM (partial marginalization, 部分边缘化) 的快速收敛优化算法。同时参考文献[6]也进行了 MPA 接收算法优化, 提出了基于 SIC (successive interference cancellation, 串行干扰消除) 的 SIC-MPA 快速收敛算法。

事实上, 接收端的数据检测依赖于具体的信道信息, 特别是对于上行链路来说, 每个用户的信号都是经历了不同的路径到达基站, 因此信道估计的准确性对系统性能有着至关重要的影响。参考文献[7-9]都是研究 SCMA 在随机接入场景下的用户识别问题和基于导频的简单信道估计, 如平坦衰落信道或块衰落信道。同样, 参考文献[10]考虑了上行 SCMA 在色散信道下不同衰落块的独立用户检测和信道估计, 从而减小导频开销。参考文献[11]则针对导频开销问题, 设计了稀疏的导频图样, 并且通过 WR (weight regularization, 权重调整) 算法来减小载波间干扰。然而, SCMA 在频率选择性衰落信道下的信道估计还没有被研究。参考文献[12]中, 只是简单讨论了可能适用于上行非正交多址的 DMRS (demodulation reference signal, 解调参考信号) 导频图样方案, 并没有进一步做仿真研究和验证。SIC-MPA 算法是目前 SCMA 的主流接收算法, 通过每次对指定的待检测用户进行 MPA 联合检测, CRC (cyclic redundancy check, 循环

冗余校验) 正确后从接收信号中减去, 直到所有用户数据检测正确。

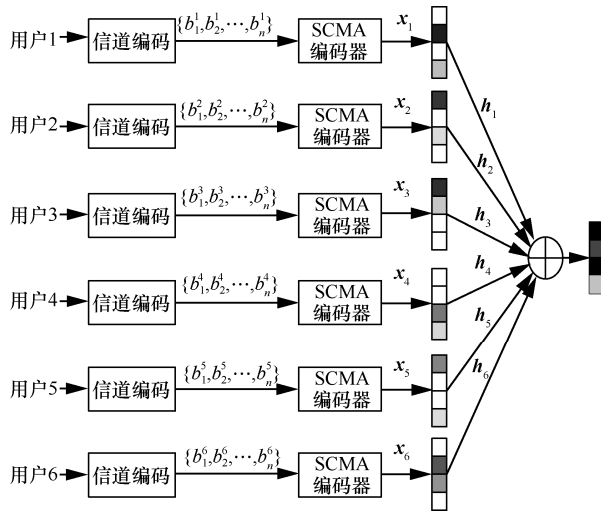
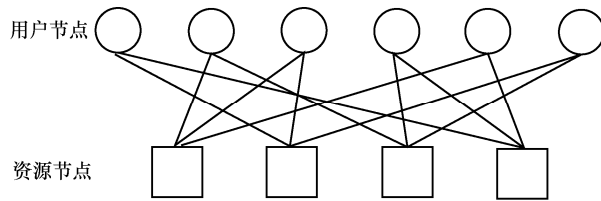
本文对 SIC-MPA 提出了目标用户选择策略, 当系统过载大于 150% 时, 接收机从所有未解码的用户中选择一部分作为目标 MPA 解码用户, 并且基于信号功率的穷尽搜索算法选取信号功率之和最大的目标解码用户组合进行 MPA 联合检测, CRC 正确后从接收信号中减去, 这样就减小了每次 MPA 联合检测需要遍历的因子图大小, 加快了迭代的收敛速度。同时, 根据资源映射单位的不同, SCMA 可分为以 RE (resource element, 资源元素) 为资源单位的 RE-Sparse SCMA 和以 RB (resource block, 资源块) 为资源单位的 RB-Sparse SCMA, 针对这两种 SCMA 资源映射方式提出适用于 RE-Sparse SCMA 的频分复用导频方案, 并且基于已有的导频图样方案, 如基于 ZC (Zadoff-Chu) 基序列循环移位实现码分复用和正交覆盖码的传统 LTE 导频图样、基于频分复用和码分复用这两者特性的稀疏导频图样, 在频率选择性衰落信道下进行一系列的实际信道估计仿真评估。

2 系统模型

2.1 SCMA 编码器

假设在 SCMA 上行链路中, 用户 J 复用相同的时频资源。每个用户发送的数据先经过 Turbo 码信道编码, 然后经过 SCMA 编码器实现符号映射和稀疏扩频。SCMA 具体过程是将 $\log_2 M$ 比特映射到大小为 M 的 K 维复数码本上, $K < J$ 。每个码本中 K 维复数码字都是一个含有 N 个非零值的稀疏向量, $N < K$ 。SCMA 上行链路模块如图 1 所示, 其中, $K=4$, $N=2$, $J=6$, SCMA 码本映射关系可通过矩阵 $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_J)$ 表示出资源节点和用户节点之间的关系。

图 2 为 SCMA 因子图表示法, 可以看出每个资源节点上碰撞用户数为 $d_f = 3$, $f = 1, \dots, K$ 。

图1 SCMA上行链路模块, $K=4, N=2, J=6$ 图2 SCMA因子图表示法, $K=4, N=2, J=6$

当式(1)中元素为1时,则因子图中用户节点和资源节点之间连接:

$$F = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \quad (1)$$

因此, J 个用户复用 K 个相同的时频资源块时,接收信号可表示为:

$$y_k = \sum_{j=1}^J h_{kj} x_{kj} + n_k \quad (2)$$

其中, x_{kj} 表示用户 j 的码字 x_j 的第 k 个元素, h_{kj} 表示用户 j 在第 k 个载波上的信道增益, n_k 是第 k 个子载波上的复高斯白噪声,均值为0,方差为 σ^2 。总接收信号 $y = (y_1, y_2, \dots, y_K)^T$ 可表示为:

$$y = \sum_{j=1}^J \text{diag}(h_j) x_j + n \quad (3)$$

其中, $x_j = (x_{1j}, \dots, x_{Kj})^T$ 是用户 j 的 SCMA 码字, $h_j = (h_{1j}, \dots, h_{Kj})^T$ 是用户 j 信道矢量。 $n \sim \mathcal{CN}(0, \sigma^2 I_K)$ 为加性高斯白噪声。通过在 K 个

正交资源上复用 J 个用户,则系统过载因子 $\lambda = J/K$ 。

2.2 SCMA 接收机

为了兼顾性能和复杂度,参考文献[6]提出了迭代接收算法 SIC-MPA。每次外环迭代包含两个内环迭代,一个是 SCMA 的 MPA 迭代解码,迭代解码次数记为 I_s ;另一个是 Turbo 码迭代解码,迭代次数记为 I_T 。外环迭代进行多次,直到所有用户都正确解码或者达到最大外环迭代次数 I_0 ,具体每次外环迭代的步骤描述如下。

首先,如果不是首次外环迭代,则要将已正确解码的用户从接收到的信号中消除。将所有 SCMA 用户的集合记为 $\xi = \{1, 2, \dots, J\}$,并将其中已正确解码的用户集合记为子集 ω ,则经过干扰消除后的信号记为 z ,具体表示为:

$$z = y - \sum_{j \in \omega} \text{diag}(h_j) \hat{x}_j \quad (4)$$

然后,从所有未解码的用户中选择一部分作为 MPA 目标解码用户,同时保证每个资源节点上碰撞用户数 $\hat{d}_f \leq d_{\max}$, $d_{\max}$ 是小于 d_f 的整数,以达到减小 MPA 计算复杂度的目的。由于参考文献[6]中未给出 SIC-MPA 具体的目标解码选择算法,因此本文提出了基于信号功率的穷尽搜索算法来选取目标解码用户。把子集 ω 在 ξ 中的补集记为 $\bar{\omega}$,用 $|\bar{\omega}|$ 表示 $\bar{\omega}$ 中元素的个数。比如在首次迭代里, ω 是空集,则 $\bar{\omega} = \xi$ 。首先设置目标解码用户数初值 $k = |\bar{\omega}|$,找出从 $\bar{\omega}$ 中取出 k 个元素的所有组合,并从这些组合里找出满足资源节点的度小于或等于 $d_{\max}$ 条件的组合。如果没有满足条件的组合,则设置 $k = k - 1$ 继续查找满足条件的组合;如果有至少一个满足条件的组合,则从中找出信号功率之和最大的组合作为目标用户并停止搜索。因此选择策略的算法复杂度为 $O\left(\sum_{k=\min(d_{\max}, |\bar{\omega}|)}^{|\bar{\omega}|} C_{|\bar{\omega}|}^k\right)$ 。用户 j 的信号功率 $P_j = \sum_{k \in \phi_j} |h_{k,j}|^2$, ϕ_j 表示用户 j 的数据符号使用的子载波集合。



把目标解码用户集合记为 ε ，把剩下的未解码用户记为 $\zeta = \bar{\omega} - \varepsilon$ 并建模为干扰，则经过干扰消除后的信号表示为：

$$z = \sum_{j \in \varepsilon} \text{diag}(\mathbf{h}_j) \hat{\mathbf{x}}_j + \mathbf{I} \quad (5)$$

其中， $\mathbf{I} = \sum_{j \in \zeta} \text{diag}(\mathbf{h}_j) \hat{\mathbf{x}}_j + \mathbf{n}$ 表示干扰和噪音之和，协方差矩阵为 $\sigma_I^2 = \sum_{j \in \zeta} \text{diag}(|h_j|^2) \text{diag}(E\{\|\mathbf{x}_j\|^2\}) + \sigma^2 \mathbf{I}$ 。

接着，通过 MPA 算法计算目标解码用户的 SCMA 码字的似然比，随后将 SCMA 码字的似然比转换为比特似然比，用作 Turbo 解码器的输入。具体对数域的 MPA 算法在参考文献[15-17]有详细描述。

最后，Turbo 解码算法输出信息比特的硬判决和编码比特的似然比。如果 CRC 正确，则用信息比特硬判决重建该用户发送的信号，并反馈给下一次外环迭代用作干扰消除处理；如果 CRC 错误，则把编码比特的似然比转换为 SCMA 码字的似然比，并反馈给下一次外环迭代中的 MPA 用作先验信息。直到所有的用户都能够全部正确或者达到最大迭代次数为止。

3 导频图样方案

SCMA 系统中数据符号在进行资源映射时，除了沿用 LTE 连续子载波映射方式外，即以 1 个 RE 为基本资源单位，每 4 个连续子载波承载复用用户，简称 RE-Sparse SCMA；还可以使用另一种资源映射方式，以 1 个 RB 为基本资源单位，每 4 个连续的 RB 来承载复用用户，简称 RB-Sparse SCMA。假设分配带宽为 4 个 RB，则 RB-Sparse SCMA 系统中，每个用户数据符号映射到所分配的 2 个 RB 时频资源上。针对这两种 SCMA 资源映射方式，可以设计不同的导频图样方案用于实际信道估计。

对于 LTE 上行链路来说，DMRS 处放置导频信号用于 PUSCH (physical uplink shared channel, 物理上行共享信道) 估计。其中，DMRS 为基于 ZC

序列生成的导频序列，占用每个子帧的第 4 个和第 11 个符号，具体 DMRS 导频序列生成式如下：

$$r_{\text{PUSCH}}^{(\lambda)}(m \cdot M_{\text{sc}}^{\text{RS}} + n) = w^{(\lambda)}(m) e^{\frac{j2\pi n_{\text{cs},\lambda}}{12}} \bar{r}(n) \quad (6)$$

其中， $0 \leq n < M_{\text{sc}}^{\text{RS}} - 1$ ， $M_{\text{sc}}^{\text{RS}}$ 为导频序列长度； λ 为 LTE 可复用数据层， $\lambda \in \{0, 1, 2, 3\}$ ； $n_{\text{cs},\lambda}$ 为 LTE 中可用的 CS (cyclic shift, 循环移位)， $n_{\text{cs},\lambda} \in \{0, 1, 2, \dots, 11\}$ ， $\bar{r}(n)$ 为 ZC 基序列； $w^{(\lambda)}(m)$ 是时域 OCC (orthogonal cover code, 正交覆盖码)， $m \in \{0, 1\}$ ，一对用户的正交覆盖码分别为 [1, 1] 和 [1, -1]，用于接收端导频分离。

3.1 RE-Sparse SCMA CDM 导频图样

在 RE-Sparse SCMA 系统中，多用户复用相同的时频资源，如图 3 (a) 所示，每个用户导频占据连续的子载波带宽，使用相同的基序列，但使用不同的序列循环移位或正交覆盖码来生成各自的导频序列，接收端可以利用它们间的正交性来区分不同用户^[13]。因此，这是一种基于 CDM (code division multiplexing, 码分复用) 的导频图样。

本文考虑了两种码分复用的导频方案，假设复用用户数 $J = 6$ ，带宽为 1 个 RB，则导频序列长度为 12。第一种方案是只用循环移位进行用户区分，则所有 J 个用户使用最大间隔的 J 个循环移位 $\text{CS}_1, \text{CS}_2, \dots, \text{CS}_J$ ，最大间隔为 $12/6=2$ ；第二种方案是将用户分成两组，采用循环移位和正交覆盖码进行用户区分，假设在一个子帧用户信道没有发生变化，将所有 J 个用户分为两组，每组 $J/2$ 个用户，两组用户重复使用最大间隔为 4 的 $J/2$ 个循环移位，且在一个子帧的两个导频符号上第一组用户使用正交覆盖码 [1, 1]，第二组用户使用正交覆盖码 [1, -1]。

3.2 RE-Sparse SCMA FDM 导频图样

基于 FDM (frequency division multiplexing, 频分复用) 的导频图样同样适用于 RE-Sparse SCMA。与 CDM 导频图样不同，FDM 导频图样是给不同的用户分配正交的导频图样，从而保证各个用户之间的信道估计互不干扰，因此不需要

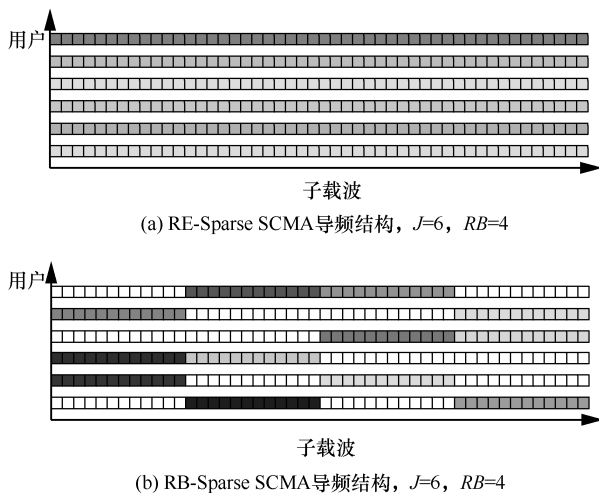


图3 SCMA 不同资源映射方式时导频结构比较

在接收端做导频分离处理。RE-Sparse SCMA FDM 导频图样设计如图 4 所示, 相同数字的部分表示同一个用户的导频分配位置, 同时为了信道估计的均匀性, 第 1 个时隙图样与第 2 个时隙图样之间错开了一半的间隔距离, 如图 4 (b) 中每个用户图样间隔距离为 12 个 RE, 则第 2 个时隙从用户 7 开始分配导频图样。从图 4 可以看出, 每个用户的导频能量只集中在少数图样点上, 而 CDM 导频图样中导频能量分摊在整个分配带宽上。



图4 RE-Sparse SCMA FDM 导频图样设计

3.3 RB-Sparse SCMA CDM+FDM 导频图样

与基本资源单位为 1 个 RE 的 RE-Sparse SCMA 不同, RB-Sparse SCMA 是以 1 个 RB 为基本资源单位, 每个用户进行稀疏扩频时只在所分配的 RB 资源块上发送数据, 因此接收端进行数据检测时, 只需要知道所分配资源块的信道信息。为了简化信道

估计算法, 只需要在每个用户所分配的 RB 上放置导频符号, 如图 3 (b) 所示。不同用户可以通过 RB 分配、ZC 序列循环移位正交性和正交覆盖码来进行区分, 如式(1)中用户 1 资源块分配 $f_1 = [0 \ 1 \ 0 \ 1]$ 与用户 2 资源块分配 $f_2 = [1 \ 0 \ 1 \ 0]$ 是频域正交的, 而与用户 3 资源块分配 $f_3 = [1 \ 1 \ 0 \ 0]$ 和用户 6 资源块分配 $f_6 = [0 \ 1 \ 1 \ 0]$ 在频域是部分重叠的, 因此重叠部分导频分离可通过不同循环移位来实现码分复用。

因此, 可以将互为正交频分复用的两个用户视为一对用户, 给它们分配使用相同的循环移位。从图 3 可以看出, (b) 中导频序列长度只有 (a) 中导频序列长度的一半。为了仿真比较的公平性, 以 LTE 导频符号总功率为基准, 保证每个子帧中 OFDM 符号的发送功率一致。这种导频相对于传统连续导频有两个优势, 一是导频符号上的信噪比增加了一倍; 二是给定用户个数, 在每个 RB 上发送导频信号的用户个数减少一半, 这样可以减少用户间的干扰。

同理, RB-Sparse SCMA 导频图样也考虑了两种码分复用方案。第一种是只用循环移位, 所有 $J/2$ 对用户使用最大间隔的 $J/2$ 个循环移位; 第二种是同时采用循环移位和正交覆盖码, 将所有 $J/2$ 对用户分为两组, 每组 $J/4$ 对用户, 两组用户重复使用最大间隔的 $J/4$ 个循环移位, 且在一个子帧的两个导频符号上第一组用户使用正交覆盖码 $[1, 1]$, 第二组用户使用正交覆盖码 $[1, -1]$ 。

4 信道估计方法

4.1 基于 DFT 逐个 RB 信道估计

对于 RE-Sparse SCMA CDM 导频图样和 RB-Sparse SCMA CDM+FDM 导频图样, 采用相同的信道估计方法, 即每个 RB 上采用基于 DFT (discrete Fourier transform, 离散傅里叶变换) 的方法逐个 RB 地做信道估计。以 1 个 RB 为例, 假设每个 RB 上有 J 个用户发送导频信号, 则第一种码分复用方案是首先需要将接收到的导频信号



乘以基序列的共轭, 然后对它做 12 点的 IDFT, 得到时域的 12 个采样 $\mathbf{s}=[s_0, s_1, \dots, s_{11}]$, 则可估计用户 j 在这个 RB 上的信道冲激响应为:

$$\hat{\mathbf{h}}_j = \left[S_{(12-CS_j) \bmod 12}, S_{(12-CS_j) \bmod 12+1}, \dots, S_{(12-CS_j) \bmod 12+\frac{12}{J}-1} \right] \quad (7)$$

最后对 $\hat{\mathbf{h}}_j$ 做 12 点的 DFT, 得到该用户在该 RB 上的信道频域响应估计^[14]。

对于第二种结合正交覆盖码的码分复用方案, 将第 g 组的第 j 用户在一个子帧的两个时隙上的循环移位序列和信道分别记为 $\mathbf{p}_j^{(g)}$ 和 $\mathbf{h}_j^{(g)}$, $g \in \{1, 2\}$ 。接收端在第 1 时隙和第 2 时隙收到的导频信号可分别表示为:

$$\mathbf{y}_1 = \sum_{j=1}^{J/2} \text{diag}(\mathbf{p}_j^{(1)}) \mathbf{h}_j^{(1)} + \sum_{j=1}^{J/2} \text{diag}(\mathbf{p}_j^{(2)}) \mathbf{h}_j^{(2)} + \mathbf{n}_1 \quad (8)$$

$$\mathbf{y}_2 = \sum_{j=1}^{J/2} \text{diag}(\mathbf{p}_j^{(1)}) \mathbf{h}_j^{(1)} - \sum_{j=1}^{J/2} \text{diag}(\mathbf{p}_j^{(2)}) \mathbf{h}_j^{(2)} + \mathbf{n}_2 \quad (9)$$

其中, $\mathbf{n}_1$ 和 $\mathbf{n}_2$ 分别表示第 4 个符号和第 11 个符号上的高斯白噪声, 功率为 σ^2 。接收端的信道估计过程分为两步。第 1 步是使用正交覆盖码分离第 1 组用户和第 2 组用户的导频信号, 分离后的信号表示为 $\mathbf{y}^{(1)}$ 和 $\mathbf{y}^{(2)}$:

$$\mathbf{y}^{(1)} = \frac{\mathbf{y}_1 + \mathbf{y}_2}{2} = \sum_{j=1}^{J/2} \text{diag}(\mathbf{p}_j^{(1)}) \mathbf{h}_j^{(1)} + \frac{\mathbf{n}_1 + \mathbf{n}_2}{2} \quad (10)$$

$$\mathbf{y}^{(2)} = \frac{\mathbf{y}_1 - \mathbf{y}_2}{2} = \sum_{j=1}^{J/2} \text{diag}(\mathbf{p}_j^{(2)}) \mathbf{h}_j^{(2)} + \frac{\mathbf{n}_1 - \mathbf{n}_2}{2} \quad (11)$$

第 2 步是利用每组中 $J/2$ 个用户的不同循环移位, 可在每个 RB 上用基于 DFT 的方法逐个 RB 地将 $J/2$ 个用户的信道分离出来。从式(10)和式(11)可以看到, 由于 $\mathbf{n}_1$ 和 $\mathbf{n}_2$ 是不相关的, $\mathbf{y}^{(1)}$ 和 $\mathbf{y}^{(2)}$ 包含的噪声功率是 $\sigma^2/2$, 这会明显提高信道估计的性能。

4.2 线性插值算法

基于 FDM 的 RE-Sparse SCMA 系统中, 每个用户的导频图样是相互正交的。因此, 接收端做信道估计时只需要对导频图样点的接收信号乘以

发送导频序列的共轭即可得到当前位置信道信息。但是, 每个用户只能获得部分导频符号位置上的信道信息, 为了获得一个子帧中完整数据符号的信道信息就需要做信道插值处理, 最简单的方法就是线性差值。

假设在每个子帧中, $\hat{\mathbf{H}}_{p1}$ 、 $\hat{\mathbf{H}}_{p2}$ 分别表示两个导频符号所估计出来的信道矩阵, 则具体插值计算式为:

$$\hat{\mathbf{H}}_d = \hat{\mathbf{H}}_{p1} \cdot \mathbf{e}^T + \mathbf{G} \cdot \mathbf{T}_\Delta^T \quad (12)$$

$$\mathbf{T}_\Delta^T = \mathbf{T}_d^T - T_{p1} \cdot \mathbf{e}^T \quad (13)$$

$$\mathbf{G} = (\hat{\mathbf{H}}_{p2} - \hat{\mathbf{H}}_{p1}) \cdot \frac{1}{T_{p2} - T_{p1}} \quad (14)$$

其中, $\mathbf{e}^T$ 为 1×12 全 1 向量, T_{p1} 、 T_{p2} 分别为第 1 个和第 2 个导频符号的中心采样点时刻, $\mathbf{T}_d^T$ 为所有数据符号的中心采样点时刻, 则 $\mathbf{T}_\Delta^T$ 为每个子帧中数据符号相对于第 1 个导频符号的时延, $\mathbf{G}$ 为每个子载波沿着时间方向的信道梯度。

5 仿真结果与分析

基于 LTE 上行链路, 仿真比较了 RE-Sparse SCMA 和 RB-Sparse SCMA 在理想信道估计和实际信道估计下的性能, 具体 LTE 系统仿真参数见表 1。

表 1 具体 LTE 系统仿真参数

参数	取值
载频/GHz	2
系统带宽/MHz	10
子载波间距/kHz	15
信道模型	TDL-C 300 ns/ETU
移动速度/(km·h ⁻¹)	3
分配 RB 数	4, 12
仿真时长/s	10
信道编码	Turbo, 1/3
调制编码	C84
发送天线配置	1
接收天线配置	2

仿真中 SCMA 使用的是参考文献[18]中提出的 C84 码, 它包含 6 个码本, 每个码本有 8 个码字, 每个码字的维度为 4。本文仿真考虑了在相同的时频资源上复用 6 个用户和 12 个用户的两种场景。用户 1 到用户 6 使用码本 1 到码本 6。当有 12 个用户时, 用户 7 到用户 12 重复使用码本 1 到码本 6。资源带宽分配 4 个或 12 个 RB 时, 对应的传输块大小分别是 120 bit 和 408 bit, 加上 24 个 CRC 比特后进行 Turbo 编码。仿真比较了时延扩展均方根为 300 ns 的 TDL-C^[19]和 ETU^[20]这两种具有不同时延长度的信道模型, 用户的移动速度为 3 km/h, 因而信道变化很缓慢。仿真图中, 横轴 SNR 的定义是用户在每个用于发送数据的子载波上的平均信号功率除以每个子载波上的噪声功率, 且保持用户的信号功率在导频符号和数据符号上是相等的。

假设接收端噪声功率已知, 当复用 6 个用户时, SIC-MPA 解码算法中 $d_{\max} = 3$, 各模块使用迭代解码次数为 $(I_o, I_s, I_T) = (3, 3, 8)$; 当复用 12 个用户时, $d_{\max} = 4$, 迭代解码次数为 $(I_o, I_s, I_T) = (6, 3, 8)$ 。

图 5~图 8 给出了在 12 个 RB 场景下信道估计的均方差 (mean squared error, MSE) 性能。可以看到, 对于 RE-sparse SCMA CDM 导频和 RB-Sparse CDF+FDM 导频来说, 第二种码分方案的性能比第一种码分方案好 3 dB, 这是由于第二种码分方案在两组导频分离之后, 信噪比增加了一倍。唯一的例外是, 图 8 中 12 个用户 ETU 信道场景下, RE-Sparse SCMA CDM 导频的第一种方案由于使用了 12 个 CS, 其性能出现错误平台, 因而远比 RE-Sparse SCMA CDM 导频的第二种方案差。从图 8 中还可以看到, CDM 导频使用 12 个 CS 或 FDM 导频只在每个 RB 中 1 个子载波上发送导频, 都不能满足 ETU 这种频率选择性很强的信道对循环移位的间隔和对导频密度的要求, 所以它们都出现了错误平台。

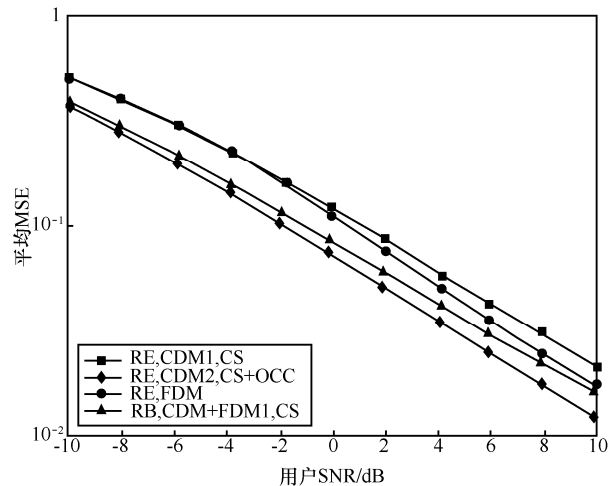


图 5 SCMA 12RB 6 用户 TDL-C 信道估计 MSE 性能

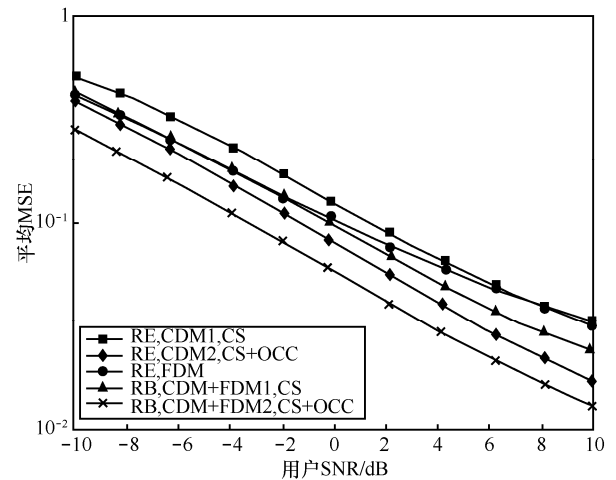


图 6 SCMA 12RB 12 用户 TDL-C 信道估计 MSE 性能

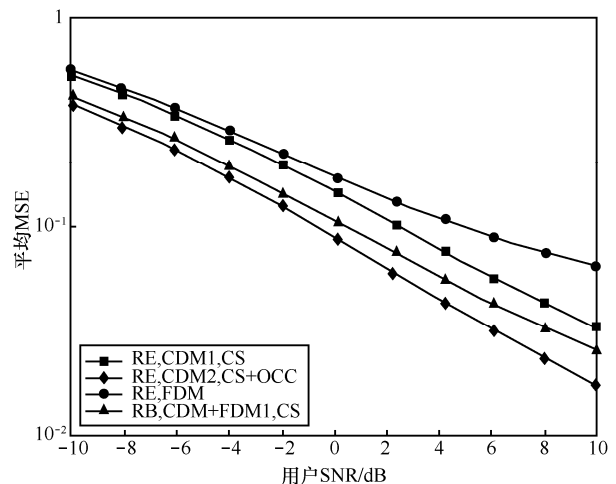


图 7 SCMA 12RB 6 用户 ETU 信道估计 MSE 性能

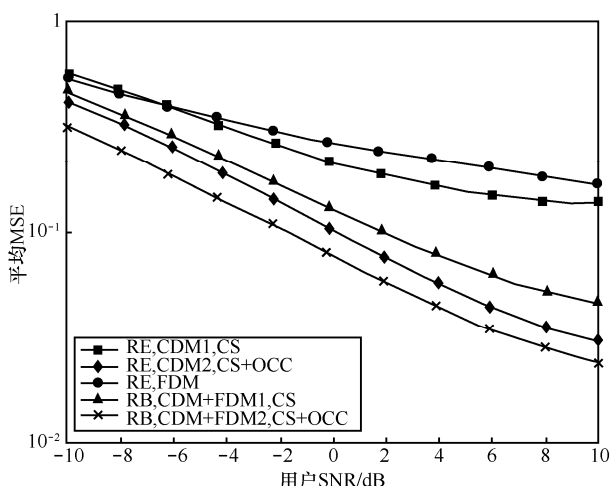


图8 SCMA 12RB 12用户 ETU 信道估计 MSE 性能

在图 5 复用 6 个用户的场景下, RE-Sparse SCMA 使用 FDM 导频比使用 CDM 导频 6 个 CS 的方案好一些。在图 6 复用 12 个用户的场景下, RE-Sparse SCMA 使用 FDM 导频比使用 CDM 导频 12 个 CS 的方案在低信噪比时好, 而在高信噪比时差, 这说明在用户数较多的场景下, FDM 导频将功率集中在少数的子载波上, 所以在低信噪比时表现出较好的抗噪声能力; 但在高信噪比时, 由于它依赖线性插值做信道估计, 从而表现出它的弱势。比较图 5 和图 7 中复用 6 个用户的场景可以看到, 在 TDL-C 信道下 RE-Sparse SCMA 使用 FDM 导频比使用 CDM 导频 6 个 CS 的方案好, 而在 ETU 信道下表现相反, 这说明采用线性插值的 FDM 导频更偏向适用于频率选择性较低的信道。

比较以上所有的导频方案可以看出, 在复用 6 个用户的场景下, RE-Sparse SCMA CDM 导频的第二种方案有最好的信道估计性能; 而在复用 12 个用户的场景下, RB-Sparse SCMA CDM+FDM 导频的第二种方案有最好的性能。如果信道是快衰落信道从而排除使用 OCC, 则 RB-Sparse SCMA CDM+FDM 导频的第一种方案在所有考虑的场景下性能都是最好的。

图 9 给出了复用 6 个用户的 SCMA 在 TDL-C

信道下理想信道估计的 BLER (block error rate, 误块率) 性能。从图 9 中可以看到, RB-Sparse SCMA 比 RE-Sparse SCMA 要差一点, 这是由于 RB-sparse SCMA 使用的子载波相对较集中, 所以它获得的频域分集增益较小; 但是随着分配带宽增加, RE-Sparse SCMA 相对于 RB-Sparse SCMA 的优势也随之减小。

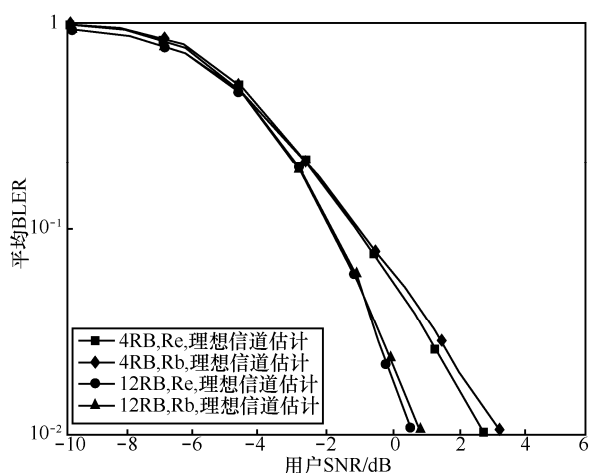


图9 SCMA 12RB 6用户 TDL-C 理想信道估计 BLER 性能

图 10 和图 11 分别给出了在 TDL-C 信道下, SCMA 复用 6 个用户和 12 个用户实际信道估计的 BLER 性能。可以看到, RE-Sparse SCMA CDM 导频的第一种方案与 RE-Sparse FDM 导频的 BLER 性能相近, 尽管它们在图 5 和图 6 中体现出的信道估计性能差异稍大一点。这说明当信道估计性能得到提升时, 不一定能得到 BLER 性能同等程度的提升。在 $BLER=0.1$ 处, 复用 6 个 UE 的场景下, RE-Sparse SCMA CDM 导频使用 6 个 CS 与理想信道估计下的 BLER 性能差距为 2.3 dB, 如果使用 3 个 CS 和 OCC, 则性能差距可缩小为 1.5 dB; RB-Sparse SCMA CDM+FDM 导频使用 3 个 CS 与理想信道估计下的 BLER 性能差距是 1.7 dB。在复用 12 个用户的场景下, RE-Sparse SCMA CDM 导频使用 12 个 CS 与理想信道估计下 BLER 性能的差距是 3 dB, 如果使用 6 个 CS 和 OCC, 则性能差距可缩小为 2 dB; RB-Sparse SCMA CDM+

FDM 导频使用 6 个 CS 与理想信道估计下 BLER 性能的差距是 2.4 dB, 如果使用 3 个 CS 和 OCC, 则性能差距可缩小为 1.6 dB。由此可见, 通过使用具有 FDM 特性的导频和使用 OCC, 都能增大导频序列循环移位的间隔, 从而提高信道估计的准确性和 BLER 性能。

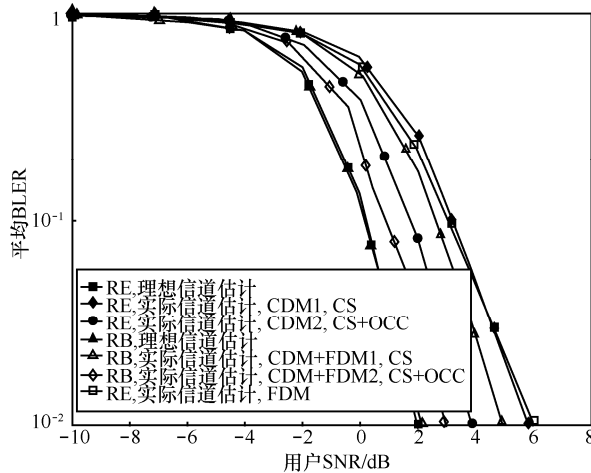


图 10 SCMA 12RB 6 用户 TDL-C 信道 BLER 性能

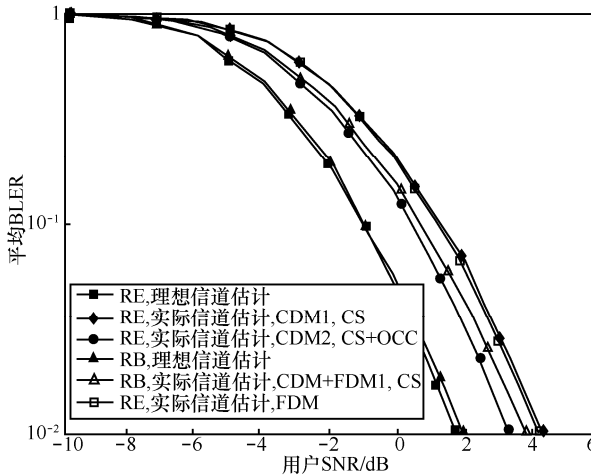


图 11 SCMA 12RB 12 用户 TDL-C 信道 BLER 性能

6 结束语

本文基于 RE-Sparse SCMA 和 RB-Sparse SCMA 这两种资源映射方式, 对适用于 RE-Sparse SCMA 的 CDM 和 FDM 导频图样以及适用于 RB-Sparse SCMA 的导频图样, 在 TDL-C 和 ETU 信道下进行仿真验证和评估, 并且基于信号功率

的穷尽搜索算法提出了适用于 SIC-MPA 的目标用户选择策略。通过对信道估计的 MSE 性能和 BLER 性能的结果分析比较可以看出, 当用户数增加到 12 个时, RE-Sparse SCMA 的信道估计性能有明显下降, 而 FDM 和 OCC 方式都可以增大导频序列循环移位的间隔, 从而提高信道估计的准确性和 BLER 性能。RB-Sparse SCMA 虽然较 RE-Sparse SCMA 在频域分集上的增益要小, 但在实际系统中只需要在所分配的 RB 上进行导频分配, 并且能够额外利用因子图的部分频域正交性, 从而减少所需要的循环移位的个数, 增大循环移位间隔来提高信道估计的性能。从仿真结果中可以看出, 在相同的导频方案下, RB-Sparse SCMA 性能都是最好的, 因此 RB-Sparse SCMA 能更灵活地适用于实际多用户传输场景。

参考文献:

- [1] 倪善金, 赵军辉. 5G 无线通信网络物理层关键技术[J]. 电信科学, 2015, 31(12): 40-45.
NI S J, ZHAO J H. Key technologies in physical layer of 5G wireless communications network[J]. Telecommunications Science, 2015, 31(12): 40-45.
- [2] NIKOPOUR H, BALIGH H. Sparse code multiple access[C]//IEEE Personal Indoor and Mobile Radio Communications (PIMRC), Sept 8-11, 2013, London, UK. New Jersey: IEEE Press, 2013: 332-336.
- [3] WANG B, WANG K, LU Z, et al. Comparison study of non-orthogonal multiple access schemes for 5G[C]//IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), June 17-19, 2015, Ghent, Belgium. New Jersey: IEEE Press, 2015: 1-5.
- [4] MU H, MA Z, ALHAJI M, et al. A fixed low complexity message pass algorithm detector for up-link SCMA system[J]. IEEE Wireless Communications Letters, 2015, 4(6): 585-588.
- [5] DU Y, DONG B, CHEN Z, et al. A fast convergence multiuser detection scheme for uplink SCMA systems[J]. IEEE Wireless Communications Letters, 2016, 5(4): 388-391.
- [6] Huawei. R1-164390 transceiver implementation and complexity analysis for SCMA[R]. 2016.
- [7] BAYESTEH A, YI E, NIKOPOUR H, et al. Blind detection of SCMA for uplink grant-free multiple access[C]//IEEE in Wireless Communications Systems (ISWCS), Aug 26-29, 2014, Barcelona, Spain. New Jersey: IEEE Press, 2014: 853-857.



- [8] WANG Y, ZHOU S, XIAO L, et al. Sparse Bayesian learning based user detection and channel estimation for SCMA uplink systems[C]//International Conference on Wireless Communications & Signal Processing (WCSP), Oct 15-17, 2015, Nanjing, China. New Jersey: IEEE Press, 2015: 1-5.
- [9] STRUMINSKY K, KRUGLIK S, VETROV D, et al. A new approach for sparse Bayesian channel estimation in SCMA uplink systems[C]//International Conference on Wireless Communications & Signal Processing (WCSP), Oct 13-15, 2016, Yangzhou, China. New Jersey: IEEE Press, 2016: 1-5.
- [10] WANG Y, ZHANG X, ZHOU S, et al. User detection and channel estimation for SCMA uplink system in dispersive channel[C]//IEEE International Conference on Communication Systems (ICCS), Dec 14-16, 2016, Shenzhen, China. New Jersey: IEEE Press, 2016: 1-5.
- [11] HEO E, KIM N, PARK H. Sparse structure-based channel estimation for uplink SCMA system[J]. IEEE Transactions on Vehicular Technology, 2017(99): 1-10.
- [12] Fujitsu. R1-166671 DM-RS considerations for UL non-orthogonal multiple access[R]. 2016.
- [13] 3GPP. Evolved universal terrestrial radio access (E-UTRA); physical channels and modulation: TS36.211[S]. 2012.
- [14] ZHAO Y, HUANG A. A novel channel estimation method for OFDM mobile communication systems based on pilot signals and transform-domain processing[C]//IEEE 47th Vehicular Technology Conference Technology, May 4-7, 1997, Phoenix, USA. New Jersey: IEEE Press, 1997, 3: 2089-2093.
- [15] ZHANG S, XU X, LU L, et al. Sparse code multiple access: An energy efficient uplink approach for 5G wireless systems[C]//IEEE Global Communications Conference, Dec 8-12, 2014, Austin, USA. New Jersey: IEEE Press, 2014: 4782-4787.
- [16] WU Y, ZHANG S, CHEN Y. Iterative multiuser receiver in sparse code multiple access systems[C]//IEEE International Conference on Communications, June 8-12, 2015, London, UK. New Jersey: IEEE Press, 2015: 2918-2923.
- [17] XIAO B, XIAO K, ZHANG S, et al. Iterative detection and decoding for SCMA systems with LDPC codes[C]//IEEE International Conference on Wireless Communications & Signal Processing, Oct 15-17, 2015, Nanjing, China. New Jersey: IEEE Press, 2015: 1-5.
- [18] Huawei Technologies Co., Ltd. SCMA pre-standardization whitepaper[R]. 2016.
- [19] 3GPP. Study on channel model for frequency spectrum above 6 GHz: TR38.900[S]. 2016.
- [20] 3GPP. Evolved universal terrestrial radio access (E-UTRA); user equipment (UE) radio transmission and reception: TS36.803[S]. 2016.

[作者简介]



胡显安（1990-），女，上海诺基亚贝尔股份有限公司算法工程师，主要研究方向为 5G 新型多址技术。



曾祥念（1976-），女，博士，上海诺基亚贝尔股份有限公司系统工程师，主要研究方向为 LTE、Turbo 均衡器、SCMA、LDPC。



暂态社区感知的 ICWN 数据转发机制

杜剑¹, 夏元轶², 赵俊峰², 王峥¹, 王鹤¹

(1. 北京智芯微电子科技有限公司, 北京 100192;

2. 国网江苏省电力公司信息通信分公司, 江苏 南京 210024)

摘要: 为了有效解决间断连接无线网络中的数据转发问题, 提出了一种暂态社区感知的数据转发机制, 运用半马尔可夫链模型描述节点在多个地理位置间的转移过程, 预测节点在未来相遇的时间概率分布, 确定节点相遇位置和时间, 为下一跳中继节点的选择提供了理论依据。实验数值表明, 与传统算法相比, 所提机制能有效提高节点相遇预测的准确性, 在数据成功投递率和传输时延等性能上都有较大的提升。

关键词: 间断连接无线网络; 数据转发; 暂态社区; 相遇预测; 半马尔可夫模型

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017262

Transient community aware data forwarding mechanism for intermittent connected wireless network

DU Jian¹, XIA Yuanyi², ZHAO Junfeng², WANG Zheng¹, WANG He¹

1. Beijing Smartchip Microelectronics Technology Co., Ltd., Beijing 100192, China

2. Information & Telecommunication Branch of State Grid Jiangsu Electric Power Company, Nanjing 210024, China

Abstract: In order to solve the problem of data forwarding in intermittent connected wireless network effectively, a transient community aware data forwarding mechanism for intermittent connected wireless network (ICWN) was proposed. Utilizing the semi-Markov chain model, the transfer process of nodes' between multiple geographic locations was described and the time probability distribution of nodes' encountering in the future was predicted, then the encountering time and locations could be obtained, which provided theoretical basis for the selection of next relay node. Experiment results show that the proposed mechanism can effectively improve the forecasting accuracy of nodes' encounter and has a great improvement in the data delivery ratio and transmission delay.

Key words: intermittent connected wireless network, data forwarding, transient community, encounter forecasting, semi-Markov model

收稿日期: 2017-07-20; 修回日期: 2017-08-31

基金项目: 国家电网公司科技基金资助项目 (No.526800150007)

Foundation Item: Science and Technology Project of State Grid (No.526800150007)



1 引言

间断连接无线网络 (intermittently connected wireless network, ICWN) 是指不需要源节点和目的节点在转发数据前建立完整的端到端路径, 利用节点移动带来的相遇机会, 以更加灵活的“存储—携带—转发”的协作方式逐跳进行传输^[1]。该网络具有自组织、节点稀疏、间歇连接、资源受限、拓扑动态变化等特性^[2]。在这种复杂多变的组网环境下, 设计能够满足间断连接无线网络通信需求的数据转发机制成为研究间断连接无线网络的关键问题^[3]。

在间断连接无线网络中, 源节点和目标节点之间并不依靠始终连通的完整链路进行数据的传输^[4], 且源节点与中继节点之间的连接也随着时间的变化而不断变化, 任意成对节点间的通信链路都随时会遭遇断裂^[5], 这种间断性的端到端连接和不断变化的网络拓扑使得传统移动自组织网络的数据转发机制难以直接应用到间断连接无线网络中^[6]。

目前, 国内外研究人员已经提出了一些应用于间断连接无线网络的数据转发机制。Becker 等人^[7]提出了传染路由 (epidemic routing) 机制, 其核心思想是源节点将数据复制、转发给任何与其相遇的节点, 一定程度上提高了数据的传输效率, 但洪泛型的数据转发机制需要占用大量的缓存空间, 消耗大量能量, 而节点自身资源很有限, 导致该机制分组丢失率较高。社会学的主要研究内容是人类在时间、空间内的交互作用, 基于社会理论, Hui 等人^[8]提出了一种名为多标签的数据转发机制, 将网络划分为若干社区, 节点在转发数据时选择通信范围内与自己同社区的节点作为下一跳节点, 但当源节点与目的节点所属社区距离较远时, 数据成功投递率不高。为此, 参考文献^[9]中提出 Bubble-Rap 算法, 在选择中继节点时考虑节点及节点所在社区的活跃度, 根据不同需求选

择下一跳节点。针对间断连接无线网络的分裂性特点及节点移动行为的趋同性, 参考文献^[10]提出 SimBet 机制, 在选择中继节点时综合考虑节点向心度和相似度, 节点向心度描述两节点之间的依赖程度, 相似度表征节点移动行为的趋同性, 根据节点向心度和相似度评估数据转发效用值, 以此选择下一跳节点。上述机制中, 节点转发数据时仅选择未来与目的节点相遇概率高或效用值大的节点作为下一跳节点, 其主要关注的是未来两个节点能否相遇, 而没有考虑节点相遇的时间和地理位置。而实际情况中, 节点不会无休止地缓存数据, 且数据自身存在时效性, 因此设计数据转发机制时既要保证数据的有效传递, 也要考虑数据的时效性。

针对上述问题, 本文提出了一种暂态社区感知的数据转发机制 (transient community aware data forwarding mechanism, TCDFM), 运用半马尔可夫模型, 准确预测节点未来相遇的时间和地理位置, 更加合理地选择下一跳中继节点, 在保障数据成功投递率的同时, 降低数据传输时延, 保证数据的时效性。

2 网络环境

间断连接无线网络中节点的社会属性使节点的移动具有一定规律性^[11], 但移动轨迹并不固定, 网络中节点以一定次序在几个地理位置之间进行访问, 节点在到达目标位置后, 以一定的概率选择在该位置滞留一定时间或者直接离开。本文假设在校园场景下, 地理位置分别为教室、食堂、图书馆等建筑, 节点为持有短距离无线通信设备的学生和老师。节点各自遵循自己的日程表进行移动。其中, 社会属性作为节点的内在属性, 则会在很大程度上影响节点的移动模式, 呈现出某时段在有限移动范围内高度活跃, 在另一时段移动相对迟缓, 将这种特性定义为节点的暂态特性。通过进一步研究发现, 节点的暂态特性主要表现

在时间和地理位置两个方面,节点往往趋向于在特定时间区间的某一地理位置聚集。那么在 ICWN 中,若干节点在某个时间区间内常驻在某个明确的地理范围内,且在此范围内的节点保持紧密的联系,从而可以将这些节点进行聚合形成稳定连通的 ICWN 子网,将其称为暂态社区(temporal community, TC)。

为了构建暂态社区,节点为数据设定一个 TTL 值, TTL 值到期,节点自动将数据删除。具体数据转发过程如下:当一个节点需要转发数据时,首先筛选邻居节点中未来能与目标节点相遇的节点作为候选中继节点;然后,节点执行本文预测算法,预测各候选中继节点与目标节点的相遇时间,并与数据的 TTL 值比对,淘汰在 TTL 到期前无法与目标节点相遇的候选中继节点;最后,计算候选中继节点中与目标节点的相遇概率,选取相遇概率最大的节点作为下一跳节点,如果候选中继节点与目标节点相遇的概率值均小于当前节点,则数据继续缓存在本地节点。图 1 为校园场景下本文数据转发机制示意,节点 A 需要转发数据给节点 E,当前时刻节点 A 与室友节点 B 和节点 C 同处于宿舍内。基于节点历史移动信息,节点 A 预测到在 TTL 到期之前,相比于节点 C,节点 B 更有可能与节点 E 相遇,所以节点 A 将数据转发给节点 B。节点 B 不久后与节点 E 相遇,进而将数据转发给节点 E。

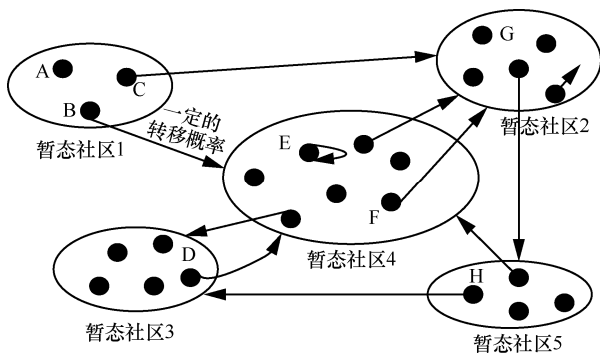


图1 校园场景下本文数据转发机制示意

3 节点行为预测

3.1 节点历史行为信息交互

为了准确预测节点的移动行为,网络中的每个节点都需要尽可能地了解其他节点的历史移动信息。本文将节点的移动信息设计为一个 5 元组 $\langle ID, LocID, P, SP, T \rangle$, 其中, P 是节点在不同地理位置之间的转移概率矩阵, SP 是节点在各个位置的滞留时间概率分布, T 是节点地理位置变化的起始时间, ID 和 $LocID$ 分别是节点 ID 和节点所在位置的 ID。当两个节点相遇时,节点间会交换历史移动信息。节点接收到其他节点的历史移动信息后,通过 5 元组内的 T 值判断数据的新旧,然后将较新的数据在本地缓存。

若节点 S 想要转发数据给节点 D,则首先查询本地数据中节点 D 的历史移动信息 $\langle ID_D, LocID_D, P_D, SP_D, T_D \rangle$ 。则节点 D 在时间 t 时所处的位置由式(1)得到:

$$\alpha_{LocID}^D(t - T_D) = P(W_t = j | W_{T_D} = LocID_D), j \in L \quad (1)$$

其中, W_t 代表节点 D 在时间 t 的地理位置, L 是地理位置集合。

3.2 行为参数评估

预测两个节点的相遇时间和地理位置需要获知两个参数:地理位置转移概率矩阵 P_{ij}^n 和滞留时间概率分布 SP_{ij}^n , 两个参数均由节点历史移动信息获知。

P_{ij}^n 是节点 n 的马尔可夫过程状态转移概率矩阵。图 2 为节点 n 的概率转移矩阵示意,节点 n 有 5 个经常访问的地理位置,分别是教室、宿舍、食堂、实验室和运动馆。节点 n 可以位于 5 个地理位置中的任何一个,并且节点 n 可以选择在当前位置滞留或根据状态转移概率矩阵选择下一个位置进行移动。节点在不同位置之间的转移概率构成节点的转移概率矩阵 P_{ij}^n , 由于每个节点偏好



不同, 每个节点的概率转移矩阵不同。

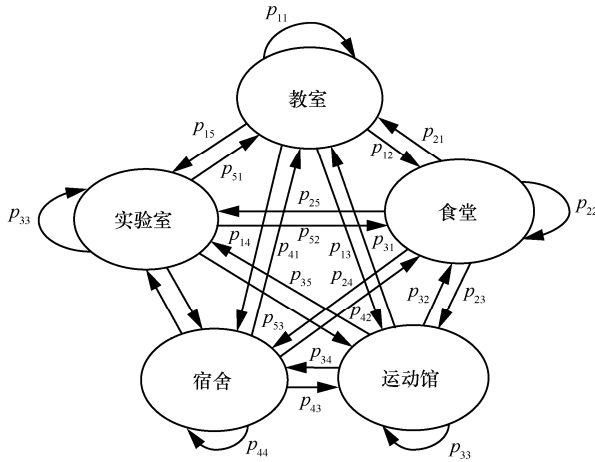


图2 地理位置转移示意

p_{ij}^n 表示节点 n 从地理位置 i 转移到地理位置 j 的概率, 计算方法为:

$$p_{ij}^n = \frac{CT_{ij}^n}{CT_i^n} \quad (2)$$

其中, CT_i^n 表示从节点 n 地理位置 i 转移到其他所有地理位置的次数, CT_{ij}^n 表示节点 n 从地理位置 i 转移到地理位置 j 的次数。显然 $CT_{ij}^n \leq CT_i^n$, 则 $p_{ij}^n \leq 1$ 。通过持续更新 CT_i^n 和 CT_{ij}^n 每个节点都能够生成完善的状态转移概率矩阵。

节点 n 当前位于地理位置 i , 随后转移到地理位置 j , 节点在位置 i 的滞留时间概率分布函数 SP_{ij}^n 可以定义为:

$$SP_{ij}^n(k) = P(t_{ij}^n < k) \quad (3)$$

其中, t_{ij}^n 表示当地理位置 j 为节点 n 的下一个访问位置时, 节点 n 在地理位置 i 的滞留时间。

当网络达到稳定状态时, 移动历史信息为滞留时间分布的计算提供依据。按照式(4)的方式计算 $P(t_{ij}^n < k)$:

$$P(t_{ij}^n < k) = \sum_{m=0}^{k-1} P(t_{ij}^n = m) \quad (4)$$

3.3 节点状态转移概率评估

为了提高间断连接无线网络中暂态社区时空

特征对节点移动行为预测的有效性, 本文使用时间齐次的半马尔可夫链模型模拟节点 n 的移动过程, 该过程可以表示为 (X_i^n, T_i^n) 。节点所处的不同状态用地理位置 ID 表示, 记作 $L=1,2,3,\dots,l$ 。一个节点在两个地理位置之间的移动对应节点在马尔可夫链中两个状态之间的转移。假设不同状态间的转移概率具有马尔可夫链的无记忆性, 即节点 n 从状态 X_i^n 转移到状态 X_{i+1}^n 的概率是独立于状态 X_{i-1}^n 的, 这个转移行为仅与节点当前状态 X_i^n 相关。基于上述特点, 随机过程 (X_i^n) 符合标准的时间离散马尔可夫链的特性。随机变量 T_i^n 表示节点 n 由状态 X_i^n 向 X_{i+1}^n 转移的时间, 随机变量 $(T_{i+1}^n - T_i^n)$ 描述了节点 n 在地理位置 i 的滞留时间。本文节点在某个地理位置的滞留时间并不包含节点在两个状态之间转移所需要耗费的时间。

假设 $P^n = [p_{ij}^n]$ 是节点 n 的关于马尔可夫链的

随机过程 (X_m^n) 的转移概率矩阵。则从状态 i 转移到状态 j 的概率如式(5):

$$p_{ij}^n = \lim_{t \rightarrow \infty} Q_{ij}^n(t), \quad i, j \in L \quad (5)$$

节点 n 在状态 i 的滞留时间的分布为:

$$SP_i^n = P(T_{m+1}^n - T_m^n \leq t | X_m^n = i) = \sum_{j=1}^l Q_{ij}^n(t) \quad (6)$$

4 转发决策

节点 n 的概率分布函数 $\alpha_{ij}^n(k)$ 表示节点 n 在未

来某个时刻 k 由当前地理位置 i 转移到地理位置 j 的概率。假设节点的移动轨迹是相互独立的, 且节点 a 进入当前所在位置 l_a 的起始时间为 k_a , 节点 b 进入当前所在位置 l_b 的起始时间为 k_b , 则节点 a 与节点 b 在时刻 k 相遇于位置 i 的概率如式(7)所示:

$$C_{ab}^i(k) = \alpha_{l_a i}^a(k - k_a) \cdot \alpha_{l_b i}^b(k - k_b), \quad k > 0 \quad (7)$$

节点 a 与节点 b 在时刻 k 在任意地理位置相

遇的概率为:

$$C_{ab} = \sum_{i \in L} C_{ab}^i(k), k > 0 \quad (8)$$

间断连接无线网络虽然是延迟可容忍网络,但是数据具有时效性,超过时效的数据即使转发到目标节点也没有意义,而且在网络中持续转发、缓存时,数据也是对网络资源的浪费。本文认为数据的最大可接受时延就是数据的生存时间 TTL。令节点 n 表示源节点的邻居节点, N 是当前节点的所有邻居节点的集合, d 表示目标节点,选取与目标节点相遇概率最大的节点作为下一跳节点:

$$U_{\text{relay}} = \max\{C_{nd}(k)\}, 0 \leq k \leq T_{\text{TTL}}, n \in N \quad (9)$$

利用上述方法,选择具有最高效用值的节点作为中继节点转发数据,即在未来的 T_{TTL} 单位时间内与目标节点相遇概率最高的节点。如果所选的中继节点未来与目标节点相遇的概率小于当前节点,则数据继续缓存在当前节点,直至更优的中继节点出现,或者待 TTL 值到期后将数据删除。

5 仿真结果分析

本文采用间断连接无线网络环境 ONE^[12] 仿真平台进行数据转发策略性能验证,并与典型的数据转发机制 Prophet 及 Bubble-Rap 进行对比,以验证所提出机制 TCDFM 的性能。仿真参数见表 1。

表 1 仿真参数设置

参数设定	参数数值
网络范围	1 500 m×1 000 m
网络仿真时间/h	24
节点通信方式	蓝牙
无线传输范围/m	10
传输速度/(kbit·s ⁻¹)	250
数据分组间隔时间/s	30
社区数/个	6
节点的缓存/MB	5
节点移动速度/(m·s ⁻¹)	0.5~1.5
停留时间/h	0.5~1

5.1 节点数量对数据转发策略的影响

间断连接无线网络是典型的节点分布稀疏网络,网络中节点的数量直接影响节点相遇机会,最终影响数据转发机制整体性能。因此,首先分析节点数量对不同数据转发机制的影响,结果如图 1 所示。

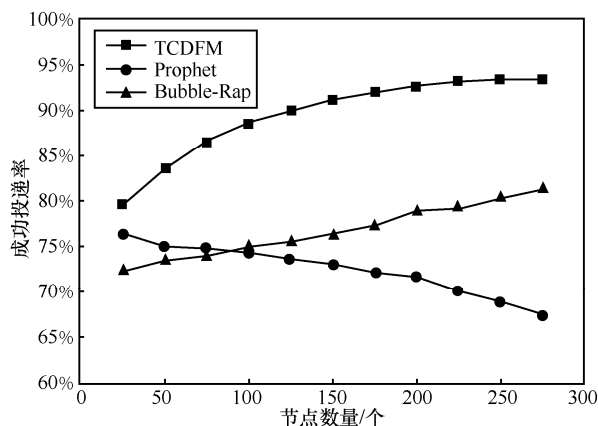


图 3 节点数量对数据成功投递率的影响

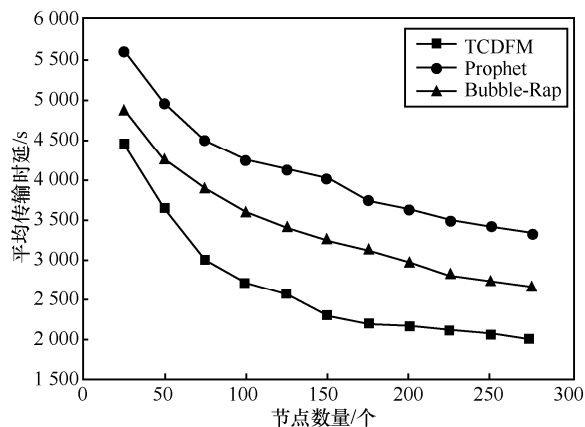


图 4 节点数量对平均传输时延的影响

由图 3、图 4 可以看出,随着网络中节点数量的增加,TCDFM 和 Bubble-Rap 机制在数据成功投递率上呈现平稳上升趋势,相比前两者,Prophet 则呈现反向趋势。TCDFM 与两种算法的数据传输时延均随着节点数量的增加呈现下降趋势。本文算法在数据成功投递率和平均传输时延所呈现出的性能均是最优的,数据成功投递率分别比 Bubble-Rap 算法和 Prophet 算法平均高出 13.2% 和 17.9%,平均传输时延分别比 Bubble-Rap 和



Prophet 平均低 26.8%和 37.3%。主要原因为随着网络中节点数量的增多,节点间相遇机会增加,为中继节点的选择提供了更多的选择,明确了中继节点进行数据投递的目标,进而提高了数据成功投递率。在 Prophet 算法中,随着数据副本在网络中大量分发,网络资源被大量占用,导致基于相遇概率的多副本分发的 Prophet 算法会进行大量的数据复制、转发,抑制了副本的有效传递,降低了网络性能,最终造成了数据传输时延较大。TCDFM 算法考虑了中继节点能否在数据时效性内将数据转发给目标节点,在中继机会增多的同时,能够挑选更优的中继节点进行数据转发,如果数据在有效时间内不能成功投递,则立即删除,避免资源浪费对其他数据的投递产生影响,从而降低了数据的传输时延。

5.2 数据 TTL 值对数据转发策略的影响

TTL 表示网络中数据的生存时长,不同的 TTL 值在一定程度上影响了网络中待转发数据的数量。因此, TTL 值对数据转发策略的性能和网络的存续时间影响很大。图 5、图 6 分别表示在不同 TTL 值下,不同算法的数据的成功投递率和平均传输时延。

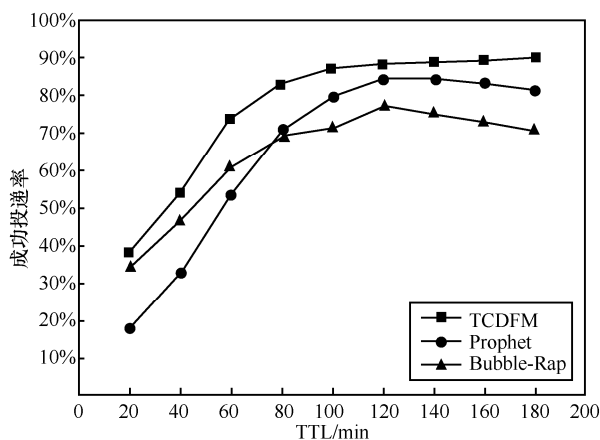


图 5 TTL 值对数据成功投递率的影响

由图 5、图 6 中可以看出,随着 TTL 值增加,3 种路由算法的数据成功投递率均先上升后下降,同时,随着 TTL 值的增加,数据传输的平均时延

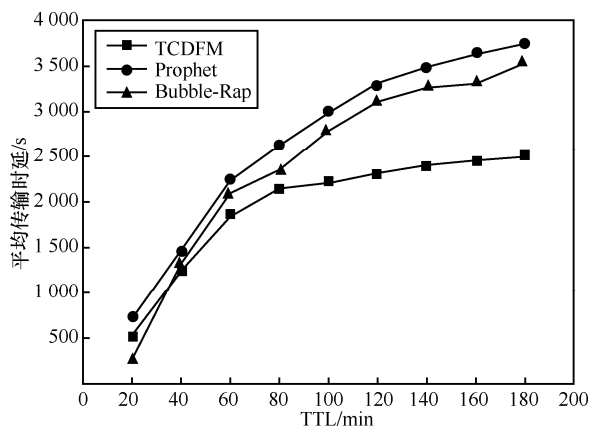


图 6 TTL 值对平均传输时延的影响

随之增加。本文 TCDFM 算法在数据成功投递率上分别比 Prophet 算法和 Bubble-Rap 算法平均高出 12.3%、17.2%,而数据的平均传输时延则比 Prophet 算法和 Bubble-Rap 算法分别低 19.6%、14.1%。主要原因为数据的 TTL 值较小时,节点间还未产生相遇机会,数据就有可能被删除,数据的成功投递率会比较低。而当 TTL 值较大时,网络中的数据会始终保存在节点的缓存空间中,即使数据成功投递到目标节点,网络中仍然包含大量数据副本,极大地浪费网络资源,影响网络的性能。因此,在 TTL 值增大的过程中,数据的成功投递率会随之增高,在达到峰值后,继续增大 TTL 值就会造成网络拥塞,数据的正常传输受阻,数据成功投递率开始下滑,同时,平均传输时延增大。

6 结束语

为了实现间断连接无线网络中数据的高效可靠传输,本文提出了一种暂态社区感知的数据转发机制,利用时间齐次的半马尔可夫链模型进行节点相遇预测,确定两个节点未来在指定时间指定位置的相遇概率。节点需要转发数据时,从邻居节点中筛选在 TTL 值到期之前与目标节点相遇的节点作为候选中继节点,选择各候选中继节点中与目标节点相遇概率最大的节点作为下一跳中继节点。仿真结果表明,与传统路由机制相比,

所提出的机制能够在保证较高数据投递率的同时,降低数据的传输时延,有效地提高网络可靠性。

参考文献:

- [1] WU D P, ZHANG H P, WANG H G, et al. Quality of protection (QoP)-driven data forwarding for intermittently connected wireless networks[J]. IEEE Wireless Communication, 2015, 22(4): 66-73.
- [2] WU D P, HE J, WANG H G, et al. A hierarchical packet forwarding mechanism for energy harvesting wireless sensor networks[J]. IEEE Communication Magazine, 2015, 53(8): 92-98.
- [3] WU D P, WANG Y Y, WANG H G, et al. Dynamic coding control in social intermittent connectivity wireless networks[J]. IEEE Transactions on Vehicular Technology, 2016, 65(9): 7634-7646.
- [4] WANG R Y, YANG H P, WANG H G, et al. Social overlapping community aware neighbor discovery for D2D communications[J]. IEEE Wireless Communications, 2016, 23(4): 28-34.
- [5] FENG M, MAO S, JIANG T. Joint duplex mode selection, channel allocation, and power control for full-duplex cognitive femtocell networks[J]. Digital Communications and Networks, 2015, 1(1): 30-44.
- [6] WU D P, ZHANG P N, WANG H G, et al. Node service ability aware packet forwarding mechanism in intermittently connected wireless networks[J]. IEEE Transaction on Wireless Communications, 2016, 15(12): 8169-8181.
- [7] BISTA B B. Improving energy consumption of epidemic routing in delay tolerant networks[C]//International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, July 6-8, 2016, Fukuoka, Japan. New Jersey: IEEE Press, 2016: 278-283
- [8] SONG L, LI Y, FAN S, et al. Social-based multi-label routing in delay tolerant networks[C]//IEEE Fourth International Conference on Big Data and Cloud Computing, Dec 3-5, 2014, Sydney, Australia. New Jersey: IEEE Press, 2014: 402-407.
- [9] GUPTA A, AGRAWAL A, NAGRATH P. Variant of BUBBLE rap forwarding algorithm for delay tolerant networks[C]//International Conference on Computational Techniques in Information and Communication Technologies, March 11-13, 2016, New Delhi, India. New Jersey: IEEE Press, 2016: 63-68.
- [10] SHRESTHA N, SASSATELLI L. Inter-session network coding-based policies for delay tolerant mobile social networks[J]. IEEE Transactions on Wireless Communications, 2016, 15(11): 7329-7342.
- [11] 潘剑飞, 徐丽丽, 董一鸿. 动态社区演化研究进展[J]. 电信科学, 2017, 33(1): 24-33.

PAN J F, XU L L, DONG Y H. Research progress of dynamic community evolution[J]. Telecommunications Science, 2017, 33(1): 24-33.

- [12] ODA T, ELMAZI D, SPAHO E, et al. A simulation system based on ONE and SUMO simulators: performance evaluation of direct delivery, epidemic and energy aware epidemic DTN protocols[C]//International Conference on Network-Based Information Systems, Sept 2-4, 2015, Taipei, China. New Jersey: IEEE Press, 2015: 418-423.

[作者简介]



杜剑(1987-), 男, 北京智芯微电子科技有限公司中级工程师, 主要研究方向为集成电路与传感芯片在电力中的应用。



夏元轶(1988-), 男, 国网江苏省电力公司信息通信分公司专职, 主要研究方向为电力信息化。



赵俊峰(1974-), 男, 国网江苏省电力公司信息通信分公司主任, 主要研究方向为电力信息化。



王峥(1983-), 男, 博士, 北京智芯微电子科技有限公司部门经理, 主要研究方向为微电子与固体电子学。



王鹤(1982-), 男, 北京智芯微电子科技有限公司中级工程师, 主要研究方向为无线网络通信及无线传感网通信技术应用。



NFV 故障关联及故障自愈方案研究

毛斌宏, 阳志明

(中国电信股份有限公司广州研究院, 广东 广州 510630)

摘 要: 利用 NFV 实现电信网络设备软硬件解耦, 摆脱了对专用硬件设备的依赖, 并能利用 NFV 资源共享、自动部署、弹性伸缩等特性, 但同时也带来了故障点多、故障定位分析困难等问题, 导致 NFV 网络的运营维护难度增加。分析了 NFV 故障处理的流程及故障关联方案, 为 NFV 场景下的故障定位处理提供指引。同时给出了 VNF 故障自愈实现方案, 利用 VNF 特性, 实现网元及业务自愈, 提升自动化运维能力。

关键词: NFV; MANO; 故障关联; 故障自愈

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017261

Research on NFV fault association and fault self-healing

MAO Binhong, YANG Zhiming

Guangzhou Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China

Abstract: NFV implements the hardware and software decoupling of telecommunications network equipment, gets rid of the dependence on dedicated hardware equipment. NFV brings benefits such as resource sharing, automatic deployment, flexibility and other features, but it also brings some problems such as more fault points, difficult fault analysis and so on. It results in NFV network operation and maintenance more difficult. The NFV fault process and fault association were analyzed, and guidance for fault location processing in NFV was provided. At the same time, the VNF fault self-healing implementation scheme was given to realize the network element and service fault self-healing, and improve capability of automatic operation and maintenance.

Key words: NFV, MANO, fault association, fault self-healing

1 引言

SDN (software defined networking, 软件定义网络)/NFV (network function virtualization, 网络功能虚拟化)是当前电信网络技术研究的热点之一, NFV 通过使用 x86 等通用性硬件以及虚拟化技术, 承载通信网络处理功能, 使网络设备功能不再依赖于专用硬件、资源可以灵活共享, 实现

新业务的快速开发和部署, 并基于实际业务需求进行自动部署、弹性伸缩等。参照 ETSI NFV 系统架构, 如图 1 所示, 虚拟化实现了底层物理设备和上层操作系统、应用程序的解耦, 而 MANO (management and orchestration, 管理和编排)则提供一个可管、可控、可运营的服务提供环境, 使得底层基础设施可以便捷地提供给应用。MANO 实现 NS (network service, 网络服务)及

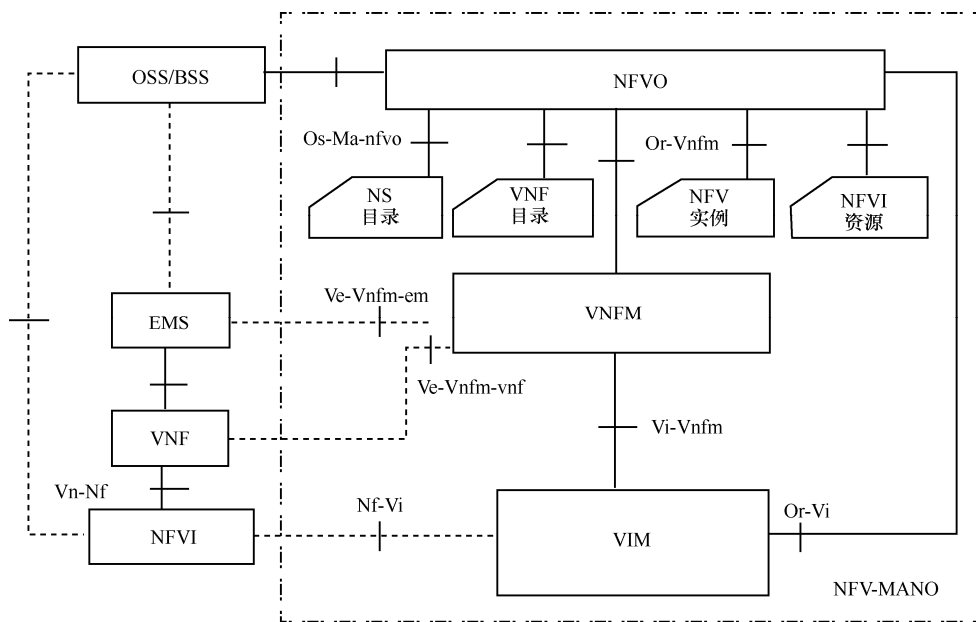


图 1 NFV 系统架构

VNF (virtual network function, 虚拟网元功能) 生命周期管理相关的业务部署、资源调度和运维管理功能。

MANO 包含 3 个功能模块：NFVO (NFV orchestrator, NFV 编排器)、VNFM (virtual network function management, 虚拟网元功能管理) 和 VIM (virtualized infrastructure management, 虚拟化基础设施管理)。

- **NFVO**: 主要负责跨 VIM 的 NFVI 资源编排及 NS 的生命周期管理。
- **VNFM**: 主要负责 VNF 的生命周期管理, 包括 VNF 实例化、VNF 弹性扩缩容、VNF 终止等。VNFM 一般分为通用 VNFM 和专用 VNFM, 其中专用 VNFM 可以部署多个, 一个专用 VNFM 可以管理同一厂商的多个 VNF。
- **VIM**: 主要负责管理一个或多个 NFVI-POP 中的 NFV 基础设施资源。NFVI-POP 是指部署了计算、存储、网络资源, 符合运营商 NFV 硬件和虚拟化技术要求, 由 NFVI 管理系统统一管理, 提供 VNF 业务承载能力的局站。

NFVO、VNFM 和 VIM 这 3 个功能模块在逻辑上独立, 可通过标准接口互通, 在实际部署时可根据需要分设或合设。此外, MANO 还定义了实体模块之间的接口, 包括: NFVO 与 VNFM 的接口 (Or-Vnfm)、NFVO 与 VIM 的接口 (Or-Vi)、VNFM 与 VIM 的接口 (Vi-Vnfm)、NFVO 与 OSS 的接口 (Os-Ma-nfvo)、VNFM 与 EMS 的接口 (Ve-Vnfm-em)、VNFM 与 VNF 的接口 (Ve-Vnfm-vnf)、VIM 与 NFVI 的接口 (Nf-Vi)。

NFV 的引入一方面带来了资源共享、网络快速部署的优势, 但同时也带来了新的问题。随着软硬件解耦, 相对于传统软硬件一体化设备而言, 网元及业务的故障定位、分析、处理的难度会增加。当 NFV 出现故障后, 故障信息可以是几个不同的故障源: 物理资源故障 (即 NFVI 的物理计算、存储和网络相关的故障)、虚拟资源故障 (如虚拟化层或 VM 相关故障) 和应用故障 (即 VNF 应用软件及功能相关故障) 等。NFV 虚拟网元相对于传统软硬件一体化网元设备, 在故障管理方面主要有两大差异: 一是故障点多, 传统网元设备软硬件故障、业务故障都通过同一厂商的设备网管 EMS 采集和上报, 故障定位比较明确; 而



NFV 虚拟网元硬件资源层、虚拟化层、应用软件层、业务逻辑层都可能产生故障。二是故障管理功能分散，传统设备通过网元—EMS—NMS (OSS) 的层次进行处理；而 NFV 网络，在原有处理功能基础上，增加了 MANO 的故障处理功能，分别负责 NFVI 层的硬件资源、虚拟资源故障以及 VNF 故障和 NS 故障。因此，本文将重点研究 NFV 场景下故障处理流程及故障关联方案，同时，由于 NFV 具备动态部署、弹性调整的特性，还要研究 NFV 故障情况下如何实现 VNF 故障自愈、网元及业务的自动恢复，提升自动化运维能力和业务可用性。

2 NFV 故障关联方案

2.1 NFV 故障处理流程

NFV 分层解耦后，每一层都会产生告警，仅靠单一层面很难进行故障定位，例如，一个网卡

出现故障，会引起相关 VIM 层的告警和操作系统层面的告警，也会产生网络层面的告警以及业务层面的告警，这些告警由不同层面采集和监控，如何进行故障关联分析和处理，需要有完善的 NFV 故障采集、故障定位和故障处理流程。

NFV 故障告警主要分为 NFVI 层物理资源告警、虚拟资源告警、虚拟网元告警和业务层告警。NFV 故障管理流程示意如图 2 所示。

NFV 故障管理流程的主要步骤如下。

- VNFM 接收与该特定 VNF 实例相关的虚拟化基础设施故障 (1a)。NFVO 接收与该特定 VNF 无关的基础设施故障 (1b)。
- VNFM 可对某些所选事件执行其自身的故障相关性分析。
- VNFM 可以将与该特定 VNF 实例的相关故障信息转发到其他不同的故障关联点，主要是 EMS (3a) 和 NFVO (3b)。

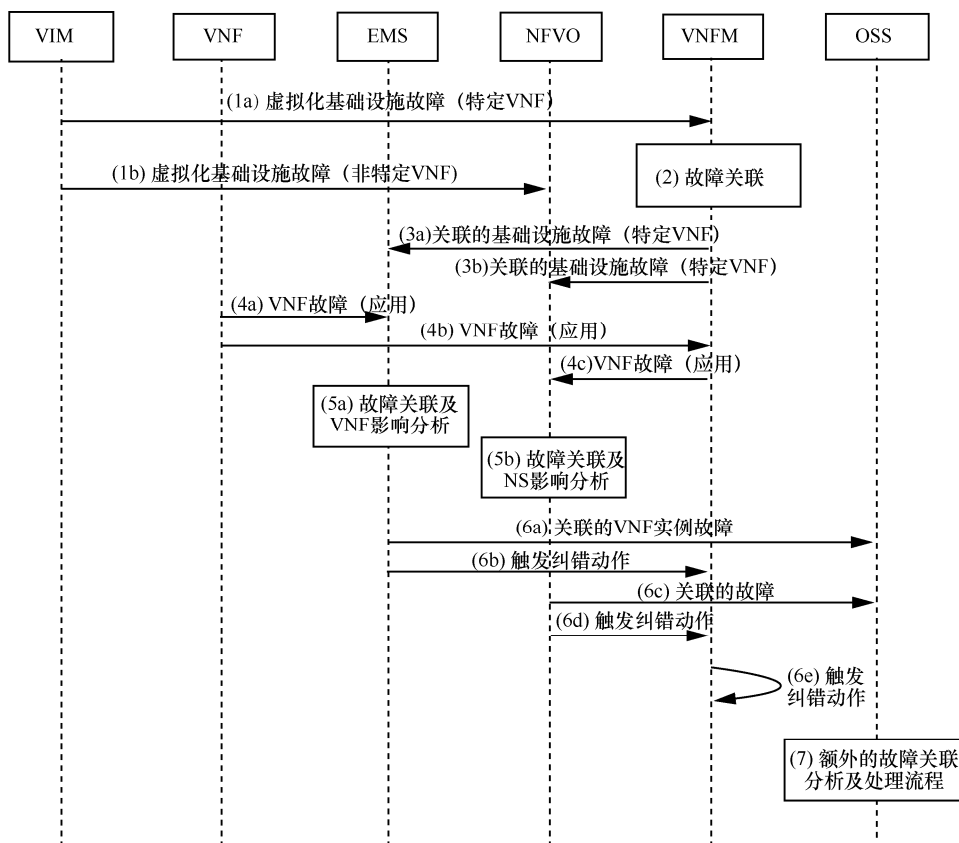


图2 NFV 故障管理流程示意

- VNF 实例可以将应用层故障发送到不同的关联点, 主要是 EMS (4a) 和 VNFM (4b)。VNFM 可以进一步将故障信息转发到 NFVO (4c)。
- 故障相关性分析可能发生在任何故障关联点。EMS 可以执行故障相关性分析以确定根原因和对 VNF 的影响 (5a)。NFVO 可以执行故障相关性分析以确定根原因和对网络服务的影响 (5b)。
- 可以在不同的故障关联点处发起解决故障动作。
- 在 OSS 中可能发生额外的相关性分析和处理流程。

当相同故障原因的告警信息由多个告警源产生时, 需要对这些信息进行关联分析。在 NFV MANO 架构中, 这种相关性分析可能发生在多个地方: EMS、VNFM、NFVO 或 OSS。一旦确定了故障关联点, 其他功能单元必须能够将故障信息转发到故障关联点, 如 VIM 负责收集 NFVI 物理资源和虚拟资源告警并进行转发; EMS 负责收集 VNF 产生的业务告警并进行转发。

2.2 NFV 故障关联分析

NFV 框架中不存在唯一的故障关联点或故障处理点。例如, 某些故障相关性分析和故障处理在 VIM 中执行, 而另外一些故障处理可能在 NFVO、VNFM 或 EMS 中执行, 具体取决于故障本身的类型和层级。NFV 架构中, 每一个管理单元负责不同层级、不同类型的告警采集和故障关联, 如图 3 所示。

NFV 框架中需要根据故障类型确定故障关联点和故障处理点, 每一个管理单元负责不同的告警采集和故障关联。

- VIM 负责收集虚拟资源告警与物理资源告警, 并进行关联, 以确定资源故障影响范围。
- VNFM 负责收集虚拟资源告警与相应的 VNF 网元告警, 并与 VNF 生命周期告警关联, 以确定网元实例影响范围。
- EMS 负责收集 VNF 业务告警与 VNFM 上报的关联后的虚拟资源告警, 并进行关联 (即基础设施与应用关联), 以确定业务影响范围。
- NFVO 负责收集虚拟资源、物理资源告警以及 VNF 生命周期相关告警, 并进行关联,

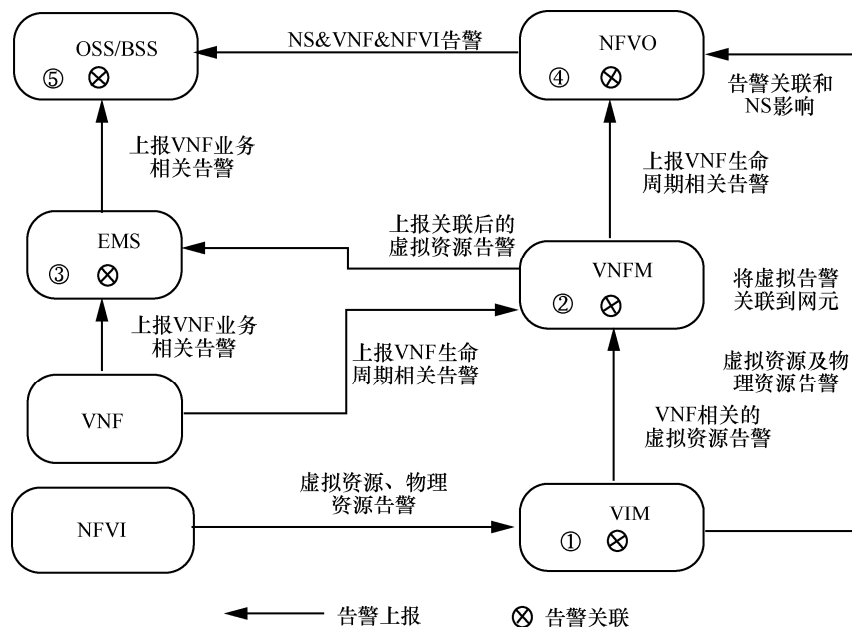


图3 NFV 故障关联示意



以确定 NS 的影响范围。

- OSS 或新一代网络运营系统负责收集 VNF 业务告警、NS 生命周期告警、VNF 生命周期告警、NFVI 重要告警，并进行关联，以确定网络影响范围、业务影响范围及客户影响范围。

基于以上分析，NFV 故障管理中采用“逐层关联、分类管理”的原则，即物理资源、虚拟资源、虚拟网元、网络服务垂直逐层关联，资源故障、业务故障分类管理，这是 NFV 故障管理应采取的基本维护管理思路，需要通过分层协同的方式处理 NFV 运行监控和故障处理。例如，某个虚拟网元的业务单元出现异常，首先进行应用进程重启，而如果是整个虚拟网元出现异常，则可能会触发 VNF 实例生命周期管理流程（例如 VNF 扩缩容、VNF 终止）或 NS 实例生命周期管理流程（例如网络服务终止）的操作以实现故障恢复。

2.3 业务与资源故障关联实现方案

参照 ETSI NFV 架构及相关实体职责分工，MANO（包括 VIM、VNFM、NFVO）主要负责

资源域的告警监控与处理，VNF-EMS-OSS 主要负责业务域的告警采集与处理。而当资源域与业务域告警信息需要关联时，目前业界一般有以下两种实现方案。

方案一：OSS 负责业务与资源故障关联

方案一如图 4 所示，EMS 负责网元层级的业务告警（从 VNF 收集）与资源告警（从 VNFM 收集）的关联。OSS 负责业务层级的业务告警（从 EMS 收集）与资源告警（从 NFVO 收集）的关联。这种方案与现有的电信网络运营维护体制相适应，能够将新的 NFV 网络融入传统 OSS 进行管理，最终实现新旧网络一体化运营维护。

方案二：NFVO 负责业务与资源故障关联

方案二如图 5 所示，EMS 负责网元层级的业务告警（从 VNF 收集）与资源告警（从 VNFM 收集）的关联。NFVO 负责业务层级的业务告警（从 EMS 收集）与资源告警的关联。这种方案需要对 ETSI 定义的 NFVO 功能框架进行功能扩展，统一管理 NFV 网络的业务编排、资源调度及运营维护，不需要与传统 OSS 发生关联，即“新网新

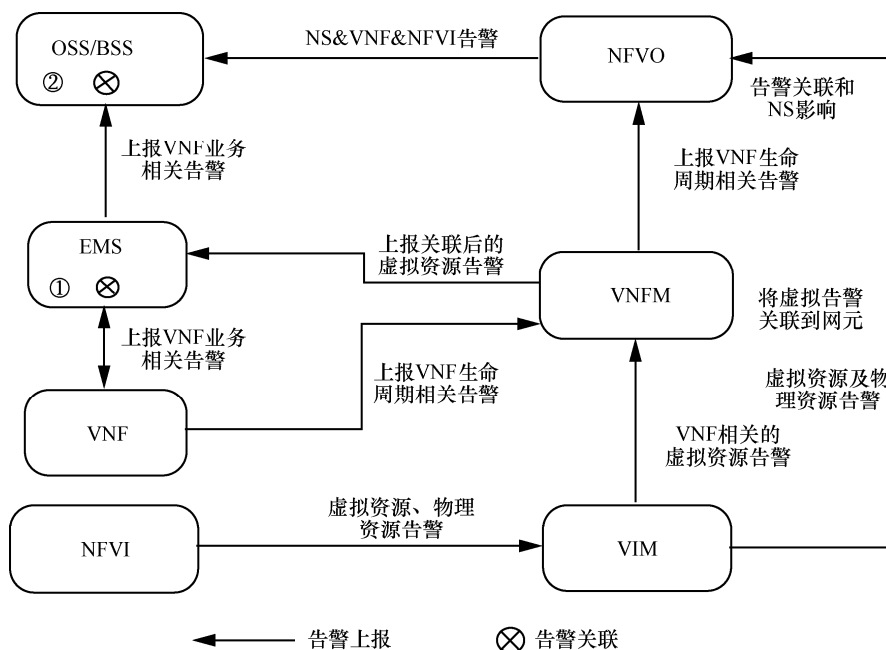


图 4 OSS 负责业务与资源故障关联

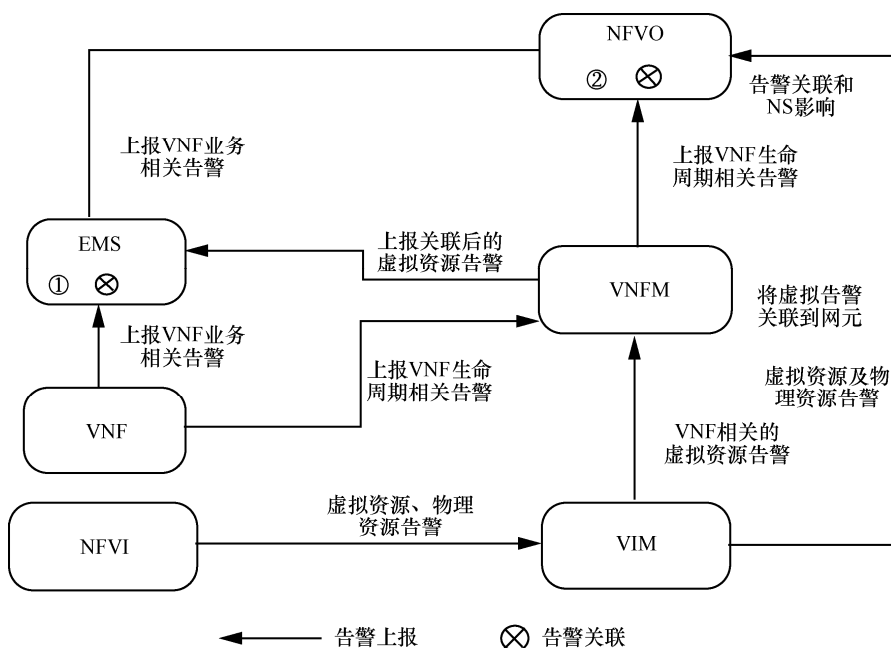


图 5 NFVO 负责业务与资源故障关联

思路”的办法。这种方案简化了故障处理流程，有利于实现故障处理流程的快速启动和基于策略的故障恢复及故障自愈。

从 NFVO 功能要求、故障处理效率、运营模式、EMS 要求等方面对上述两种方案进行对比分析，见表 1。

上述两种方案各有优劣，如果需要兼顾传统网络与 NFV 新网络一体化运营，则建议采用方案一；如果需要提升 NFV 网络告警处理效率，则建议采用方案二。

2.4 分层故障处理方案

故障采集上报、关联分析需要逐层分解、逐层关联，同样，故障处理也需要采用逐层处理的

方案，如图 6 所示。

NFV 架构中各功能实体分别负责相关故障的处理。

- EMS 负责 VNF 业务故障处理。
- EMS 根据策略向 VNF 发起配置变更操作。
- OSS 负责业务、网络故障处理。
- OSS 通知 EMS 对相应 VNF 网元进行调整。
- OSS 通知 NFVO 对相应 NS、VNF 生命周期及其资源进行调整。
- NFVO 负责 NS 故障处理以及与网络相关的物理资源、虚拟资源故障处理。
- NFVO 基于策略调整虚拟资源或物理资源。
- VNFM 负责 VNF 生命周期相关故障处理。

表 1 业务与资源告警故障关联方案对比分析

对比分析	方案一：OSS 负责业务与资源故障关联	方案二：NFVO 负责业务与资源故障关联
NFVO 功能要求	遵循 ETSI MANO 框架所定义的要求，NFVO 只负责资源层的告警处理	NFVO 需要进行功能扩展，增加 VNF 业务告警处理和关联分析
故障处理效率	MANO 需要与 OSS 协同处理，降低告警处理效率	故障处理流程缩短，MANO 内部实行告警闭环，处理效率高
运营模式	新旧网络一体化运营	NFV 网络基于 MANO 统一运营维护，与现有传统网络分离
对 EMS 要求	对厂商网管 EMS 有明确要求	可以不需要 EMS，直接通过 VNF 收集业务告警数据

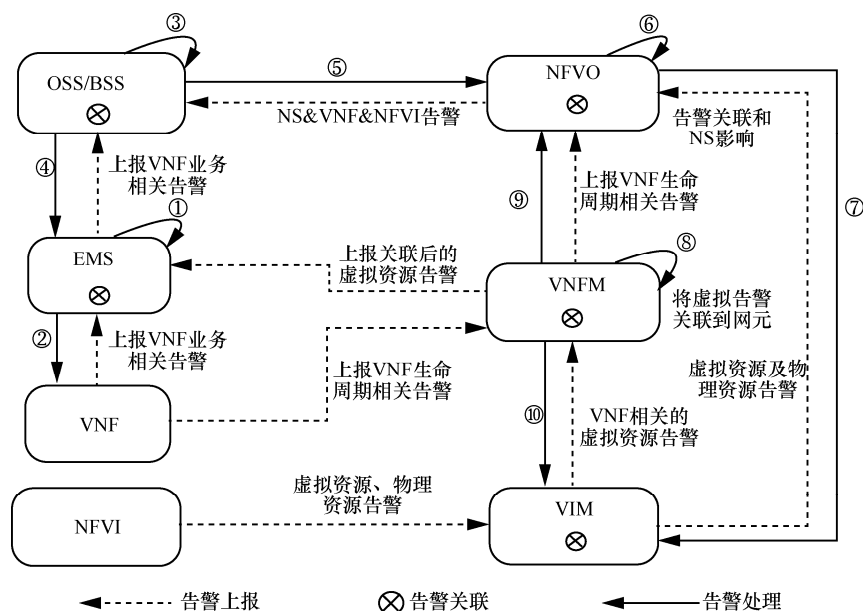


图6 NFV 分层告警处理方案

- VNFM 基于策略需要对虚拟资源进行调整时，通知 NFVO（间接模式）。
- VNFM 基于策略需要对虚拟资源进行调整时，通过 VIM 实现（直接模式）。

因此，NFV 在故障处理时，首先需要分析故障类型（资源故障或业务故障）和故障影响范围（单网元故障或业务级故障），再确定合适的故障处理流程，快速定位故障，并基于策略实现故障修复或故障自愈。

3 故障自愈方案

传统电信网络设备的告警采集监控是自动化处理过程，而故障修复一般需要转入人工线下处理，因此故障监控和故障处理是两个相对独立的处理流程。而 NFV 的引入使得虚拟化网元设备的故障监控与故障处理流程实现了闭环融合，利用虚拟云化资源特性为 NFV 故障自愈创造了条件。

NFV 将软件功能和专用硬件分离，且采用通用的硬件，当硬件出现故障时，应用软件可以自动转移到备份的硬件上重新部署。具体而言，硬

件故障时，该硬件上承载的 VM 转移到备用硬件上，也就是将应用软件重新部署在新的通用硬件上，维持业务的连续性。当 VM 出现故障时，重建 VM 即可。

当 VNF 出现异常时，可以通过 VNF 自愈流程实现自愈，保障 VNF 网元和业务的连续性。

如图 7 所示，基于策略的 VNF 故障自愈有 3 种常见模式。

- VNF 故障自愈：当 VNF 出现异常无法恢复时，则通过 VNFM 发起 VNF 自愈恢复流程，重新实例化部署该 VNF，恢复历史数据，进行 VNF 恢复。
- VNFC 故障自愈：当某个 VNFC 出现异常无法恢复时，则通过 VNFM 删除原有的 VNFC，新建 VNFC 进行恢复。
- VM 故障自愈：当 VNFC 对应的 VM 出现异常时，则通过 VNFM 发起新建 VM，并安装部署相应的 VNFC，实现 VNFC 功能恢复。

以 vIMS 分层解耦部署为例，要求网元及 VNFM 具备故障自动恢复功能，备用模块对应

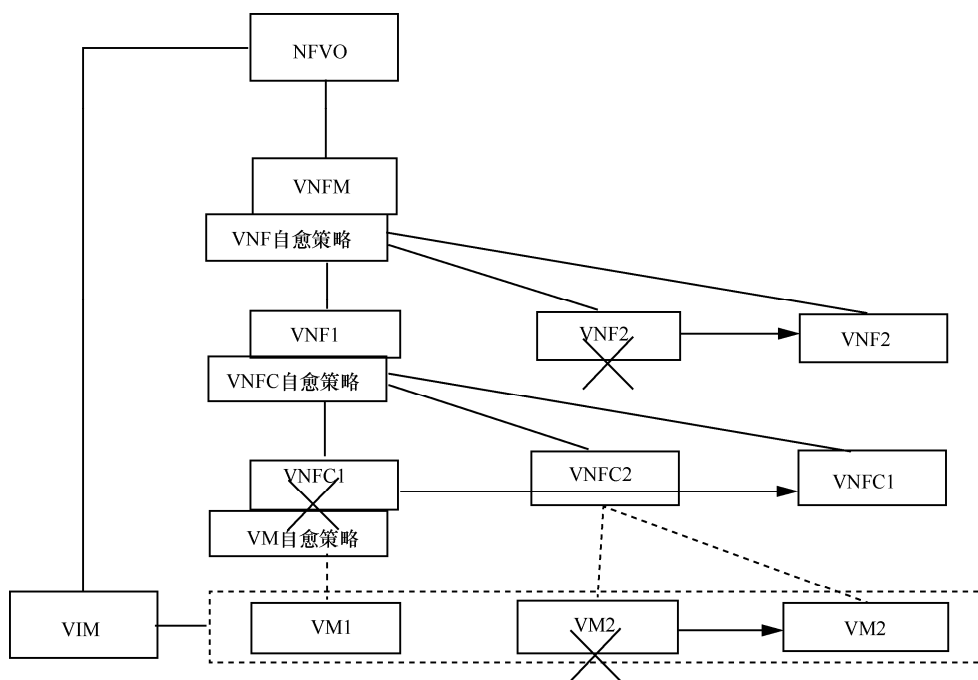


图7 基于策略的 VNF 故障自愈示意

虚拟机恢复后，vIMS 网元应能恢复对应模块的备用工作状态。在设定时间内，备用模块虚拟机由于物理故障等无法恢复的情况下，VNFM 应能启动虚拟机重建将备用模块对应虚拟机迁移至其他物理机上，并启动部署及配置业务功能。在故障自愈过程中，也对各个环节的处理时限提出了要求。

- 从故障产生到被 VIM 检测出来并上报告警的间隔时间小于 10 s。
- 故障硬件对应的模块业务倒换至备用模块，业务损失时间小于 2 s。
- 在设定时间内虚拟机无法自动恢复，VNFM 能启动虚拟机重建流程，虚拟机能够重建成功，重建时间小于 10 min。

尽管虚拟化基础设施管理平台 VIM 一般具备 VM 迁移的能力（如 OpenStack），但 VIM 主动发起的虚拟机迁移无法实现业务网元连接、业务数据恢复，无法保障业务的连续性，因此一般不建议采用 VIM 发起的主动迁移虚拟机，而是通过 VNFM 或 EMS 根据告警及性能策略触发，由业务层发起

业务级的虚拟机迁移，并恢复相应的业务连接和业务配置，保证业务的连续性。

4 结束语

NFV 为通信网络变革带来了机遇，同时也为未来的网络运维提出了新的挑战。本文分析了 NFV 故障处理的流程及故障关联方案，提出了 VNF 故障自愈实现方案，为 NFV 场景下的故障定位处理、故障自愈提供了基本思路和方法。总体来说，NFV 故障处理需要采用“逐层关联、分类管理”的原则，既分工负责，又统筹协同，充分发挥云化部署弹性可扩展的特性，实现虚拟网元故障自愈，提升自动化运维能力。在此基础上，下一步还需深入研究策略引擎、规则引擎等关键技术，制定分层协同故障处理模式下故障快速定位的方法和机制；研究制定故障自愈策略，提高虚拟网元的自愈恢复能力。

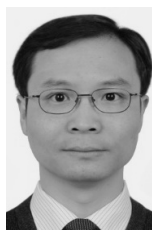
参考文献：

- [1] ETSI. NFV management and orchestration[S]. 2014.

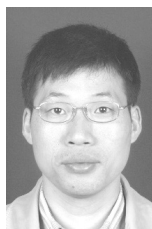


- [2] ETSI. NFV MANO functional requirements specification[S]. 2017.
- [3] ETSI. NFV MANO Or-Vi reference point - interface and information model specification[S]. 2016.
- [4] ETSI. NFV MANO Vi-Vnfm reference point-interface and information model specification[S]. 2016.
- [5] ETSI. NFV MANO Or-Vnfm reference point-interface and information model specification[S]. 2016.
- [6] 赵慧玲, 史凡. SDN/NFV 的发展与挑战[J]. 电信科学, 2014, 30(8): 13-18.
ZHAO H L, SHI F. Development and challenge of SDN/NFV[J]. Telecommunications Science, 2014, 30(8): 13-18.
- [7] 王海宁, 赵慧玲. NFVO 标准和实践[J]. 电信科学, 2017, 33(4): 1-7.
WANG H N, ZHAO H L. NFVO standards and practice[J]. Telecommunications Science, 2017, 33(4): 1-7.

[作者简介]



毛斌宏（1975-），男，中国电信股份有限公司广州研究院高级工程师、高级企业信息管理师，主要研究方向为 NFV MANO 及电信运营支撑系统。



阳志明（1972-），男，中国电信股份有限公司广州研究院高级工程师，主要研究方向为信息与通信技术、运营支撑系统理论与实践。



小基站部署方案规划算法

刘旸

(中国移动通信集团设计院有限公司河北分公司, 河北 石家庄 050021)

摘要: 随着移动网络的飞速发展, 小基站在 LTE 网络建设中的地位越来越重要。结合实际建网经验提出一种最优小基站部署方案规划算法——关联层次分析法。该算法结合层次分析法和灰色关联分析法的思想, 确定各部署方案和各指标参数之间的关系, 并对各方案进行排序, 最终得到最优小基站部署方案。仿真结果表明, 关联层次分析法在投资收益均衡方面优于其他传统算法, 并且建网以后的用户体验也有所提升。

关键词: LTE 网络; 小基站部署; 层次分析法; 灰色关联分析法; 关联层次分析法; 投资收益均衡

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017253

Planning algorithm of small cell deployment scheme

LIU Yang

Hebei Branch of China Mobile Group Design Institute Co., Ltd., Shijiazhuang 050021, China

Abstract: With the rapid development of mobile network, small cell is more and more important in the construction of LTE network. A best small cell deployment scheme named association-AHP was proposed. Combining the analytic hierarchy process with the gray relational analysis method, the relationship between each deployment plan and index parameter was determined, and each plan was ranked. Finally, the best small cell deployment scheme was obtained. Simulation results indicate that association-AHP performs better than traditional algorithms in ROI, and it can improve user satisfaction.

Key words: long term evolution network, small cell deployment, analytic hierarchy process, grey correlation analysis method, association-AHP, return on investment

1 引言

随着移动互联网的发展, 人们对流量的需求呈爆炸式增长。移动网络经过 30 年的发展, 已经发展到 4G 阶段。虽然 4G 系统相比以前的通信系统可以承载更多业务, 但是仍然不能满足人们飞速增长的业务需求^[1]。且随着移动网络的发展, 低频资源越来越少, 后续网络大有向高频发

展的趋势, 但是频段越高, 宏基站的穿透能力就越弱, 且建设和选点难度越大。小基站经过多年的技术演进, 组网结构多样化、部署方便快捷, 通过提升网络密度和深度覆盖范围达到更高的传输率和网络容量, 能更加有效地解决应对快速增长的流量需求。基于此, 本文提出一种可以选择出适合不同场景的最优小基站部署方案规划算法——关联层次分析法。通过层次分析法和灰



色关联度分析法处理实际影响方案规划中的各指标元素,最终规划出最符合场景的小基站部署方案。

2 小基站部署方案

据统计,目前 70%以上的流量发生在室内。而小基站主要用于解决室内弱覆盖和容量压抑的问题,小基站根据其功率大小可以分为微基站、微微基站和毫微微基站等多种类型。目前现网规划应用比较多的有微基站(5~20 W)和新型室分(50~200 mW)。微基站根据其设备形态不同又可以分为一体化微基站和分布式微基站,两种方案都主要用于补盲。以上 3 种部署方案针对不同场景的不同部署环境各有其优势,本文将新型室分、一体化微基站和分布式微基站 3 种小基站部署方案作为选择对象,验证所提算法的准确性和优越性^[2,3]。

2.1 新型室分

分布式设备通常由 BBU、RHUB (remote HUB) 以及微型 pRRU (picoRRU) 组成,信源为 BBU,BBU 与 RHUB 间可由光纤连接,RHUB 通过网线与微型头端 pRRU 相连,最终信号由 pRRU 进行发射。一个 RHUB 支持多个 pRRU,RHUB 通过 CPRI 把多路信号汇聚到 BBU。

与传统室分(DAS)相比,新型室分有部署快、维护方便、性能好和容易进行小区分裂合并等优点。新型室分因为其组网灵活,在各类室分场景都比较适用。例如华为的 LampSite,可以支持 900 MHz、1 800 MHz、2.3 GHz 和 2.6 GHz 共 4 个频段,并支持 4CC(载波聚合),可以灵活进行小区合并和分裂。一台 BBU 最多可以支持 24 个 RHUB 和 192 个 pRRU,其灵活的组网和超强的性能使其适用于各类室内场景^[4]。

2.2 一体化微基站

一体化微基站将射频单元、天线和基带等传统基站的模块集成到一起成为一体化基站。一体

化微基站通过 S1/X2 接口直接与核心网和其他基站进行连接,并且可以支持微波回传,在传输受限或者机房稀缺的场景比较适用。

一体化微基站有选址简单、安装方便快捷、美观易伪装等特点,主要用于补盲,与宏基站和室分相结合打造立体深度覆盖网络。特别是在传输受限的场景下,一体化基站有天然优势。

2.3 分布式微基站

分布式微基站和一体化微基站的区别在于,其分为射频处理与天线合一的微 RRU 和基带处理模块 BBU。通过分离这两个模块,可以很大程度上缩小模块的体积和重量,使微 RRU 更易于伪装,且组网更加灵活^[5]。

两种微基站的应用场景比较类似。但由于分布式微基站可以与宏基站共用基带单元,因此它具有更多宏微协同的功能,譬如宏微小区合并、宏微 CoMP、宏微载波聚合等,这些都是一体化微基站所不具备的功能。

3 关联层次分析算法

3.1 指标参数的选择

基于网络规划部署中的实际经验,初步选择把成本、容量、覆盖、传输和物业协调难度 5 个参数作为网络规划中的指标参数。

其中,成本是备选方案在完成覆盖和容量条件的情况下,所花费总成本;容量是考虑是否满足目标场景的容量需求以及后续的扩容需求;覆盖是考虑是否满足目标场景覆盖需求,是否存在弱覆盖区域;传输是附近机房到小基站传输是否通畅。

3.2 指标参数的处理

所选指标参数中包含较多抽象参数,如物业协调难度、容量和覆盖等。对于这类参数,可以借鉴模糊逻辑的思想。根据指标参数的不同特点确定相应的隶属函数 $\mu(x)$,然后通过去模糊法,将抽象的指标参数化作可度量的量。

具体表达式如下:

$$\bar{x} = \frac{\int x \cdot \mu(x) d_x dx}{\int \mu(x) d_x dx} \quad (1)$$

其中, x 是其指标参数隶属函数中的变量; $\mu(x)$ 为相应指标参数的隶属函数; d_x 为指标参数权重; $\bar{x}$ 是最后得到的可度量的值。隶属函数, 是用于表征模糊集合的数学工具, 根据参数的不同类型可以选择不同的隶属函数, 为了方便运算, 本文选择正态分布作为隶属函数。

经过去模糊化后, 所有的指标参数都化作可度量的值。为了对其进行处理需要把它们进行归一化, 这样所有的指标参数即可处于同一量纲, 方便后期处理。参数可以分为越大越好型(容量等)和越小越好型(成本等), 要分开处理。

假设有 m 个候选小基站部署方案, 影响方案选择的指标参数有 n 个。首先需要建立多目标参数矩阵, 归一化后得到如下矩阵:

$$B = (b_{ij})_{m \times n} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2n} \\ b_{31} & b_{32} & b_{33} & \cdots & b_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & b_{m3} & \cdots & b_{mn} \end{bmatrix} \quad (2)$$

其中, b_{ij} 表示第 i 个部署方案的第 j 个指标参数。

3.3 指标参数的筛选

选择的指标参数之间的关联性过高会增加该类指标参数的权重, 加重主观性对规划结果产生的影响, 从而影响最终的规划结果。本文通过灰色关联分析法去除关联度过高的指标参数, 使各参数之间保持独立性, 确保规划结果的准确性, 具体计算式如下:

$$\gamma_{ij} = \frac{1}{m} \sum_{k=1}^m \frac{\min_i \min_j |b_{ki} - b_{kj}| + \rho \max_i \max_j |b_{ki} - b_{kj}|}{|b_{ki} - b_{kj}| + \rho \max_i \max_j |b_{ki} - b_{kj}|} \quad (3)$$

其中, γ_{ij} 代表第 i 个与第 j 个指标参数的相关性; 可以设定的结果分辨率, 最终影响各参数的关联性, 如无特殊要求, 通常 ρ 取 0.5。

通过计算, 可以得到各指标参数之间的关联矩阵:

$$(\gamma_{ij})_{m \times n} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \cdots & \gamma_{2n} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \cdots & \gamma_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \cdots & \gamma_{nn} \end{bmatrix} \quad (4)$$

把相关性过高的指标参数进行筛选剔除。

3.4 指标参数权重的确定

不同场景在选择小基站部署方案时对关键参数的偏重项是有所差异的。本文将适合小基站部署的场景分为居民区、商业街、人流量密集区(如学校、医院和商场等)三大典型场景, 具体见表 1。

表 1 小基站部署场景分类

场景类型	指标特征
居民区	物业协调难, 对基站较排斥, 需伪装
商业街	对传输要求高, 容量也有一定需求
人流密集区	对容量要求高, 且物业协调有难度

本文借鉴层次分析法思想^[6], 根据不同业务对各类指标参数的不同要求, 分别确定出 3 种典型场景下指标参数的权重。

首先, 根据各场景的指标特征, 建立判断矩阵:

$$C = (c_{ij})_{n \times n} \quad (5)$$

其中, c_{ij} 表示第 i 个和第 j 个指标参数对部署方案规划的重要程度的比值, 且 $c_{ij} = 1/c_{ji}$, c_{ij} 的取值可以参考表 2。

依据判断矩阵的值可以分别计算出各场景类

表 2 c_{ij} 参考值

i 与 j 的重要性比较	相等	较强	强	很强	绝对强	表示相邻判断的中间值
c_{ij}	1	3	5	7	9	2、4、6、8



型下对各指标参数的权重值, 计算过程如下:

$$W_i = \prod_{j=1}^n c_{ij} (i=1, 2, \dots, n) \quad (6)$$

计算 W_i 的 n 次方根, 方法如下:

$$\bar{W}_i = \sqrt[n]{W_i} \quad (7)$$

对 $W = (\bar{W}_1, \bar{W}_2, \dots, \bar{W}_n)$ 做归一化处理, 计算式如下:

$$d_i = \frac{\bar{W}_i}{\sum_{i=1}^n \bar{W}_i} \quad (8)$$

$D = (d_1, d_2, \dots, d_n)$ 即相应的决定部署方案规划的指标参数的权重。

因为上述计算过程包含一定的主观性, 因此为了保证权重的准确性, 需要对该结果进行一致性检验。只有通过一致性检验的结果才可以被应用。

一致性检验是将一致性指标 C_1 与平均随机一致性指标 R_1 之比 C_R 作为当前的两两比较判断矩阵是否满足一致性检验的一个标准, C_R 和 C_1 的计算式分别如下:

$$C_R = \frac{C_1}{R_1} \quad (9)$$

$$C_1 = \frac{\lambda_{\max} - n}{n - 1} \quad (10)$$

式 (9) 和式 (10) 中, $\lambda_{\max}$ 代表相互比较得出的判断矩阵 C 的最大特征值; n 为矩阵 C 的阶数; R_1 代表平均一致性指标, 它的取值可以参考表 3^[7]; C_1 代表一致性指标。

只有当得到的 C_R 不大于 0.1 时, 才能表明前面计算的结果满足一致性标准, 可以作为后续计算参考。本文根据前述方法求出用户在 3 种典型场景下的指标参数权重^[8]。

在得到每个典型场景下的指标参数矩阵和指标参数的权重矩阵后, 通过矩阵点乘即可得到最终的规划判决矩阵如下:

$$E = \begin{bmatrix} e_{11} & e_{12} & e_{13} & \cdots & e_{1n} \\ e_{21} & e_{22} & e_{23} & \cdots & e_{2n} \\ e_{31} & e_{32} & e_{33} & \cdots & e_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e_{m1} & e_{m2} & e_{m3} & \cdots & e_{mn} \end{bmatrix} = \begin{bmatrix} d_1 b_{11} & d_2 b_{12} & d_3 b_{13} & \cdots & d_n b_{1n} \\ d_1 b_{21} & d_2 b_{22} & d_3 b_{23} & \cdots & d_n b_{2n} \\ d_1 b_{31} & d_2 b_{32} & d_3 b_{33} & \cdots & d_n b_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_1 b_{m1} & d_2 b_{m2} & d_3 b_{m3} & \cdots & d_n b_{mn} \end{bmatrix} \quad (11)$$

本文借鉴 TOPSIS 算法思想, 依据 key 判决矩阵对用户在使用该业务的情况下的部署方案进行排序。最后通过 TOPSIS 算法对判决矩阵进行处理。其中, x_i 表示第 i 个参数的理想最大值, y_i 表示第 i 个参数的理想最小值:

$$G = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, i=1, 2, \dots, n \quad (12)$$

规划判决矩阵的每行可以代表一种小基站部署方案, 通过计算矩阵中每个值和其每一列的最大值和最小值的差值 G_+ 和 G_- , 其中最大值和最小值的计算也要根据参数的属性 (越大越好或越小越好) 分开处理, 最终可以得到判决距离矩阵 G :

$$G = \begin{bmatrix} G_+^1 G_-^1 \\ G_+^2 G_-^2 \\ \dots \dots \\ G_+^m G_-^m \end{bmatrix} \quad (13)$$

然后通过式 (14) 得到最终的判决值:

$$P = \frac{G_-}{G_- + G_+} \quad (14)$$

判决值越大, 代表该部署方案越适合作为相应目标区域。并且可以根据该值对候选部署方案进行排序, 最终规划出最佳部署方案。

至此, 本文完成了关联层次分析法的运算过程, 算法流程如图 1 所示。

4 仿真及分析

分场景选择河北需要规划的区域, 用本文所

表 3 R_1 参考值

阶数	1	2	3	4	5	6	7	8	9	10	11
值	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46	1.49	1.54

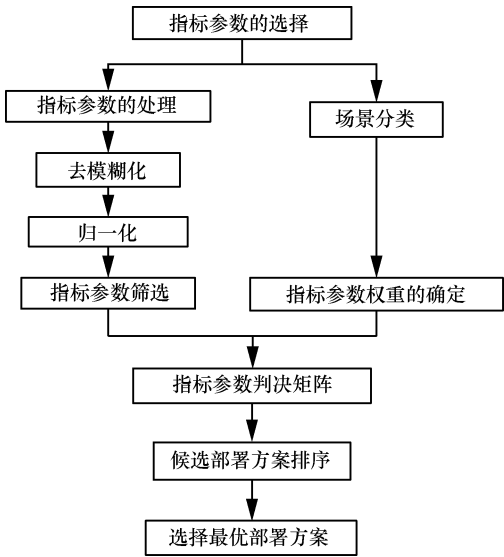


图1 关联层次分析算法流程

提算法进行部署方案选择。分别假设用 3 种小基站方案进行部署，可以得到部署 3 种方案的关键参数表，表 4 是其中一个规划区域的参数示例。

首先对指标参数的原始数据进行去模糊化和归一化处理，可得到表 5 的结果。

根据式 (3) 得到指标参数间关联度矩阵：

关联度矩阵=

$$\begin{bmatrix} 1 & 0.826 & 0.603 & 0.909 & 0.586 \\ 0.826 & 1 & 0.643 & 0.559 & 0.885 \\ 0.603 & 0.643 & 1 & 0.641 & 0.678 \\ 0.909 & 0.559 & 0.641 & 1 & 0.653 \\ 0.568 & 0.885 & 0.678 & 0.653 & 1 \end{bmatrix} \quad (15)$$

从以上关联矩阵可以看出，覆盖和容量两个治疗参数的关联度超过 0.9，这两个指标的相关性

会对规划结果产生影响，降低其准确性。依据标准，结合小基站的特性，把容量指标筛选掉，只留下覆盖指标。最后，可以通过算法得到该区域最适合部署的是新型室分。通过对多个区域的规划，经过归一化处理，得到 3 种场景下各个部署方案的贴近度分别为：

居民区： $P_1 = \{0.21, 0.37, 0.42\}$
商业街： $P_2 = \{0.18, 0.46, 0.36\}$
人流密集区： $P_3 = \{0.58, 0.18, 0.24\}$ (16)

最终可以得出 3 个场景下的平均选择率，如图 2 所示。

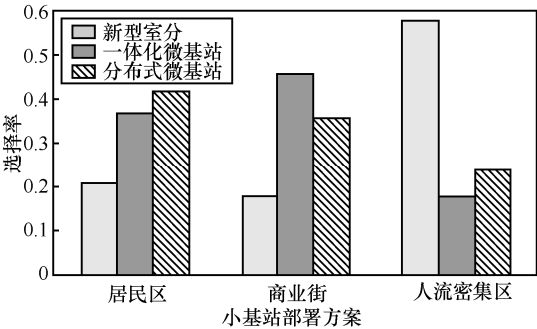


图2 3种典型场景下部署方案的选择

由图 2 可知，居民区、商业街和人流密集区 3 种典型场景的最佳小站部署方案分别是分布式基站、一体化微基站和新型室分。根据河北移动往年小基站部署数据，该方案排序比较准确。

为了体现关联层次分析算法在规划小基站部署方案的优越性。从 2017 年河北需要新建的小基站站点中随机选取 500 个站点进行仿真。并以规

表 4 某规划区域指标参数原始数据

部署方案	覆盖	成本/万元	物业协调难度	容量	传输
新型室分	易	12.5	难	易	较易
一体化微基站	较难	12.1	较难	较难	易
分布式微基站	较难	11.7	较难	较难	易

表 5 某规划区域指标参数处理后数据

部署方案	覆盖系数	成本/万元	物业协调难度系数	容量系数	传输系数
新型室分	0.9	0.34	0.2	0.9	0.75
一体化微基站	0.4	0.33	0.4	0.4	0.9
分布式微基站	0.4	0.32	0.4	0.4	0.9



划站点平均收益均衡值为性能指标,对比本文所提算法和层次分析法(AHP)、传统的以成本为导向的小基站部署方案选择算法,结果如图3所示。

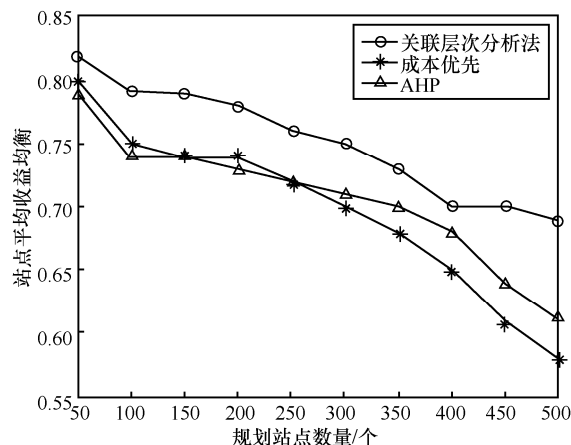


图3 站点平均收益均衡示意

从图3中可以看到,刚开始规划站点数量少,3种方案的收益均衡值比较相近,但是随着规划站点数量的增加,另外两种方法的收益均衡值下降速度明显比关联层次分析法快,而且在最终的平均收益均衡值方面,关联层次分析法也远远领先于另外两种规划算法。说明关联层次分析法在收益均衡方面优于传统的规划算法。另外,相对于传统算法关联层次分析法规划的部署方案可以最大程度降低投诉率,有效提升用户体验。

5 结束语

本文针对现网中小基站的规划部署问题,提出一种新的规划算法。结合层次分析法和灰色关联分析法,通过对影响规划结果的指标参数和其在每种场景下的权重分开处理的方式,得到规划区域各候选部署方案的排名,最终得到规划区域的最优小基站部署方案。仿真结果表明,层次分析法在投资收益均衡方面优于传统算法,并且可以有效提升用户体验。宏微结合进行立体组网已经成为运营商的必然选择^[9],小基站的部署能够有效提升用户体验,但同时带来同频干扰、移动性管理等问题,需要进一步研究解决。

参考文献:

- [1] HOADLEY J, MAVEDDAT P. Enabling small cell deployment with HetNet[J]. IEEE Wireless Communications, 2012, 19(2): 4-5.
- [2] WU G, YANG C, LI S, et al. Recent advances in energy-efficient networks and their application in 5G systems[J]. IEEE Wireless Communications, 2015, 22(2): 145-151.
- [3] 陈晓冬, 吴锦莲, 王庆扬. LTE 异构网络技术研究[J]. 电信科学, 2012, 28(11): 13-18.
CHEN X D, WU J L, WANG Q Y. Study on heterogeneous network technologies[J]. Telecommunications Science, 2012, 28(11): 13-18.
- [4] BAROLLI L, MINO G, IKEBATA T, et al. Performance evaluation and comparison of fuzzy-based intelligent CAC systems for wireless cellular networks[C]//2010 2nd International Conference on Intelligent Networking and Collaborative Systems(INCOS), Nov 24-26, 2010, Thessaloniki, Greece. New Jersey: IEEE Press, 2010: 177-184.
- [5] PORJAZOSKI M, POPOVSKI B. Radio access technology selection algorithm for heterogeneous wireless networks based on service type, user mobility and network load[C]//2011 10th International Conference on Telecommunication in Modern Satellite Cable and Broadcasting Services (TELSIKS), Oct 5-8, 2011, Nis, Serbia. New Jersey: IEEE Press, 2011: 475-478.
- [6] 张玉胜. TD-LTE 无线网络规划技术发展与挑战[J]. 电信工程技术与标准化, 2015, 28(2): 1-6.
ZHANG Y S. Technological development issues and challenges of TD-LTE wireless network planning[J]. Telecom Engineering Technics and Standardization, 2015, 28(2): 1-6.
- [7] 贝斐峰, 陈旭奇, 李新, 等. 小基站在 LTE 异构网络中的应用发展研究[J]. 电信快报, 2014(7): 26-30.
BEI F F, CHEN X Q, LI X, et al. Application development of small cell in LTE heterogeneous network[J]. Telecommunications Information, 2014(7): 26-30.
- [8] 许灵军, 程广辉. 灵活小基站在 TD-LTE 网络中的应用规划[J]. 通信世界, 2012(31): 30.
XU L J, CHENG G H. Flexible small cell in TD-LTE network application planning[J]. Communications World, 2012(31): 30.
- [9] 田峰, 刘翰焘, 周亮. 分层异构网络无线资源管理技术探讨[J]. 电信科学, 2013, 29(6): 32-38.
TIAN F, LIU H T, ZHOU L. Research on radio resource management for hierarchical heterogeneous wireless network[J]. Telecommunications Science, 2013, 29(6): 32-38.

[作者简介]



刘阳(1977-),男,中国移动通信集团设计院有限公司河北分公司咨询设计总监、高级工程师,主要研究方向为移动通信网络规划设计、4G/5G 无线网络技术及演进。