信息通信领域产学研合作特色期刊　十佳皖刊
第三届国家期刊奖百种重点期刊　中国科技核心期刊

ISSN 1009-6868
CN 34-1228/TN

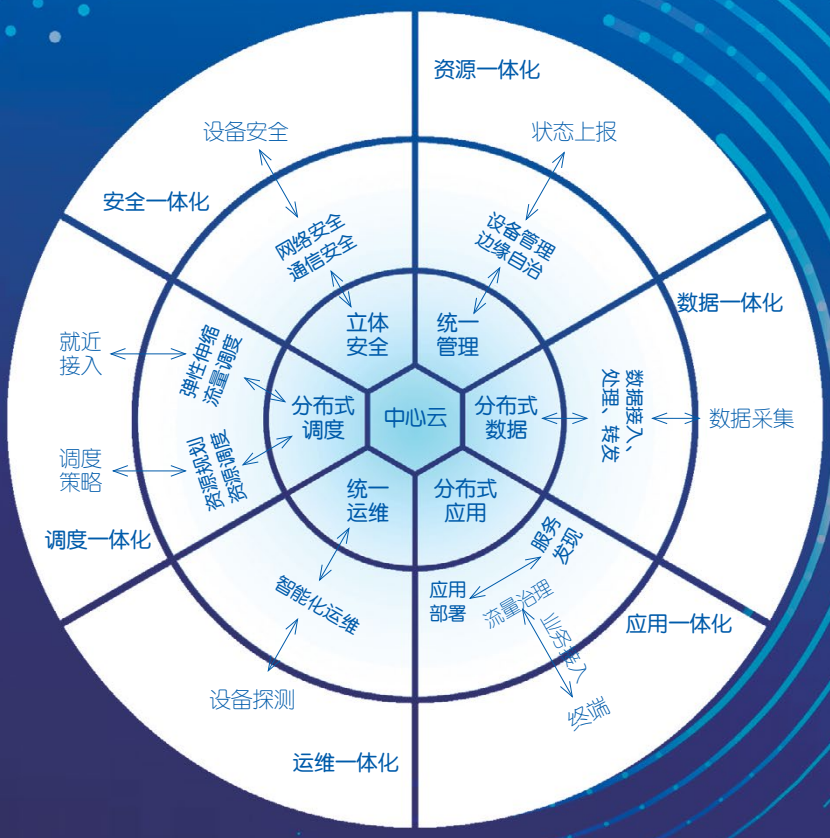
中兴通讯技术

ZTE TECHNOLOGY JOURNAL

http://tech.zte.com.cn

2023 年 2 月 · 第 1 期

专题：面向云网安全的新型防护技术



（封面图片详解见 P14）

ISSN 1009-6868
9 771009 686236 02>

《中兴通讯技术》第 9 届编辑委员会成员名单

目　次

中兴通讯技术（ZHONGXING TONGXUN JISHU）
总第168期　第29卷　第1期　2023年2月

信息通信领域产学研合作特色期刊　第三届国家期刊奖百种重点期刊　中国科技核心期刊　工信部优秀科技期刊　十佳皖刊　中国五大文献数据库收录期刊　1995年创刊

《中兴通讯技术》2023年热点专题名称及策划人

1. 面向云网安全的新型防护技术
中国电信研究院教授级高工　解冲锋
北京邮电大学教授　杨义先

2. 语义通信技术
清华大学教授　陶晓明
中国科学院院士　陆建华

3. 数字孪生技术
重庆邮电大学教授　陈前斌

4. 算力网络和东数西算
工业和信息化部通信科技委
专职常委　赵慧玲

5. 6G网络技术
北京邮电大学教授　王文东

6. 面向双碳的新一代无线通信网络
华中科技大学教授　葛晓虎
西安电子科技大学教授　李建东

MAIN CONTENTS

ZTE TECHNOLOGY JOURNAL
Vol. 29 No. 1 Feb. 2023

期刊基本参数：CN 34-1228/TN*1995*b*16*78*zh*P*¥20.00*6500*13*2023-02

算力网络研究与探索

Research and Exploration of Computing Power Network

张宏科
中国工程院院士、本刊编委

权　伟
北京交通大学教授

刘　康
北京交通大学在读博士研究生

摘要：算力网络研究尚处于起步阶段，在架构、标准以及技术方面尚未达成共识。通过分析国家与社会各行业的算力需求，并结合算力网络发展现状，分别从总体建设目标、理论体系与架构、关键核心技术3个方面为算力网络研究提出相关建议。认为未来算力网络研究与建设要立足于中国算力基础设施现状，着眼于算力与网络的融合发展趋势，突破关键核心技术，建立算力网络服务平台，促进国家数字经济的发展。

关键词：算力网络；算网深度融合；算力网络理论体系

Abstract: Research on the computing power network is still in its infancy with no consensus on architecture, standards and technologies. By analyzing the demand for computing power in the country and all sectors of society and combining with the development status of the computing power network, relevant suggestions on the research of the computing power network are put forward from three aspects: the overall construction goal, theoretical system and architecture, and key core technologies. In the future, the research and the construction of the computing power network should be based on the existing of China´s computing power infrastructure, focus on the development trend of the integration of computing power and networks, make breakthroughs in key core technologies, establish a computing power network service platform, and promote the development of the national digital economy.

Keywords: computing power network; deep integration of computing power and network; theoretical system of computing power network

在国家数字经济发展战略与“十四五”发展规划的推动下，加快信息网络基础的协同化、服务化、智能化进程，深化国家新型基础设施建设（“新基建”），已成为中国进行大国博弈的重要基础。在“新基建”中，5G、大数据中心以及人工智能等相关技术对新一代信息网络提出了新的大算力、大模型处理等算力需求。这推动现有网络从基本的信息数据通信向信息数据智能化处理转变。2021年5月，国家发展和改革委员会、中央网信办、工业和信息化部等四部委联合印发了《全国一体化大数据中心协同创新体系算力枢纽实施方案》[1]，强调要推动中国数据中心网络算网一体化、智能化的发展。“东数西算”工程同样强调构建以算力和网络为核心的体系、优化全国算力整体布局的重要性[2]。在此背景下，算力网络应运而生。算力网络旨在通过泛在算力与网络的融合，突破数据中心、超算中心、云计算、边缘计算等“孤岛”状态下的计算能力限制，构建算网一体的新型智能、高效、按需的算力服务体系，满足国家与行业急需，促进国家数字经济的发展。

1 算力网络的现状与挑战

算力网络作为中国提出且主导的科研技术，已得到业界的广泛认可。诸多产学研团队包括中国科学院、北京交通大

DOI：10.12142/ZTETJ.202301001
网络出版地址：https：//kns.cnki.net/kcms/detail//34.1228.tn.20230224.1457.004.html
网络出版日期：2023-02-27
收稿日期：2022-12-26

学、中国移动、中国联通、中国电信等，已开展算力网络的研究。各单位基于已有的设备、系统、平台以及应用场景，经过长期的积累已取得诸多成果，例如《中国电信云网运营自智白皮书》《中国移动算力网络白皮书》《中国联通算力网络白皮书》《中国通信学会算力网络前言报告》的陆续发布[3–6]，算力感知网络概念的提出等[7]。在国际上，互联网研究工作组（IRTF）设立的在网计算研究组（COIN）致力于算网融合的新型传输架构的研究，互联网工程任务组（IETF）提出分布式方案架构，国际电信联盟（ITU）开展算力网络架构和场景的研究。相较于中国算力网络的蓬勃发展，国际上的研究进展相对缓慢。

目前，算力网络的研究呈现百花齐放的繁荣景象，但相关架构、标准的设计依赖于传统网络技术，尚未形成统一的标准体系。目前，算力网络的研究面临着诸多新需求和新挑战，例如：算力如何一体化？算网如何融合？数据与算力如何满足应用服务需求等。因此，算力网络研究需要明确新需求与新挑战所带来的问题，例如：算力网络中算力主要服务哪些主体？算力如何实现计算？算力依托哪些实体进行计算等。此外，算力网络作为一种新的网络架构，更需要从根本上明确算力网络研究与建设过程中基础理论体系、架构设计、关键核心技术等方面的问题与挑战。

2 算力网络的建设与建议

从算力网络的建设目标与技术发展理念来看，算力网络是通信、计算、存储以及智能化调度的高度融合。算力网络以泛在算力资源为基础，网络通信为纽带，智能化调度为核心，实现网、云、边、端、业务的高效协同与适配，满足行业高差异化算力服务需求。算力网络在实际建设中通常存在两种方向："网中有算"和"算中有网"。"网中有算"是指以网络为中心，算为网用，算力作为基础资源嵌入网中，网络利用算力来提升网络感知、资源调度以及服务功能的编排能力，实现智能高效的网络算力服务。"算中有网"是指以云为中心，网为算用，网络作为连接纽带将离散的数据中心、超算中心等泛在算力进行融合，实现以云为中心的算力资源运营。基于以上分析，面对算力网络的建设需求与挑战，我们从总体建设目标、理论体系架构、关键核心技术3个方面提出研究建议。

2.1 总体建设目标

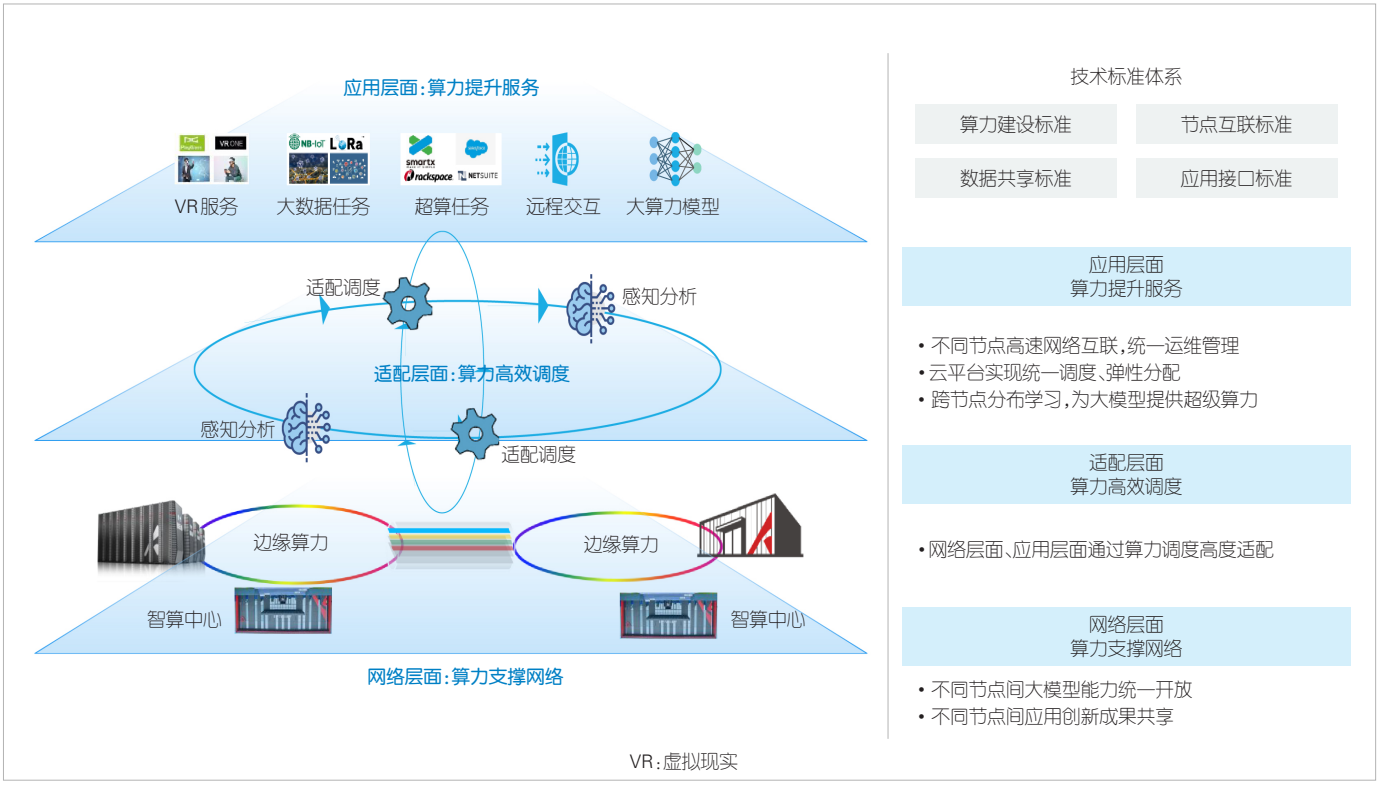
算力网络作为中国率先提出的新型网络架构，相关研究应以技术自主可控、功能性能国际领先为目标，实现智能、高效、灵活的算力资源融合调度，满足行业的差异化算力服务需求，为国家算力网络发展与实施提供支撑。具体来讲，在"算中有网"和"网中有算"两个主要研究方向中，网络是不可或缺的一部分，是算力网络的重要基础支撑与纽带。算力作为一种高效的计算资源，可以提高网络的资源管理、传输调度、路由规划等性能。网络可以连接、协同更多算力资源，提升算力的大数据、大模型处理效率。"网算"与"算网"相辅相成。因此，算力网络的建设应统一融合算力与网络，同时突破算力与传统网络的技术限制，构建"统一调度、弹性适配"的算力网络平台（如图1所示），实现全国范围内算力的高效协同调度与应用，为中国数字经济打下算力基础。

算力网络平台可分为应用层面、适配层面和网络层面。应用层面利用算力来提升服务质量，通过建立应用层面的融合资源池，将超算中心、数据中心等云平台算力进行融合。应用层面的算力服务单元依据资源池进行划分，并实现了统一的调度和弹性分配，满足超算任务、人工智能（AI）任务等分布式与大模型的算力需求。网络层面利用算力来支撑整个网络的融合，强化节点的计算能力以及节点间的主动智能融合与协同能力。适配层面利用算力强化调度方法，实现应用层服务与网络层资源的动态适配调度。此外，算力网络建设需要建立完善的技术标准体系，包括算力建设标准、节点互联标准、数据共享标准、应用结构标准等，为算力网络平台建设与应用提供支撑。

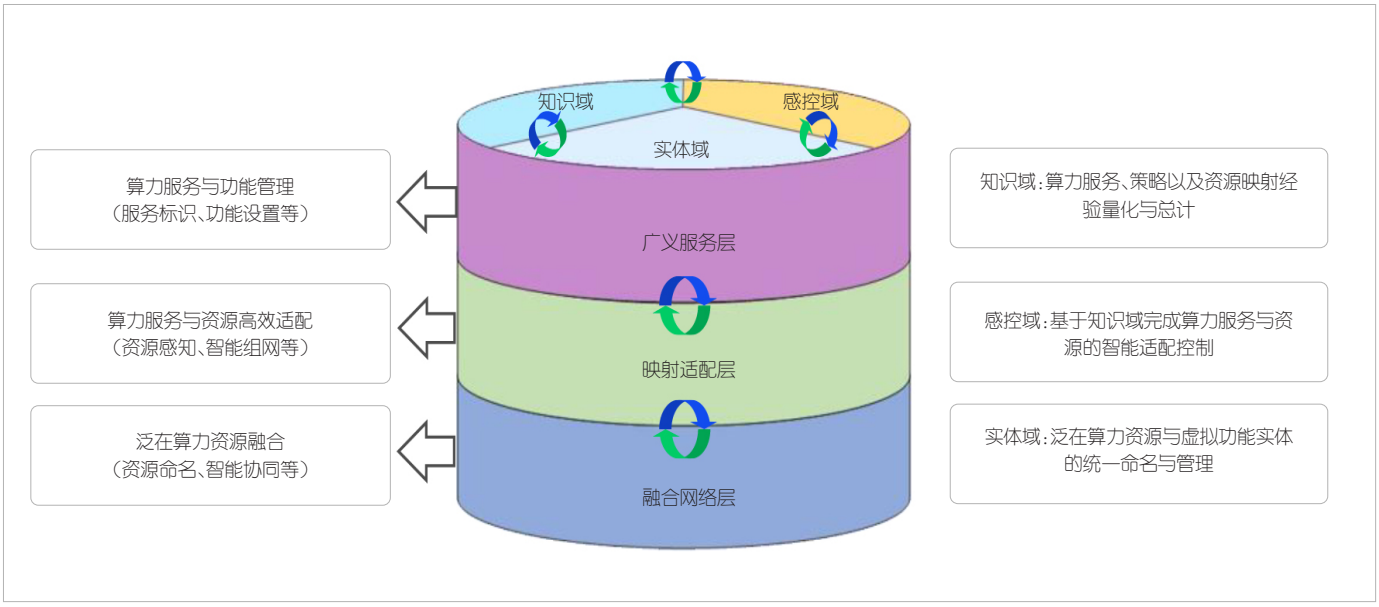
2.2 理论体系与架构

"算中有网"与"网中有算"都表明网络是泛在算力的纽带，是算力网络不可或缺的一部分。然而，当前网络面临架构静态僵化、异构并存、智能受限的状况，行业"高移动、高可靠、高安全、确定性"等差异化服务，成为算力网络建设的新需求与新挑战。此外，新型网络建设正处于谋求网络深度融合、提升网络智慧的新发展趋势中。此趋势与算力网络研究与建设方向不谋而合。因此，算力网络研究不仅要考虑算力，更要关注新型网络，算力与网络不能只是"算中有网"或"网中有算"的分离式协同，而是要实现"算力+网络"的融合突破。目前来看，算力网络研究刚好与新型网络建设相呼应：一方面网络融合可以更好地实现异质异构、分布不均的泛在算力资源的互联；另一方面，算力可以满足大数据、大模型、AI任务等高性能计算需求，实现应用服务、网络以及基础算力之间更高效、更智能的适配调度。算力网络与新型网络研究相辅相成。针对以上需求与挑战，我们提出"三层三域"算力网络架构（如图2所示）。

在"三层三域"算力网络架构中，"三层"包括广义服



▲图1 算力网络平台设计与技术标准体系



▲图2 “三层三域”算力网络架构

务层、映射适配层、融合网络层。广义服务层主要负责服务与功能的标识和描述，具体服务包括：虚拟计算平台、虚拟存储平台、计算容器等虚拟服务以及传输服务功能单元、安全防护服务功能单元等功能服务；映射适配层主要负责服务需求与网络资源的动态适配，通过感知网络状态与服务需求实现服务与算力资源的动态适配；融合网络层主要负责网络与算力资源的协同自组管理，主要包括卫星网络、数据中心网络、超算中心以及泛在算力单元（计算、存储）、通信设备等。此外，架构在“三层”之间还设计了层间解析映射，以强化层间交互性。广义服务层与映射适配层的解析映射，

是将用户的服务需求映射转化为对算力资源的需求。映射适配层与融合网络层的解析映射，是将用户对网络的资源需求映射转化为对实体算力资源的调度，指导算力资源的协同与运行过程。“三层”与层间解析映射的设计既实现了用户与网络的解耦，服务与资源的解耦，又为算力服务与资源的高效适配奠定了基础。

“三域”包括实体域、感控域、知识域。实体域用于格式化描述网络实体组件以及服务功能虚拟实体，实现资源与虚拟服务功能的统一命名；知识域用于服务、策略、网络对象三者的映射经验信息收集与量化，生成拓扑知识库、状态知识库、功能知识库等；感控域对服务功能、执行策略以及网络对象进行数字抽象，以知识域的经验知识为基础，利用算力对服务、执行策略以及网络对象的适配进行动态模拟，生成最优适配策略并指导实体域完成服务。此外，架构在“三域”之间设计域间解析映射，强化各域之间的交互性：知识域与感控域的解析映射是为了将知识域中各类知识库与感控域中的各类策略进行映射连接，便于在感控域策略生成过程中对知识域中的知识进行提取与借鉴，提高感控域策略的准确性；感控域与实体域的解析映射是为了使感控域高效感知实体域资源状态以及属性变化，便于策略调整以及策略下发，实现对实体域资源的精确调度；知识域与实体域的解析映射是为了将知识域中的各类知识库与实体域资源进行对应，根据实体域中资源的属性变化来调整、更新对应知识库。“三域”与域间解析映射的设计既实现了知识、策略、资源的动态解耦，又为用户服务、网络以及泛在算力资源的智能高效处理提供逻辑支撑。

2.3 关键核心技术

算力网络研究与建设要实现“算力+网络”的深入融合目标，建立智能、高效、按需的算力服务平台，从而满足用户高差异化算力服务需求。针对当前网络与泛在算力资源异质异构、分布不均、资源跨网调度困难、智能化程度不足等问题，算力网络研究与建设应从多维标识、智能映射、按需组网、协同传输、智能计算、系统安全6个方面进行关键核心技术突破。

1）多维标识关键技术。算力网络建设集计算、存储、传输资源为一体，关联卫星网络、数据中心、超算中心、云平台等多种网络资源及平台。网络与设备的异质异构，导致算力网络资源调度困难，融合受限。因此，研究需要突破多维标识关键技术，建立算力网络一体化标识体系，实现对泛在算力资源的计算、存储、传输能力以及其他功能属性的统一命名。

2）智能映射关键技术。算力网络是多种平台、网络以及泛在算力资源的深度融合，但融合后的网络资源数量繁多、服务能力差异大，在进行统一的多维标识后需要实现用户服务需求与网络资源的高效动态适配。因此，研究需要突破智能映射关键技术，设计建立完备的解析映射体系，实现用户与网络、服务与资源的智能、高效映射。

3）按需组网关键技术。算力网络建设是为了满足国家与社会产业的发展需求。高铁、工业互联网以及智能制造等行业的发展对网络提出“高移动、高可靠、高安全、确定性”的差异化算力需求。因此，算力网络需要突破按需组网关键技术，根据差异化需求进行网络资源的智能高效编排，将融合后的网络资源进行动态组网调度与管理，满足用户服务需求。

4）协同传输关键技术。算力网络是多种平台、网络以及泛在算力资源的融合，各平台、网络以及设备存在配置差异大、分布不均衡等问题，面对大规模、大模型的计算需求，算力资源需要进行分布式跨平台协作。因此，研究需要突破协同传输关键技术，根据计算服务需求对算力资源的数量、类型、位置以及互联传输设备进行协同传输管理，保障数据在各算力平台、网络以及资源间的高效交互，为算力服务的计算执行提供高效的传输通信支撑。

5）智能计算关键技术。算力网络面对高差异化计算服务需求，不仅需要考虑计算、存储、传输资源的选取问题，还要考虑资源费用、节能等问题。因此，研究需要突破智能计算关键技术，根据服务需求、资源配置、资源费用、节能等进行资源选取、任务分配、路由规划的综合考虑，提升算力网络计算、存储以及传输的智能性，减少服务资源消耗并保障算力服务的高效性。

6）系统安全关键技术。算力网络作为多种平台、网络以及泛在算力资源的融合，多种异质异构网络、资源、平台的互联，使得整体算力网络的安全风险呈指数级增长。因此，研究需要突破系统安全关键技术，在满足算力网络大范围、跨平台、分布式协同计算需求的同时，解决算力网络系统安全防护问题，实现服务与安全的双重保障。

3 结束语

算力网络作为中国率先提出的新型网络架构，是推动信息产业发展、支撑 “十四五”发展规划中“网络强国、数字中国”发展战略的重要基础。当前算力网络领域的研究呈现出繁荣的景象，但在架构、标准设计等方面尚未达成共识。未来算力网络的研究与建设要立足中国算力基础设施现状，着眼于算力与网络的融合发展趋势，研究探索算力网络

基础理论体系，突破关键核心技术，建立算力网络服务平台，满足国家与行业急需，促进国家数字经济的发展。

参考文献

[1] 中华人民共和国国家发展和改革委员会. 关于印发《全国一体化大数据中心协同创新体系算力枢纽实施方案》的通知 [EB/OL]. (2021-05-24)[2022-01-09]. http://www.gov.cn/zhengce/zhengceku/2021-05/26/content_5612405.htm

[2] 中华人民共和国国家发展和改革委员会. 东数西算 [EB/OL]. [2022-01-09]. https://www.ndrc.gov.cn/xwdt/ztzl/dsxs/?code=&state=123

[3] 中国电信. 中国电信云网运营自智白皮书 [R/OL]. [2022-01-12]. https://www.sgpjbg.com/luodi/194069.html? tg=1&bd_vid=9765405328762200487&plan=%E6%B3%A8-13&sdclkid=A52D15fGxSDNArDDx

[4] 中国移动. 中国移动算力网络白皮书 [R/OL]. [2022-01-12]. https://www.waitang.com/report/

[5] 中国联通. 中国联通算力网络白皮书 [R/OL]. [2022-01-12]. https://www.sgpjbg.com/luodi/224295.html? tg=1&bd_vid=12217316302776406816&plan=O-17&sdclkid=A52D15fGb6D6bLApA5-6

[6] China Institute of Communications. Computing network frontier report [R/OL]. [2022-01-12]. https://www.vzkoo.com/document/a9255c0ed9e4db735d93e42b1f3ab839.html? keyword= % E7%AE% 97%E5%8A% 9B% E5%89%8D%E6%B2%BF

[7] China Mobile, Hua Wei. Computing-aware networking white paper [R/OL]. [2022-01-12]. https://wenku.baidu.com/view/9db54352a717866fb84ae45c3b3567ec102ddcee.html

作　者　简　介

张宏科，中国工程院院士，现任北京交通大学电子信息工程学院教授、博导，移动专用网络国家工程研究中心主任，IEEE Fellow，曾任两期国家“973”计划首席科学家，享受国务院政府特殊津贴，首批全国高校黄大年式教师团队带头人；长期从事专用通信网络理论与工程技术研究，建立了标识网络功能结构及解析映射机制，有效解决了复杂场景下网络高移动支持和高可靠传输难题；获得国家技术发明二等奖2项、省部级一等奖4项；出版专著6部。

权伟，北京交通大学电子信息工程学院教授、博导，移动专用网络国家工程研究中心新型网络系统研究所副所长，中国通信标准化协会TC1 WG4工作组副组长；主要研究方向为新型网络体系架构、高可靠网络传输关键技术；发表论文80余篇。

刘康，北京交通大学电子信息工程学院在读博士研究生；主要研究方向为新型网络体系架构、可靠协同传输关键技术、AI与网络传输融合关键技术；擅长AI智能算法的模型设计、理论分析以及AI融合网络系统的研究工作。

面向云网安全的新型防护技术专题导读

专题策划人

解冲锋，中国电信集团高级技术专家，博士，教授级高工，欧洲ETSI IPE工作组副主席、中国互联网协会学术委员会副主任委员、北京市IPv6重点实验室主任；曾在美国加利福尼亚大学洛杉矶分校做政府公派访问学者一年；长期从事网络架构、IPv6下一代互联网、物联网、网络安全、云网融合等方面的研究；在IETF合作发布RFC 5项，拥有授权发明专利50余项。

杨义先，北京邮电大学教授、“长江学者”特聘教授、国家杰出青年基金获得者、国家教学名师、国家教学团队带头人、全国百篇优秀博士学位论文指导教师、国家精品课程负责人；长期从事网络与信息安全方面的科研和教学工作；创立了网络空间安全的统一理论并编写了《安全通论》《博弈系统论》《黑客心理学》等书籍。

近年来，云网融合逐渐打破云和网相对独立和隔离的局面，融合人工智能（AI）、算力、大数据、安全、绿色等多种要素，从而为各行各业数字化转型提供强大的基础设施支撑。云网融合在基础架构、底层设施和资源调度等方面使得云和网趋于一体化，成为中国信息基础设施的核心特征。与此同时，随着国际形势的变化和安全攻防技术的演进，信息基础设施面临的安全形势也日趋复杂和严峻，云网攻击呈现出自动化强、复杂性高、隐匿性深、破坏性大等新的阶段性特性。数据泄露、勒索软件、高级持续性威胁（APT）攻击、路由劫持等安全事件频发，企业用户对安全防护的多样化需求正在与日俱增。信息基础设施快速发展与安全防御能力不足的矛盾日益突出。如何利用新技术、新手段应对云网面临的安全威胁，是业界非常关注的问题。

本专题中的多篇论文论述了云网场景下的安全风险挑战，并对各种新型防护技术展开讨论。《面向云网融合的网络安全互操作》分析了云网时代的主要安全挑战，认为网络安全互操作是云网融合安全发展的重要路径，从标准规范、能力验证、行业示范等方面指出云网安全互操作的重点发展方向；针对5G网络中云、终端、协同机制中的安全缺陷威胁，《基于超级SIM的5G端云安全体系架构与关键技术》论述了“云、端、卡”协同运作的完整安全体系，对端云安全体系关键技术的发展提出新思路；《云网融合下的安全能力池关键技术与应用》提出了云网安全一体化的安全能力池技术方案，实现了网络安全防护能力的快速扩展、灵活定制和可编排，从而满足用户对安全防护能力多层次可定制的需求；《未来网络内生安全通信技术》基于网络可信身份的轻量化密钥验证机制，提出了网络可信通信技术，该技术具备近源协同防护、无状态随路验证等特征，为未来网络安全可信保障提供参考；《云平台DNS安全体系研究》在介绍域名系统（DNS）技术和业态演进过程的基础上，梳理了云平台DNS的安全风险和特征，提出了云平台DNS安全体系框架，并介绍了所在企业的安全实践；《构建可扩展的RPKI依赖方系统部署机制》梳理了影响互联网码号资源公钥基础设施（RPKI）依赖方系统运行效能的4对矛盾，探讨了RPKI依赖方系统部署机制以及对应的运行机制；《大型企业SASE解决方案及应用实践》在分析安全访问服务边缘（SASE）架构基础上，设计了企业一体化安全运营系统Q-SASE及其技术方案，分享了其为大型企业客户提升安全防护方面的实践。《关于发展中国安全浏览器的建议》通过对全球安全浏览器发展趋势的洞察和重要性分析，论述了中国在安全浏览器领域所面临的机会与挑战，提出中国发展安全浏览器的对策建议。

本专题的作者来自各知名高校、企业与科研机构，文章聚焦于云网安全面临的新挑战及当前主要的防护技术。作者们从行业需求分析、系统设计、理论分析、性能评估、实际运营等方面，介绍了云网安全最新的研究成果和经验。希望本期的内容能为读者提供有益的借鉴与启示，并在此对所有作者的大力支持表示由衷的感谢！

解冲锋　杨义先

2023年1月18日

DOI：10.12142/ZTETJ.202301002
收稿日期：2023-01-19

面向云网融合的网络安全互操作

Network Security Interoperability Towards Cloud-Network Convergence

魏亮/WEI Liang，查选/ZHA Xuan，戴方芳/DAI Fangfang

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

DOI：10.12142/ZTETJ.202301003
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230222.1751.008.html
网络出版日期：2023-02-23
收稿日期：2022-12-18

摘要：云网融合通过网络与云的主动适配、协同和融合，为数字化应用提供灵活部署的云网资源。分析了云网时代的主要安全挑战，认为网络安全互操作是云网融合安全发展的重要路径。基于网络安全互操作发展现状分析，从标准规范、能力验证、行业示范3方面指出了云网安全互操作的重点发展方向。

关键词：云网融合；网络安全；安全互操作

Abstract: Cloud-network convergence provides flexible deployment of cloud network resources for digital applications through active adaptation, collaboration, and integration of information network and cloud environment. The main security challenges faced by cloud-network convergence are analyzed, and the security interoperability is proposed as an important way for the cloud-network convergence security development. With the development status analysis of security interoperability, the future development directions of security interoperability are pointed out from the aspects of standards, capability verification and industry demonstration.

Keywords: cloud-network convergence; network security; security interoperability

在数字时代，云网融合的智能化数字基础设施建设，打通了经济社会发展的信息“大动脉”，为数字产业化和产业数字化发展注入新动能。全方位、全链条的数字化升级加速传统行业转型。围绕新生产要素的攻击风险日益突出，传统网络安全威胁与新型网络安全威胁相互交织，保障融合领域安全的能力需求不断提高。网络安全已成为全球各国面临的共同挑战。

自党的十八大以来，以习近平同志为核心的党中央高度重视网络安全，从发展中国特色社会主义、实现中华民族伟大复兴中国梦的战略高度，统筹发展和安全两个大局，系统部署和全面推进网络安全工作。习近平总书记在《关于〈中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议〉的说明》中指出，“安全是发展的前提，发展是安全的保障”。习近平总书记还指出，“过去分散独立的网络变得高度关联、相互依赖，网络安全的威胁来源和攻击手段不断变化，那种依靠几个安全设备和安全软件就想永保安全的想法已不合时宜，需要树立动态、综合的防护理念”。在数字化发展新阶段下，要深刻落实习近平总书记讲话精神，通过打造网络安全互操作新体系，深度融合各领域安全能力，打破割裂的、静态的、封闭的安全，打造一体的、动态的、开放的网络安全屏障，推动网络安全综合保障能力再上新台阶，为加快建设网络强国提供有力支撑。

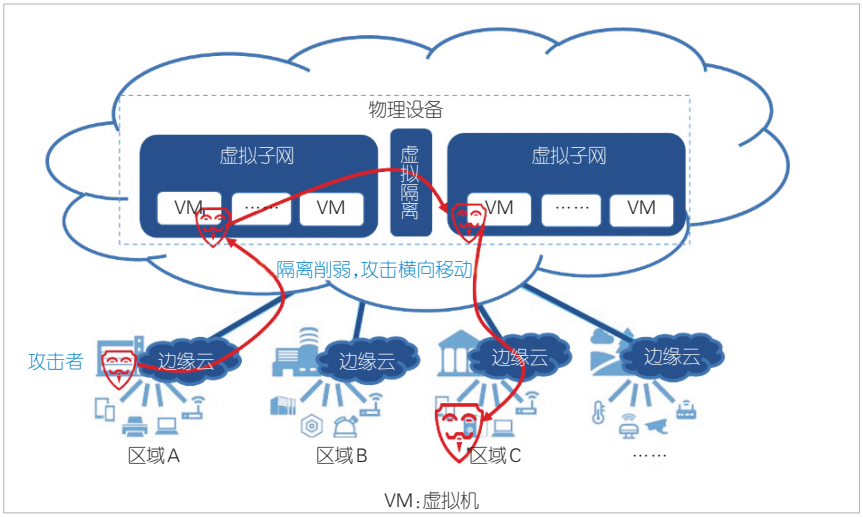
1 云网时代的安全新挑战

通过虚拟化、软件化、云化和人工智能（AI）等信息化技术，云网与通信技术深度融合。这将深刻变革信息基础设施的技术架构、业务形态和运营模式，有助于实现云网边端的智能互联、统一调度和智能化运维，打造新型信息基础设施底座[1-2]。与此同时，云网融合技术变革、融合应用、开放式生态新特性也引入了新的安全挑战[3]。

1）从技术变革看，虚拟化的架构给传统基于边界的防护带来挑战。

云网融合采用统一的虚拟化技术架构，将网络功能从硬件设备中分离，使网络架构从传统固态封闭向动态开放改变。网络的边界感和隔离感随之削弱。一方面，网络布局从传统的接入、汇聚、核心3层网络架构转向围绕数据中心的扁平部署，难以在分层架构的边界提供传统的物理安全隔离；另一方面，存储虚拟化、计算虚拟化、网络虚拟化打破设备单一物理机形态，通过池化的方式对外提供动态服务，不同业务可能共享相同的物理、计算、存储资源，传统的物理隔离保护在资源池内部失效。利用虚拟化网络架构下被削弱的网络隔离，威胁可在边界突破单点后向内扩散，引入内

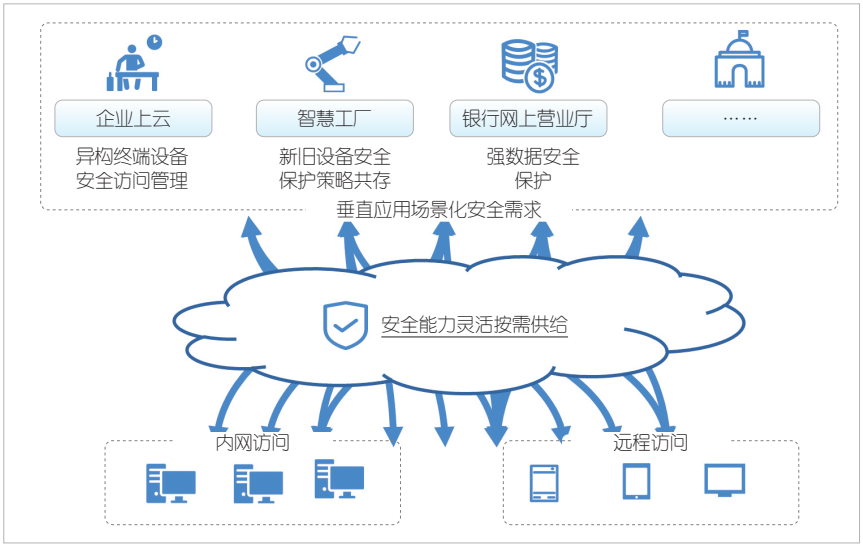
网横向移动攻击威胁，甚至潜藏或传播多处后形成攻击链，发动分布式拒绝服务（DDoS）、高级可持续威胁（APT）等更高级别的攻击，如图1所示。为应对攻击内部横向移动风险，需要从单点防御转向全网联防联动，实现全域安全能力的按需编排和弹性调度。



▲图1 虚拟化网络新架构引入攻击横向传播风险

2）从应用场景看，差异化的安全需求带来按需安全供给新挑战。

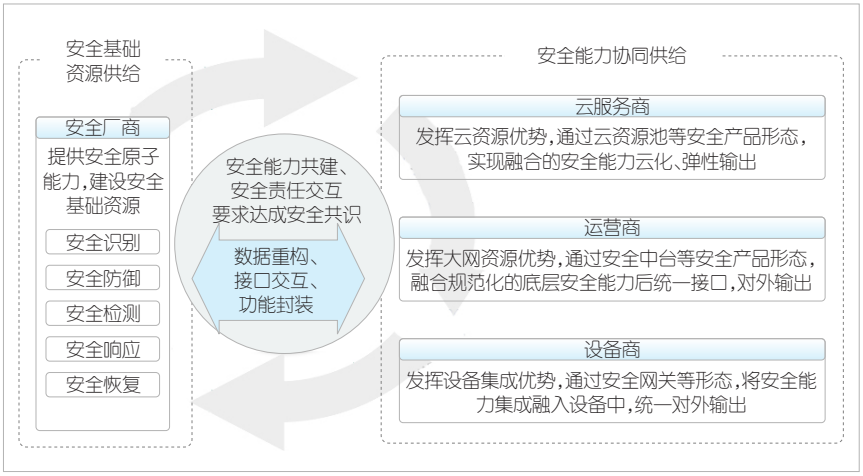
随着云网在数字化场景中的融合应用，云网上承载的垂直应用具有场景化安全需求。应用终端设备安全能力各异，组网架构更新迭代、周期各不相同，数据流量类型千差万别，网络性能要求高低不一，网络安全保障需求各不相同。垂直应用场景的差异化使安全特性从“通用安全”向“按需安全”转变，对安全能力的灵活性提出更高要求。例如：大型信息通信技术（ICT）企业上云存在多类不同终端异地安全接入的场景需求，要求基于身份与设备的分级别安全访问机制能够实现终端识别与安全防御，并具备对不同类终端的异常检测和响应能力，通过识别、防御、检测、响应能力联动实现对异构设备的全闭环安全管理，避免恶意终端对企业内部系统的破坏。智慧工厂中既有适用于早期封闭式生产环境的传统生产设备，又有新型的智能化生产设备。前者安全能力缺乏，需要部署外挂式安全设备来提供防御检测等安全能力，后者可与外部安全机制进行数据联动以实现安全保护。因此，我们需要多类不同的安全机制共存联动以满足新旧设备不同的安全需求。银行等金融系统对数据安全具有极高要求。银行有大量的分支机构和合作伙伴，要求基于业务、网络的分层分级逻辑隔离实现对数据的高效保护。如图2所示，为了满足不同应用场景差异化的安全需求，云网环境下的安全能力需要实现灵活编排重组，根据业务场景需求按需提供安全保障。



▲图2 满足差异化安全需求的安全能力

3）从产业生态看，纵向深度协同的新生态打开了云网安全协同新局面。

云网融合以运营商与云服务商主导的基础通信网络云化为核心，向下带动运营商与设备商开展云网一体硬件体系创新，向上为各类ICT数字化平台提供云网资源，推动产业生态融合发展。在云网上下协同产业的新生态下，安全角色也面临新变化，传统由安全厂商单一供给的状态发生变化，呈现出各方安全共建的网络安全协同新局面，如图3所示。安全厂商作为网络安全基础能力的建设者，由上至下提供安全能力供给；运营商与云服务商利用云网基础资源优势与入口优势，将安全厂商的成熟安全能力进行融合重组后赋能产业界。例如：运营商、云服务商等与安全厂商优势互补，合作推动安全访问服务边缘（SASE）等基于云网的安全能力供给服务，或在云网环境中融合抗DDoS、流量清洗等安全能力，通过安全即服务的方式，对外提供可按需订购的网络攻击防护服务；下游的设备商在云网融合硬件创新过程中，将安全能力封装在设备中，如通过安全网关等在硬件设备中融入安全能力。在各参与方深度参与云网安全的规划、建设和运营的过程中，安全责任界面需要各方共同构建。这不仅涉及安全厂商在供给安全原子能力时的安全责任，还涉及数据重构、接口调用、功能封装等环节引入的新安全责任，因此需



▲图3 云网安全协同新局面

要各参与主体在网络安全能力共建的不同环节（规划、建设、运营、使用等）达成共识，共担安全保护职责。

面对云网时代下的威胁横向移动风险、按需安全能力供给新要求以及云网安全协同新生态，单一的安全能力、单一的安全相关方无法适应云网融合新安全态势。技术发展、垂直应用以及产业生态3个层面都要求安全能力之间实现融合协同。打通网络安全能力间的互操作通道，可实现全域安全能力联动、安全能力灵活编排、安全能力高效集成。

2 网络安全互操作发展现状

2.1 标准化现状

1）国际标准化工作起步较早，多维度推动规范实践。

其他国家对安全互操作研究已久，政产学研高度协同：既有美国国土安全部（DHS）、美国国家安全局（NSA）等诸多国家部门引导，又有IBM、思科、戴尔、McAfee、Fortinet、Mandiant等各领域头部企业发力推进，同时还有国际电信联盟（ITU）、国际互联网工程任务组（IETF）等国际标准化组织推进标准规范研制。它们共同从安全协同顶层模型、基础类规范、语义类规范、接口类规范、数据类规范等诸多维度，推动安全互操作的研究应用[4]，如表1所示。

a）顶层模型。2014年，DHS、NSA等联合提出集成自适应网络防御框架（IACD），从顶层定义了安全协同参考架构、互操作规范草案、用例和实施案例，通过将安全产品抽象为“感觉–理解–决策–行动”工作流，共享威胁情报、编排协调响应和行动，实现采集、分析、决策、执行、恢复、信息共享的全自动化，提高了响应处理效率。根据金融服务信息共享和分析中心（FS–ISAC）的统计，基于IACD的系统将调查和响应事件的时间从11 h缩减至10 min，有效提高了安全效能。IACD已形成由FireEye、Splunks、Microsoft、VMware等主流安全厂商、机构、系统和产品构成的生态体系，被试用于美国FS–ISAC和美国能源部等多个部门。

b）基础类规范。安全能力作为安全互操作的基础元素，定义了为抵抗安全攻击所提供的能力。ITU–T、IETF等国际

▼表1 安全互操作国际标准化及应用情况

互操作层面		项目名称	提出年份	主导及参与方	应用情况
顶层模型		IACD	2014	DHS、NSA、约翰·霍普金斯大学应用物理实验室	试用于美国金融和能源等行业政府部门
基础规范		安全能力定义	2022	ITU–T SG17	—
		I2NSF工作组	2014	IETF	—
语义规范		OpenC2	2017	NSA、美国银行和奥斯陆大学等	多用于美国军方，未实现大规模商业应用
接口规范	威胁情报	TAXII	2012	DHS、OASIS网络威胁情报技术委员、MITRE等	被DHS、IBM、微软、惠普、思科、戴尔及大型金融机构等广泛应用
	通用安全	OpenDXL、OpenDXL Ontology	2016、2020	IBM、McAfee、Fortinet等	至2020年，OpenDXL已被4 000多个组织使用
数据规范		STIX	2012	DHS、OASIS网络威胁情报技术委员、MITRE、CTIN等	被DHS、IBM、微软、惠普、思科、戴尔及大型金融机构等广泛应用
		OpenIOC	2011	Mandiant	在威胁情报中心及相关产品有广泛应用，如启明星辰、安天等威胁情报中心，CounterCraft网络欺骗平台等

DHS：美国国土安全部
I2NSF：网络安全功能接口
IACD：集成自适应网络防御框架
IETF：国际互联网工程任务组
ITU–T SG17：国际电联安全研究组
NSA：美国国家安全局
OASIS：结构化信息标准促进组织
OpenIOC：开放威胁指标
OpenDXL：开放数据交换层
STIX：结构化威胁信息表达
TAXII：情报信息自动化交换

标准化组织启动了安全能力相关的标准化研究。2022年5月，ITU-T安全研究组SG17启动国际标准《X.secadef：安全能力定义》的制定，旨在为信息系统、网络、应用程序生命周期的每个阶段定义一组通用的安全能力。2014年，IETF成立网络安全功能接口（I2NSF）工作组，旨在通过提供架构、软件接口规范和数据模型，实现对物理和虚拟的网络安全能力的统一监控和管理。目前IETF已形成3个请求评论（RFC）、10余个工作组草案，包括《RFC8329: Framework for Interface to Network Security Functions》《RFC8129: Interface to Network Security Functions (I2NSF): Problem Statement and Use Cases》《RFC9061: A YANG Data Model for IPsec Flow Protection Based on Software-Defined Networking (SDN)》等。

c）语义类规范。2017年，NSA牵头形成了网络防御互操作性规范OpenC2，并制定了配套语言结构、配置文件、传输协议相关规范，通过将防火墙、沙箱等安全防御响应类产品间的交互抽象为“动作—目标”类标准化编排语言，以实现不同厂家和不同编程语言的安全产品间的自动化联动。受到各方保护自身威胁情报、漏洞库等安全信息商业价值的影响，OpenC2尚未实现大规模商业应用。中国奇安信、绿盟、启明星辰等在各自安全编排自动化与响应（SOAR）平台上提供相应适配和支持。

d）接口类规范。安全互操作接口类规范较为成熟，包括典型的情报信息自动化交换（TAXII）、开放数据交换层（OpenDXL）、OpenDXL Ontology等。基于指标信息的可信自动化交换TAXII是威胁信息共享领域的典型接口规范。该规范定义了网络威胁情报共享的协议、服务和信息格式，得到了美国国防部（DoD）、NSA等国家机构及IBM、思科、戴尔等的支持，具有成熟且广泛的应用。OpenDXL与OpenDXL Ontology是基于开源思路的通用型安全互操作接口通信类标准。2016年，McAfee发布了开源工具OpenDXL。OpenDXL通过提供软件开发工具包（SDK）来创建或连接基于数据交换层（DXL）的应用程序，协调不同供应商应用程序间的数据和操作，完成安全情报共享。截至2020年，OpenDXL已被4 000多个组织使用。2020年，在OpenDXL的基础上，由IBM、McAfee、Fortinet等联合成立的开放网络安全联盟（OCA）发布了开源的消息传递框架OpenDXL Ontology。该框架结合了OpenC2与结构化威胁信息表达（STIX）等通用消息内容开放标准，进一步定义了安全互操作消息格式。

e）数据类规范。典型的安全互操作数据类规范包括STIX、开放威胁指标（OpenIOC）等。STIX为威胁分析、威胁情报交换、检测和响应等安全行为提供描述威胁信息的语言和序列化格式，包括对威胁对象、威胁活动、威胁属性等威胁情报的多方面特征，被DHS、IBM、微软、惠普、思科、戴尔及大型金融机构等广泛使用。OpenIOC是Mandiant公司发布的情报共享规范。通过建立威胁指标（IOC）的逻辑分组，OpenIOC以可扩展标记语言（XML）文档类型描述捕获多种威胁的事件响应信息，包括病毒文件的属性、注册表改变的特征、虚拟内存等，在机器中以可读的格式进行通信，从而实现威胁情报的交流共享。该规范在威胁情报中心及相关产品中得到广泛支持。

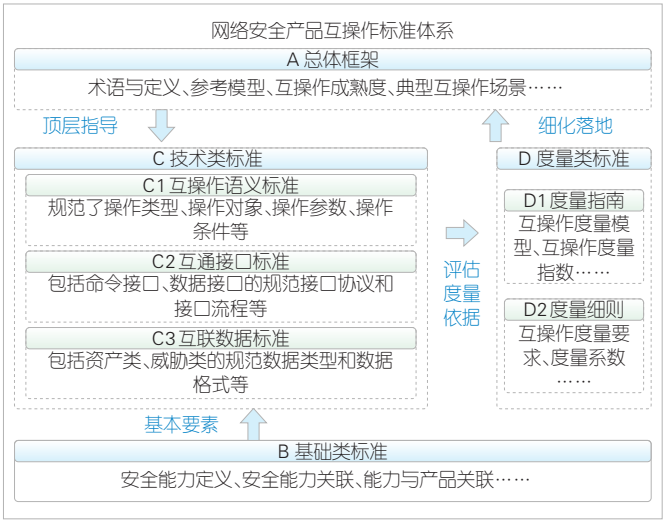
2）中国标准化工作发展不均衡，体系化标准建设迫在眉睫。

中国标准化工作起步较晚，各领域发展不均衡的特征较为明显。其中，整体框架和标准路线图研究初现成效，威胁领域互操作的数据规范化程度较高。安全日志、恶意程序等部分领域的网络安全互操作也形成了相关接口规范，但缺乏语义领域标准规范研究。

a）顶层框架。2022年3月，全国信息安全标准化技术委员会（TC260）发布网络安全国家标准需求，从国家标准层面提出网络安全产品互联互通框架、接口和数据格式等方面的标准制定的急迫性。2022年9月CCSA TC8 WG1（中国通信标准化协会工作组）结项的行业标准研究课题2022B72《网络安全产品互操作标准体系研究》，是中国首个通用的网络安全互操作标准项目。该项目解答了网络安全产品互操作发展现状、术语定义、参考模型、标准体系、标准路线等关键问题，为系统推动网络安全互操作系列标准研制提供顶层指导，如图4所示。2022年9月，CCSA TC8 WG1新立项的行业标准《网络安全产品互操作 第1部分：总体框架》，是安全产品互操作系列标准的首部标准，这标志着中国正式启动网络安全互操作标准体系化研制。

b）接口类规范。中国已在安全日志、恶意程序样本等信息互通方面发布了接口行业规范，如YD/T 3496-2019《Web安全日志格式及共享接口规范》、YD/T 2849-2015《移动互联网恶意程序疑似样本报送接口规范》等，定义了接口的名称、协议、流程、字段等信息。此外，中国通信标准化协会（CCSA）还启动了SOAR、SASE、安全中台等新安全技术产品与其他类安全功能间的接口规范化研制，使新技术在应用落地之初便具备与其他安全能力进行接口层面互操作的能力。

c）数据类规范。中国在网络安全互操作数据共享领域的标准化起步较早，已面向威胁情报、WEB漏洞、终端漏洞、源代码漏洞等形成了国家标准和行业标准，为安全产品间交互共享各类漏洞威胁信息提供了规范化格式。相关标准



▲图4 网络安全产品互操作标准体系结构图[4]

包括GB/T 28458–2012《信息安全技术 安全漏洞标识与描述规范》、GB/T 36643–2018《信息安全技术 网络安全威胁信息格式规范》、YD/T 3448–2019《联网软件源代码漏洞分类及等级划分规范》、YD/T 3667–2020《移动智能终端漏洞标识格式要求》、YD/T 3955–2021《WEB漏洞分类与定义指南》等。

2.2 行业应用现状

运营商、云服务商、安全厂商、设备商等发挥各自优势，以云安全资源池、安全中台、安全访问服务边缘等新网络安全产品为服务对象，打通网络安全能力互操作通道，向下对底层安全识别、分析、检测、响应等安全能力进行统一规范、编排、集成，向上通过云化方式、规范化接口等为上方数字化应用提供灵活、按需的安全能力，实现上层数字化应用需求与下层网络安全产品能力供给按需对接，如图5所示。

在云网环境下，运营商依托大网资源优势，通过实现多个安全厂商的抗DDoS安全产品互操作，以安全即服务等方式，为客户提供可定制、防护能力秒级生效、超大防护流量的抗DDoS服务。我们以此为例进行详细说明。

1）部署位置。运营商在城域网出口、IDC边界、骨干网络等重要位置部署多个安全厂商的抗DDoS检测设备和抗DDoS流量清洗设备，并集中部署抗DDoS管理平台，以管理全网抗DDoS检测设备和抗DDoS流量清洗设备。

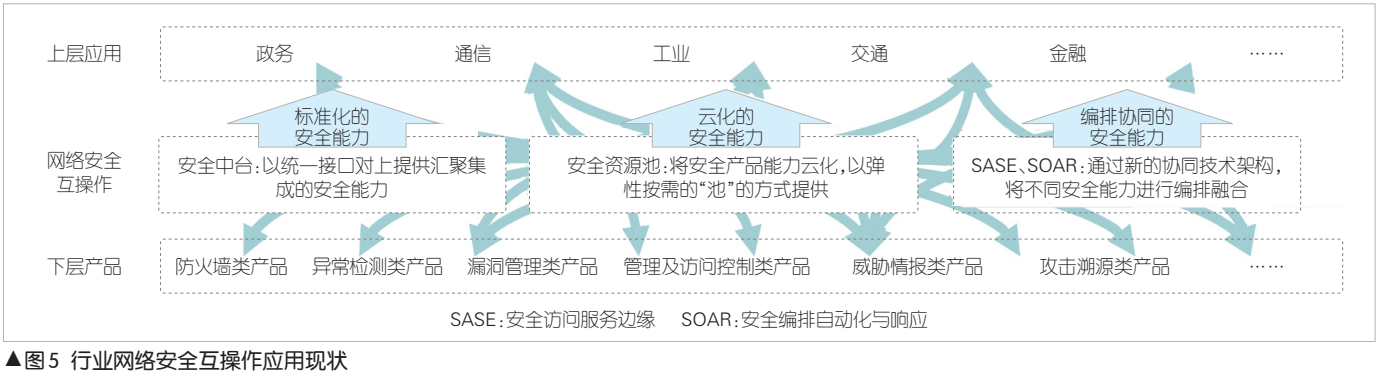
2）互操作过程。运营商统一规范各安全厂商的抗DDoS检测设备和流量清洗设备的DDoS攻击类型、异常流量信息、清洗策略等相关数据定义，并统一各安全厂商抗DDoS设备与平台间的接口，实现抗DDoS检测设备、流量清洗设备以及管理平台间的互操作，如图6所示。具体协同运作流程如下：

a）抗DDoS检测设备将检测出的异常流量信息上报给抗DDoS管理平台。上报信息包含异常流量攻击类型、异常流量攻击目标、异常流量五元组信息等。抗DDoS管理平台基于汇聚的异常流量信息，可形成对异常流量更加精准的判断，依此形成更精准的流量清洗策略。

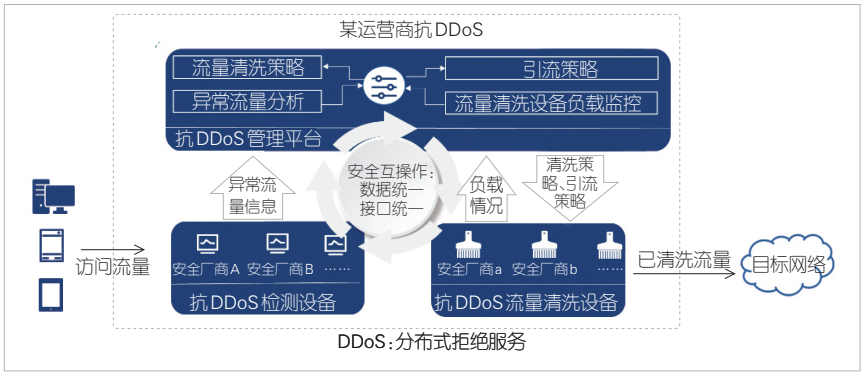
b）抗DDoS管理平台可监控抗DDoS流量清洗设备负载情况。根据汇聚的异常流量信息与抗DDoS流量清洗设备负载情况，遵循近源、流量均衡的原则，抗DDoS管理平台制定引流策略和流量清洗策略，并下发给相应的抗DDoS流量清洗设备，即引流至近源的抗DDoS流量清洗中心进行引流清洗，或在异常流量带宽过高时将流量牵引至多个抗DDoS流量清洗中心进行分级引流清洗，以此提升抗DDoS效能。

3 未来工作展望

网络安全互操作是云网时代下打造高效、灵活、全方位安全能力的技术基石。为实现高质量发展和高水平安全的良性互动，提升网络安全互操作实力，中国政府需发动产学研各界力量，开展标准规范、能力验证、行业示范3个方面工作。



▲图5 行业网络安全互操作应用现状



▲图6 抗DDoS网络安全互操作示例

1）规范先行，推动标准体系化建设。体系化的标准规范是推动网络安全互操作实践的前提。中国需要加快网络安全互操作标准体系化建设，梳理现有国家和行业标准，遵循急用先行、基础先行的原则，加快重点空白领域的网络安全互操作标准研究，如基础标准、语义类标准、评估类标准；鼓励运营商、云服务商、安全厂商、设备厂商等共同参与标准的研制与落地验证，确保标准的可操作性；推动在国际标准组织中牵头或参与网络安全互操作相关标准的立项研制，争取网络安全互操作领域的国际标准话语权。

2）能力验证，以实效评估推动能力提升。面向典型网络安全互操作场景、新型协同技术平台、典型安全产品等开展网络安全互操作评估验证，中国需要了解网络安全产业界的网络安全互操作现状，梳理网络安全互操作的优势领域与短板领域，形成未来网络安全互操作实践重点突破方向；推动网络安全互操作实验测试床建设，打造网络安全互操作能力验证硬能力，突破“现网级”安全实效验证。

3）打造标杆，遴选示范案例引领行业实践。依托网络安全试点示范等工作，遴选行业网络安全互操作优秀案例，打造单类网络安全产品、网络安全互操作平台类等不同类型的示范标杆，为威胁共享、身份验证、策略编排、运营管理等典型网络安全互操作场景下的互操作实践提供参考模板；通过网络安全互操作案例集等方式，引领行业开展网络安全互操作实践。

4 结束语

云网融合为数字化发展提供高性能网络连接、海量数据存储与多形态计算能力。为进一步夯实数字化中国发展的安全基础，本文从云网融合技术变革、融合应用、开放式生态3个方面梳理了安全新挑战，指出打通网络安全互操作通道是实现全域安全能力高效联动、提供灵活按需协同的安全能力的关键所在，可有效应对云网时代安全新挑战。未来，网络安全产业各方应携手并进，通过标准建设、能力验证、示范引领3个方面的工作，提升中国网络安全互操作水平，为云网时代打造扎实的安全基石。

参考文献

[1] DUAN Q, WANG S G, ANSARI N. Convergence of networking and cloud/edge computing: status, challenges, and opportunities [J]. IEEE network, 2020, 34(6): 148–155. DOI: 10.1109/MNET.011.2000089
[2] 柯瑞文. 立足科技自立自强全面推进云网融合 [J]. 人民论坛, 2021(36): 6–8. DOI: 10.3969/j.issn.1004–3381.2021.36.001
[3] 张鉴, 唐洪玉, 刘文韬, 等. 面向云网融合的电信网安全防护体系参考架构 [J]. 电信科学, 2020, 36(5): 10–15. DOI: 10.11959/j.issn.1000–0801.2020140
[4] CCSA. 网络安全产品互操作标准体系研究: 2022B72 [S]. 2022

作　者　简　介

魏亮，中国信息通信研究院副院长、ITU-T SG17副主席，教授级高级工程师；研究领域包括下一代电信网、网络架构、网络与信息安全等。

查选，中国信息通信研究院安全研究所高级工程师；研究领域包括无线网络安全、区块链安全等。

戴方芳，中国信息通信研究院安全研究所网络安全研究部副主任、ITU-T SG17 Q8报告人、CCSA TC8 WG1副组长、高级工程师；研究领域包括5G网络安全、云计算网络安全等。

基于超级SIM的5G端云安全体系架构与关键技术

Security Architecture and Key Technologies for Super SIM-Based 5G End-Cloud System

李佩源/LI Peiyuan，刘建伟/LIU Jianwei
（北京航空航天大学，中国 北京 100191）
(Beihang University, Beijing 100191, China)

DOI：10.12142/ZTETJ.202301004
网络出版地址：https://kns.cnki.net/kcms/detail//34.1228.TN.20230224.1443.002.html
网络出版日期：2023-02-24
收稿日期：2022-12-24

摘要：基于超级用户身份识别模块（SIM）卡的增强型安全功能，以及终端、云、端云协同3个方面的安全需求，结合区块链、雾计算、轻量认证等技术，设计了适配于5G大规模物联网场景下的安全高效且去中心化的端云鉴权认证机制。该机制以端云联动、主动立体防御为核心实现了终端应用全生命周期的安全防护，构建起云、端、卡协同运作的安全体系，为5G+业务安全赋能。

关键词：端云安全；超级SIM；访问控制；隐私保护；终端安全

Abstract: Based on the enhanced security function of the super subscriber identity module (SIM) card, and the security requirements of terminal, cloud, and terminal cloud collaboration, combined with technologies such as blockchain technology, fog computing, and lightweight authentication, a secure, efficient, and decentralized end cloud authentication mechanism suitable for 5G large-scale Internet of Things scenarios is designed. The mechanism takes end-cloud linkage and active three-dimensional defense as the core to realize the security protection of the whole life cycle of the terminal application, and builds a security system of collaborative operation of cloud, end, and card, which prompts the security of 5G business.

Keywords: end-cloud security; super SIM; access control; privacy protection; terminal security

随着物联网和云计算技术的发展，各类小型化智能终端能够通过“云”进行大容量数据存储和高速复杂计算，并将所需结果下载至终端本地。过去，由于网络带宽及可靠性有限，端云协同难以在实际应用中部署。5G网络因其“高带宽、高可靠、低时延、海量连接”特性，使端云协同的网络架构在实际应用部署中成为可能。在端云协同系统的运行过程中，云、终端、协同机制任一部分的安全影响的不仅是其自身，还会是整个系统。攻击者可以以任意一个模块的安全缺陷为跳板，对整个系统进行攻击。端云协同体系作为国家信息基础设施的重要底座，直接影响国家信息基础设施安全，因此亟需构建一套5G环境下安全的端云协同体系。

1 5G端云网络架构及其安全挑战

过去几年，中国云计算产业呈现出高速增长的态势。根据中国信息通信研究院（后文简称信通院）《云计算白皮书（2021）》中的相关数据，2020年中国云计算整体市场规模达2 091亿元，企业“上云用云”进程加快。同时，物联网技术的发展进入快车道。根据信通院预测，到2025年中国物联网连接数将达到80.1亿个。万物互联将进一步释放数据驱动力，推动各行业数字化转型发展。两种技术在各自快速发展的过程中又互相渗透、彼此融合，形成端云一体化协同运作的网络架构[1]。

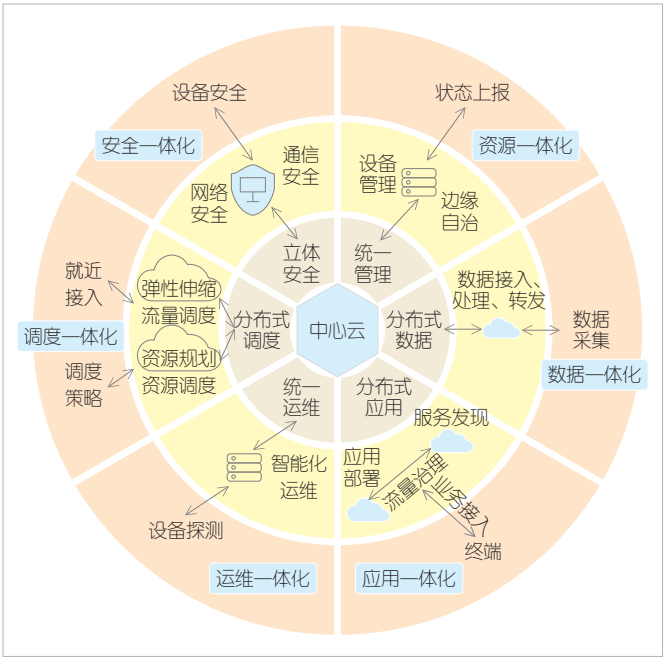
1.1 5G端云网络架构简介

5G通信技术使得海量终端和云基础设施的一体化融合成为可能。终端设备具有大量传感器，可以实时捕获大量数据，但由于其受到存储能力、计算能力、通信能力的限制，无法对海量大数据进行集中统一的整合处理。这使得数据价值得不到充分发挥，出现数据孤岛的窘境。云计算将海量物联网终端设备在感知层获取的数据信息，通过网络层传输到一个标准平台上，再利用高性能的中心云进行处理，并赋予

基金项目：国家重点研发计划（2021YFB2700200）；国家自然科学基金（U21B2021、61972018、61932014）

这些数据智能，最终转换成对终端用户有用的信息。同时，云端也可以长期存储海量数据，并统一管理广泛分布的终端，终端设备间也可以通过云端进行远距离交互和信息共享。5G通信技术的出现，极大促进了云计算和物联网技术的融合与优势互补，5G端云一体化网络架构（如图1所示）的形成有力推动了中国信息化产业的高质量发展[2]。

5G环境下的端云融合是信息技术（IT）与通信技术（CT）融合的新阶段，是新型信息基础设施的底座，是赋能数字化转型的基础，同时也是电信运营商、互联网公司和各类信息与通信技术（ICT）制造商和供应商共同追逐的目标。端云旨在屏蔽云、端分布式异构基础设施资源，提供统一视角进行资源的管理和使用，实现数据自由流通、业务应用统一运行，构建立体化安全保障能力，满足多样化、实时敏捷、安全可靠业务需求。如图2所示，端云融合可以实现资源一体化、运维一体化、数据一体化、应用一体化、调度一体化、安全一体化的目标[1]。



▲图2 端云一体化融合架构

1.2 5G端云网络架构的安全风险

现有的端云体系在终端、云、端云协同机制3个方面都存在一定的安全隐患：

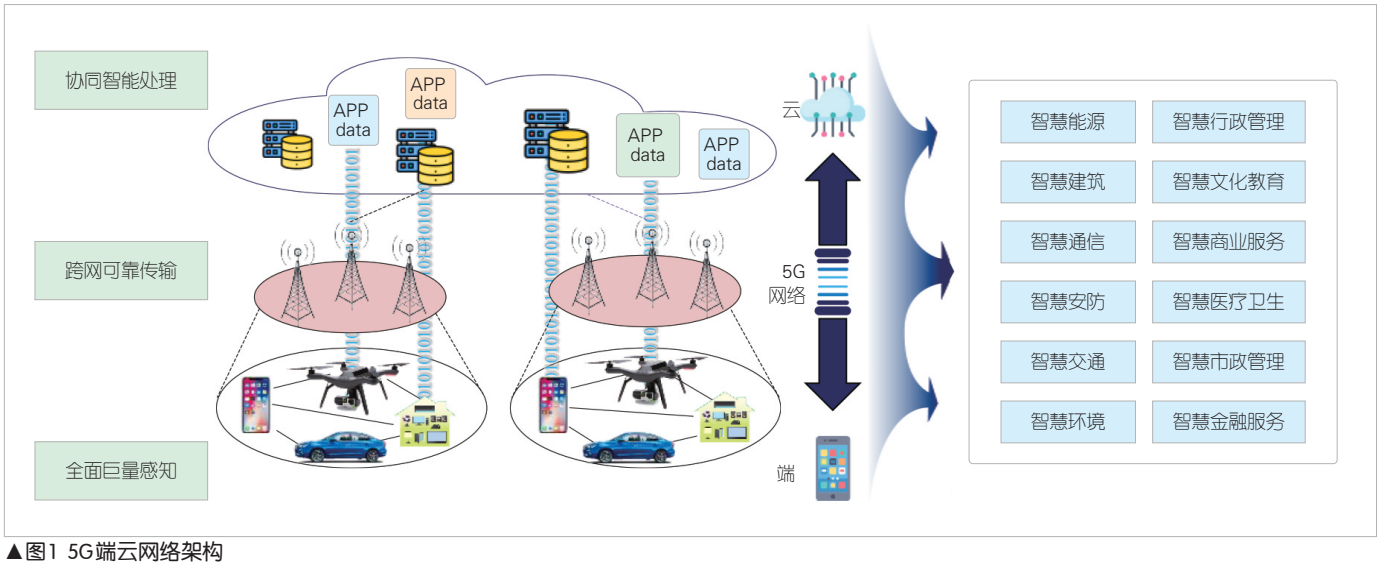
1）在终端应用安全方面，缺乏有效的应用安全检测和防御技术，并且缺少高效率的终端应用合规检查和运行质量检测系统。

2）在云端数据存储方面，主要存在隐私保护方面的问题。终端与服务器之间的通信过程存在数据窃听、数据泄露以及数据完整性破坏的风险，海量用户数据的云端存储和动态更新将会使服务器数据隐私保护面临挑战[3]。

3）在端云协同机制方面，主要存在身份认证方面的风险。现有终端认证服务吞吐量低、可扩展性较差；由于连接设备种类多样，认证面临劫持终端、冒用身份等安全接入风险。

1.3 5G端云网络架构的安全需求

在5G场景下，海量终端存在大规模的接入认证等需求。同时，用户对云计算、云存储等云服务的需求越来越广泛，而端云网络架构的安全性是确保用户能够正常使用网络服务的关键。针对上述安全威胁，我们从终端、云、端云协同机



▲图1 5G端云网络架构

制3个层面提出对应的安全需求：

1）在终端层面，存在终端应用安全防御、多应用安全隔离、应用安全评测等需求；

2）在云层面，需要有安全、层次化的密钥管理措施，高效的数据隐私保护机制和分布式、可扩展的安全服务；

3）在5G网络端云协同机制层面，网络需要满足低时延的身份认证、大规模的访问控制和高效加解密等需求。

2 超级SIM的安全优势与挑战

2.1 超级SIM简介

5G超级用户身份识别模块（SIM）的出现，能够赋能端云安全体系，如图3所示。超级SIM增强了安全能力，实现了机卡接口升级，并通过空间开放和多应用安全隔离为各行业的合作伙伴提供了多业务承载的大容量高安全等级的优质容器，是5G用户网络身份认证和应用敏感数据的安全存储空间。

超级SIM创建了安全信任根，能够提供安全可信的计算能力：在存储方面，融合了存储卡和SIM卡，支持GB级别的大容量安全可信存储。与此同时，超级SIM进行了卡机接口升级，其高开放性能够为设计和开发人员提供更便捷的应用。在超级SIM的可信存储和计算之上，能够实现增强的接入认证等功能。

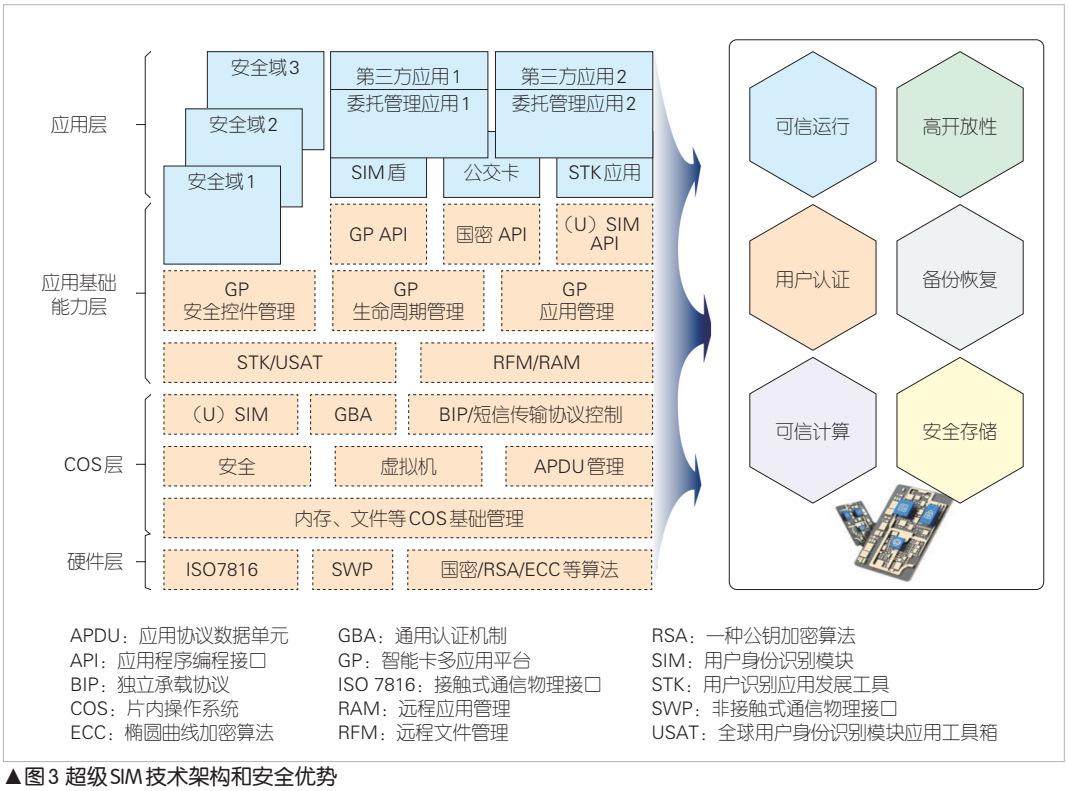
2.2 超级SIM的安全优势

SIM卡作为运营商认证用户身份的硬件载体，从2G时代就作为终端接入网络的主要凭证。在SIM卡中存储用户的证书密钥，能够实现终端与网络的双向认证接入。随着接入认证场景的安全需求越来越高，传统SIM卡用于终端与网络的认证已难以满足5G时代大连接、多应用场景需求，而超级SIM为业务合作带来无限可能，将成为承载各类业务的高安全性优质容器。基于超级SIM实现端云一体化鉴权认证是技术发展的大趋势，有助于实现真正的万物互联，构建可信的网络空间。

1）目前弱口令认证机制存在易破解、效率低的缺陷。基于超级SIM的端云一体化鉴权机制充分发挥了超级SIM具备的安全可信运行环境、安全可信存储环境、安全可信计算能力等优势，可有效弥补现有鉴权认证方式的不足，提升系统身份鉴别与登录认证能力，从技术上解决信息安全风险，提升网络安全整体水平。

2）超级SIM能为多行业大连接终端提供增强接入认证。此外，与2G/3G/4G的SIM对比，超级SIM除提供基础电信服务外，还可凭借高安全性、高开放性特点发挥更大作用，赋能千行百业。丰富的业务能力需要多层架构、多组件的支撑。一方面，超级SIM通过将加密认证机制与终端硬件层、片内操作系统（COS）层、应用层紧密耦合，保障接入终端的硬件可信；另一方面，引入统一的SIM卡鉴权认证机制，能够有效屏蔽物联网场景下多终端接入安全性差异，实现可编程、动态的授权接入，也能够基于终端位置、标识、可信根等多种维度对终端接入网络和云上应用进行鉴权认证，从而有效防范劫持终端、冒用身份等安全接入风险。

3）建设超级SIM生态能够有效提高中国相



▲图3 超级SIM技术架构和安全优势

关技术自主化可控水平。超级SIM主要依赖的网络终端访问控制机制是以美国为主导的，并不是开源的。如果出现针对超级SIM的技术限制，那么将影响中国基础电信服务的正常运作，也会影响接入5G网络使用超级SIM的各行业的正常运作。因此，我们要开展超级SIM生态的研究，构建基于SIM卡的跨终端多层接入认证和加密通信架构设计，形成具备中国自主化技术的超级SIM终端生态安全，构建基于5G超级SIM的安全端云完整体系架构。

2.3 基于5G超级SIM的端云体系的安全问题

基于5G超级SIM的端云体系尚存的安全问题具体如图4所示。在5G网络异构、终端异质、海量连接的场景下，在终端设备鉴权过程中，传统中心化身份认证、访问控制等方式存在单点故障、易遭受拒绝服务攻击等安全问题。

5G超级SIM中的信息涉及用户的隐私，具有高度敏感性。用户数据安全与隐私保护问题是制约5G超级SIM网络应用普及的重要因素。协同系统提供多种手段来支持运营商合作伙伴间的信息共享，这和云端服务器的数据安全与用户的隐私保护需求之间形成了激烈的冲突[4–5]。

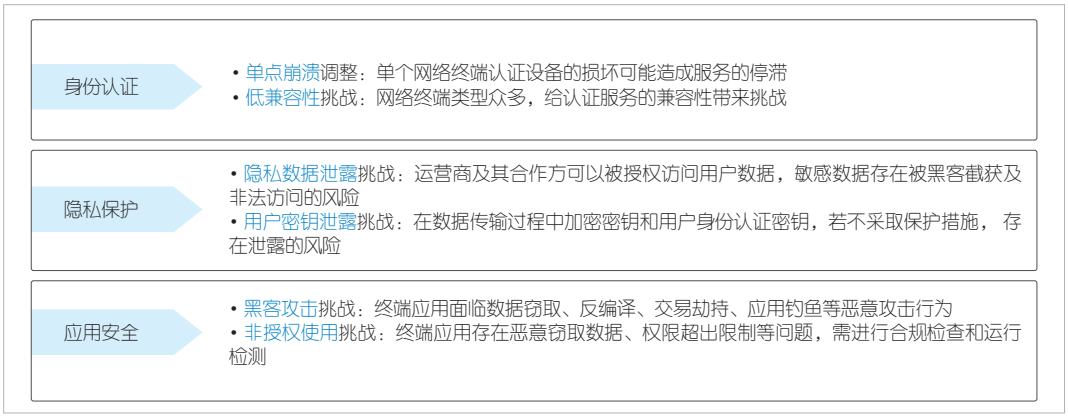
随着移动互联网场景化需求的快速发展，移动端应用的数量和业务种类呈现出爆发式增长。与此同时，移动终端面临的安全威胁种类和数量也在不断增多，例如：手机操作系统漏洞存在不可预知的业务逻辑缺陷，动态攻击、虚假设备等攻击手段和网络黑产链严重威胁用户的资金和隐私安全。现有的防御方法主要有对终端应用的静态特征检测、动态特征检测、动静结合的特征检测、基于深度学习的特征检测等，这些防御方法大都是从被动防御的角度在应用正式使用前或使用时进行测评，只考虑了在终端应用的部分生命周期，且仅局限于单个软件安全的层面。这造成了长期以来针对终端系统的恶意攻击层出不穷、“治标不治本”的后果[6–9]。

3 基于5G超级SIM端云安全体系的关键技术

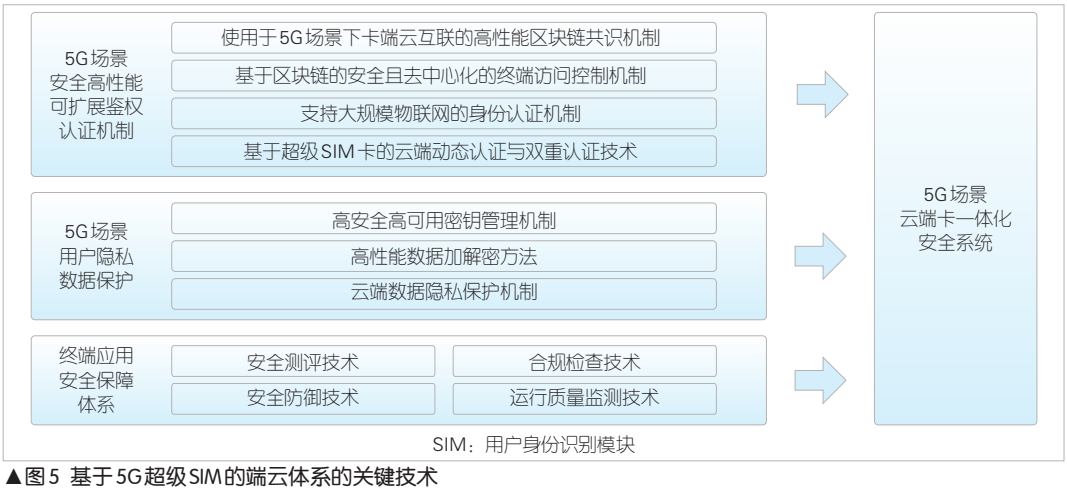
针对上述基于5G超级SIM的端云体系尚存的安全问题，结合区块链、雾计算、零信任等新兴信息技术，我们创新性地提出以下关键保障技术（如图5所示），进一步完善安全机制，构建基于5G超级SIM的安全端云体系[10]（如图6所示）。

3.1 5G场景下基于区块链的终端访问控制机制

由于5G网络规模庞大，5G终端设备从硬件类型到操作系统均有较大差异，因此如何能够在5G海量网络终端的环境下以超级SIM为核心，实现安全且去中心化的终端访问控制机制成为首要问题。这个机制能够使网络设备进行统一的鉴权验证，确保入网终端的安全可靠，并实现安全且去中心化的认证。该机制要具备如下一些特点：首先，要保证抗损



▲图4 基于5G超级用户身份识别模块(SIM)的端云体系尚存的安全问题



▲图5 基于5G超级SIM的端云体系的关键技术

毁效果好，稳定性强，单个或多个网络终端备鉴权认证设备的损坏不会造成重大影响；其次，要兼容性好，对网络终端设备的硬件类型、操作系统、业务应用有较好的兼容性；同时，还要保证维护性强，在网络终端设备出现故障时能够实现快速维修。

采用区块链技术能够实现去中心化的SIM卡终端鉴权认证和访问控制，防止鉴权服务节点发生单点故障，从而提供安全的身份认证和访问控制。5G场景下安全高效及可扩展的访问控制机制以云、端、卡互联的高性能区块链共识为基础，基于区块链构建去中心化的网络终端鉴权认证体系框架，对各终端进行安全去中心化的访问控制。

适用于5G场景下云、端、卡互联的高性能区块链共识机制，以可验证随机函数、可验证秘密分享和门限签名等重要密码工具为基础，主要包括3个模块：云端分布式随机数生成模块，用于委员会成员分配、委员会领导节点选举以及重配置过程；5G云节点重配置模块，用于筛选和替换参与共识的节点；云节点委员会内共识模块，用于处理交易并对新区块达成共识。

在5G海量终端背景下，需要以区块链为基础、超级SIM为核心，构建基于区块链的网络终端访问控制机制。区块链中的区块应包含加密后的终端详细信息及对应的访问控制权限。网络中任意一台入网终端都能够访问区块链上的信息，但只有合法终端能通过解密算法对区块链上的信息进行读取，从而判断该终端的类型及访问行为是否合法。可以采用基于区块链的访问控制模型，将权限绑定到角色，并将角色分配给5G网络终端以促进权限管理。

5G网络终端设备在注册和激活后可接入系统，设备之间、客户端和设备间进行交互前需要进行身份认证和权限验证，以确认设备或用户的身份以及对相关资源的访问权限。设备注册、设备激活、设备身份认证和设备访问权限认证以及访问控制策略管理都需要和区块链平台进行交互。使用基于区块链技术的分布式鉴权认证模式，可实现在网络状态不稳定、入网终端分散、维护力量薄弱的情况下的入网终端的鉴权工作。

3.2 支持大规模物联网场景的身份认证技术

针对大规模物联网场景，构造基于身份信息的区块链能够保证物联网设备身份的完整性和可靠性，同时可以基于端、雾、云提供不同设备的身份认证机制。为了保证物联网安全性，同时对用户敏感信息进行最大程度的保护，基于超级SIM的物联网设备的合法身份凭证只有经过加密和承诺后才能注册到区块链上，区块链平台再使用智能合约对设备进行访问权限控制。在不同的物联网设备交互前，发起设备需先向物联网平台和相应接入设备节点发送鉴权请求和接入请求，经平台调用后区块链系统对接入设备双方进行身份标识认证和管理，保证各设备在不泄露敏感信息的前提下，实现数据的安全交互。具体技术路线如下：

1）在线快速身份（FIDO）。FIDO可用于SIM卡端身份信息认证。该技术的核心是将身份认证手段与身份认证协议解耦，以密码技术为基础，采用非对称密码算法机制，以密钥作为用户凭证，通过签名验签的方法完成对用户的身份鉴别。

2）轻量级设备安全认证技术。大规模物联网中存在大量计算能力极低且存储力极低的终端设备，如电力感知层设备、车间监控设备等。NTRU（公钥密码体制名）是一种基于格理论的轻量级公钥密码算法，可根据不同的安全级别进行不同的参数集选择，具有内存和计算量少、加解密和签名/验证速度快、安全性高等优势，适用于海量终端的物联网场景下资源受限的轻

▲图6 基于5G超级用户身份识别模块(SIM)的端云体系架构

量级节点的无线通信环境。

3）融合端雾云架构与区块链技术的物联网系统。在物联网中选取满足一定计算和存储能力的设备或高性能节点作为控制中心雾节点，以其为中继进行多层次计算任务卸载和大数据处理。网络边缘的雾计算节点实时感知多个SIM卡设备终端的计算资源富余信息，雾控制节点能够以整体效率最大化的方式有效管理终端自组织微集群的构建。一方面，雾计算节点可以实时检测多终端计算卸载的任务，合并用户计算任务中的同质任务，实现多个同质任务一次卸载而多用户共享，从而减少网络通信开销和云资源占用；另一方面，雾控制中心可利用雾计算节点充当计算预处理节点，按需调配云计算资源，与公有云协同应对大型计算任务。

3.3 基于超级SIM的云端动态认证与双重认证技术

端云协作鉴权认证机制以超级SIM的安全能力为访问主体（包括用户、设备、应用等）建立可信身份标识，结合零信任安全架构，并基于终端SIM卡标识、位置、网络地址等多维度因子，构建可信应用代理、可信应用程序编程接口（API）代理、可信访问控制台、智能身份分析系统，从而实现基于会话连接粒度的动态访问控制。

在终端接入5G网络时，SIM卡网络层与应用层的双重认证可以构建“端–网–云”的可信通信。针对5G终端业务低时延、大并发、高可靠等不同场景，基于物联网轻量级认证要求、轻量级安全认证协议、轻量级网络认证机制，形成针对“终端–终端”“终端–边缘云”“终端–核心云”“终端–接入网”等多段超级SIM认证方案。

3.4 高安全高可用密钥管理机制

针对5G网络环境中大数量级终端、复杂接入场景等新情况所带来的密钥管理问题，需建立新的全套密钥管理方案，以适应云、端、卡协同体系中的密钥管理需求。

在密钥的生成、分发和认证模块中，根据用户所需的不同密钥安全等级，按照不同的密钥产生方式生成不同等级的密钥；根据不同的网络架构，利用多种密码技术，选择相应的密钥分配模式以实现密钥分发。

在密钥分发过程中，我们需要对密钥进行认证，以确保密钥被正确、完整地送达。在密钥分级保护、存储与备份模块中，根据密钥的使用场景和安全等级对密钥进行分级保护，并在密钥存储时选择合适的方法，保证密钥的机密性、可认证性和完整性，以防止密钥泄漏和被篡改。在密钥保护和存储过程中需要考虑密钥的备份问题，以避免密钥因意外而丢失而造成的损失。

在密钥更新与销毁模块中，当密钥泄漏或丢失时，设计合理的更换密钥方法，可以使损失最小化；根据密钥的作用和安全等级，设置合理的密钥使用期限，实现分级管理；对已泄漏、已过期的密钥及时进行销毁，设计安全可靠的销毁方式，以避免攻击者通过旧密钥寻找有关的秘密信息。

3.5 高性能数据加解密方法

面向5G场景下大带宽实时传输的通信需求和日益强大的攻击者，我们需设计基于超级SIM的高性能数据加解密算法，以及算法的相应逻辑电路实现方式和应用模式。

在算法层面，结合国产密码算法系统和密码评价标准，以及5G网络对密码算法轻量级、低功耗、抗侧信道攻击等要求，我们需要设计可应用于超级SIM及整个5G网络系统的密码算法系列。

在硬件层面，根据所选密码算法逻辑结构特点，结合超级SIM平台，我们要选择适宜的并行电路架构，并采用流水线技术等，缩减电路规模，提高密码算法电路的工作频率和数据吞吐率。在接口设计方面，采用容错技术、握手机制与端口数据寄存技术等，并兼顾算法自检电路设计，确保密码算法电路异常工作状态能够被实时检测，从而提升整个密码模块的工作可靠性。采用动态电压调节、门控时钟和可变频率时钟等技术，可以降低密码算法实现电路的功耗。

在应用层面，我们需要结合5G场景具体应用需求，设计密码算法的智能化调用策略；并引入多级安全、域隔离等思想，制定不同需求下的密码算法的调用规则，包括对称密码与非对称密码算法的高效运用策略、密码强度分级策略等，实现整个系统效率与安全性的统一。

3.6 云端数据隐私保护机制

为了应对5G网络下海量的个人数据与用户终端有限的存储和计算能力之间的矛盾，我们往往需要借助于云端来辅助用户数据的处理，从而带来数据所有权与管辖权的分离。在上述条件下，我们需要在保护云端数据安全的同时，保证用户对数据合法、灵活、高效的访问。

1）数据安全共享机制。该机制利用层次身份基的可撤销数据访问权限管理方案，以免交互的方式撤销无效用户的访问凭证，动态地管理用户数据的访问权限，为有效用户免去权限撤销操作中繁复的计算和通信开销。同时，考虑到超级SIM协同体系的巨大规模，基于分层的用户结构可以减轻私钥生成中心（KGC）为所有用户生成访问控制凭证的负担，提高系统的工作效能。

2）跨系统的云端隐私数据保护机制。针对计算能力较

弱的移动设备，该机制利用轻量级的基于身份的广播加密系统，根据授权用户集合对用户隐私数据实施大范围、灵活的控制。引入代理重加密机制，在身份基广播加密和身份基加密这两种系统之间搭建超级SIM数据直接共享的通道，使得不同加密系统下的超级SIM用户可以快速安全地共享个人隐私数据。

3.7 终端应用安全保障机制

基于当前应用市场的安全现状和现实需求，亟需提出一套能满足以下需求的全新终端应用安全管理方案：多平台通用、主动被动防御结合、在系统层面进行预防监测、涵盖终端应用全生命周期。具体应包括：

1）终端应用安全性检测。该方案利用了被动防御的思想。首先，检测对象覆盖范围要全面，针对终端平台上的所有应用类型，包括Android应用、iOS应用、Web应用、开发包、函数库等；其次，检测类型要全面，包括软件恶意行为检测，软件漏洞检测与修复，软件行为、权限、隐私策略等。针对上述目标，结合深度静态检测、动态监测、源代码扫描、人工智能、自然语言处理等技术，构建一套高效、自动化的测评系统。

2）终端应用安全加固。该方案利用了主动防御的思想，针对移动应用面临的反编译、二次打包、内存注入、动态调试、数据窃取、交易劫持、应用钓鱼等恶意攻击行为，将针对各种应用安全缺陷的保护技术集成到应用客户端内，构建全面保护软件安全的主动防御体系。这些技术主要包括：代码防逆向技术、应用防篡改技术、反调试技术、数据防泄漏技术、运行环境保护技术等。

3）运行时监控与态势感知。该方案利用系统防御、应急处置、恢复溯源的思想，针对某些恶意应用采用高级反检测技术逃过安全性检测的情况，或者某些恶意软件隐蔽性极高的高级可持续威胁（APT）攻击，从系统安全的角度对移动应用上线后的动态运行安全问题及运行稳定性问题进行实时监控，充分挖掘软件运行模式，识别其安全属性，同时为软件检测提供丰富的数据支持。基于上述数据，可以进一步建立企业端和用户端联动的立体化安全态势感知体系，这样既弥补了企业业务反欺诈、风控等业务系统对终端风险监测的短板，又为用户快速建立事前预警、事中处置、事后恢复的自动化安全体系。

4 结束语

本文中，我们认为需要以高性能国产密码SIM为核心，充分发挥区块链、雾计算、端云联动立体防御等新兴技术的优势，构建“云、端、卡”协同运作的完整安全体系。这一举措既可以适应5G场景下大规模物联网设备接入认证的新场景，也能覆盖芯片硬件、芯片操作系统、终端芯片、终端硬件设计、终端应用等的研发和应用，为5G+业务安全赋能，推动数字中国关键基础设施的构建。

参考文献

[1] 云边协同产业方阵. 云边端一体化发展报告(2022年) [R]. 2022
[2] 徐恩庆, 李昂, 王蕴婷. 云边端一体化创新助推算力泛在化发展 [J]. 通信世界, 2022, 4(18): 44-45. DOI: 10.13571/j.cnki.cww.2022.18.006
[3] 欧阳雪, 徐彦彦. IaaS 云安全研究综述 [J]. 信息安全学报, 2022, 7(5):39-50
[4] ZHAO H Y, YANG X, LI X L. cTrust: trust aggregation in cyclic mobile ad hoc networks [C]//Proceedings of the 16th International Euro-Par Conference on Parallel Processing: Part II. ACM, 2010: 454-465. DOI: 10.5555/1885276.1885325
[5] ZHANG C, SUN J Y, ZHU X Y, et al. Privacy and security for online social networks: challenges and opportunities [J]. IEEE network, 2010, 24(4): 13-18. DOI: 10.1109/MNET.2010.5510913
[6] LI J, SUN L C, YAN Q B, et al. Significant permission identification for machine-learning-based android malware detection [J]. IEEE transactions on industrial informatics, 2018, 14(7): 3216-3225. DOI: 10.1109/TII.2017.2789219
[7] ALZAYLAEE M K, YERIMA S Y, SEZER S. DL-droid: deep learning based android malware detection using real devices [J]. Computers & security, 2020, 89: 101663. DOI: 10.1016/j.cose.2019.101663
[8] DIMJASEVIC M, ATZENI S, RAKAMARIC Z, et al. Android malware detection based on system calls [R/OL]. [2022-12-10]. https://www-old.cs.utah.edu/docs/techreports/2015/pdf/UUCS-15-003.pdf
[9] YU L, LUO X P, QIAN C X, et al. Enhancing the description-to-behavior fidelity in android apps with privacy policy [J]. IEEE transactions on software engineering, 2018, 44(9): 834-854. DOI: 10.1109/TSE.2017.2730198
[10] 天融信：构建“云、管、边、端”协同防御的5G端到端安全闭环 [EB/OL]. [2022-07-05]. https://zhuanlan.zhihu.com/p/537693680

作　者　简　介

李佩源，北京航空航天大学在读硕士研究生；主要研究领域为网络攻防、系统安全、软件安全。

刘建伟，北京航空航天大学网络空间安全学院教授、博导、院长，享受国务院政府特殊津贴，现任国务院学位委员会第八届学科评议组成员、教育部高等学校网络空间安全专业教学指导委员会委员、中国密码学会常务理事、中国指挥与控制学会常务理事、中国电子学会网络空间安全专委会副主任委员、中国指挥与控制学会网络空间安全专委会副主任委员、中关村智能终端操作系统联盟副理事长；曾获国家技术发明一等奖、国防技术发明一等奖、中国指挥与控制学会科技进步一等奖等，所编写的教材获全国普通高校优秀教材一等奖、国家网络安全优秀教材、国家精品教材、全国优秀科普作品奖、第四届中国科普作家协会优秀科普作品金奖等；出版教材7部、专著2部、译著1部。

云网融合下的安全能力池关键技术与应用

Key Technologies and Application of Security Capability Pool for Cloud-Network Convergence

余启明/YU Qiming，吴爽/WU Shuang，黄帅/HUANG Shuai，刘紫千/LIU Ziqian

（天翼安全科技有限公司，中国 北京 100020）
(China Telecom Cybersecurity Technology Co., Ltd., Beijing 100020, China)

DOI：10.12142/ZTETJ.202301005
网络出版地址：https://kns.cnki.net/kcms/detail//34.1228.TN.20230224.1459.006.html
网络出版日期：2023-02-24
收稿日期：2022-12-15

摘要：设计并实现了云网安一体化的安全能力池系统。该系统将重要的网络安全能力虚拟化、原子化后沉入边缘云节点进行部署，支持多种流量型与非流量型的安全服务场景。系统基于IPv6的段路由（SRv6）与Flowspec技术实现了安全服务链编排与流量调度，配置简单，灵活高效，并通过标准化接口实现了跨厂商安全能力统一纳管。系统通过工程方法解决了生产运行过程中存在的运行速度慢、业务中断等问题，实现了安全能力的集中管理与智能调度。目前，安全能力池已进入商用阶段，服务客户上千家，累计防御次数达到百万级。

关键词：云网融合；安全能力池；近源防护；弹性扩展；按需组合

Abstract: A security capability resource pool system that integrates cloud, network, and security is proposed. The system deploys major security capabilities at the edge cloud nodes and supports a variety of traffic-mode and non-traffic-mode service scenarios. Based on Segment Routing IPv6 (SRv6) and Flowspec technology, the resource pool achieves service chain orchestration and flow scheduling with simple configuration and high efficiency. Meanwhile, the system manages cross-vendor security capabilities through standardized interfaces. In addition, the whole system solves the problems such as slow running speed and service interruption in operation and achieves unified management and intelligent scheduling of security capabilities. At present, with the commercial use of the security capability resource pool, thousands of customers have been successfully protected from millions of cyber-attacks.

Keywords: cloud-network convergence; security capability resource pool; near-source protection; elastic expansion; integration by requirements

互联网信息技术发展日新月异，大数据、云计算等新兴技术加快了进入政务、金融、医疗、教育等行业的步伐[1-2]。与此同时，网络安全问题也日渐凸显，如何高效满足各行业企业对网络安全防护的多种需求成为产业界亟待解决的问题。传统企业具有相对固定的网络安全边界，因此安全厂商常采用串接或者并接硬件设备的方式为用户提供安全解决方案。此种方式的弊端日益凸显，例如：部署繁琐，功能单一，可维护性差等[3-4]。随着网络安全威胁的持续变化和企业自身数字化转型的深入，企业对安全防护手段的多样化需求与日俱增，对各种手段的综合防护能力和效果的要求也不断提高。传统的依靠堆叠安全硬件的方法已经无法满足用户对安全能力按需快速组合和防护能力弹性扩容的需求。因此，通过可运营升级的云化软件即服务（SaaS）安全防护方案来解决各类网络安全问题已经成为了行业新的重要趋势[5-6]。

为满足企业用户对安全防护能力多层次可定制的需求，在运营商云网融合的技术驱动下，云资源池应运而生[7-10]。本文提出了一种云网安一体化的安全能力池技术方案，提高了网络安全防护的灵活定制和可编排能力，满足网络安全能力的可快速扩展的要求。安全能力池技术方案主要基于边缘云技术、自动化云端部署安全防护能力，通过软件化、服务化的安全能力为用户提供实时安全保障，并可直接借助安全管理平台对所需安全能力进行统一管理与配置，极大地提升了安全能力的使用效率，降低了使用成本。

1 系统整体设计方案

根据中国电信网络安全统一规划，安全能力池计划覆盖中国电信所有主要的城域网，为客户提供复合安全防护能力。这些安全能力可分为流量型和非流量型两类：流量型指基于业务流量行为进行实时检测阻断的安全能力，这类能力

往往需要在业务流量流经路径中进行干预才能起到效果，例如：防火墙、入侵检测和Web应用防护等；非流量型指不依赖业务流量路径干预也能发挥作用的安全能力，例如：漏洞扫描、日志审计等。安全能力池系统需要解决复杂流量调度、多业务场景编排、跨厂商应用维护等技术问题。此外，安全能力池需要作为多生态能力的承载平台，与上层应用低耦合，并具备高扩展能力。安全能力池分为上、中、下3层架构，如图1所示，分别为安全能力管理平台层、安全业务中台层和资源池层。

上层安全能力管理平台（后文简称为安管平台）在多租户场景中为角色和权限各异的用户提供统一的管理访问入口；中间层为安全业务中台，主要实现对资源池的纳管、原子能力的适配和服务之间的安全认证。从图1中可以看到，业务中台将安管平台与底层资源池分割开来，并将复杂的安全业务场景逐步拆分，这样降低了系统耦合度，同时提高了系统的高扩展性。最下层为资源池层，具备多种安全原子能力、流量调度与服务链编排能力和大规模数据存储能力，解决了核心的安全防护和大规模流量安全调度问题。资源池层的流量调度与服务编排技术解决大流量传输效率和自动化编排问题。业务中台层的原子能力统一纳管解决多厂商原子能力适配问题，以及整体的系统性能优化升级。

安全能力池的系统架构具有集中、近源、共享等特点：

1）集中。安全能力池通过安管平台对所有线上资源池集中统一纳管，提供SaaS化服务，提高运维效率。

2）近源。资源池结合边缘云技术与电信运营商的大网优势实现了近源流量安全防护，将需要防护的业务流量引导到距离防护目标最近的资源池内，经过多种安全防护能力检测、网络安全威胁清除后将流量回注给被防护对象。此种方式为用户提供更加实时、高效的防护。

3）共享。SaaS化的原子能力的共享模式实现了同一服务点下多租户共享虚拟机（VM）集群。每个VM集群部署一类服务，不同租户间通过逻辑隔离，共享模式具有成本低、效率高和资源利用最大化等优势。

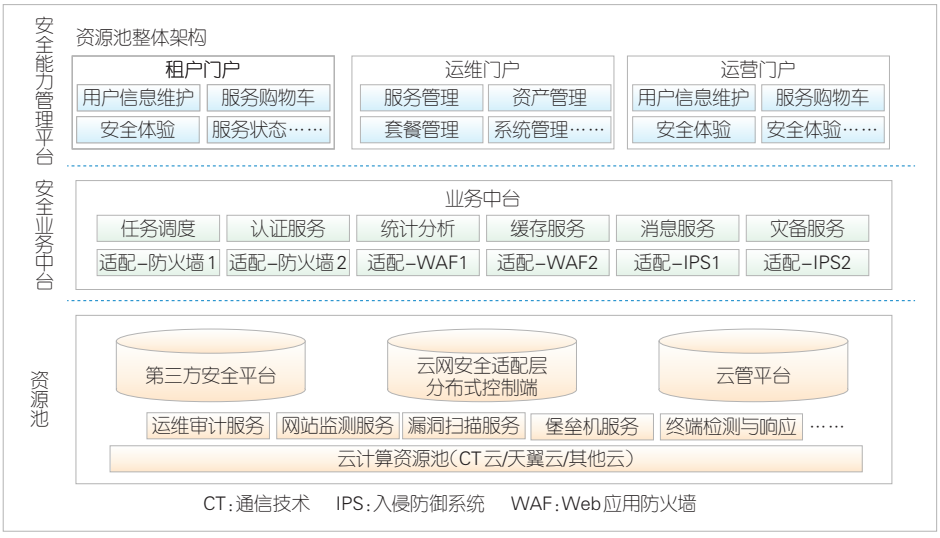
1.1 管理平台层设计

安管平台层重点解决资源池分布广、管理难度大、运维难度高等问题。作为整个系统架构的最顶层，安管平台为租户门户、运维门户、运营门户提供三位一体的管理功能，其整体结构如图2所示。

安全原子能力服务管理模块提供原子能力的统一配置、管理及安全策略下发服务，通过原子能力编排服务，实现资源池内安全原子能力的整合与联动。订单管理服务组、用户/租户管理服务组、主机管理服务组等为租户提供了良好的用户体验。告警监控服务组可以实时查看资源池健康状态、内存使用率，提升了系统整体透明度，增强了网络故障定位能力。安管平台作为资源池的集中管理入口，通过三大门户实现资源集中配置，从而降低产品集成风险，提升工作效率。管理层平台三大门户的具体功能如下：

1）租户门户系统主要为租户提供安全服务订阅功能；

2）运维门户系统采用自动一体化形式，可自动部署、启动安全原子能力服务并完成资源池内服务自动化编排；



▲图1 安全能力池整体架构图

▲图2 安全能力管理平台整体结构图

3）运营门户系统负责自定义安全服务策略、资源池运营等服务。

1.2 业务中台层设计

业务中台层重点解决两大问题：适配性问题和安全性问题。适配性问题聚焦于如何纳管各种厂商的多种安全原子能力，如何适配不同的云上资源池；安全型问题重点考虑平台直接调用底层原子能力对资源池的较高侵入性，以及未经监管的流量所带来的安全隐患。能力池系统在安全能力管理平台与底层资源池之间引入了安全业务中台层，功能如图3所示。

安全业务中台主要具备4个方面能力：1）通过制定安全组件接口规范，实现对多云底座的纳管和对多个厂家异构资源的统一纳管，并对第三方安全平台中间件、云管中间件进行统一管理；2）作为南北向通信的桥梁，将南北向解耦；3）作为业务信息、运维、运营信息的管控入口，承担着应用程序编程接口（API）统一管理的职责，通过一套标准化的接口规范，支持多种业务的水平扩展，提高了系统的可用性；4）承担整个系统架构的安全访问认证职责，实现整个系统的自主可控，提升了云网安全运营的自动化、智能化。

1.3 资源池层设计

安全能力池底层是基于轻量化边缘云技术的资源池底座实现的，其结构如图4所示。资源池封装了多厂商优势产品的安全能力，极大地提升了安全防护能力的多样性，而且安全场景覆盖性高，已覆盖10多类共计30项安全原子能力。资源池中的物理设备资源主要为安全能力提供底层的路由交换、流量调用以及数据存储。其中，安全流量调度网关具有虚拟化管理和网络流量编排功能，承担整体调度职责，分别和新型城域网控制器、安全能力池控制器对接，完成用户流量的路径编排。目前，资源池的部署方式有两种：一是采用安全能力集中部署的方法，同时在云平台核心网络设备上部署近源安全能力，主辅双线并行共同满足用户需求；二是在专线用户网络中部署近源安全能力，满足专线用户需求，充分利用运营商大网优势，将集中能力复用给专线客户，减少专线用户安全能力的部署工作。

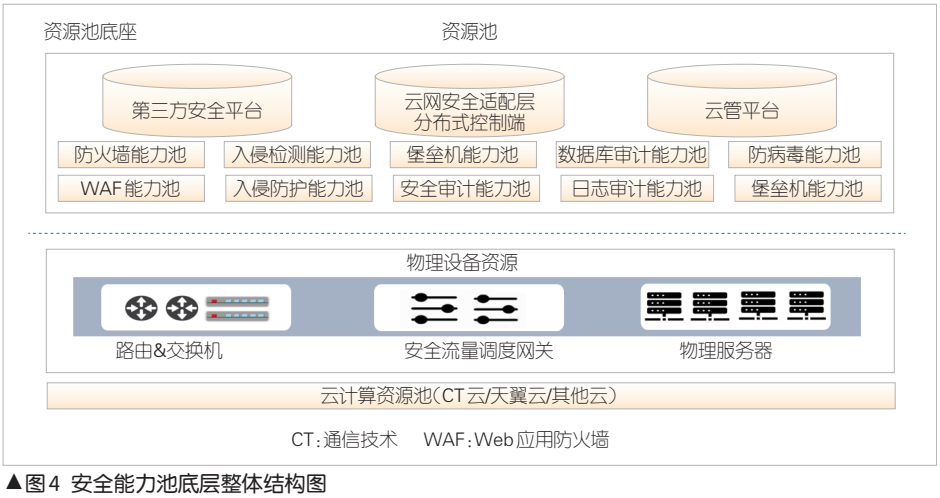
安全能力池采用统一的SaaS化架构来提供安全服务，具有按需组合、弹性扩张的特点。用户在使用安全服务时，在租户平台输入所需的安全能力和需防护的目标资产。根据运营人员的配置情况，资源池会通过Flowspec控制器将流量从IP承载网核心路由器（CR）牵引至离用户最近的资源池，在资源池内为用户流量进行安全防护，之后会将处理后的流量通过路由原址送回给用户。用户可以实时自主选择所需的安全能力。这种方式更加灵活、方便。

2 关键问题与技术

2.1 流量调度与服务编排

传统资源池使用基于策略的路由（PBR）进行流量调度。针对运营商级别的大网流量牵引调度，该方式会占用路由条目项，配置繁琐，速度慢，规模部署将难以维护。为使网络与云池有机融合和统一调度，安全能力池采用了一种新颖的流量调度与服务链编排技术。在系统需要流量调度的情况下，资源池通过自主研发的Flowspec控制器来修改路由的下一跳地址，从而将流量牵引至距用户最近的池子，进而提供安全防护。流量进入云池后，针对云资源池内的原子能力编排，安全能力池使用基于IPv6的段路由（SRv6）的端到

▲图3 安全业务中台功能图



▲图4 安全能力池底层整体结构图

端引流方案。资源池将原子能力所在虚拟机的IPv6地址定义成实例化的段标识（SID），通过操作不同的SID，实现路径规划。SRv6将128 bit的IPv6地址作为SID。如果段路由扩展报文头（SRH）封装较多的SID，则会造成SRv6报文头开销过大、传输效率低。针对该问题，资源池利用SID包头压缩技术，在保持对128 bit SID兼容的同时，删除SID的冗余信息并将其压缩为32 bit。压缩后的方案中引入了SI字段来控制32 bit SID的目的地址更新。32 bit的压缩方案在支持原有硬件设备的同时具有更高的传输效率和转发性能。基于SRv6的服务链动态编排技术效率是传统编排方式的3倍。在传递相同大小的数据块时，优化后的编排相比于原SRv6，流量传输效率能提高约30%。

安全能力池通过实验验证了SRv6端到端引流方案的有效性。在城域网某支持SRv6的多服务边缘设备（MSE）的节点上部署了1台服务器并将其作为靶机。另外，在云安全池内部署了2台服务器，每台服务器各虚拟化部署2台虚机、1台防火墙（FW）和1台入侵防御系统（IPS）。每台服务器宿主机系统采用开源虚拟化路由器（VR），并接入所有虚拟FW和IPS。VR作为池内与池外流量的锚点，也是SRv6路由调度策略的实施点。靶机与互联网的流量交互通过MSE与VR之间规划的SRv6策略实现引流，实验环境如图5所示。

实验首先对SRv6 SID进行规划，每一个SID代表一个引流行为，如表1所示。由于SRv6的流量策略在头端生效，因此将靶机的流量引入FW1时，只需头端多服务边缘（MSE）设备通过SRv6策略将流量引入靶机，并将上行流量segment list设置为A::1__C::1，下行流量segment list设置为A::1__B::1。如需将流量引入IPS1，上行segment list则设置为A::2__C::1，下行segment list设置为A::2__B::1；如需将流量依次引入FW1、IPS1，上行segment list则设置为A::1__A::2__C::1，下行segment list设置为A::2__A::1__B::1。

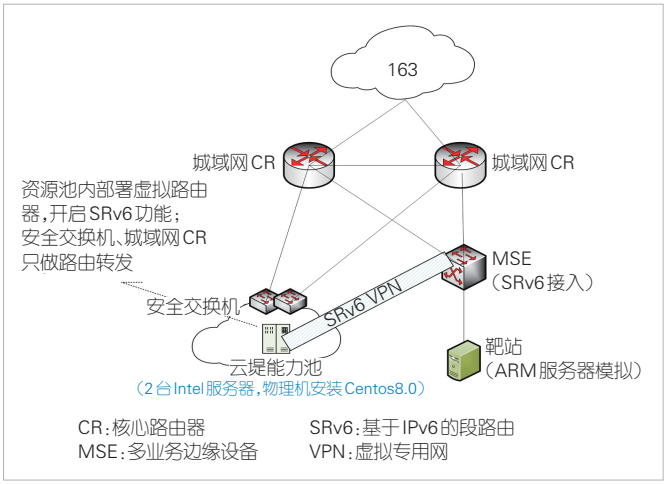
实验测试了流量编排能否按需穿过不同安全网元以及安全网元功能是否正常的情况。通过在FW1、IPS1处进行抓包，我们发现本方案可以实现基于SRv6的云网融合的流量编排。通过防火墙访问控制技术（ACL）功能测试，我们发现安全网元可以在该场景下正常工作。

2.2 系统性能优化升级

安全能力池将10余款安全能力的核心引擎、平台管理、数据分析进行解耦，融合微服务化、容器化等云原生技术，实现了从威胁监测、威胁分析到威胁处置，以及威胁审计等全流程自动化秒级联动。基于Lambda大数据技术架构，建立了统一的安全数据采集、分析、存储、查询和可视化能力，在保障平台健壮性、易于水平扩展等特性的同时，又满足查询的便利性，以及海量数据查询的低时延。

另外，针对实际的业务中断、运行速度慢等问题，本文给出以下两个典型解决方案：

1）在对资源池进行多线程压力测试时，我们发现程序运行速度远低于单线程。该问题导致系统无法满足实时性需求，用户体验下降。通过分析发现，流量型原子能力存在内存消耗大且访问具有随机性的问题。为了保证资源池系统稳定同时提升用户整体体验，方案引入大页内存技术。原始的小页内存通过页表来定位真实的物理内存空间，这使得中央处理器（CPU）在存取一个数据时，需要2次访问内存空间：第1次访问页表，然后根据页表计算出物理内存地址；第2次访问物理内存地址。为了提高地址变换速度，操作系统会在高速缓冲存储器中增设一个快表，来缓存部分经常使用的页表。这样通过访问一次高速缓冲存储器和一次内存就可以完成地址映射，实现速度的提升。但对于资源池来说，快表通常只能缓存几百页。如果页很小而程序占用内存很大，那么快表无法命中某页表的概率很大，缓存功能就失去



▲图5 引流测试环境

▼表1 SID规划

配置点	引流目标	SID	作用
VR1	FW1	A::1	将流量引入FW1
VR1	IPS1	A::2	将流量引入IPS1
VR2	FW2	A::3	将流量引入FW2
VR2	IPS2	A::4	将流量引入IPS2
MSE	靶机	B::1	将流量引入靶机
VR1 VR2	CR	C::1	将SRv6流量解封装将IPv4流量回送CR

CR：核心路由器　IPv4：互联网通信协议第4版　SRv6：基于
FW：防火墙　MSE：多服务边缘设备　IPv6的段路由
IPS：入侵防御系统　SID：段标识　VR：虚拟路由器

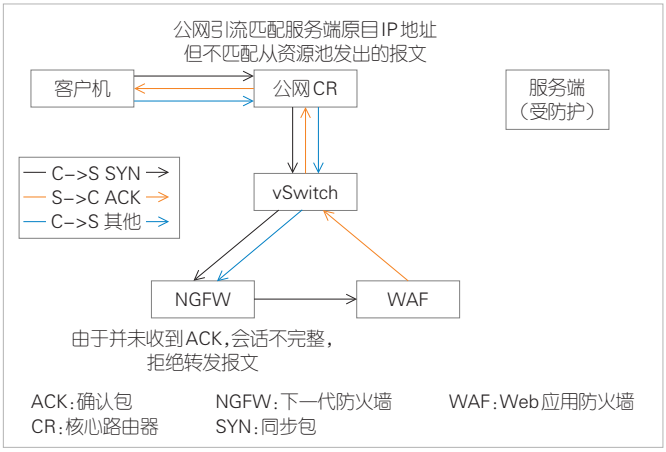
了效果。因此，资源池根据实际运行情况，调整了页的大小从而减少页表项，这使得快表尽可能完全缓存页表，从而提高程序性能。实验测得，大页内存在配置时采用对半配置原则效果较好。如果总内存为512 GB，则分配256 GB的大页内存。而当每一页大小为1 GB时，系统性能最佳。经过测试，当安全能力池配置16核32 GB内存的Web应用防火墙（WAF）时，大页技术可将程序性能提升50%左右。

2）在安全能力池多类型防护网元编排测试过程中，会出现某些业务中断的现象。例如，下一代防火墙（NGFW）和WAF进行流量编排时，出现流量经过WAF后业务中断的情况，如图6所示。客户端将与传输控制协议（TCP）3次握手的1号包同步包（SYN）引流至NGFW，NGFW再将SYN包转发给WAF（如图6中黑色箭头），WAF接收到客户端的SYN后，直接向客户端响应同步包–确认包（SYN–ACK）（TCP握手2号包，如红色箭头所示），客户端接收到SYN–ACK后，发起ACK（TCP握手3号包）。当ACK数据转发到NGFW时，此前NGFW只转发了1号包SYN，而2号包即WAF代理响应客户端的ACK未经过NGFW转发，因此当客户端发起的3号包到NGFW后，NGFW默认不放行（如蓝色箭头所示）。这导致客户端与服务器3次握手无法建立，业务异常。因此，通过优化TCP握手的数据传输验证机制，解决了此业务中断问题。

针对某些安全原子能力，例如堡垒机只能单租户独享问题，安全能力池会深入其内部结构，实现了堡垒机SaaS化。

2.3 原子能力统一纳管

安全能力池通常涵盖大量资源池底座与多种原子能力。传统资源池每引入一种新型原子能力都需要大幅度改动系统，因此存在交付效率低、系统维护困难等问题。安全能力池有上百个资源池、数千台硬件服务器，原子化后的安全能力组合方式呈指数级增长，因此如何快速纳管安全原子能力以及不同云下的资源池成为系统能否高效运行的关键。



▲图6 多网元流量编排异常问题示意图

针对以上问题，安全能力池在安全能力管理平台和底层资源池之间引入分布式高可用的安全业务中台。中台通过统一规范的接口为上层平台侧的扩展提供便利，并面向多安全厂商异构设备提供标准化接口，例如：《中国电信原子能力组件接口规范》将私有协议解耦，对南北向接口进行标准化。建立这种网络安全产品互联互通的标准，使得安全能力池能够实现跨厂商安全原子能力的统一纳管。所有厂商安全原子能力只要符合该标准，均可接入安全业务中台。统一纳管使得用户无须关注底层架构，这提升了业务的灵活性和高效性，打破了各厂商间安全设备难以互通的孤岛形态，建立了可信任的安全联动体系。标准化接口规范的推进和安全能力的适配，为云上安全可信生态的构建奠定了基础。接口规范目前已迭代至第5版。

3 安全能力池应用

3.1 典型应用场景

1）等级保护二级、三级认证应用场景

《中华人民共和国网络安全法》明确规定，网络运营者应当按照网络安全等级保护（后简称为等保）制度的要求，履行相应的安全保护义务。中国各行业大量重要信息系统已经或正在按照规定进行等保定级备案并定期接受测评，等保合规应用场景的需求不断增多。针对上述情况，安全能力池通过提供“等保二级套餐”“等保三级套餐”等产品解决方案，为各个政企用户提供安全咨询、定级、备案、建设、测评和监督检查。例如：通过安全管理中心和NGFW、端点检测响应（EDR）、日志审计、堡垒机等安全原子能力，满足用户等保二级认证的需求，并达到基础安全防护的效果；在上述基础上增加WAF、数据库审计、漏洞扫描3项原子能力可满足用户等保三级认证的需求。

2）用户增加内部安全防护能力场景

企业用户希望增强内部安全防护能力，保护内部系统正常安全运转。针对这一人群，可通过网络入侵防护、主机安全检测和综合审计这3类安全原子能力来满足用户需求。首先通过网络引流的方式将流量牵引至网络入侵防护平台，然后通过在主机上部署EDR客户端，进行漏洞扫描、基线检查以及病毒查杀，最后通过日志审计原子能力对各类设备的日志进行收集、解析和关联分析，从而达到全方位内部安全防护效果。

3.2 应用案例

目前，安全能力池已投入生产应用，为政府、金融机构和云服务提供商等政企用户提供了可定制、全面和深层次的安全防护服务，例如：某客户公司互联网出口使用简单堆叠式硬件安全防护，面临硬件设备升级困难、维护成本高等问题。另外，该客户在将本地业务系统迁移上云过程中也存在新的安全挑战。本文所提出的新型安全架构，通过安全管理平台调用了电信某省级安全能力池中的10余种能力。流量型原子能力通过Flowspec控制器将流量从客户公司网络出口牵引至资源池，流经池内安全检测和服务链自动化编排后，再次回注到该公司出口；非流量型原子能力则直接使用资源池内SaaS化能力提供安全检测或分析。最终，该安全能力池具备按需组合、流量编排等技术优势，为用户提供了全面、纵深的安全防护能力，实现其云上等保三级需求，现网安全攻击防护成功率高达99%。

4 结束语

本文提出了云网安一体化的安全能力池技术架构。该架构基于Flowspec控制器与SRv6技术实现了大网与资源池内部流量的牵引调度和服务链编排，增强了整体防护性能，提高了流量编排效率；基于多云底座、跨厂商安全能力统一纳管，实现了服务的按需组合，灵活扩张；基于大页内存等技术，实现了资源池系统的性能的优化。安全能力池为安全行业探索出一种高效、智能、稳定的云安全防护服务模式。

未来，安全能力池可在两个方向上进行升级和迭代：

1）动态授权访问与持续安全监测。当各个资源池内部网络安全边界防护逐渐模糊时，需要改变传统的边界防护模式。针对这一问题，可研究基于零信任的动态授权访问系统，确保安全能力的动态授权访问与持续安全监测，更好保证资源池内部的安全性[11-12]。

2）提出面向云网融合的新型城域网技术方案。通过SRv6、以太虚拟专用网络（EVPN）等新协议的支持能力，实现流量的自动化调度和网络的自动化配置，解决流量转发迟滞的问题。

参考文献

[1] ALLAM Z, DHUNNY Z. On big data, artificial intelligence and smart cities [J]. Cities, 2019, 89: 80-91. DOI: 10.1016/j.cities.2019.01.032

[2] TANG Y, DANANJAYAN S, HOU C, et al. A survey on the 5G network and its impact on agriculture: challenges and opportunities [J]. Computers and electronics in agriculture, 2021, 180: 105895. DOI: 10.1016/j.compag.2020.105895

[3] ARZIEVA J, NUKUSBAEV N. Network security issues and effective protection against network attacks [J]. Bulletin of science and practice, 2021, 7(9): 479-485. DOI: 10.33619/2414-2948/70/45

[4] SERROR M, HENZE M, HACK S, et al. Towards in-network security for smart homes [C]//Proceedings of the 13th International Conference on Availability, Reliability and Security. ACM, 2018: 1-8. DOI: 10.1145/3230833.3232802

[5] SUBRAMANIAN N, JEYARAJ A. Recent security challenges in cloud computing [J]. Computers & electrical engineering, 2018, 71: 28-42. DOI: 10.1016/j.compeleceng.2018.06.006

[6] HERARDIAN R. The soft underbelly of cloud security [J]. IEEE security & privacy, 2019, 17(3): 90-93

[7] 回红秀. W公司安全资源池项目风险管理研究 [D]. 北京: 北京邮电大学, 2021

[8] 吴晨花, 王瑶, 李映壮. 基于SDN安全云资源池提升中小企业安全防护能力 [J]. 科技创新导报, 2019, 16(5): 140-143. DOI: 10.16660/j.cnki.1674-098x.2019.05.140

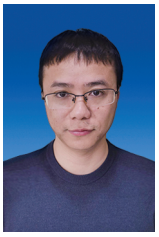
[9] 乔延臣, 张结辉, 陈晓帆. 基于安全资源池的云安全解决方案 [J]. 信息技术与标准化, 2018(9): 57-62

[10] 才宏. 云资源池网络安全策略的分析与设计研究 [J]. 网络安全技术与应用, 2021(6): 79-80. DOI: 10.3969/j.issn.1009-6833.2021.06.049

[11] WARD R, BEYER B. Beyondcorp: a new approach to enterprise security [J]. USENIX & SAGE, 2014, 39(6): 6-11

[12] BUCK C, OLENBERGER C, .SCHWEIZER A, et al. Never trust, always verify: a multivocal literature review on current knowledge and research gaps of zero-trust [J]. Computers & security, 2021, 110: 102436. DOI: 10.1016/j.cose.2021.102436

作 者 简 介



余启明，天翼安全科技有限公司开发工程师；负责安全能力池相关技术架构和开发工作，研究方向为安全服务SaaS化；已发表论文1篇。

吴爽，天翼安全科技有限公司研发工程师；现从事网络安全产品研发；已发表论文1篇，申请专利4项。

黄帅，天翼安全科技有限公司研发工程师；现从事网络安全产品研发和强化学习等；已发表论文4篇，拥有软著2项。

刘紫千，天翼安全科技有限公司总经理，正高级工程师；主要从事网络安全技术研究和安全产品研发运营工作；获得省部级科技进步奖4次；已发表论文10余篇，获发明专利10余项。

未来网络内生安全通信技术

Intrinsic Security Technology for Future Network

闫新成/YAN Xincheng[1,2]，周娜/ZHOU Na[1,2]，
蒋志红/JIANG Zhihong[1,2]
（1. 移动网络和移动多媒体技术国家重点实验室，中国 深圳 518055；
2. 中兴通讯股份有限公司，中国 深圳 518057）
(1. State Key Laboratory of Mobile Network and Mobile Multimedia, Shenzhen 518055, China;
2. ZTE Corporation, Shenzhen 518057, China)

DOI：10.12142/ZTETJ.202301006
网络出版地址：https://kns.cnki.net/kcms/detail//34.1228.tn.20230227.0939.002.html
网络出版日期：2023-02-28
收稿日期：2022-12-23

摘要：网络安全技术是保证未来云网发展的关键技术。针对未来网络安全挑战及现有技术缺陷，提出了一种基于网络可信身份的轻量化密钥验证技术——网络可信通信（NISC）技术。该技术具备近源协同防护和无状态随路验证等特征，目前已在试验网络上验证了其防御网络攻击的有效性和可行性，可以为未来网络安全可信保障提供参考，加速未来云网安全技术研究和产业化进程。

关键词：未来网络；NISC；内生安全；主动防御

Abstract: Network security is a key technology to ensure the future cloud network. To address future network security challenges and the shortcomings of existing technologies, a lightweight key authentication mechanism based on network trusted identity – Network Trusted Communication (NISC) technology is proposed. The technology features such as near-source cooperative protection and stateless follow-the-road authentication. The effectiveness and feasibility of its defense against network attacks is verified on China Environment for Network Innovation (CENI) network, which can provide reference practice for future network security trustworthiness assurance and accelerate the research and industrialization process of future cloud network security technology.

Keywords: future network; NISC; intrinsic security; active defense

1 未来网络的安全挑战

网络的演进是一个开放性和动态性不断增加的过程。5G/5G-A面向医疗、交通、工业等领域，促进通信技术（CT）与信息技术（IT）/运营技术（OT）的融合；6G网络与算力融合，尝试对网络服务进行感知，实现泛在的接入和服务访问。未来的网络除了要提供更快的传输速率、更精准的服务、更智能的连接外，还需要提供更安全可信的传输能力。这就不仅需要安全适应网络，解决由开放性和灵活性所造成的安全问题[1]，还要使网络具备内生的安全能力，基于网络的新特性，更好地发挥网络的安全潜能。

1.1 安全需求

随着网络的演进和发展，融合体系、通信模式和防护主体都发生了变化，这也促使了网络安全架构的发展。

首先，新的网络融合体系对传统互联网协议（IP）安全体系提出了挑战。数字经济发展需要云边端协同的强大算力和广泛覆盖的网络连接做支撑，算网融合已成为重要趋势。算网融合衍生出新的网络结构，但角色多样泛在化、连接多变动态化、信任关系多元复杂化等特点为攻击提供了更多的条件，这会严重加剧攻击程度，因此需要人们重新审视算力网络的安全防护架构与能力[2]。由于IP缺乏安全设计，未来网络需要从架构上解决IP安全问题[3]。

其次，新的网络通信模式对以网络服务为主体的传统防护模型提出了挑战。园区生产网络采用工业总线技术，经IP化改造后，将普遍采用L2/L3层点对点（D2D）的通信模式[4-5]。由于传统的OT网络通信协议缺乏严格的权限管理和验证机制，IP化改造后基于静态配置的网络访问控制策略难以奏效，攻击者可能会假冒合法用户身份进行越权访问[6]，并利用系统漏洞发起网络攻击。现有的IT安全以保护客户端-服务端（C-S）通信模式为主，难以解决OT局域网IP化后访问不受控的问题。

最后，防护主体的变化对传统孤立的防护模式提出了挑战。协同制造使移动通信网络与先进制造技术深度融合：园区内OT设备通过多种方式混合接入企业生产制造网络，园区外不同企业间通过广域网动态构建专网进行协同生产。园区内的局域应用向广域化转变，网络风险由消费领域向产业领域持续渗透，这需要进行多信任主体间的协同防护。行业终端和边缘节点因安全能力不足，也需要端到端地设计安全

方案，提供多点多域的协同防护机制。由于网络广域互联、攻击各个层面容易扩散，海量异构节点存在安全能力差异，针对行业应用的恶意攻击也将不断增加[5]。

在网络架构融合开放的发展趋势下，传统外挂式、补丁式的被动安全防御机制已无法有效支撑未来网络安全性需求，因此需要基于网络“内生安全”的理念去解决网络安全问题[7-9]。目前业界对网络内生安全的研究主要包括网络通信的可信和网络基础设施的可信两方面。本文中的研究主要聚焦于前者，简称为可信通信，即将安全功能作为基本要素耦合到体系结构中，在不借助外力（安全软件、防火墙等）的情况下，实现对网络通信的攻击防范和内生安全保障[15]。

当前业界非常重视网络内生安全技术研究[10]：科技部专项研究《1.3 内生安全支撑的新型网络体系结构与关键技术》涵盖了网络体系结构内生安全机理、未知网络攻击免疫方法等内容；产业界也在近年来开展了IP网络内生安全的研究，网络5.0产业联盟在《网络5.0技术白皮书》[11]中提出网络需要具备可信管理、可信接入以及可信路由能力等可信通信能力，并强调了抗网络攻击的网络内生安全能力需求；全球标准组织积极研究和制定攻击防御技术相关安全标准；运营商一致认同网络安全可信的重要性，并针对攻击问题提出内生安全能力需求。

1.2 网络攻击分析

网络安全架构设计的目的是建立一个安全的网络环境，保护网络系统免受攻击。要达成网络可信通信的目标，需要先分析网络攻击。传统的IP网络体系设计以设备间通信互联为导向，难以在网络层实现针对网络攻击的检测控制。未来网的典型攻击的特征和原理分析如表1所示。

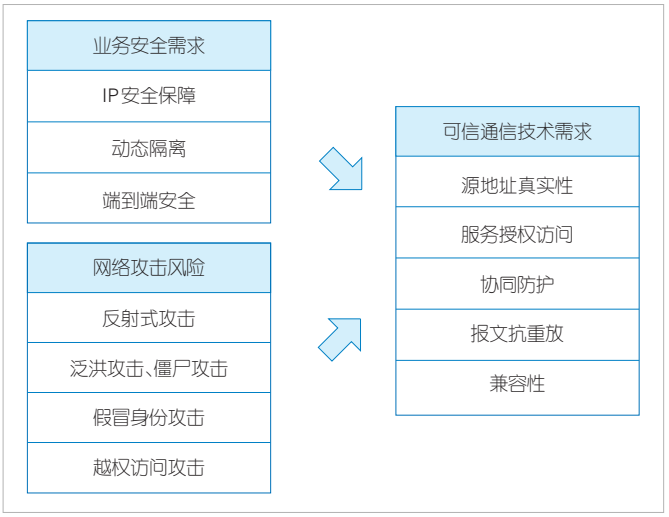
基于网络业务安全需求和所面临的攻击风险的综合分析，我们提出了未来网络可信通信技术需求，如图1所示。为及时、高效、系统地实现可信通信，需要网络尽早检查终端源地址的真实性，并有效验证终端用户访问应用的合法性，确保只有真实终端得到合法授权才能被允许访问行业应用。与此同时，还应考虑跨域场景下的协同防护机制，针对报文重放攻击的防范机制以及对传统终端和网络的兼容性方案，全方位、多角度增强系统安全可信通信能力。

2 可信通信的设计原则与技术体系

网络可信通信（NISC）体系是专门针对未来网络的关键安全需求及现有技术的缺陷而构建的。基于可信的网络架构和IP协议，NISC应具备近源协同防护、无状态随路验证等特征，无须依赖攻击先验知识，能及时、主动识别控制异常的通信数据流，减轻攻击造成的系统危害。在设计可信通信机制时，为了提供体系化的可信通信能力，应考虑具有以下特性的安全防御机制：

1）实时性：借助网络层为业务提供接入域、传输域、目的域等多点防范，尽可能实现自动化接入和靠近攻击源的防护。

2）高效性：打造高效防御阻断系统，减少业务冲击，为未来网络赋予高性价比、精准化攻击检测能力。



▲图1 未来网络可信通信技术需求

▼表1 未来网络典型攻击分析

攻击类型	典型场景	攻击特征	网络固有缺陷
反射式攻击	服务感知网络、协同制造网络	攻击者修改报文中的源地址，利用收发主机或报文数量的放大效应进行攻击	IP通信中报文发送或转发时没有检查源地址的真假
泛洪或僵尸网络攻击	服务感知网络、协同制造网络	攻击者不改源地址，借助网络分布式的特点在多点同时发送正常报文，这会造成受害者链路拥塞或处理能力不足	源主机发送报文时，无须获得接收端的允许，就可随意发送
假冒身份攻击	园区生产网络、协同制造网络	攻击者通过修改报文中的源地址，发起攻击或假冒合法用户身份	IP通信中报文发送或转发时没有检查源地址的真假
越权访问攻击	园区生产网络、协同制造网络	系统检查业务授权时存在漏洞，导致攻击者可绕过该权限检查，访问或操作原本无权访问的高权限功能	利用网络无法感知业务授权，无法进行应用层访问控制

3）系统性：面向全系统构建攻击防护方案，设计系统性的信任链传递机制，多维度、多层次、多位置提供服务节点可信可控防护机制。

基于上述设计原则，NISC技术具体应满足下列要求：

1）源地址真实性要求。针对因IP地址假冒所引发的攻击问题，需要对通信发起端的身份或标识，通过轻量化访问控制技术、密码算法技术进行真实性校验。这样能够增强网络业务通信的可信度，弥补端到端安全访问能力方面存在的不足。

2）服务授权访问要求。在地址真实性得以保障的基础上，端到端通信业务应根据业务认证与访问授权情况，借助密码学机制识别合法报文并实施服务可访问控制，确保业务获取目的端的认证和授权，从而阻止网络攻击行为的发生。

3）协同防护要求。对于通信业务经过不同信任主体的情况，可以基于密钥派生、认证信息共享等机制来实现安全域间的互信传输，并尽量在靠近报文发起源、目标域检测数据报文的合法性，保障多信任主体场景的高效攻击阻断、端到端系统性防御。

4）报文抗重放要求。系统应基于时间校验子、序列号等动态因子对重放报文进行轻量化主动识别和检查，有效避免因非法截获报文引起的重放攻击。

5）兼容性要求。可信通信技术应对传统终端、网络以及传输设备提供兼容性支持。

图2为NISC可信通信体系架构，包括控制面和转发面，分别实现网络可信身份和基于凭证的可信转发。控制面负责用户设备认证、业务授权、网络可信身份生成、可信凭证分配等管理功能，并作为通信系统信任锚点，为转发面源端身份可信检验及业务可访问控制等提供验证依据。转发面基于控制面传递的可信凭证，负责全系统通信过程中的数据随路识别、验证和控制。通过控制面和转发面的信任关联，NISC为未来网络系统安全可信通信提供了系统性保障。

1）控制面包含目标域认证服务器、授权服务器、接入认证服务器等网元：

a）目标域认证服务器由企业或目的应用方部署，对用户设备进行认证并生成验证信息，实现自动化可信企业接入认证，为用户设备安全接入企业应用提供保障。

b）授权服务器也由企业或目的应用方部署，基于细粒度的服务防控策略对成功认证的用户进行服务访问授权，并控制授权有效期，避免因永久授权造成的安全攻击隐患。

c）接入认证服务器由接入网络运营商部署，对用户设备接入网络进行认证，提供源地址真实性验证和抗重放验证依据。

2）转发面除了用户设备、应用以外，还包含接入网关、服务网关和路由器等传输节点：

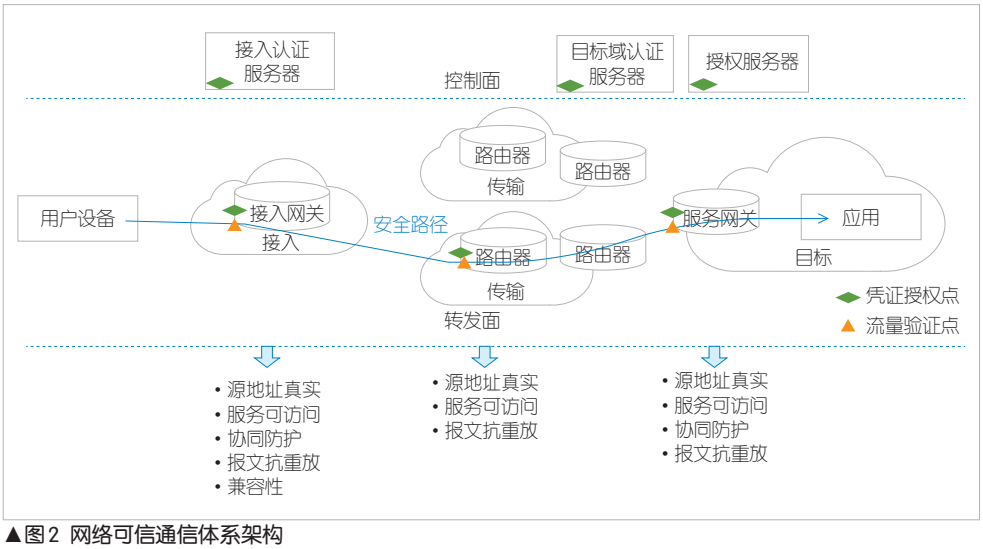
a）接入网关所涉及的功能：用户设备的地址真实性验证；按照运营商或企业需求，针对业务的可访问性执行授权、验证和控制功能；提供通信流量的抗重放验证，从近源端有效避免攻击者因非法截获或恶意构造报文所导致的重放威胁。

b）服务网关控制用户对企业的访问行为：识别并控制服务授权请求报文；按照运营商或企业需求，针对业务的可访问性执行授权、验证和控制功能。

c）路由器针对业务的可访问性实现授权、验证和控制功能：该项功能无须通信路径中所有路由器参与，需要根据业务需求、运营商需求以及网络能力在域边界路由器上部署并启动。

针对跨域通信的情况，除了在接入网关和服务网关执行数据流验证功能以外，中间域也可在其边界路由器部署、启用数据流验证功能，对域间发生的攻击场景形成安全屏障，从而减轻目的端网关处理负担。

面向业务的网络层可信通信通过终端业务身份认证、服务授权、服务访问控制机制全方位构建目标域可访问能力，利用验证信息的统一化生成、动态更新管理以及轻量化抗重



▲图2 网络可信通信体系架构

放机制，为业务访问提供高效的合法性验证依据，增强网络自身抵御攻击的能力。同时，在网络层面实现近源以及近目的的多点攻击防范，及时阻止无合法访问权限的真实地址用户非法访问业务行为，达成流量实时高效检测控制的安全防护系统。

3 可信通信关键技术

网络可信通信关键技术包括源地址真实性检查、服务动态授权机制、跨域协同防护、重放攻击主动检测、终端兼容性，如图3所示。

3.1 源地址真实性检查

传统IP网络缺乏基本的安全性设计，因此仿冒源地址引发的攻击层出不穷，而现有的可信通信技术验证开销大，保护机制不够健全，难以满足多样化应用、海量终端的泛在网络安全需求，因而需要考虑如何系统性地构建高效的业务真实源验证安全机制[12]。

作为信息隐私保护的一种典型技术，基于对称密钥的验证机制具有统一信任锚点，可以利用控制面集中生成或多方协商、派生出具有私密属性的共享密钥，具体如图4所示。统一信任锚点借助该共享密钥对需要验证的信息进行密码学运算，生成相关的通信验证凭证，再下发给报文发起端。发起端在每报文中携带通信验证凭证。收到业务报文之后，转发面验证节点基于事先获取的共享密钥和报文中待验证信息，利用与信任锚点一致的密码学算法生成通信验证凭证，以此对报文合法性进行识别和区分。相对于非对称密钥，对称密钥运算性能更好，可提供更高效的验证和控制机制。

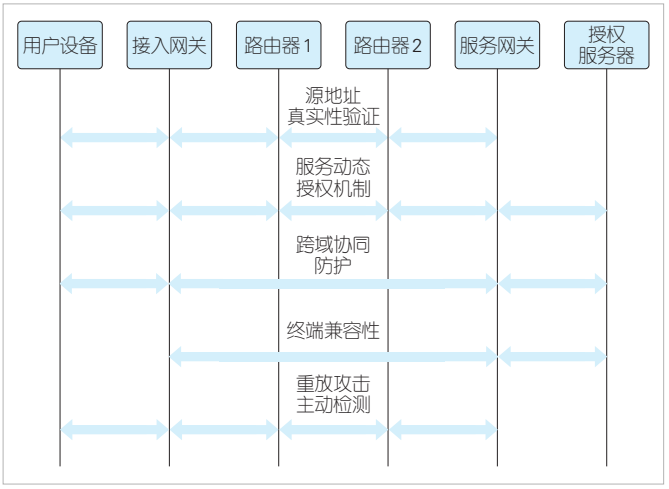
在用户设备接入网络执行认证的过程中，接入认证服务器基于该设备的标识信息（ID）对用户进行认证，采用散列消息认证码（HMAC）算法，并根据共享密钥生成验证码，之后再返回给用户设备。经过地址可信分配后，用户设备获取含有设备标识的地址信息。在业务通信过程中，每数据报文将携带验证码信息，然后由接入网关根据事先获得的共享密钥以及报文中的标识和验证码信息执行用户设备的源地址校验。

基于业务源IP的真实性控制机制提供了面向用户可信身份的认证增强、基于密码学的接入网验证技术，避免因地址假冒引发的网络攻击、信息非法获取等异常操作，实现用户接入云网系统时的轻量化实时验证和网络安全可信传输。

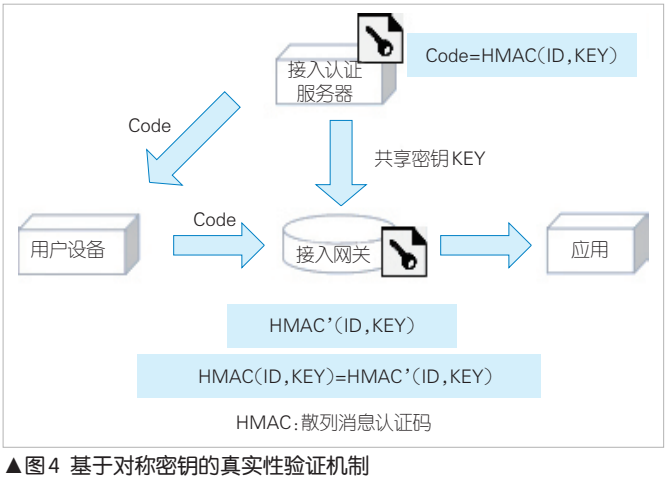
3.2 服务动态授权机制

为了及时阻止未经许可的流量恶意注入，避免非法访问对应用系统带来的不利影响，在充分保证终端用户源地址真实的同时，目标应用系统的可访问性也应得到关注。端到端通信业务中如何确保终端用户获取目的端访问授权、如何高效识别数据报文的合法与否等问题均值得深入研究[13]。

服务动态授权全方位构建了目标域鉴权、授权和网络验证机制。如图5所示，服务端基于终端的信任度、应用系统资源占用情况、应用访问控制列表等策略，对合法访问应用的终端进行动态授权，并在数据转发层面由网络节点对其业务流量进行合法性验证，及时阻止无合法访问权限的真实地址用户非法访问业务的恶意行为，实现对应用系统的有效防护。服务动态授权利用验证信息的统一化生成、一致性表达、动态更新管理机制，为用户访问业务提供高效的合法性验证依据，构建流量实时高效检测控制的安全防护系统。该机制兼具动态授权和无状态过滤的优点，既能主动防范针对主机身份的网络攻击，又能有效抵御反射性攻击、泛洪攻击和中间人攻击等各种分布式拒绝服务（DDoS）攻击，增强了系统主动抵御攻击的可信通信能力。



▲图3 网络可信通信的关键技术



▲图4 基于对称密钥的真实性验证机制

3.3 跨域协同防护

在传统防护模式中，当终端用户访问企业应用系统时，运营商网络对用户进行接入认证，并作为管道承载用户与应用间的业务认证。当运营商网络与企业应用系统处于不同信任域时，用户和接入网络、用户和应用系统分别建立信任关系，独立进行认证、授权和验证。这种基于二元信任的攻击防御模式[14]传输开销大，验证效率低。这种攻击防御架构无法尽早在接入网络对针对应用系统的攻击进行近源检测和防护，因此防护效果滞后，给应用侧的防御系统带来较大压力。

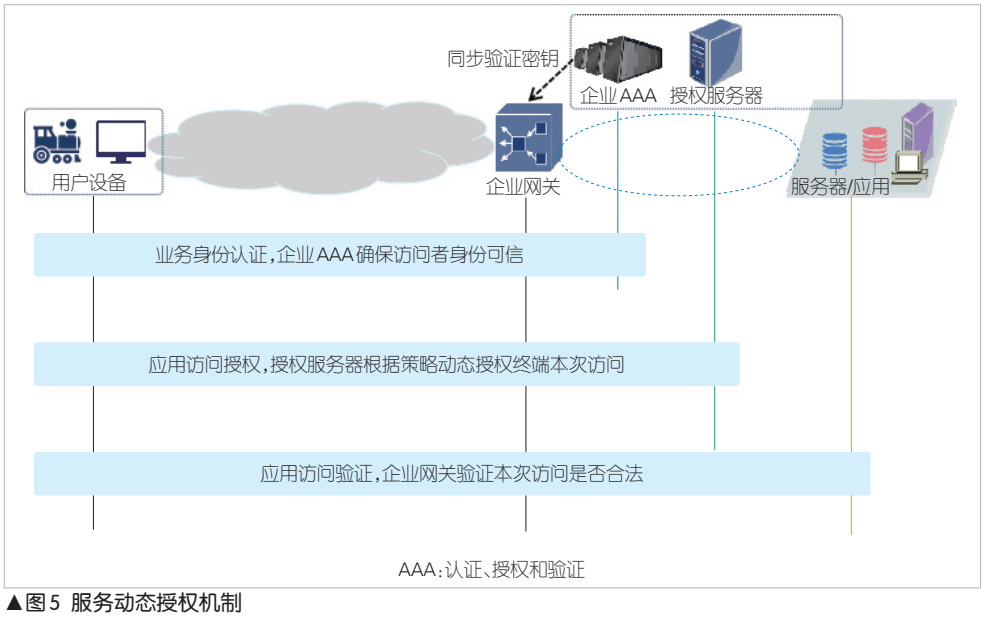
跨域协同防护机制的工作原理如图6所示。该机制通过不同信任域间的信任协商，以运营商网络为锚点派生出企业的验证密钥。以此为基础，企业网络在用户业务认证时，派生出用户的企业应用会话密钥。终端用户基于终端身份、企业身份及企业会话密钥生成验证码，并在运营商网络、企业网络中对用户访问企业应用进行多点随路验证，以从近源端抵御对企业应用的攻击。其中，密钥派生和验证码生成机制如图7所示。

协同防护方法通过在不同信任域间建立信任协同机制，共享认证结果，协商信任凭证，构建基于用户、网络和应用的跨信任域协同防护和无状态随路验证机制。一套凭证即可验证用户身份与访问合法性，实现多点验证、近源防护，提高了防护系统的实时性、高效性与系统性。

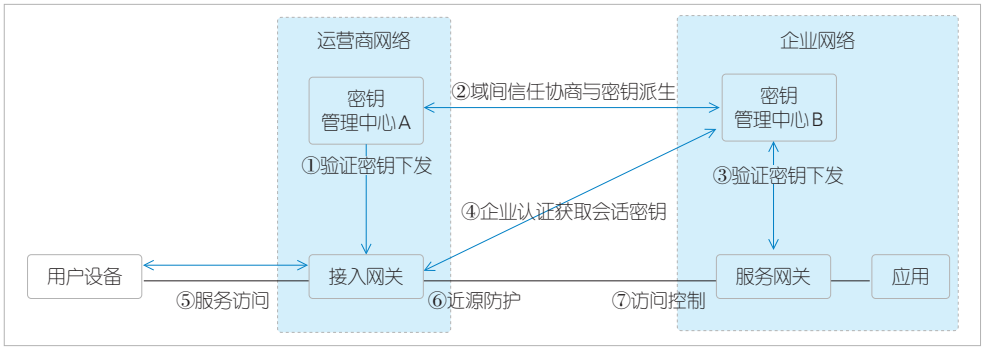
3.4 重放攻击主动检测

在可验证信息中添加时间校验子、序列号等动态信息，便于接入网关在用户设备发起业务流程时，及时基于动态信息对数据包进行检查与控制，有效抵御报文重放攻击。

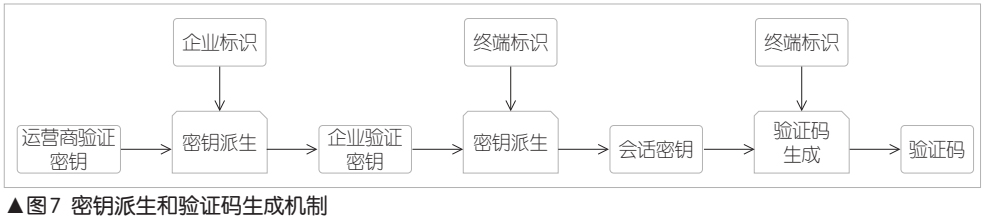
在如图8所示的重放攻击主动检测机制中，用户设备向接入认证服务器发起接入认证请求。认证成功后，接入认证服务器为用户设备分配终端密钥和系统时间，接入网关记录该系统时间，并结合网关本地时间计算、保存时间差。用户设备根据系统时间和设备本地时间计算时间差。当发送报文时，用户设备先根据时间差和设备本地时间生成时间校验子，再根据获取的终端密钥、时间校验子、序列号等生成校验码，最后发送数据包，包括校验码、时间校验子和序列号等信息。接入网关接收到用户设备发送的数据包后，根据保存的时间差和网关本地时间来确定系统时间，然后检验数据包中动态信息与校验码，以此识



▲图5 服务动态授权机制



▲图6 跨域协同防护机制



▲图7 密钥派生和验证码生成机制

别、控制因重放攻击引发的非法报文。

3.5 终端兼容性

未来网络中海量终端、产业弱终端均为攻击者提供了更多的攻击条件，严重加剧攻击程度，因此需要考虑终端兼容的可信通信机制，以减弱对性能差、安全能力不足或者传统终端所造成的影响。在终端无须感知的情况下，可考虑由接入网关代替终端完成可信通信相关安全功能：一方面需验证报文所经接入网关的真实可信；另一方面，对终端是否可访问服务进行控制。

当终端经传输设备发出报文时，接入网关应确定待转发数据报文的类型。如果是服务授权请求报文，则使用接入网关的密钥对接入网关的信息进行密码学计算并生成验证值，这便于后续的传输设备对报文的源身份进行验证。在保障身份真实性的基础上，后续设备为该报文生成预授权，并转发该报文。给用户设备返回的服务授权响应报文应携带授权检验信息。接入网关代替用户设备存储相关的授权检验信息。如果传输设备确定待转发的数据报文类型是服务请求报文，则对该报文添加存储的授权检验信息。接入网关之后的传输设备对数据报文携带的授权检验信息进行验证。若验证通过，则转发该数据报文。

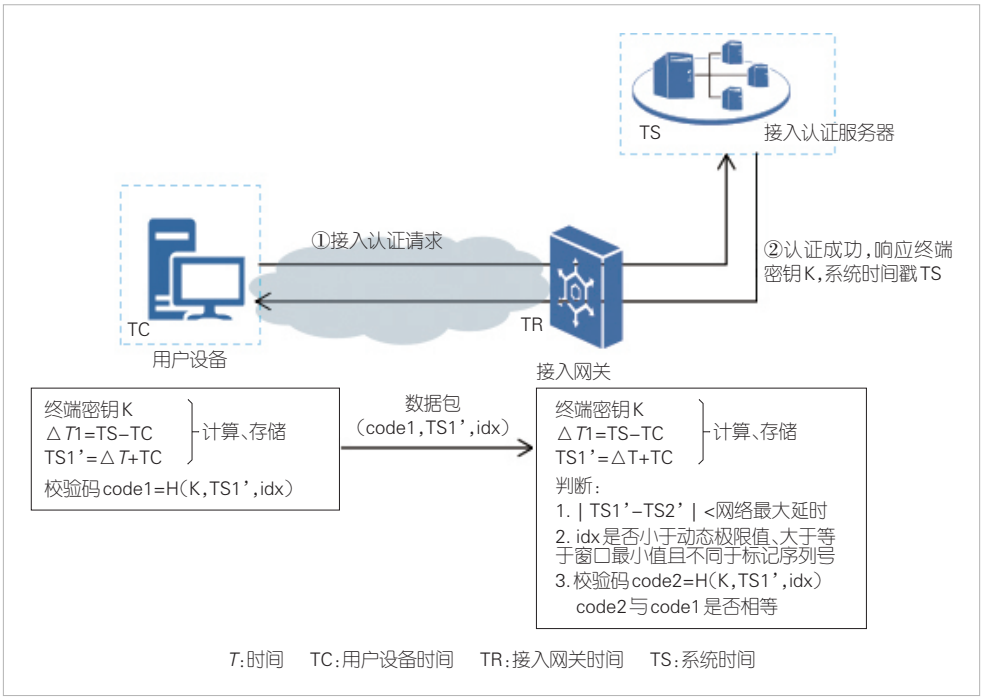
4 技术应用实例

基于NISC体系，中兴通讯进行了原型研制，并在中国信息通信研究院的协助下，在未来网络试验设施（CENI）深圳分系统开展了源地址真实性和服务动态授权技术跨域试验。

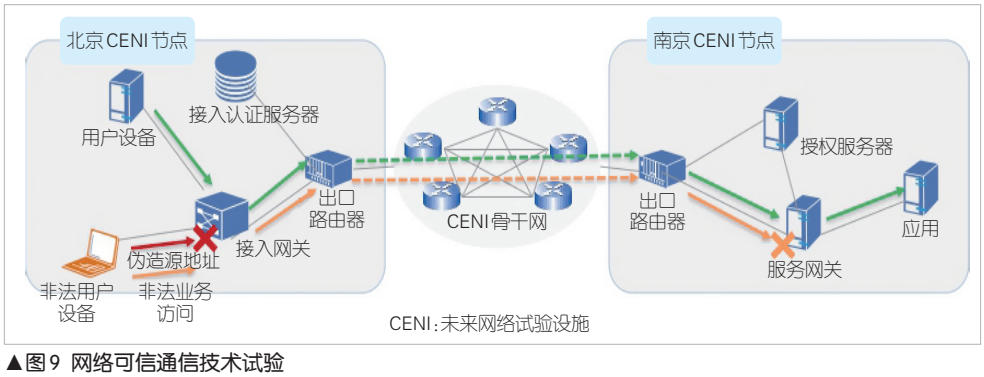
如图9所示，在NISC技术试验中，接入网关为用户设备提供源地址真实性检查功能，服务网关和授权服务器提供业务动态授权防御功能。整个系统利用认证服务器进行基于用户设备标识的认证，实现自主式便捷身份可信认证；支持基于对称密码学的身份可信机制，实现增强型高效地址真实性验证，以及面向应用业务的可信通信功能。经验证，本防御体系可实时检测并防范未认证用户，非法地址终端导致的泛洪、反射、中间人、越权访问等网络攻击，有效增强内生的系统性可信通信能力，实现未来云网系统的高效可信。通过对现有路由器和交换机等网络设备进行软件升级，可以有效防御绝大多数典型网络攻击，避免因额外部署流量清洗系统或防火墙等专用安全设备所带来的网络复杂、流量迂回和资本性支出（CAPEX）、运营成本（OPEX）增加等问题，可以在简化网络部署和运营的前提下实现网络安全性能的提升。

5 结束语

网络安全已成为社会发展、国家安全的基础需求。随着未来网络的不断发展，基于先验知识的被动防御模式已无法



▲图8 重放攻击主动检测机制



▲图9 网络可信通信技术试验

满足新型信任关系下的安全需求。因此，本文分析了未来网络面临的安全挑战，探讨了网络内生安全的技术要求和设计原则，提出了NISC技术。该技术具备近源协同防护、无状态随路验证等特征。未来，我们将继续探索分布式服务授权和精细化防控方案，促进可信通信技术在实际应用中的发展，如服务感知网络、园区生产网络、协同制造网络等。

致谢

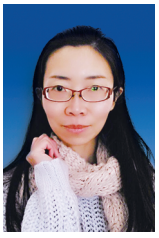
本研究得到中兴通讯股份有限公司谭斌、罗鉴、周继华、宋琳、武天元等专家的帮助，在此表示感谢！

参考文献

[1] 中国移动研究院. 5G-Advanced安全技术演进白皮书 [R]. 2022
[2] 中国移动. 算力网络安全白皮书 [R]. 2022
[3] CLARK D D, PARTRIDGE C, RAMMING C J, et al. A knowledge plane for the Internet [J]. Computer communication review, 2003, 33(4): 3-10. DOI: 10.1145/863955.863957
[4] SIST. Industrial communication network-network and system security part3-3: system security requirements and security assurance levels: IEC 62443-3-3 [S]. 2013
[5] MANOJ R, TRIPTI C. An effective approach to detect DDos attack [M]. Advances in Computing and Information Technology. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013: 339-345. DOI: 10.1007/978-3-642-31600-5_33
[6] IETF. Source address validation Improvement (SAVI) threat scope: RFC 6959 [S]. 2013
[7] 徐恪, 朱亮, 朱敏. 互联网地址安全体系与关键技术 [J]. 软件学报, 2014, 25(1): 78-97. DOI: 10.13328/j.cnki.jos.004509
[8] 闫新成, 周娜, 蒋志红. 未来网络可信通信技术 [J]. 中兴通讯技术, 2021, 27(5): 52-59. DOI: 10.12142/ZTETJ.202105011
[9] 中兴通讯. IP网络未来演进技术白皮书 [R]. 2021
[10] WU J P, WU Q, XU K. Research and exploration of next-generation Internet architecture [J]. Chinese journal of computers, 2009, 31(9): 1536-1548. DOI: 10.3724/sp.j.1016.2008.01536
[11] 网络5.0产业和技术创新联盟. 网络5.0技术白皮书 [R]. 2021
[12] AAZHANG B, AHOKANGAS P, ALVES H, et al. Key drivers and research challenges for 6G ubiquitous wireless intelligence [EB/OL]. [2023-01-05]. https://www. researchgate. net/publication/336000008_Key_drivers_and_research_challenges_for_6G_ubiquitous_wireless_intelligence_white_paper
[13] YANG X W, WETHERALL D, ANDERSON T. TVA: a DoS-limiting network architecture [J]. ACM transactions on networking, 2008, 16(6): 1267-1280. DOI: 10.1109/tnet.2007.914506
[14] 闫新成, 毛玉欣, 赵红勋. 5G典型应用场景安全需求及安全防护对策 [J]. 中兴通讯技术, 2019, 25(4): 6-13. DOI: 10.12142/ZTETJ.201904002
[15] 徐恪, 冯学伟, 李琦, 等. 安全可信的互联网体系结构与端到端传送关键技术 [J]. 中兴通讯技术, 2022, 28(6): 17-22. DOI: 10.12142/ZTETJ.202206004

作 者 简 介

闫新成，中兴通讯股份有限公司网络安全首席系统架构专家，正高级工程师，江苏省“333高层次人才”；曾主持或参与国家科技重大专项课题，获多项省部级科技奖励；拥有专利40余项。



周娜，中兴通讯股份有限公司技术预研系统工程师；主要负责网络安全、无线通信安全和未来网络安全等技术研究工作。

蒋志红，中兴通讯股份有限公司技术预研系统工程师；主要负责IP网络安全、移动通信安全和未来网络安全等技术研究工作。

云平台DNS安全体系研究

Security Framework for Cloud DNS

宋林健/SONG Linjian，马永/MA Yong，梁卓/LIANG Zhuo
（阿里云计算有限公司，中国 北京 100102）
(Alibaba Cloud Computing Co. Ltd., Beijing 100102, China)

DOI：10.12142/ZTETJ.202301007
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230301.1441.002.html
网络出版日期：2023-02-28
收稿日期：2022-12-05

摘要：域名系统（DNS）是互联网的基础设施服务，对数字经济发展的安全和稳定至关重要。结合阿里云DNS的安全实践，提出了适应融合云时代发展的DNS安全体系，包括全链路和融合云DNS的应用场景，DNS南北向、东西向服务接口能力，以及在数据保密性、数据一致性、服务高可用、软件质量、安全运维、服务测量等多个维度的安全能力框架。

关键词：DNS；安全体系；互联网基础设施

Abstract: The domain name system (DNS) is the infrastructure service of the Internet and is crucial to the continuous security and stability of the digital economy. Based on the security practice of Alibaba Cloud DNS, a DNS security framework suitable for the development of a converged cloud era is proposed, including the application scenarios of the all-link and converged cloud DNS, the south-north and east-west service interface capabilities of DNS, as well as a security capability framework in multiple dimensions such as confidentiality, Integrity, availability, software quality, operation, and service measurement.

Keywords: DNS; security framework; Internet infrastructure

互联网域名系统（DNS）提供了互联网域名和互联网协议（IP）地址两种网络标识符体系之间的衔接转换。随着新技术新场景新模式的涌现，企业信息架构在持续升级，数字化建设朝着云管边端一体化演进，万物互联格局已然显现。有IP的地方就有DNS寻址。寻址的形态渗透在云管边端各个场景中，从以南北流量为主的互联网通用互联场景到以东西流量为主的企业机房场景，再到云技术服务的云内IP寻地和多云的云间寻址。作为互联网的中枢神经，DNS在网络安全和企业数字化治理体系中扮演至关重要的角色，例如：2021年6月，美国政府要求域名注册局对36个伊朗媒体域名进行“查封”[1]，引起了国际社会的关注。2021年10月，Facebook DNS服务不可用导致其旗下很多应用发生了故障，持续了6个多小时[2]。“十四五”是中国推进信息通信行业高质量发展、建设网络强国和数字中国的关键时期，DNS作为核心网络基础设施的重要地位也正在得到业界愈发广泛的认可。

随着互联网发展和技术演进，DNS技术和产品形态不断丰富，机遇与挑战并存。一方面，通过安全扩展协议和新技术的引入，DNS不断增强安全能力。互联网工程任务组（IETF）不断发布新的DNS安全扩展协议，从底层协议标准层面完善DNS安全，如DNS Cookie、DNS安全扩展（DNSSEC）。加密传输技术开始广泛应用在DNS领域，增强了数据一致性和隐私保护。进入现代的互联网时代，新型的移动互联网服务模式为DNS提供了新的服务架构，大型云计算平台为DNS服务提供了全链路自研的更可控的服务、更高的弹性、更高的可用能力，以及更及时的软件和服务漏洞更新。新技术、新模式提高了DNS抗攻击的安全加固能力。

另一方面，进入云计算时代，云平台DNS的服务架构和形态正在发生变化，以适应复杂的多应用场景的互联互通和新业务形态的规模化增长。尤其是在多云异构的融合场景下，DNS成为部署在公有云、私有云、本地互联网数据中心（IDC）、应用和智能终端等多场景的IP地址寻址和统一调度平台服务。这对DNS软件质量、安全运维和体系化服务能力提出了新的挑战。

1 DNS的演进和各阶段特征

1.1 网络协议和分布式IP数据库

传输控制协议（TCP）/IP被发明并普及后，互联网规模迅速扩展。基于Host.txt集中式的名字解析已无法满足日益扩大的网络规模和主机名字解析的需要。20世纪80年代初，为了解决名字解析服务的扩展性问题，DNS的基本概念和实现框架（RFC882/883）被提出。DNS引入了树状的域名空间，按照分层的域名结构划分管理域，数据和管理权限的下

放实现了分级的分布式结构。该时期，DNS支撑了TCP/IP初期的互联网商业化。

在该阶段，作为网络基础组件和协议，DNS提供了分布式“查询-响应”的IP查找。域名拥有者或网络管理员在自己的网络中独立部署和运行DNS解析服务，并依赖开源的DNS软件，如BIND（软件名）。该阶段，DNS安全能力依赖于DNS协议的安全和开源DNS软件的质量，任何DNS协议漏洞或开源DNS软件的漏洞都会影响DNS服务。所以该时期大部分DNS安全的讨论集中在DNS协议标准层面，例如DNS安全扩展（DNSSEC）。但DNSSEC安全扩展协议未能快速进行全球部署，一些例如DNS劫持之类的安全风险到今天仍然普遍存在[5]。

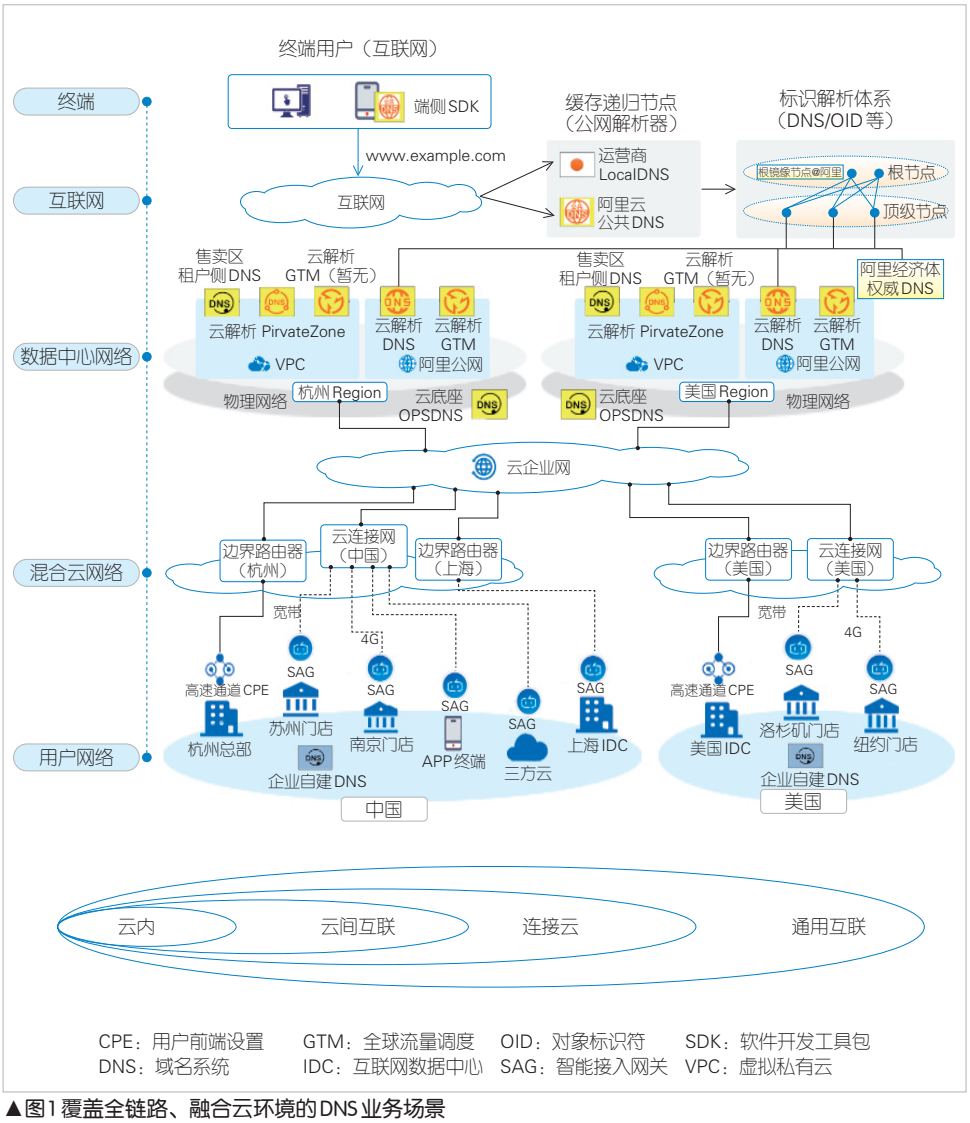
1.2 平台型IP寻址和流量调度服务（SaaS/PaaS）

随着移动互联网、在线视频服务的兴起，金融、民生、政务等重要行业加速了数字化进程，通用DNS IP解析功能已无法满足不同应用的性能、功能和安全需求。DNS服务对象和服务部署模式发生了改变：用户从PC转移到智能手机、智能终端物联网（IoT）；应用服务开始共享超大型第三方公共域名解析平台，并出现了超过百万级别的DNS权威解析托管服务，如Amazon Route 53、Alibaba Cloud DNS；公共递归DNS的出现也让用户流量集中在少数DNS递归服务平台，如Google的8.8.8.8和Cloudflare的1.1.1.1。

从业务功能上看，DNS不仅是一个简单查询静态的域名IP地址库，更是一个基于用户位置、资源状态、容灾等需求的动态智能的流量调度系统。从业务形态上看，DNS不再只是网络协议和基础组件，而是面向业务和应用，对外提供软件即服务（SaaS）或应用程序编程接口（API）的平台即服务（PaaS），具有更强的智能化、安全性和可扩展性。大型企业和云平台将更多的网络、研发运维资源投入到DNS服务，增强了服务的高可用和安全性。从全球范围来看DNS的故障变少了，但少数的故障的影响却更大了。平台型DNS的稳定和高可用仍然是挑战。

1.3 融合云DNS：面向IT数字资产的DNS融合管理架构

针对信息技术（IT）数字资产，出于服务高可用、风险控制和数据治理的考虑，越来越多的企业采用融合云部署的方式，即将企业的数字化业务资产同时部署在公有云、私有云、本地IDC、应用和智能终端等多场景（如图1所示）。在融合云场景下，IT数字资产高效管理和运维成为一个痛点。融合云DNS也应运而生，开始承担面向融合云的IT数字资产融合管理的角色。用户可以跨平台地统一管理、配置、维护DNS，达到统管、统维、统防的目标。



▲图1 覆盖全链路、融合云环境的DNS业务场景

在该阶段，平台型DNS企业已经具备全链路解析资源和信息，包括APP/智能终端解析器、递归解析、云上/云下权威解析、网络质量探测，提供DNS全链路安全可控的服务，同时，平台型DNS企业还采用IT视角用软件定义DNS服务，提供端-递归-权威融合的云端一体技术架构，摆脱了DNS协议固有安全限制。

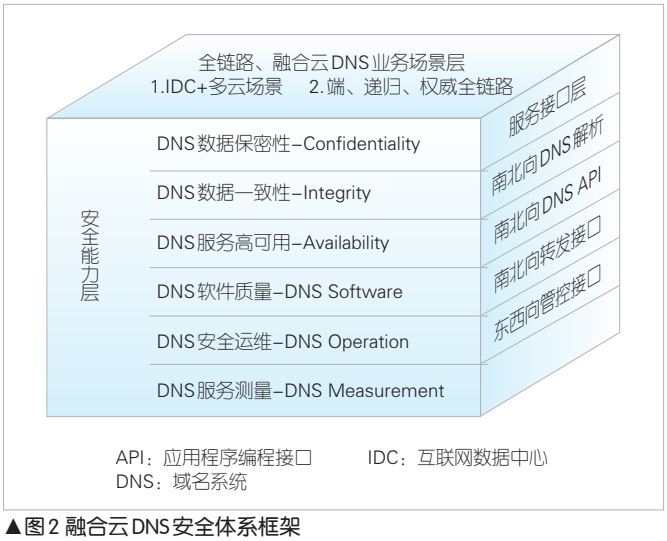
2 融合云DNS安全体系研究

结合该领域的相关工作和阿里云在融合云DNS方面的实践，我们提出了一个融合云DNS安全体系框架，如图2所示。该框架具体包括全链路的融合云DNS业务场景层、DNS业务服务接口层，以及最重要的安全能力层。融合云DNS的安全能力应该充分考虑南北向、东西向的业务和管控接口，覆盖多个融合云DNS业务场景。

2.1 DNS数据保密性

近年来，人们也提高了对DNS的关注度。如RFC9076所描述，DNS甚至被认为是互联网隐私泄露最严重的领域。传统DNS协议没有专门的安全机制来保障隐私安全。DNS查询响应数据揭示了特定用户和设备访问的网络行为，包括所访问域名和和位置信息。参考文献[11]指出，IoT设备会查询少数固定的域名，而DNS查询数据包信息的泄露会暴露智能的厂家、信号以及潜在的设备漏洞。

针对DNS隐私的安全防护，目前有两种基本思路：一种是尽量减少对外发送的信息，如RFC7816中提到的QNAME Minimization技术，该技术可以通过递归服务器修改查询的名字，以达到减少信息泄露的目的；另一类是将DNS数据通过安全传输层协议（TLS）加密信道进行传输，以达到数据隐私保护的效果，例如RFC7858中提到的基于TLS的DNS（DoT）、RFC8484中提到的基于安全超文本传送协议的DNS（DoH）。近几年，业界主流的浏览器、操作系统平台都已对外宣布支持DoT和DoH。



▲图2 融合云DNS安全体系框架

另外，在融合云场景下，除了DNS查询过程中的数据具有保密性以外，有研究表明DNS zone数据也可能包含敏感信息，因此也需要考虑数据保密性[12]。例如，在递归、权威服务器存储、与多个服务器通过完全区域传输（AXFR）/增量区域传输（IXFR）同步zone数据场景下，zone数据信息都有可能被泄露。因此，在融合云DNS设计和使用中，需要充分考虑必要的传输和存储的加密，为各个功能接口和应用场景预留DNS数据保密能力。

针对DNS数据保密性，仍有如下的3个方面内容需要进一步关注：

1）除了静态配置服务外，如何在家庭、企业不同场景下支持动态DoH/DoT地址和加密证书服务发现机制。目前，IETF的ADD工作组正在讨论制定新的技术标准。

2）当DNS递归成为数据加密传输的中心化节点后，如何保证递归服务商不泄露数据。曾经有Oblivious DNS和Oblivious DoH（ODoH）方案，隔离查询域名和用户地址的对应关系，但是并没有得到技术社群的一致认可。

3）递归-权威的加密传输机制研究。递归需要在面对大量权威服务器具备加密证书服务发现的能力。

2.2 DNS数据一致性

DNS协议固有的缺陷会使DNS易遭受缓存数据篡改，从而引导用户访问攻击者设置的恶意网站和文件内容。典型的有Kaminsky攻击[13]，以及2020年出现的因DNS侧信道漏洞带来的缓存“投毒”[14]。早在2000年，DNSSEC的概念被提出，旨在通过携带签名保障数据传输的一致性和可验证性。然而，直到2023年1月，全球递归支持DNSSEC校验的比例仅有31%。

除了缓存投毒攻击以外，数据一致性的安全风险也会发生在数据源头。网络黑客可以攻陷域名注册账号，通过域名注册商平台直接给DNS zone文件注入恶意数据，这类攻击也被称为“注册劫持”。例如：2019年发生的海龟攻击[16]，即攻击者通过系统漏洞获取用户权限后，恶意篡改用户的DNS注册信息。在融合云环境中，需要引入“域名注册锁”或者双因子认证机制来保护DNS注册信息不被攻击者单方面篡改。

DNS数据一致性的挑战在于：DNSSEC由于其复杂性难以大规模部署，也没有新的签名保障机制。另外，由于大规

模量子计算的发展，现在广泛使用的加密算法如RSA（算法名）和椭圆曲线加密（ECC）会更容易破解。因此，我们需要考虑量子安全算法的研究，以进行快速替换。更大的密钥和签名可能会给现有的DNSSEC和DNS数据的传输带来挑战。

2.3 DNS服务高可用

DNS是互联网基础服务，在互联网业务的高可用和稳定性方面起到至关重要的作用。互联网中心化的趋势也体现在DNS领域中，因此越来越多的网络应用与服务共享超大型第三方公共域名解析平台。研究表明，在全球排名前10万的流行域名之中，有89%的域名使用了公共域名解析服务[3-4]。为了高效管理，基于云平台的DNS域名托管服务往往为其托管的大量域名配置共用的解析服务和统一的安全策略。一旦域名托管服务遭遇大型分布式拒绝服务（DDoS）攻击，则会导致服务不可用[17]。还有一些研究表明，域名托管服务所管理的域名存在被攻击者接管的风险[9-10]。因此，平台型DNS需要具备应对各种故障的弹性能力，以提供极致的高可用服务，也就是说能够在任意时间、任何地方都能提供间断的服务。

DNS的高可用能力可以从表1中的7个方面来加强。

2.4 DNS软件质量

《“十四五”软件和信息技术服务业发展规划》要求强化基础组件供给，推进域名、标识等基础资源管理与服务的软件研发。然而，中国仍有不少企业和网络运维人员使用和集成其他国家开源软件，为国家数字经济基础设施安全带来潜在软件供应链安全风险。

中国互联网信息中心2021年发布的《中国域名服务安全状况与态势分析报告》[8]指出，中国二级以下权威域名服务器主要使用互联网系统联盟（ISC）维护的开源软件BIND，占比达到59%。其中，超过40%的BIND开源软件仍旧开启版本应答功能，为漏洞扫描和针对性攻击留下安全隐患。据互联网安全大会（ISC）官网统计，2016—2021年的5年，BIND共有69个软件bug被曝出，仅2022年就新增11个软件漏洞，这些漏洞主要为DDoS安全威胁漏洞[18]。

大型的商业化DNS公司和机构出于高性能和高安全性的考虑，都会专注于自主研发DNS软件和安全测试技术。这样可以在保障DNS软件质量的同时，在遇到问题时能够快速定位软件故障并进行服务恢复。为了缓解一款软件带来的质量和安全风险，企业通常也会考虑采用至少两个不同开发者的DNS软件来增加系统的多样性[20]，避免单一软件可能带来的漏洞，但相应的代价是需要增加运营和维护成本。总归DNS软件的质量对服务安全至关重要。

DNS软件安全风险主要体现在两个方面：一方面，从软件开发的角度看，很难100%保证软件没有漏洞。对此，业界有通用的软件自动化测试方法，例如Fuzz测试和symbolic测试，但是它们很难对复杂的语意和交互式行为的DNS软件进行可扩展的测试。另一方面，从网络通信软件方面看，DNS软件的实现需要严格遵守协议，对于协议定义不明确的领域DNS软件，要考虑安全异常的情况。目前，业界有不少针对域名协议的安全漏洞分析，如缓存投毒、DNSSEC等。然而，对于域名协议的载体、域名解析软件实现代码的安全性和正确性研究工作不多，在最近一两年才逐渐引起研究者的重视[19]。

2.5 DNS安全运维

超大规模的分布式、平台型的DNS安全运维是DNS服务中至关重要的一环，直接影响用户体验和服务质量。当云平台DNS进入融合云DNS阶段时，安全运维场景更加复杂。

▼表1 提升DNS高可用能力的7种方案

机制	权威	递归/缓存	需求描述
性能/资源	适用	适用	在DNS服务器处理能力和网络带宽两个方面预留足够的资源来抵御超过3～10倍的攻击流量。对于无差别网络洪泛攻击，可结合特殊的DDoS防御机制做流量清洗
多个NS	适用	不适用	权威设置多组NS，通过zone文件分发同步数据
组播Anycast	适用	适用	每个NS可以在不同的地理位置的站点通过Anycast组播机制设置镜像和备份服务器
服务检测	适用	适用	大规模的DNS服务需要对多站点部署的服务可用性进行监控，如服务器状态、业务流量、数据一致性等
缓存	不适用	适用	当某一权威服务器不可达时，可以采用本地历史或缓存数据来应答服务（RFC8767）
安全控制	适用	适用	当遭遇攻击时，可以采用RRL以及ACL安全控制等手段来减少应答长度，降低高可用风险
多样性	适用	适用	在网络部署、软件硬件选型方面充分考虑多样性，增加系统性的冗余

注：表1中的方案能够增强DNS的服务能力，但是无法百分之百保证DNS服务的高可用。在互联网中心化趋势的背景下，超大规模的DDoS攻击一直会是DNS服务可用性的挑战之一。

ACL：访问控制列表　DDoS：分布式拒绝服务　DNS：域名系统　NS：名字服务器　RRL：响应速率限制

有研究认为，大量的DNS故障来自于DNS配置变更[22]。例如：2019年，微软因NS配置变更错误导致了其在线服务全球故障，微软Azure云宕机3个小时[21]；2021年Facebook自动化运维漏洞导致了路由故障和DNS服务不可用，引起了其旗下的各种应用故障，持续6个多小时[2]。

学术和产业界都很重视DNS安全运维的研究和建设。针对潜在的运维漏洞和DNS变更故障，微软开发了GRooT工具[22]，通过分析DNS配置文件来排查潜在的DNS服务风险[22]；作为中国最大的DNS云平台，阿里云也从管理保障、设计与开发、测试与评估、发布与变更、监控与应急、基础设施保障等各方面保障DNS服务安全稳定[23]。

2.6 DNS测量

域名解析服务异常有可能隐藏在正常业务中，虽没有造成大规模故障但却存在安全隐患，因此我们需要对DNS服务进行安全测量来揭示中国乃至全球DNS运行的规律和安全风险。近期一项研究工作表明，互联网中13.5%的域名解析查询均会以失败告终[6]。另外，引发域名解析服务异常的原因不尽相同，其中包括域名权威服务器配置错误、网络通信链路存在劫持、网络中间件缓存不一致性等等[5,7]，这些都需要通过DNS安全测量来定位根因。

DNS安全测量通常有主动发探测包的主动测量和收集DNS业务数据的被动测量两种方案。但由于探测节点的广度、测量手段的局限，现有测量方法还很难还原完整的DNS解析链路全局信息。尤其是对复杂的云平台DNS，以及融合云DNS而言，当域名解析发生故障时，通常难以准确获知用户终端网络、IDC侧网络的环境信息，也难以有效排查并追溯故障原因。总之，DNS服务测量、故障定位一直是业界的难点。

3 阿里云DNS的安全实践

面对融合云DNS的发展趋势和安全挑战，阿里云在业界首次提出了全链路安全可控的融合云DNS技术。该技术不再局限于单一设备、单一服务运行场景，而是从DNS业务的全局视角出发，实现融合云环境下的统一管理、统一运维、统一防护。技术服务覆盖了公网域名解析、内网域名解析、全球流量调度、移动解析、专有云和客户IDC的域名解析场景。

为了提供高可用的永远在线解析服务，同时保障云生用户的域名解析安全稳定，阿里云DNS将安全实践主要在下面集中于以下几个方面：

1）基础资源和能力：基于阿里云全球覆盖的基础设施，阿里云在全球28个地理区域内运营着86个可用区，部署了243个DNS集群，日解析量超过2万亿次。

2）安全攻击防护：基于云计算的弹性和安全运维体保障了DNS服务的弹性和安全抗攻击能力，具备全球10T+带宽储备和多个大型流量清洗中心，提供大规模DDoS流量攻击防护。

3）自研软件：具备全链路的融合云DNS软件自研能力，能够自研高性能解析服务器集群，且单集群每秒过亿防护能力。

4）安全研究：对DNS服务异常测量、软件安全漏洞测试、正确性验证进行深入研究。

5）数据一致性：提供DNSSEC在线签名服务，避免DNS劫持/缓存投毒，保障网站访问安全。

6）数据隐私：提供了DoT、DoH和HTTP(s) DNS件开发工具包（SDK）服务，保障DNS传输的数据隐私性。

7）数据本地化：引入了DNS根镜像、.CN./.COM/.NET镜像，以及本地备份重要的热点DNS数据，预防因网络中断导致的DNS服务不可达故障。

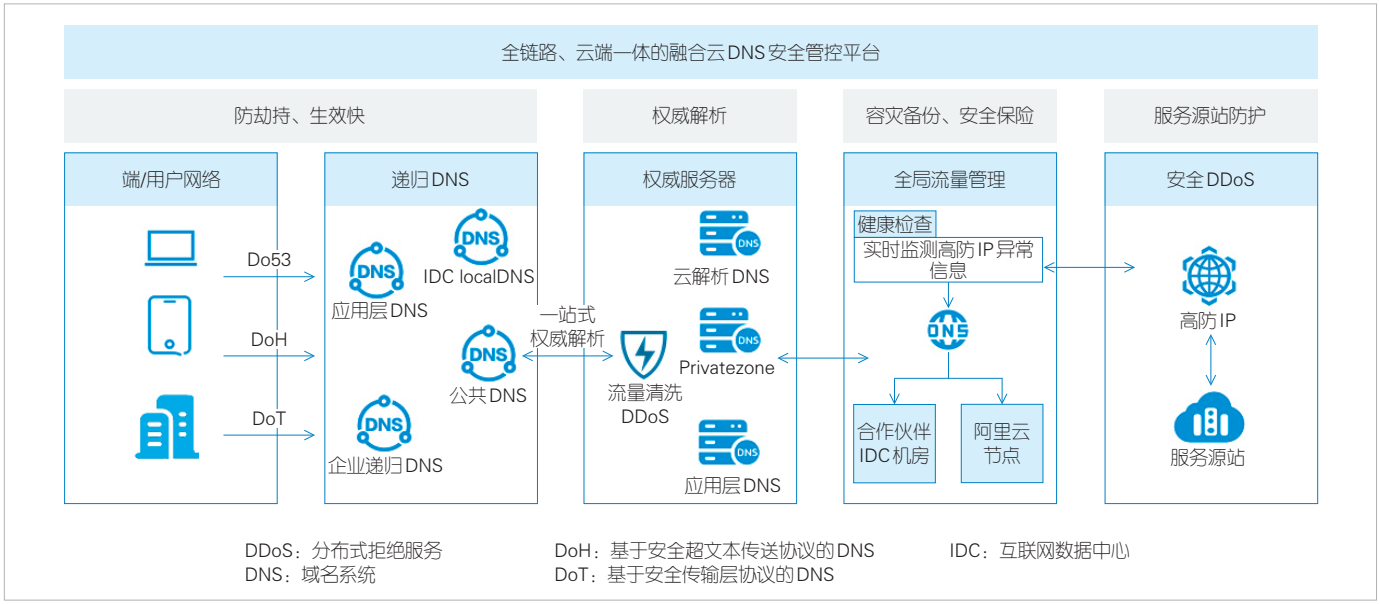
8）云端一体安全运维体系：阿里云全系列的DNS产品和技术能力覆盖了用户端（通过软件和SDK部署）、企业/公共递归解析、权威解析、DNS服务检测和调度、DDoS防护，从而能够从全链路视角来服务客户，减少中间链路的不确定性，保障DNS业务的安全可控、可预期，具体信息如图3所示。

4 结束语

DNS是全球互联网也是中国数字经济的重要基础设施，其安全稳定至关重要。DNS从一个简单的网络IP数据查找的基础组件，发展到智能算力和流量调度的平台型服务（SaaS/PaaS）。步入融合云时代，DNS在网络协议和平台型服务的基础上，又增加了新的功能场景，成为更高效的现代企业IT数字资产和流量调度管理平台。DNS的功能和应用场景更加丰富，承担的流量也进一步集中化和平台化，这其中机遇和挑战并存。

基于阿里云的DNS安全研究和运维经验，本文提出了融合云DNS安全体系框架，也介绍了我们在安全和稳定性方面所做工作。我们虽然在该领域收获了一些成果，但对DNS的安全和稳定性心存敬畏，因为一个小的故障就可能引发大量用户、大面积的业务受损。

DNS是国家关键信息基础设施的组成部分。当前，中国仍存在依赖外部关键域名解析资源（根和顶级服务器）、开源软件核心组件，DNS软件国产化水平不足，安全运维总体



▲图3 全链路、云端一体的融合云DNS安全运维体系

能力不高，对DNS安全体系缺少顶层设计等一系列问题。DNS是一个生态，需要全产业链参与其中，共同分享、协作，并统筹行动，应对各种安全风险。

参考文献

[1] The United States Department of Justice. United States seizes websites used by the Iranian Islamic radio and television union and Kata'ib hizballah [EB/OL]. [2022-12-04]. https://www. justice. gov/opa/pr/united-states-seizes-websites-used-iranian-islamic-radio-and-television-union-and-kata-ib

[2] Facebook. Update about the October 4th outage [EB/OL]. [2022-11-22]. https://engineering.fb.com/2021/10/04/networking-traffic/outage/

[3] KASHAF A, SEKAR V, AGARWAL Y. Analyzing third party service dependencies in modern web services: have we learned from the mirai-dyn incident? [C]//Proceedings of the ACM Internet Measurement Conference. ACM, 2020: 634-647. DOI: 10.1145/3419394.3423664

[4] MOURA G C M, CASTRO S, HARDAKER W, et al. Clouding up the Internet: how centralized is DNS traffic becoming? [C]//Proceedings of the ACM Internet Measurement Conference. ACM, 2020: 42-49. DOI: 10.1145/3419394.3423625

[5] LIU B J, LU C Y, DUAN H X, et al. Who is answering my queries: understanding and characterizing interception of the DNS resolution path [C]//Proceedings of the Applied Networking Research Workshop. ACM, 2019: 15-16. DOI: 10.1145/3340301.3341122

[6] LU C Y, LIU B J, ZHANG Y M, et al. From WHOIS to WHOWAS: a large-scale measurement study of domain registration privacy under the GDPR [EB/OL]. [2022-12-03]. https://www. researchgate. net/publication/349216875_From_WHOIS_to_WHOWAS_A_Large-Scale_Measurement_Study_of_Domain_Registration_Privacy_under_the_GDPR

[7] LU C Y, LIU B J, LI Z, et al. An end-to-end, large-scale measurement of DNS-over-encryption: how far have we come? [C]//Proceedings of the Internet Measurement Conference. ACM, 2019: 22-35. DOI: 10.1145/3355369.3355580

[8] 中国互联网络信息中心. 中国域名服务安全状况与态势分析报告[R/OL]. [2022-12-03]. https://www.cnnic.cn/NMediaFile/2022/0827/MAIN16615908654649L2MS5ZEJS.pdf

[9] LIU D P, HAO S, WANG H N. All your DNS records point to us: understanding the security threats of dangling DNS records [C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2016: 1414-1425. DOI: 10.1145/2976749.2978387

[10] ALOWAISHEQ E, TANG S Y, WANG Z H, et al. Zombie awakening: stealthy hijacking of active domains through DNS hosting referral [C]//Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2020: 1307-1322. DOI: 10.1145/3372297.3417864

[11] RASMUSSEN R. ICANN SAC105: The DNS and the Internet of things: opportunities, risks, and challenges [EB/OL]. [2022-11-18]. https://www.icann. org/en/blogs/details/dns-and-the-internet-of-things-opportunities-risks-and-challenges-18-7-2019-en

[12] SKWAREK M, KORCZYNSKI M, MAZURCZYK W, et al. Characterizing vulnerability of DNS AXFR transfers with global-scale scanning [C]//Proceedings of 2019 IEEE Security and Privacy Workshops (SPW). IEEE, 2019: 193-198. DOI: 10.1109/SPW.2019.00044

[13] KAMINSKY D. Black ops 2008: it's the end of the cache as we know it, in: Black Hat USA, 2008 [EB/OL]. [2022-11-15]. https://www.blackhat.com/presentations/bh-jp-08/bh-jp-08-Kaminsky/BlackHat-Japan-08-Kaminsky-DNS08-BlackOps.pdf

[14] MAN K Y, QIAN Z Y, WANG Z J, et al. DNS cache poisoning attack reloaded: revolutions with side channels [C]//Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2020: 1337-1350. DOI: 10.1145/3372297.3417280

[15] WWDC. Improve DNS security for apps and servers [EB/OL]. [2022-11-16].https://developer.apple.com/videos/play/wwdc2022/10079/

[16] Cisico Talos. A DNS hijacking called sea turtle [EB/OL]. [2022-12-03]. https://blog.talosintelligence.com/seaturtle/

[17] ABHISHTA A, VAN RIJSWIJK-DEIJ R, NIEUWENHUIS L J M. Measuring the impact of a successful DDoS attack on the customer behaviour of managed DNS service providers [J]. ACM SIGCOMM computer communication review, 2019, 48(5): 70-76. DOI: 10.1145/3310165.3310175

[18] CVE. ISC BIND CEV 漏洞库 [EB/OL]. [2022-12-04]. https://www.cvedetails.com/product/144/ISC-Bind.html?vendor_id=64

[19] KAKARLA S K R, BECKETT R, MILLSTEIN T, et al. SCALE: automatically finding RFC compliance bugs in DNS nameservers [EB/OL]. [2022-12-28]. https://www.usenix.org/conference/nsdi22/presentation/kakarla

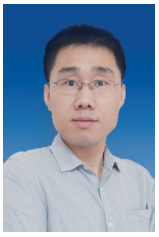
[20] SIGARAM A. Implementing dual stack recursive DNS at Microsoft: challenges and learning [EB/OL]. [2022-12-04]. https://indico. dns-oarc.net/event/42/contributions/904/

[21] TUNG L. Azure global outage: our DNS update mangled domain records, says Microsoft [EB/OL]. [2022-11-15]. https://www. zdnet. com/article/azure-global-outage-our-dns-update-mangled-domain-records-says-microsoft/

[22] KAKARLA S K R, BECKETT R, ARZANI B, et al. GRooT: proactive verification of DNS configurations [EB/OL]. [2022-11-25]. https://dl.acm.org/doi/10.1145/3387514.3405871
[23] 阿里云基础设施. 阿里云DNS荣获信通院2022首批“分布式系统稳定性保障能力评估”最高等级证书 [EB/OL]. (2022-04-28)[2022-12-01]. https://mp.weixin.qq.com/s/f8AFlr8EyModp_C78e7_CA

作 者 简 介

宋林健，阿里云计算有限公司高级技术架构师、ICANN根服务器咨询委员会专家组ICANN TEG技术专家、国家OID注册中心受邀专家；长期从事互联网体系架构、互联网地址标识领域等研究工作；曾参与多个下一代网络体系结构国家“973”和“863”计划项目，参与制定IETF RFC8483、ITU-T X.672等；发表多篇论文，拥有发明专利20项。



马永，阿里云计算有限公司基础设施网络研发高级技术专家、阿里云DNS解析研发&运维负责人；负责阿里DNS解析服务平台的技术架构设计、功能研发、建设运维工作，主导了阿里DNS系统的多次架构升级。

梁卓，阿里云计算有限公司基础设施网络研发技术总监、DNS产研负责人；长期从事网络基础服务关键技术和应用体系开发和设计工作；曾先后负责和参与工业和信息化部的相关项目、国家“863”计划项目等，并积极参与物联网标识解析国际标准ISO/IEC 29168-2的制定。

构建可扩展的RPKI依赖方系统部署机制

Scaling RPKI Relying Party System

马迪/MA Di
（互联网域名系统国家地方联合工程研究中心，中国 北京 100102）
(Internet Domain Name System National Engineering Research Center, Beijing 100102, China)

DOI：10.12142/ZTETJ.202301008
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230227.1253.004.html
网络出版日期：2023-02-27
收稿日期：2022-12-15

摘要：互联网码号资源公钥基础设施（RPKI）依赖方系统是各类网络运行机构开展RPKI应用实践的一个关键环节。RPKI依赖方系统的研发和部署，既需要处理RPKI核心功能的“普遍性”问题，又需要兼顾网络互联互通特征的“特殊性”问题。相关解决方案需要考虑RPKI依赖方系统应当有哪些组件，各个组件如何在网络上分布，以及以何种逻辑关系分布。面向RPKI依赖方系统的核心功能，梳理了影响RPKI依赖方系统运行效能的4对矛盾，并提出了一种可扩展的RPKI依赖方系统部署机制，包含软件层面的解耦机制和硬件层面的部署机制。

关键词：RPKI；路由安全；互联网码号资源管理

Abstract: The resource public key infrastructure (RPKI) relying party system is key to network operations with regard to the RPKI in practice. The development and deployment of the RPKI relying party system involves both the essential functionality of the RPKI universally and the networking condition where it operates particularly. The very resolution calls for the design of modularizing the RPKI relying party system and deploying those modules physically and logically. Four contradictions regarding the operation efficiency of the RPKI relying party system are summarized and a scheme of scaling the RPKI relying party system is proposed with respect to both the decoupling mechanism of software and the deployment principle of network hardware.

Keywords: RPKI; routing security; Internet number resource management

1 研究背景

自2012年国际互联网工程任务组（IETF[1]）完成基础协议的标准化工作以来，历经国际互联网体系结构委员会（IAB[2]）的背书[1]以及国际互联网路由安全自律协定（MANRS[3]）项目面向全球网络运行机构的推广倡议，互联网码号资源公钥基础设施（RPKI）已成为解决当前互联网域间路由安全问题的技术路线共识。RPKI的理念肇始于互联网安全协议专家S. KENT博士的论文[2]，并通过对传统的基于X.509公钥基础设施进行扩展[3]，进入到IETF的工业标准体系。

RPKI的部署和运行是一个复杂的系统工程，需要网络运行机构（网络运营商、互联网交换中心、内容分发网络服务商等）、RPKI数据服务机构（互联网IP地址注册机构、RPKI依赖方系统服务商等）以及路由器制造商等角色的配合和协调。其中，RPKI依赖方（RP）系统充当了地址管理系统和路由控制系统之间RPKI数据传递的桥梁，是连接RPKI数据供给侧和RPKI数据需求侧的RPKI生态关键环节。

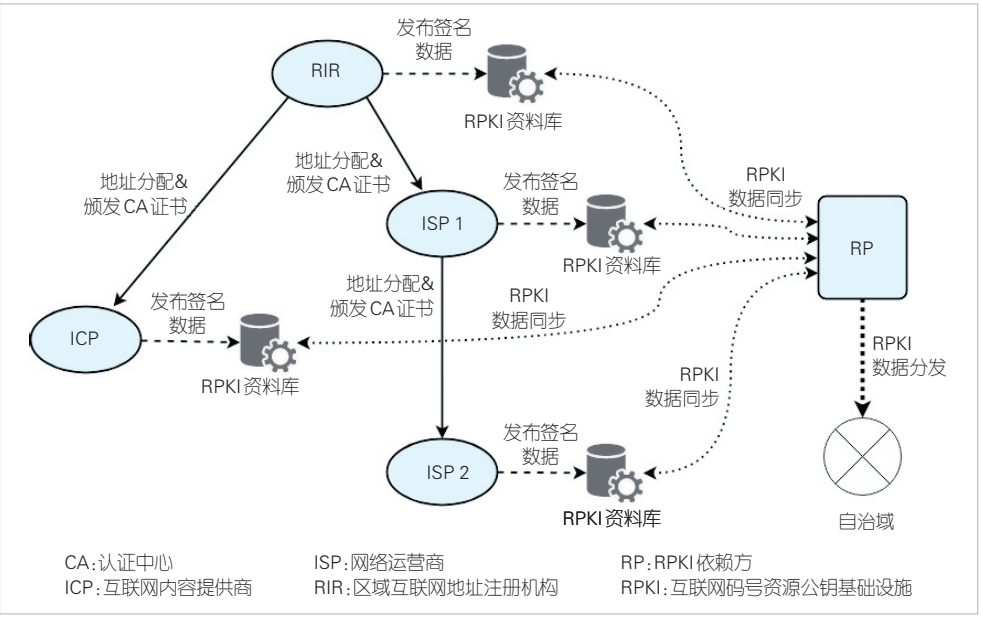
RPKI依赖方系统的部署，涉及网络运行机构的路由控制策略、安全保障策略和地址分配策略，需要统筹网络规模、拓扑结构、互联互通策略以及地址资源分配格局等要素。这些要素会因应业务和技术的演进而随之变化。设计一个既可以处理RPKI依赖方系统功能诉求的“普遍性”问题，又能兼顾具体网络（及其变化）“特殊性”问题的可扩展部署机制，是RPKI技术在网络运营商、互联网交换中心等网络运行机构落地应用的关键。

2 RPKI原理简介

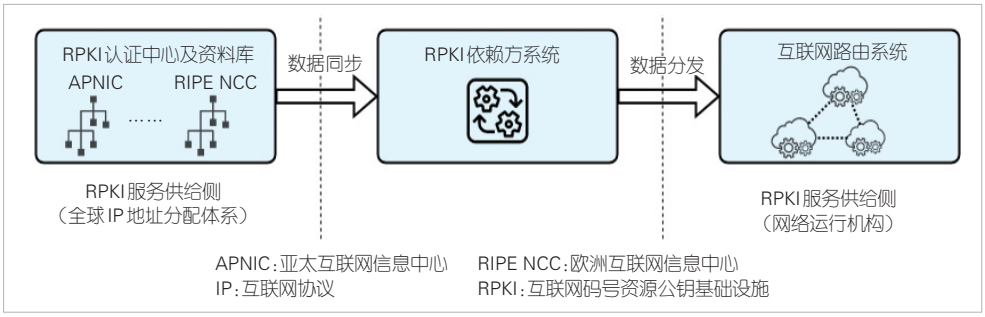
RPKI是一种公钥证书基础设施。基于当前全球的互联网码号资源分配关系，RPKI构建了一个面向IP地址及互联

1 IETF: Internet Engineering Task Force，国际互联网工程任务组（https://www.ietf.org/）
2 IAB: Internet Architecture Board，国际互联网体系结构委员会（https://www.iab.org/）
3 MANRS: Mutually Agreed Norms for Routing Security，国际互联网路由安全自律协定（https://www.manrs.org）

网自治系统（AS）号码的授权认证体系，并以X.509证书扩展及若干签名对象的形式加以呈现。如图1所示，RPKI体系中的码号资源分配者在分配资源的同时，为下游节点签发资源证书（基于X.509证书的扩展）。依托RPKI提供的认证功能，互联网码号资源（IP地址及AS号码）的最终用户单位（资源持有者）通过签发相关数据对象，来完成路由通告相关信息的发布（例如路由起源授权等）。作为RPKI认证体系的依赖方，参与域间路由交互的网络运行机构（例如网络运营商、互联网交换中心等）定期从RPKI资料库系统（基于码号资源分配关系组织起来的分布式数据存储体系）同步资源证书以及包括各类基于RPKI的数据对象，并将经过验证的信息推送给边界路由器，供其在接收路由通告时进行真伪判断。



▲图1 RPKI的工作原理



▲图2 RPKI的生态结构

概括地讲，RPKI生态有3个组成部分：RPKI供给侧、RPKI依赖方、RPKI需求侧。如图2所示，RPKI供给侧包含全球分布的RPKI认证中心（CA）以及用于存储RPKI资源证书和各类RPKI数据对象的分布式RPKI资料库系统。RPKI需求侧是当前互联网的域间路由系统，由部署在不同网络自治域的边界网关协议（BGP）路由器组成。RPKI依赖方是连接“供给”和“需求”的桥梁，负责收集RPKI供给侧产生的数据并加以验证后交付给RPKI需求侧参考使用。

RPKI数据尽可能快速、完整、准确地从供给侧扩散至需求侧的关键在于RPKI依赖方。全盘考察RPKI的运行机制，并结合相关RPKI依赖方系统的运行实践，笔者归纳了影响RPKI依赖方系统效能的4对矛盾：

矛盾1：RPKI资料库（发布点）越来越多，与实时感知全球RPKI数据更新情况之间的矛盾；

矛盾2：RPKI数据对象数量越来越多，与快速同步全球RPKI数据之间的矛盾；

矛盾3：RPKI数据授权链（深度和广度）越来越复杂，和快速构建全球RPKI数据认证路径之间的矛盾；

矛盾4：RPKI依赖方系统集中化趋势（远离网络边缘），和路由器快速获得RPKI认证数据之间的矛盾。

以上矛盾既有在“RPKI基本原理范畴”的普遍性，又有在“网络互联互通特征范畴”的特殊性，因此需要构建一个能够因应网络规模和信任模型变化而灵活调整的可扩展RPKI依赖方系统部署机制。该机制映射至解决方案层面，即RPKI依赖方系统应当有哪些组件，组件如何在网络上分布及以何种逻辑关系进行分布。

3 RPKI依赖方系统组件的功能解耦机制

可扩展RPKI依赖方系统部署机制在“RPKI基本原理范畴”的首要任务是，通过对RPKI依赖方系统核心功能实施解耦，形成彼此“正交”的RPKI依赖方系统的组件布局。按此原则，基于IETF RFC8897[4]，笔者从工程实践的角度梳理了RPKI依赖方系统在理论上的最低技术要求，包括：同

步RPKI资料库的数据、处理RPKI资源证书、处理RPKI数据签名对象、分发验证过的RPKI认证信息以及本地化控制。一个具备IETF RFC8897所列举功能的系统，可以称为RPKI依赖方系统。

RPKI依赖方系统在全球各个网络内的部署已逾10年，形成了一些运行实践和讨论。笔者在起草IETF标准和进行RPKI系统设计的工作中，有两点体会：在部署层面，RPKI依赖方系统的功能仍需要进一步模块化（正交化）；在运行层面，IETF RFC8897所列举的功能无法满足商用路由控制系统对RPKI依赖方系统的需求。为此，构建可扩展的RPKI依赖方系统的第一步是将其核心功能模块化，并给出各个模块彼此解耦之后的逻辑关系（接口关系）。图3是笔者对RPKI依赖方系统的设计思考和建议。

1）更新感知系统

将“更新感知”功能同“数据同步”功能进行解耦，是互联网内容分发范畴常见的工程设计思路。当这一思路被应用到RPKI体系时，更新感知系统便成为了一个独立的RPKI依赖方系统组件。该系统采用实时或不定时的方式获得RPKI资料库（RPKI数据发布点）的数据更新情况，并将这些更新信息传递给数据同步系统。

2）数据同步系统

数据同步系统以“全量”或“增量”的方式，将RPKI资料库内发布的各类RPKI资源证书及RPKI数字签名对象下载到本地网络，以形成与RPKI资料库一致的且具有一定时效的数据副本。在当前的RPKI生态中，数据同步系统和RPKI资料库之间的接口已经在IETF形成标准[5]。

3）数据验证系统

数据验证系统先后对相关证书及数据签名对象进行语法检查和RPKI逻辑验证。其中，语法检查包括检查相关数据格式是否符合技术标准、是否在有效期内等，RPKI逻辑验证包括验证PKI数字签名、验证相关的互联网码号资源包含关系[6]以及其他与RPKI授权体系相关的逻辑验证等。

4）数据分析系统

RPKI是一个分布式系统，因此位于不同RPKI授权体系子树上的数据可能存在冲突关系（码号资源分配、授权信息等）。数据分析系统旨在根据一定的算法，辅之以WHOIS数据、BGP广播存档数据等带外数据，对潜在的冲突关系进行检测，给出可能的（本地化）修正方案，并输出至本地控制系统。

5）本地控制系统

出于网络管理和安全保障的需求，网络运行机构可能希望以“本地过滤和添加”的形式建立RPKI路由认证数据的本地视图，对来自全球RPKI的数据进行覆盖。本地控制系统对“验证缓存”直接进行操作，增加或删减相关的路由认证数据条目。工业界将这种操作称为RPKI本地化控制（SLURM）。SLURM配置文件格式已经在IETF形成标准[7]。

6）数据分发系统

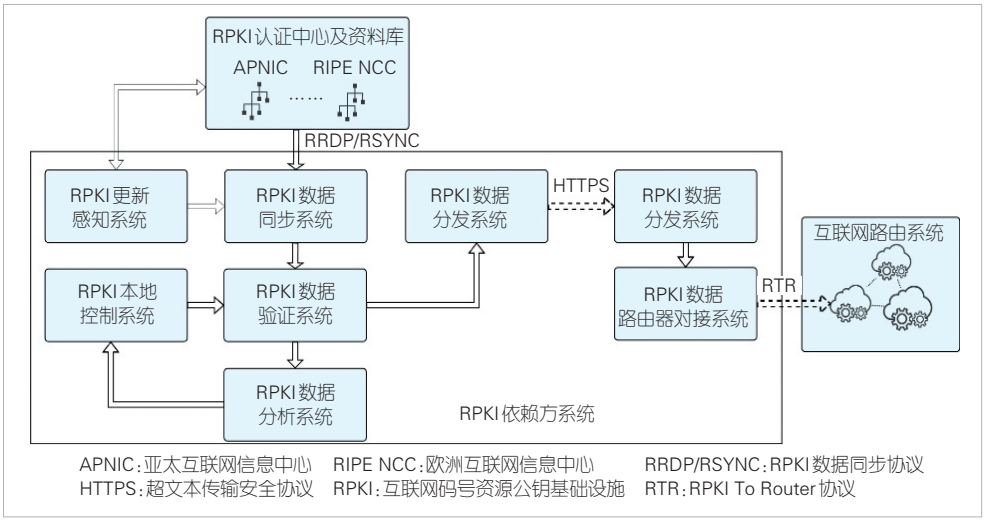
经过验证并最终可以供给路由器进行RPKI路由认证的数据称为“验证缓存”。基于“主从模型”，数据分发系统将“验证缓存”从一个网络节点分发至其他一个或多个有信任关系的网络节点，实现RPKI验证数据的共享。对于数据分发系统，RPKI意义上的数字签名已不复存在，其完整性依赖于传输信道（如超文本传输安全协议）。

7）路由器对接系统

RPKI供给侧形成的路由认证数据，通过同步、验证、分发、本地化处理，最终形成副本，然后经路由器对接系统，注入至BGP路由器。基于“客户端–服务器”操作模型，维护“验证缓存”的网络节点充当服务器，同时路由器充当客户端。两者之间的通信由一种RTR的IETF标准化协议承载[8]。

4 RPKI依赖方系统组件在规模网络上的编排机制

基于RPKI依赖方系统核心组件的解耦机制，本节探讨可扩展RPKI依赖方系统部署机制在“网络互联互通特征范

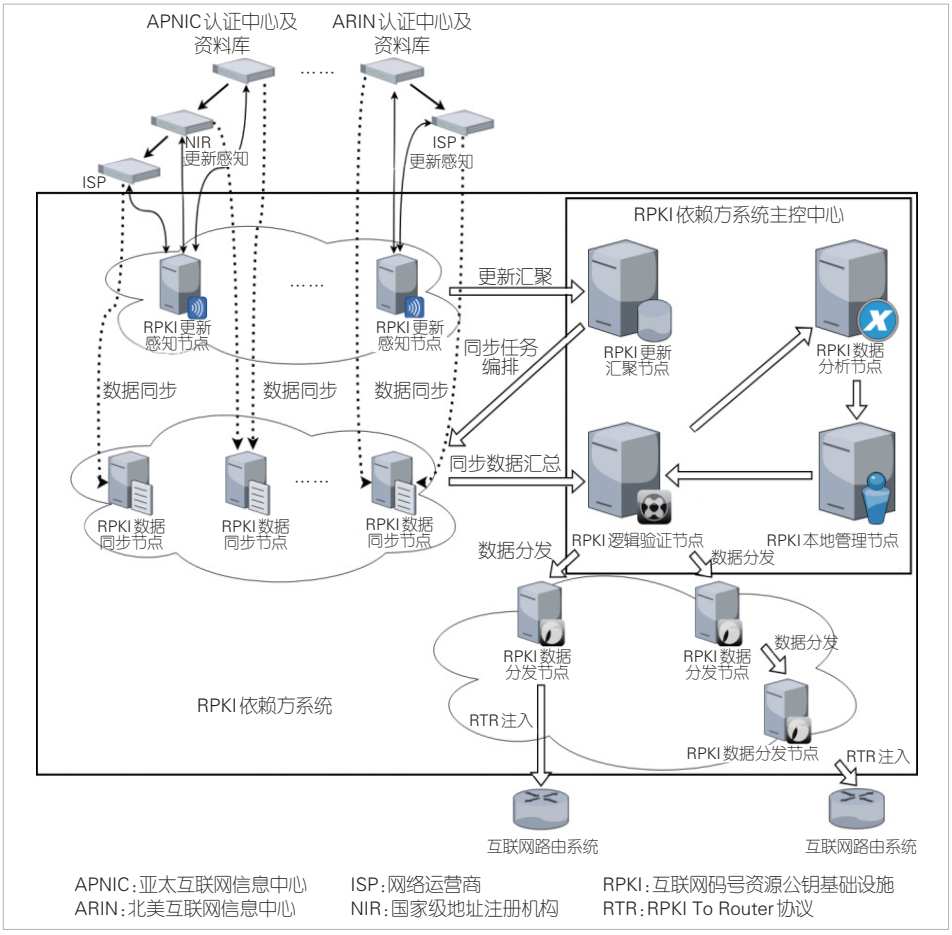


▲图3 RPKI依赖方系统组件

畴”的特殊性，就如何将相关组件编排在网络运行机构所管理的云和网的不同位置，设计一个能够适配各类规模网络（骨干网运营商、互联网交换中心、内容分发网络服务商等）的RPKI依赖方系统组件部署机制（框架）。

针对前文所述RPKI依赖方系统的4对矛盾，面向一个规模网络的一般特征，笔者建议在RPKI依赖方系统的组件颗粒度上展开相关设计，包括：一个RPKI依赖方系统北向分布式节点群组、一个RPKI依赖系统方南向分布式节点群组和一个RPKI依赖方系统主控中心。如图4所示，北向分布式节点群组面向RPKI供给侧，从分布式的RPKI资料库获取RPKI原始数据，包含RPKI更新感知节点和RPKI数据同步节点；南向分布式节点群组面向RPKI需求侧，将验证过的RPKI路由认证数据分发给分布式网络的边界路由器，即RPKI数据分发节点的集合；主控中心负责RPKI数据的验证和其他综合处理任务，包含RPKI更新汇聚节点、RPKI逻辑验证节点、RPKI本地管理节点、RPKI数据分析节点等。

RPKI依赖方系统的各个组件在该架构下的部署机制如下：



▲图4 规模网络上的RPKI依赖方系统组件部署示例

1）更新感知系统

鉴于RPKI资料库的全球分布特征，更新感知系统相应地采用分布式的更新获取方法。该系统拥有“更新感知模块”和“更新汇聚模块”。前者部署在分布式的“更新感知节点”之上。全体“更新感知节点”按照一定的编排算法各自分工，完成对RPKI资料库的遍历，并传递给负责整合更新信息的“更新汇聚节点”。“更新汇聚节点”再将更新信息传递至数据同步系统。

全球大型运营商（例如Tier 1 ISP）或头部流量的互联网交换中心，可以考虑将更新感知系统的寻址信息（域名、IP地址等）和接口方式公布出去，供相关的RPKI发布点主动推送更新信息。

2）数据同步系统

面向全球RPKI资料库的分布特征，数据同步系统也采用分布式的部署形态，根据一定的编排算法，将同步任务分散至不同的“数据同步节点”。“更新汇聚节点”负责运行该编排算法，在统筹更新任务来源、同步节点数量、同步节点分布位置等要素的前提下，实现同步任务的动态分配。多个“数据同步节点”的分布采用和“更新感知节点”类似的策略。

3）数据验证系统

面向RPKI的数据验证系统的核心是建立数据对象之间的关联，包括经典公钥基础设施（PKI）体系下的数字签名验证路径构建，以及RPKI特有的码号资源包含关系验证。这种“关联性”的验证任务（证书路径验证、资源包含关系验证）由数据验证系统的RPKI逻辑验证模块负责。该模块站在全局视角，以集中化的方式对来自不同“数据同步节点”的合规数据进行综合处理，并以“RPKI逻辑验证节点”的形式部署在网络运行机构的网运中心（NOC）。数据验证系统的语法检查模块不涉及对数据关联性的验证，可以部署在“数据同步节点”之上，对相关数据进行语法合规检查，实现“边同步，边语法检查”的高效机制。合规数据会汇聚至“RPKI逻辑验证节点”。

4）数据分析系统

数据分析系统承担的是RPKI数据同步、验证等任务之外的旁路功能，宜独立部署在专用的RPKI数据分析节点之中。

5）本地控制系统

本地控制系统将连同其他一些面向RPKI的互联网码号资源本地管理支撑系统（可视化、运行监控等），部署在专用的RPKI本地管理节点之中。

6）数据分发系统

鉴于数据分发系统的核心任务是将RPKI验证缓存从集中管理的RPKI依赖方系统主控节点分发至分布式部署的路由控制系统，其部署节点宜根据网络运行机构所辖自治域的数量和管理机制进行规划，以方便BGP边界路由器在就近获取RPKI验证缓存的同时，在网络运行机构的网络管理边界之内形成一致的RPKI数据视图。数据分发系统部署在数据分发节点之上，并根据该系统所定义的“主从模型”使相关节点（“分发服务器模块”与“分发客户端模块”）形成一个有序的数据共享体系。

7）路由器对接系统

路由器对接系统部署在面向路由器服务的末梢数据分发节点之上。

RPKI依赖方系统主控中心可部署在网络运行机构的NOC之中。北向分布式节点群组的节点数量和分布规则，可结合网络拓扑以及去RPKI资料库之“远近”（路由及寻址）情况量体裁衣。南向分布式节点群组的节点数量和分布规则，可参考网络运行机构的网络互联互通情况和管理机制进行规划设计。

5 总结与展望

RPKI依赖方系统连接RPKI供给侧和RPKI需求侧，是各类网络运行机构开展RPKI应用实践的一个关键环节。RPKI依赖方系统的研发和部署，既需要关注RPKI核心功能的“普遍性”问题，又需要兼顾网络互联互通特征的“特殊性”问题。相关解决方案需要考虑RPKI依赖方系统有哪些组件，各个组件如何在网络上分布，以及以何种逻辑关系分布。因此，各类网络运行机构使用RPKI依赖方系统实施路由认证，不仅是简单的软硬件集成，更需要设计能够“因地制宜”涵盖功能编排、部署方法及运行机制的一揽子解决方案。

面向RPKI依赖方系统的核心功能，本文梳理了影响RPKI依赖方系统运行效能的4对矛盾，并提出了一种可扩展的RPKI依赖方系统部署机制，包含软件层面的解耦机制和硬件层面的部署机制。本文相关论述是对RPKI依赖方系统在规模网络运行机构内进行服务模式设计的宏观思考。骨干网运营商、CDN服务商和互联网交换中心，在互联互通格局和码号资源管理等范畴具有不同特征。面向这些特征，探索如何在现有网络运维管理系统上增量部署RPKI依赖方系统组件以及对应的运行机制，是RPKI路由认证领域下一步值得深入研究的问题。

参考文献

[1] IAB. IAB statement on the RPKI [EB/OL]. [2022-11-25]. https://www.iab.org/documents/correspondence-reports-documents/docs2010/iab-statement-on-the-rpki/
[2] KENT S, LYNN C, SEO K. Secure border gateway protocol (S-BGP) [J]. IEEE journal on selected areas in communications, 2000, 18(4): 582-592. DOI: 10.1109/49.839934
[3] LYNN C, KENT S, SEO K X. 509 extensions for IP addresses and AS identifiers [EB/OL]. [2022-11-25]. http://mirror. cogentco. com/pub/rfc/pdf/rfc3779.txt.pdf
[4] MA D, KENT S. Requirements for resource public key Infrastructure (RPKI) relying parties [EB/OL]. [2022-11-25]. https://www.ietf.org/rfc/rfc8897.pdf
[5] BRUIJNZEELS T, MURAVSKIY O, WEBER B, et al. The RPKI repository delta protocol (RRDP) [EB/OL]. [2022-11-25]. https://www.rfc-editor.org/rfc/pdfrfc/rfc8182.txt.pdf
[6] HUSTON G, MICHAELSON G, MARTINEZ C, et al. Resource public key infrastructure (RPKI) validation reconsidered [EB/OL]. [2022-11-25]. https://www.rfc-editor.org/rfc/pdfrfc/rfc8360.txt.pdf
[7] MA D, MANDELBERG D, BRUIJNZEELS T. Simplified local Internet number resource management with the RPKI (SLURM) [EB/OL]. [2022-11-25]. https://www.rfc-editor.org/rfc/pdfrfc/rfc8416.txt.pdf
[8] BUSH R, AUSTEIN B. RPKI to router protocol [EB/OL]. [2022-11-25]. https://www.rfc-editor.org/rfc/pdfrfc/rfc8210.txt.pdf

作 者 简 介

马迪，互联网域名系统国家地方联合工程研究中心（ZDNS）首席研究员、中国科学院大学研究生导师、国家（杭州）新型互联网交换中心高级顾问、中国电信集团“基于IPv6的电信网络技术北京市重点实验室”学术委员会委员、国际互联网码号资源理事会（NRO NC/ICANN ASO AC）委员、亚太互联网信息中心（APNIC）路由安全SIG联合创始人、国际互联网工程任务组（IETF）域名领域技术标准评审专家委成员、国际互联网名称与码号分配机构（ICANN）域名根服务器咨询委员会核心专家组成员；撰写了3项RPKI相关IETF RFC（RFC8211、RFC8416、RFC8897）；研究方向包括互联网体系结构、互联网寻址定位技术等。

大型企业SASE解决方案及应用实践

SASE Technology Solution and Implementation Practice for Large Enterprise

王茜/WANG Qian，陈晨/CHEN Chen，井俊丰/JING Junfeng，季家震/JI Jiazhen

（奇安信科技集团股份有限公司，中国 北京 100032）
(QI-ANXIN Technology Group Inc. , Beijing 100032, China)

DOI：10.12142/ZTETJ.202301009
网络出版地址：https://kns.cnki.net/kcms/detail//34.1228.TN.20230223.1047.002.html
网络出版日期：2023-02-23
收稿日期：2022-11-25

摘要：针对大型企业的数字化转型和提升安全防护水平的需求，基于安全访问服务边缘（SASE）架构设计了企业一体化安全运营系统Q-SASE及其技术实现方案。Q-SASE不仅可以对分支机构众多的企业提供统一安全防护水平、统一安全访问策略和统一安全运营能力，还可以在后疫情时代支持移动办公场景的终端身份管理和可信安全接入。在大型企业中的落地实践验证了Q-SASE技术方案的可行性和有效性。

关键词：SASE；云安全资源池；零信任；网络安全态势感知；安全运营中心

Abstract: In response to the digital transformation and improvement of the security protection level of large enterprises, based on the secure access service edge (SASE) architecture, an enterprise integrated security operation system Q-SASE and its technical scheme are designed, which can not only provide a unified security protection level, unified security access policies, and unified security operations for enterprises with many branches , but also support identity management and secure access of user equipment in mobile office scenarios in the post-epidemic era. The feasibility and effectiveness of this Q-SASE technical solution have been verified through the implementation practice in large enterprises.

Keywords: SASE; cloud security resource pool; zero trust; network security situational awareness; security operations center

近年来，企业业务系统向云化迁移，信息数据日趋集中化。各种信息化系统趋于集中式建设和分布式服务。新型基础设施如新型广域网、移动互联网、混合云、泛终端、大数据平台的出现，也让信息化系统的建设和运维模式发生变化。同时，网络安全形势日趋复杂，网络攻击手段更为多样。数据泄露、勒索软件、高级可持续威胁（APT）攻击等安全事件频发。相应地，针对这些安全威胁的实战化、体系化、常态化要求也变得越来越高。信息技术（IT）、网络、安全需要统筹管理。同步规划、同步建设、同步运营已成为企业数字化转型的必然要求。

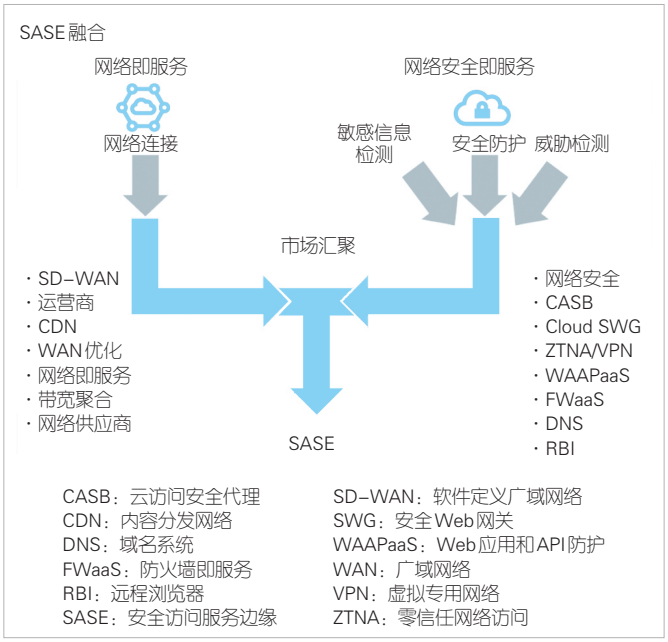
但是，中国的企业信息化网络依然面临着防护不全、投入不足、能力不够、效率不高等问题。尤其是那些具有众多分支机构的大型企业，其分支机构分布广、防护范围广、防护点多，且各分支机构的安全防护能力参差不齐，难以实现统一管理和安全防护。另外，全球疫情的蔓延使办公环境从局域网延伸到居家办公（SOHO）场景。各类自带设备（BYOD）终端已成为企业办公环境的接入边界。漏洞、后门、僵尸木马等针对终端设备的安全威胁日益严重。黑客更容易入侵各类智能设备，渗透企业网络和重要的业务系统，窃取用户数据或者企业经营数据。这将给企业带来直接和间接的经济损失。

基于Gartner提出的安全访问服务边缘（SASE）架构，我们设计了针对大型企业的一体化安全运营系统Q-SASE，通过“软件定义安全”“软件定义网络”“零信任动态评估”等技术实现了整体解决方案，并通过在大型企业的落地实践，对Q-SASE的一体化安全运营的可行性和有效性进行了验证。本研究可作为行业推广的经验参考。

1 基于SASE架构的解决方案

1.1 SASE架构的由来

2019年Gartner在《未来的安全在云端》中提出了SASE的概念。Gartner官方对SASE的定义如下：SASE通过将网络和网络安全的功能融合为统一服务的模式，为企业客户提供一个新的网络安全架构，如图1所示。SASE能够使分支机构人员和移动办公用户高效、安全地就近接入安全节点（部署在云端或者数据中心的PoP点），以访问互联网应用、公



▲图1 SASE技术概念图

有云软件即服务（SaaS）、公司内部应用等。

根据Gartner的定义，SASE是一种基于实体的身份、实时上下文、企业安全/合规策略，以及在整个会话中持续评估风险/信任的服务。实体的身份可与人员、人员组（分支机构）、设备、应用、服务、物联网系统或边缘计算场地相关联。SASE架构将使安全运营以一致和集成的方式提供一组丰富的安全网络服务，从而支持企业数字化转型和业务向云计算的迁移，并满足员工移动办公的需求。

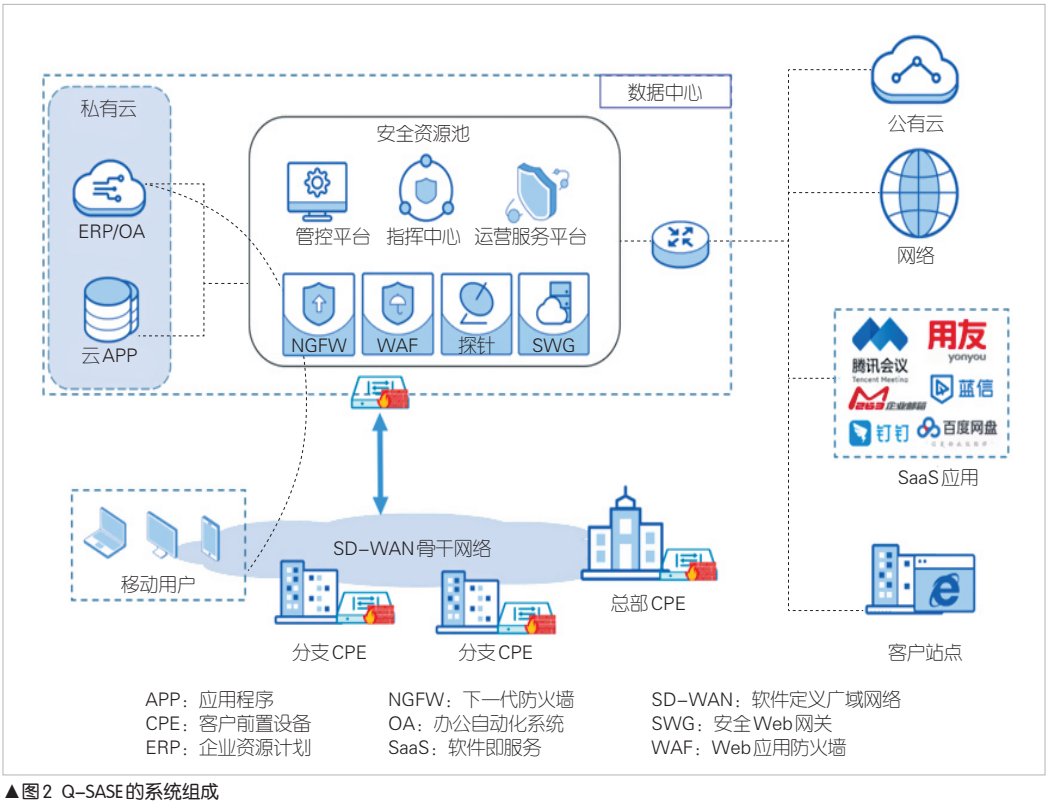
1.2 Q-SASE解决方案

基于SASE架构和零信任理念，Q-SASE解决方案采用“软件定义广域网络（SD-WAN）+软件定义安全+零信任动态信任”的技术路线，实现了针对大型企业的分支机构和移动办公场景下访问互联网、公有云、私有云内部应用的整体安全防护。其中，SD-WAN技术对分支机构的各类访问流量进行组网编排和引流，软件定义安全的云安全资源池对多种访问流量进行安全防护，零信任技术对接入的用户终端进行身份认证和动态访问控制。因此，Q-SASE能够实现组网和安全功能相融合的整体解决方案。

Q-SASE解决方案包括一套安全运营管理服务平台，以及为企业客户建设的云安全资源池，通过在分支机构部署SD-WAN安全网关，以及移动办公终端安装零信任客户端，将访问互联网、内网业务系统的流量引流到安全资源池进行安全防护和安全威胁检测分析，并在威胁检测分析的基础上，提供“安全运行闭环管理并持续监测响应为核心”的安全运营，如图2所示。

1）Q-SASE的安全运营管理服务平台。该平台能够对整个SASE架构中的系统模块进行持续运行监测，以保障整个系统的持续稳定运行。此外，该平台还可对运营人员的权限和工单进行管理，实现对安全告警日志和安全事件的持续跟踪与管理，并基于企业需求针对安全资源池和安全网关中的组网策略和安全策略进行持续优化。

2）Q-SASE的安全资源池。安全资源池采用虚拟化镜像的方式来部署不同的安全组件。安全资源池的安全组件按需配置。安全组件的创建、初始化、激活等操作都由安全资源池来支撑系统自动完成。在不同分支机构的SD-WAN安



▲图2 Q-SASE的系统组成

全网关接入资源池前，安全资源池将支撑系统，使系统按照不同租户角色申请来部署安装相应的安全组件。安全访问服务的云安全资源池采用虚拟化方式部署，可基于接入的分支机构数量和互联网流量规模实现弹性扩容。根据不同企业的需求，安全资源池可部署丰富的安全组件，包括虚拟化防火墙安全组件、上网行为审计安全组件、零信任接入安全组件、虚拟化Web应用防护（WAF）安全组件、日志审计安全组件、态势感知云探针安全组件等。

3）SD-WAN组网及引流。采用SD-WAN技术，安全网关与安全资源池之间可实现快速灵活组网，并支持将分支机构访问互联网应用和内网应用的流量引流到安全资源池以进行安全防护和安全运营。安全网关设备支持零配置开局部署，并支持自动注册及从运营管理平台获取初始化网络配置和安全策略配置，还可通过预配置向导、批量脚本导入、邮件零配置上线（ZTP）、无线网络ZTP等多种方式，实现分钟级零配置上线。安全网关还支持灵活接入能力，可以支持专线接入、互联网以及4G/5G移动网接入。

4）零信任客户端接入。基于零信任客户端对可信访问控制台和可信应用代理的访问，从身份风险、终端风险、网络风险、权限和数据风险5个维度，全面构建从终端到应用访问的端到端安全防护信任评估能力。便捷的运维管理能力和动态访问控制机制，可确保在业务访问的各个阶段都能拥有较好的零信任防护效果。零信任可信客户端对接入终端的用户进行身份认证，支持账号的统一管理与单点登录，拥有权限管理与多因子认证等安全能力；支持对终端的应用环境进行实时监测，即只有通过终端环境信任评估的才能接入政企客户内部网络，例如是否安装杀毒软件、是否升级到最新版本和最新病毒库；基于终端的身份管理，可以依托企业的4A（包括认证、账号、授权、审计）、身份识别与访问管理（IAM）、Windows服务器的活动目录（AD）、轻量目录访问协议（LDAP）、公钥基础设施（PKI）等基础设施，也可以基于企业自建的身份认证中心和应用访问会话，对所有访问请求建立动态访问控制策略。

2 Q-SASE的关键技术实现

基于SASE的创新型架构和内生安全框架，Q-SASE方案采用“软件定义网络”“软件定义安全”和“零信任动态评估”3种技术，不仅实现了SD-WAN技术的灵活组网和引流，还实现了云安全资源池的按需交付和分布式部署，以及零信任的身份管理和信任评估动态控制，并基于实时威胁检测实现了安全风险分析与协同处置。

2.1 软件定义网络技术方案

Q-SASE采用SD-WAN的技术路线，而SD-WAN是基于软件定义网络（SDN）的技术体系发展而来的。Q-SASE实现了SDN管控平台与安全网关的协同工作机制。

SDN采用与传统网络截然不同的控制架构，将网络控制平面和转发平面分离，采用集中控制替代原有分布式控制，并通过开放和可编程接口实现软件定义。SDN技术架构如图3所示。

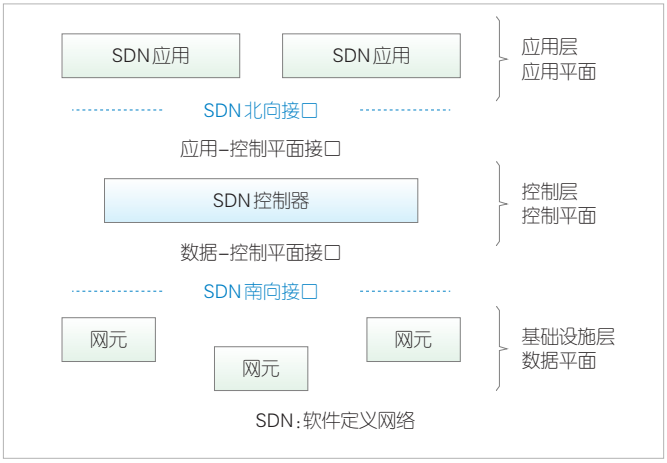
从网络架构层次上看，SDN典型的网络架构包括转发层（基础设施层）、控制层和应用层。该新技术会对组网技术产生以下积极的影响：

1）降低设备复杂度。转发和控制的分离，使得网络设备转发平面的能力要求趋于简化和统一，硬件组件趋于通用化而且便于不同厂商设备的互通。这些都有利于降低设备的复杂度和硬件成本。

2）提高网络利用率。集中的控制平面可以实现海量网络设备的集中管理，使得网络运维人员能够基于完整的网络全局视图实施网络规划，优化网络资源，提高网络利用率，降低运维成本。

3）加速网络创新。一方面，SDN通过控制平面可以便捷地为网络设备制定各种策略，提升网络灵活性；另一方面，SDN提供开放的北向接口，允许上层应用直接访问所需的网络资源和服务，使得网络可以差异化地满足上层应用需求，提供更灵活的网络服务，加速网络创新。

SD-WAN是将SDN技术应用到广域网场景中的一种实践方案。这种方案用于连接广阔地理范围的企业网络、数据中心、互联网应用及云服务，旨在帮助企业降低广域网的开支，提高网络连接灵活性。SD-WAN作为SDN技术体系中的一种可落地的门类，为企业带来了低成本、高可用带宽的



▲图3 软件定义网络技术方案

组网方式。

2.2 软件定义安全技术方案

“软件定义”作为一种理念，可以从网络领域沿用到安全领域。云安全资源池作为Q-SASE解决方案实现的重要载体，其背后的技术支撑正是软件定义安全（SDS）。云安全资源池也是软件定义安全技术的核心应用方向之一。

云安全资源池技术方案的目标在于“随需而变”，而这正符合软件定义安全敏捷、高效、开放的特点。云安全资源池需要运行在云计算环境中，不仅要解决传统安全能力落地的问题，还要能够充分发挥云计算基础设施的功能与优势，实现快速交付、分布式部署、多云（含信创）环境支持、服务链编排等。

在软件定义安全的技术实现中，安全管理控制是重中之重。这是因为安全管理控制承担了所有安全能力的服务抽象、服务编排及调度、策略管理、策略交付等安全核心功能。此外，安全管理控制还需要实现与云平台的深度集成，基于应用程序编程接口（API）获取云上租户资产的关键信息，以便安全管理员部署和管理所需的安全资源。

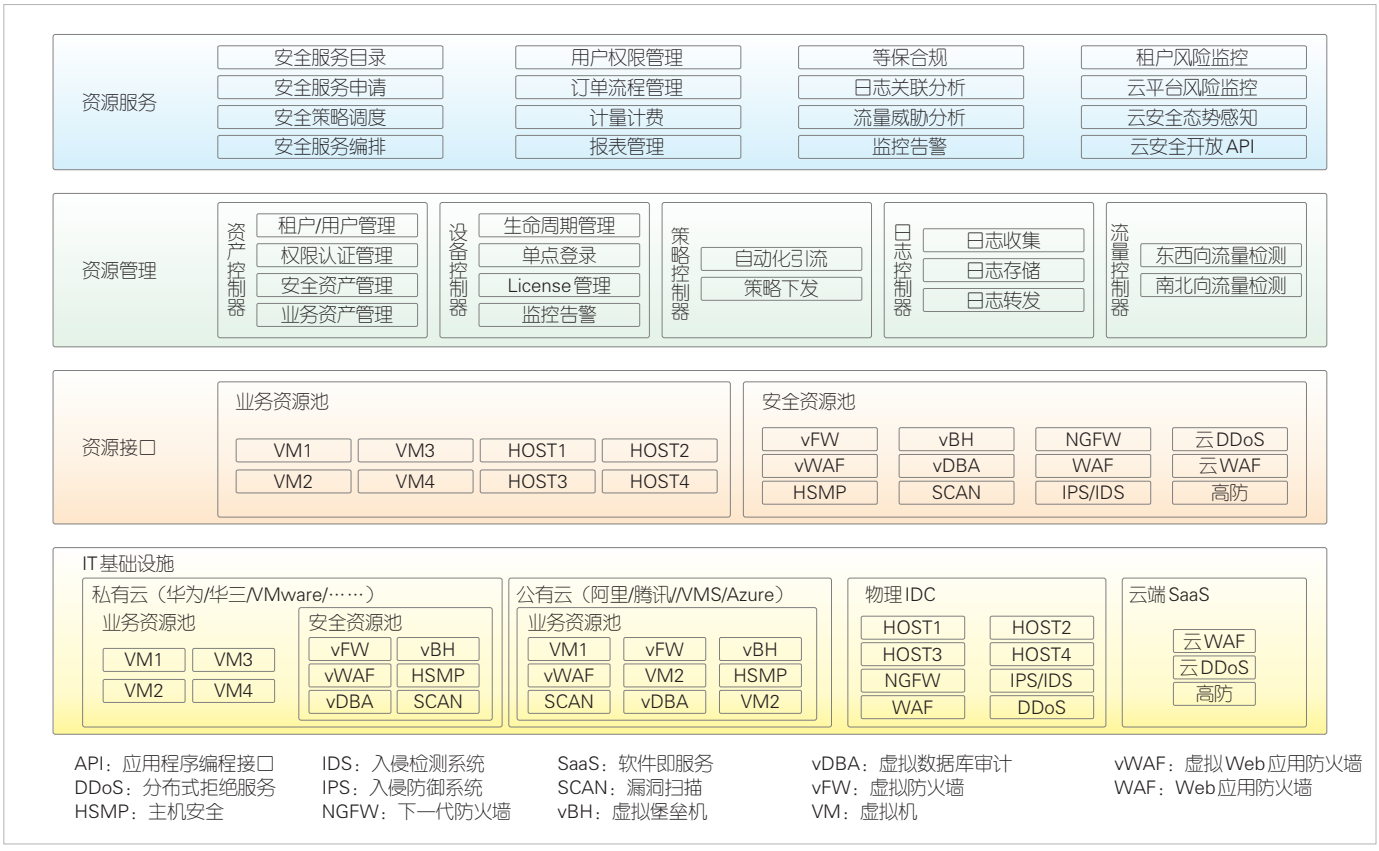
Q-SASE中的云安全资源池，技术方案架构如图4所示。

此外，云安全资源池南向对安全能力完全开放，可通过定义安全组件的统一接入规范来支持各类安全能力，包括第三方安全能力的接入；北向通过开放API支持平台的能力和用户的业务系统深度融合；西向通过定义标准的服务链编排接口支持各种引流设备，包括传统的交换机引流和SDN控制器引流；东向则通过定义标准的云平台对接技术规范来统一支持各种私有云、公有云等云平台的对接，实现云平台租户、资产、用户的同步和管理。

2.3 零信任身份管理及信任评估技术方案

零信任技术方案关注业务保护面的构建，通过业务保护面实现对资源的保护。在零信任方案中，应用、服务、接口、数据都可以作为业务资源。该方案通过构建保护面实现对暴露面的收缩，要求所有业务默认隐藏，并根据授权结果进行最低程度的开放。所有的业务访问请求都应该进行全流量加密和强制授权。业务安全访问相关机制需要尽可能工作在应用协议层。

基于身份而非网络位置来构建访问控制体系，首先需要为接入网络的人和设备赋予数字身份，将身份化的人、设备和应用进行运行时组合构建访问主体，并为访问主体设定其



▲图4 软件定义安全技术方案

所需的最小权限，以进行全面数字化管理。其中，访问主体由用户、设备和应用组合而成。系统会在身份管理的基础上进行持续信任评估，并通过信任评估模型和算法，实现基于身份的信任评估能力，同时需要对访问的上下文环境进行风险判定，对访问请求进行异常行为识别，并对信任评估结果进行调整。在身份信任的基础上，系统还需要评估主体信任。主体信任是对身份信任在当前访问上下文中的动态调整，和认证强度、风险状态和环境因素等相关。身份信任相对稳定，而主体信任和网络代理一样，具有短时性特征，是一种动态信任。

基于主体的信任等级进行动态访问控制是零信任技术方案的本质所在。动态访问控制采用基于角色授权（RBAC）和基于属性授权（ABAC）的组合授权模式。这样便于系统实施灵活的动态访问控制。基于安全基线叠加信任等级可实现分级的业务访问。同时，当访问上下文和环境存在风险时，系统需要对访问权限进行实时干预，并评估是否需要对访问主体的信任进行降级。

2.4 安全威胁分析及协同处置技术方案

在日常的安全运营工作中，真正的威胁往往会被淹没在大量的未确认安全事件中，如低危的防火墙、IDS和WAF告警等。然而，这些告警的分析确认和处置往往会成为令人头疼的问题。传统的安全检测能力主要依托特征库匹配的检测机制。虽然这样能够有效地检测并拦截普通的低级威胁，但也会产生大量的冗余和误报告警。如果不对安全策略和检测机制进行优化，安全运营人员就无法在发生威胁的第一时间判断出哪些威胁会造成严重影响，哪些威胁需要优先处置。

基于大数据架构设计的流式关联分析引擎，能够实时关联多维度数据，结合云端的威胁情报样本，可以针对使用不同日志数据（如入侵防御日志、上网行为日志等）检测内部主机连接攻击者远程命令和控制服务器，进而发现失陷主机的安全威胁，防止由失陷带来的数据泄密、系统破坏等关键风险。基于云安全资源池的本地威胁情报，配合云端威胁情报分析平台进行进一步的分析，了解安全威胁的背景信息，以及攻击者的相关网络资源和历史攻击行为，并进行深入追踪，通过多数据关联分析和威胁溯源，实时提供攻击者上下文信息，提升威胁分析、溯源和协同处置的效率。

3 Q-SASE的应用实践

基于中国电子信息产业集团有限公司（简称中国电子）的一体化安全运营需求，结合企业数字化转型的业务发展目标，我们构建了包括Q-SASE运营管理服务平台，云安全资源池的安全防护组件、零信任组件，以及SD-WAN安全网关和零信任客户端在内的Q-SASE一体化安全运营系统，为中国电子26家二级企业的240家分支机构和超过12万员工的公有云、行业云终端访问业务，提供覆盖云网端的安全防护和安全运营能力。

在中国电子的3类企业信息系统访问场景中，Q-SASE重点实现以下系统建设和安全防护：

1）采用新型SD-WAN技术，建设覆盖全部二三级企业的广域网，并将各分支机构的互联网访问流量进行汇聚，统一实现互联网出口集中管理和安全防护；

2）根据数字中国电子自身业务发展和未来业务系统集中上云的规划，在北京、武汉、深圳等云数据中心部署分布式的安全资源池，具备针对统一互联网出口流量和内部业务系统访问流量的安全防护能力和零信任安全访问能力，对集团总部、所属二三级企业以及新建云数据中心之间的网络通信安全、公有云和行业云业务系统安全、办公访问安全进行有效保障；

3）依托云安全资源池的安全防护组件和零信任组件的能力，以及态势感知的威胁发现能力，通过专业的安全运行团队进行持续巡检监测、故障发现、处置保障、策略优化等安全运行闭环，周期性安全评估安全防护系统平台自身的安全性，提升网络安全攻防演练期间的统一安全防护效果；

4）结合数字中国电子的实际组织架构现状，建立全集团统一安全运营服务中心，将原有纯建设的防护交付模式，演进为以安全服务保安全效果的服务交付模式，从“集团业务全应用场景”的角度出发，全面考虑“集团网络安全职能落地”“各单位网络安全职责落地”所需的工作内容，贴合设计、服务保障。

中国电子是以网络安全和信息化为主业的国有信息技术（IT）企业，也是兼具计算机中央处理器（CPU）和操作系统关键核心技术的中国企业。Q-SASE提供的安全防护和安全运营不仅能够覆盖公共通信和信息服务业，计算机、通信和其他电子设备制造业，还覆盖专用设备制造业、商务服务业、批发业等多个国民经济行业。在基于Q-SASE方案进行一体化安全运营过程中，系统累计发布42期安全周报，下发110份安全事件通告，累计处理74.3万条告警（其中有危急告警8.4万条、高危告警22.7万条）。前10位的攻击类型和所占比例分别为：SQL注入占23.11%，信息泄露占11.82%，代码执行占9.50%，命令执行占4.67%，弱口令占4.04%，暴力猜解占4.03%，跨站脚本攻击占2.59%，网络扫描占2.39%，配置不当/错误占2.29%，非授权访问占1.34%。Q-SASE通过及时分析监测发现威胁及安全事件，同步开展

响应和处置工作，并通过策略编排及时阻断病毒文件、间谍软件横向传播等风险，形成预警及处置报告。Q-SASE方案在中国电子集团总部南迁、重大活动网络安全保障、挖矿行为自查自纠、国家级实战攻防演练等活动中，均取得明显成效。

4 总结和展望

Q-SASE的整体解决方案，将原有分散在各分支机构的网络安全设备集中到云安全资源池，以进行统一建设，实现分支机构的互联网和内网访问流量的收口和统一的安全防护与安全运营。Q-SASE的应用实践表明，Q-SASE方案可以系统性、工程化地实现安全防护和安全运营能力的集中部署、集中运行和统一运营，使大型企业的分支机构快速具备网络安全能力，在疫情常态化后的困境中，让灵活、安全的办公和安全上云的访问成为可能，为企业数字化转型保驾护航。当然，目前的Q-SASE整体方案还处于初级阶段，仍需要通过不断的研究及技术实现，将现有的安全能力以及未来可能增加的安全能力通过编排集成到一起，并且实现安全能力编排化、安全流程自动化、安全运行智能化的升级演进。

参考文献

[1] Gartner. The future of network security is in the cloud [R]. 2019
[2] Gartner. SASE will improve your distributed security everywhere [R]. 2020
[3] Gartner. 2021 strategic roadmap for SASE convergence [R]. 2021
[4] MEF70. SD-WAN service attributes and services [EB/OL]. [2022-11-25]. https://www. mef. net/wp-content/uploads/2019/07/MEF-70. pdf? id=122&fileid=file1
[5] 云安全联盟大中华区. SASE安全访问服务边缘白皮书 [R]. 2022
[6] CCSA. 安全访问服务边缘(SASE)能力成熟度: T/ZGTXXH 048 [S]. 2022

作 者 简 介

王茜，奇安信科技集团SASE产品部总经理、中国通信学会第八届信息通信网络专家委员会委员；先后在中国电信北京研究院、中国电信集团公司、奇安信科技集团从事网络规划与优化、新技术新产品研究等工作，曾参与ITU-T SG13、MEF、IETF等标准工作；有40余篇标准文稿被采纳，发表论文30余篇，出版专著5本，获授权专利6项。

陈晨，奇安信科技集团助理总裁、中国软件行业协会信息主管（CIO）分会副主任委员；负责奇安信科技集团信息管理部的管理工作，先后在中国国航、香港航空、奇安信科技集团从事信息化与数字化管理工作，曾参与《数字化管理师能力评估标准》团体标准的编写与制定工作。

井俊丰，奇安信科技集团总体部网络安全架构师；从事网络安全规划与架构设计、网络安全解决方案与落地支撑等工作，主持或参与多个“十四五”网络安全规划设计与工程落地建设工作。

季家震，奇安信科技集团SASE产品部产品经理；从事SASE产品规划及设计工作。

关于发展中国安全浏览器的建议

Suggestions on Developing China's Secure Browsers

魏小强/WEI Xiaoqiang，张义荣/ZHANG Yirong，
黄亚洲/HUANG Yazhou

（360数字安全集团，中国 北京 100102）
(360 Digital Security Group, Beijing 100102, China)

DOI：10.12142/ZTETJ.202301010
网络出版地址：https://kns.cnki.net/kcms/detail//34.1228.TN.20230222.1735.006.html
网络出版日期：2023-02-23
收稿日期：2022-11-05

摘要：安全浏览器是数字工作空间的安全入口。通过对全球安全浏览器发展趋势的洞察，分析中国在安全浏览器领域所面临的机遇与挑战，提出通过发展中国安全浏览器来构建中国关键数字安全能力的建议。认为应鼓励政企单位使用国产安全浏览器，培育安全浏览器社区文化，加强安全浏览器人才培训。

关键词：安全浏览器；风险；数字工作空间；微软；谷歌；人才培训

Abstract: The secure browser is a secure portal of digital workspaces. Through the insight into the development trend of global security browsers, the opportunities and challenges faced by China in the field of security browsers are analyzed, and suggestions for building China's key digital security capabilities by developing China's security browsers are put forward. It is believed that government and enterprise units should be encouraged to use domestic security browsers, cultivate security browser community culture, and strengthen security browser talent training.

Keywords: secure browser; risk; digital workspace; Microsoft; Google; talent training

安全浏览器已经成为数字工作空间的安全入口。安全浏览器也将成为在操作系统之上保护数字工作空间的新操作系统。微软放弃开发26年之久的网页浏览器（IE），转而拥抱竞争对手谷歌的Chromium开源浏览器内核项目并推出新Edge浏览器（目前排名全球第4）。这进一步刷新了行业对Chromium的认知，也使得跨平台、标准化、统一性、拥抱开源的安全浏览器迅速成为热点。从以色列安全浏览器初创公司Talon Cyber Security获得2022年RSAC创新沙盒冠军，到刚刚成立两年的企业级安全浏览器厂商Island一跃成为估值最高的独角兽企业，以及资本市场对安全浏览器的热烈追捧，可以看出，安全浏览器将推动安全行业的深刻变革。

在俄乌冲突之际，美国网络安全和基础设施安全局（CISA）在2022年2月份发布的“屏蔽”计划（Shields-UP）指出，“每个组织，无论大小，都必须准备好应对破坏性的网络事件……更新手机、平板电脑和笔记本电脑上的操作系统，并更新所有设备上的应用程序——尤其是网络浏览器”[1]。这进一步证明了安全浏览器不仅是数字工作空间的安全入口，也已经成为国家网络防御的前沿阵地。此外，CISA还表示，“利用Web浏览器中的漏洞已成为攻击者破坏计算机系统的一种主要方式”。由此可见，Web浏览器的战略意义已经从为消费者提供互联网“冲浪”的入口变成构建国家网络防御能力的前沿阵地。随着全球加速进入数字时代，数字工作空间无处不在，保护数字工作空间的第一道门户——安全浏览器很可能是未来安全行业的游戏规则改变者。

1 现有网络浏览器存在巨大安全风险

网络浏览器是人们日常使用的用来检索、展示以及传递Web信息资源的应用程序。随着互联网的飞速发展，网络浏览器已成为人们使用最多的应用程序。它不仅是一个互联网访问入口，还是用户业务、资产、信息的接入枢纽。目前，网络浏览器已经发展为新一代数字工作空间，即用户进入互联网空间的访问入口。但是，它并不是为安全而设计的。Chrome、Firefox等浏览器已成为黑客渗透攻击的重要目标，也是许多高级持续威胁（APT）组织常用的攻击入口。近年来，受新冠疫情影响，自带设备（BYOD）办公方式被众多组织大量使用，导致网络浏览器所面临的安全问题更加突出。

1.1 网络浏览器充斥大量易被黑客利用的漏洞

谷歌 Chrome、微软 Internet Explorer、Mozilla Firefox 和 Apple Safari 等网络浏览器安装在几乎全球所有的计算机上。据 Statcounter 数据显示，谷歌 Chrome 浏览器以 63.63% 的份额（拥有 33 亿用户）引领网络浏览器市场；Apple Safari 紧随其后，份额为 19.37%；Mozilla Firefox、微软 Edge（Chromium）、Opera、Internet Explorer 的份额分别为 3.65%、3.24%、2.16%、0.81%[2]。

操作系统附带的网络浏览器往往采用默认的安全设置，例如：Chrome 和 Safari 这样的网络浏览器很容易受到跟踪、恶意工具栏和插件的影响，在其外挂的工具栏和插件中，充满了公共漏洞（CVE），从而使得这些浏览器成为黑客最喜欢的攻击目标。网络攻击者经常利用浏览器上的安全漏洞来窃取机密数据，安装勒索软件并策划网络攻击。

CVE 详情报告显示，所有 4 个主要浏览器在 CVE 方面均排在前 25 位。其中 Chrome 浏览器排在第 9 位，有 2 346 个 CVE；Firefox 排在第 13 位，有 1 933 个 CVE；Internet Explorer 排在第 24 位，有 1 168 个 CVE；Safari 排在第 25 位，有 1 136 个 CVE。Microsoft Edge 于 2015 年亮相，迄今有 250 个 CVE，因此没有进入前 25 位[3]。

统计数据显示，超过 9% 的 Chrome 浏览器扩展功能权限具有高风险或极高风险。2021 年的 Chrome 因被发现 308 个 CVE 而被评为最易受攻击的浏览器。其中，24% 的零日漏洞与 Chrome 有关。由通用漏洞评分系统（CVSS）评估的 Chrome CVE 的风险值（6.4）高于所有其他应用的 CVE 平均值（6）。其他网络浏览器的情况和 Chrome 类似，这里不再赘述。

综上所述，网络浏览器已经成为最容易被攻击的应用程序。然而，企业或机构正在广泛使用网络浏览器来访问其业务应用环境。也就是说，全球有数十亿人的数字工作空间门户正面临着巨大的安全风险。

1.2 网络浏览器缺乏企业级应用所需的安全性设计

Chrome 等网络浏览器往往是为消费者、广告商、内容生产者和开发者而构建的，它也会通过广告、用户跟踪和搜索优化等方式来盈利。因此，面向消费者的浏览器很容易受到网络攻击。比如，攻击者可以遍历浏览器标签，以了解有关受病毒感染主机的丰富信息；通过读取保存用户名和密码的特定文件来获取用户账号信息；利用浏览器安全漏洞和固有的浏览器功能来更改内容，修改用户行为，获取终端系统的最高访问权限并完成横向移动，拦截信息并进行会话劫持；滥用浏览器扩展功能来实现对被攻击者应用系统的持续访问与机密数据获取等。

网络浏览器之所以容易受到上述网络攻击，是因为它并不是为企业级安全性需求而设计的。企业在使用浏览器时会遇到 3 个非常重要而且经常被低估的问题：可见性、可控性和可管理性问题，即企业无法获知用户的访问行为、访问的目的地、访问的风险性、具体的操作行为等。

1.3 非受控端广泛采用网络浏览器进一步加剧企业安全风险

根据美国市场研究公司 Forrester 的一项调查，69% 的受访者声称其网络上一半或更多的设备是 BYOD 等非管理设备或物联网设备[4]。这些设备往往通过网络浏览器接入和访问企业的各种资源。

当 BYOD 访问企业资源时，企业无法在员工自带设备上部署企业级安全策略，而网络浏览器是此类非受控端点接入和访问企业应用的主要途径。这将进一步加剧安全风险，使得企业无法完全对访问内容进行安全控制。事实上，大多数企业的安全团队根本无法验证访问企业资源的非企业设备的安全状况。

但是，BYOD 工作的方式越来越受欢迎，物联网设备正在爆炸性增长。这是数字时代的基本特征。那么如何帮助企业 IT 和安全部门对这些设备实现安全可见性和控制能力？比较好的方法就是采取更轻量级的方式，比如使用安全浏览器等手段来增强可见性和可控性。

2 安全浏览器成为数字工作空间的安全入口和各国关注焦点

如果把网络浏览器比作一辆汽车，那么安全浏览器就是汽车安全气囊。全球数字化转型正在加速发展，为应对混合工作环境下的安全挑战，人们需要一种新的安全工具来保护数字工作空间。安全浏览器很可能将顺应历史潮流担当此任。随着浏览器正在朝着标准化、平台化和系统化演进，安全浏览器正在成为数字工作空间的安全入口和新防线。微软、谷歌以及资本市场等均在加速布局安全浏览器战略，抢占未来数字工作空间安全入口的控制权。

2.1 数字化转型需要新的安全工具来保护用户数字工作空间入口

数字化转型重塑了企业的交付服务和访问应用的方式。比如：BYOD 的广泛采用、企业上下游供应链之间的相互访问以及大量物联网设备的出现，导致企业或机构在混合工作方式下面临的网络攻击面急剧扩大。网络攻击者很容易采用多种方式针对分布式工作的用户发起攻击，例如：利用合作

方未修补的浏览器漏洞，通过浏览器下载的恶意软件，利用浏览器实施欺诈、网络钓鱼以及零日攻击等。传统安全手段仅仅保护可管理设备，无法收敛攻击面，同时还带来高延迟、高成本以及用户体验差等问题。

因此，企业必须找到并实施正确的解决方案，既要实现数据安全性、完全可见性和可控性，还要考虑在混合工作环境下不断提升用户体验等问题。这意味着需要创建一个跨平台、标准化、统一化、易于操作且可实施一致的安全态势管理策略的安全环境，并将其作为保护数字工作空间入口的关键。

随着企业不断把业务迁移到云上，越来越多的企业开始采用基于浏览器的软件即服务（SaaS）应用程序。那么，将安全性集成到浏览器中并创建基于安全浏览器的企业环境就成为一个轻量级的、用户体验良好的选择。安全浏览器的作用方式相当于在传统操作系统之上构建了一个新的基于浏览器的安全操作系统。因此，企业安全浏览器有望成为网络安全行业的“游戏规则改变者”。据统计，到2026年，支持分布式工作的网络解决方案市场规模预计将达到420亿美元[5]。毫无疑问，安全浏览器将是该解决方案中的核心组件之一。

2.2 微软推动Chromium发展，加大数字工作空间入口控制权的争夺

以微软为代表的巨头在Chromium开源内核方面不断发力。微软于2022年6月15日停止IE浏览器服务。该科技巨头果断放弃IE（截至2021年3月市场份额为1.7%），转而与竞争对手谷歌联合推出基于Chromium的Edge浏览器。很显然，该事件将极大地推动Chromium成为浏览器标准，加速行业关于Chromium的共识。

微软于2018年转向Chromium开源内核之后，并不是一个简单的Chromium的跟随者，而是投入其Edge团队的核心力量在Chromium上进行持续开发，并迅速成为全球仅次于谷歌的第二大Chromium开发团队[6]。虽然Edge和Chrome都基于开源Chromium内核，但是微软做了很多差异化的能力开发，基于Chromium底层技术做了很多性能优化，并且引入了人工智能和机器学习技术等，例如：其在Edge中构建所谓的“图灵图像超分辨率引擎”[7]，允许用户增强他们在Web上看到的图像，这使得使用微软的Edge浏览器比Chrome看起来更清晰；在安全能力方面，增加了Smart Screen的功能，能够帮助用户免受网络钓鱼、恶意软件站点和软件的侵害；为用户提供过滤机制，用机器学习智能提醒、过滤、防御钓鱼网站、有潜在危险的网站等[8]。微软Edge团队正在为Chromium开源社区做出巨大贡献，迅速成长为该社区的第二大力量，正在影响并可能引领该社区的发展[9]。

此外，微软还在资本市场收购安全浏览器领域的新锐企业和技术，如对Talon Cyber Security的扶植。以色列企业安全浏览器初创公司Talon Cyber Security获得了2022年RSAC创新沙盒冠军。该公司基于Chromium推出自己的企业安全浏览器产品。Talon Cyber Security在2022年6月与微软签署了排它性创业公司孵化计划，将获得微软的技术、能力和资源的支持。可以看出，微软正在借力Talon的安全浏览器技术扩大其企业安全浏览器战略，争夺数字工作空间入口的控制权。

事实上，目前资本市场对安全浏览器领域表现出极高的兴趣。我们注意到，最近两年涌现出了一些独立的安全浏览器厂商，比如：另一家以色列安全浏览器公司Island，在2022年3月完成B轮融资，共筹集到1.15亿美元，以13亿美元的估值成为历史上首个估值最高的安全浏览器独角兽企业。该公司的主打产品也是基于谷歌Chromium内核的企业安全浏览器。

值得注意的是，当前，中国非常依赖微软IE的浏览器插件，例如：利用基于IE等标准浏览器兼容和扩展的一些国密产品的安全加密功能，来保护金融领域电子交易等。因此，微软放弃IE浏览器将对中国原有基于微软IE扩展的兼容加密技术的升级等产生重要影响。

2.3 谷歌发布Chrome企业连接器框架，加速布局安全浏览器战略

为了应对混合工作环境下的网络安全挑战，谷歌在其Chrome浏览器和Chrome OS的基础上，于2022年5月26日推出了Chrome浏览器和Chrome OS企业连接器框架（Chrome Enterprise Connectors Framework）。该框架主要在Chrome OS中为企业增强数据控制能力，从而更好地保护企业环境中用户和设备的安全，同时为安全团队提供更多工具来报告和管理安全事件。

目前约有10家硬件和安全头部企业响应谷歌这一框架计划，例如：英特尔推出vPro设备管理工具包，以加密Chrome OS设备的内存；Palo Alto Networks、CrowdStrike Holdings、BlackBerry统一终端管理（UEM）、VMware、Splunk以及三星Knox Manage等均承诺会尽快推出基于谷歌企业连接器框架的安全方案。

可以看出，谷歌正在借助合作伙伴的力量构建从芯片到云的全链路安全方案。企业安全浏览器作为数字工作空间的

入口具有举足轻重的战略意义。

3 相关建议

安全浏览器是构建在操作系统之上的操作系统，具有巨大的平台价值和极其重要的战略意义，需要我们高度重视。目前，针对中国浏览器使用情况的调研显示，中国安全浏览器在开发方面具有先发优势，但是也面临着很大挑战：人们还未认识到安全浏览器的重要战略意义，企业对安全浏览器的付费愿望不足，安全浏览器研发人才短缺等因素正在阻碍中国安全浏览器的发展。承载中国党政军企各类业务的系统均通过网络浏览器进行访问，浏览器的安全性已关乎到国家网络空间的安全。为加速推动我国安全浏览器的发展，我们提出3个相关建议。

3.1 发挥先发优势，鼓励政企单位使用国产安全浏览器

在安全浏览器领域，中国企业具有一定的先发优势。如360公司在2018年提出了企业安全浏览器的概念，并最早完成了规模化商业产品落地。该安全浏览器支持Windows全系列、MacOS全系列、信创平台全系列的操作系统。在信创平台的应用级产品层面，360企业安全浏览器与中国100多家应用厂商的产品实现了全面兼容。此外，中国企业具有的先发优势还表现在：

1）积极跟进了与浏览器相关的国际标准编制工作。2012年，以360为代表的中国安全公司加入万维网联盟（W3C）中最重要的超文本标记语言（HTML）工作组，与W3C共同研究和制定HTML5等新互联网标准；随后，加入全球CA/B根证书信任联盟。

2）发起了安全浏览器根证书计划。早在2018年，360公司就正式启动了公开密钥加密算法RSA（由发明者Rivest、Shmir和Adleman姓氏首字母缩写而来）根证书计划，与其他国家操作系统中的根证书库认证体系脱离，构建了自有的根证书审查机制，目前已完成全球100%权威CA机构的入根工作。在2020年360公司又启动了国密根证书计划。目前该计划已获得数十家中国CA机构的入根工作。这一举措是具有前瞻性的，这是因为在发生极端冲突时，一旦其他国家CA机构吊销用户网站证书，用户还可以利用360安全浏览器内嵌的安全可信根证书计划，立即重新建立网站访问信任，防止被“卡脖子”。

3）在技术上具有一定的创新性。和国际安全企业浏览器厂商相比，360安全浏览器具备大量的内置安全功能，包括数据丢失预防（DLP）功能、远程浏览器隔离（RBI）功能、安全访问控制功能、文件加密功能以及细化的认证和授权控制。该浏览器可面向Windows、Mac、信创等多个平台，提供统一的跨平台集约化管理方案，内置兼容性检测和修复工具，扩展应用商店、数据开放接口，提供安全方案咨询及定制化开发服务。

我们建议有关部门创造有利条件，发挥中国浏览器厂商已经取得的上述优势，从战略上重视并帮助安全浏览器厂商扩大生存空间，在数字经济、数字政府、数字城市等重要项目建设中，明确要求采用企业安全浏览器并限期全员部署，以保护中国数字工作空间入口。

3.2 发挥新型举国体制优势，打造自主战略级安全浏览器

浏览器是数字空间入口。一旦“断供”事件发生，中国数字空间的门户将敞开，因此我们建议应尽快制定中国企业安全浏览器的B计划（在积极跟踪Chromium浏览器开源内核的基础上，防止在未来极端情况下所面临的“卡脖子”风险），发挥新型举国体制优势，支持以科技领军企业为龙头，推动政府、市场与社会有机结合，集中各方力量进行攻关突破，打破技术垄断，从根本上解决中国各类数字应用的入口安全问题。

中国安全浏览器要获得长远发展一定离不开安全社区的建设。谷歌在开源社区方面的成功经验很值得中国借鉴：不论是实力强大的Chromium开源社区，还是开源供应链社区，均展现出娴熟利用社区力量推动技术和产业创新的卓越实践能力。未来，创新一定源于群体的力量，也一定驱动于不断提升用户体验的使命感。所以，培育安全浏览器社区文化，鼓励开发者创新，在创新中再回报开发者，才能获得真正的技术竞争优势。

3.3 加强安全浏览器人才培训，鼓励开展浏览器安全基础技术研究

目前，中国的安全浏览器是基于谷歌Chromium内核来开发的。Chrome浏览器拥有约2 300万行代码，因此构建企业浏览器是一件相当复杂的工作，需要大量安全专业人员投入和协同攻关。比如，微软Edge团队目前拥有1 000多名开发者，谷歌Chrome浏览器拥有一个2 000多人的开发团队。据粗略估算，中国从事浏览器开发工程师的数量不会超过500人，而具有浏览器内核开发能力的人才更少。实际上，目前中国安全浏览器开发方面的技术水平仅仅能跟上Chromium开源代码更新的速度。安全浏览器开发人才资源的严重不足将极大制约中国在该领域的创新和发展。因此，我们建议国家应加快制定针对安全浏览器的人才培养计划，并从科研项目、应用示范、

力量整合等方面加强浏览器安全基础技术研究，尽快实现中国在浏览器安全基础技术领域的突破。

参考文献

[1] CISA. Be cyber smart: get your “Shields Up” simple steps for safety online [EB/OL]. (2022-02-14) [2022-11-20]. https://www. cisa. gov/sites/default/files/publications/CISA_fact_sheet_4_things_Cyber_English_508.pdf

[2] Statcounter. Browser market share worldwide [EB/OL]. [2022-11-20]. https://gs.statcounter.com/browser-market-share

[3] MITRE corporation. CVE details [EB/OL]. [2022-06-06] [2022-11-20]. https://www.cvedetails.com/top-50-products.php

[4] Forrester. State of enterprise IoT security in north America: unmanaged and unsecured [EB/OL]. [2022-11-20]. https://info.armis.com/rs/645-PDC-047/images/State-Of-Enterprise-IoT-Security-Unmanaged-And-Unsecured.pdf

[5] Team8. Talon cyber security raises $26 million to develop next-generation cyber security for a distributed workforce [EB/OL]. [2022-11-20]. https://team8. vc/press-release/talon-cyber-security-raises-26-million-to-develop-next-generation-cyber-% E2%80%8B% E2%80%8Bsecurity-for-a-distributed-workforce/

[6] Wikipedia. Chromium (web_browser) [EB/OL]. [2022-11-20]. https://en.wikipedia.org/wiki/Chromium_(web_browser)

[7] Microsoft. Microsoft edge and bing maps [EB/OL]. [2022-11-20]. https://blogs. bing. com/search-quality-insights/may-2022/Turing-Image-Super-Resolution?s=09

[8] Microsoft. How can SmartScreen help protect me in Microsoft Edge [EB/OL]. [2022-11-20]. https://support. microsoft. com/en-us/microsoft-edge/how-can-smartscreen-help-protect-me-in-microsoft-edge-1c9a874a-6826-be5e-45b1-67fa445a74c8

[9] Microsoft. Chromium [EB/OL]. [2022-11-20]. https://microsoft. fandom. com/wiki/Chromium#Contributors

作 者 简 介

魏小强，360集团天枢智库研究员、国际云安全联盟（CSA）大中华区多云工作组组长、以色列Trusteer亚太区总经理、IBM大中华区高级安全专家、加拿大Entrust亚太区技术总监；拥有20年产品开发、运营管理、投融资、创业等经验；在边缘计算、零信任、多云安全、SASE、XDR、行为意识安全等领域拥有软著、专利、编著等20余个，发表论文多篇。

张义荣，360集团天枢智库负责人、高级工程师，中国网络空间安全协会个人信息保护专家组、网络安全产业统计调研专家组专家；具有16年以上行业经验，主要从事网络安全总体规划、体系设计、技术跟踪和产业研究等工作；完成国家自然科学基金、部委预先研究及重点工程建设等项目40余项；获部委级科技进步奖一等奖1项、二等奖4项、三等奖2项；发表学术论文30余篇。

黄亚洲，360数字安全集团浏览器业务线负责人；拥有17年产品开发和团队管理经验，擅长团队管理、战略规划、产品定义、市场布局等，开辟中国企业级浏览器品类赛道。

新型家庭全光网技术

New Home All-Optical Network Technology

王新余/WANG Xinyu，孔雪/KONG Xue，贺峰/HE Feng
（中兴通讯股份有限公司，中国 深圳 518057）
(ZTE Corporation, Shenzhen 518057, China)

DOI：10.12142/ZTETJ.202301011
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230227.1353.006.html
网络出版日期：2023-02-27
收稿日期：2023-02-10

摘要：提出了光纤到房间（FTTR）新型家庭全光网相关技术。围绕底层支撑、网络管理和业务保障3类技术，中兴通讯取得一系列技术创新和应用改进：研发FTTR无源光网络（PON）芯片，实现主从网关的点到多点（P2MP）架构功能；编制家庭光纤施工技术指导书，解决FTTR落地推广中光纤分布式网络（ODN）布线施工难的问题；打造智能管理云平台，实现FTTR网络的管理，包括网络拓扑管理和性能数据采集；提出五维评分体系和无线接入终端切换模型分析，使全屋漫游“选得好”且“切得快”，漫游切换时延低至20 ms。指出家庭网络在飞速演进，FTTR会面临更多挑战，中兴通讯将持续进行FTTR技术攻关，推动家庭网络新基建。

关键词：FTTR；PON；管理；深度包检测

Abstract: New home all-optical network technologies related to Fiber to The Room (FTTR) are proposed. Based on three categories of technologies covering the bottom infrastructure, network management, and service assurance, ZTE has achieved a series of technological innovations and application improvements: The FTTR passive optical network (PON) chipsets are developed to realize the point-to-multipoint (P2MP) architecture functions of the master and slave gateways; the technical guide of home optical fiber construction is formulated to solve the difficulties in optical fiber distributed network (ODN) cabling construction in FTTR implementation; a smart management cloud platform is developed to implement manageability of the FTTR network, including network topology management and performance data collection; the five-dimensional rating system and STA handover model analysis are proposed to realize the whole-home roaming "selected well" and "cut quickly", building the roaming handover delay as low as 20 ms. It is pointed out that the home network is evolving rapidly, and FTTR will face more challenges. ZTE Corporation will continue to tackle the key problems of FTTR technology and promote the new infrastructure of home network.

Keywords: FTTR; PON; management; deep packet inspection

1 FTTR背景和概念

近几年，居家办公、在线网课、网红直播、4K/8K超高清视频、增强现实（AR）/虚拟现实（VR）[1-2]、云游戏等新业务层出不穷，对网络的带宽、时延、抖动等都提出了更高的要求。基本宽带已无法满足现代家庭的多元业务需求。

在中国“双千兆”战略和“十四五”规划的明确指引下，光纤入户要向用户端进一步延伸，以实现光纤到房间（FTTR）、到桌面、到机器[3]。

在业务和政策的双轮驱动下，家庭全光网络FTTR相关技术应运而生。FTTR将光纤部署进一步延伸到家庭和小微企业内部的房间，实现全屋超千兆的网络覆盖。基于无源光网络（PON）点到多点（P2MP）物理架构，FTTR由主网关、从网关和室内光纤分布式网络（ODN）3部分组成。主网关上行连接光线路终端（OLT）的PON口，下行经分光器及光纤连接多个从网关，为每个区域提供高速可靠的网络，实现全屋高质量的网络覆盖。

2 FTTR技术创新研究

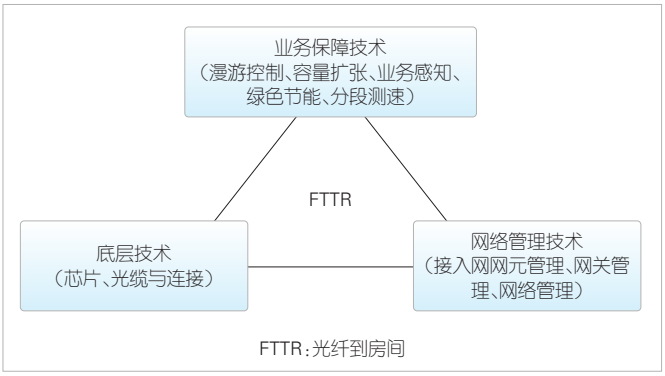
中兴通讯围绕FTTR开展了深入研究，目前已实现多项技术创新。其中，有些技术创新已取得专利认证，有些创新技术已经在实际产品中落地应用。

FTTR技术主要分为底层技术、网络管理技术、业务保障技术三大类，如图1所示。基于核心软硬件能力和算法创新，我们做了一系列的技术创新和方案改进。

2.1 底层技术

2.1.1 芯片技术

FTTR是基于P2MP架构的家庭PON网络技术。相比于传统家庭网关，主网关具备类OLT功能，对下挂从网关具有光路拓展、从网关注册和业务分配等局端主控功能。这对主网关内核心芯片的架构设计、处理能力、功能扩展等都提出



▲图1 FTTR技术架构

了更高的要求。

中兴通讯研制的PON芯片能够匹配主从网关的P2MP架构和技术功能。目前中兴通讯已研发多款FTTR PON芯片，是首家支持XGS-PON/XG-PON/10G-EPON/GPON全系列上行制式的厂商。当前，FTTR主网关采用现场可编程门阵列（FPGA）芯片，实现主网关的OLT化功能，但这种方式存在功耗大、成本高等问题。因此，中兴通讯已启动新一代芯片的开发。新一代芯片设计功耗更低、成本更优，可帮助FTTR技术实现大规模推广应用。

2.1.2 光缆与连接技术

相比于传统的宽带入户，FTTR需整体考虑室内ODN的网络规划和布线施工，这也是FTTR技术得以落地商用的重要环节。ODN网规主要分为普通场景和大户型场景。普通场景不超过4个房间，采用1∶4的单级ODN，可支持4个从网关的接入；大户型场景包括大平层、复式、别墅等，房间数大多超过4个，采用不等比分光的多级级联ODN（例如1∶5分光器级联，最多可支持4级级联、16个从网关接入）。

为了满足不同户型的ODN布线需求，可采用弯曲损耗不敏感的G.657.A2或G.657.B3光纤。施工方式具体包括暗管和明线两种：对于新装修房屋，可采用蝶形光缆进行暗管施工，利用管道内旧线缆抽拉引导布放，或者借助橄榄头弹簧穿管器进行穿管；对于已装修房屋，可视情况采用暗管或明线施工，明线施工采用隐形光缆和短线槽部署，以确保整体家居的美观度。

据此，中兴通讯编制了《家庭光纤施工技术白皮书》，详拟了FTTR标准化的网络规划和施工步骤，提供了工程物料、辅助工具的详细清单，以及施工常见问题（FAQ）等，解决了落地推广中ODN布线施工难的问题。

2.2 网络管理技术

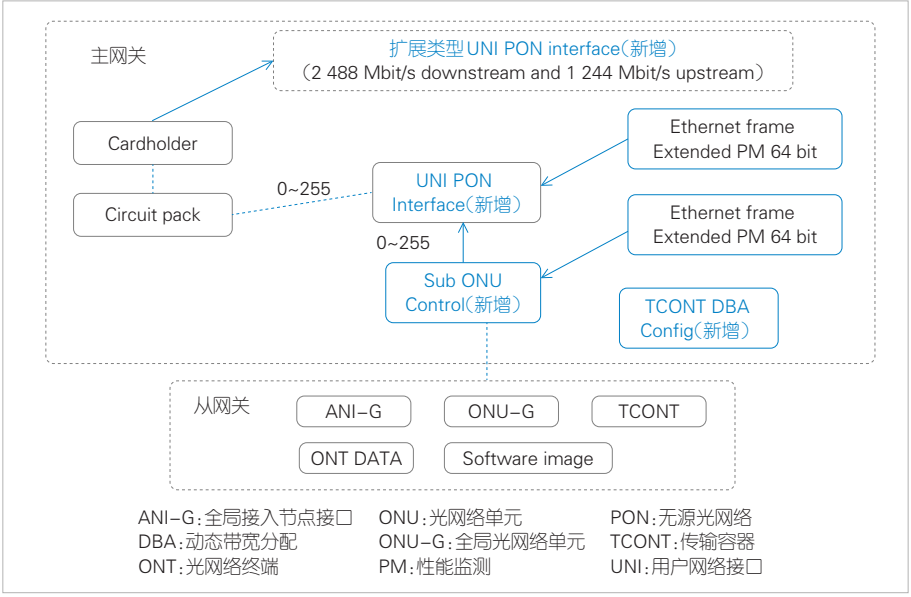
FTTR技术架构使家庭组网演变成了一个小型网络接入系统。如何实现对从网关的可管理、可检测、可运维，是FTTR能否成功推向市场的关键。

FTTR远程管理由3个管理单元组成：接入网网元管理单元、网关管理单元、网络管理单元。其中，接入网网元管理单元负责FTTR网络PON层的管理，网关管理单元负责主网关的网关业务功能管理，网络管理单元负责FTTR网络的管理。三者协同管理，实现FTTR网络的主动管理、主动诊断、主动质差和智能调优。

2.2.1 接入网网元管理单元

接入网网元管理单元基于OLT代理方式实现PON层的管理，采用光网络单元管理控制接口（OMCI）/操作管理维护（OAM）协议[4]。传统光纤到户（FTTH）的OLT管理模型只能管理主网关，无法管理从网关。对此，中兴通讯提出两大创新技术，使OLT可直接管理从网关，并拥有网关的注册、动态带宽分配（DBA）、告警、性能统计、组播、版本升级等功能。

1）定义了主网关下联PON口与从网关的OMCI管理模型，如图2所示。通过光网络单元管理控制通道（OMCC）



▲图2 主网关下联PON口与从网关的光网络单元管理控制接口管理模型

（PON原生OAM通道），OMCI消息直通从网关，从而实现OLT对主网关下联PON口和从网关的全管理。

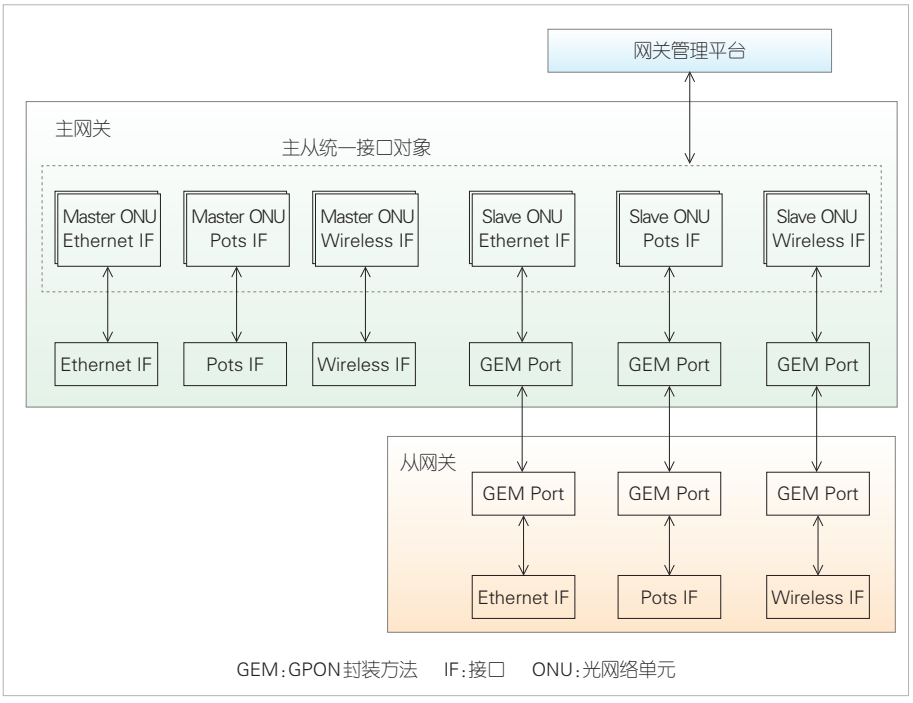
2）定义了用户网络接口（UNI）PON Interface ME和Sub ONU Control ME。OLT发现从网关并建立直接管理通道。OLT负责下行，主网关负责上行，在发送信息前，先发送点灯消息，以告知后续信息所属，如图3所示。此项创新技术不仅降低了FTTR PON层的管理复杂度，还提升了管理效率。相关成果已形成专利[5]。

▲图3 信号灯消息

2.2.2 网关管理单元

和接入网网元管理单元一样，网关管理平台默认只管理主网关的业务，无法管理从网关。基于TR-069协议，我们创建了从网关上报接口（IF）对象与主网关分配GEM（GPON封装方法）Port一对一的映射关系表。主从网关自身的虚拟IF对象和实际IF对象一对一映射，按照统一的命名和编码规则对外呈现。因此，通过主网关上报IF虚拟对象，网关管理单元可以管理从网关的业务。

如图4所示，该创新技术使FTTR主从网关的二级架构简化为一级架构，不仅优化了主从网关管控和转发架构，而且降低了管控复杂度和业务转发时延。该架构兼容现有管理协议和接口，有利于未来FTTR的互通开放和规模化部署。相关成果已形成专利[6]。



▲图4 FTTR网关管理平台架构创新

2.2.3 网络管理单元

上述两个管理单元主要是单个网关设备层面的管理，无法进行FTTR网络的管理运维。中兴通讯研发的智能管理云平台可实现FTTR网络两个方面的管理：一是FTTR主从一张网的网络管理，该管理主要基于消息队列遥测传输（MQTT）/JSON（JS对象简谱）协议，包括网络拓扑管理、Wi-Fi信息管理和配置；二是FTTR主从一张网的性能数据采集管理，该管理基于MQTT/GPB（gRPC ProtoBuf编码格式）协议，包括信息协同收集、性能与告警上报、协同漫游与调优、交互式网络电视（IPTV）业务配置和版本升级。

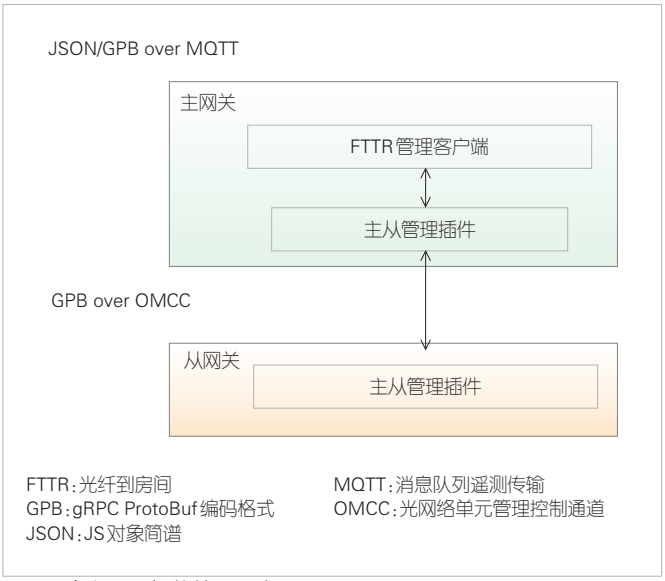
如图5所示，我们创新定义了主从网关管理插件，通过主网关代管从网关的方式，实现智能管理平台对主从一张网的管理，采用OMCC和高性能编码协议GPB，自定义结构化数据传递，保证了同步的实时性。GPB协议对采集的带宽需求仅为JSON协议的1/3，因此大大减少了对主网关下联PON口带宽的占用。

此项创新技术填补了FTTR网络管理的空白，提高了管理能力和效率，相关成果已形成专利[7]。

2.3 业务保障技术

2.3.1 漫游控制技术

无线接入终端（STA）在Wi-Fi网络中移动时，可以在



▲图5 中兴通讯智能管理平台

各网关接入点（AP）之间无缝切换。这解决了大户型家庭和小微企业场景下的Wi-Fi全覆盖问题。当用户终端在家中随意移动时，如何确保网络不掉线和业务不中断？对此，基于多AP协同与整网感知技术，中兴通讯创新性地提出了智能决策技术与快速切换技术，实现STA的全域无缝漫游，业务不中断，用户无感知。

智能决策技术主要解决“何时漫游”与“漫向何处”的问题，采用动态检测技术，结合STA漫游历史，综合判断漫游时机，并预防漫游乒乓现象的发生，如图6（a）。漫游目标的决策是FTTR漫游技术的核心，会直接影响用户的接入带宽。对此，我们提出五维评分体系，以决策出最佳漫游目标。

1）接收信号强度指示（RSSI）策略。如图6（b），当在网络中移动时，STA与不同AP之间的信号强度持续变化。该策略基于整网感知技术，综合多个AP与STA的测量信息，选择信号质量较好的AP作为切换目标。

2）负载均衡策略。如图6（c），当AP或射频的接入用户数较多时，单用户的上网速率会大幅下降。负载均衡策略会平衡STA在AP及射频上的分布，提升网络的利用率。

3）流量分布均衡。如图6（d），部分用户在进行大数据传输时会占据大量的带宽，此时选择流量较小的AP进行漫游，可避免单AP上的传输竞争。

4）层级策略。如图6（e），FTTR存在多级组网或无线组网的情况。对于处于尾端层级的无线组网AP，接入体验相对较差。因此在进行漫游决策时，应尽量避免选择此类AP。

5）5 GHz频段优先策略。大量的实测和分析研究表明，5 GHz频段的接入体验在多数时候都优于2.4 GHz频段。因此，在同等条件下可优先选择漫游到5 GHz频段。

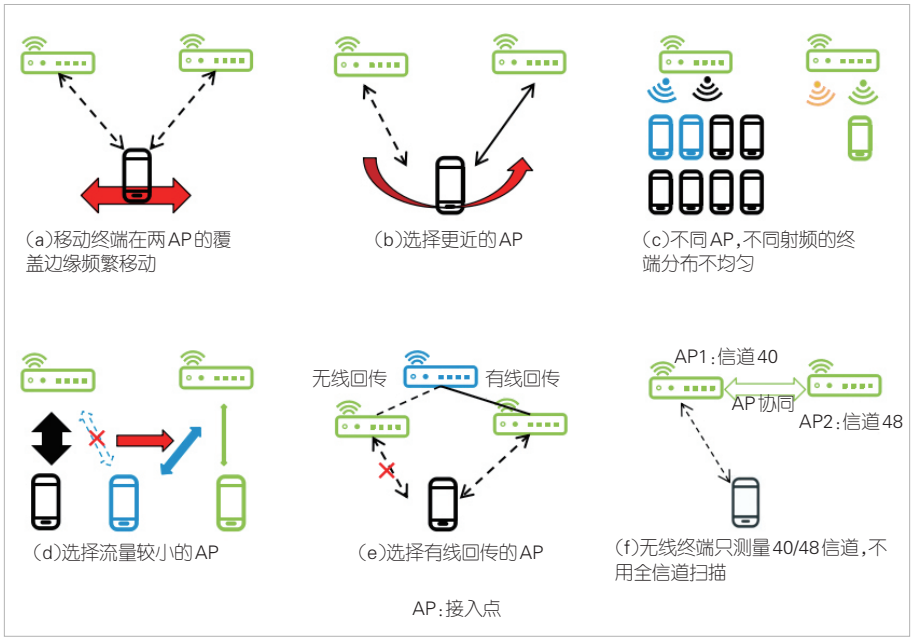
快速切换技术主要解决漫游过程中“不切换”和“切换慢”的问题。独创的STA切换模型能够分析STA频段、802.11 k/v/r、Wi-Fi模式等，判断最优的漫游切换方式；结合组网协同，选取最小的信道集合进行802.11k测量扫描，如图6（f），将漫游切换时延降低到20 ms。

上述漫游控制技术方案已经在实际产品中落地应用，其中部分研究成果已形成专利[8-9]。

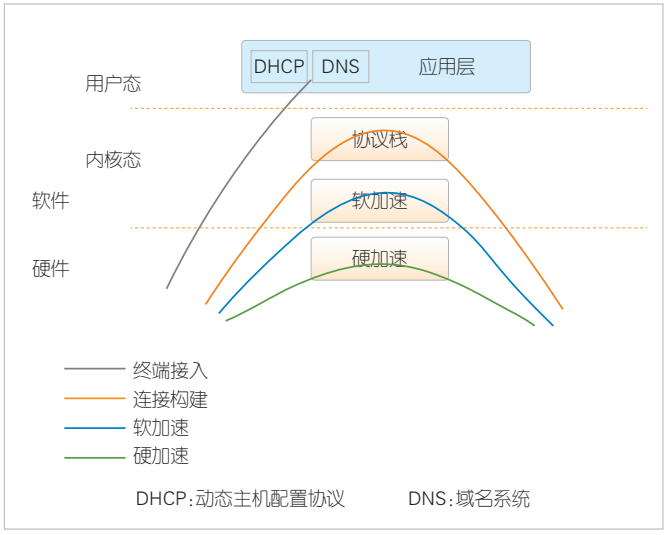
2.3.2 容量扩张技术

未来，智慧家庭场景会有多终端同时接入网络的情况，例如中小企业中两三百个终端同时接入网络。这就需要扩展传统网络的容量，在大量数据报文高并发下，保障FTTR的网络性能。

终端接入流程如图7所示。在第一阶段，系统在终端接入后进行动态主机配置协议（DHCP）地址获取和域名系统（DNS）处理，通过协议栈处理业务报文，构建连接；在第



▲图6 漫游决策与切换



▲图7 终端接入及报文转发流程

二阶段，系统对报文进行加速处理（先软加速后硬加速）。

1） 连接构建优化

终端接入时以IPv4网络为主。我们对DHCP地址分配和DNS转发两个关键技术进行改进，优化了接入流程，提升了效率。

当大量终端接入时，DHCP地址分配过程会出现大量discovery消息堆积，导致超时且无法完成DHCP地址分配。为此，我们提出了两项研究改进：

一是采用报文分类队列调度技术。我们将DHCP报文按类型分类（如discovery、request等）并将其纳入不同队列，随后按照一定权重出队列（优先处理request，然后处理discovery），以此解决消息堆积导致的超时问题。当DHCP地址池容量超过C类地址段时，可采用跨段分配机制，扩充地址容量。

二是采用三重重复地址检测技术。相比于传统的单种重复检测技术，我们使用地址解析协议（ARP）检测、网际报文控制协议（ICMP）检测和邻居表信息对要分配的地址进行三重重复检测。该方法能大幅提高地址可用性的检测准确度，兼容有线和无线多种链路，匹配原先网络的地址状态。

在DNS转发处理上，我们通过3个方面的优化来提升DNS转发性能：

一是DNS程序的flowcache和dnscache容量扩展：flow - cache容量扩大到原先的4倍，单DNS请求记录调整为原先1/4，并进行Hash查找优化；dnscache容量也扩大到原来的4倍，采用环形队列和Hash查找方式。

二是DNS转发路径优化：配置数据预读，降低频繁读取文件，提升查找效率。

三是DNS功能模块化标准化：注入DNS解析流程，减少DNS重复解析。

2）报文加速优化

我们改进了加速条目容量和快速查询算法，实现了FTTR网络的容量扩充和性能提升。

软加速将报文直接转发，避免走协议栈。我们将软加速的连接跟踪数和条目增加到原来的8倍，同时引入环形队列和Hash算法以实现快速查找。软加速学习到连接条目后，将其配置到硬加速。报文直接经硬件转发，不通过CPU。将硬加速条目数也扩大8倍，采用Hash算法快速查询，可避免通过五元组在整个表中进行对比。

2.3.3 业务感知应用识别技术

面对日益丰富的家庭应用，应用流量的精确控制是网络管理面临的重大挑战。精确识别是实现精确控制的必要前提。业务感知应用识别技术[10]通过提取报文中特定字段或报文的行为特征，与业务感知特征库进行匹配，进而识别应用。相比于传统的协议识别技术，该技术适用范围更广，智能化程度更高。

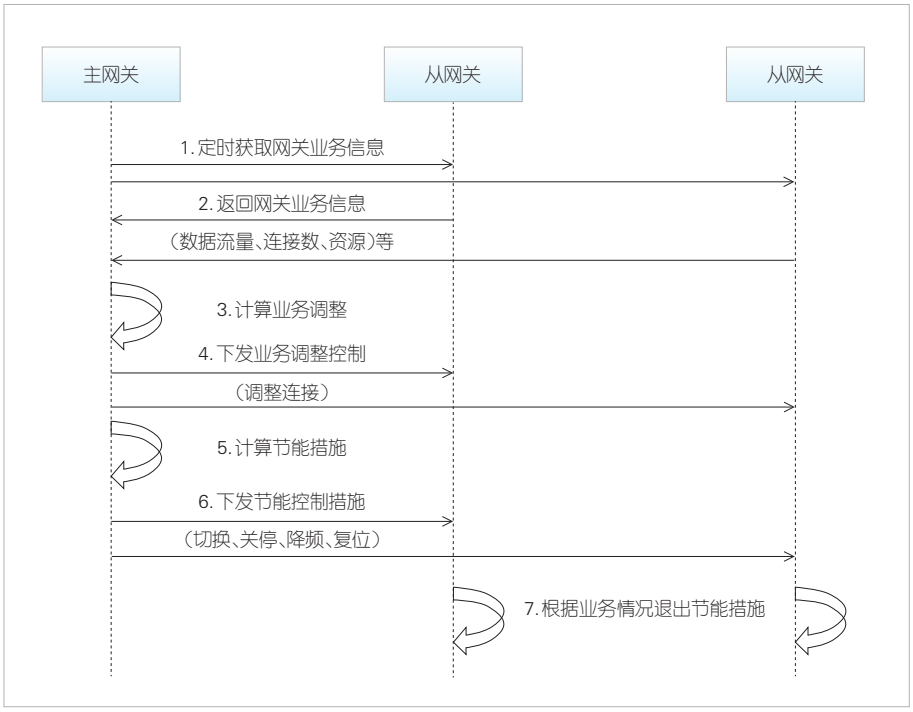
中兴通讯研发了业务感知应用识别插件ZXDPI。ZXDPI持续从互联网应用中提取常见应用的特征，并将其存储到智能管理云平台的业务感知特征库中。特征库保持动态更新，避免APP应用对深度包检测（DPI）插件的反制。FTTR网关也定期同步，动态加载新增的特征库。当业务流量进入网络后，ZXDPI进行应用分类和业务流分析，将分析结果和网关的特征库进行对比，识别出该应用程序，进而实施精细化服务质量（QoS）策略控制。

目前该技术实现了游戏、视频、办公、支付、社交5类应用的智能管控，已经在实际产品中落地应用。

2.3.4 绿色节能

FTTR组网网关较多，容易出现流量不均衡的情况。有些从网关的业务流量大，有些从网关的流量较小甚至没有流量。对此，我们开发了绿色节能机制，对整网流量进行动态智能分析，在用户无感知的情况下实施节能控制，实现网络负载均衡和整体能耗最低，如图8所示。主网关定时获取所有从网关的业务信息，计算出整网的业务分配和从网关的资源占用情况，并制定业务平衡调整策略；从网关调整好自身业务，并执行节能控制。如果从网关的业务情况发生变化，从网关可及时退出节能控制，以确保不影响用户的业务。

基于大量的研究测试，我们形成了3个层面的节能控制措施：



▲图 8 绿色节能机制流程图

1）LAN 口层面：进行快速以太网切换或关停操作。

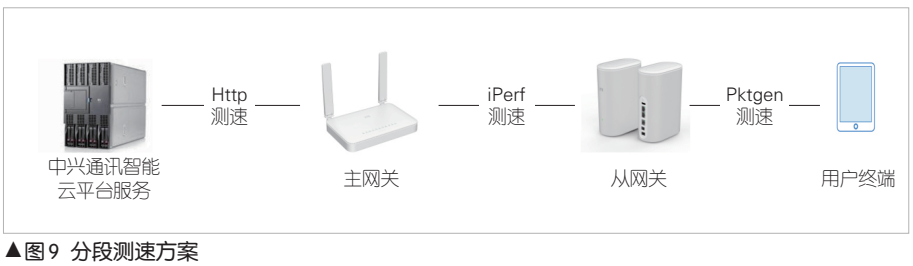
2）Wi-Fi 层面：进行多输入多输出（MIMO）切换或关停操作。

3）芯片层面：进行降频或复位操作。通过多轮优化与验证，我们将 FTTR 网关的功耗降低了 50%。

2.3.5 分段测速

目前，多数家庭网络报障多为“上网慢”“卡顿”等，运营商无法定位具体的故障原因。我们研发了 FTTR 分段测速方案，将网络分为主网关出口至服务器、主网关至从网关和各网关至用户终端，以测试各分段网络的速率。通过速率对比，该方案能够实现故障快速定界，从而找出故障原因。

分段测速方案如图 9 所示，主从网关之间采用 iPerf 测速。原先的方案需要经过整个协议栈，并且中央处理器（CPU）全程参与。这导致 CPU 满载，只能测到线速的 25%，影响网络诊断的准确性。对此，我们设法减少 CPU 的参与，使主网关到从网关的测速能达到下行线速值，从网关到主网关的测速能到上行线速值。



▲图 9 分段测速方案

FTTR 分段测速方案已经在实际产品中落地应用，实现了运维零上门、用户零操作。该方案不仅能够实现对故障的快速定界定位，还可以实现全网批量测速、问题批量整改。

3 FTTR 的机遇与挑战

随着各类新兴应用的发展，家庭网络也在飞速演进，FTTR 势必会面临诸多挑战。中兴通讯在持续思考和探索，不断推动 FTTR 技术的创新发展。

在底层技术方面，我们已经启动新一代 FTTR PON 芯片的研发，以更好适配未来 FTTR 的产品竞争力和网络技术要求。针对落地推广中面临的 ODN 布线施工难题，我们已研发出面板式从网关和吸顶式从网关，并有多种安装部署方案可供选择。

在网络管理方面，基于 FTTH 管理平台，我们研发了中兴智能管理云平台，实现了 FTTR 网络管理。目前该平台还存在标准规范上的不足，无法管理到 FTTR 的每一个层面。因此，对于 FTTR 的管理标准，我们将逐步细化和优化。

在业务保障方面，当前 FTTR 技术实现了 300 STA 并发场景，未来会突破 300 STA 接入能力。因此我们还需要进一步研究网络扩容和性能优化方案，不断提高智能业务感知应用技术的识别效率，扩大应用范围，以实现更多家庭和企业复杂网络中业务流量的精确管控。

在网络演进方面，10G PON 向 50G PON 演进，会给 FTTR 提供更高的带宽。在 Wi-Fi 6 向 Wi-Fi 7 的演进中，除了具有更高的速率外，Wi-Fi 7 引入了多链路设备（MLD）技术，把传统的 Wi-Fi 单链路改造成多链路，在射频链路上增加了冗余，未来升级到 Wi-Fi 7 的 FTTR 可以更好地实现无缝漫游。

4 结束语

继 FTTH 的巨大成功之后，FTTR 开启了第二次光纤革命。中兴通讯在芯片和连接技术、网络管理平台、业务保障机制等方面实现了多项技术创新，为用户提供高速网络和极致体验，

助力运营商打造高质量的家庭千兆网络。未来，我们将持续进行FTTR技术攻关，探索与FTTR相结合的前沿技术和创新应用，推动家庭全光新基建，筑基智慧家庭新未来。

致谢

本文得到中兴通讯股份有限公司李二洁、羊兆磊、武云飞、董志华、石宏宇的帮助，在此表示感谢！

参考文献

[1] 吕达, 郑清芳 . 构建智能实时网络, 使能5G视频业务繁荣 [J]. 中兴通讯技术, 2021, 27(1): 60-67. DOI:10.12142/ZTETJ.202101013
[2] 徐代刚, 姜磊, 梅君君 . 面向视频云微服务系统的智能运维技术 [J]. 中兴通讯技术, 2021, 27(1): 68-76. DOI:10.12142/ZTETJ.202101014
[3] 宽带发展联盟.FTTR光纤到房间白皮书(2022年) [R]. 2022
[4] ITU. ONU management and control interface [EB/OL]. [2022-12-15]. https://www.itu.int/rec/T-REC-G.988-202206-S!Amd5/en
[5] 贺峰 . 管理从光网络单元的方法、光线路终端、主光网络单元: 202210858652.3 [S]. 2022
[6] 贺峰 . 光网络系统的管控方法及装置: 202211458430.9 [S]. 2022
[7] 贺峰 . 网元管理方法及其系统、网元、存储介质: 202111579075.6 [S]. 2021
[8] 羊兆磊 . 多AP组网设备漫游后快速稳定更新转发路径的一种方法: 201911218498.8 [S]. 2019
[9] 羊兆磊 . 一种Mesh网络中的端到端QoS管理方法: 202111138376.5 [S]. 2021
[10] 饶翔, 张顺颐, 周筠 . 基于业务感知的下一代互联网QoS控制架构及关键技术 [J]. 电信科学, 2010(12): 88-95. DOI: 10.3969/j.issn.1000-0801.2010.12.018

作 者 简 介

王新余，中兴通讯股份有限公司智能家庭PON ONU研发总工；主要从事FTTR产品开发和技术标准研究。

孔雪，中兴通讯股份有限公司固网综合方案经理；主要从事固网综合方案和FTTR市场推广。

贺峰，中兴通讯股份有限公司智能家庭PON ONU领域专家；主要从事FTTR管理方案开发和标准研究。

基于两跳IRS辅助的下行无线能量和上行信息传输WPCN性能优化

Performance Optimization for Dual-Hop IRS Assisted Downlink Energy Transfer and Uplink Information Transmission for WPCN

冯璇/FENG Xuan，吕斌/LYU Bin，杨震/YANG Zhen

（南京邮电大学宽带无线通信与传感网技术教育部重点实验室，中国 南京 210003）
(Key Lab of Broadband Wireless Communication and Sensor Network Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

DOI：10.12142/ZTETJ.202301012
网络出版地址：https://kns.cnki.net/kcms/detail//34.1228.TN.20230222.1701.004.html
网络出版日期：2023-02-23
收稿日期：2022-06-18

摘要：面向无线供电通信网络（WPCN），混合接入点（HAP）和无线设备之间常存在障碍物的遮挡，这导致HAP和无线设备之间无法直接通信。针对该问题，提出了两跳智能反射面（IRS）辅助的传输方案，即在HAP和无线设备附近分别放置IRS，使得HAP、IRS和无线设备之间构成反射链路，从而实现HAP到无线设备的下行能量传输和无线设备到HAP的上行信息传输。为了最大化系统吞吐量，建立了关于能量和信息传输的时隙分配、IRS的相位和无线设备信息传输功率的联合优化问题。为了解决该非凸优化问题，提出了两阶段交替优化算法，以有效求得该问题的局部最优解。仿真结果表明，相较于参照方案，所提出的最优方案可以有效提升系统性能。

关键词：两跳智能反射面；无线供电通信网络；相移优化；能量收集

Abstract: In wireless powered communication networks (WPCN), there may exist obstacles between the hybrid access point (HAP) and wireless devices, which makes it impossible for them to communicate directly. To address this issue, two intelligent reflecting surfaces (IRSs) are used to establish a dual-hop relaying transmission link between the HAP and wireless devices. In particular, IRSs are deployed near HAP and wireless devices respectively, which realize the downlink energy transfer from the HAP to wireless devices and the uplink information transmission from wireless devices to the HAP. The system throughput is maximized by jointly optimizing the time scheduling of energy and information transmission, the phase shifts of IRSs, and the transmit power for information transmission of wireless devices. In order to tackle the non-convexity of the formulated problem, a two-stage alternating programming method is proposed to obtain the local optimal solution. Numerical results demonstrate the proposed scheme can achieve significant system throughput gain compared with the benchmark scheme.

Keywords: dual-hop intelligent reflecting surface; wireless powered communication network; phase shift optimization; energy harvesting

5G推动了物联网的发展，人们也越来越享受物联网服务带来的便利生活。物联网是由相互连接的设备、传感器和通信网络实现的[1-2]。物联网中这些相互连接的无线设备在运行时需要持续消耗能量，而其能量储备无法满足这一巨大的能耗需求[3]。为了有效解决无线设备的能量受限问题，研究人员提出了无线功率传输（WPT）新方式。WPT因能够持续为低功耗无线设备提供能量而受到工业界和学术界的广泛关注[4]。无线设备接收来自功率站的射频信号，从中获取更加可控且相对稳定的能量，以此为无线设备的电池充电。这有效延长了电池和无线设备的使用期限[5-6]。基于WPT，无线供电通信网络（WPCN）[7-8]引起了广泛的关注。在WPCN中，无线设备先从混合接入点（HAP）发送的射频信号中收集能量，再利用收集的能量向HAP传输信息[9]。

然而，WPCN在实际应用中依旧面临诸多挑战。例如：当HAP与无线设备之间的距离较远时，无线设备的能量收集效率随之降低。这导致无线设备收集的能量较小，从而影

基金项目：国家自然科学基金资助项目（62071242、61671252、61901229）；江苏省博士后科研基金资助项目（SBH20002）

响WPCN的系统性能[10]。针对该问题，研究者们开展了广泛的研究。例如：文献[11–12]在发射端和接收端布置天线阵列，通过发射端和接收端的波束成形增益有效提高WPCN的性能。多天线技术虽然可以提高WPCN系统的性能，但也会增加收发端的处理复杂度和硬件成本。文献[13–14]研究了主动中继辅助的WPCN，其中中继将来自接入点的能量信号转发到无线设备，并将无线设备的信息反向转发到接入点。但是，主动中继基于射频模板实现能量和信息信号的转发，需要消耗一定的能量来维持其自身的运行，这增加了系统的能耗。因而，如何设计更为有效的方案来解决上述问题值得深入研究。

近年来，智能反射面（IRS）被认为是提高无线通信系统的有效方案。IRS由许多成本较低的反射单元组成，可以自适应地调节反射信号的相位，从而提高反射信号的强度[15–16]。与传统的中继相比，IRS不需要将信号放大和再生，从而降低了硬件成本和系统能耗[17]。因此，在未来无线通信网络中，IRS被视为一种提高系统频谱和能量效率的关键技术[18–19]。

近年来，基于IRS的WPCN的研究引起了学者们的关注。在文献[20]中，当HAP发送的射频信号经由IRS反射到用户后，用户先从射频信号中收集能量，然后再通过非正交多址（NOMA）的方式向HAP上传信息。文献[21]提出了自供电IRS辅助的混合中继方案，其中IRS被用以提高下行能量传输和上行信息传输的性能。在文献[22]中，基站利用IRS将能量传输到多组集群用户，集群用户以时分多址（TDMA）和NOMA的混合方式向基站传输信息。作者通过联合优化功率传输，以及不同用户集群信息传输的IRS的反射波束成形相位、时间分配和功率分配，最大化了网络的吞吐量。

现有的关于IRS辅助WPCN的研究中，HAP和无线设备间的通信由HAP和无线设备间的直接链路或者IRS辅助的单跳中继链路实现。然而，由于网络环境复杂，HAP与无线设备间的通信无法通过直接链路或者单跳中继链路实现。文献[23]研究了基于两个IRS辅助的单用户无线通信系统，解决了复杂网络环境下用户与基站（BS）间无法通信的难题。需要说明的是，文献[23]只关注了系统的信息传输，没有考虑系统中无线设备能量储备不足的情况。因此，面向复杂网络环境的WPCN，同时解决下行能量传输和上行信息传输的高效中继方案仍有待研究。基于此，针对WPCN，本文提出了基于两跳IRS辅助的传输方案，即在HAP和无线设备间部署两个IRS，以此构建两跳的中继传输链路，实现HAP与无线设备间的下行能量和上行信息的中继传输。值得注意的是，相较于传统的中继方案，基于IRS辅助的中继传输方案可以为下行能量传输和上行信息传输提供大量的传输链路，由此可以实现客观的传输效率。本文的主要研究工作包括3个方面：

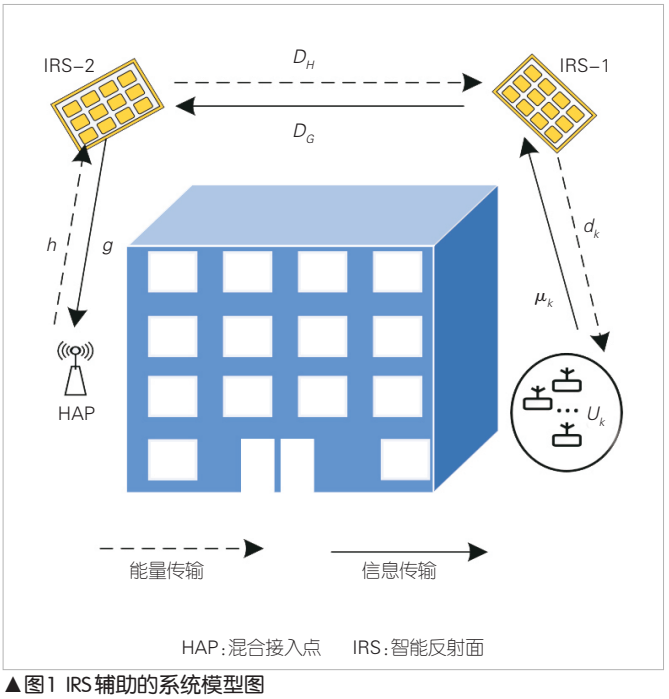
1）针对小区内HAP与无线设备间存在较多障碍物的复杂网络环境，分别在HAP和无线设备附近放置IRS，使得HAP、IRS–2、IRS–1和无线设备之间构成中继链路，从而实现HAP到无线设备的能量传输和无线设备到HAP的信息传输。

2）基于上述模型，在满足能量约束和时间约束的情况下，为了使系统的吞吐量达到最大，考虑了关于能量和信息传输的时间调度、无线设备的发送功率和IRS相位的联合优化问题。由于所定义的优化问题为非凸优化问题，本文提出了一种高效的两阶段交替优化算法，即将该问题分解为两个子问题，然后通过交替优化算法分别获得子问题的次优解。

3）通过仿真实验，验证本文所提算法的可行性。仿真结果表明，相较于参照方案，本文提出的最优传输方案可以获得更大的系统吞吐量。

1 系统模型

如图1所示，该系统由单天线的HAP、IRS–1、IRS–2以及K个单天线的无线设备$(U_k, k = 1,2,3,\cdots,K)$组成。IRS–1和IRS–2的反射单元数量分别是$M_1$和$M_2$，反射单元分别放置在$U_k$和HAP附近。其中，$U_k$表示能量受限的无线设备，需



▲图1 IRS辅助的系统模型图

要用从HAP收集能量来维持自身的能量消耗。假设IRS-1和HAP、IRS-2和U_k之间的反射链路，以及HAP和U_k之间的直射链路被障碍物遮挡，HAP和U_k只能通过IRS-2和IRS-1构建的两跳中继链路实现通信[1]。HAP作为整个网络的控制中心，不仅能够为U_k提供稳定的能量供应，还可以协调K个无线设备的信息传输。HAP、IRS和U_k都需执行严格的同步机制。

HAP与IRS-2、IRS-2与IRS-1、IRS-1与U_k的下行信道系数分别用$\boldsymbol{h}\in\mathbb{C}^{M_2\times 1}$、$\boldsymbol{D}_H\in\mathbb{C}^{M_1\times M_2}$和$\boldsymbol{d}_k\in\mathbb{C}^{M_1\times 1}$来表示，上行信道系数分别用$\boldsymbol{g}\in\mathbb{C}^{M_2\times 1}$、$\boldsymbol{D}_G\in\mathbb{C}^{M_2\times M_1}$和$\boldsymbol{u}_k\in\mathbb{C}^{M_1\times 1}$表示。当前存在诸多技术能够估计IRS系统的信道状态信息（CSI）[17]。因而，假设HAP与IRS-2、IRS-2与IRS-1和IRS-1与U_k间链路的信道状态信息是可以提前获知的。

时长为T的总时隙分为两个阶段，分别为能量收集阶段和信息传输阶段。在能量收集阶段，IRS-2先将来自HAP的能量信号反射到IRS-1处，再通过IRS-1反射到U_k；在信息传输阶段，IRS-1可以将U_k发送的信号反射到IRS-2，进而反射到HAP处。系统具体的时隙分配如图2所示。根据该时隙分配方案，可以有效避免下行能量传输和上行信息传输间的干扰。

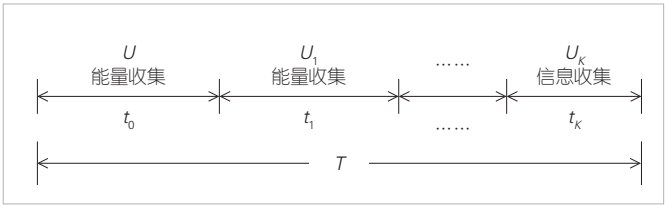
1） 能量收集阶段

在HAP到U_k的能量传输阶段，HAP在时长为t_0的子时隙内发射射频信号[24]，经由IRS反射到U_k，然后U_k从中收集能量。下行IRS的反射相移矩阵为：$\boldsymbol{\Phi}_{d\mu}=\sqrt{\rho}\,\mathrm{diag}\left\{\boldsymbol{\theta}_{d\mu}\right\}$，$\mu\in\{1,2\}$，$\boldsymbol{\theta}_{d\mu}\triangleq\left[\theta_{d\mu,1},\theta_{d\mu,2},\cdots,\theta_{d\mu,M_\mu}\right]^T$，$\theta_{d\mu,m}=\beta_{\mu,m}e^{j\omega_m}$，$m=\left\{1,\cdots,M_\mu\right\}$。其中，$\rho\in(0,1)$表示IRS的反射效率，其值通常被设置为常数[21]；$\beta_{\mu,m}\in[0,1]$和$\omega_m\in[0,2\pi)$分别表示第μ个IRS的第m个反射单元的振幅和相移系数。为了最大化下行能量传输和上行信息传输的效率，IRS的反射振幅$\beta_{\mu,m}$可设为1[21]。U_k接收到来自HAP的能量信号如下：

$$y_{U_k}=\left(\boldsymbol{d}_k^H\boldsymbol{\Phi}_{d,1}\boldsymbol{D}_H\boldsymbol{\Phi}_{d,2}\boldsymbol{h}\right)\sqrt{P_B}\,s_B+n_k,\tag{1}$$

其中，P_B表示HAP的发送功率，s_B表示HAP发射的能量信号且$s_B\sim\mathbb{CN}(0,1)$，n_k表示U_k处的噪声。

为表征非线性能量收集模型的特点，本文采用两阶段的线性能量收集模型[25-26]，该模型在WPCN的相关研究中被广泛使用。基于该模型，U_k的接收功率为：



▲图2 系统时隙分配图

$$P_k=\begin{cases}\eta P_B\left|\boldsymbol{d}_k^H\boldsymbol{\Phi}_{d,1}\boldsymbol{D}_H\boldsymbol{\Phi}_{d,2}\boldsymbol{h}\right|^2, & \eta P_B\left|\boldsymbol{d}_k^H\boldsymbol{\Phi}_{d,1}\boldsymbol{D}_H\boldsymbol{\Phi}_{d,2}\boldsymbol{h}\right|^2<p_{f,k}\\ p_{f,k}, & \eta P_B\left|\boldsymbol{d}_k^H\boldsymbol{\Phi}_{d,1}\boldsymbol{D}_H\boldsymbol{\Phi}_{d,2}\boldsymbol{h}\right|^2\geqslant p_{f,k}\end{cases},\tag{2}$$

其中，η表示U_k的能量收集效率，$p_{f,k}$表示U_k的饱和功率。在t_0时隙内，U_k收集的能量为：

$$E_{d,k}=t_0\min\left(\eta P_B\left|\boldsymbol{d}_k^H\boldsymbol{\Phi}_{d,1}\boldsymbol{D}_H\boldsymbol{\Phi}_{d,2}\boldsymbol{h}\right|^2,p_{f,k}\right)。\tag{3}$$

2） 信息传输阶段

在U_k到HAP的信息传输阶段，无线设备利用收集的能量发送信息[24]，并经由IRS反射到HAP。上行IRS反射相移矩阵与$\boldsymbol{\Phi}_{d\mu}$类似，$\boldsymbol{\Phi}_{u\mu}^k=\sqrt{\rho}\,\mathrm{diag}\left\{\boldsymbol{\theta}_{u,\mu}^k\right\}$，$\boldsymbol{\theta}_{u,\mu}^k\triangleq\left[\theta_{u,\mu,1}^k,\theta_{u,\mu,2}^k,\cdots,\theta_{u,\mu,M_\mu}^k\right]^T$，$\theta_{u,\mu,m}^k=\beta_{\mu,m}^k e^{j\omega_m}$，$\beta_{\mu,m}^k=1$，$\omega_m\in[0,2\pi)$。无线设备$U_k$在时长为$t_k\in[0,T)$的子时隙内向HAP传输信息，即无线设备采用时分多址的方式进行信息传输。则HAP收到的信号表示为：

$$y_{BS}=\left(\boldsymbol{g}^H\boldsymbol{\Phi}_{u,2}^k\boldsymbol{D}_G\boldsymbol{\Phi}_{u,1}^k\boldsymbol{u}_k\right)\sqrt{p_k}\,x_k+n_h,\tag{4}$$

其中，p_k与n_h分别表示U_k的信息传输功率和HAP处的噪声。则U_k在时间t_k内的吞吐量为：

$$R_k=t_kB\log_2\left(1+\frac{p_k\left|\boldsymbol{g}^H\boldsymbol{\Phi}_{u,2}^k\boldsymbol{D}_G\boldsymbol{\Phi}_{u,1}^k\boldsymbol{u}_k\right|^2}{\delta^2}\right),\tag{5}$$

其中，δ^2表示HAP处噪声的功率。此外，U_k信息传输阶段消耗的总能量不应该超过收集的能量，即满足：

$$\left(p_k+p_{c,k}\right)t_k\leqslant t_0\min\left(\eta P_B\left|\boldsymbol{d}_k^H\boldsymbol{\Phi}_{d,1}\boldsymbol{D}_H\boldsymbol{\Phi}_{d,2}\boldsymbol{h}\right|^2,p_{f,k}\right),\tag{6}$$

其中，$p_{c,k}$表示U_k到HAP信息传输过程中的电路功耗。

2 系统吞吐量最大化

为最大化系统吞吐量，本节构建了关于能量传输和信息传输阶段的IRS反射相移矩阵、时隙调度和无线设备传输信

1 值得注意的是，上述模型也可拓展为多个IRS辅助传输场景，即HAP和U_k间部署多个IRS以构成多跳中继传输链路。然而，考虑到网络的实际控制难度和布设IRS成本，本文只关注两个IRS辅助传输场景。

息的发送功率联合优化问题。该问题可以表述为：

$$\max_{\boldsymbol{p},\boldsymbol{t},\boldsymbol{\Phi}_{d,\mu},\boldsymbol{\Phi}_{d,\mu}^{k}} \sum_{k=1}^{K} t_k B \log_2\left(1+\frac{p_k\left|\boldsymbol{g}^H\boldsymbol{\Phi}_{u,2}^{k}\boldsymbol{D}_G\boldsymbol{\Phi}_{u,1}^{k}\boldsymbol{u}_k\right|^2}{\delta^2}\right)$$

$$\text{s.t. C1: } p_k t_k + p_{c,k} t_k \leqslant t_0 \min\left(\eta P_B\left|\boldsymbol{d}_k^H\boldsymbol{\Phi}_{d,1}\boldsymbol{D}_H\boldsymbol{\Phi}_{d,2}\boldsymbol{h}\right|^2, p_{f,k}\right)$$

$$\text{C2: } \sum_{k=0}^{K} t_k \leqslant T, 0 \leqslant t_k \leqslant T, \text{C3: } p_k \geqslant 0$$

$$\text{C4: } \left|\theta_{u,\mu,m}^{k}\right| = 1, \left|\theta_{d,\mu,m}\right| = 1,\ m = \{1,\cdots,M_\mu\}, \mu \in \{1,2\}, \quad \text{(P1)}$$

其中 $\boldsymbol{p} = [p_1,\cdots,p_K]$，$\boldsymbol{t} = [t_0,\cdots,t_K]$，$k = \{1,\cdots,K\}$。约束条件C1表示无线设备到HAP信息传输阶段的能量约束，C3为无线设备信息传输功率约束，C4是IRS在能量和信息传输阶段的相移约束。在优化问题（P1）的目标函数和约束条件中，由于优化变量 t_k、p_k、$\boldsymbol{\Phi}_{u,\mu}^{k}$ 和 $\boldsymbol{\Phi}_{d,\mu}$ 间存在耦合的情况，因而（P1）是非凸优化问题，很难求得最优解。为此，本文提出一种两阶段的交替优化算法，可以有效求得其次优解：首先，通过分析系统吞吐量最大化问题（P1）的结构，将问题（P1）分解为两个子问题，即上行信道增益最大化问题和系统资源分配优化问题；然后，分别采用交替优化算法获得两个子问题的次优解。

2.1 上行信道增益最大化

通过分析问题（P1）的结构可以看出，R_k 关于 $\left|\boldsymbol{g}^H\boldsymbol{\Phi}_{u,2}^{k}\boldsymbol{D}_G\boldsymbol{\Phi}_{u,1}^{k}\boldsymbol{u}_k\right|^2$ 单调递增，即当 $\left|\boldsymbol{g}^H\boldsymbol{\Phi}_{u,2}^{k}\boldsymbol{D}_G\boldsymbol{\Phi}_{u,1}^{k}\boldsymbol{u}_k\right|^2$ 最大时，R_k 也是最优的。因此，当给定任意可行的 $\boldsymbol{p}$、$\boldsymbol{t}$ 和 $\boldsymbol{\Phi}_{d\mu}$ 时，求解问题（P1）就等价于求解 K 个上行信道增益最大化问题（P2），其定义如下：

$$\max_{\boldsymbol{\Phi}_{u,\mu}^{k}} \quad \left|\boldsymbol{g}^H\boldsymbol{\Phi}_{u,2}^{k}\boldsymbol{D}_G\boldsymbol{\Phi}_{u,1}^{k}\boldsymbol{u}_k\right|^2$$

$$\text{s.t.} \quad \text{C5: } \left|\theta_{u,\mu,m}\right| = 1, \mu \in \{1,2\},\ m = \{1,\cdots,M_\mu\}。 \quad \text{(P2)}$$

在优化问题（P2）的目标函数中，$\boldsymbol{\Phi}_{u,2}^{k}$ 和 $\boldsymbol{\Phi}_{u,1}^{k}$ 间存在耦合的情况，故（P2）不是凸优化问题，很难求得最优解。为此，本节采用交替优化算法来求得问题（P2）的次优解。

1）给定 $\boldsymbol{\Phi}_{u,1}^{k}$，优化 $\boldsymbol{\Phi}_{u,2}^{k}$

在 U_k 到HAP的信息传输阶段，给定任意可行IRS-1的相移矩阵 $\boldsymbol{\Phi}_{u,1}^{k}$，令 $\bar{\boldsymbol{e}}_2 = \left[\theta_{u,2,1}^{k},\theta_{u,2,2}^{k},\cdots,\theta_{u,2,M_2}^{k}\right]^H$ 且 $\boldsymbol{v}_k = \text{diag}\left(\boldsymbol{g}^H\right)\boldsymbol{D}_G\boldsymbol{\Phi}_{u,1}^{k}\boldsymbol{u}_k$，则 U_k 到HAP的上行信道系数 $h_{u,k} = \boldsymbol{g}^H\boldsymbol{\Phi}_{u,2}^{k}\boldsymbol{D}_G\boldsymbol{\Phi}_{u,1}^{k}\boldsymbol{u}_k = \bar{\boldsymbol{e}}_2^H\boldsymbol{v}_k$。然后，引入变量 $\boldsymbol{E}_{u,2} = \bar{\boldsymbol{e}}_2\bar{\boldsymbol{e}}_2^H$ 和 $\boldsymbol{V}_k = \boldsymbol{v}_k\boldsymbol{v}_k^H$，则 $\left|\bar{\boldsymbol{e}}_2^H\boldsymbol{v}_k\right|^2 = \text{tr}\left(\boldsymbol{E}_{u,2}\boldsymbol{V}_k\right)$，其中 $\boldsymbol{E}_{u,2} \geqslant 0$ 且 $\text{rank}\left(\boldsymbol{E}_{u,2}\right) = 1$。由于秩为1的约束条件是非凸的，此处通过半正定松弛技术（SDR）将问题（P2）松弛为如下问题（P2.1）：

$$\max_{\boldsymbol{E}_{u,2}} \quad \text{tr}\left(\boldsymbol{E}_{u,2}\boldsymbol{V}_k\right)$$

$$\text{s.t.} \quad \boldsymbol{E}_{u,2} \geqslant 0$$

$$\boldsymbol{E}_{u,2}(m,m) = 1, m = 1,\cdots,M_2。 \quad \text{(P2.1)}$$

问题（P2.1）是半正定规划问题，可以用现有的凸优化工具CVX直接求解。但CVX求得的 $\hat{\boldsymbol{E}}_{u,2}$ 不一定满足秩为1的约束条件，此处可以利用高斯随机的方法解决这一问题：首先对矩阵 $\hat{\boldsymbol{E}}_{u,2}$ 进行奇异值分解，即 $\hat{\boldsymbol{E}}_{u,2} = \boldsymbol{U}_e\boldsymbol{\Sigma}_e\boldsymbol{U}_e^H$，其中，$\boldsymbol{U}_e \in \mathbb{C}^{M_2\times M_2}$ 是酉矩阵，$\boldsymbol{\Sigma}_e \in \mathbb{C}^{M_2\times M_2}$ 是对角矩阵；然后生成 D 次随机向量 $\boldsymbol{r}_e \sim \mathbb{CN}\left(0,\boldsymbol{I}_{M_2}\right)$，则（P2.1）的近似解用 $\hat{\boldsymbol{e}}_2$ 表示且 $\hat{\boldsymbol{e}}_2 = \boldsymbol{U}_e\sqrt{\boldsymbol{\Sigma}_e}\boldsymbol{r}_e$，其目标函数值可以表示为 $\text{tr}\left(\hat{\boldsymbol{e}}_2\hat{\boldsymbol{e}}_2^H\boldsymbol{V}_k\right)$，同时使目标函数值最大的 $\hat{\boldsymbol{e}}_2$ 可用 $\tilde{\boldsymbol{e}}_2$ 来表示；最后，问题（P2.1）的次优解为 $\bar{\boldsymbol{e}}_2^* = e^{j\arg\left([\tilde{\boldsymbol{e}}_2]_{1:M_2}\right)}$，$\boldsymbol{\Phi}_{u,2}^{k} = \text{diag}\left(\bar{\boldsymbol{e}}_2^*\right)$，其中，$[\omega]_{1:M_2}$ 表示 ω 的首个 M_2 元素。

2）给定 $\boldsymbol{\Phi}_{u,2}^{k}$，优化 $\boldsymbol{\Phi}_{u,1}^{k}$

在 U_k 到HAP的信息传输阶段，给定任意可行IRS-2的相移矩阵 $\boldsymbol{\Phi}_{u,2}^{k}$，令 $\bar{\boldsymbol{e}}_1 = \left[\theta_{u,1,1}^{k},\theta_{u,1,2}^{k},\cdots,\theta_{u,1,M_1}^{k}\right]^H$ 且 $\boldsymbol{\psi}_k = \text{diag}\left(\boldsymbol{g}^H\boldsymbol{\Phi}_{u,2}^{k}\boldsymbol{D}_G\right)\boldsymbol{u}_k$，则 $h_{u,k} = \boldsymbol{g}^H\boldsymbol{\Phi}_{u,2}^{k}\boldsymbol{D}_G\boldsymbol{\Phi}_{u,1}^{k}\boldsymbol{u}_k = \bar{\boldsymbol{e}}_1^H\boldsymbol{\psi}_k$。然后，引入变量 $\boldsymbol{E}_{u,1} = \bar{\boldsymbol{e}}_1\bar{\boldsymbol{e}}_1^H$ 和 $\boldsymbol{\Psi}_k = \boldsymbol{\psi}_k\boldsymbol{\psi}_k^H$，则 $\left|\bar{\boldsymbol{e}}_1^H\boldsymbol{\psi}_k\right|^2 = \text{tr}\left(\boldsymbol{E}_{u,1}\boldsymbol{\Psi}_k\right)$，其中 $\boldsymbol{E}_{u,1} \geqslant 0$ 且 $\text{rank}\left(\boldsymbol{E}_{u,2}\right) = 1$。同样地，我们使用SDR技术秩为1约束条件进行松弛，问题（P2）可转换为（P2.2）。问题（P2.2）的定义如下：

$$\max_{\boldsymbol{E}_{u,1}} \quad \text{tr}\left(\boldsymbol{E}_{u,1}\boldsymbol{\Psi}_k\right)$$

$$\text{s.t.} \quad \boldsymbol{E}_{u,1} \geqslant 0$$

$$\boldsymbol{E}_{u,1}(m,m) = 1, m = 1,\cdots,M_1。 \quad \text{(P2.2)}$$

问题（P2.2）也是半正定规划问题，我们同样可以利用CVX进行求解，然后再通过高斯随机获得秩为1的约束条件的解，其中使得问题（P2.2）的目标函数值最大的次优解用 $\bar{\boldsymbol{e}}_1^*$ 来表示，则 $\boldsymbol{\Phi}_{u,1}^{k} = \text{diag}\left(\bar{\boldsymbol{e}}_1^*\right)$。

2.2 系统资源分配优化

通过求解 K 个上行信道增益最大化问题可获得上行IRS反射波束成形矩阵 $\boldsymbol{\Phi}_{u\mu}^{*}$。随后引入辅助变量 y_k，并令 $y_k = t_k p_k$，$h_k = \left|h_{u,k}^{*}\right|^2/\delta^2$。问题（P1）转换为系统资源分配优化问题（P3）：

$$\max_{\boldsymbol{y},\boldsymbol{t},\boldsymbol{\Phi}_{d,\mu}} \quad \sum_{k=1}^{K} t_k B \log_2\left(1+\frac{y_k h_k}{t_k}\right)$$

$$\text{s.t. C6: } y_k + p_{c,k} t_k \leqslant t_0 \min\left(\eta P_B \left|\boldsymbol{d}_k^H \boldsymbol{\Phi}_{d,1} \boldsymbol{D}_H \boldsymbol{\Phi}_{d,2} \boldsymbol{h}\right|^2, p_{f,k}\right)$$

$$\text{C7: } y_k \geqslant 0, \text{C2}$$

$$\text{C8: } \left|\theta_{d,\mu,m}\right| = 1, \mu \in \{1,2\}, m = \{1,\cdots,M_\mu\}。 \quad \text{(P3)}$$

由于约束条件C6中优化变量t_0与$\boldsymbol{\Phi}_{d\mu}$存在耦合的情况，故问题（P3）仍然是非凸优化问题，其最优解的求解较为困难。为此，本节采用交替优化的方法，以有效求得系统资源分配优化问题（P3）的次优解。

1）给定$\boldsymbol{\Phi}_{d,1}$和$\boldsymbol{\Phi}_{d,2}$,优化$\boldsymbol{y}$和$\boldsymbol{t}$

在HAP到U_k的能量传输阶段，给定任意可行的IRS的相移矩阵$\boldsymbol{\Phi}_{d,1}$和$\boldsymbol{\Phi}_{d,2}$，系统资源分配优化问题（P3）可以转换为问题（P3.1）：

$$\max_{\boldsymbol{y},\boldsymbol{t}} \quad \sum_{k=1}^{K} t_k B \log_2\left(1+\frac{y_k h_k}{t_k}\right)$$

$$\text{s.t.} \quad \text{C6, C7, C2} \quad 。 \quad \text{(P3.1)}$$

问题（P3.1）中$R_k = t_k B \log_2\left(1+\frac{y_k h_k}{t_k}\right)$。可以看出，$R_k$是关于$y_k$和$t_k$的函数。$R_k$的海森矩阵如下：

$$G_k(t_k, y_k) = \begin{bmatrix} -\dfrac{y_k^2 h_k^2 B}{t_k \ln 2 \left(t_k + y_k \mathrm{h_k}\right)^2} & \dfrac{y_k h_k^2 B}{\mathrm{In}2\left(t_k + y_k h_k\right)^2} \\ \dfrac{y_k h_k^2 B}{\mathrm{In}2\left(t_k + y_k h_k\right)^2} & -\dfrac{t_k h_k^2 B}{\mathrm{In}2\left(t_k + y_k h_k\right)^2} \end{bmatrix}。 \quad (7)$$

$G_k(t_k,y_k)$矩阵半负定，因此R_k是关于y_k和t_k的凹函数。同时，给定$\boldsymbol{\Phi}_{d,1}$和$\boldsymbol{\Phi}_{d,2}$时，不等式约束C2和C6都是线性的，故问题（P3.1）是凸优化问题，可以用凸优化工具CVX直接求解。

2）给定$\boldsymbol{y}$、$\boldsymbol{t}$和$\boldsymbol{\Phi}_{d,1}$,优化$\boldsymbol{\Phi}_{d,2}$

给定$\boldsymbol{y}$、$\boldsymbol{t}$和$\boldsymbol{\Phi}_{d,1}$，问题（P3）就转换为关于$\boldsymbol{\Phi}_{d,2}$的可行性验证问题：

$$\text{Find} \quad \boldsymbol{\Phi}_{d,2}$$

$$\text{s.t.} \quad \text{C6, C9: } \left|\theta_{d,2,m}\right| = 1, m = \{1,...,M_2\}。 \quad \text{(P3.2)}$$

引入辅助变量$\boldsymbol{x} = [x_1,\cdots,x_k,\cdots,x_K]$，令$\boldsymbol{e}_2 = \left[\theta_{d,2,1},\theta_{d,2,2},\cdots,\theta_{d,2,M_2}\right]^H$且$\boldsymbol{a}_k = \mathrm{diag}\left(\boldsymbol{d}_k^H \boldsymbol{\Phi}_{d,1} \boldsymbol{D}_H\right)\boldsymbol{h}$，则HAP到$U_k$的下行信道系数$h_{d,k} = \boldsymbol{d}_k^H \boldsymbol{\Phi}_{d,1} \boldsymbol{D}_H \boldsymbol{\Phi}_{d,2} \boldsymbol{h} = \boldsymbol{e}_2^H \boldsymbol{a}_k$。然后，令$\boldsymbol{E}_{d,2} = \boldsymbol{e}_2 \boldsymbol{e}_2^H$且$\boldsymbol{A}_k = \boldsymbol{a}_k \boldsymbol{a}_k^H$，则$\left|\boldsymbol{e}_2^H \boldsymbol{a}_k\right|^2 = \mathrm{tr}\left(\boldsymbol{E}_{d,2} \boldsymbol{A}_k\right)$，其中$\boldsymbol{E}_{d,2} \geqslant 0$且$\mathrm{rank}\left(\boldsymbol{E}_{d,2}\right) = 1$。利用SDR技术将秩为1的约束条件进行松弛后，问题（P3.2）可变换为：

$$\max_{\boldsymbol{E}_{d,2},\boldsymbol{x}} \quad \sum_{k=1}^{K} x_k$$

$$\text{s.t. C10: } \boldsymbol{E}_{d,2} \geqslant 0$$

$$\text{C11: } \boldsymbol{E}_{d,2}(m,m), m = 1,\cdots,M_2$$

$$\text{C12: } x_k \geqslant 0, k = 1,\cdots,K$$

$$\text{C13: } y_k + p_{c,k} t_k + x_k \leqslant t_0 \min\left(\eta P_B \,\mathrm{tr}\left(\boldsymbol{E}_{d,2} \boldsymbol{A}_k\right), p_{f,k}\right)。$$

$$\text{(P3.3)}$$

问题（P3.3）是半正定规划问题，可以先用凸优化工具CVX直接求解，再使用高斯随机方法得到满足秩为1约束的解，最后选取使得K个用户接收功率总和最大的解作为$\boldsymbol{e}_2^*$，则$\boldsymbol{\Phi}_{d,2} = \mathrm{diag}\left(\boldsymbol{e}_2^*\right)$。

3）给定$\boldsymbol{y}$、$\boldsymbol{t}$和$\boldsymbol{\Phi}_{d,2}$,优化$\boldsymbol{\Phi}_{d,1}$

给定$\boldsymbol{y}$、$\boldsymbol{t}$和$\boldsymbol{\Phi}_{d,2}$，问题（P3）就转换为类似（P3.2）的关于$\boldsymbol{\Phi}_{d,1}$的可行性问题。引入辅助变量$\boldsymbol{x} = [x_1,\cdots,x_k,\cdots,x_K]$，令$\boldsymbol{e}_1 = \left[\theta_{d,1,1},\theta_{d,1,2},\cdots,\theta_{d,1,M_1}\right]^H$且$\boldsymbol{q}_k = \mathrm{diag}\left(\boldsymbol{d}_k^H\right)\boldsymbol{D}_H \boldsymbol{\Phi}_{d,2} \boldsymbol{h}$，则$h_{d,k} = \boldsymbol{d}_k^H \boldsymbol{\Phi}_{d,1} \boldsymbol{D}_H \boldsymbol{\Phi}_{d,2} \boldsymbol{h} = \boldsymbol{e}_1^H \boldsymbol{q}_k$。然后，令$\boldsymbol{E}_{d,1} = \boldsymbol{e}_1 \boldsymbol{e}_1^H$和$\boldsymbol{Q}_k = \boldsymbol{q}_k \boldsymbol{q}_k^H$，则$\left|\boldsymbol{e}_1^H \boldsymbol{q}_k\right|^2 = \mathrm{tr}\left(\boldsymbol{E}_{d,1} \boldsymbol{Q}_k\right)$，其中，$\boldsymbol{E}_{d,1} \geqslant 0$且$\mathrm{rank}\left(\boldsymbol{E}_{d,1}\right) = 1$。同上，利用SDR技术将秩为1的约束进行松弛，关于$\boldsymbol{\Phi}_{d,1}$的可行性问题变形为如下问题：

$$\max_{\boldsymbol{E}_{d,1},\boldsymbol{x}} \quad \sum_{k=1}^{K} x_k$$

$$\text{s.t. C14: } \boldsymbol{E}_{d,1} \geqslant 0, \text{C12}$$

$$\text{C15: } \boldsymbol{E}_{d,1}(m,m), m = 1,\cdots,M_1$$

$$\text{C16: } y_k + p_{c,k} t_k + x_k \leqslant t_0 \min\left(\eta P_B \,\mathrm{tr}\left(\boldsymbol{E}_{d,1} \boldsymbol{Q}_k\right), p_{f,k}\right)。 \quad \text{(P3.4)}$$

问题（P3.4）也是半正定规划问题，因此可以先用凸优化工具CVX直接求解，再使用高斯随机方法得到满足秩为1约束的解最后选取使得K个用户接收功率总和最大的解作为$\boldsymbol{e}_1^*$，则$\boldsymbol{\Phi}_{d,1} = \mathrm{diag}\left(\boldsymbol{e}_1^*\right)$。

3 算法分析

1） 复杂度分析

算法1详细总结了求解（P1）的两阶段交替优化算法。首先，步骤1—6可以求解上行信道增益最大化问题（P2），步骤7—12可以求解系统资源分配优化问题（P3）。下面我们对于算法1的计算复杂度进行分析。上行信道增益最大化问题（P2）的计算复杂度主要由步骤3和4决定。根据文献[21]可知，问题（P2.1）和（P2.2）的算法复杂度分别为$O\left(M_2^{4.5}\log(1/\varepsilon)+D\right)$和$O\left(M_1^{4.5}\log(1/\varepsilon)+D\right)$，其中$\varepsilon$表示

CVX内部使用内点法求解凸优化问题时的计算精度，D表示高斯随机的次数。故问题（P2）的算法复杂度为$O\left(Kl\left(M_2^{4.5}\log\left(1/\varepsilon\right)+M_1^{4.5}\log\left(1/\varepsilon\right)+2D\right)\right)$，其中$K$表示无线设备的数量，$l$表示（P2）的中迭代次数。问题（P3）的计算复杂度主要源于步骤9和10。根据文献[14]，问题（P3.3）和（P3.4）的计算复杂度分别为$O\left(\max\left(K,M_2\right)^4M_2^{0.5}\log\left(1/\varepsilon\right)+DK\right)$和$O\left(\max\left(K,M_1\right)^4M_1^{0.5}\log\left(1/\varepsilon\right)+DK\right)$，故（P3）的计算复杂度为$O\left(n\max\left(K,M_2\right)^4M_2^{0.5}\log\left(1/\varepsilon\right)+n\max\left(K,M_1\right)^4M_1^{0.5}\log\left(1/\varepsilon\right)+2nDK\right)$，其中$n$表示求解问题（P3）的迭代次数。因而，算法1的复杂度为：$O\Big(Kl\big(M_2^{4.5}+M_1^{4.5}\big)\log\left(1/\varepsilon\right)+n\Big(\max\left(K,M_2\right)^4M_2^{0.5}+\max\left(K,M_1\right)^4M_1^{0.5}\Big)\log\left(1/\varepsilon\right)+2KD(n+l)\Big)$。

2）收敛性分析

算法1的收敛性主要取决于步骤2—12。步骤3—5可以求得上行信道增益最大化问题（P2）的次优解。令$\boldsymbol{\Phi}_{u,2}^{k(n)}$和$\boldsymbol{\Phi}_{u,1}^{k(n)}$分别表示问题（P2）第$n$次迭代的解，则（P2）第$n$次迭代的目标函数值为$h_u{}^{(n)}=f\left(\boldsymbol{\Phi}_{u,2}^{k(n)},\boldsymbol{\Phi}_{u,1}^{k(n)}\right)$。给定$\boldsymbol{\Phi}_{u,2}^{k(n-1)}$时，可以得到公式（8）：

$$h_u{}^{(n-1)}=f\left(\boldsymbol{\Phi}_{u,2}^{k(n-1)},\boldsymbol{\Phi}_{u,1}^{k(n-1)}\right)\leqslant f\left(\boldsymbol{\Phi}_{u,2}^{k(n-1)},\boldsymbol{\Phi}_{u,1}^{k(n)}\right)。\tag{8}$$

给定$\boldsymbol{\Phi}_{u,1}^{k(n)}$时，公式（9）成立：

$$f\left(\boldsymbol{\Phi}_{u,2}^{k(n-1)},\boldsymbol{\Phi}_{u,1}^{k(n)}\right)\leqslant f\left(\boldsymbol{\Phi}_{u,2}^{k(n)},\boldsymbol{\Phi}_{u,1}^{k(n)}\right)=h_u{}^{(n)}。\tag{9}$$

故$h_u{}^{(n-1)}\leqslant h_u{}^{(n)}$。每次迭代后，问题（P2）的目标值（上行信道增益）是非减的。同时，信道增益存在一个有限值的上界，故步骤3—5是收敛的。通过步骤9—11可得系统资源分配优化问题（P3）的次优解。分别用$\boldsymbol{y}^{(n)}$、$\boldsymbol{t}^{(n)}$、$\boldsymbol{\Phi}_{d,2}^{(n)}$和$\boldsymbol{\Phi}_{d,1}^{(n)}$表示问题（P3）在第$n$次迭代的解，则（P3）的目标函数值为：$R^{(n)}=f\left(\boldsymbol{y}^{(n)},\boldsymbol{t}^{(n)},\boldsymbol{\Phi}_{d,2}^{(n)},\boldsymbol{\Phi}_{d,1}^{(n)}\right)$。通过步骤9可以求得子问题（P3.3）的局部最优解。给定$\boldsymbol{\Phi}_{d,1}^{(n-1)}$、$\boldsymbol{y}^{(n-1)}$和$\boldsymbol{t}^{(n-1)}$时，公式（10）成立：

$$R^{(n-1)}=f\left(\boldsymbol{y}^{(n-1)},\boldsymbol{t}^{(n-1)},\boldsymbol{\Phi}_{d,2}^{(n-1)},\boldsymbol{\Phi}_{d,1}^{(n-1)}\right)\leqslant f\left(\boldsymbol{y}^{(n-1)},\boldsymbol{t}^{(n-1)},\boldsymbol{\Phi}_{d,2}^{(n)},\boldsymbol{\Phi}_{d,1}^{(n-1)}\right)。\tag{10}$$

通过步骤10可以求得子问题（P3.4）的局部最优解，给定$\boldsymbol{\Phi}_{d,2}^{(n)}$、$\boldsymbol{y}^{(n-1)}$和$\boldsymbol{t}^{(n-1)}$时，公式（11）成立：

$$f\left(\boldsymbol{y}^{(n-1)},\boldsymbol{t}^{(n-1)},\boldsymbol{\Phi}_{d,2}^{(n)},\boldsymbol{\Phi}_{d,1}^{(n-1)}\right)\leqslant f\left(\boldsymbol{y}^{(n-1)},\boldsymbol{t}^{(n-1)},\boldsymbol{\Phi}_{d,2}^{(n)},\boldsymbol{\Phi}_{d,1}^{(n)}\right)。\tag{11}$$

通过步骤11可以求得问题（P3.1）的全局最优解，给定$\boldsymbol{\Phi}_{d,1}^{(n)}$和$\boldsymbol{\Phi}_{d,2}^{(n)}$时，不等式（12）成立：

$$f\left(\boldsymbol{y}^{(n-1)},\boldsymbol{t}^{(n-1)},\boldsymbol{\Phi}_{d,2}^{(n)},\boldsymbol{\Phi}_{d,1}^{(n)}\right)\leqslant f\left(\boldsymbol{y}^{(n)},\boldsymbol{t}^{(n)},\boldsymbol{\Phi}_{d,2}^{(n)},\boldsymbol{\Phi}_{d,1}^{(n)}\right)=R^{(n)}。\tag{12}$$

因此，不等式$R^{(n-1)}\leqslant R^{(n)}$成立。故每次迭代后问题（P3）的目标值（系统吞吐量）是非减的。同时，系统的吞吐量存在一个有限值的上界。因而步骤9～11是收敛的。最终，算法1是收敛的。

算法1：求解问题(P1)算法

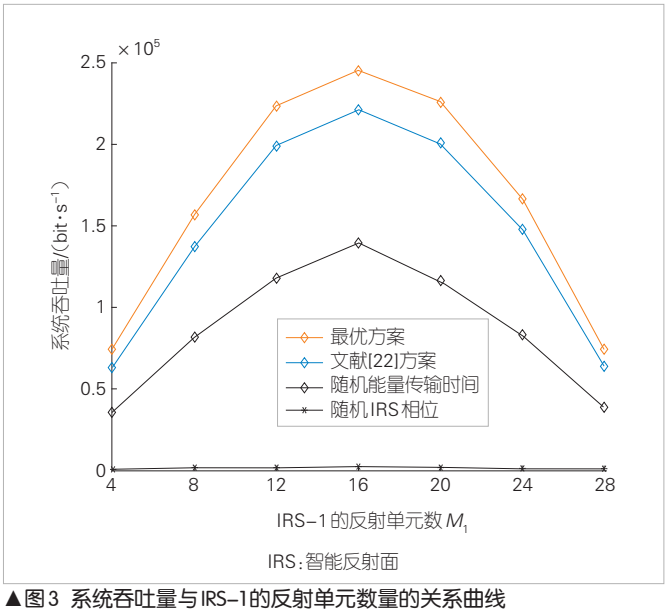
1:初始化$\boldsymbol{\Phi}_{u,1}^{k}$和$\boldsymbol{\Phi}_{u,2}^{k}$，令$l=0$，计算$h_u{}^{(0)}=\left|\boldsymbol{g}^H\boldsymbol{\Phi}_{u,2}^{k}\boldsymbol{D}_G\boldsymbol{\Phi}_{u,1}^{k}\boldsymbol{u}_k\right|^2$
2:重复
3:　给定$\boldsymbol{\Phi}_{u,1}^{k}$，求解问题(P2.1)得到$\boldsymbol{\Phi}_{u,2}^{k}$；
4:　给定$\boldsymbol{\Phi}_{u,2}^{k}$，求解问题(P2.2)得到$\boldsymbol{\Phi}_{u,1}^{k}$；
5:　给定$\boldsymbol{\Phi}_{u,1}^{k}$和$\boldsymbol{\Phi}_{u,2}^{k}$，令$l=l+1$，计算$h_u{}^{(l)}$；
6:循环直至$\left|h_u{}^{(l)}-h_u{}^{(l-1)}\right|\leqslant\varepsilon_1$
7:初始化$\boldsymbol{\Phi}_{d,1}$和$\boldsymbol{\Phi}_{d,2}$，令$n=0$，求解问题(P3.1)得到$R^{(0)}=f\left(\boldsymbol{y},\boldsymbol{t},\boldsymbol{\Phi}_{d,1},\boldsymbol{\Phi}_{d,2}\right)$
8:重复
9:　给定$\boldsymbol{y}$、$\boldsymbol{t}$和$\boldsymbol{\Phi}_{d,1}$，求解问题(P3.3)得到$\boldsymbol{\Phi}_{d,2}$；
10:　给定$\boldsymbol{y}$、$\boldsymbol{t}$和$\boldsymbol{\Phi}_{d,2}$，求解问题(P3.4)得到$\boldsymbol{\Phi}_{d,1}$；
11:　给定$\boldsymbol{\Phi}_{d,1}$和$\boldsymbol{\Phi}_{d,2}$，令$n=n+1$，求解问题(P3.1)得到$R^{(n)}=f\left(\boldsymbol{y},\boldsymbol{t},\boldsymbol{\Phi}_{d,1},\boldsymbol{\Phi}_{d,2}\right)$；
12:循环直至$\left|R^{(n)}-R^{(n-1)}\right|\leqslant\varepsilon_2$

4 仿真分析

本节中，我们通过仿真实验对提出的方案进行性能分析。本文中，我们考虑小区内无线供电通信系统的载波频率为750 MHz[19]，仿真的网络拓扑为3维坐标系，HAP、IRS-2和IRS-1的坐标分别为（5,0,0）、（0,1,4）和（0,19,4）；无线设备随机分布在以中心坐标为（4,20,0）、半径为2 m的圆形区域内。根据文献[20-22]，大尺度衰落信道建模为$\varsigma\left(d/d_0\right)^{-\alpha}$，其中$\varsigma$表示当参考距离$d_0=1$ m时的路径损耗且$\varsigma=-10$ dB，d是两个节点之间的距离，α表示路径损耗指数。IRS间反射链路的路径损耗指数设置为2.4，其他反射链路的路径损耗指数设置为2.2。各反射链路的小尺度衰落信道建模为莱斯衰落[21-23]，例如HAP和IRS-2间的小尺度信道表示为：$\bar{h}_r=\sqrt{\dfrac{\beta_{hap,irs_2}}{\beta_{hap,irs_2}+1}}\bar{h}_r^{LoS}+\sqrt{\dfrac{1}{\beta_{hap,irs_2}+1}}\bar{h}_r^{NLoS}$，其中，$\beta_{hap,irs_2}$表示HAP与IRS-2间反射链路的莱斯因子，$\bar{h}_r^{LoS}$表示视

距信道分量，$\bar{h}_r^{NLoS}$表示瑞利衰落分量。各反射链路的莱斯因子分别为β_{hap,irs_2} = 10 dB、β_{U_k,irs_2} = 10 dB和β_{irs_1,irs_2} = −10 dB。如无特别说明，其他参数设置如下：B = 1 MHz，δ^2 = −70 dBm，$p_{c,k}$ = 8 mW，η = 0.7，P_B = 44 dBm，$p_{f,k}$ = 2 mW，M = 32。这里我们将文献[22]中的等分信息传输时间方案、IRS随机相位方案和随机能量传输时间方案作为参照方案。对于文献[22]中等分信息传输时间方案，无线设备的信息传输时间相等，需要联合优化能量和信息传输时隙、IRS相位以及无线设备的发送功率；对于随机IRS相位方案，在下行能量传输和上行信息传输过程中，IRS的相位随机生成，只对时隙调度以及无线设备发射功率进行优化；对于随机能量传输时间方案，能量传输阶段的时长随机产生，只联合优化IRS的相位、信息传输的时间以及功率分配。

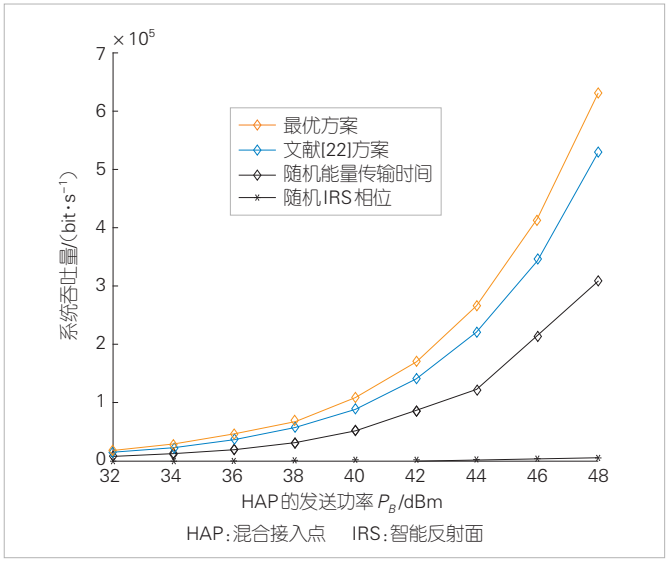
图3中展示了当$M_1 + M_2$ = 32时，IRS–1的反射单元数对系统吞吐量的影响。由图3可知，当总的反射单元数量给定时，随着IRS–1反射单元数增加，系统吞吐量先增后减；当$M_1 = M_2$ = 16，即IRS–1和IRS–2的反射单元数相等时，系统的吞吐量最大。同时，最优方案对应的系统吞吐量明显优于其他参照方案。其中，随机IRS相位方案的系统吞吐量最小。原因如下：当HAP的发送功率P_B不变时，相较于文献[22]中的传输方案，合理分配信息传输时间可以进一步提高系统性能；相较于随机IRS相位方案，IRS的相位优化可以提高能量传输阶段无线设备接收的能量值和信息传输阶段的信道增益，因此最优方案与随机能量传输时间方案下的系统吞吐量明显优于随机IRS相位方案；相较于随机能量传输时间方案，能量传输时间的优化可以均衡能量和信息传输的时间，从而提高系统的吞吐量。



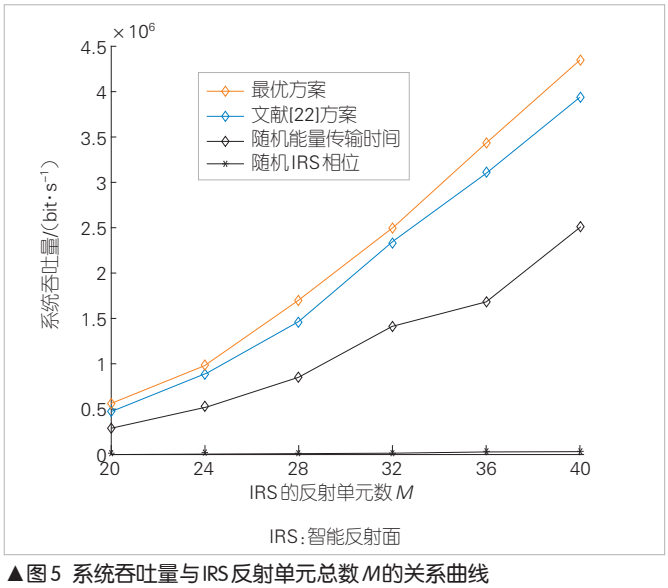
▲图3 系统吞吐量与IRS–1的反射单元数量的关系曲线

图4研究了HAP的发射功率P_B对系统吞吐量的影响。从图4中可以观察到，随着P_B的增大，系统吞吐量逐渐增加。这是因为当P_B增大时，无线设备的接收功率会随之增加，从而提高了系统性能。另外，从图中可以观察出，随着HAP发送功率的增加，最优方案与其他参照方案下的系统性能差异逐渐增加。相较于其他参照方案，当HAP接收功率较高时，优化IRS相位和能量传输时间以及协调多个无线设备间的信息传输时隙能够进一步提高系统性能。

图5为IRS的反射单元数量M对系统吞吐量的影响曲线。随着M的增加，IRS可以为HAP和无线设备之间能量和信息传输提供更多的反射链路，从而使系统的吞吐量增加。优化



▲图4 系统吞吐量与HAP发射功率关系曲线



▲图5 系统吞吐量与IRS反射单元总数M的关系曲线

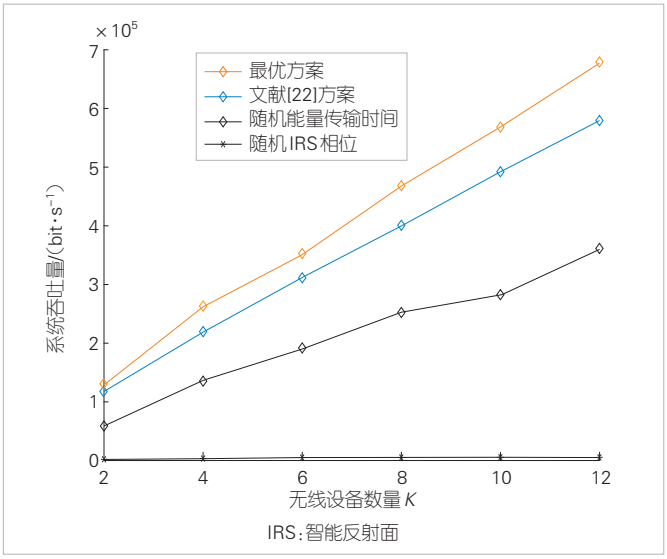
IRS的相位可以提高无线设备的接收功率以及信息传输阶段的信道增益，因此，最优方案、文献[22]传输方案和随机能量传输时间方案下的系统性能优于随机IRS相位方案，并且最优方案可以取得最佳的系统吞吐量。优化能量传输时间可以平衡无线设备的能量收集与信息传输，因而最优方案下的系统性能优于随机能量传输时间方案。优化信息传输的时间可以充分协调多个无线设备的信息传输，进一步提高系统性能。

图6展示了无线设备数量K对系统吞吐量的影响。从图6可知，随着K的增多，系统吞吐量也在增加。这是因为随着K的增多，无线设备收集的能量也会增加，系统吞吐量也随之增加。另外，随着无线设备数量的增加，最优方案与其他参照方案下的系统吞吐量差异逐渐增加。这表明，相较于参照方案，本文提出的最优方案在无线设备数量较多的情况下也可以保证较高的系统性能。

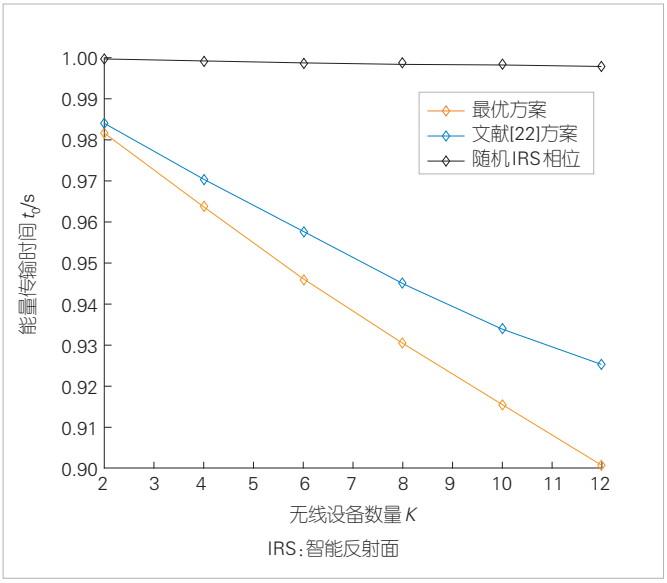
图7展示了下行能量传输时间t_0与无线设备数量K的关系。从图7可知，随着K的增加，系统能量传输的时间t_0逐渐减少。这是因为K增加时，每个无线设备所需信息传输的时间增加，那么系统总时间T在一定的情况下，下行能量传输的时间就会减少。另外，最优方案与其他参照方案中，当K相同时，最优方案下的t_0会更短。这是因为优化IRS的相位可以提高能量的传输效率，使得系统达到最佳性能时所需能量传输的时间更短。

5 结束语

本文中，我们提出了两跳IRS辅助的无线供电通信网络传输方案，有效地解决了HAP和无线设备无法直接通信的难题。为了最大化系统吞吐量，我们还研究了关于时间分配、IRS的相位和无线设备信息传输功率的联合优化问题，通过提出的两阶段迭代优化算法对该问题进行了求解，并获得了高精度的次优解。仿真结果表明，本文提出的最优方案可以明显提升系统的性能。



▲图6 系统吞吐量与无线设备数量K的关系曲线



▲图7 下行能量传输时间t_0与无线设备数量K的关系曲线

参考文献

[1] ZIEGLER S. Considerations on IPv6 scalability for the Internet of Things—towards an intergalactic Internet [C]//Proceedings of 2017 Global Internet of Things Summit (GIoTS). IEEE, 2017: 1–4. DOI: 10.1109/GIOTS.2017.8016238

[2] CHETTRI L, BERA R. A comprehensive survey on Internet of Things (IoT) toward 5G wireless systems [J]. IEEE Internet of Things journal, 2020, 7(1): 16–32. DOI: 10.1109/JIOT.2019.2948888

[3] LU X, WANG P, NIYATO D, et al. Wireless networks with RF energy harvesting: a contemporary survey [J]. IEEE communications surveys & tutorials, 2015, 17(2): 757–789. DOI: 10.1109/COMST.2014.2368999

[4] AOKI T, YUAN Q W, QUANG-THANG D, et al. Maximum transfer efficiency of MIMO- WPT system [C]//Proceedings of 2018 IEEE Wireless Power Transfer Conference (WPTC). IEEE, 2019: 1–3. DOI: 10.1109/WPT.2018.8639417

[5] ZHANG J Y, DAI L L, SUN S Y, et al. On the spectral efficiency of massive MIMO systems with low-resolution ADCs [J]. IEEE communications letters, 2016, 20(5): 842–845. DOI: 10.1109/LCOMM.2016.2535132

[6] LEE D G, KIM T, KIM S, et al. A CMOS rectifier with 72.3% RF-to-DC conversion efficiency employing tunable impedance matching network for ambient RF energy harvesting [C]//Proceedings of 2018 International SoC Design Conference (ISOCC). IEEE, 2019: 259–260. DOI: 10.1109/ISOCC.2018.8649983

[7] JU H, ZHANG R. Throughput maximization in wireless powered communication networks [J]. IEEE transactions on wireless communications, 2014, 13(1): 418–428. DOI: 10.1109/TWC.2013.112513.130760

[8] KIM M U, YANG H J. Min-sinr maximization with dl swipt and ul wpcn in multi-antenna interference networks [J]. IEEE wireless communications letters, 2017, 6(3): 318–321. DOI: 10.1109/LWC.2017.2682248

[9] WU Q Q, CHEN W, NG D W K, et al. Spectral and energy-efficient wireless powered IoT networks: NOMA or TDMA? [J]. IEEE transactions on

vehicular technology, 2018, 67(7): 6663-6667. DOI: 10.1109/TVT.2018.2799947
[10] RAMEZANI P, JAMALIPOUR A. Toward the evolution of wireless powered communication networks for the future Internet of Things [J]. IEEE network, 2017, 31(6): 62-69. DOI: 10.1109/MNET.2017.1700006
[11] BOSHKOVSKA E, NG D W K, ZLATANOV N, et al. Robust resource allocation for MIMO wireless powered communication networks based on a non-linear EH model [J]. IEEE transactions on communications, 2017, 65(5): 1984-1999. DOI: 10.1109/TCOMM.2017.2664860
[12] LI Z R, CHENG L, LEI W, et al. Research on precoding performance optimization of MU-MIMO system based on WPCN slot allocation [C]// Proceedings of 2021 5th International Conference on Automation, Control and Robots (ICACR). IEEE, 2021: 58-63. DOI: 10.1109/ICACR53472.2021.9605197
[13] CHEN H, LI Y H, LUIZ REBELATTO J, et al. Harvest-then-cooperate: wireless-powered cooperative communications [J]. IEEE transactions on signal processing, 2015, 63(7): 1700-1711. DOI: 10.1109/TSP.2015.2396009
[14] ZENG Y, CHEN H B, ZHANG R. Bidirectional wireless information and power transfer with a helping relay [J]. IEEE communications letters, 2016, 20(5): 862-865. DOI: 10.1109/LCOMM.2016.2549515
[15] WU Q Q, ZHANG S W, ZHENG B X, et al. Intelligent reflecting surface-aided wireless communications: a tutorial [J]. IEEE transactions on communications, 2021, 69(5): 3313-3351. DOI: 10.1109/TCOMM.2021.3051897
[16] DI RENZO M, DEBBAH M, PHAN-HUY D T, et al. Smart radio environments empowered by reconfigurable AI meta-surfaces: an idea whose time has come [J]. EURASIP journal on wireless communications and networking, 2019, 2019(1): 1-20. DOI: 10.1186/s13638-019-1438-9
[17] WU Q Q, ZHANG R. Towards smart and reconfigurable environment: intelligent reflecting surface aided wireless network [J]. IEEE communications magazine, 2020, 58(1): 106-112. DOI: 10.1109/MCOM.001.1900107
[18] PAN C H, REN H, WANG K Z, et al. Multicell MIMO communications relying on intelligent reflecting surfaces [J]. IEEE transactions on wireless communications, 2020, 19(8): 5218-5233. DOI: 10.1109/TWC.2020.2990766
[19] WU Q Q, ZHANG R. Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints [J]. IEEE journal on selected areas in communications, 2020, 38(8): 1735-1748. DOI: 10.1109/JSAC.2020.3000807
[20] SONG D, SHIN W, LEE J. A maximum throughput design for wireless powered communication networks with IRS-NOMA [J]. IEEE wireless communications letters, 2021, 10(4): 849-853. DOI: 10.1109/LWC.2020.3046722
[21] LYU B, RAMEZANI P, HOANG D T, et al. Optimized energy and information relaying in self-sustainable IRS-empowered WPCN [J]. IEEE transactions on communications, 2021, 69(1): 619-633. DOI: 10.1109/TCOMM.2020.3028875
[22] ZHANG D C, WU Q Q, CUI M, et al. Throughput maximization for IRS-assisted wireless powered hybrid NOMA and TDMA [J]. IEEE wireless communications letters, 2021, 10(9): 1944-1948. DOI: 10.1109/LWC.2021.3087495
[23] YOU C S, ZHENG B X, ZHANG R. Wireless communication via double IRS: channel estimation and passive beamforming designs [J]. IEEE wireless communications letters, 2021, 10(2): 431-435. DOI: 10.1109/LWC.2020.3034388
[24] 谢天怡. 基于非正交多址接入的无线供电通信网络吞吐量优化问题研究 [D]. 南京: 南京邮电大学, 2019
[25] EL SHAFIE A, NIYATO D, AL-DHAHIR N. Security of an ordered-based distributive jamming scheme [J]. IEEE communications letters, 2017, 21(1): 72-75. DOI: 10.1109/LCOMM.2016.2615043
[26] PEJOSKI S, HADZI-VELKOV Z, SCHOBER R. Optimal power and time allocation for WPCNs with piece-wise linear EH model [J]. IEEE wireless communications letters, 2018, 7(3): 364-367. DOI: 10.1109/LWC.2017.2778146

作 者 简 介

冯璇，南京邮电大学在读硕士研究生；主要研究方向为无线供电通信网络、智能反射面。

吕斌，南京邮电大学副教授；长期在无线通信网络和智能物联网领域从事教学和科研工作，主要研究方向包括无线通信供电网络、反向散射通信、共生无线电、中继协同通信和可重构智能表面等无线通信理论前沿及关键技术；发表论文30余篇。

杨震，南京邮电大学教授、工业和信息化部科技委委员、国家发展改革委通信与网络技术国家地方联合工程研究中心主任；长期从事信号与信息处理、通信理论与技术的教学和科研工作；主持国家科技支撑计划重点项目、国家“973”课题、国家“863”重点课题及面上项目、国家科技重大专项、国家自然科学基金等项目近30项；发表论文300余篇，出版专著2部，获得国家发明专利30余项。

面向卫星通信系统的寻呼方法

Paging Method for Satellite Communication System

毛玉欣/MAO Yuxin[1]，闫新成/YAN Xincheng[2,3]
（1. 小米通讯技术有限公司，中国 北京 100085；
2. 移动网络和移动多媒体技术国家重点实验室，中国 深圳 518055；
3. 中兴通讯股份有限公司，中国 深圳 518057）
(1. Xiaomi Corporation, Beijing 100085, China;
2. The State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China;
3. ZTE Corporation, Shenzhen 518057, China)

DOI:10.12142/ZTETJ.202301013
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20230222.1655.002.html
网络出版日期：2023-02-23
收稿日期：2022-10-15

摘要：分析了卫星通信中的固定追踪区（TA）和卫星移动小区的配置特性，以及由此导致的通信寻呼效率低下问题。针对卫星通信系统提出了一种基于动态追踪区（DTA）的寻呼方法，包括基于用户位置、移动模式、签约信息为用户设备（UE）定制DTA以及基于UE-DTA进行UE寻呼。该方法可实现寻呼信令载荷和TA更新信令频率之间的平衡，减轻寻呼信令负荷，提高寻呼效率。

关键词：卫星通信；非地面网络；寻呼；动态追踪区

Abstract: The configuration characteristics of earth-fixed tracking areas (TA) and earth-moving cells in satellite communication systems and the resulting of paging inefficiency are analyzed. A paging optimization method based on dynamic tracking area (DTA) is proposed for the satellite system, which includes determining a customized DTA for user equipment (UE) by the network based on UE position, mobility patterns, and subscription information, and performing paging request based on UE-DTA. It achieves a better trade-off between paging signalling load and registration update signalling, reduces paging signalling load and improves paging efficiency.

Keywords: satellite telecommunication; non-terrestrial network; paging; dynamic tracking area

无线移动通信系统最初可以满足人们的语音通信需求，经过30年的发展，如今可以提供高速数据通信服务。目前陆地无线移动通信系统已经为全球大多数人口所在区域提供较为完善的网络覆盖服务，全球80%的人口都可以享受到移动通信服务。随着5G技术的普及，网络应用从面向人的通信扩展到面向人和物以及物与物的通信。这使得通信应用场景不断丰富，通信服务的范围不断扩大。在促进社会全面数字化的同时，5G技术也因受制于经济成本、技术、自然条件等因素，在人口密度较低的偏远地区以及人迹罕至的高山、荒漠、远洋等区域难以实现5G蜂窝接入的普遍覆盖。而卫星通信因为具有广域覆盖的特点，可以以陆地蜂窝通信系统难以比拟的成本优势提供广域甚至全球通信覆盖服务，从而对陆地蜂窝通信覆盖形成有效补充。因此，构筑天地一体化通信网络，提供无缝覆盖的网络服务，是5G和6G技术研究的重点领域之一[1-5]。随着卫星通信技术的发展，单星服务能力和星链技术都得到了有效提升，服务的业务场景以及部分技术指标也越来越接近陆地蜂窝移动通信。这些均使得天地一体化通信深度融合的紧迫性进一步加强。作为全球移动通信技术标准制定的主要组织，第3代合作伙伴计划（3GPP）在5G标准的第16版中启动了5G支持卫星通信的标准研究和制定工作[6]。

1 卫星通信简介

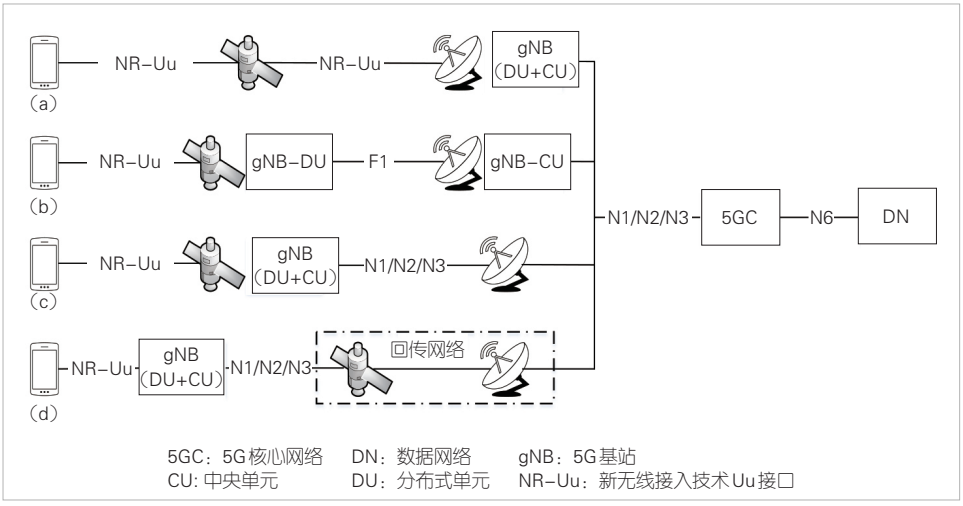
1.1 卫星通信类型和架构

受空气密度、太空碎片以及已经在轨运行的低轨卫星的限制，同时考虑到2 000~8 000 km高度的范艾伦辐射带上的高能粒子的影响，通信卫星通常运行在下列几类轨道上[7]：

1）地球同步轨道（GEO）卫星。这类卫星运行在高度为35 786 km的赤道平面上，相对地球保持静止，其卫星轨道周期与地球自转周期相同，因此可以为信号覆盖区域提供持续信号覆盖。

2）非地球同步轨道（NGSO）卫星。这类卫星相对于地面移动。某个指定区域如果需要长期稳定的信号覆盖，就需要若干卫星组建星群来满足这一需求。NGSO运行高度越低，提供持续稳定信号覆盖所需的卫星数就越多。按照运行高度的不同，NGSO通常又可以分为运行在500~2 000 km的近地轨道（LEO）卫星，运行在8 000~20 000 km的中地轨道（MEO）卫星和运行在7 000~45 000 km的高偏心轨道（HEO）卫星。

当前针对天地一体化通信的研究主要以陆地移动通信技术标准为基础，结合卫星通信的技术特点做出适应改进。在天地融合通信的实现过程中，卫星通信通常作为地面接入覆盖的延伸，或者作为在基站（gNB）和核心网络（5GC）之间的回程传输。按照卫星通信的作用以及不同层面的融合类型，卫星通信可分为如图1所示的几种场景[8]。



▲图1 卫星通信部署场景

按照卫星在通信系统中发挥作用的不同，卫星通信可以分为弯管模式和再生模式。图1中的（a）方式是卫星工作在弯管模式，该模式下卫星仅负责通信数据的透明转发；（b）和（c）两种方式是再生模式，该模式下卫星承担了基站的部分或者全部功能，并在发送数据之前，需要承担数据的处理功能；（d）是卫星作为回程传输使用，在gNB和5GC之间转发上下行数据。

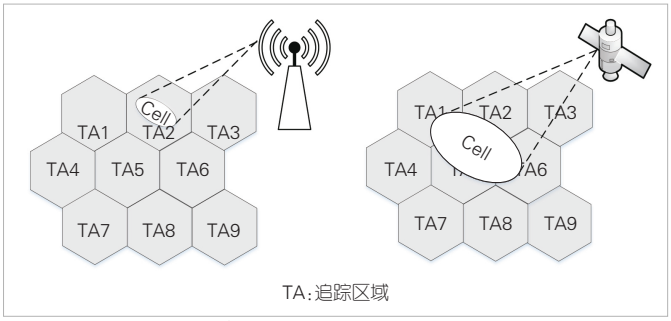
1.2 卫星通信的移动性管理

追踪区域（TA）是移动通信系统为用户设备（UE）位置的移动管理设立的概念。当UE处于空闲态时，5GC能够知道UE所处的TA。同时当处于空闲态的UE需要被寻呼时，必须在UE注册的TA的所有小区中进行寻呼。在UE进行网络注册时，网络会根据UE的位置分配注册区（RA）。一个RA包含TA列表，并且一个TA列表可以包含一个TA或者多个TA。UE在该TA列表包含的TA内移动时不需要执行TA更新过程，从而较少与网络的频繁交互。当UE移动并进入不被RA包含的新TA时，注册更新过程会被触发执行以更新TA。此时5GC会给UE分配一组新的TA。新分配的TA可以完全不包含或者部分包含原来RA中的一些TA。

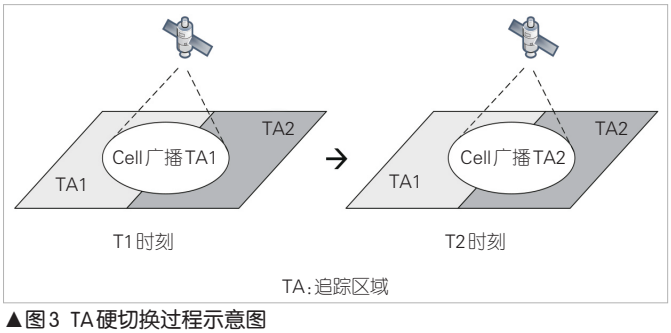
在陆地蜂窝通信系统中，一个TA区域通常大于小区（Cell）的范围。TA是Cell级别的配置。多个Cell可以配置相同的TA，但一个Cell只能属于一个TA[9]。在卫星通信系统中，Cell由卫星发射的波束覆盖地面形成。一个Cell可由单个或者多个卫星波束构成。Cell的大小和卫星波束半径相关。卫星波束半径随着卫星高度的增加而增大，例如：LEO的波束半径可达数十公里，MEO/GEO的波束半径可达数百公里。因此，与地面蜂窝通信的Cell和TA规划不同，卫星通信中一个Cell的覆盖范围就可能涵盖多个TA[10]。图2说明了陆地蜂窝通信和卫星通信中TA和Cell关系的区别。

NGSO通信中的卫星相对地面是移动的，因此向地面广播的Cell也在不断移动。这就要求Cell不断更新向地面广播的TA。在卫星通信中有两种TA广播方式[11]。

一种TA广播方式是卫星Cell广播单个TA，如图3所示。当卫星从西向东移动并跨越两个TA区域时，在T1时刻卫星波束对TA1的覆盖范围大于对TA2的覆盖范围，此时Cell广播TA1；随着Cell的移动，在TA2时刻，卫星波束对TA2的覆盖范围大于对TA1的覆盖范围，此时Cell需要进行TA更新，并向地面广播TA2。这种由于卫星移动引发的TA1向TA2的切换就像是接入到该Cell的UE发生了跨TA的移动一样。如果此前TA2不包含在UE的RA中，上述广播TA变化



▲图2 Cell和TA关系示意图



▲图3 TA硬切换过程示意图

的过程就需要触发UE执行注册更新，并将TA2更新到RA中。这种广播单个TA下的切换称为硬切换。

另一种TA广播方式是软切换。卫星Cell广播多个TA，例如：被卫星波束覆盖的TA都会在Cell中广播，如图4所示。在T1、T2时刻，Cell总是广播{TA1,TA2}。在这两个时刻，UE的RA只要至少包含上述TA的一个，就会允许UE接入该Cell，并且在卫星从T1到T2的移动过程中不需要触发UE来发起注册更新过程。在T3时刻，Cell覆盖已经完全移出了TA1，此时Cell需要对广播的TA进行更新，并向地面广播{TA2,TA3}。如果T3时刻UE在TA2区域内，且RA仅包含TA1，则UE发起注册更新过程并将RA更新为{TA2,TA3}。

受卫星部署高度的影响，LEO和MEO的卫星波束覆盖范围通常在100~1 000 km，而地面无线通信系统使用的TA划分范围通常在100~200 km，因此卫星通信的Cell覆盖范围是TA的数十倍。这就意味着卫星通信如果使用现有的TA规划，硬切换会导致UE频繁进行注册更新，进而大大增加终端和网络设备的信令处理开销和网络通信负担，而软切换可以减少UE在TA边界处发生TA更新的次数，从而有效减少和网络交互信令的频度。

2 卫星通信中的寻呼方法

2.1 无线移动通信的寻呼

UE在空闲态时会释放与网络的连接。当网络需要与空闲态的UE进行通信时，例如网络收到发往UE的下行数据，首先需要对UE发起寻呼过程。当寻呼到UE并恢复网络连接后，网络才可以向UE发送数据。对UE寻呼的过程需要知道UE所处的TA，并向所述TA所在的小区发起对UE的寻呼；如果不知道UE当前所处的TA，则需要向UE注册的所有TA所在的小区发起寻呼。

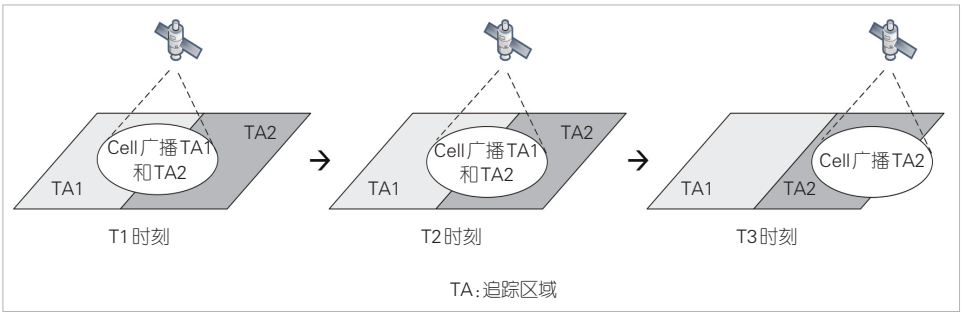
以5G网络寻呼为例，如图5所示，如果UE处于CM_IDLE态或者处于CM_CONNECTED with RRC_INACTIVE态，此时N2、N3连接的资源会因空闲而被释放。当5GC侧有下行数据到达用户面功能（UPF）时，由于与UE之间无可用连接，UPF无法将所述下行数据发送给UE。在UPF根据指令缓存所述数据或者将所述数据发送给会话管理功能（SMF）缓存后，SMF将向接入和移动性管理功能（AMF）触发寻呼流程。AMF通过gNB向UE发送寻呼消息，通知UE恢复会话上下行数据链路。

根据TA进行寻呼过程的示意如图5所示。网络会记录UE最近访问的TA并将其作为UE的位置，以便根据该TA信息进行UE寻呼。如果UE仍然在所述TA内，则寻呼成功，恢复连接；如果UE在转换为空闲态后发生了位置移动（已经不在所述TA区域内），则寻呼失败，接下来需要根据UE的RA进行寻呼。寻呼会根据RA中的TA列表以及TA和Cell的关系进行。

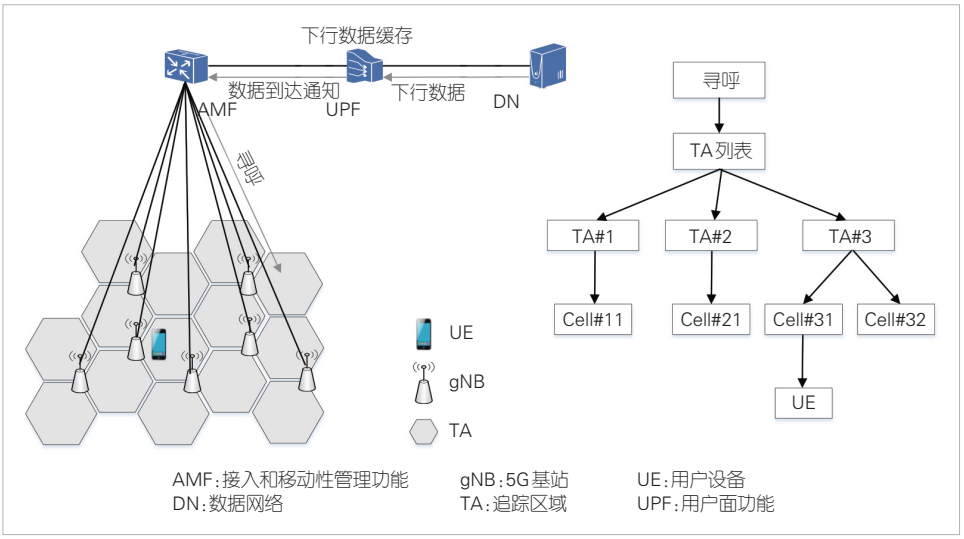
从寻呼示例可以看出，寻呼效率的高低取决于网络是否知道寻呼时刻UE所处的TA。寻呼策略可以优先使用UE最近使用的TA进行寻呼，以尽量缩短寻呼时间。当基于精确TA寻呼失败时，网络才会基于RA寻呼。

2.2 卫星通信中的寻呼问题和改进

在移动通信系统中，网络基于TA对UE发起寻呼。TA



▲图4 TA软切换过程示意图



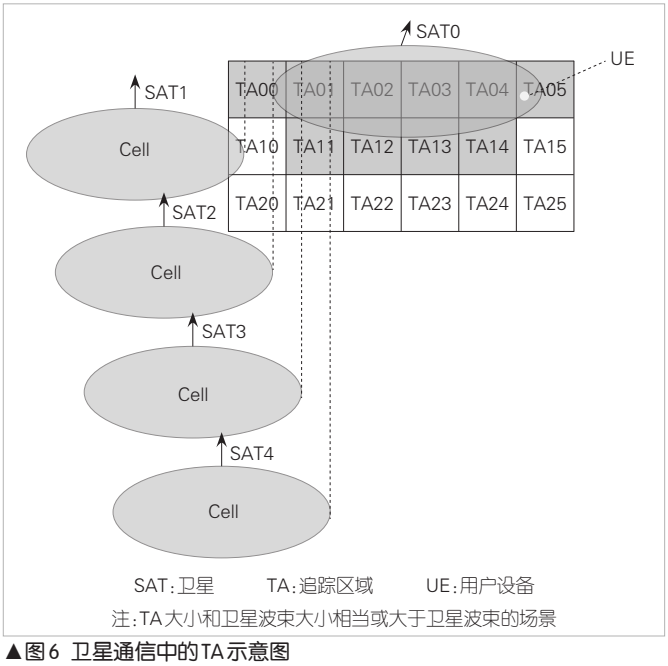
▲图5 寻呼过程示意图

没有严格的地理定义，而是由构成TA的一组Cell的位置和覆盖范围决定的。无论卫星通信中的卫星波束相对地面是固定的，还是移动的，对TA的定义都被认为相对地面固定。1.2节描述了卫星通信的软切换模式。在该模式下，卫星Cell需要广播多个TA编码（TAC）。UE的RA中会包含所述广播TAC。由于TA边界无法被UE感知，UE只能根据从卫星Cell的广播消息中接收TAC并判断是否发生了TA更新，即当任一广播的TAC都不被RA包含时，UE才触发发起TA更新。这种模式可以有效减少由UE的TA更新导致的和网络信令交互的频率。但是在软切换模式中，在没有从终端接收到有用位置信息的情况下，每当与该UE相关联的任何潜在TAC被覆盖时，网络必须尝试对该UE进行寻呼。这实质上增加了寻呼负荷。如图6所示，UE位于卫星Cell初始飞越覆盖的边缘TA05区域。如果gNB没有获取到UE的有用位置信息，系统便将公共陆地移动网（PLMN）下所有黑色标记相关的TAC发送给5GC。由于地球自转和卫星轨道之间的周期不同，后续飞越的卫星对同一区域的覆盖可能会出现一些偏移，例如：随后的卫星Cell飞越覆盖所述TAC中另外一端的TA00。从SAT0到SAT4的覆盖过程中，如果发起寻呼，在网络未得到UE的精确位置信息时，只要黑色标记TAC中的任何一个被覆盖，所有黑色标记的TAC都需要尝试寻呼UE。

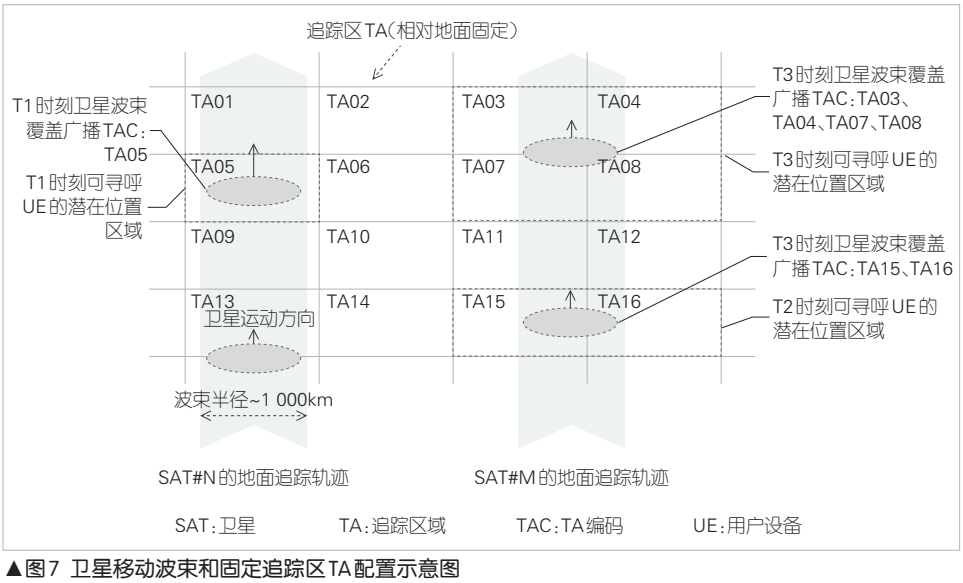
随着时间的推移，网络可以完全排除没有收到UE寻呼应答对应的TAC，从而减少寻呼开销。但这个过程意味着在找到正确的TAC之前，不同卫星的多次飞行会持续产生寻呼开销。

这种相对地面移动的卫星Cell增加寻呼信令开销的情况将随卫星波束的扩大而加剧。这种情况可通过图7进一步说明。为了简化表示，但又不失代表性，在地表上我们用矩形网格表示TA，且假设TA大小和卫星波束大小相当。图7还说明了由于地球自转和卫星轨道之间的周期不同，卫星随后经过同一区域时可能会出现一些偏移。因此，将TA布局安排为与卫星地球轨道重合是不可行的。从图7可以看出，如果仅根据TA信息进行寻呼，卫星波束覆盖最多可与4个TA重叠，如图7中T3时刻的卫星波束覆盖。在这种情况下，寻呼UE的次数将明显增多，随后无线接入网络（RAN）上的寻呼资源开销也会变大，寻呼失败的次数也会增加。

为了提高寻呼效率，可以考虑使TA小于卫星波束覆盖范围并且基于更准确的UE位置信息进行寻呼。卫星通信中定义了一个用Mapped Cell ID标识且与Cell ID相对应的固定地理区域[12]。这个区域与卫星轨道或者UE和卫星之间的连接（Service Link）的类型无关，如图8所示。Mapped Cell ID和地理区域的映射关系可以根据运营商策略预先配置在RAN和核心网（CN）上。RAN负责从UE接收的位置信息，例如全球导航卫星系统（GNSS）信息，以构建Mapped Cell ID。当UE触发注册或注册更新流程时，RAN将广播TAC上



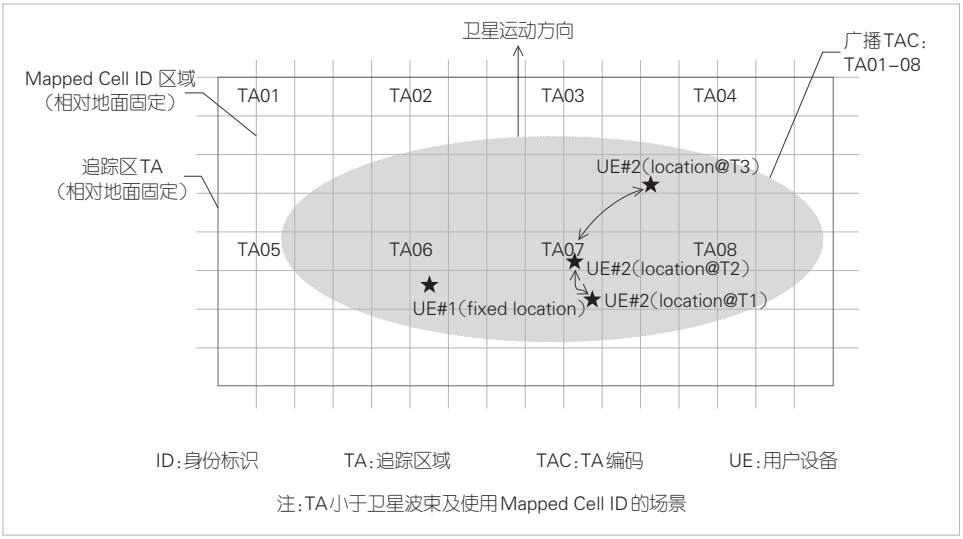
▲图6 卫星通信中的TA示意图



▲图7 卫星移动波束和固定追踪区TA配置示意图

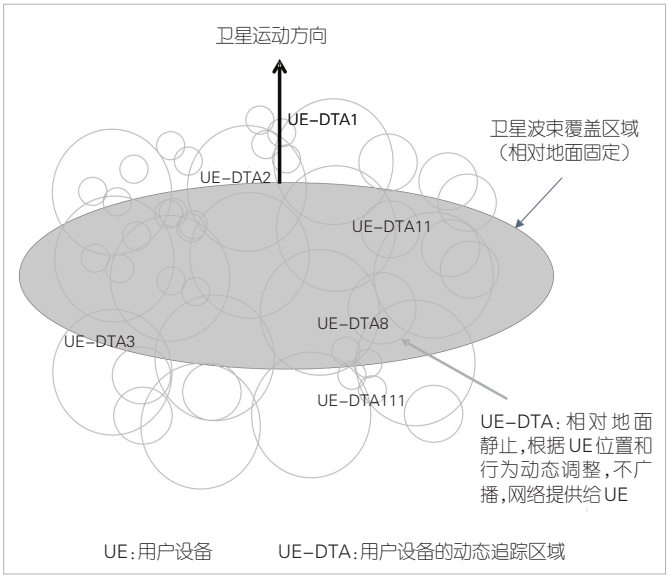
报给5GC。同时，如果RAN获知UE的具体位置信息，还需要基于所述位置信息确定UE当前所处TA的TA标识（TAI），并将该TAI也发送给CN。

对于静止UE，如图8中的UE#1，当UE#1的RA、Mapped Cell ID被卫星波束所覆盖时，依据UE所处的TA以及Mapped Cell ID就可以快速寻呼到UE#1。但对于处于移动的UE，如图8中的UE#2，沿着卫星的移动轨迹移动跨越了Mapped Cell ID对应区域甚至TA对应的区域。在所述UE#2的移动过程中，只要不尝试接入网络，网络就不会感知到所述移动过程。而在这个移动过程中，由于卫星波束覆盖范围很大，并且卫星Cell会广播覆盖内的多个可达的TAC，因此在所述广播TAC对应的TA区域内移动时，UE#2本身也无法感知到Mapped Cell ID对应区域的变更，甚至无法感知TA的变更。例如：如图8所示，UE#2在T1时刻从TA07注册到网络，当在T3时刻移动到TA03时，UE#2仍然认为在TA07。这种场景下，如果寻呼策略基于Mapped Cell ID或者基于注册时所处的TA将被证明不再高效。解决这个问题的一个方法就是增加定期注册的频度以获取近乎实时的UE位置，但这显然会增加信令开销。



▲图8 卫星移动波束和固定追踪区TA配置示意

对此，一种改进寻呼效率的方法是定义针对UE的动态追踪区域（UE–DTA）。UE–DTA可以基于UE位置的地理围栏和地理区域规范，为每个UE建立定制化的追踪区域。UE–DTA区域的定义有多种方式：可以通过参数化来定义一个几何图形，例如圆、椭圆或多边形，也可以通过作为UE–DTA区域顶点的一组地理坐标来构成电子围栏。为便于表述，这里UE–DTA使用圆形区域，并用中心点和半径表示。另外，在构建UE–DTA区域时，除了参考UE的位置信息之外，UE–DTA的大小和形状也可以基于UE的移动性、UE部署密度等其他条件进行动态调整，例如：快速移动的UE可以定义范围较大的UE–DTA，而静态或者准静态UE则需要定义范围较小的UE–DTA。当为UE分配UE–DTA时，多个UE的UE–DTA可以相同。图9展示了动态卫星波束下的UE–DTA。网络将为UE制定的UE–DTA提供给UE保存。UE基于UE–DTA、UE位置信息和地理区域描述（GAD）来监测位置移动是否超出了UE–DTA区域。



▲图9 卫星移动波束和UE–DTA配置示意

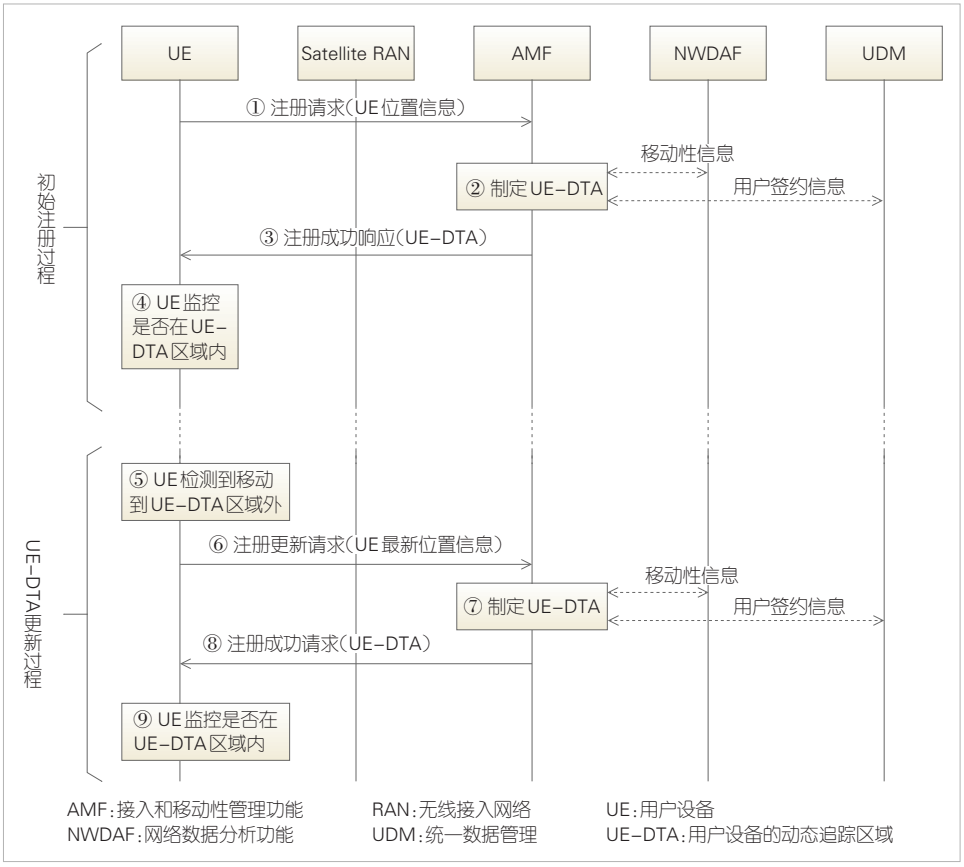
2.3 基于UE–DTA的寻呼实现和分析

UE–DTA依赖于网络根据UE的位置等信息动态变化并提供给UE。UE在移动过程中会根据UE–DTA判断是否需要触发位置更新。网络可以根据UE–DTA进行UE寻呼。基于UE–DTA的寻呼过程可分为UE–DTA制定过程和基于UE–DTA的寻呼过程，如图10和图11所示。

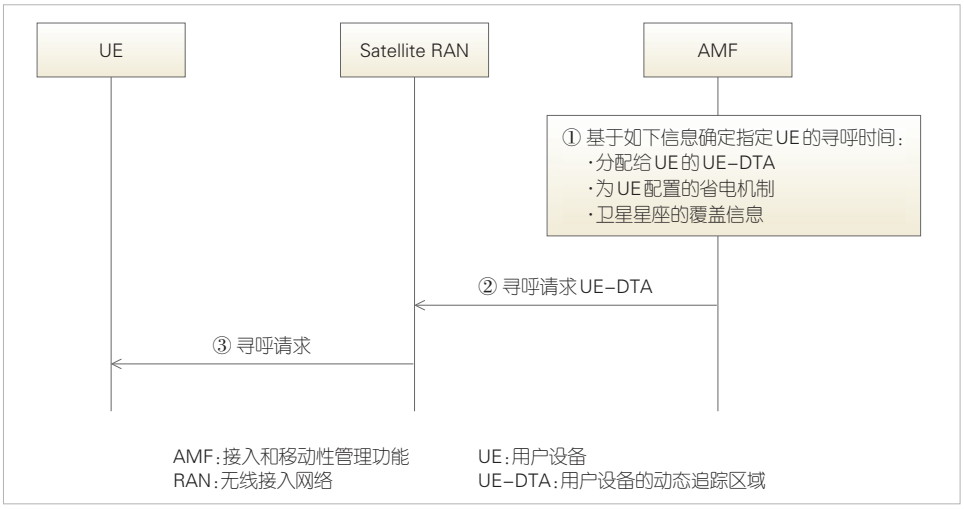
1）基于注册/注册更新过程的UE–DTA制定

① 当UE接入网络发起注册请求时，UE向网络提供位置信息，例如GNSS信息。

② AMF为所述UE接入确定UE–DTA。UE–DTA的计算过程需要结合UE当前位置信息、签约信息、移动模式等。



▲图10 UE-DTA制定和更新流程



▲图11 AMF基于UE-DTA触发的寻呼过程

这些信息可以从签约数据库、网络数据分析功能（NWDAF）中获取。

③ AMF通过注册成功消息将UE-DTA提供给UE保存。

④ UE监控自身的移动，判断自身位置是否在UE-DTA区域内。

⑤ 由于位置移动，UE检测到位于UE-DTA之外

⑥ UE触发注册更新流程，要求进行TA更新。同时UE将当前位置提供给AMF。

⑦ 根据UE提供的新的位置信息，AMF计算并产生新的UE-DTA。

⑧ AMF将更新的UE-DTA提供给UE。

⑨ UE使用新的UE-DTA覆盖原有的UE-DTA，并根据最新的UE-DTA对位置移动进行监测。

2）基于UE-DTA的UE寻呼

① AMF确定触发UE寻呼过程的时间或者时间窗。这个确定过程基于分配给UE的UE-DTA和卫星星座的信号覆盖信息，例如星历信息、卫星波束信息、配置的省电机制。

② AMF向给UE当前位置提供覆盖的RAN节电发起寻呼过程。另外，AMF可以将UE-DTA提供给RAN节点，由RAN结合卫星覆盖信息决定发起UE寻呼的时机。

③ RAN通过Service link向UE发起寻呼过程。

基于UE-DTA实现寻呼的方案需要对现有系统进行如下两个方面的增强：

1）在UE侧，需要UE在注册阶段，向5GC提供当前位置信息，以便网络结合UE提供的位置信息进行UE-DTA的制定。UE需要保存网络提供的UE-DTA信息，并根据UE-DTA信息实时监测位置是否位于UE-DTA范围内。如果UE检测到位置位于UE-DTA范围之外，则需要触发注册更新过程，请求网络进行UE-DTA更新。

2）在核心网络侧，需要基于UE提供的位置信息，并结合签约信息、UE移动模式的分析数据、地理区域描述等信息为UE制定UE-DTA，随后向UE提供UE-DTA。此外网络根据UE-DTA参数以及卫星星座的覆盖信息，计算并确定

UE–DTA是否被卫星星座覆盖。如果存在覆盖，就可根据UE–DTA向UE发起寻呼。

动态确定UE–DTA并基于UE–DTA进行UE寻呼的方法主要具有3个功能特点：

1）基于UE–DTA进行终端寻呼。由于UE–DTA的范围小于TA范围，只有被卫星波束覆盖的UE–DTA才会被寻呼，寻呼请求的调度可以根据卫星沿着UE–DTA的移动进行调整，这样可以对卫星波束覆盖下的终端寻呼进行优化，有效减少寻呼信令载荷。

2）平衡寻呼信令载荷和信令交互频率。得益于UE–DTA的范围可以根据UE的位置特性（UE静止、移动、高速移动）、部署环境（例如部署密度）等进行动态调整，基于UE–DTA的寻呼方法可以在寻呼信令载荷和TA更新信令之间实现更好的平衡。

3）可实现更灵活的寻呼和移动性管理解决方案。对于基于TA的寻呼，由于TA的范围较大，且不与任何UE的位置关联，TA的地理边界也不被UE感知，因此很难确定寻呼UE的范围。对于基于UE–DTA的寻呼，由于UE–DTA可以根据需要动态定义，UE可以根据GNSS测量出UE是停留在UE–DTA内，还是移动到UE–DTA外。这使得寻呼更加高效灵活。

3 结束语

卫星通信作为5G地面通信的补充，可以为地面蜂窝覆盖难以到达的区域提供补充覆盖，从而使万物互联变为可能。卫星通信在提供通信便利的同时，也因为卫星高度、星座密度、波束半径等具体的因素，在通信效率方面还需要不断发展和改进。基于UE–DTA的寻呼方法通过动态制定UE的寻呼范围，有效地提高卫星通信过程中的寻呼效率，减少寻呼负荷。然而，制定UE–DTA所需的UE位置信息的可靠性、位置信息的粒度，以及位置信息作为隐私如何保证安全，均需要做进一步的研究。另外，除了应用于寻呼过程，UE–DTA是否还可应用于注册区管理、服务域限制管理等，以进一步增强移动性管理的效率，也需要做进一步的探索验证。

3GPP非地面通信（NTN）标准研究的推进和卫星能力的增强，推动着卫星通信技术逐步迈向成熟。5G NTN未来应用场景广阔，将支持更多频段，使星上处理能力不断提高。这将推动卫星产业和蜂窝技术的深度融合，极大拓展卫星通信的应用范围，催熟相关产业链条。卫星通信将广泛应用于个人领域和垂直行业应用领域。在打造偏远地区、海洋、民航等全域泛在连接，丰富应急通信等方面，卫星通信技术将发挥巨大的商用和社会价值。

参考文献

[1] 田开波，杨振，张楠. 空天地一体化网络技术展望 [J]. 中兴通讯技术, 2021, 27(5): 2–6. DOI: 10.12142/ZTETJ.202105002
[2] 孙智立，李天儒. 大规模低轨星座卫星通信网发展展望 [J]. 中兴通讯技术, 2021, 27(5): 48–51. DOI: 10.12142/ZTETJ.202105010
[3]YAN X C, TENG H Y, PING L, et al. Study on security of 5G and satellite converged communication network [J]. ZTE Communications, 2021, 19(4): 79–89. DOI: 10.12142/ZTECOM.202104009
[4] MA Y Y, MA G Y, WANG N, et al. OTFS enabled NOMA for mMTC systems over LEO satellite [J]. ZTE Communications, 2021, 19(4): 63–70. DOI: 10.12142/ZTECOM.202104007
[5] IMT–2030. 6G总体愿景和关键潜在技术白皮书 [R]. 2021
[6] 3GPP. New WID on integration of satellite access in 5G (5GSAT) [EB/OL]. [2022–11–20]. https://www. 3gpp. org/ftp/tsg_sa/TSG_SA/TSGS_80/Docs/SP–180326.zip
[7] 3GPP. Study on using satellite access in 5G, stage 1: 3GPP TR 22.822 [S]. 2018
[8] 3GPP. Study on architecture aspects for using satellite access in 5G, stage 2: 3GPP TR 23.737 [S]. 2019
[9] 3GPP. General packet radio service (GPRS) enhancements for evolved universal terrestrial radio access network (E–UTRAN) access (release 17): 3GPP TS 23.401 [S]. 2020
[10] 3GPP. System architecture for the 5G system (5GS), stage 2 (release 17): 3GPP TS 23.501 [S]. 2021
[11] 3GPP. Support of mobility registration update for 5G satellite access [EB/OL]. [2022–11–20]. https://www. 3gpp. org/ftp/tsg_sa/TSG_SA/TSGS_93E_Electronic_2021_09/Docs/SP–210916.zip
[12] 3GPP. Summary for NB–IoT/eMTC support for non–terrestrial networks (NTN) [EB/OL]. [2022–11–20]. https://www. 3gpp. org/ftp/tsg_ct/TSG_CT/TSGC_96_Budapest/Docs/CP–221272.zip

作 者 简 介

毛玉欣，小米通讯技术有限公司标准技术高级工程师；主要研究方向为6G网络架构、天地一体化通信和网络安全；参与多项国际技术标准的制定，拥有发明专利90余项、国际标准提案70余篇。

闫新成，中兴通讯股份有限公司网络安全系统架构首席专家，正高级工程师，江苏省“333高层次人才”；曾主持或参与国家科技重大专项课题，获多项省部级科技奖励；拥有专利40余项。

中兴通讯技术杂志社
促进产学研合作青年专家委员会

刊物相关信息

投稿须知　投稿平台　过刊下载　论文索引与引用指南

中兴通讯技术

(ZHONGXING TONGXUN JISHU)

办刊宗旨：

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

双月刊　1995 年创刊　总第 168 期
2023 年 2 月　第 29 卷 第 1 期

主管：安徽出版集团有限责任公司
主办：时代出版传媒股份有限公司
　　深圳航天广宇工业有限公司
出版：安徽科学技术出版社
编辑、发行：中兴通讯技术杂志社

总编辑：王喜瑜
主编：蒋贤骏
执行主编：黄新明
编辑部主任：卢丹
责任编辑：徐烨
编辑：杨广西、朱莉、任溪溪
设计排版：徐莹
发行：王萍萍
编务：王坤

《中兴通讯技术》编辑部
地址：合肥市金寨路 329 号凯旋大厦 1201 室
邮编：230061
网址：tech.zte.com.cn
投稿平台：tech.zte.com.cn/submission
电子信箱：magazine@zte.com.cn
电话：(0551)65533356

发行方式：自办发行
印刷：合肥添彩包装有限公司
出版日期：2023 年 2 月 28 日
中国标准连续出版物号：ISSN 1009-6868
CN 34-1228/TN
定价：每册 20.00 元