专 题 导 读

NewIP 继承传统 IP 的成功基因，进行数据网络协议架构创新，通过持续增强 IP 网络的能力，解决万网互联与万物互联，使能更多的新服务接入网络，丰富人们的沟通与生活。

本专题由 7 篇文章组成，其中，《NewIP：开拓未来数通网络的新连接和新能力》整体介绍万网互联的统一网络协议——NewIP 的设计理念以及几个关键使能技术，包括确定性 IP、内生安全、面向万网互联的新寻址与控制机制、用户可定义及新传输层等。《大规模确定性网络转发技术》详细介绍了确定性 IP 的技术背景、原理及工作机制等，并通过仿真实验对比了在相同网络环境下，传统 IP 及确定性 IP 在端到端最差时延及抖动上的差异，证明了确定性 IP 技术的有效性。《内生安全网络架构》详细介绍了具有内生安全特性的网络架构，并重点阐述具有内生安全的隐私 ID/Loc、安全验证和审计协议、跨域联合防御机制等核心安全技术。《基于微服务架构的下一代 IP 网络测试体系框架》提出了在基于微服务架构的基础上强调对业务的赋能，综合不同应用场景特征，建立基于下一代 IP 网络场景的全过程领域驱动模型。《面向计算网络融合的下一代网络架构》根据对应用、网络技术、计算技术等发展趋势的分析，提出计算网络融合的一体化网络架构应基于新型 IP 网络体系，支持网络可编程、函数能力寻址、确定性网络传输，算力感知的协同调度和控制等功能，向各相关产业提供网络能力、计算能力及数据能力服务，并使其更加有效地满足万物互联、万物智能、万物感知的需求。此外，《基于云、网、边融合的边缘计算新方案：算力网络》提出了基于云、网、边深度融合的算力网络方案，该方案能够有效应对未来业务对计算、存储、网络甚至算法资源的多级部署以及在各级节点之间的灵活调度。《温敏网络的关键能力和架构体系》提出了温敏网络的几种关键技术能力，能够很好地应对低时延、高带宽、易流量微突发的未来网络，可作为复杂多变网络的基础能力。

网络 5.0 产业和技术创新联盟接口与协议组负责万物互联和智能社会下的 IP 协议的研究与制定。作为一种协议体系创新研究，NewIP 尚有大量的技术细节待业界共同研究和完善。

本专题文章作者来自于数据通信领域的高等院校、科研院所及设备制造商等，他们在该领域有着丰富的研究成果和科研经验。希望他们的观点能够帮助读者了解数据网络的协议体系创新进展，并有更多的读者关注网络 5.0 的研究动态。在此对各位作者对本专题的大力支持和辛勤付出表示衷心的感谢！

专题策划人：

蒋林涛，男，中国信息通信研究院科技委员会主任，工业和信息化部通信科技委员会常务委员，原信息产业部电信研究院总工程师，长期从事多媒体技术、数据通信网、IP 网络技术的标准研究和系统开发工作。获得国家科学技术进步奖 2 项，工业与信息化部科学技术进步奖 3 项，中国通信标准化协会科学技术奖 4 项。1992 年获得国务院颁发的政府特殊津贴，1996 年获得“中华人民共和国有突出贡献的中青年科学技术专家”称号。

NewIP：开拓未来数据网络的新连接和新能力

郑秀丽，蒋胜，王闯
（华为技术有限公司，北京 100095）

摘　要：互联网应用深刻地影响着人们的工作与生活，纷至沓来的新应用对数据网络提出了新的挑战。基于未来应用对数据网络提出的需求，剖析了数据网络需要基于 IP 进行继承式发展，提出了一种新型的网络协议体系——NewIP，并介绍了 5 种关键使能技术，包括确定性 IP、内生安全、面向万网互联的新寻址与控制机制、用户可定义、新传输层等。

关键词：NewIP；数据网络；确定性；异构；万网互联；内生安全；高吞吐
中图分类号：TP3
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019208

NewIP: new connectivity and capabilities of upgrading future data network

ZHENG Xiuli, JIANG Sheng , WANG Chuang
Huawei Technologies Co., Ltd., Beijing 100095, China

Abstract: Internet applications deeply affect people's work and life. As the booming of internet applications, the data network faces more and more challenges. Based on the requirements of future applications, the trend for inheritance development based on traditional IP of data networks was analyzed and a new network protocol suite——NewIP was proposed. The core technologies including deterministic IP, intrinsic security, new control and addressing mechanisms for ManyNets, user definable and new transport were also introduced.

Key words: NewIP, data network, deterministic, heterogeneous, ManyNets, intrinsic security, ultra-high throughput

1　引言

互联网发展 40 多年来，随着 IP 技术的不断演进，其承载的应用越来越丰富，邮件、云计算、社交网络、在线购物、电子银行、视频直播等正在深刻影响着人们的学习、工作与生活，取得了巨大的成功。AR/VR、远程医疗、工业互联网、车联网等已悄然而至，全息通信、意识通信、空天地一体化通信等在不久的将来也将揭开神秘的面纱，人们正在快速进入一个万物感知、万物互联的智能世界。纷繁的新应用对 IP 网络提出了新的需求与挑战。网络技术创新是新应用创新的基

收稿日期：2019−08−10；修回日期：2019−09−10
基金项目：国家重点研发计划基金资助项目（No.2018YFB180079）
Foundation Item: The National Key Research and Development Program of China (No.2018YFB180079)

础与使能条件，IP 网络的能力亟待增强，以允许更多的新业务接入网络。

然而，与互联网业务快速更新相比，TCP/IP 协议作为互联网的基石，40 多年来一直没有发生实质性的变革。从 1996 年美国提出 NGI（Next Generation Internet Program）、Internet2 计划，到 2007 年欧盟发起第七框架计划（7th Framework Programme，FP7），虽然各国一直在努力探索数据通信的下一代技术，TCP/IP 协议自身也经历了一些优化改进（如引入 IPv6 解决地址耗尽问题、引入 IPSec 解决安全型问题等），但始终没有改变 TCP/IP 技术的内核，其本身的固有缺陷一直没有得到解决。为解决该问题，国内外学术界与工业界都在进行未来网络技术相关的研究，研究领域包括新型体系结构、路由机制、网络管理、故障诊断、网络感知与测量、移动性、安全与隐私等。虽然 IETF（全球互联网标准协议制定组织）针对部分问题给出了很多补丁式方案，但是缺少自顶向下的设计，且网络整体发展目标不明确。

基于以上现状，网络技术代际发展的思路——网络 5.0 被提出。网络 5.0 技术研究，聚焦未来 8~10 年典型应用对数据网络的需求，在演进思路上采取“分代目标、有限责任”的策略，在继承无连接统计复用的优势能力下，通过持续增强 IP 自身能力，以解决万物互联的智能世界所需要的内生安全可信、网络可规划和性能可预期、大连接下的感知与管控、泛在移动性支持等能力，连通多种异构接入网络，实现万网互联，使能更多的新服务接入网络，丰富人们的沟通与生活。

本文聚焦网络 5.0 典型应用场景对数据网络提出的关键需求，提出了一种新型的网络协议体系——NewIP，旨在通过顶层设计，提供“万网互联、万物互联”的新连接能力、确定性传输及大吞吐量传输的新服务能力、安全可信及用户可定义的新内生能力，使能更多的新服务接入网络。本文首先介绍了未来典型应用场景对数据网络的需求，剖析了数据网络需要基于 IP 进行继承式发展，对未来数据网络的“新腰”——NewIP 进行了顶层设计，并介绍了以下 5 种关键使能技术，以解决确定性时延、超巨吞吐量、内生安全、海量异构通信主体及异构网络的互联互通及用户可定义等关键问题。

（1）确定性 IP 技术

通过引入异步的周期调度机制来严格避免微突发的存在，从而保证确定性的端到端时延及抖动。

（2）内生安全技术

面向未来网络业务对安全可信的需求以及当前网络的安全可信脆弱性等问题，设计“端到端通信业务安全可信”和“网络基础设施的安全可信”方案，实现网络的内生安全能力。

（3）面向万网互联的新寻址与控制机制技术

设计变长网络地址、多样化寻址、面向服务的路由等机制，实现海量异构通信主体、异构网络的互联互通。

（4）用户可定义技术

通过报文携带指令及可修改的元数据，支持用户感知网络状态及用户定义网络行为。

（5）新传输层技术

基于已有的 TCP/IP 协议基础，通过获知上层业务传达的传输策略并感知网络性能等技术，对传输参数进行调整，增强信息本身的抗损和传输能力，以支撑未来新型媒体通信模式和潜在高吞吐业务的需求。

2　未来典型应用场景对数据网络的需求

移动承载、空间网络、全息通信及工业互联网、远程医疗与车联网等应用场景对数据网络提出的新需求主要包括确定性时延、超巨吞吐量、万网互联、内生安全及用户可定义等。未来典型应用场景及其对数据网络的需求如图 1 所示。

（1）万网互联

新垂直行业应用、数字化个人、行业数字实体等场景对应海量通信主体及多种异构接入

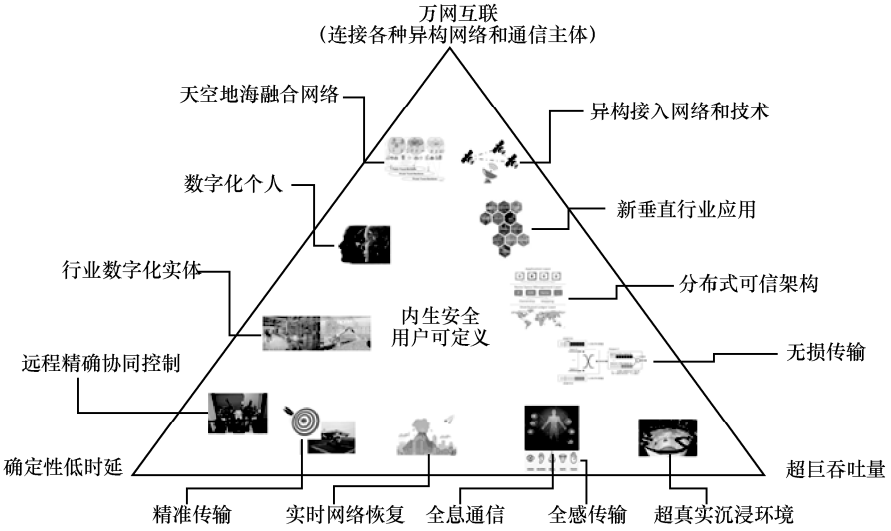


图 1　未来典型应用场景及其对数据网络的需求

网络，亟需天、空、地、海融合的万网互联技术拥抱海量通信主体、多种异构网络接入互联网。未来数据需要支持灵活多样化寻址，以支持使用不同标识、不同长度网络地址的通信主体间的互通，进而实现多种异构网络间的互联互通。

（2）确定性低时延

工业互联网、远程医疗、全息通信、车联网、电网继电保护等交互性高的场景，需要实现精准传输，网络不仅要提供“及时”服务，还要提供“准时”服务。

（3）内生安全

随着互联网深入渗透进人类的生产和生活，工业互联网、车联网、远程医疗等对于网络安全可信提出了更高的要求。要求与当前互联网的“补丁式”安全方案不同，未来网络需要一整套完整的、内生的安全可信机制，不仅要保证通信双方和网络基础设施的可信性，还要保证端到端通信的真实性、可审计性、隐私性、完整性、机密性以及面临网络故障和网络攻击下的可用性等。

（4）用户可定义

应用天然具有优先级差异和传输性能需求差异，亟待研究用户可定义技术支持终端/用户感知网络状态、表达需求及定义网络行为。

（5）超巨吞吐量

全息通信、自动驾驶、远程医疗、AR/VR 等应用需要网络提供超大带宽、超高吞吐量的传输。

3　数据网络需要基于 IP 进行继承式发展

3.1　传统 IP 瘦腰

TCP/IP 早在 20 世纪 70 年代被提出，在 IP 架构的沙漏型“瘦腰”结构（如图 2 所示）支持下，上层网络应用只要支持简单的 IP 协议，即可在互联网上运行，所以大量的互联网创新应用随之而来。这个“瘦腰”架构也是互联网发展 40 多年来的关键成功因素之一。

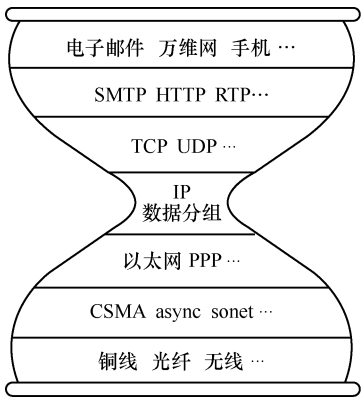


图 2　传统 IP 瘦腰结构

基于“瘦腰”结构，传统 IP 具有全球可达、

高生存性等特征。TCP/IP 设计之初，网络的主要需求为固定主机间端到端的可靠通信，所以主要以可达性为目标，仅能提供尽力而为转发，无法提供精细化的确定性质量保障，且缺乏顶层设计，难以支撑更多对确定性时延、超巨吞吐量、万网互联、内生安全及用户可定义等有更高要求的复杂应用，亟待设计未来数据网络的“新腰”以满足新应用场景的需求。

3.2　NewIP—— 未来数据网络新腰

本文提出了一种新型网络协议体系——NewIP，如图 3 所示。 NewIP 保留了传统 IP 网络统计复用和上下兼容的优势，将在确定性时延、超巨吞吐量、内生安全、海量异构通信主体、异构网络的互联互通及用户可定义等方面跨代提升 IP 网络的能力，以满足未来业务的新需求。NewIP 旨在成为未来数据网络的“新腰”，互联设备、内容、服务与人等海量异构通信主体，联通空、天、地、海等多种新型异构网络，使能新业务成为全息通信、远程医疗、工业互联网、车联网等业务生长的“黑土地”。

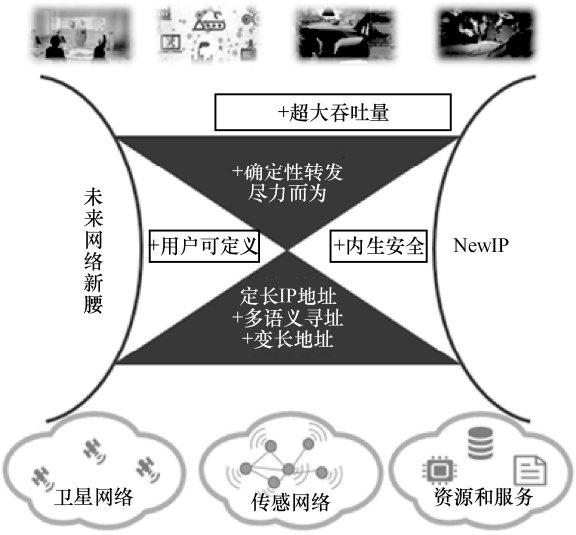


图 3　NewIP——未来数据网络“新腰”

4　NewIP 关键技术

NewIP 将在保留传统 IP 全球可达、高生存性、尽力而为转发等能力的基础上，进一步提供“万网互联、万物互联”的新连接能力、确定性传输及大吞吐量传输的新服务能力、安全可信及用户可定义的新内生能力。NewIP 关键技术包括确定性 IP、内生安全、面向万网互联的新寻址与控制机制、用户可定义、新传输层等，如图 4 所示。

4.1　面向万网互联的新寻址与控制机制

数据网络的服务边界从互联网业务逐渐扩张到工业制造、交通运输和农业生产等传统行业。未来，空、天、地、海等多种新型异构网络将不断融合与协同以支撑丰富的新应用。随着新应用越来越丰富，接入网络的通信实体的种类和数量也越来越多。海量通信主体不再局限于传统的主机，设备、内容、服务与人都可以作为通信主体。亟待研究统一的网络协议，通过实现变长网络地址、多样化寻址、面向服务的路由等，以支持海量异构通信主体、异构网络间的互联互通。

4.1.1　变长网络地址

数据网络经历了持续 40 余年的快速发展，作为其核心的 TCP/IP 协议均是定长、定界、定序的。

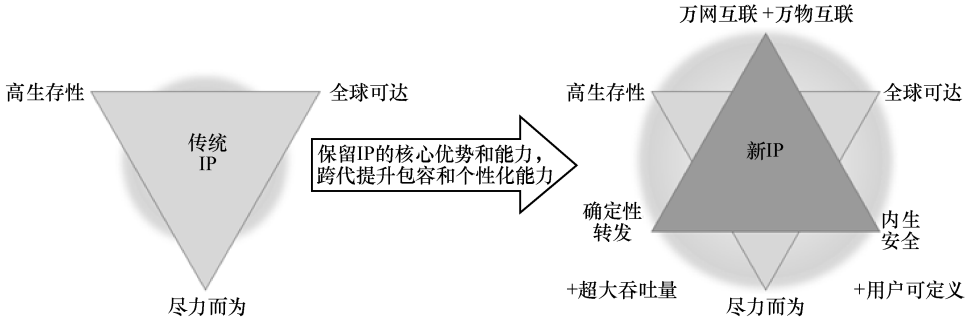


图 4　NewIP——开拓未来数据网络的新连接和新能力

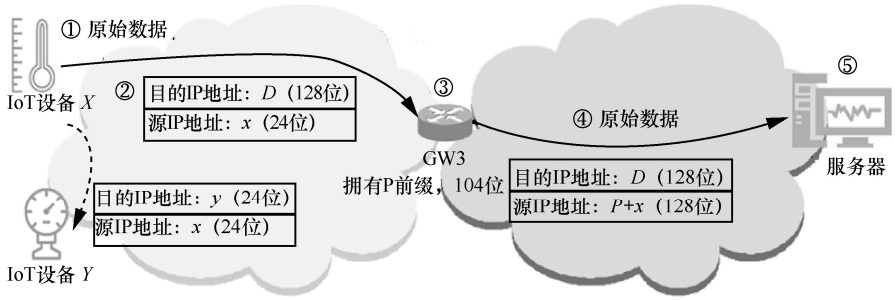


图 5　NewIP 变长地址通信示例

随着未来互联网业务的更加繁荣，各种异构网络、异构终端都需要连接互联网，并且具有迥异的通信需求。此时迫切需要打破网络协议定长、定界、定序的设计约束，提出一种新型的、支持地址长度可变的网络协议。

NewIP 网络层协议将采用变长的、结构化的地址设计。网络设备可以为不同长度的地址建立统一的路由转发表项。不同的网络地址将共存于数据报文中，网络设备则根据任意长度的地址进行路由表查找操作，从而决定数据报文的下一跳。据此，可根据网络规模平滑扩充地址空间，而无需修改旧有的网络地址配置。网络互联和扩容不依赖于协议转换或者地址映射网关设备，使组网方案更加灵活。因此，未来的数据网络可以同时满足海量通信主体引起的长地址需求及异构网络互联带来的短地址需求。

NewIP 局域网络内部的设备使用变长地址进行通信，并可与外部其他地址空间设备直接互通，如图 5 所示。NewIP 局域网络内部使用短地址通信，IoT 设备 *X* 与 *Y* 使用 24 位地址进行通信。IoT 设备 *X* 与外部服务器通信时，构造的数据分组目的 IP 地址是 128 位 IPv6 地址，源 IP 地址则为 24 位地址，经过 NewIP 网关时，网关会将数据分组的源 IP 地址补全到 128 位，数据分组经由 NewIP 骨干网络路由至服务器。

4.1.2　多样化寻址

NewIP 支持多样化的寻址方式，网络地址不仅标识主机，还可标识各种虚拟实体及异构节点，如人、内容、计算资源、存储资源等。路由器既可支持传统的拓扑寻址，又可支持主机 ID 寻址、内容名字寻址、OTT 私有名字寻址、计算名称和参数寻址等，如图 6 所示。

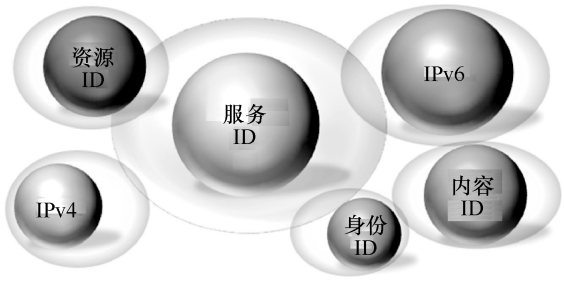


图 6　多语义标识

通过引入多样化的寻址实体，将主机、用户、内容、计算资源等与拓扑解耦，通过各自的地址空间进行路由（如图 7 所示，用户观看电影可以基于内容 ID 进行路由，一些信息可以基于身份 ID 进行推送等）。这种打破传统网络单一拓扑寻址的设计可以带来两点优势。第一，多样化的寻址方式可以消除对额外映射系统的依赖，进而消除映射系统所引入的时延、隐私以及单点故障等问题。第二，新网络中的多样化地址与拓扑解耦，能够有效支持各类物理、虚拟通信实体的泛在移动。

4.1.3　面向服务的路由

面向服务的路由（service-oriented routing）旨在改变传统 IP 基于拓扑的单一路由寻址机制，直接以服务标识或类型作为寻址依据以优化服务获取时延，并可根据各种通信实体差异化的需求，对服务标识、实体 ID 等实施路由策略。

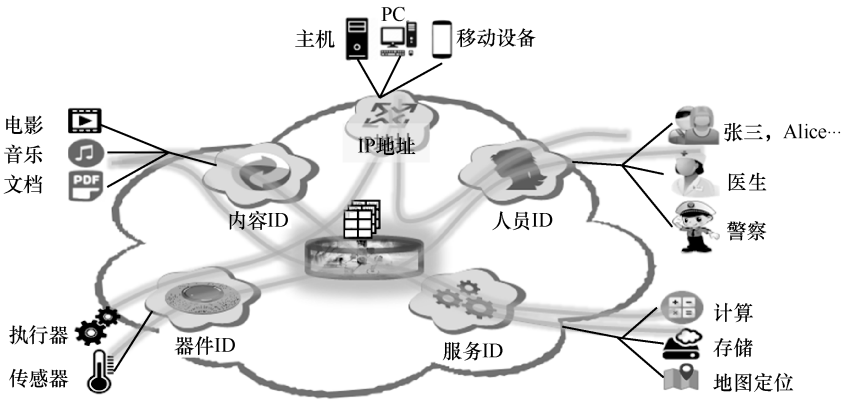


图 7　NewIP 多样化寻址转发

面向服务的路由机制需要服务侧、网络侧和用户侧的协作来完成，如图 8 所示。

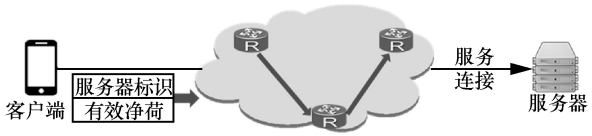


图 8　面向服务路由

用户端和服务端将根据网络设定达成服务标识生成规则的共识，如按照算法 F 为服务名/域名生成服务 ID 以及必要的冲突检测和解决机制。据此，用户端将无需通过映射系统就可以获知网络可路由的服务 ID，因此可以跳过当前比较耗时的 DNS 过程，缩短服务获取时延。

由通信实体端侧（用户端和服务端）生成并维护的标识系统，需要网络设备进行路由通告并在网络中形成该标识的转发表项。推而广之，不同的通信实体端侧系统可以生成并维护多种标识，网络设备将使用标准化的协议为多样化的标识提供路由转发能力支撑。

在上述过程中，网络可以对上述通信过程实施首个分组服务寻址、时延服务器绑定、用户体验寻址、泛在移动支持等优化措施。因此，面向服务的路由机制可以用于边缘计算、CDN 等场景，对于时延敏感的服务可以带来明显的收益，提升用户服务体验。并且方案可基于 IPv6 以最小的代价部署于现有网络体系。

4.1.4　天空地海网络使用统一网络协议进行多样化寻址

天地一体化网络，是实现多系统、多信息融合和协同的重要平台，整合空、天、地、海等多维资源信息，互通万物、互联万网，充分发挥空、天、地信息技术的各自优势，实现功能互补，扩大可处理事件的范围，实现时空复杂网络的一体化综合处理和最大有效利用，为各类用户提供可靠、按需的服务。

天地一体化网络中，空间网络与地面互联网最好使用相同的协议体系，以便于地面上的网元和卫星之间可以根据应用需求建立星间链路，进行数据交换，并简化网络管理。

NewIP 支持多样化寻址，其中包括传统 IP 地址寻址与地理地址寻址，通过此特性可实现空地网络一体组网，完成地面网元和卫星间的数据交换。NewIP——统一的新型网络协议体系如图 9 所示。

4.2　确定性 IP 技术

确定性 IP 的目标就是在现有 IP 转发机制的基础上提供确定性的时延及抖动保证。确定性 IP 的主要使能技术为大规模确定性网络（large-scale deterministic network，LDN）。通过引入周期调度机制来严格避免微突发的存在，从而保证了确定性时延和无拥塞分组丢失。LDN 技术的异步调度、支持长距链路、核心节点无逐流状态等特点使其适用于大规模网络可部署。

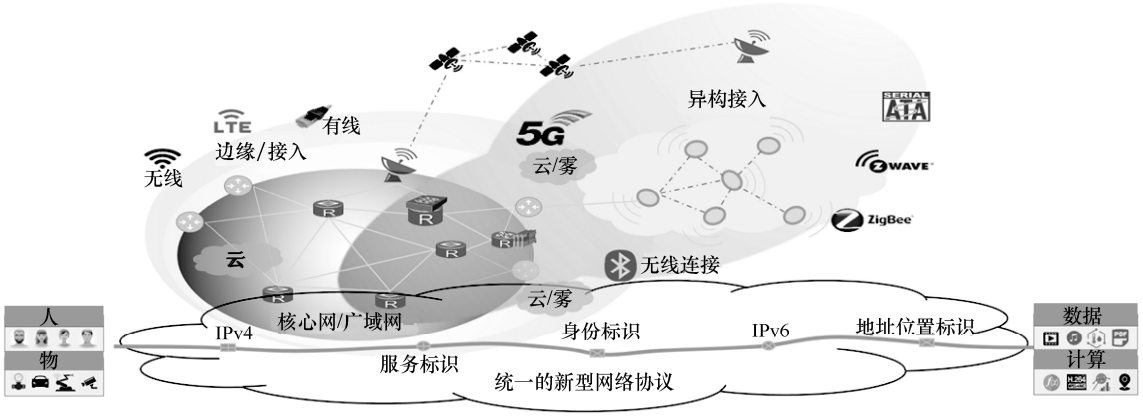


图 9　NewIP——统一的新型网络协议体系

LDN 首先要求全网设备频率同步，如图 10 所示，所谓的频率同步即各设备将自己的时间轴划分为等长的周期，不同设备的周期可以从不同的时间开始在不同的时间结束。并且任意两个设备的周期边缘之差 *D* 保持不变。

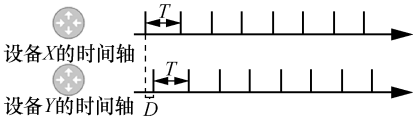


图 10　频率同步

任意两个邻居设备之间都维持着一个稳定的周期映射关系。如图 11 所示，*X* 设备的“周期 1”“周期 0”分别映射到 *Y* 设备的“周期 3”“周期 2”。该周期映射关系约束了两跳设备之间的数据分组转发行为，数据分组需要且只能在规定的周期内发送，从而保证了单跳数据传输的时延确定性。从源节点到目标节点经过逐跳的周期约束转发，保证了端到端的时延确定性。基于确定性的时延上界，选择一个满足业务需求的确定性服务管道。

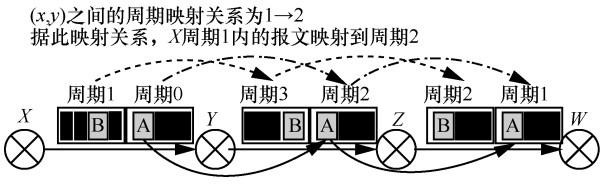


图 11　基于映射周期关系的数据分组转发

设备之间的周期映射关系可以通过控制面配置的方式，也可以通过自适应分布式学习的方式得到。构造出的周期映射关系可以分布存储在转发设备上，也可以集中存储在少量控制设备上。后续用户数据报文只需要携带周期相关信息，通过查表转发或者其他方式即可实现确定性转发。

4.3　内生安全技术

考虑到未来网络业务对安全可信的需求以及当前网络的安全可信脆弱性，希望借鉴当前的经验和教训，自顶向下地设计一套完整的、内生的网络安全架构。把网络需要解决的安全可信问题归纳为“端到端通信业务安全可信”和“网络基础设施的安全可信”两大类，并分别提出相应的使能技术。

（1）端到端通信业务安全可信

端到端网络通信在 IP 地址真实性、隐私保护与可审计性的平衡、密钥安全交换、拒绝服务攻击等方面存在较大的安全威胁。面对以上安全威胁，未来网络可根据安全目标及需求划分不同的安全域，将不可信、攻击流量阻断在安全域外，将域内安全问题控制在安全域内，限制安全问题的扩散。在划分安全域的基础上，通过在不同安全域中的网络元素及协议中内嵌关键安全技术，提供可信身份管理、真实身份验证、审计追踪溯源、访问控制、密钥管理等安全模块，实现端到端通信的身份/IP 地址真实

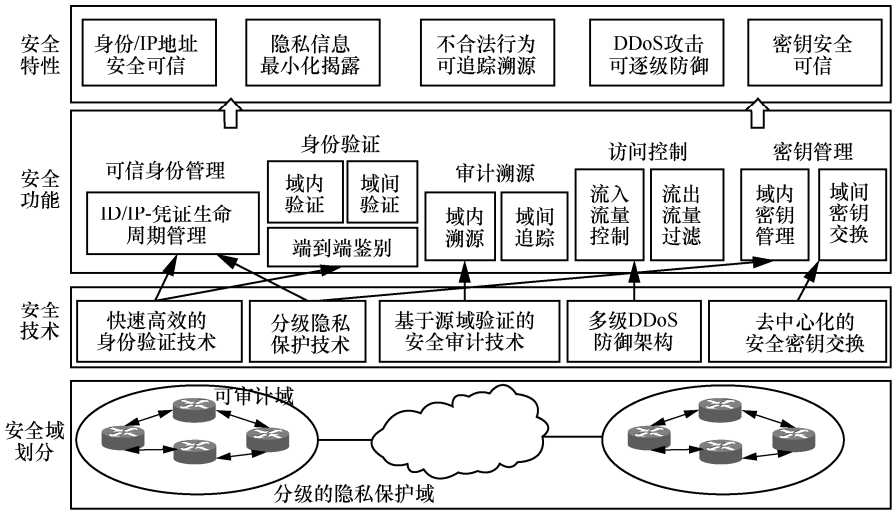


图 12　端到端通信业务安全技术框架

可靠性、个人隐私信息最小化、不合法行为可追踪溯源、DDoS 攻击可追级防御、密钥安全可信等特性，如图 12 所示。

（2）网络基础设施安全可信

除了端到端通信业务安全外，支撑全球互联网运行的基础设施的安全性和可信性也需要增强。目前，互联网最重要的两大基础设施是路由系统和域名系统。这两大基础设施和其背后的安全可信模型都是中心化的，以某个可信第三方作为整个系统的单一信任锚点。由于中心化的模型存在着中心节点权限过大、单点失效等脆弱性，这些基础设施存在安全可信隐患，同时也大大降低了互联网的平等性和可靠性。为了构建一个更加平等、可靠和开放的互联网，未来互联网的基础设施需要以某种去中心化的方式来作为安全可信基础，以摆脱中心化模型导致的安全隐患和信任危机。

在未来网络中，可以采用以分布式账本技术为代表的去中心化技术来构建基础设施的可信根。分布式账本等去中心化技术不存在单一可信锚点，所有节点平等，并且有全部信息副本，因此更加可信和安全。在此基础之上可以构建统一的资源管理平台，实现网络核心资源（如 IP 地址、域名、AS 号及其他未来可能的资源类型）的申请和管理，并提供不依赖于第三方的资源所有权证明。进一步地，资源所有者可以发布其所拥有资源相关的映射信息，如 AS/IP 地址映射、IP 地址/域名映射等，基于资源间的映射信息，可以进一步实现 BGP 宣告和 DNS 查询的基本能力。由于资源所有权不依赖于单一信任实体，基于此上的所有信息均可信可验证，基于去中心化技术的网络基础设施安全可信架构如图 13 所示。

4.4　用户可定义技术

现代芯片技术的发展使得终端的能力越来越强大，同时网络的能力也在随之增长。这种技术的长足发展促使新的网络应用爆发式增长。然而作为终端与网络的唯一接口的 IP 协议，常年来几乎没有发生太大变化。这导致了用户侧的需求无法完整、及时地传递给网络侧，同时终端也缺乏感知必要的网络状态的能力。NewIP 希望通过对 IP 协议进行优化，解决上述问题。

用户可定义技术，将控制指令、信息封装在数据报文中，由控制面进行网络功能与协议的部署配置，数据面进行报文级的用户可编程的功能支撑，一方面支持用户感知网络状态，包括报文传输路径、是否发生拥塞、中间设备处理信息等，另一方面可支持用户定义网络的行为，包括低时延转发、大带宽转发、订阅分组丢失通告、细粒度的随路测量等，进而达成在网络层的多种新特

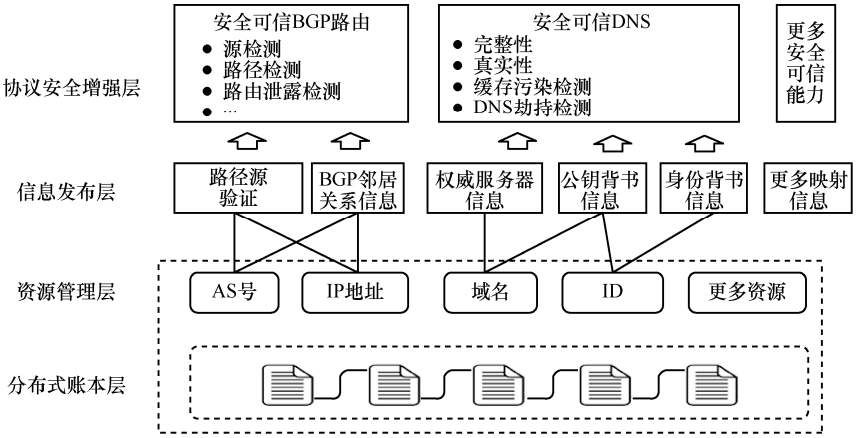


图 13　基于去中心化技术的网络基础设施安全可信架构

性，满足未来场景需求。

用户终端发送数据分组，数据分组携带目标、资源、服务等方面的需求或操作，比如网络需要满足哪些特定的 SLA（service level agreement）需求等，如图 14 所示，服务 A 的 SLA 需求体现在无分组丢失、低时延方面，服务 B 的 SLA 需求体现在高带宽、低时延方面。网络根据数据分组携带的 SLA 信息进行相应操作，包括随路资源预留等，图 14 中，网络基于服务 A 的 SLA 需求为其预留资源以实现无分组丢失、低时延转发，基于服务 B 的 SLA 需求为其预留资源以实现高带宽、低时延转发。新 UNI 支持动态性，用户基于感知到的网络状态，可以动态调整网络行为。

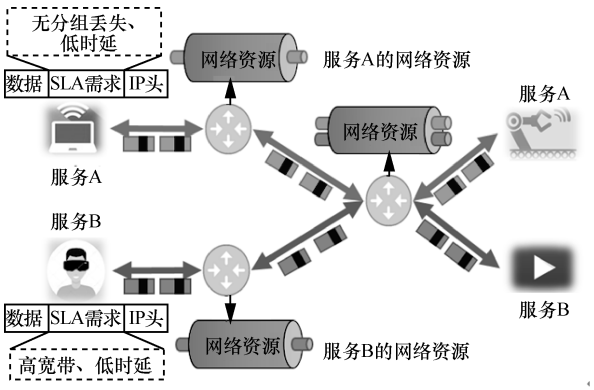


图 14　新 UNI 使用示例

4.5　新传输层技术

新传输层技术（X-transport）的目标是支撑未来新型媒体通信模式和潜在高吞吐业务（如全息通信）的需求。基于已有的协议基础与技术能力，新传输技术的演进方向主要集中在 3 个层面：

- 结合上层业务特征进行对传输策略的表达；
- 感知下层网络性能进行对传输参数的调整；
- 内生地结合其他技术（如编码技术）增强信息本身的抗损和传输能力。

新传输层提供向上与向下的能力接口 ALA+（application layer assistant）和 NLA+（network layer assistant），如图 15 所示。上层应用结合业务特征，如损失可容忍、服务等级约束等，向传输层表达传输策略的需求。同时，终端侧程序通过带内或带外的信令实时获取网络状态与各个链路的关键性能信息，实时调整传输策略的具体参数。面向 NewIP 所关注与支持的重点场景，新传输层的关键技术包括以下内容。

（1）超大吞吐量

结合网络编码技术，适应复杂多变的网络环境，动态反馈式调整编码冗余效率，避免重传，提供超大吞吐，大幅减少流完成时间，提高传输效率。

（2）并发多路高带宽

网络感知接口（network aware interface）通过带内带外信令及网络设备的配合，规划并发多路径，避免单一瓶颈链路，精准掌握网络多路径状态参数，为多路传输的策略优化、调度算法、编

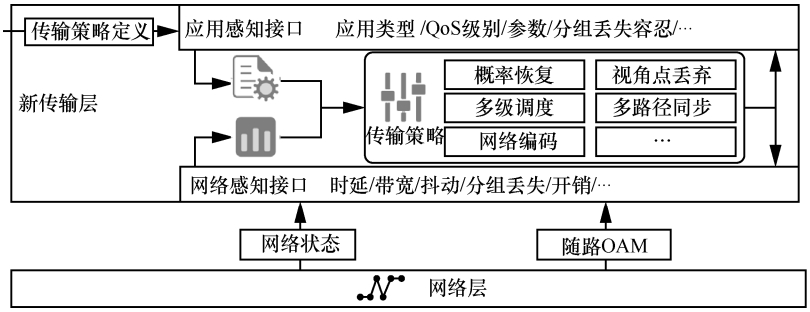


图 15　新传输层技术框架

码策略等提供决策支撑。

（3）传输可定义

应用感知接口（application aware interface）通过上层应用对传输内容特征的描述，定义传输策略，表达对数据优先级、服务质量、损失容忍的能力，选择匹配传输策略。

5　结束语

本文基于新应用场景对数据网络提出的新需求，提出了一种统一的新型网络协议体系——NewIP。NewIP，作为未来数据网络的“新腰”，继承了传统 IP 的成功基因，保留了传统 IP 网络统计复用和上下兼容的优势，跨代提升了传统 IP 网络的能力。通过引入异步的周期调度机制来严格避免微突发的存在，保证确定性低时延传输能力。基于“端到端通信业务安全可信”和“网络基础设施的安全可信”方案，实现网络的内生安全能力。通过变长网络地址、多样化寻址、面向服务的路由等机制，提供“万网互联、万物互联”的新连接能力。通过用户可定义技术，提升用户感知网络状态及用户定义网络行为的能力。通过新传输层技术，提升网络的高吞吐传输能力。作为一种协议体系创新研究，NewIP 尚有大量的技术细节还有待业界的共同研究和完善。

参考文献：

[1] 刘韵洁. 未来网络的发展及前景[C]//2016 全球 SDN/NFV 技术大会, 6 月 1-2 日, 2016, 北京, 中国. [出版地不祥:出版社不详], 2016.
LIU Y J. Future development and prospect of network[C]//2016 Global SDN/NFV Technology Conference, June 1−2, 2016, Beijing, China. [S.l.:s.n.], 2016.

[2] 蒋林涛. 数据网的现状及发展方向[J]. 电信科学, 2019, 35(8): 2-14.
JIANG L T. Current situation and development trend of data network[J]. Telecommunications Science, 2019, 35(8): 2-14.

[3] 郑秀丽, 蒋胜, 王闯, 等. 对网络技术跨代发展的思考——网络 5.0[J]. 信息通信技术, 2017(6): 37-44.
ZHENG X L, JIANG S, WANG C, et al. A potential direction of next generation data communication network – network 5.0[J]. Information and Communications Technologies, 2017(6): 37-44.

[4] 郑秀丽, 谭佳瑶, 蒋胜, 等. 未来数据网络需求分析[J]. 电信科学, 2019 (8): 16-25.
ZHENG X L, TAN J Y, JIANG S, et al. Analysis on the requirements of future data network[J]. Telecommunications Science, 2019, 35(8): 16-25.

[5] XU S, PEREZ M, YANG K, et al, Determination of the latency effects on surgical performance and the acceptable latency levels in telesurgery using the dv-trainer simulator[J]. Surgical Endoscopy, 2014, 28(9): 2569-2576.

[作者简介]

郑秀丽（1984− ），女，华为技术有限公司高级工程师，主要研究方向为网络架构及协议等。

蒋胜（1978− ），男，博士，华为技术有限公司主任工程师，主要研究方向为网络架构设计、IPv6 等。

王闯（1975− ），男，华为技术有限公司高级技术专家，主要研究方向为未来网络探索、网络架构设计等。

大规模确定性网络转发技术

强鹏，刘冰洋，于德雷，王闯
（华为技术有限公司，北京 100095）

摘　要：提出了一种适用于大规模网络部署的3层转发技术——LDN（large-scale deterministic network），在保留传统IP转发技术统计复用的优势基础之上，LDN技术可实现对端到端时延上界及抖动上界的严格保证，为5G uRLLC（ultra-reliable low-latency communication）切片、工业互联网等未来应用场景提供网络服务支持。通过仿真实验对比了在相同网络环境下，传统IP及确定性IP在端到端最差时延及抖动上的差异，证明了LDN技术的有效性。

关键词：确定性网络；有界时延；抖动
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019213

Large-scale deterministic network forwarding technology

QIANG Li, LIU Bingyang, YU Delei, WANG Chuang
Huawei Technologies Co., Ltd., Beijing 100095, China

Abstract: A layer 3 forwarding technology LDN (large-scale deterministic network) which was scalable for wide area deployment was proposed. LDN not only keeps the advantage of traditional IP's statistic multiplexing, but also be able to guarantee end-to-end bounded latency and bounded jitter required by 5G uRLLC (ultra-reliable low-latency communication) slice, industrial network, and other future scenarios. A simulation experiment was taken between traditional IP and deterministic IP (LDN enabled) on end-to-end worst case latency and delay variance (i.e., jitter), the comparison results successfully proved the effective and efficient of LDN technology.

Key words: deterministic network, bounded latency, jitter

1　引言

面对诸如工业互联网、远程医疗等时延敏感型业务[1]，传统网络尽力而为的转发方式已无法满足此类业务对网络时延的需求，确定性网络技术受到越来越多的关注。本文提出了一种支持确定性IP服务的技术——大规模确定性网络（large-scale deterministic network，LDN）技术。LDN技术旨在扩展传统3层转发技术，以提供对传输时延上界及时延抖动上界有保障的确定性转

收稿日期：2019−07−15；修回日期：2019−09−12
基金项目：国家重点研发计划基金资助项目（No. 2018YFB180079）
Foundation Item: The National Key Research and Development Program of China (No. 2018YFB180079)

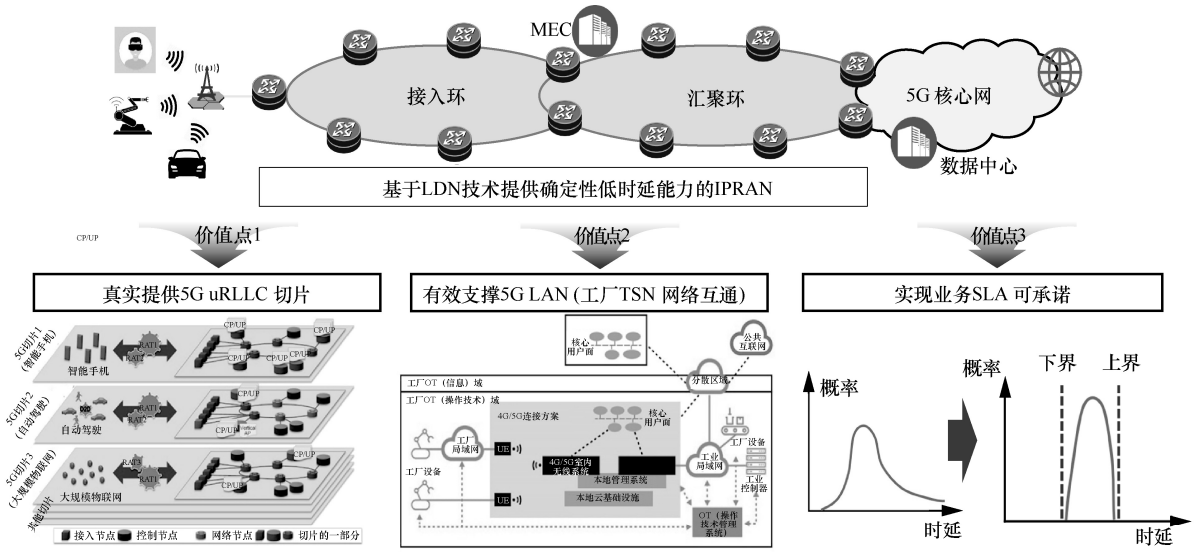


图 1　LDN 用于 5G 承载场景

发技术。本文将有界的时延及时延抖动统称为确定性时延。

随着 3GPP Release 15[2]的封版，5G 技术面临着全面部署。其中网络切片，尤其是 uRLLC（ultra-reliable low-latency communication, 超可靠低时延通信）切片[3]，对于时延的保证有严格的要求。uRLLC 切片要求不论在什么情况下 100%地保证服务水平协议（service level agreement, SLA），即对于最差时延的上界是有保证的。LDN 技术可以有效支撑 5G uRLLC 业务，如图 1 所示。此外 LDN 技术也可支持 5G LAN，即利用 5G 网络连通工厂时延敏感网络（time-sensitive network, TSN）孤岛，实现端到端精准控制。除 5G 承载之外，LDN 还可替代企业专线支持传统网络中的时延敏感型业务。

2　确定性 IP 技术背景

传统 IP 网络的时延概率分布曲线如图 2 所示，在传统 IP 网络中由于存在时延长尾效应[4]而无法实现确定性时延。为了消除长尾效应实现确定性时延，首先需要分析时延的组成，了解长尾效应的产生原因。

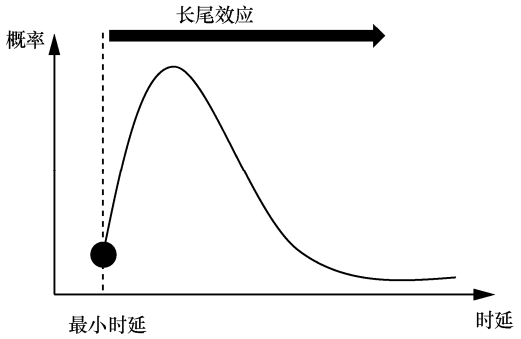


图 2　传统 IP 网络的时延概率分布曲线

2.1　时延分析

以图 3 所示的单跳场景来分析时延的组成，单跳时延=节点内时延+链路时延。其中，链路时延是指数据分组在链路上传输的时延，主要取决于网络设备之间的链路距离及链路传输速率。在一个稳定的网络拓扑中，链路距离及链路速率相对稳定，因此链路时延几乎没有什么变化的空间。节点内时延是指设备内部操作的耗时，主要为排队时延。在不同负载等情况下，节点内部时延的变动很大，图 2 中的长尾效应主要就是由节点内部时延引起的。

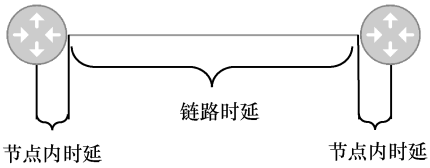


图 3　单跳时延分析

基于以上分析可以知道一直以来通信行业追求的低时延与确定性时延的技术目标其实是一致的。

- 为了实现确定性时延，需要减少节点内时延以消除长尾效应。
- 为了实现低时延，由于链路时延在固定的网络拓扑下没有多少变动空间，因此能做的也是设法减少节点内时延。

确定性 IP 的技术目标如图 4 所示，即压低最差时延上界使其无限逼近最小时延。

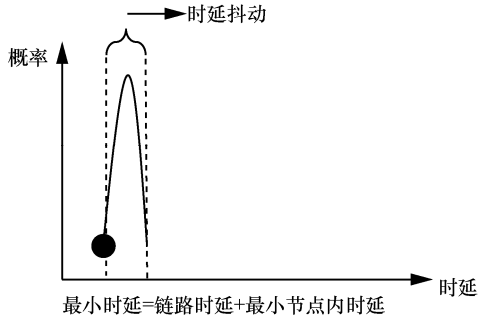


图 4　DIP 网络的时延概率分布曲线

2.2　现有技术分析

近几十年来有很多的技术探索去保障 IP 的服务质量（quality of service，QoS），比如优先级队列调度、资源预留等。可是这些技术依然无法实现确定性时延，其根本原因是微突发。以一个简单的例子来进行解释，如图 5 所示，假设节点 A 有 10 个入接口，1 个出接口，所有接口的速率都是一样的。假设每个入接口上都有一条流，并为此流预留了 5%的接口带宽资源。假设 10 个接口的 10 条数据流都是同一优先级，10 条流在节点 A 的出接口汇聚时总共占用出接口最多 50%的带宽资源，因此不存在拥塞的情况。然而由于对每个数据分组的到达时间没有控制，因此最差的情况是 10 条流的数据同时到达，在出接口排队等待调度。排队的产生会挤压原本流内报文的间隙，形成微突发。假设下游节点 B 有 10 个像 A 一样的上游节点，则微突发的情况会进一步累积，形成微突发迭代。多跳之后，确定性时延将无从保证。

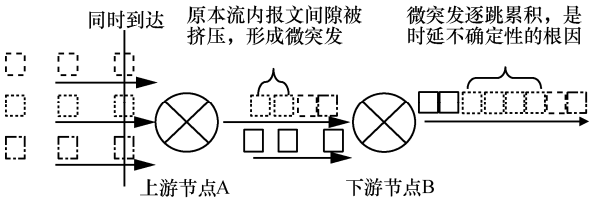


图 5　微突发及突发迭代

图 5 中的例子说明，传统 IP 转发即使配合了资源预留及优先级调度，也无法实现确定性时延。其主要原因是由于对每个数据分组的行为缺乏控制，导致了排队形成了微突发。为此需要控制每个数据分组的行为，不能容许数据分组随意发送（接收），而且为每个数据分组分配一个特定的发送（接收）时间周期。通过控制每个数据分组的行为来避免冲突，从而控制节点内部排队时延，最终消除长尾效应实现确定性时延。

3　确定性 IP 的使能技术 LDN

3.1　基于 LDN 的承载解决方案

基于 LDN 的承载解决方案如图 6 所示，其中包括终端 CE、入口 PE 节点、支持 LDN 能力的 P 节点、控制器等，整体流程如下：

（1）所有网络设备（PE 节点及 P 节点）需要保持微秒级周期相对固定；

（2）对每个准许接入的确定性业务需要进行路径计算、资源预留等操作；

（3）（1）（2）完成之后，可以开始传送确定性业务的报文。用户报文的流量模型需要满足资源预留的约束；

（4）入口网关需要对用户报文进行流量整形，

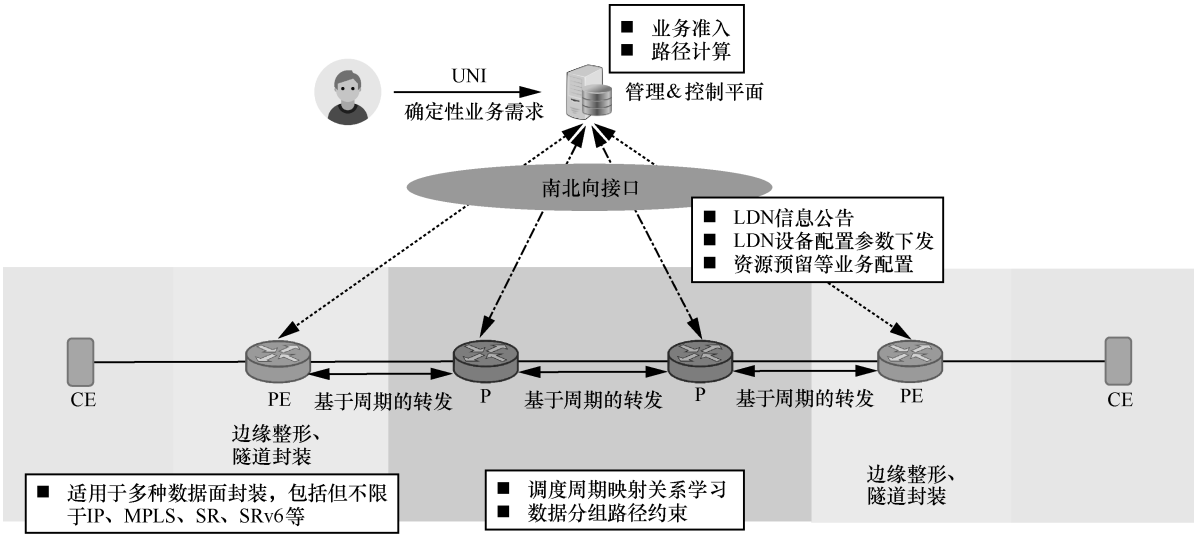


图 6　基于 LDN 的承载解决方案

并为报文打上初始周期标签，正式进入 LDN 周期转发流程；

（5）数据分组携带周期标签发往下游设备，当数据分组抵达 LDN 路由器后，路由器根据本地维护的周期映射关系表，替换数据分组中的周期标签并将数据分组送入相应的队列等待转发；

（6）每个 P 节点及 PE 节点都维护着特定数量的 LDN 队列，并对这些队列进行周期门控调度。

下面将对整体架构中涉及的技术细节进行描述。

3.2　等长周期划分

为了避免盲目转发而造成的微突发，LDN 技术需要控制每个报文在每一跳的转发行为。如果将网络设备视为一个黑盒子，那么报文经过一跳网络设备的行为即可描述为：报文进入该跳设备的时间及报文出该跳设备的时间。控制报文在每一跳的转发行为则相应地可以描述为：一旦给定报文进入某跳设备的时间，则该报文出该跳设备的时间即可确定。将报文进、出网络设备的时间控制到一个精确的时刻是很困难的，但是控制到一段时间区间要相对容易。基于这个思想，LDN 设备将各自的时间以 T 为单位划分为等长周期，并为每个数据报文合理安排进、出本跳的周期。

3.3　资源预留

每条确定性流在正式发送数据报文之前，都需要为其预留沿途的所有资源。资源预留的过程可以利用现有技术完成，例如通过 SDN 控制器进行集中算路及配置的方式，或者 RSVP-TE 的方式等。特殊之处在于，每条确定性流都有一个最小预留资源量。

最小预留资源量的目的是确保每条流在每个周期内都能发送至少一个数据报文。假设数据报文的大小为 U（例如 U=1 250 byte），时间周期的长度记为 T（例如 T=10 μs），则最小资源预留量为 U/T（例如 U/T=1 250 byte/10 μs=1 Gbit/s）。所有确定性流的资源预留量都不得小于最小资源预留量。例如有一条流只需要 0.5 Gbit/s 的带宽资源，由于在此例中最小预留资源量为 1 Gbit/s，则依然需要为这条流预留 1 Gbit/s 的带宽。

为某条确定性流所预留的资源，哪怕在某段时间没有被使用，也不得为其他的确定性流所用，然而这部分空闲资源却可以用于转发尽力而为（best effort，BE）报文。

3.4　周期映射关系学习

在 LDN 系统中，每对邻居 LDN 路由器之间都有一个稳定的周期映射关系。该周期映射关系

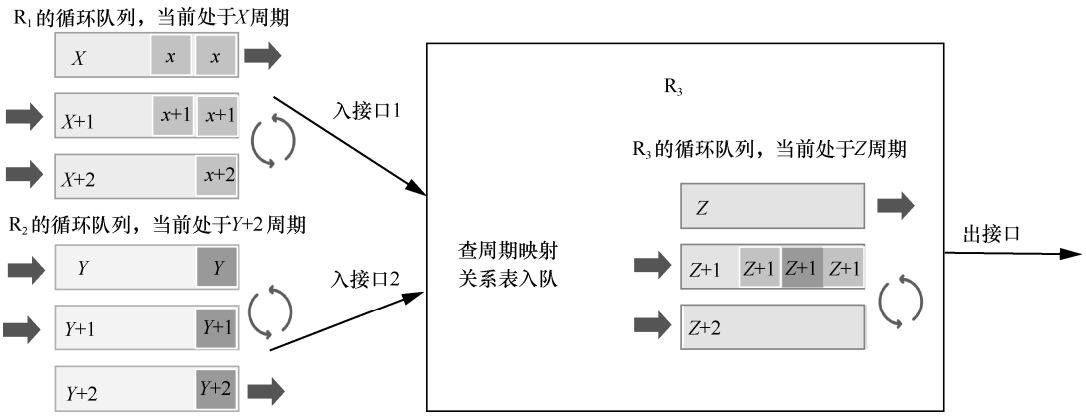


图 7　循环队列调度

指导着后续 LDN 路由器的数据分组转发行为。周期映射关系的构建可以通过SDN控制器进行配置，也可以通过自适应的方式学习得到。通过学习报文获得周期映射关系之后，后续的数据分组均可以简单地据此映射关系转发。

数据报文在每一跳发送之前只需要携带发送设备的当前周期编号。下游节点收到数据报文后根据本地维护的周期映射关系表，即可确定收到的报文需要在本地哪个周期再次转发出去。

3.5　循环队列调度

如第 3.2 节所述，每个 LDN 路由器将自己的时间划分为等长的周期。LDN 路由器对周期进行循环编号，每个周期对应着一个队列。在一个周期内只有其对应的队列会打开，存储于该队列中的数据报文得以发送，其余队列处于关闭的状态，只能用于接收数据分组。

图 7 给出了一个循环队列调度的例子，R_1 与 R_2 是 R_3 的两个上游节点。所有节点都有相同数量 Q 个循环队列，在本例中假设 Q 等于 3。当下游节点 R_3 收到数据报文之后，读取报文上携带的周期编号信息，并据此查找周期映射关系表。随后 R_3 将报文中的周期信息替换为查表所得到的出标签信息（比如入周期标签 X 对应的出周期标签 Z+1），并将报文送入本跳相应的队列中等待发送。

4　理论分析

4.1　时延

首先分析一跳之内数据报文的时延，图 8 和图 9 分别给出了 LDN 的最差单跳时延及最小单跳时延分析。若上游节点最后 1 bit 到达下游节点的时间刚好超出了前一个周期一点，则出现最差情况。最差情况下数据报文需要在下游节点中等待约 $2T$，才能继续向下转发，如图 8 所示。若上游节点最后 1 bit 到达下游节点的时间刚好落在某一周期的后边沿，则在紧随其后的一个周期内即可向下游转发，此为最好情况，如图 9 所示。在最好情况下数据报文需要在下游节点中等待 T 的时间。

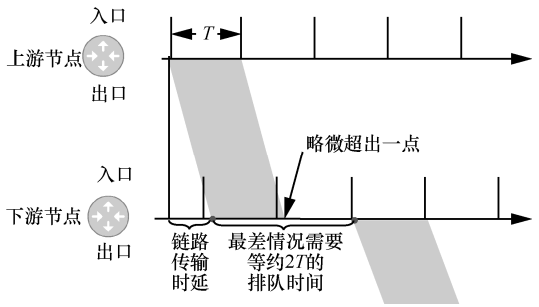


图 8　LDN 最差单跳时延

由上述分析可知 LDN 系统中，单跳的排队时间为 T~$2T$，因此：

- LDN 系统中端到端的最差时延=总链路传输时延+$2T$×跳数；

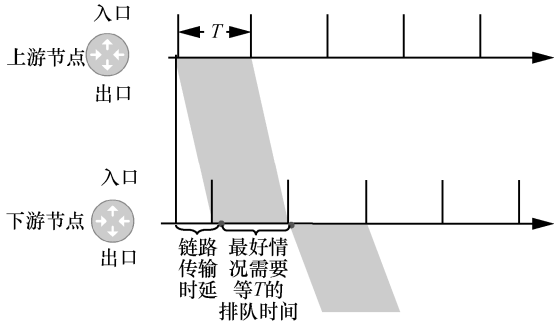


图 9　LDN 最好单跳时延

- LDN 系统中端到端的最好时延=总链路传输时延+T×跳数。

4.2　抖动

由于每对邻居节点之间都有一个周期映射关系，因此一旦数据报文在源点的发送周期确定了，则该报文在目标点的接收周期也就可以确定下来。

唯一无法确定的是，数据报文在发送周期的具体哪个时刻发送，并在接收周期的具体哪个时刻接收。图 10 给出了最好及最差情况：

- 最好情况是数据报文在发送周期的最末尾才发出，在接收周期的一开始就收到了；
- 最差情况是数据报文在发送周期的最开始就发出，在接收周期的最末尾才收到。

则端到端时延抖动≤最差情况−最好情况=2T。

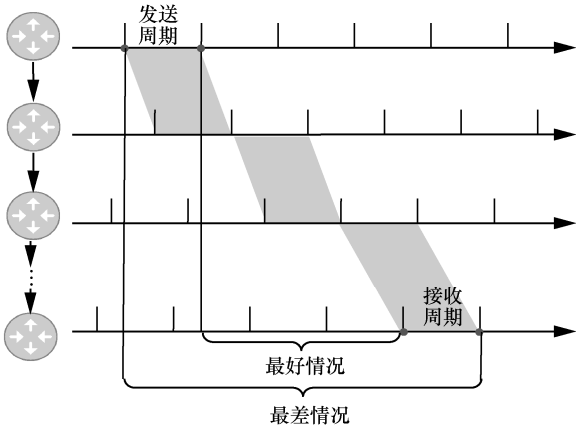


图 10　LDN 端到端抖动

4.3　队列开销

由第 3.5 节可知每个 LDN 设备维护着若干 LDN 循环队列，每个 LDN 队列对应一个周期，用于存储一个周期内的数据报文。假设 LDN 出接口的接口速率为 S，则每个 LDN 队列的存储开销为 S×T。因此每个 LDN 节点只需要维护 Q×S×T 大小的队列空间，其中 Q 为一台设备内 LDN 循环队列的数目。

5　仿真实验

为了验证 LDN 技术的性能，本文对比了传统 IP 与 DIP（LDN enabled）在相同网络环境下的端到端最差时延及抖动。实验采用如图 11 所示的拓扑[5]，共涉及 315 个网络节点。假设每个网络节点下挂一台主机，且任意两台主机之间都可以建立一条业务流，则最多有 315×314=98 910 条流同时存在。考虑到实际情况中网络大多轻载，因此实验假设网内仅有 5 000 条流共存，其中 1 000 条目标待测流、4 000 条背景流。实验参数见表 1。实验开始阶段首先将图 11 所示拓扑导入 OMNest 中，随后利用 INET Framework 安装协议栈完成路由发现，实验对比结果如图 12 与图 13 所示。

假设目标待测流在 DIP 模式下以每周期 1 个报文的节奏进行转发，DIP 与传统 IP 在不同突发条件下的端到端最大抖动对比如图 12 所示。突发数据报文均一次性发送，例如“100 个数据分组/ms”即意味着每 1 ms 一次性发送 100 个数据报文，“200 个数据分组/2 ms”即每 2 ms 一次性发送 200 个数据报文，以此类推，因此不同突发条件下的背景流速其实是不变的。由图 12 可知，在 DIP 模式下端到端抖动始终小于 20 μs（即 2T），且不随突发的加剧而改变，而 IP 模式下的最差抖动随着突发的加剧而显著增加。在每 10 ms 有 1 000 个报文的突发情况下，IP 模式的最差抖动是 DIP 模式的最差抖动的近 300 倍。由此可见 DIP 可以有效降低数据报文 3 层转发的时延抖动。

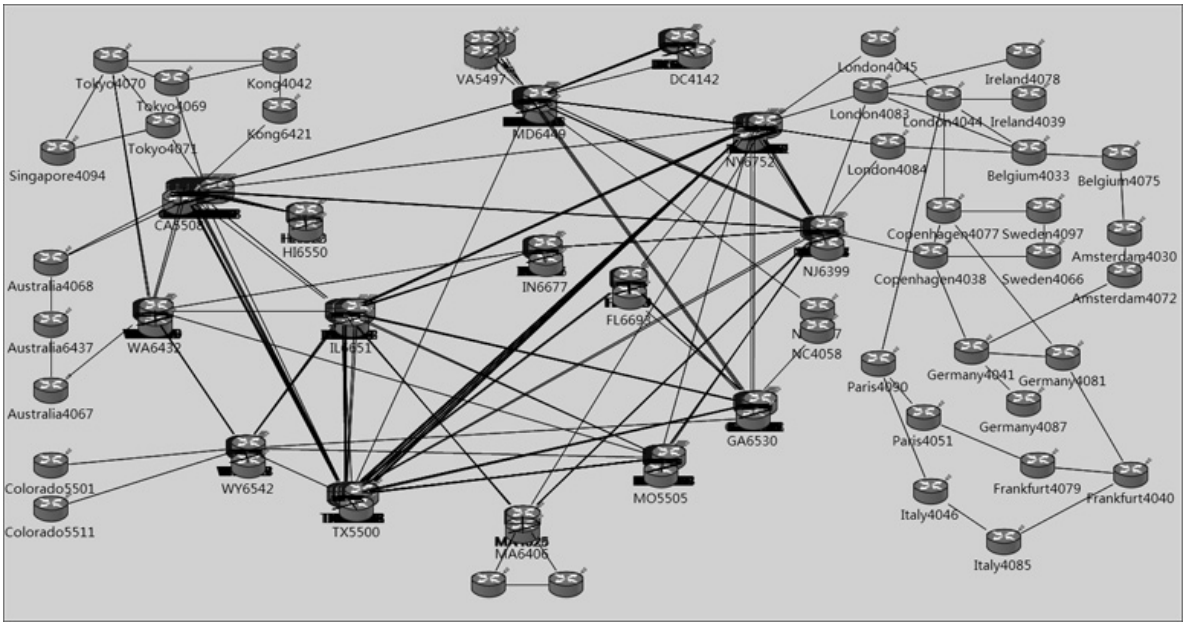


图 11　实验拓扑

表 1　实验参数

参数类型	参数值
周期长度 T	10 μs
数据报文大小 P	64 byte≤P≤1 500 byte
每周期发送报文数 N	1
链路速率	100 Gbit/s
实验环境	OMNest + INET Framework

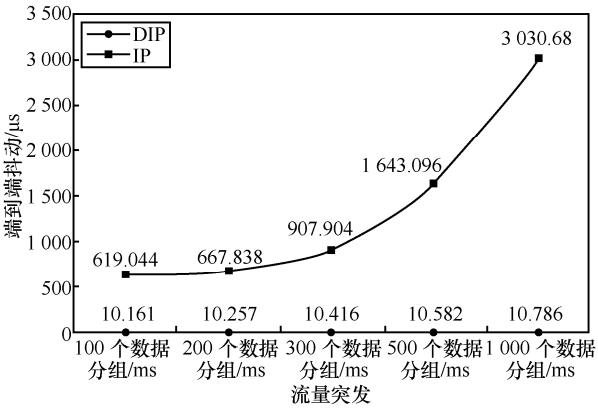


图 12　DIP 与 IP 在不同突发条件下的端到端抖动对比

DIP 与传统 IP 在不同突发条件下的端到端最差时延对比如图 13 所示。由图 13 可知，在 DIP 模式下的端到端最差时延几乎不受突发的影响，始终维持在一个相对稳定的数值。而传统 IP 模式下的端到端最差时延随着突发的加剧而显著增加。在每 10 ms 有 1 000 个报文的突发情况下，IP 模式的最差时延是 DIP 模式的最差时延的近 30 倍。由此可见 DIP 可以有效地减少时延长尾效应，提供 SLA 保证。

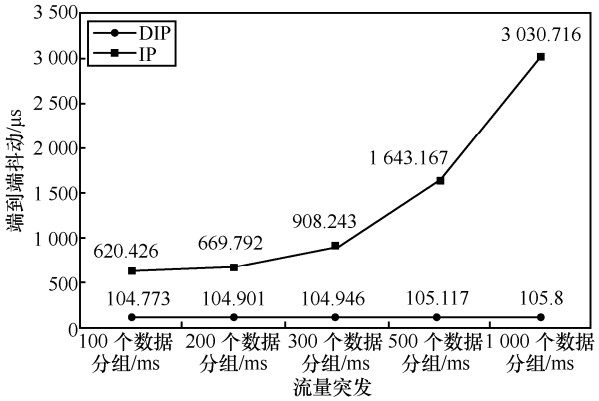


图 13　DIP 与 IP 在不同突发条件下的端到端最差时延对比

6　结束语

LDN 技术是确定性 IP 的主要使能技术。LDN 在基于传统 IP 的基础之上引入周期转发的思想，通过控制每个数据分组在每跳的转发时

机来减少微突发，消除长尾效应，最终实现端到端时延确定性。LDN 技术可以保证在最差情况下的端到端时延依然有界，且最差时延与最好时延之间的差距仅为 2 *T*。以 μs 级周期转发为例，则最差时延无限逼近于最好时延，两者之差仅为微秒级。

此外 LDN 技术中核心节点无逐流状态，设备之间不需要精准时间同步，因此具有良好的大网可扩展性。同时，LDN 技术的芯片开销低，以 4 个循环队列为例，对于 10GE 接口，10 μs 周期调度下每出接口仅需要 50 KB 的 LDN 队列资源。最后报文的开销较小，数据报文中仅需要携带 2 bit 额外信息，用于区分不同的循环队列即可。

综上，LDN 技术不仅可以实现对端到端时延及抖动的有界保障，且适用于大规模部署，是应对 5G 承载、工业互联网等未来场景时极具竞争力的一项技术。

参考文献：

[1] IETF. Deterministic networking use cases: RFC 8578[S]. 2019

[2] 3GPP. Summary of rel-15 work items: TR 21.915[S]. 2019.

[3] POPOVSKI P, TRILLINGSGAARD K F, SIMEONE O, et al. 5G wireless network slicing for eMBB, uRLLC, and mMTC: a communication-theoretic view[J]. IEEE Access, 2018(6): 155765-55779.

[4] TUCKER C, ZHANG J . Long tail or steep tail? a field investigation into how online popularity information affects the distribution of customer choices[J]. Working Papers, 2007, 6(1): 6.

[5] MAHAJAN R, SPRING N, WETHERALL D, et al. Inferring link weights using end-to-end measurements[C]//ACM Sigcomm Internet Measurement Workshop, Nov 6-8, 2002, Marseille, France. New York: ACM Press, 2002: 231-236.

[作者简介]

强鹏（1988− ），女，博士，华为技术有限公司高级工程师，主要研究方向为网络切片与确定性网络。

刘冰洋（1985− ），男，博士，华为技术有限公司主任工程师，主要研究方向包括网络架构、网络安全及可行、路由和命名解析、确定性网络等。

于德雷（1977− ），男，华为技术有限公司技术专家，主要研究方向为未来网络架构、网络切片、确定性网络等。

王闯（1975− ），男，华为技术有限公司技术专家，主要研究方向为未来网络架构与技术探索。

内生安全网络架构

江伟玉，刘冰洋，王闯
（华为技术有限公司，北京 100095）

摘　要：IP 网络通过连接全球大量的网络设备给人类带来了便利，但网络面临持续性的安全和隐私问题令人担忧。由于网络缺乏内生安全的设计，IP 地址伪造、隐私泄露、中间人攻击、分布式拒绝服务（DDoS）攻击等顽固安全问题难以根治，传统的补丁式解决方案补不胜补。在研究 IP 网络面临的各类安全威胁及相关安全技术的基础上，剖析了 IP 网络固有的安全缺陷，提出了具有内生安全特性的网络架构，包括具有内生安全的隐私 ID/Loc、安全验证和审计协议、跨域联合防御机制等，能够为端到端通信保驾护航。

关键词：IP 网络；内生安全；隐私；可审计性；DDoS
中图分类号：TN915.08
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019215

Network architecture with intrinsic security

JIANG Weiyu, LIU Bingyang, WANG Chuang
Huawei Technologies Co., Ltd., Beijing 100095, China

Abstract: IP network brings big benefits for human's life by connecting most devices all over the world, but its security and privacy issues make people frustrating when using end to end communication. Without intrinsic security design of the network, it is difficult for patch-like solutions to cure stubborn security issues (IP spoofing, privacy leakage, MITM attack, DDoS, etc.). On the basis of surveying different kinds of security threats and related security techniques, an overview of the security weakness analysis was given, and network architecture with intrinsic security (NAIS) was presented, including dynamic ID/IP with intrinsic security, security verification and audit protocols, and cross-domain cooperation defense mechanism, which could provide security and trustworthiness for end to end communication.

Key words: IP network, intrinsic security, privacy, accountability, DDoS

1　引言

基于 IP 协议的网络作为数据世界的管道，在构建万物互联的世界中扮演重要角色。然而，频发的安全事故显示，当前的网络面临巨大的信誉挑战，用户很难相信 IP 网络能够为端到端通信提供安全和隐私保护。IP 地址伪造是攻击者隐藏身份、发动攻击的惯用手段，然而应用网络数据分

收稿日期：2019−08−10；修回日期：2019−09−10
基金项目：国家重点研发计划基金资助项目（No. 2018YFB180079）
Foundation Item: The National Key Research and Development Program of China (No. 2018YFB180079)

析中心（Center for Applied Internet Data Analysis，CAIDA）数据显示，23.5%的 IPv4 和 25.4%的 IPv6 自治域系统依然允许携带虚假 IP 的数据分组流出[1]。暴露在头部的 IP 地址使用户面临较大的隐私泄露风险，已成为各网站识别用户身份、关联用户行为的重要指纹，WebRTC 的数据显示 80%以上的主流网站会通过提取用户数据分组中的 IP 地址追踪用户[2]。另外，网络层各类 DDoS 攻击依然是网络安全的一大顽疾，服务中断事件时有发生，严重破坏了网络服务的可用性。

由于 IP 网络设计之初缺乏安全考虑，外挂的补丁式安全技术难以为网络通信提供保障。在耦合了多种语义的 IP 地址基础上，现有方案难以平衡地址伪造带来的真实性破坏和 IP 地址追踪带来的隐私泄露问题，可审计性和隐私保护矛盾重重。虽然密码技术可为通信内容提供机密性和完整性保护，但是由于 IP 缺乏可信的自验证机制，作为安全锚的密钥交换过程存在欺骗威胁。DDoS 攻击源也已不再局限于外部网络，来自内部的攻击流量同样不可小觑。由于缺乏协议面内生的防御机制，外挂的防御方案无法低成本自适应迅速增长的攻击流量规模，联合作战也难以在现网部署。在边缘计算、雾计算、5G 等技术兴起的未来，基于安全域内部设备可信的假设更加难以立脚，依靠安全域隔离的防御方案已无法应对内部攻击。随着大量异构设备接入，网络环境将变得更加复杂，凸显的安全问题将成为网络技术广泛应用的主要障碍。IP 网络固有的安全缺陷使得外挂式安全方案无力对抗各类网络攻击，亟需内生的安全技术来构建安全可信的未来网络。

本文工作以解决顽固安全问题为出发点，以安全视角重新设计通信协议和安全机制，旨在为未来网络提供内生安全特性。本文首先给出了 IP 网络的安全分析总览，剖析了 IP 网络固有的安全脆弱性。在研究现有技术的基础上，设计了内生安全网络架构，通过将身份 ID 与位置 Loc 从 IP 地址中分离，提出了 4 种核心安全技术。

- 动态可审计的隐私 ID/Loc 技术。面向隐私保护和可审计性平衡难题，使用动态变化的隐私 ID 支持真实性验证和审计追踪，使用非对称的最小隐私揭露 IP 地址进行路由寻址。
- 去中心化的 ID 内生密钥技术。构建去中心化信任锚点，为域级身份及密钥管理提供可信保障，免除单点权威沦陷危机。基于 ID 内生密钥技术，规避 ID 与密钥脱绑带来的中间人攻击问题。
- 基于最小化信任模型的真实性验证技术。域内域间多步快速验证，过滤虚假流出/流入流量。基于最小化信任模型，防止接触网络流量的内部节点作恶，可辅助快速定位失效节点。
- 跨域联合审计和多级攻击阻断技术。新增安全审计组件，通过审计协议实现跨域追溯、源域定位和多级阻断非法流量。在源域过滤失效情况下，依然可赋予目的域有效的攻击防御能力。

2　安全脆弱性及技术需求

端到端通信安全脆弱性分析如图 1 所示，用户在使用 IP 协议进行端到端传输数据分组的过程中，存在源 IP 地址伪造、IP 地址头部隐私泄露、审计逃避、密钥交换欺骗、DDoS 攻击等顽固安全威胁。本节将具体分析现有的安全问题和脆弱性，提出确保端到端通信真实性、隐私性、可审计性、机密性、完整性和可用性等多种安全特性的技术需求。

（1）真实性需求分析

伪造源 IP 地址是破坏通信真实性的持久威胁。IP 地址伪造常被攻击者采用，用于发动会话劫持、中间人攻击和 DDoS 攻击等网络攻击，也是反射放大型 DDoS 攻击的必要攻击手段。在缺乏有效验证的情况下，攻击者可以轻易地生成大

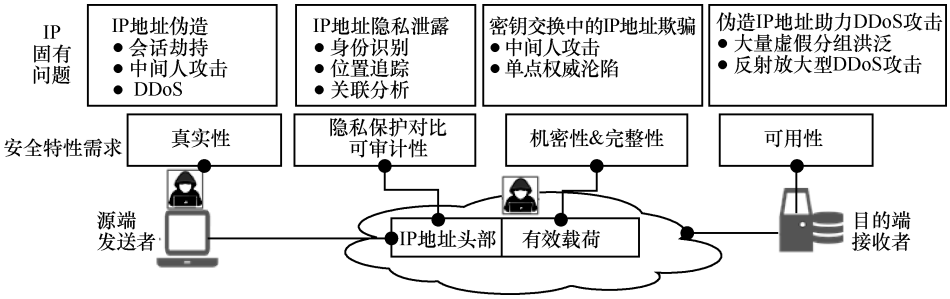


图 1　端到端通信安全脆弱性分析

量包含虚假 IP 地址的攻击流量，欺骗无辜接收者，逃避过滤和审计追踪。

为了对抗 IP 地址伪造，多种防御技术均被提出，但依然无法消除 IP 地址伪造威胁。由于 IP 地址缺乏内生的自验证机制，大量伪造 IP 地址的流量依然可在网络中畅行。即使部署了流入权/流出权过滤等技术[3-6]，受害主机依然难以识别经过精心构造的虚假 IP 分组，也无法有效处理 IPSec 和 Tor 等加密流量。例如，可在路由器上广泛部署的 uRPF 技术[7]，难以识别来自同一个源域的伪 IP 地址流量。实际上，到目前为止，仍有 20%以上的 AS 都没有部署源过滤机制，允许其流出的流量伪造 IP 地址，AS 在没有遭受攻击时部署自滤机制动力不足。因此，仅依靠源域自滤的流出权过滤技术（如（SAVI[8]）难以凑效，大量的虚假数据分组依然可到达受害主机。虽然基于密码学的路径验证技术（Passport[9]）在理论上有效，依赖数据分组传输路径上的多个 AS 添加密码学标记，缺乏激励多个 AS 协作的机制，很难直接在现网上快速部署。可见，由于缺乏内生安全的顶层设计，补丁式的安全方案难以根治 IP 地址伪造带来的真实性问题。

因此，对于持续的 IP 地址伪造威胁，未来网络需要从顶层设计真实性验证机制。一方面，随着网络规模的进一步扩大和网络接入设备的倍增，不仅需要防止外部虚假流量在受害主机侧汇聚，也需要验证来自内部的虚假数据分组。另外，在目的网络边界甚至近源对流出流入数据分组进行多级验证，可以免受单点验证失效威胁，大大削减基于 IP 地址伪造的 DDoS 攻击流的破坏力。另一方面，在真假流量混杂的大量数据分组转发背景下，网络层真实性验证机制应该不仅准确且高效。在确保准确性的前提下，需要考虑采用高效轻量级的密码验证技术以支持高速大吞吐的分组转发能力。

（2）隐私保护与可审计性需求对比分析

提供以真实性为基础的可审计性是确保网络可信的必要条件，但是可审计性的实现也给隐私保护带来巨大挑战。通过验证标识符是否由合法权威颁发可以保证真实性，但不一定能实现合法权威对非法流量进行快速精准的在线追踪和溯源。在真实性保障的前提下，可审计性需要一个静态且可验证的持久实名身份标识符，来实施非法流量的在线快速追踪溯源。然而，持久实名身份标识符的使用又会给合法用户的隐私带来威胁。近年来，隐私的关注率一直在上升，特别是随着欧盟 GDPR（General Data Protection Regulation）和我国网络安全法的发布。用户为了限制非授权的设备、链路窃听者以及访问的网络服务，根据数据分组中携带的持久标识符，识别发送者的身份、跟踪位置和关联发送者的网络行为。因此，未来网络需要平衡隐私保护与可审计性的关键技术。

尽管网络隐私保护与可审计性被研究了多年，但是依然缺乏非常有效的解决方案。例如，Passport[10]和 ISP 隐私方案[11]能够提供源 IP 地址

的隐私，但是不能隐藏发送者 ISP 的位置信息。Mailbox[12]通过提供多个代理位置来为多个 IP 地址接收数据分组，可以隐藏接收者的真实 IP 地址，但是并没有提供发送者的隐私保护。洋葱网络[13]被认为是一种有效的隐私保护技术，但在支持高速分组转发方面存在性能弱势。虽然 IETF 也有多项针对 IPv6 地址隐私保护的工作，但是大部分工作（如临时地址[14]、SLAAC[15]）聚焦在后 64 位的接口标识符隐私，没有考虑前缀隐私。实际上，5G 网络的 IPv6 PDU session 使用唯一的前缀作为发送者的标识符，并且在共用同一 IPv6 前缀的用户比较少的情况下，前缀隐私同样需要提供保护。为了提供可审计性，学术界也提出了多种方案，如 AIP[16]、APIP[17]、APNA[18]等。然而，AIP 仅考虑可审计性，却忽略了隐私保护；APIP 仅考虑源 IP 地址隐私，由于没有对所有分组进行验证，伪造的数据分组依然可以绕过审查。APNA 通过引入临时 ID 及证书，能够很好地平衡隐私保护和可审计性，但是需要频繁的请求、计算、颁发临时 PKI 证书来抵御长时间的关联分析。

实际上，由于当前 IP 网络固有的设计缺陷，耦合了身份（身份标识）和位置（定位符）语义的 IP 地址使外挂式安全方案很难兼顾隐私保护与可审计性。出于路由寻址目的，一方面需要在报文头部暴露精确的定位符，另一方面移动性也使得定位符会动态变化。因此，使用一种必然要暴露、动态变化的定位符以唯一标识用户，显然是不合适的。原因有两点：出于网络提供商对可审计性的考虑，需要一个可被合法实体唯一识别的、静态持久标识符，而定位符不满足静态持久特性；出于用户的隐私考虑，需要一个非授权实体无法识别的、足够匿名、动态不可关联的标识符，暴露在头部的定位符不满足匿名不可关联特性。显然，对于耦合了定位符语义的 IP 地址而言，难以身兼身份标识符语义，无法满足可审计性和隐私保护需求。因此，需要打破当前 IP 地址的安全设计缺陷，为未来网络提供一个动态不可关联、可审计追踪的隐私 ID 机制。

（3）机密性与完整性需求分析

人们普遍认为 IP 数据分组有效负载的机密性和完整性可以采用基于密码技术的安全协议来保证。例如，可以采用 TLS/SSL 或其他上层安全协议提供机密性和完整性，通过 IPSec 隧道模式甚至还可以实现 IP 报头信息的机密性。但是，所有这些协议的安全性都建立在安全的密钥协商基础上。如果密钥协商过程不可信，则数据安全性也就无从谈起。

在端到端通信场景中，动态的密钥交换协议广泛应用在密钥协商过程中，然而动态密钥交换又面临诸多安全威胁。由于静态的密钥配置方法复杂且不灵活，难以适用大量连接下的密钥交换需求。因此，基于 Diffie-Hellman 的动态密钥交换协议得到了广泛的应用。然而，在缺乏对 IP 地址与密钥进行绑定和验证的情况下，Diffie-Hellman 密钥交换过程容易遭受中间人攻击，中间人可以使用伪造的 IP 地址欺骗合法实体与其进行密钥交换，从而可以解密、窃听甚至篡改合法实体之间的加密流量。虽然通过 PKI 证书机制可以将 ID 与身份公钥进行绑定来防止中间人攻击，但是又会引入单点中心化权威作恶或者单点故障问题。由于 IP 缺乏内生的密钥生成和验证机制，现有外附的多级中心化 PKI 机制又存在可信和多级证书链验证带来的大计算开销等问题，难以满足大量多样化异构设备之间的安全通信需求。

因此，未来网络需要解决当前密钥交换中的欺骗问题，保障数据分组有效负荷的机密性和完整性。针对单点权威失效问题，去中心化的公钥基础设施已成为构建安全信任锚的必然选择，具体方案需结合场景设计以满足性能和安全的差异化需求。在此基础上，未来网络还需要具备内生的密钥自验证功能，以根治密钥交换过程中的欺骗问题。

（4）可用性需求分析

DDoS 攻击依然是破坏网络可用性的顽疾。在

2018 年，DDoS 攻击流量已突破 1.7 Tbit/s 量级[19]。2016 年广为人知的 Mirai 僵尸网络[20]、通过沦陷大量 IoT 设备，给网络带来了多起巨大安全风暴。随着越来越多的异构 IoT 设备的接入，DDoS 攻击威胁将会加剧，并且会打破现有基于防火墙的安全防御基线。作为 5G 的目标，未来 5G 网络将支持每平方千米百万量级的设备连接，因此，来自同一管理域内部的 DoS 攻击流量也不容小觑。

由于网络在设计之初缺乏安全考虑，现有主流的 DDoS 防御思想类似一场基于网络资源的军备竞赛。广泛应用的黑洞技术，虽然将攻击流量引入了黑洞，同样也致使了合法流量不能被处理。并且，大部分外附的 DDoS 防御技术通常实现在昂贵的专用硬件设备上（如部署在企业内部的 DDoS 防火墙设备[21]或部署在云中的 DDoS 流量清洗设备[22]），依赖于对深层数据分组的解析，滞后有时延，无法在网络层精准快速过滤。虽然，基于 BGP Spec Flow 的方案使得受害主机能够通过协议消息请求其他的 AS 协助 DDoS 流量过滤，但缺乏有效的激励机制来促动或约束非受害 AS 主动部署和协助 DDoS 防御。此外，现有大多数易于部署的 DDoS 防御方案无法独立于攻击流量规模，也难以做到有效的近源阻断，当大量的 DDoS 流量汇聚到受害主机侧时，正常的服务已受到影响。虽然 SIBRA [23] 提供的 DDoS 防御能力可以独立于攻击流量规模，但是其预留带宽的机制依赖于复杂的分级层次化的 AS 关系设计，难以被广泛接受。总而言之，如果没有大量的计算或带宽等资源优势，现有的补丁式防御方法仍然难以对抗压倒性的 DDoS 攻击流量。

因此，在未来的网络架构及协议设计中，需要内嵌多层次的 DDoS 防御技术，在可能的攻击流量路径上，尽可能早地发现攻击流量、在靠近攻击源侧进行阻断，同时使目的端具备区分合法和非法流量的能力，使得网络节点或实体具备 DDoS 攻击免疫特性，从而确保网络服务或节点的可用性。

3 系统方案设计

本文针对网络面临的顽固安全问题，通过重新设计 IP 网络协议及安全机制，提出了一种具有内生安全特性的网络架构（network architecture with intrinsic security，NAIS）。NAIS 基于最小信任模型，不完全信任管理域（如自治域系统）内部和外部的实体。所有来自内外部的流量真实性都将被验证。仅根据功能需要向少部分网络节点揭露 IP 地址分组头部的隐私信息，对于内部不可信的网络节点同样采取了保护措施避免隐私泄露。NAIS 充分利用了：

- 身份标识符和位置标识符分离技术，将身份标识符 ID 与位置标识符 Loc 从当前 IP 地址中解耦，以兼顾真实性、可审计性和隐私性；
- 临时身份标识，短期有效的主机标识符和身份凭证，用于对抗标识符伪造攻击和身份隐私泄露；
- 动态的位置标识符，频繁变化、部分信息加密的 Loc 以防止识别、关联分析和位置追踪。

为了确保端到端通信安全，NAIS 提供了 5 类安全功能组件：身份管理者（identity manager，IDM）、审计代理（accountability，AA）、本地 DHCP 服务器、ID 验证者和边界路由器。身份管理者负责管理 AS 域内的身份以及签发动态临时的匿名身份标识符（ephemeral and encrypted identifier，EID）及其凭证给终端主机。审计代理负责对非法流量进行追踪审计。ID 验证者是一种具备真实性验证功能的路由器，负责验证发送者的真实性，过滤虚假或恶意的数据分组。本地 DHCP 服务器管理和分发动态变化的位置标识符（ELoc）给终端主机。内生安全网络架构如图 2 所示，数据分

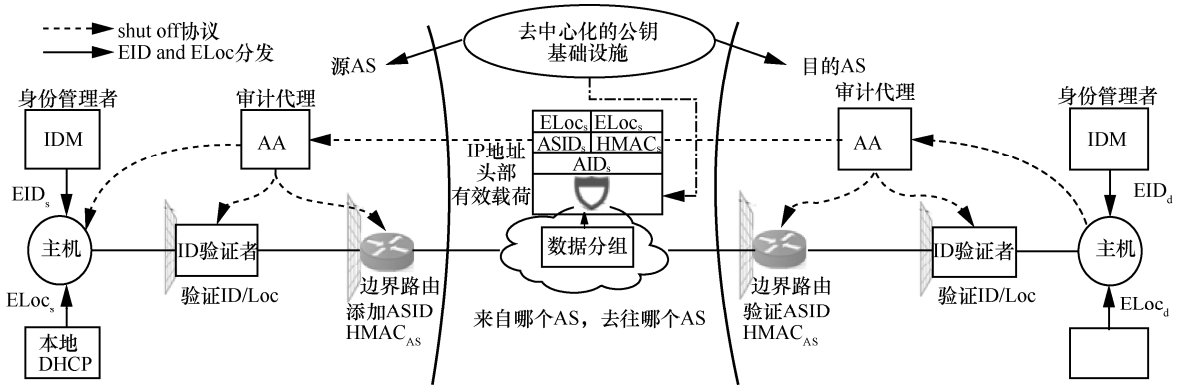


图 2　内生安全网络架构

组在源端主机生成、经网络传输、到达目的主机的过程中，发送者数据分组中的标识符信息，会被 ID 验证者检查以确保流量的真实性和合法性，源 AS 边界路由器在转发数据分组到互联网时同样会校验源的真实性、并添加一个 $HMAC_{AS}$ 作为可验证的域标记，数据分组在到达目的 AS 边界路由器时，AS 边界路由器通过验证 $HMAC_{AS}$ 可以过滤掉虚假来源的数据分组。

为了确保端到端通信安全，NAIS 通过设计和修改具备安全特性的网络协议，新增和部署具备安全功能的网元，提出了 4 种核心技术，具体如下。

（1）动态可审计的隐私 ID/Loc

具有迷惑攻击者的动态 ID 和 Loc 可用于防止隐私泄露和关联追踪。源 AS 域外部的网络节点仅能揭露数据分组来源于哪个 AS，而无法获知发送者具体所在的子网和位置信息以及发送者的长久标识符。路由器仅能从 IP 地址头部获得最小必要信息进行数据分组转发。当攻击发生时，受害主机可以追溯非法流量到其所属的源域，只有源 AS 域的 AA 才能打开发送者的真实身份，根据其安全策略，发送控制命令进行流量过滤。NAIS 通过对永久 ID 和真实 Loc 进行加密，生成的 EID 和 ELoc 不仅具有匿名性，并且通过动态变化标识符可以防止关联分析和追踪。

（2）去中心化的 ID 内生密钥

在本安全架构中，采用了去中心化的公钥基础设施来确保信任锚的安全域可信，AS 域级别的身份 ID 及密钥管理基于去中心化的基础设施，可以防止单点权威失效或作恶带来的安全问题。终端主机的密钥管理可被隔离在一个管理域内，即域密钥的沦陷仅会影响域内终端主机的可信性。而实际上，当前 PKI 权威在分发和撤销证书时，缺乏具体且清晰的边界，某个权威 CA 的沦陷带来的安全问题具有蔓延扩散特性。在 NAIS 中，用于协商会话密钥的身份公钥内生于主机标识符，即接收端可以根据主机 ID 直接派生出信息验证公钥，赋予了 ID 自验证公钥的特性，从而能够防止密钥交换过程中的欺骗。

（3）基于最小信任模型的真实性验证

去除安全域划分机制中对域内流量的信任弊端，域间和域内流量都将被验证。对于需要跨域传输的网络流量，当数据分组从源端发送后，源 AS 域的 ID 验证者和边界路由器均都会对其流出流量进行验证。NAIS 不假设 ID 验证者完全可信，根据不同场景，通过部署不同的匿名验证方法，如零知识证明技术、盲签名技术等，可以在 ID 验证者处实现匿名验证功能，兼顾真实性和隐私保护。当域间流量到达目的 AS 时，目的端的边界路由器根据域间共享的密钥，对流入流量进行验证。因此，跨域传输流量的真实性可以通过流出和流入权多步验证得到保障。对于来自 AS 域内部的网络流量，接收者侧的 ID 验证者同样需要验

证其来源真实性，从而使得当 AS 域内部发生故障和攻击时，网络管理面能够快速定位错误进行恢复。

（4）跨域联合审计和多级攻击阻断

本方案赋予了目的域区分合法流量和非法流量的能力，设计的审计协议可以实现跨域的追踪溯源和攻击切断。目的 AS 域的边界路由器基于源域边界路由添加的 $HMAC_{AS}$，对每个流入的数据分组进行验证，识别和阻断所有携带虚假域 ASID 的 DDoS 攻击流量。同时，根据不同目的端需求，可在 IP 地址头部携带可被目的端验证的 AID，从而使目的端网络具备区分合法用户流量与攻击者流量的能力。当一个具体的攻击流被识别后，受害主机通过发送一个 shut off 审计协议消息给其所在域的 AA，再到达目的域的 AA，进行非法流量的溯源和阻断。源 AS 域的 AA 验证该审计请求、提取出原始数据分组中的隐私标识，打开发送者的真实身份和位置，最后发送过滤非法流量的控制命令给源主机或验证者或边界路由。在源主机沦陷的情况下，审计者可以根据安全策略和指控同一主机的 shut off 请求数量，请求验证者过滤掉虚假流量。即使在验证者沦陷的情况下，边界路由器也具备过滤非法流量的能力。对于源 AS 不诚实执行攻击阻断的情况，目的端依然可以通过限流等措施减轻 DDoS 攻击的危害。

4 应用探讨和案例分析

随着 5G 网络和边缘计算[24]的发展，为了提高响应效率和充分利用资源，大量的连接请求处理和计算操作将下沉到靠用户侧的网络节点。本文提出的内生安全架构基于最小化信任模型，在隐私保护、真实性、DDoS 防御等方面，不完全信任管理域内外部的所有节点，可应用在边缘计算、5G 网络等存在不完全可信内部节点的场景。

边缘计算应用探讨如图 3 所示，在主机接入和使用网络的过程中，某些计算操作将从运营商的核心机房走出到位于开放物理环境的不可信设备。例如公共可用的 Wi-Fi 设备、基站以及其他下沉的路由设备等。由于暴露在开放式环境中、安全防护措施难以部署等原因，此类设备容易遭受攻击、泄露隐私或被攻击者控制。一方面，出于路由目的，此类不完全可信设备依然需要接触身份标识符或位置标识符等敏感信息；另一方面，针对未来小空间大规模的网络连接需求，存在将某些验证终端主机的计算下放到此类设备的需求，以减小集中化处理带来的大计算开销和 DDoS 攻击风险。如何确保这些设备在保护隐私的情况下，依然能够诚实地执行验证计算是一个挑战。因此，采用本文提出的动态可审计隐私 ID/Loc 技术，可面向不完全可信计算节点，实现基于隐私 ID 的匿名鉴别技术，不仅可以防止内部节点和外部 Internet 节点泄露隐私，减轻集中式处理大量连接请求带来的 DDoS 攻击风险，同时能确保真实性和可审计性。

5G 网络应用案例如图 4 所示。5G 网络在支持合法监听和追踪溯源等可审计性时，也非常重视隐私。在 5G 控制面的主鉴别流程中，为了防止永久标识符 SUPI 在空口泄露，采用了非对称密码

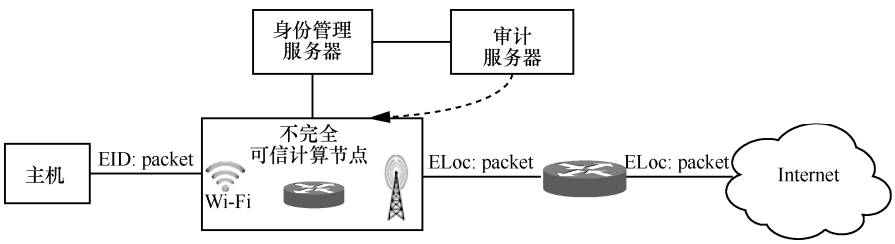


图 3 边缘计算应用探讨

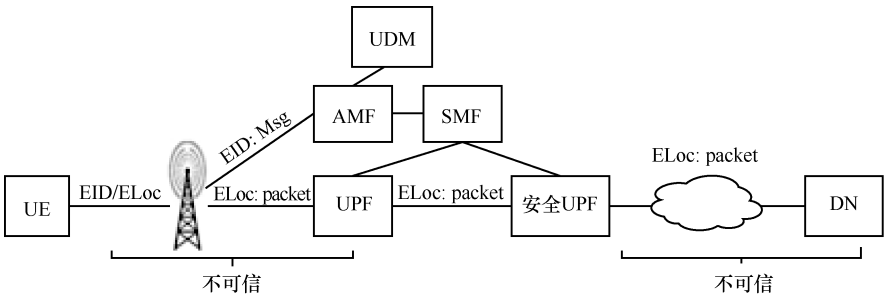


图 4　5G 网络应用案例

技术加密 SUPI，但也给集中式的身份数据管理服务器 UDM（unified data manager）带来了 DDoS 攻击威胁。本文的隐私 ID 技术可实现在 UDM 中，而匿名验证技术可实现在分布式的 AMF 中。通过 UDM 分发控制面的隐私 ID，不仅可以防止空口窃听者、不可信基站甚至漫游场景中不同安全域的 AMF 获得永久标识符，也能支持必要的审计追踪。通过将身份验证计算下发给 AMF 或者基站，可以在确保接入用户真实合法的情况下，规避针对 UDM 的 DDoS 攻击。另外，隐私 Loc 技术也可以应用在 5G 数据面。由于 5G 的 IPv6 类型流量使用唯一的 IPv6 前缀识别用户身份和位置，采用本文的动态 ELoc 替代数据面静态的位置标识，可以大大地减小隐私泄露风险。

5　结束语

本文通过分析当前网络面临的顽固、持续性安全威胁，对网络固有的安全缺陷进行了剖析，提出了新型的具有内生安全特性的网络架构。该安全架构面向未来网络，基于最小化信任模型，提出了 4 种核心技术，能够为端到端通信提供内生的安全与可信能力。通过将身份和位置从 IP 地址中解耦，在协议层面，提出的动态可审计的隐私 ID/Loc 技术，以安全视角设计新的 IP 地址头部，不仅能够确保网络流量的真实性，也能很好地兼顾隐私保护与可审计性。内生安全网络架构依赖于去中心化公钥基础设施、验证者、审计代理等新增功能组件的实现和部署，不仅能够打造坚实可信的通信协议信任锚，也能防止通信过程中的欺骗和隐私泄露问题。同时，跨域审计协议和多级攻击阻断技术也进一步加固了网络的安全性。最后，本文探讨了边缘计算应用场景和 5G 应用案例，阐述了本文技术的安全优势。

参考文献：

[1] Caida spoofer program report[R]. 2019.

[2] ENGLEHARDT S, NARAYANAN A. Online tracking: a 1-million-site measurement and analysis[C]//the 2016 ACM SIGSAC Conference on Computer and Communications Security, Oct 12-16, 2015, Denver, Colorado, USA. New York: ACM Press, 2016: 1388-1401.

[3] FERGUSON P, SENIE D. Network ingress filtering: defeating denial of service attacks which employ IP source address spoofing: RFC 2827[S]. 2000.

[4] DUAN Z, YUAN X, CHANDRASHEKAR J. Constructing interdomain packet filters to control IP spoofing based on BGP updates[C]//INFOCOM, April 23-29, 2006, Barcelona, Spain. Piscataway: IEEE Press, 2006.

[5] BREMLER-BARR A, LEVY H. Spoofing prevention method[C]//IEEE INFOCOM, 2005(1):536–547.

[6] JIN C, WANG H, SHIN K. Hop-count filtering: an effective defense against spoofed DDoS traffic[C]//the 10th ACM Conference on Computer and Communications Security, Oct 27 - 30, Washington D.C., USA. New York: IEEE Press, 2003.

[7] BAKER F, SAVOLA P. Ingress filtering for multihomed networks: RFC 3704[S]. 2004.

[8] BI J, LIU B. Problem statement of SAVI beyond the first hop, Internet Draft[R]. 2012.

[9] LIU X, YANG X, WETHERALL D, et al. Efficient and secure source authentication with packet passports[C]//2nd conference on Steps to Reducing Unwanted Traffic on the Internet, San Jose, CA, USA. New York: ACM Press, 2006.

[10] LIU X, YANG X, WETHERALL D, et al. Passport: secure and adoptable source authentication[C]//5th USENIX NSDI, April

16-18, 2008, San Francisco, California, USA. New York: ACM Press, 2008.
[11] RAGHAVAN B, KOHNO T, SNOEREN A C, et al. Enlisting ISP to improve online privacy: IP address mixing by default[C]//PETS '09, August 5-7, 2009, Seattle, WA, USA. New York: ACM Press, 2009.
[12] CHAUM D. Untraceable electronic mail, return address, and digital pseudonyms[J]. Communications of the ACM, 1981, 24(2):84-88.
[13] DINGLEDINE R, MATHEWSON N, SYVERSON P F. Tor: the second-generation onion router[J]. Journal of the Franklin Institute, 2004, 239(2):135-139.
[14] NARTEN T, DRAVES R, KRISHNAN S. Report from the IAB workshop on routing and addressing: RFC4941[S]. 2007.
[15] GONT F. Method for generating semantically opaque interface identifiers with ipv6 stateless address autoconfiguration (slaac), RFC 7217, Internet engineering task force, request for comments[S]. 2014.
[16] ANDERSEN D G, BALAKRISHNAN H, FEAMSTER N, et al. Accountable internet protocol[C]//ACM SIGCOMM 2008 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, August 17-22, 2008, Seattle, WA, USA. New York: ACM Press, 2008.
[17] NAYLOR D, MUKERJEE M K, STEENKISTE P. Balancing accountability and privacy in the network.[J]. ACM Sigcomm Computer Communication Review, 2014, 44(4).
[18] LEE T, PAPPAS C, BARRERA D, et al. Source accountability with domain-brokered privacy[C]// the 12th International on Conference on emerging Networking Experiments and Technologies, December 12 - 15. 2016, Irvine, California, USA. New York: ACM Press, 2016.
[19] US service provider survives the biggest recorded DDoS in history[Z]. 2018.
[20] ANTONAKAKIS M, APRIL Y, BAILEY M, et al. Understanding the mirai botnet[C]//26th USENIX Security. [S.l.:s.n.], 2017.
[21] CloudFlare. Advanced DDoS attack protection[Z]. 2018.
[22] FAYAZ S K, TOBIOKA Y, SEKAR V, et al. Bohatei: flexible and elastic DDoS defense[C]// Usenix Conference on Security Symposium, August 12 - 14, 2015, Washington, D.C., USA. New York: ACM Press, 2015.
[23] BASESCU C, REISCHUK R M, SZALACHOWSKI P, et al. SIBRA: Scalable internet bandwidth reservation architecture[J]. arXiv preprint arXiv:1510.02696, 2015.
[24] 李子姝,谢人超,孙礼,等.移动边缘计算综述[J].电信科学,2018,34(1):87-101.
LI Z S, XIE R C, SUN L, et al. A survey of mobile edge computing[J]. Telecommunications Science, 2018, 34(1): 87-101.

[作者简介]

江伟玉（1987− ），女，博士，华为技术有限公司 2012 实验室中央研究院网络技术实验室高级工程师，主要研究方向为网络安全架构及协议、可信身份管理、数据安全等。

刘冰洋（1985− ），男，博士，华为技术有限公司 2012 实验室中央研究院网络技术实验室主任工程师，主要研究方向包括网络架构、网络安全及可行、路由和命名解析、确定性网络等。

王闯（1975− ），男，华为技术有限公司 2012 实验室中央研究院网络技术实验室技术专家，主要研究方向为未来网络架构与技术探索。

基于微服务架构的下一代 IP 网络测试体系框架

徐竟祎，赵泽宇，沈敏虎，应奕彬，周伟强
（复旦大学，上海 200433）

摘　要：下一代 IP 网络是未来新型公共网络，融合有线无线一体化、宽带窄带一体化、传输接入一体化以及有源无源一体化的全新型业务网络，其面临的应用场景也变得复杂化和多样化。然而，每个应用场景的需求及标准各不相同，对标现有的网络技术、协议、标准规则仍有较大差距。为保障新应用场景下各技术、设备等方面的完备性及安全性，满足未来网络技术发展产生的新应用特征，需建立全方位统一的测试体系。提出在基于微服务架构的基础上强调对业务的赋能，综合不同应用场景特征，建立基于下一代 IP 网络场景的全过程领域驱动模型。以构建多渠道、全方位的微服务测试框架，为网络技术的发展提供支持手段。

关键词：测试体系框架；微服务；下一代 IP 网络
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019218

Next generation IP network test system framework based on microservices architecture

XU Jingyi, ZHAO Zeyu, SHEN Minhu, YING Yibin, ZHOU Weiqiang
Fudan University, Shanghai 200433, China

Abstract: The next generation IP network is a new type of public network. It integrates wire and wireless, broad-band and narrow-band, transmit and receive, passive and active, which leads to the diverse and complex service scenarios. However, different scenario has different needs and standards, and there remains gaps with the current techniques, protocols and standards. In order to ensure the completeness and security of all kinds of techniques and facilities in the new application scenarios, and the upcoming new characteristics in the process of network development, a comprehensive test system framework needs to be established. The next generation IP network test framework based on micro services architecture was proposed, combined with different new application characteristics, which provided support for the development of network technology.

Key words: test system framework, micro service, next generation IP network

1　引言

下一代 IP 网络是未来新型公共网络，以 IP 技术为基础，融合有线无线一体化、宽带窄带一体化、传输接入一体化以及有源无源一体化的新型业务网络。随着技术的发展，新业务层出不穷，

收稿日期：2019−08−10；修回日期：2019−09−12

互联网与车联网、远程医疗、AR/VR 等新技术领域深度融合，要求下一代 IP 网络具有高速率、低时延、高带宽、高并发等特征，对传统的工业领域、生活领域、技术领域带来了广泛影响，导致应用场景泛化，对传统的 IP 网络技术以及测试方式方法都带来新的挑战。传统测试以单一测试场景为主，主要对场景功能、设备性能、平台压力进行测试。在下一代 IP 网络场景下，还应从场景需求和特征出发，考虑设备的异构性、场景的组合性，从服务需求角度构建测试框架。传统测试方法主要面临以下几个挑战。

（1）下一代 IP 网络涌现众多创新应用

网络技术的发展，产生了纷繁的新应用，互联网承载的应用越来越丰富，远程医疗、工业互联网、智慧校园、AR/VR 等新应用已悄然而至，全息通信、泛在移动、边缘计算等应用场景也即将到来，人们正在进入一个万物感知、万物互联的智能化世界。

（2）多元化测试指标

网络的新应用对 IP 网络提出了新的需求和挑战。而随着技术的发展，性能指标早已不是限制技术发展的关键因素，而是以面向服务的方式，对场景的符合性测试才是行业内普遍关注的重点。

（3）设备的异构性

由于新引用场景本身特征和运行环境的不同，其设备所支持的网络通信协议、依赖的传输网络、数据解析和处理的过程都截然不同，且除了传统的网络设备，越来越多物联设备、通信设备的涌现，无法通过统一的测试体系框架对所有设备的链接性、安全性、一致性等进行测试。

（4）传统测试模型无法满足新业务需求

在传统的测试方法上，主要测试流程先分解测试目标、定义测试用例、制定测试预期结果，从而反复迭代；而当被测对象或需求场景有变动时，就必须修改其源测试方案，导致测试周期变长；且由于每个测试项目的测试场景及目标不同，已有的测试方案的重复使用率较低，导致测试成本升高。

针对以上挑战，本文提出采用基于微服务的测试体系框架可以有效地提高测试效率，通过搭建各类微服务模块，借助其服务独立、易扩展性等特征，整合各类应用场景，建立基于场景特征的全过程驱动模型，制定统一的测试体系框架，使得在未来 IP 网络新应用的场景下，对其整体的应用测试过程能够灵活快速地适应新业务的持续发展状况，降低测试方案的复杂度及测试成本。

2 研究现状

当前，下一代 IP 网络技术是一个重点研究的方向，不仅影响着产业发展前景，同时给国家利益及网络安全带来改变，因此各个国家都展开了针对下一代 IP 网络的研究计划[1]，且大数据和人工智能的发展，也为下一代 IP 网络向智能化发展提供了新的契机。从现阶段来看，目前网络基础架构、网络协议等方面在面对下一代 IP 网络的新应用需求时还存在诸多问题。由于业务环节、业务范畴及运营需求的改变，现有 IP 网络正面临着一系列的挑战，很多需要网络核心支持的协议功能（例如移动 IP、InterServ、DiffServ 等控制协议）都难以成功部署，约束了网络的创新能力[2]。蒋林涛[3]认为在网络层面，网络能力和性能的缺失和不足、网络智能管理策略不足、网络自动部署、自愈能力不足，都成为信息业发展的瓶颈。

目前常用的针对网络层面的测试框架大部分集中在对设备性能及传输协议的测试。参考文献[4]提出互操作性测试，用于检测在两个或两个以上的协议实现关于能否互联互通互操作的测试。为了提高测试效率，参考文献[5]采用了分层体系结构，设定任务模板，规范测试任务的输入域输出，有效提升了整体测试效率。

但传统的测试框架已无法满足未来多应用场

景下的特征和模式需求。因此，很多学者开始关注采用微服务的架构与新型网络场景结合，参考文献[6]提出了采用微服务和物联网的云平台系统来构建智慧城市管理平台，Sun 等人[7]提出了通用型的物联网微服务中间件，有效解决通信设备在传输协议上的差异性。在测试层面中，参考文献[8]提出用微服务架构实现对软件的自动化测试；参考文献[9]提出基于场景特征的测试体系架构。

但上述诸多研究成果均经过在未来 IP 网络中场景的多样性、终端的海量性以及传输的异构性等特征方面的综合考虑，缺乏统一的测试框架可以适用于未来多变的业务场景，且不够重视场景的性能特征以及业务请求之间的差异，对不同测试服务的互操作支持都缺少详尽的考虑。

3　下一代 IP 网络典型应用场景及特征

3.1　移动互联网

移动互联网的主要应用场景包括 5G、6G、物联网等。很多传统的行业例如能源、市政、交通等都将参与移动互联网生态环境的建设。和以往移动通信技术相比，速率更高、终端更多；未来移动互联网需要满足更加多元化的场景，面临更多极致的性能挑战。

（1）海量物联网终端通信

以智慧交通、智慧校园、智慧城市[10]为典型应用场景，主要以数据采集和传感为目标，具有低功耗、低流量、海量链接等特点，且要求支持高密度链接。

（2）泛在移动性

在下一代高铁与磁悬浮列车应用场景中，需要保证在高速移动下的车载业务不中断。需要基于物理位置所分配的 IP 地址不断变化，因此针对多种业务，网络需要提供按需、泛在的移动性支持。

（3）网络安全性

在特定应用场景中，未来 IP 网络在安全防护中，应具备对网络攻击的自动识别以及自我防御能力，具有网络故障的自动检测和自动恢复能力。

3.2　产业互联网

传统实体行业与当今网络技术的结合，催生了产业互联网。产业互联网将重点关注生产智能化、定制个性化以及网络协同化等场景特征。产业互联网的未来新应用场景包括智能工厂、车联网、远程医疗、智能电网等多种引用场景；综合产业互联网海量终端、时延低等新特性，其普遍对下一代 IP 网络的需求如下。

（1）网络体系需求

需要建设高可靠、高覆盖、高带宽、可定制的企业网络基础设施；具备 IPv6 支持能力及数以百计的产业物联网终端设备。

（2）平台体系需求

协同传统产业协同发展产业互联网平台，满足企业的云化需求，保障企业工艺的流程控制，优化时间敏感度、实时性、稳定性及低功耗。

（3）安全体系需求

对场景下设计的所有设备、数据、平台、网络建立安全态势感知及安全保障能力。

（4）网络自适应性

在车联网等新应用中，要求节点之间通信链路的切换时间大幅缩短，导致网络拓扑将频繁变换，针对此场景，网络应具备自适应的能力。

（5）组网方式不同

包括实体组网、虚拟组网的混合组网方式。

3.3　全息通信

当前通信网络传输最多为二维信息，而人类从日常行为到复杂操作都高度依赖人们的视觉感知系统。全息通信则可获取三维场景，让人们感知视觉信息的全面性、真实性及沉浸感。将互联网技术与全息通信技术相结合可以突破远程医疗、赛事直播、游戏娱乐[11]等应用场景中三维视觉信息丢失问题，具有十分广阔的应用前景。相比于传统二维通信，预估其传输量是传统方式的

40 000 倍。全息通信场景下对下一代 IP 网络的需求如下。

（1）大带宽高吞吐

全息通信可提供双目视差、运动视差等视觉信息，其数据更全面、更真实，其带宽传输要求通常达到 8K 高清视频的传输要求。

（2）低时延

以保障全息通信的实时性要求。

3.4 下一代 IP 网络体系架构

下一代 IP 网络的体系架构如图 1 所示，包括基础设施链路层、数据层、管理层、控制层、网络层、传输层和应用层。具体包含以下。

（1）下一代 IP 网络基础设施链路层

为下一代 IP 网络提供安全可信的基础设施内核，是整个体系架构的基础。

（2）数据层

涵盖下一代 IP 网络中的所有元素，包括基础设施资源、网络数据资源、终端数据等。真正实现万物互联，打通存储、计算、内容资源的连通性。

（3）网络层

支持下一代 IP 网络架构中的网络层应能实现面向万物互联提供网络技术，包括新寻址协议、新路由协议、确定性低时延协议、内生安全协议。

（4）传输层

能面向高通量低时延场景提供新型传输协议。

（5）控制层

支持超高通量、网络编码，对多种网络和资源提供高效路由和控制协议策略。

（6）管理层

通过资源优化、差异服务及策略执行，实现网络从传输、控制到管理的全面智能化管理，构建基于人工智能、机器学习的大数据管控平台。

（7）应用层

基于底层协议支持上层开展符合下一代 IP 网络的服务业务，支持全息通信、远程医疗、智慧工场等重要场景。

综上的整体网络体系架构，其对应的测试框架要涵盖下一代 IP 网络体系的方方面面，针对网络基础设施链路层，需要验证在设备的异构性以及保障设施终端的互联互通和可操作性；数据层，则作为整个体系框架中的所有数据的汇总层，需要保障数据的安全性以及数据的共享性；

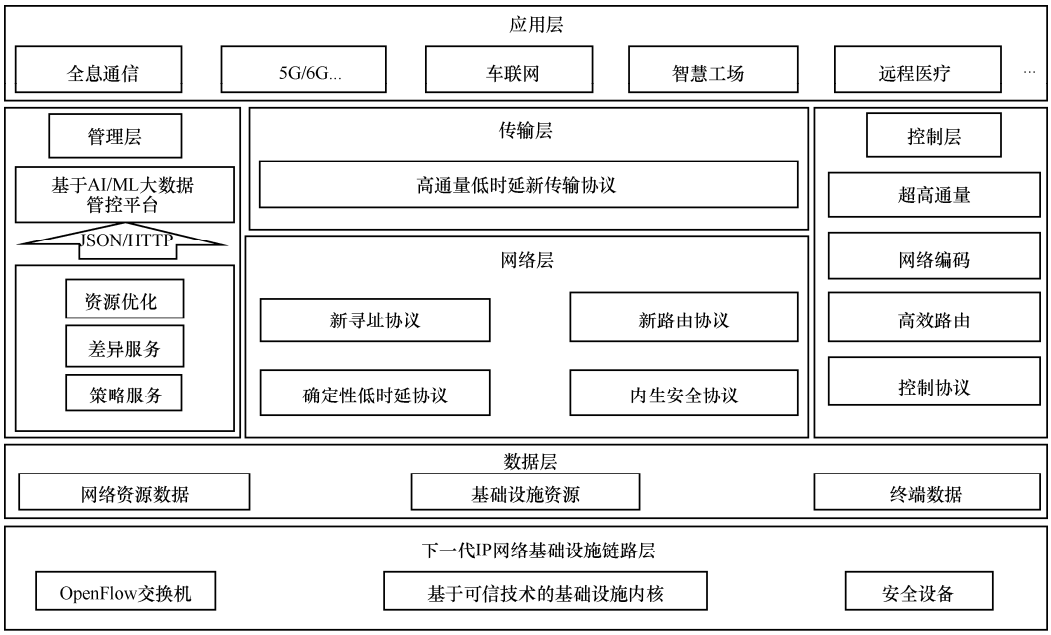


图 1　下一代 IP 网络体系架构

网络层及传输层是整体架构中的核心，负责稳定及高性能的传输，一方面要验证协议的一致性，另一方面也要保障整体核心层级的安全性；控制和管理层则是实现对整体框架的智能化管理，验证整体网络结构的顽健性；最后，针对应用层，要针对场景特征进行黑盒测试，验证服务的正确性。综上，提出了基于下一代 IP 网络的测试体系框架。

4　测试体系及框架

下一代 IP 网络的可变性、多样性及复杂性的特点决定了网络测试系统的复杂性，为了验证其安全性、可靠性，对于不同的应用场景应当提供相应的测试框架及手段。因此基于下一代 IP 网络的测试框架应当具备随网络技术、网络拓扑多样性发展而持续更新、不断迭代扩展、融合多场景应用的能力。此外，随着下一代 IP 网络技术的成熟，业务领域及场景不断扩展和细分，针对典型应用场景设计的通用型测试框架势在必行。而微服务架构具有灵活的扩展性、部署速度快等特点，成为整个测试体系框架的架构首选。

测试的总体架构如图 2 所示，分为基础设施资源层、数据共享层、服务层、应用层。基础设施资源层包括自适应的测试硬件平台，提供测试的基础设施及组件；数据共享层则基于场景业务的统一数据存储层，提供测试数据的存储与解析；服务层由不同的微服务应用所组成，为场景提供对应的场景分类模型、业务流程处理以及数据共享等服务；应用层为测试者提供一个交互接口。整体测试框架提供标准的负载均衡策略、监控报警功能以及扩展性能的服务组件，支撑所有微服

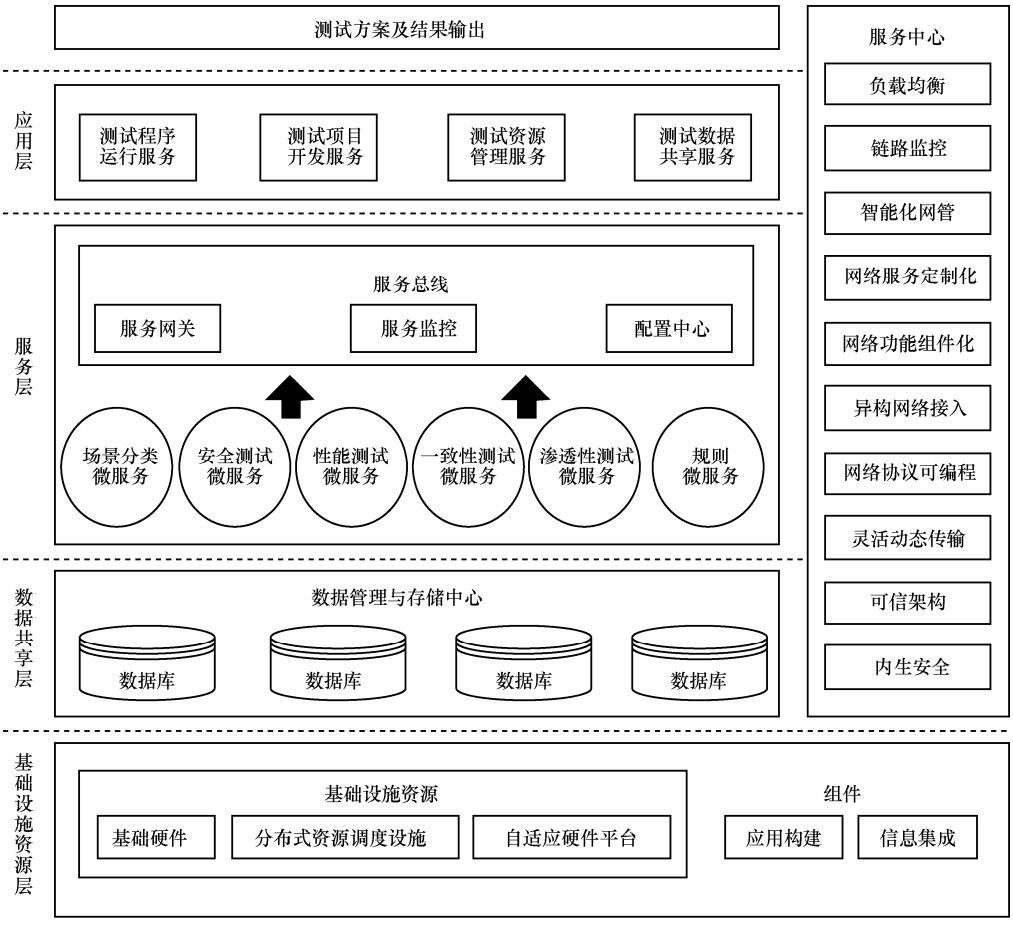


图 2　下一代 IP 网络测试框架

务相互组合运行。

基于总体测试框架的设计，充分考虑各业务场景下的性能特征，设计了一个适用于多场景化的综合测试体系框架，使其能面向多渠道提供公共测试服务，通过对业务场景的抽象及特征提取，共享数据和事务组件，对纵向业务场景资源进行划分，构建完整的测试用例。

基于场景的领域模型如图 3 所示。首先基于场景配置文件，按照特征、性能及相关场景特性进行分类，通过规则微服务，对各类微服务应用进行组合调度；每个微服务又包含多个测试实例，采用容器技术，实现整体测试框架的快速部署和水平扩展的能力。通过不同微服务的组合，针对不同的应用领域产生特定的测试手段及策略，有效解决下一代 IP 网络体系中，场景的多样性、设备的异构性以及体系的扩展性等问题。

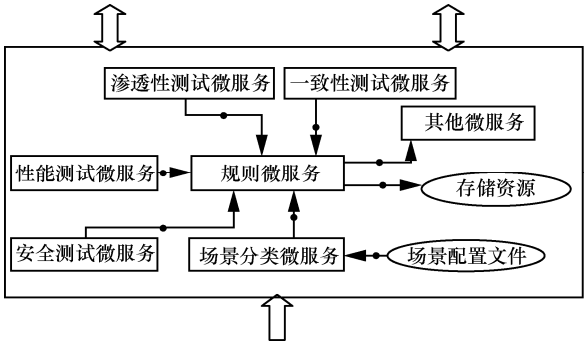


图 3　基于场景的领域模型

5　关键服务设计

为保障不同的应用场景能有效接入整体测试框架，采用微服务架构实现快速部署、灵活扩展。但如何有效组合不同的测试微服务，都依赖于具体的、不同的微服务组件设计。

5.1　场景分类微服务

场景分类微服务负责将不同的应用场景进行特征提取，并抽象成服务调用与资源调度之间的相互转化。要完成上述机制，首先要从不同应用场景中抽象出服务的机制，能够屏蔽场景使用的传输网络、通信协议等方面的差异特征。针对应用场景概念包含的 3 个方面——业务主题、业务特性、传输通信方式，依据相关理论进行分类。在保证分类原则的系统性、综合性、稳定性、唯一性和可扩展性的前提下，构建出一个有层次的、逐渐展开的分类体系。主要配置文件包含：场景描述文件、场景需求及特征文件。

场景分类微服务依据场景提供的服务特征创建场景描述文件，包括其场景特征、传输网络、属性等异构特点。具体的场景描述文件中的关键字段见表 1。

表 1　场景描述文件关键字段

关键字段	描述
time	实时性
mobility	移动性
quality	QoS 保障
flow	数据流量
concurrency	单位区域连接数
protocol	通信协议
attribute	属性

根据场景描述文件的不同，会选择合适的测试资源。例如在物联网应用中，智能水表使用的是低功耗、窄带广域网技术；而在智能家居中，往往采用的是无线局域网，不同的传输特性在连接数、QoS 等方面都相差甚远。

5.2　性能测试微服务

RFC2544 提供了一个测试基准，采用测试帧反映性能特征，包括吞吐量测试、时延测试、分组丢失率测试等，每个测试指标反映了不同的侧重点。在性能测试微服务中，根据性能需求而设计测试方案。为验证线路质量，应对不同长度的以太帧进行至少 15~30 min 的传输测试。在被测线路上以特定速率传送规定数量的以太帧，统计传输返回帧的数目。当两者数目相同时，此时最大的传输速

率则是吞吐量。在车联网场景中，考虑到存在多种以太网通信方式（单播、多播、全双工或自协商等），应结合以上场景特征，进行设计。

负载测试、压力测试、可靠性测试、恢复性测试也是性能测试子类，针对不同场景的不同特点及要求进行合理选取。通过配置文件的场景特征，验证性能指标是否符合服务场景的要求。例如在车联网场景中，负载测试可以测试在网络切换时间低于 1 ms 的情况下车的最大速率，获取通信转化能力的极限时间，评测不同速率下的网络切换性能表现。压力测试则是在面向海量终端时，在一定吞吐量和分组丢失率下，明确该场景下的最大支持终端数，以此方法判断系统通信的稳定性。可靠性测试则是在一定终端和速率下，车联网关键性能指标是否均能满足要求，可提早发现未知故障。恢复性测试，则是通过针对 ARP、TCP 等协议的故障注入，模拟错误的以太网帧内容，检测能否在制定时间内修正、自动恢复，以保证通信的稳健性。

5.3　一致性测试微服务

一致性测试是一种黑盒测试，它按照标准协议，观察被测试结果是否和预期结果一致，解决设备的异构性，保障不同的设施之间可以正确互联互通。随着物联网、5G 等新应用的诞生，推动了网络终端需求的增加；不同的终端厂商都在竞相开发支持下一代 IP 网络的终端。而在市场化的过程中，必须要对其进行一致性测试。一致性测试不仅是检验终端性能的有效途径，其系统的全面性与准确性也被认为是无线移动通信技术不断完善、通信行业稳步发展的标志。

常见的一致性测试方法有 RRM 一致性测试、射频测试、协议一致性测试。RRM 测试主要是从各方面检验终端的性能，包括 3 个方面：测量精度（判断是否满足未来网络通信协议的规范）、时间精度（从事件被触发到成功发送的时延和通信协议的规定的时延范围是否满足）、功能测试（从行为的正确性和及时性来判断是否满足入网要求）。

5.4　安全测试微服务

安全测试微服务主要负责对场景终端及用户进行验证，依据不同的终端性能以及场景对安全等级的要求，提供不同级别的安全测试服务。另外，除了负责对外的安全测试服务，对内也负责相应的性能指标管理，提示报警信息。

在下一代 IP 网络的自身架构特征来看，虚拟化及可编程技术等都需要更智能化的管控手段和更高效的安全机制。安全性测试是下一代 IP 网络的整体架构的基础。鉴于下一代 IP 网络的开放性和实时性，网络访问不仅仅具有实时可达的特点，同时也是一个动态的变化过程；除了在考虑传统网络安全的基础上，更应该实现对网络访问或入侵的实时检测；通过构建正常模式和异常模式的知识库，采用聚类的算法，自动利用增量熵值对攻击进行检测判断；这种安全监测不仅能监测已知的攻击，对新的攻击也能起到预警作用。

5.5　渗透测试微服务

渗透性测试指通过模拟恶意的方法来评估被测对象的稳健性。在此过程中可发现挖掘系统中存在的漏洞，知晓系统的安全隐患。渗透性测试可用来测试下一代 IP 网络核心主干的顽健性：例如常见的 OSPF RFC2328，模拟发送错误报文，对整体网络架构的容错性进行验证。主干网的总体测评指标应包含：先进性（新型网络是否能在采用高速 IP 线路及交换技术下，与原有网络互联，并成功交换语音、图像等多种信息）、实用性（新型网络是否按照统一标准进行设计，并满足运营所必须达到的要求）、统一性（新型网络是否遵循统一标准，保证开放性）、可扩展性（新型网络是否能支持可扩充性和冗余性，充分留有扩展的余地）、安全可靠性和可管理性能（新型网络是未来公用带宽网络的重要组成部分，应保证整个系统的可管理性

和安全可靠性）。

5.6 规则微服务

规则微服务负责在通过测试模型时，协调多个微服务交互的机制。支持在进行测试时，实现不同微服务之间的自由搭配，并对测试数据进行统一处理。

每一条规则代表了一次完整的测试用例，对应在实际测试场景中不同的测试策略。规则微服务统筹其他几个测试微服务可产生许多实际生产测试中的重要功能，例如与安全微服务相统筹可以实现安全监测的报警测试，与渗透测试微服务相统筹可以构造错误报文，验证稳健性与容错性。

6 结束语

本文设计了一种基于下一代IP网络场景下的微服务架构测试框架。该测试框架能满足在新型网络技术发展的潮流下，业务场景的扩展能力。使用场景描述微服务可以对不同场景按照不同特征进行分类描述，生成具有针对性的、独特的场景测试微服务，有效解决了在未来网络发展趋势下网络传输、通信协议、海量终端等方面异构的特点。此外，基于微服务的测试框架，能大幅提升测试数据的处理能力，实现海量终端、可扩展的测试体系框架。总体测试框架在保障现有场景的测试能力下，综合未来新应用的业务能力的扩展性，充分考虑了下一代IP网络的多变性、灵活性以及复杂性，以便适应传统应用场景向未来新型业务场景演进时，在终端数量、场景特征、服务模式、性能要求上带来的变化，支撑实现创新的建设目标。

参考文献：

[1] 刘韵洁, 黄韬, 姚海鹏, 等. 未来网络的研究进展与展望[M]. 北京: 社会科学文献出版社, 2014: 132-173.
LIU Y J, HUANG T, YAO H P, et al. Research progress and prospects of future networks(2014)[M]. Beijing: Social Science Literature Publishing House, 2014: 132-173.

[2] 裘晓峰. 基于 Web 资源的未来网络安全服务研究[D]. 北京: 北京邮电大学, 2014.
QIU X F. On sercurity services of web resource oriented future network[D]. Beijing: Beijing University of Posts and Telecommunications, 2014.

[3] 蒋林涛. 网络进入 5.0 时代 顶层设计必不可少[EB]. 2017.
JIANG L T. The network enters the 5.0 era, the top design is essential[EB]. 2017.

[4] 李华, 叶新铭. 协议互操作性测试综述[J]. 内蒙古大学学报(自然科学版), 2008, 39(5): 590-596.
LI H, YE X M. A survey on protocol interoperability testing[J]. Journal of Inner Mongolia University (Natural Science Edition), 2008, 39(5): 590-596.

[5] 袁利, 王磊. 星载软件可测试性设计方法[J]. 中国空间科学技术, 2010, 30(4): 31-37.
YUAN L, WANG L. Testable designing methods of Satellite Software[J]. Chinese Space Science and Technology, 2010, 30(4): 31-37.

[6] KRYLOVSKIY A, JAHN M, PATTI E. Designing a smart city internet of things platform with microservice architecture[C]// 3rd International Conference on Future Internet of Things and Cloud (FiCloud 2015), Aug 24-26, 2015, Rome, Italy. Piscataway: IEEE Press, 2015.

[7] SUN L, LI Y, MEMON R A. An open IoT framework based on microservices architecture[J]. China Communications (English), 2017(2).

[8] 张圆冰. 基于微服务架构的自动化测试[J]. 电子技术与软件工程, 2019, 150(4): 135-136.
ZHANG Y B. Automated testing based on microservice architecture[J]. Electronic Technology and Software Engineering, 2019, 150(4): 135-136.

[9] 王万金, 韩成柱. 一种基于场景的装备软件测试用例设计方法研究[J]. 舰船电子工程, 2018, 38(10): 139-142, 164.
WANG W J, HAN C Z. A method study based on the scene of equipment software test designing[J]. Ship Electronic Engineering, 2018, 38(10): 139-142, 164.

[10] 金剑锋, 姚元, 王明晓, 等. 窄带物联网技术在现代城市精细化管理中的研究与应用[J]. 电信科学, 2018, 34(S2): 198-204.
JIN J F, YAO Y, WANG M X, et al. Research and application of narrowband internet of things technology in fine management of modern cities[J]. Telecommunications Science, 2018, 34(S2): 198-204.

[11] 罗莹, 宋利, 解蓉, 等. 全景媒体的系统架构研究综述[J]. 电信科学, 2018, 34(2): 88-98.
LUO Y, SONG L, XIE R, et al. Review of system architecture of omnidirectional media[J]. Telecommunications Science, 2018, 34(2): 88-98.

[作者简介]

徐竟祎（1992− ），女，复旦大学助理工程师，主要从事日志数据的采集与分析、网络性能优化测试及网络 5.0 等前沿研究相关工作。

赵泽宇（1978− ），男，复旦大学高级工程师，主要从事方向网络架构、网络管理与测试等方面的工作，主要研究方向为参与网络 5.0 等。

沈敏虎（1986− ），男，复旦大学工程师，主要从事网络建设与运维、网络性能优化测试、日志数据的采集与分析等相关研究工作。

应奕彬（1986− ），男，复旦大学工程师，主要从事方向运维管理、无线网络建设与测试工作。

周伟强（1982− ），男，复旦大学工程师，主要从事校园一卡通建设与管理、网络建设维护与测试工作，主要研究方向为智慧校园校园卡数据赋能、智慧支付等。

面向计算网络融合的下一代网络架构

姚惠娟，耿亮
（中国移动通信有限公司研究院，北京 100053）

摘　要：以工业互联网为代表的产业互联网的大发展，促进未来网络从以信息传输为核心的信息基础设施，向以融合感知、传输、存储、计算、处理为一体的智能化信息基础设施发生转变 。从应用、网络技术、计算技术等的发展趋势分析，推导出未来数据网络需要从计算、网络和存储融合重新设计网络架构，以满足未来新业务和新场景的需求。

关键词：计算网络融合；ICT 融合；未来网络架构；数据通信网络；IP 网络架构
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019212

Trend of next generation network architecture: computing and networking convergence evolution

YAO Huijuan, GENG Liang
China Mobile Research Institute, Beijing 100053, China

Abstract: The great development of the industrial internet represented by the industrial internet will promote the transformation of the future network from the information infrastructure with information transmission as the core to the intelligent information infrastructure integrating perception, transmission, storage, computing and processing. Based on the development trend of application, network technologies and cloud computing technologies, the future data network needs to redesign the network architecture, which were proposed from the integration of computing, network and storage to meet the needs of new business and new scenarios in the future.

Key words: computing and networking convergence, ICT convergence, future network architecture, data network, IP network architecture

1　引言

随着 5G、人工智能、云计算、产业互联网技术的快速发展，面向不同垂直行业多样化的新型应用不断涌现。与消费类互联网的应用不同，垂直行业应用对网络服务质量有着极高的要求。自动驾驶、远程医疗、工业控制、全息通信等诸多业务对网络的需求千差万别，有的应用例如全

收稿日期：2019−07−15；修回日期：2019−09−11
基金项目：国家重点研发计划基金资助项目（No. 2018YFB180079）
Foundation Item: The National Key Research and Development Program of China (No. 2018YFB180079)

息通信，对于网络的带宽资源需求非常高，人均需要消耗 Tbit/s 级带宽；与 IoT 相关的应用的数据传输呈高频小分组的特征，其带宽消耗量并不大，但是需要持续在线；工业控制领域，为了确保各个模块之间完美配合，需要在正确的时间点执行正确的操作，数据传输要求保证确定性时延。传统尽力而为的数据通信网络在提供更加灵活、定制化的带宽、时延、可靠性等网络指标的 QoS 保证时，面临着巨大的挑战。

目前学术界和产业界呈点状涌现出众多数据通信网络新技术，确定性网络、网络切片、IBN（基于意图的网络）和在网计算等，均聚焦于解决现有网络某一方面的痛点，但尚未形成体系化的，具备跨代能力的未来网络架构总体视图。以国际互联网工程任务组（The Internet Engineering Task Force，IETF）为代表的数通领域标准组织大多采取自下而上的发展模式，网络整体被强行拆分为路由、管理、安全、应用等诸多方面。各方面的技术独立向前发展。由于缺乏整体架构的指导，各方面发展进度参差不齐，彼此之间难以形成合力。这也是 IP/MPLS 主导的数据通信网络后，数据通信网络领域一直没有出现重大技术变革的重要原因之一。面对未来垂直行业互联网应用苛刻的质量保证需求，考虑到云计算、边缘计算以及在网计算等技术正在大力推进的 ICT 一体化网络融合趋势，数据通信网络有必要借鉴移动互联网跨代发展的方式方法，在兼顾现有 IP/MPLS 主导的网络体系的同时，尽快构建整体网络架构统一牵引，加快数据通信网络不同领域新技术的跨代发展。

2　计算网络融合趋势分析

2.1　全行业数字化/智能化给网络和计算提出了巨大挑战

全球已经掀起行业数字化转型的浪潮，数字化是基础、网络化是支撑、智能化是目标。智能化社会的一个典型特征即物理世界和数字世界的深度融合，未来数字世界通过 IoT、AR 等技术提供的传感器、执行器，与真实世界产生互动。网络作为物理世界和数字世界的连接的桥梁实现数据流动。网络连接的一侧是 IoT 的传感器和执行器的 I/O，作为物理世界和数字世界的接口产生海量数据，对网络提出更高带宽、更低时延、更强安全的需求；另一侧是人工智能运算所需要的数据、算力、算法，实现数据价值化。根据思科云指数预测，截至 2021 年，出现终端设备数量将大于 500 亿，每年产生数据达 847 ZB，其中超过 50% 的数据需要在网络边缘侧分析、处理与存储。海量数据的传输需求、分析和存储对传统网络和云计算提出了巨大挑战，使云计算和网络面临“传不畅、算不动、存不下”的局面，所以为了满足大数据传输需求，或者低时延、高安全的业务需求，驱动计算从云端下移到接近数据源的边缘近处理和分析数据。

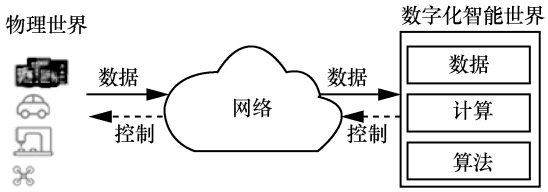


图 1　物理世界和数字化智能世界融合

2.2　5G 网络架构已全面走向服务化

5G 网络从 IT 化、互联网化、极简化、服务化 4 个系统设计理念出发，借鉴业界成熟的 SOA、微服务架构等理念，结合电信网络的现状、特点和发展趋势，对网络架构进行了革新性的设计，定义了全新的基于服务的网络架构（service-based architecture，SBA）作为统一基础架构，如图 2 所示。这种设计有助于网络快速升级、提升资源利用率、加速新能力引入、便于网内和网外的能力开放，使得 5G 系统从架构上全面云化。这意味着从 2G 到 5G，网络已经从烟囱式网络走向服务化网络，为不同垂直行业提供快速响应和灵活部

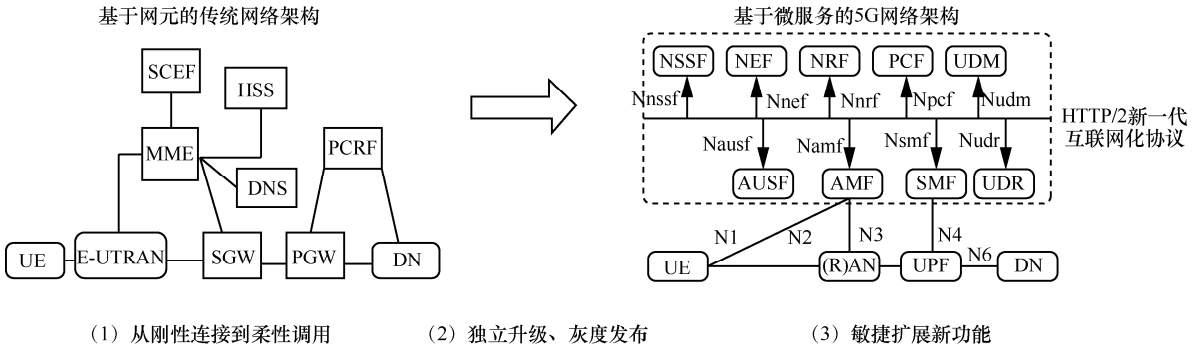


图 2　5G 服务化网络架构

署。尤其是 5G 对 NFV、端到端网络切片及边缘计算的支持，要求网络与计算基础资源协同发展。

2.3　边缘计算助力计算从网络中心走向边缘

边缘计算利用 4G/5G、FTTx、企业专线等多种网络接入方式，在靠近数据源或用户的地方提供计算、存储等基础设施，并为边缘应用提供云服务和 IT 环境服务，助力计算从网络中心走向边缘。过去的 20 年，中国移动通信集团打造了一张卓越的覆盖无线和固定连接的网络基础设施平面，如图 3 所示。NFV 技术的演进发展也促使中国移动开始建设服务于虚拟化网元的电信云设施。面向未来工业互联网、人工智能等新兴业务，运营商需要在端到端的网络平面的基础上，借助边缘计算打造一张面向全连接的算力平面，形成算力的全网覆盖，为垂直行业就近提供智能连接基础设施。在这个新的算力平面中，无处不在的现场级边缘计算为用户提供智能化接入和实时数据处理，实现业务的灵活接入，实现为数据生态的赋能；触手可及的网络侧边缘计算则就近为用户提供丰富的算力，承载人工智能、图像识别和视频渲染等新业务，实现为应用生态的赋能。丰富的网络资源与算力资源将不断地融合互补，为垂直行业业务提供极致的用户体验。

2.4　互联网应用向函数即服务演进

传统基于客户端/服务器客户端与长生存周期的服务端进程通信，服务端处理大部分业务逻辑。随着微服务的发展，传统的客户端/服务器模式被解构，服务器侧的应用解构成功能组件部署在云平台上，由 API 网关统一调度，可以做到按需动态实例化，服务器中的业务逻辑转移到客户侧，客户只需要关心计算功能本身，而无需关心服务器、虚拟机、容器等计算资源，聚焦业务逻辑，从而实现函数即服务。基于此

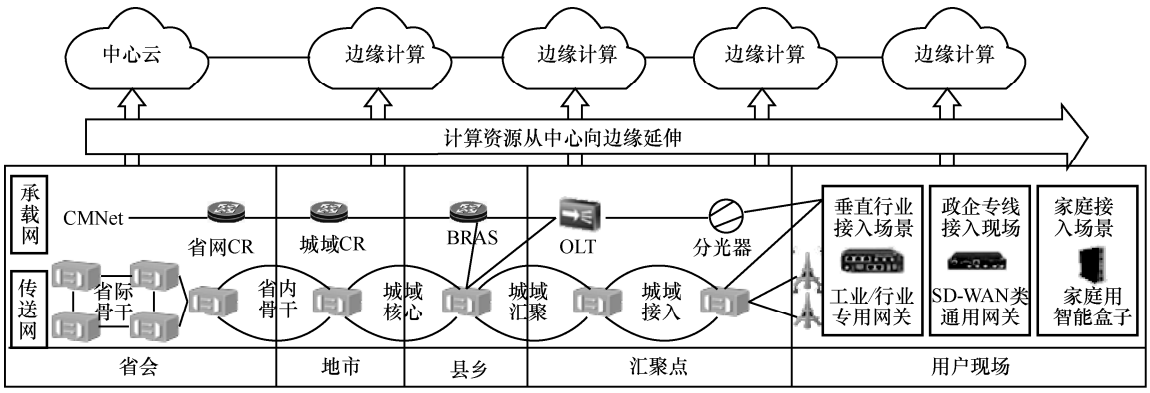


图 3　边缘计算助力计算从网络中心走向边缘

新型微服务应架构，结合网络中遍布的无处不在的计算能力和存储能力，每一个网络节点都可以成为资源的提供者，用户的请求可以通过调用最近的节点资源来满足，不再局限于某一特定节点。下一代网络架构需要支持更加灵活的调度机制和路由机制，支持计算资源节点之间具备互动调度的能力，或者计算任务动态路由的能力。基于函数即服务的互联应用发展架构如图 4 所示。

3　计算网络融合改变网络参考模型

在云计算、边缘计算乃至普适计算的发展大趋势下，从云计算走向边缘计算，再从边缘计算走向泛在算力。随着网络和计算的深度融合，算力由外延向内生演进、由通用向异构持续演进，如图 5 所示。未来社会中会在靠近用户的不同距离遍布许多不同规模的算力，通过全球网络为用户提供各类个性化和智能化服务。

计算资源融入网络使得架构的拓扑假设也发生变化。传统互联网架构的基本拓扑抽象是端到端模型：网络在中间、计算在外围，主机通过网络实现逻辑虚拟的全连接，如图 6 所示。网络计算融合的一体化网络架构改变了当前计算在网络边缘的端到端模型，计算像葵花籽一样嵌在网络

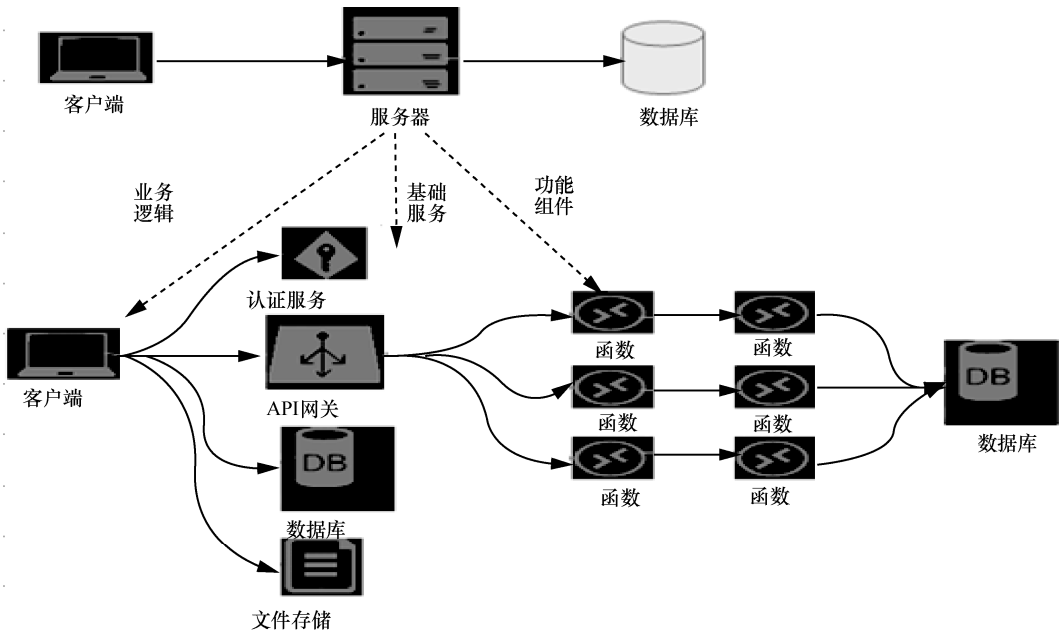


图 4　基于函数即服务的互联应用发展架构

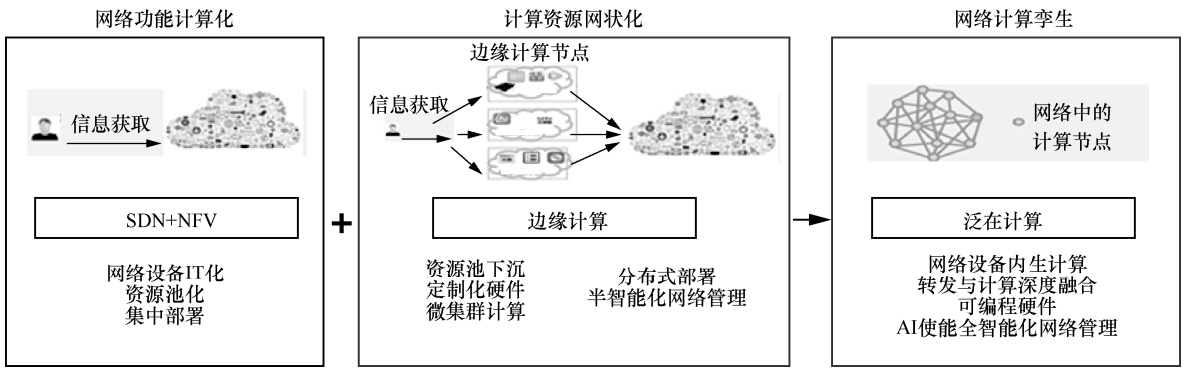
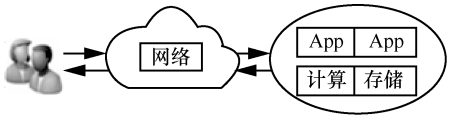
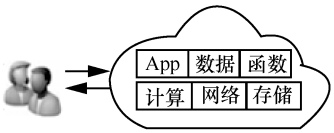


图 5　计算在网络中的演进

中间，是动态、分布式的计算与网络深度融合的网络模型。因此新一代网络架构设计需协同考虑网络和计算融合演进的需求，实现“泛在连接+计算+智能”网络的全局优化、算力的灵活调度、业务的合理分布。



(a) 信息社会：用户使用互联网获取信息



(b) 智能社会：用户使用互联网获取计算结果

图 6　网络参考模型

4　计算网络融合的架构思考

面对新业务新需求，计算网络融合的一体化网络架构如图 7 所示，应基于新型 IP 网络体系，支持网络可编程、函数能力寻址、确定性网络，算力感知的协同调度和控制等功能，向各相关产业提供网络能力、计算能力及数据能力服务，并使其更加有效地满足万物互联、万物智能、万物感知的需求。

（1）算力感知的协同管理与控制

网络可感知算力，并协同调度算力资源和网络资源，实现网络算力的可用、可管和可控。根据当前的网络状况和计算资源，网络依据功能/业务的标识将报文转发到相应的计算节点，实现用户体验最优、计算和网络资源利用率最优。

（2）函数能力寻址

未来网络自支持更加灵活的服务化架构，通过服务器侧的应用解构成“功能组件”实现原子化功能的按需实例化，向函数即服务演进。

（3）网络可编程

未来网络根据不同的应用场景灵活配置网络协议字段以及字段长度等，并支持计算、安全、隐私和确定性时延等不同能力的集成，支持 IoT 大跨度的差分服务。同时可平滑演进，不影响已有业务应用。

（4）确定性网络传输

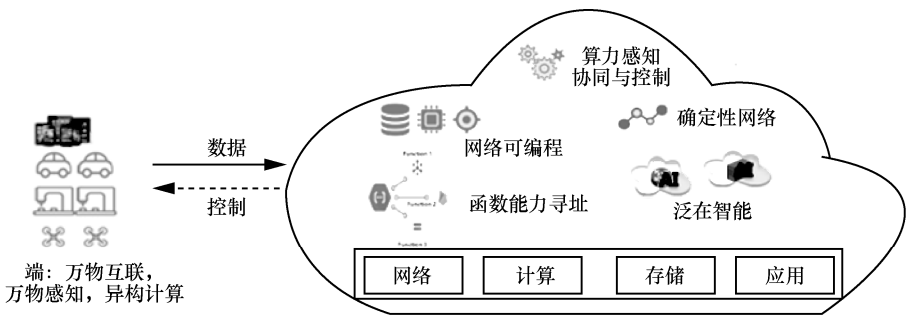
网络中通过引入资源预留、时钟同步、流量排定等机制提供确定性时延、极低分组丢失率以及虚拟化运营管理等确定性服务。

（5）泛在智能

计算网络一体化的基础设施，不但实现网络无所不达，还为人工智能提供了无处不在的算力，实现无所不及的泛在智能。

5　结束语

面对全行业数字化和智能化大发展的浪潮，业务需求和技术创新并行驱动带来了网络架构深



“连接+计算+智能”ICT一体化基础设施

图 7　“连接+计算+智能”计算网络融合的一体化网络架构

刻变革。计算网络深度融合的下一代网络架构应以连接为基础，以计算为依托，以智能化为目的构建计算网络融合的一体化新型网络架构，使得计算和网络高度协同、互为支撑、相互融合，形成“云在网上、网在云中、网随云动”的新型网络模式。

参考文献：

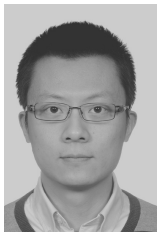
[1] ZILBERMAN N, MOORE A W, CROWVROFT J A. From photons to big-data applications: terminating terabits[J]. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2016, 374(2062): 20140445.

[2] 郑秀丽, 蒋胜, 王闯, 等. 对网络技术跨代发展的思考——网络 5.0[J]. 信息通信技术, 2017(6): 37-44.
ZHENG X L, JIANG S, WANG C, et al. A potential direction of next generation data communication network–network 5.0[J]. Information and Communications Technologies, 2017(6): 37-44.

[3] 网络 5.0 技术白皮书[R]. 2019.
Network 5.0 technology white paper[R]. 2019.

[作者简介]

姚惠娟（1980− ），女，博士，中国移动通信有限公司研究院项目经理，主要研究方向为新一代无线接入网络架构和协议设计、未来互联网安全研究、5G 安全研究、边缘计算及下一代 IP 网络架构等。



耿亮（1987− ），男，博士，中国移动通信有限公司研究院研究室主任，CCSA TC614 网络 5.0 技术标准推进委员会架构组组长，主要研究方向为新型数通网络架构、未来网络协议、产业互联网网络技术以及 OICT 网络融合技术等。

基于云、网、边融合的边缘计算新方案：算力网络

雷波，刘增义，王旭亮，杨明川，陈运清
（中国电信股份有限公司北京研究院，北京 102209）

摘　要：边缘计算已经成为5G时代重要的创新型业务模式，尤其是其低时延特性，被认为是传统方案所不具备的，因此边缘计算能够提供更多的服务能力且具有更为广泛的应用场景。但边缘计算与处于中心位置的云计算之间的算力协同成为新的技术难题，即需要在边缘计算、云计算以及网络之间实现云网协同、云边协同，甚至边边协同，才能实现资源利用的最优化。在研究边缘计算算力分配和调度需求的基础上，提出了基于云、网、边深度融合的算力网络方案，并针对AI类应用给出了一个典型实施系统，该方案能够有效应对未来业务对计算、存储、网络甚至算法资源的多级部署以及在各级节点之间的灵活调度。

关键词：边缘计算；云计算；云网融合；算力网络
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019209

Computing network: a new multi-access edge computing

LEI Bo, LIU Zengyi, WANG Xuliang, YANG Mingchuan, CHEN Yunqing
Beijing Research Institute of China Telecom Co., Ltd., Beijing 102209, China

Abstract: Edge computing has become an important innovative business model in the 5G era, especially its low latency characteristics, which are considered to be unavailable in traditional solutions. Therefore, edge computing can provide more service capabilities and more application scenarios. However, the synergy of computing resources between edge computing and cloud computing has become a new technical problem, so it is necessary to realize cloud network collaboration, cloud edge collaboration and even edge collaboration between edge computing, cloud computing and network, so as to achieve the optimization of resource utilization. A computing network solution based on cloud, network and edge depth fusion was introduced, and a typical system for AI application was proposed, which could effectively cope with the future.

Key words: edge computing, cloud computing, cloud network convergence, computing network

1　引言

随着5G时代的到来，业界在畅想5G将改变社会时，经常会提及边缘计算，并将其视为改变通信信息服务模式的关键创新之一。据IDC发布的《数据时代2025》报告预测，到2025年50%的数据将

收稿日期：2019−07−15；修回日期：2019−08−20
基金项目：国家重点研发计划基金资助项目（No. 2018YFB1800100）
Foundation Item: The National Key Research and Development Program of China (No. 2018YFB1800100)

在网络边缘侧分析、处理与存储，与此同时边缘计算也被认为是 5G 与工业互联网、物联网等的重要结合点，能够推动相关产业带来飞跃性发展。但随着研究和实践的深入，边缘计算的概念已经不限于 5G 领域，扩展到了专线、PON、Wi-Fi、4G 等，即多接入边缘计算（multi-access edge computing）。

但目前业界对边缘计算的定义与内涵并没有形成一致意见，各标准组织或企业分别从自己的角度提出了不同的认识，如 ISO 认为边缘计算是一种将主要业务处理和数据存储放在网络边缘节点的分布式计算形式，ETSI 定义边缘计算是在靠近数据源或用户的地方提供计算、存储等基础设施，并为边缘应用提供云服务和 IT 环境服务，而国内的边缘计算产业联盟（ECC）则定义边缘计算是靠近物或数据源头的网络边缘侧，融合网络、计算、存储、应用核心能力的开发平台等。虽然大家都认可边缘计算是在网络边缘上提供计算服务这个基本观点，但对什么是网络边缘、边缘计算的功能包含什么、如何部署和实现都存在较大争议。邬贺铨院士在 2018 年年底提出了“十问”边缘计算，系统化地总结了边缘计算发展过程中遇到的多个关键重大问题。

在“十问”中，邬院士连续追问了两个关于计算能力部署和调度的问题，即第二问“计算能力是一级设置还是多级设置？”和第三问“计算能力如何在边缘计算和云计算之间优化配置？”。这里直指了一个边缘计算部署过程中最常见的问题，即算力的分配与调度问题。没有人认为边缘计算会完全替代云计算，因此自然存在这个疑问。笔者在推进边缘计算现场试点过程中，被一线生产运营部门问得最多的也是这个问题。

为了解决这个问题，一种方案是采用“云边协同”的方案，即将边缘计算同样划分为 IaaS（基础设施即服务）、PaaS（平台即服务）、SaaS（软件即服务）等多层，然后将 EC-IaaS（边缘计算的基础设施即服务）与云端 IaaS 对接实现对网络、虚拟化资源、安全等的资源协同；EC-PaaS（边缘计算的平台即服务）与云端 PaaS 对接实现数据协同、智能协同、应用管理协同、业务管理协同；EC-SaaS（边缘计算的软件即服务）与云端 SaaS 对接实现服务协同。但在实践过程中，发现这种方案需要边缘计算节点具备复杂的云计算管理平台，但边缘计算所在的边缘机房一般环境受限，能够容纳的服务器资源有限，此方案需要将过多的资源用于管理和协同，其建设与维护成本可观，存在一定的局限性。

因此本文提出了一种新的解决思路，即利用云网融合技术以及 SDN/NFV 等新型网络技术，将边缘计算节点、云计算节点以及含广域网在内的各类网络资源深度融合在一起，减少边缘计算节点的管控复杂度，并通过集中控制或者分布式调度方法与云计算节点的计算和存储资源、广域网的网络资源进行协同，组成新一代信息基础设施，为客户提供包含计算、存储和连接的整体算力服务，并根据业务特性提供灵活、可调度的按需服务。采用这种方案构建新型信息基础设施架构，被称为“算力网络”，其能够根据客户需求，在云、网、边之间按需分配和灵活调度计算资源、存储资源以及网络资源。

2　边缘计算与算力调度需求

边缘计算并不是简单地将服务器放到边缘机房即可，目前普遍认为边缘计算应该具备三大关键指标，即“低时延、大带宽和低成本”，只有这样的方案，才能让客户和平台运营方双赢。因此针对这三大指标逐一进行分析。

首先，针对低时延指标。究竟时延需要多低，各方给出的建议数值不一致，有 2 ms、4 ms、6 ms 或者 10 ms 等。但业界主流认为至少客户流量不需要在广域网上绕行，即无论是 4G/5G 的移动接入方式，还是以 PON 为主的光接入方式，都希望

客户流量能够直接从接入位置就近送入边缘计算节点，无须到运营商原有的部署在汇聚或核心机房的业务接入控制网关处绕行。以光接入为例，理想的接入路径是流量直接从客户站点到OLT再到边缘计算节点，但现有的实践路径是从客户站点到OLT后，先上行到MSE（综合业务网关，也可以是SR（业务路由器），然后再回到OLT进入边缘计算节点，如图1所示。因此在边缘计算提出之时，就考虑采用5GUPF下移的方案来解决相关问题，同样采用光接入方式也需要vBRAS、vCPE、vSR分布式部署等解决方案。

其次，针对大带宽指标，随着5G网络建设的开展，移动接入和固网接入的双吉比特已经成为下一阶段网络服务的标准配置，因此接入带宽并不是困难所在。

最后，对于低成本指标，这里所指低成本不是说边缘计算能够实现无条件的低成本。事实上，单以建设和运营来看，传统的网络边缘机房改造困难很多，比如电力引入、空调改造、承重加固等都存在很多问题，通常需要投入大量的资金进行改造，使得边缘计算节点在单位建设成本上要远高于集中建设的云计算节点。但另一方面，考虑在相同服务质量（如低时延和低抖动）要求下，传统云计算方案需要网络提供高品质的专线传输（如OTN等传输专线），才能按性能指标要求将业务流量送达集中部署的云计算节点，这样算来，高品质传输专线成本与云计算节点成本之和可能高于边缘计算成本，因此有部分观点从这个角度得出了边缘计算成本更低的结论。

综上所述，边缘计算能够提供低时延、大带宽的高品质服务，但由于它的单位算力的成本高于云计算的算力成本，因此在不计入网络连接产

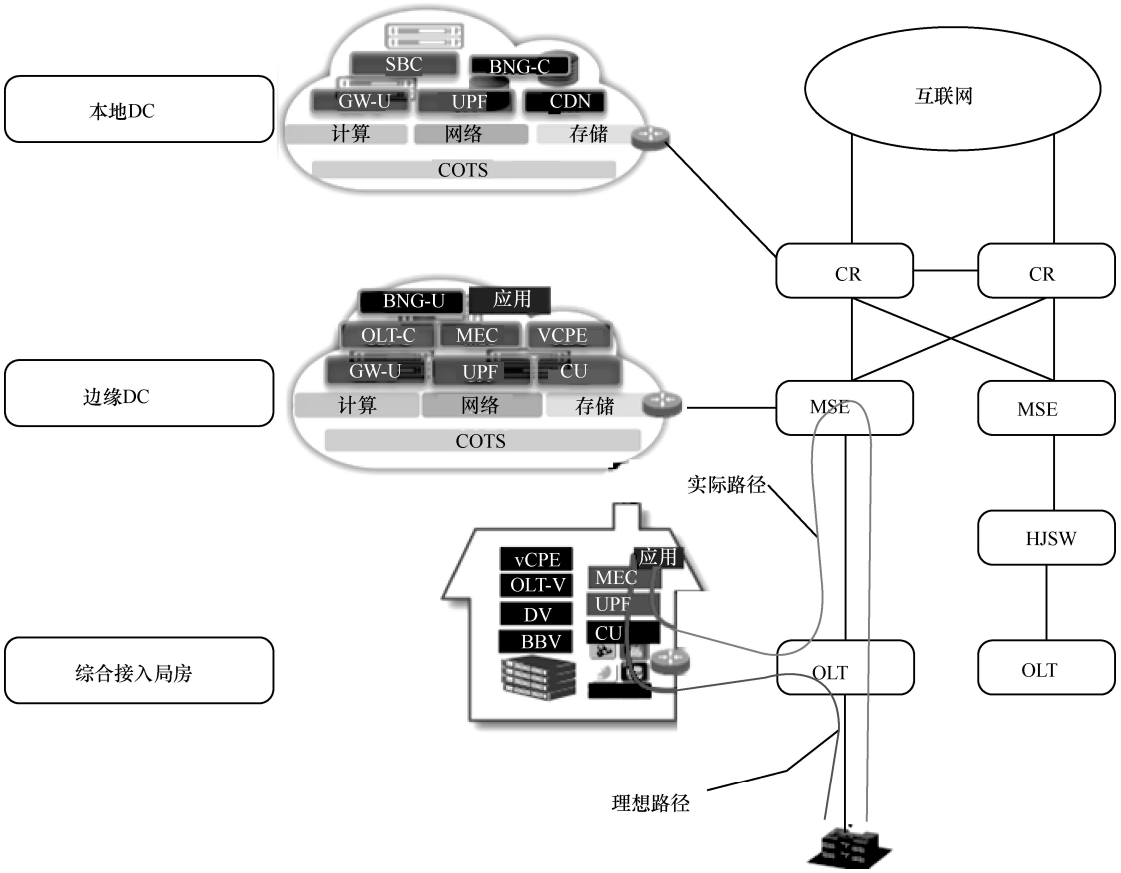


图1　基于传统网络架构的边缘计算路径迂回示例

品价格时，边缘计算产品的定价应该高于云计算产品，所以边缘计算并不适用所有客户，更适用于那些愿意以一定费用来换取其业务所需的大带宽和低时延等高品质服务的客户。因此从业务角度来看，客户有合理安排其应用部署位置的需求，甚至采用混合部署模式，比如将非实时、不重要的运算放在云计算节点，将实时、重要的运算放在边缘计算节点。

以一个典型的 AI 应用为例，如图 2 所示，边缘计算节点负责数据的实时采集、计算和处理以及 AI 推理计算；而位置较远的云计算节点则负责大数据分析挖掘、数据共享，同时进行 AI 算法模型的训练和迭代以及用户个性化功能塑造等非实时工作；并且，云计算节点将迭代升级后的算法模型推送到边缘计算节点，使边缘计算节点更新和升级，完成自主学习闭环。

3　基于云、网、边深度融合的算力网络

从第 2 节分析中可以看出，多级算力部署是边缘计算乃至云计算发展的必然选择，因此在多级算力之间进行合理的算力分配与灵活调度，也就成为了边缘计算实施与部署过程中必不可少的一环。但现有的云计算服务体系中，尤其 IaaS 层面，还停留在让客户自行选择应用部署位置的阶段，没有提供按需分配和调度算力的手段。究其原因，因为现有的云计算体系并未将广域网（如电信运营商的接入网、城域网、骨干网）纳入整体的管控中，而各方所提的云网融合还处在一个相对简单的初级阶段，通常需要建设一个横跨云网和网管的超级协同编排系统，运营难度相对复杂。因此有必要从底层架构开始，重新考虑和设计云、网、边深度融合方案，用以实现算力等基础信息资源的分配与调度，构建“算力网络”，成为一种新的技术发展方向。

目前，业界尚无对算力网络的标准定义，但本文认为算力网络需要满足以下 4 个特征要求。

（1）资源抽象

算力网络需要将计算资源、存储资源、网络资源（尤其是广域范围内的连接资源）以及算法资源等都抽象出来，作为产品的组成部分提供给客户。

（2）业务保证

以业务需求划分服务等级，而不是简单地以地域划分，向客户承诺诸如网络性能、算力大小等业务 SLA，屏蔽底层的差异性（如异构计算、不同类型的网络连接等）。

（3）统一管控

统一管控云计算节点、边缘计算节点、网络资源（含计算节点内部网络和广域网络）等，根据业务需求对算力资源以及相应的网络资源、存储资源等进行统一调度。

（4）弹性调度

实时监测业务流量，动态调整算力资源，完成各类任务高效处理和整合输出，并在满足业务需求的前提下实现资源的弹性伸缩，优化算力分配。

总结起来，算力网络是“一种根据业务需求在云、网、边之间按需分配和灵活调度计算资源、

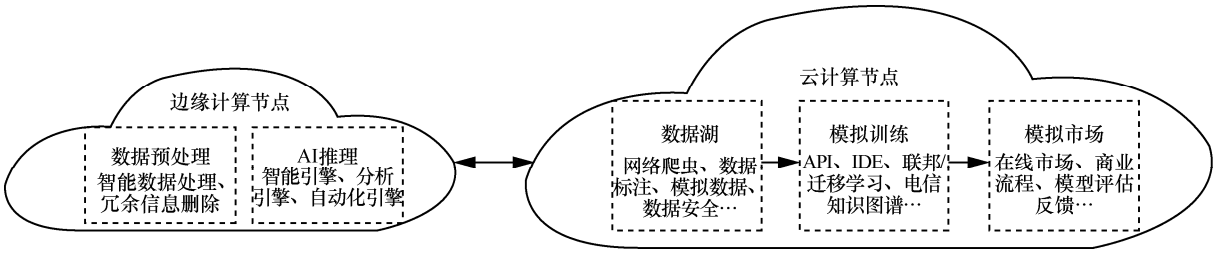


图 2　AI 应用在边缘计算与云计算节点之间的混合部署方案示例

存储资源以及网络资源的新型信息基础设施”。典型的算力网络示意图如图 3 所示。

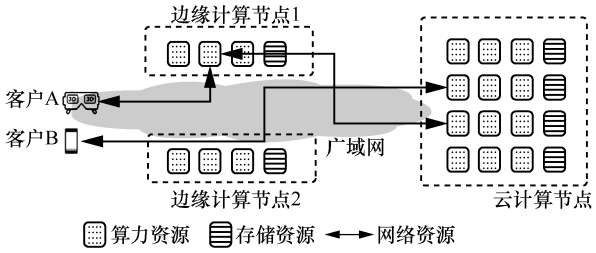


图 3　算力网络示例

在图 3 中，客户 A 需要低时延、大带宽的 VR 应用，因此算力网络在客户 A 的终端与边缘计算节点 1 之间分配低时延、大带宽的网络资源，并在边缘计算节点 1 上分配相应的算力资源和存储资源，另一方面，考虑到该 VR 应用的一些行为记录需要上传至个人中心，但此项记录可以是非实时上传，因此算力网络在边缘计算节点 1 和云计算节点之间分配一条 SLA 相对较低的网络资源。而客户 B 用手机终端查看私人视频，需要加密通道，但考虑到手机终端有一定的缓存能力，因此只需要在客户 B 的终端与云计算节点之间建立不保证 SLA 的加密连接即可。

当客户 A 处于移动状态时，比如从靠近边缘计算节点 1 的位置移动到了靠近边缘计算节点 2 的位置上，这时算力网络通过探测与计算发现由边缘计算节点 2 来提供服务更好，此时通过广域网建立一条从客户 A 到边缘计算节点 2 的通道，相应的应用也从边缘计算节点 1 迁移到边缘计算节点 2，从而继续为客户 A 的 VR 应用提供低时延和大带宽的服务。客户位置发生变化后，重新部署算力资源如图 4 所示。

综上所述，通过采用云计算技术与网络领域技术最新的成果，如 SDN/NFV 等技术，算力网络能够为客户提供云、网、边深度融合的整体解决方案，并能够在网络范围内实现灵活可控的算力及各类资源的调度，既能满足客户的高性能要求，又能有效降低建设与维护成本，提升整网运营效率。

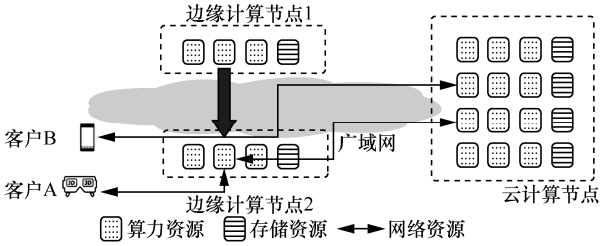


图 4　客户位置发生变化后，重新部署算力资源示例

4　面向 AI 应用需求的算力网络实践

算力网络中的算力调度实现可以有多种方案，比如通过集中管控平台实现统一调度，也可以通过分布式路由协议来实现，如在 BBF 计划立项的城域算力网络项目。因此为了进一步验证算力网络的能力，在前期研究的基础上组建了基于 SDN、NFV、AI、云计算等新型技术的试验环境，并且结合 AI 赋能平台，为 AI 应用需求提供灵活的算力调度系统，这套系统被取名为“AI 算力网络”，其架构如图 5 所示。

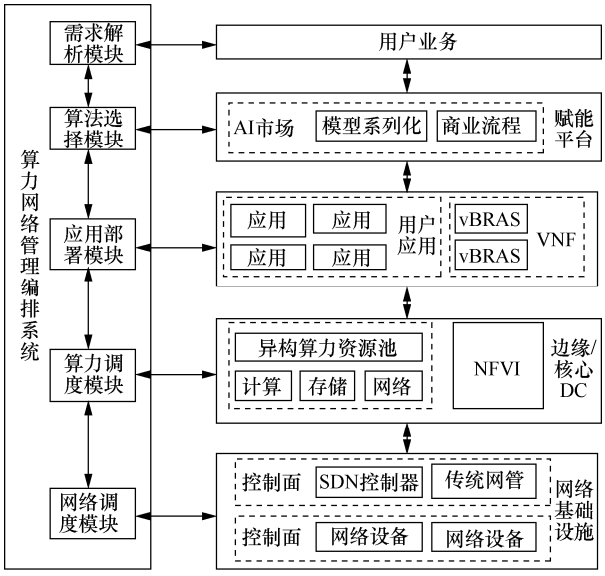


图 5　AI 算力网络框架

系统架构中包含的主要部分如下。

（1）算力网络管理编排系统

算力网络的资源管理和调度系统，根据业务需求对算力资源进行弹性调度，在满足业务实时需求的同时，提高算力利用率。

（2）赋能平台

为用户的 AI 业务提供基于平台的 AI 服务。

（3）边缘/核心 DC

包含算力资源基础设施和 NFV 基础设施，其中，用户应用部署在异构算力资源池之上，vBRAS、vCPE 等虚拟网关部署在 NFVI 之上。

（4）网络基础设施

连接用户、边缘云、核心云的网络基础设施，包括控制面的 SDN 控制器、传统网管以及转发面的网络设备。

算力网络管理编排系统的主要模块功能如下。

（1）需求解析模块

分析用户业务需求，根据不同场景将用户业务需求转化为算力资源需求，根据算力需求划分业务等级，以确定业务的部署位置、资源等信息。

（2）算法选择模块

由用户指定或根据需求解析模块的结果，在 AI 赋能平台中选择用户业务的 AI 服务，确定用户业务部署的规格。

（3）应用部署模块

将 AI 赋能平台中的服务部署到指定的节点上。

（4）算力调度模块

管理核心云和边缘云的算力资源，根据业务需求为用户分配相应的计算、存储、网络资源，并根据策略对业务部署位置、业务算力进行弹性调整。

（5）网络调度模块

管理用户、边缘云、核心云的网络，在确定用户业务部署位置后，联合算力调度模块将业务网关下沉到用户业务同一位置，将业务流量路由到处理节点。

客户可以分级提出不同的需求，AI 算力网络能够自动分析需求，并分配合适的基础资源。针对 AI 类应用，客户需求可以细分为以下 4 个指标。

（1）业务需求

用户 AI 业务所要达到的实际效果，如处理时延、数据规模等。

（2）算法需求

针对同种 AI 业务具有多种处理算法，侧重点不同，用户可以指定使用何种 AI 算法和模型。

（3）算力需求

部署用户业务所需要的算力资源，用户可以指定需要的算力资源规模。

（4）网络需求

用户业务接入处理节点的网络需求，用户可以指定接入的网络节点。

针对用户提出的不同场景，AI 算力网络都有对应的处理流程，具体可以分类以下 4 类场景。

（1）第一类场景，客户给出了 4 类需求的指标，AI 算力网络将根据客户输入，直接分配相应的基础资源并建立连接，如图 6 所示。

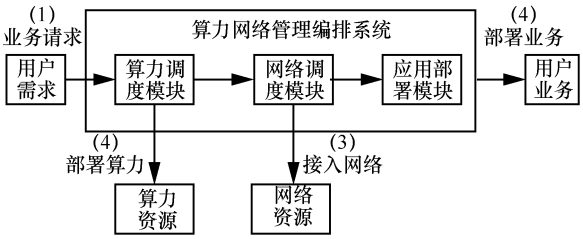


图 6　场景一的处理流程

（2）第二类场景，客户根据经验选择算法并明确算力资源需求，但不了解网络资源需求，因此 AI 算力网络将根据用户需求自动解析用户业务的网络接入点，并进行业务部署，如图 7 所示。

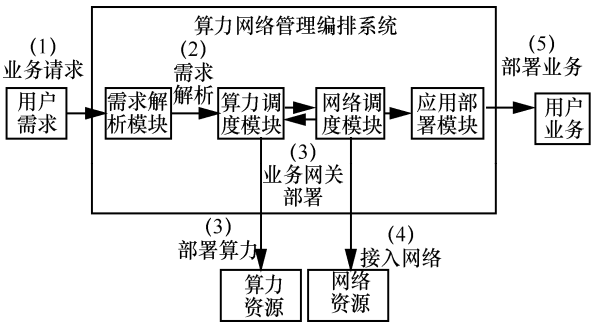


图 7　场景二的处理流程

（3）第三类场景，客户选择了算法，但不清楚所需的算力和网络的资源需求，AI 算力网络将

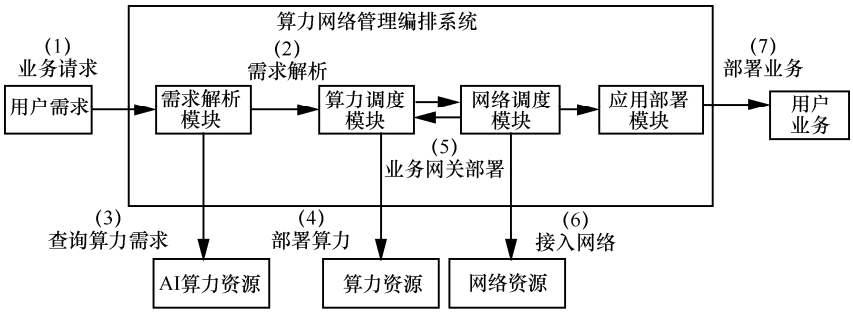


图 8　场景三的处理流程

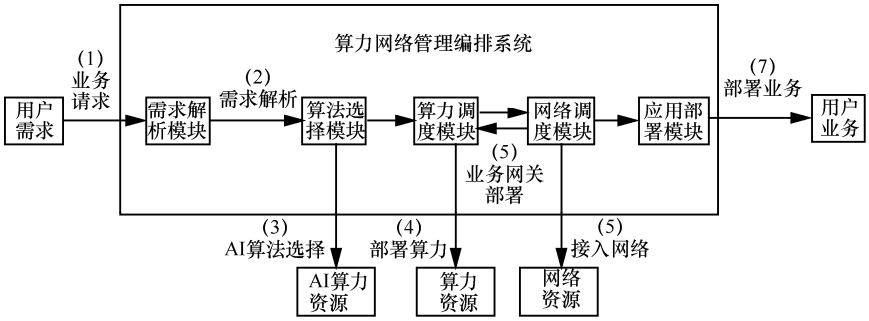


图 9　场景四的处理流程

根据 AI 赋能平台中记录的算法所需的算力资源，为用户进行算力分配和网络调度，如图 8 所示。

（4）第四类场景，客户只是提出了 AI 应用需求，因此 AI 算力网络将自动选择最匹配的算法，并分配对应的算力资源和网络资源，如图 9 所示。

作为实验验证，在实验室模拟了针对人脸识别业务的算力调度。实验室搭建了两套环境不同的云平台，可以分别用于处理人脸识别业务，识别员工的身份、电话等信息，以图片的形式输出并标记处理时延，用户业务要求处理时延在 1 s 以内，实验拓扑和识别结果如图 10 和图 11 所示。

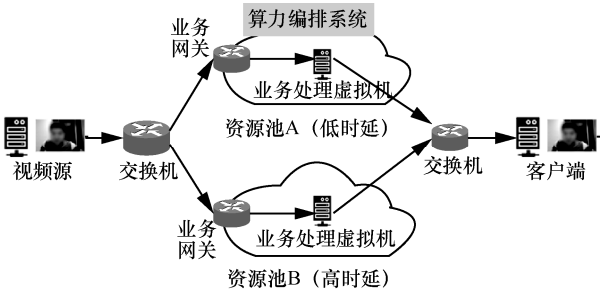


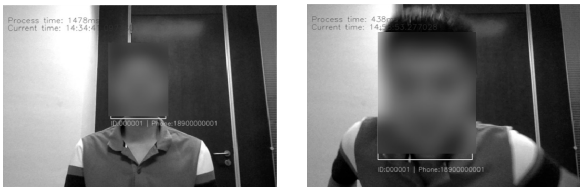
图 10　实验拓扑

图 11　实验结果说明

当用户业务处于闲时，用户业务将部署到资源池 B 进行处理，在满足业务需求的同时，节省高性能资源池 A 的资源；而当用户处于忙时，资源池 B 的处理时延将到达 1 s 以上，难以满足用户需求，此时需要将用户业务调度到资源池 A 进行处理，利用 NFVO 在资源池 A 进行业务网关和处理虚拟机的部署。

由于资源池 A 可以提供更低的时延和更多的算力，满足忙时的业务需求，将时延重新降低到 1 s

以内，调度前和调度后的结果对比如图 12 所示。



（a）调度前　　（b）调度后

图 12　调度前和调度后的结果对比

5　结束语

边缘计算作为 5G 时代最重要的创新场景，能够为客户提供低时延、大带宽等多种业务保障，但随着研究和部署的深入，边缘计算与云计算、网络（尤其是广域网）之间的协同成为新的研究点。本文针对在多级计算节点之间按需部署与灵活调度算力的需求，提出了一种基于云、网、边深度融合的方案，通过构建算力网络，满足不同类型业务的需求。

参考文献：

[1] ETSI ISG MEC. Multi-access edge computing (MEC) framework and reference architecture[R]. 2016.

[2] 边缘计算产业联盟，工业互联网产业联盟(AII). 边缘计算参考架构 3.0[R]. 2016.
ECC, AII. Multi-access edge computing(MEC) reference architecture 3.0[R]. 2016.

[3] SDN/NFV 产业联盟. MEC 行业应用白皮书[R]. 2018.
Alliance of SDN/NFV Industry. White paper on industry applications of MEC[R]. 2018.

[4] 阿里云计算有限公司, 中国电子技术标准化研究院. 边缘云计算技术及标准化白皮书[R]. 2018.
Alibaba Cloud Computing Co., Ltd,. CESI. White paper on technologies and standardization of edge cloud computing[R]. 2018

[5] 网络 5.0 产业和技术创新联盟. 网络 5.0 技术白皮书(2019)[R]. 2019.
N5A. White paper on network 5.0 technologies(2019)[R]. 2019.

[6] 百度, 中国联合网络通信集团有限公司, 中国电信集团有限公司, 等. AI 边缘计算技术白皮书(2018-2019)[R]. 2019.
Baidu, CUCC, CTCC, et al. AI-oriented edge computing technical white paper[R]. 2019.

[7] 云计算开源产业联盟. 云网融合发展白皮书(2019 年)[R]. 2019.
OSCAR. White paper on development of integration of cloud and network(2019)[R]. 2019.

[8] GAO Y, GUAN H B H, QI Z Q, et al. Service level agreement based energy-efficient resource management in cloud data centers[J]. Computers and Electrical Engineering, 2013: 1621-1633.

[9] 李林哲，周佩雷，程鹏，等. 边缘计算的架构、挑战与应用[J]. 大数据, 2019(2): 3-16.
LI L Z, ZHOU P L, CHENG P, et al. Architecture, challenges and applications of edge computing[J]. Big Data Research, 2019(2): 3-16.

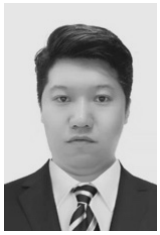
[10] 施巍松, 张星洲, 王一帆, 等. 边缘计算:现状与展望[J]. 计算机研究与发展, 2019, 56(1): 69-89.
SHI W S, ZHANG X Z, WANG Y F, et al. Edge computing: state-of-the-art and future directions[J]. Journal of Computer Research and Development, 2019, 56(1): 69-89.

[11] 张开元，桂小林，任德旺，等. 移动边缘网络中计算迁移与内容缓存研究综述[J]. 软件学报, 2019, 30(8): 2491-2516.
ZHANG K Y, GUI X L, REN D W, et al. Survey of computation offloading and edge caching in mobile edge Networks[J]. Journal of Software, 2019, 30(8): 2491-2516.

[作者简介]

雷波（1980− ），男，中国电信股份有限公司北京研究院新兴信息技术研究所 IP 与未来网络研究中心主任、高级工程师，CCSA“网络 5.0 技术标准推进委员会”管理与运营组组长，主要研究方向为未来网络架构、新型 IP 网络技术等。

刘增义（1992− ），男，中国电信股份有限公司北京研究院新兴信息技术研究所 IP 与未来网络研究中心工程师，主要研究方向为网络功能虚拟化、未来网络等。



王旭亮（1986− ），男，中国电信股份有限公司北京研究院新兴信息技术研究所 IP 与未来网络研究中心工程师，主要研究方向为网络功能虚拟化、未来网络等。

杨明川（1973− ），男，中国电信股份有限公司北京研究院副院长，主要研究方向为云计算、大数据和人工智能、区块链的研究与融合技术等。

陈运清（1964− ），男，中国电信股份有限公司北京研究院院长、教授级高级工程师，CCSA“网络 5.0 技术标准推进委员会”副主席，主要研究方向为未来网络架构、新型 IP 网络技术等。

温敏网络的关键能力和架构体系

孙嘉琪[1]，杨广铭[1]，党娟娜[2]，刘文杰[2]
（1. 中国电信股份有限公司智能网络与终端研究院，广东 广州 510630;
2. 华为技术有限公司，北京 100095）

摘　要：温敏网络给网络安装了高精度的智能传感器，以快速感知网络的质量，挖掘网络的最大能量，能满足业务 SLA 保障、网络更高可靠和最大通量。温敏网络提出的 nTouch、xRecognition 和 iX 关键技术能力，能够很好地应对低时延、高带宽、易流量微突发的未来网络，可作为复杂多变网络的基础能力。

关键词：温敏网络；网络测量；分布式网络优化；集中式网络优化
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019216

Critical capabilities and architecture of the iCAN

SUN Jiaqi[1], YANG Guangming[1], DANG Juanna[2], LIU Wenjie[2]
1. Research Institute of Intelligent Network and Terminal of China Telecom Co., Ltd., Guangzhou 510630, China
2. Huawei Technologies Co., Ltd., Beijing 100095, China

Abstract: The intelligent capability-aware network (iCAN) has installed intelligent sensors with high accuracy to quickly perceive the quality of the network, tap the maximum energy of the network, and meet the business SLA guarantee network reliability and maximum throughput. The key technology capabilities including nTouch, xRecognition and iX in iCAN were proposed, which could cope well with future network with low latency, high bandwidth and traffic micro-burst, and could be used as the basic capability of complex and variable network.

Key words: iCAN, network measurement, distributed network optimization, centralized network optimization

1　引言

1.1　新的业务挑战

5G 的引入[1-2]促进传统业务向消费者生活渗透，为车联网、对战类游戏、VR 和全息通信等业务带来很大的发展空间。新型业务也对未来网络提出如下新的要求。

（1）毫秒级低时延

车联网[2-4]和对战类游戏属于典型的毫秒级时延业务。车联网自动驾驶场景中两辆车通过网络通信要求 IP 承载网 0.5～1 ms，对战类游戏要求 IP 承载网 5～15 ms。

（2）大带宽

VR 和全息通信属于典型大带宽业务，例如一个

收稿日期：2019−08−10；修回日期：2019−09−04
基金项目：国家重点研发计划基金资助项目（No. 2018YFB180079）
Foundation Item: The National Key Research and Development Program of China (No. 2018YFB180079)

VR 业务带宽可高达 4.2 Gbit/s、一个手机屏幕大小的全息影像带宽可高达 12.6 Gbit/s。

（3）流量微突发

新型业务最典型的特征是交互式业务，易流量微突发，即持续时间很短的剧烈突发流量，如在韩国的某类 CloudVR 业务属于脉冲式流量模型，在毫秒级时间之内 50 Mbit/s 的平均视频流的脉冲峰值可达 750 Mbit/s。根据调研，现网存量业务越接近末端，其流量微突发现象越严重。微突发流量的特点为不确定性和公共性。不确定性指措不及防地发生，公共性指可能危害公共的传输业务。所以，交互类新型业务叠加和规模达到一定的量后，未来网络的突发问题会越来越严重，且将带来网络拥塞引发额外时延并干扰其他公共业务。

综上所述，新型业务的低时延、高带宽和流量微突发的特征，对未来的网络产生新的挑战。

1.2　传统网络的技术问题

众所周知，传统的 IP 承载网的特征是根据目的 IP 地址 best effort（尽力而为）地转发，特点是灵活、易于部署，所以在网络发展初期 IP 网络得以快速普及。但是 IP 网络的尽力而为特征不感知网络流量大小，在收敛型架构的网络中易局部拥塞。而今大部分运营商的网络平均利用率为 30%～40%[5]，目的是通过低负载来缓解局部拥塞的问题，以确保用户业务体验。

平均利用率为 30%～40%的网络面对低时延、高带宽、流量微突发的未来网络时也显得力不从心，按照传统思路只能通过提升 CAPEX 而进一步地降低网络平均利用率。显然，这是一种被动防守的高成品的举措。

所以，温敏网络的概念来临了。

2　温敏网络理念

温敏网络，“温”指网络质量，“敏”指反应速度。温敏网络相当于给网络安装了高精度的智能传感器快速地感知网络质量，挖掘网络的最大能量，以满足业务质量保证、网络高可靠和最大通量。

（1）业务质量保证。网络拥塞导致业务传输时延增大甚至分组丢失，有了温敏网络以后就可以解除网络拥塞，降低因为局部拥塞传输而出现的路径额外时延。

（2）网络高可靠。网络故障引发的网络分组丢失可以在最精准的时间内被捕获，触发流量可用的备份路径切换，极大地提升了网络的可靠性。

（3）最大通量。网络最大程度地分担负载，网络的通量自然而然提升。

温敏网络是在全网资源允许的情况下可开启的关键能力。

3　温敏网络关键能力

温敏网络包含 nTouch、xRecognition 和 iX 3 个关键能力单元，如图 1 所示。

3.1　nTouch

高精度的智能传感器取名为 Network Touch（nTouch）。nTouch 包含毫秒级精度的网络质量测量和网络拥塞评估两个关键能力。

3.1.1　毫秒级精度的网络质量测量

传统的 OWAMP[6]和 TWAMP[7]测量技术精度和效率上已经满足不了新型业务要求。

（1）分钟级检测。设备工作原理已经决定其大于分钟级的精度，原因是设备控制面发起的测量分组，通过设备数据面进行测量后再由设备数据面上送给设备控制面进行测量数据收集。

（2）统计流量速率为测量周期内的流量平均速率和流量峰值平均速率。当初这么设计，主要结合当时的设备能力，尽可能减少网络测量对设备的资源消耗。到后面，考虑平均流量满足不了对网络拥塞判断的需求，例如提出 5 min 的检测结果在携带流量平均速率的同时必须携带秒级峰值。

随着设备能力的提升和新业务的逐步发展，传统分钟级的检测技术成为检测精度不够的硬

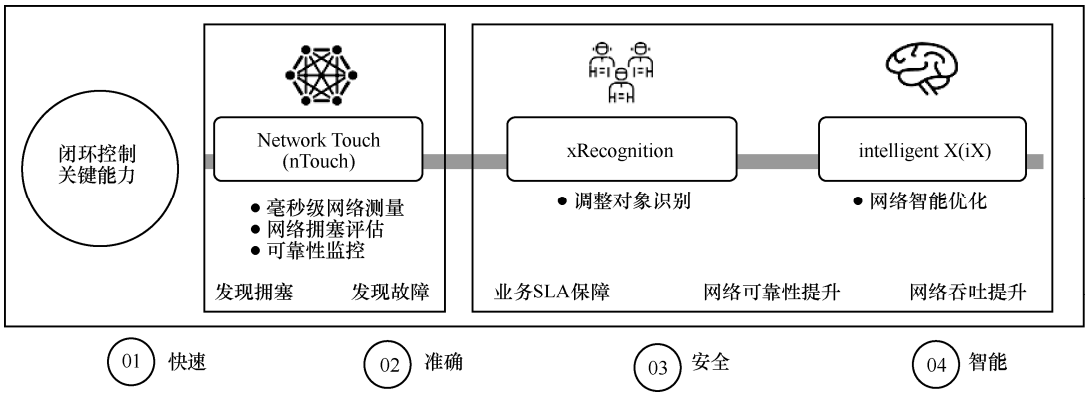


图 1　温敏网络关键能力

伤，毫秒级的网络性能检测也被提上日程。

温敏网络 nTouch 能力单元的网络质量测量技术精度为毫秒级，要求测量发起、测量数据收集都在设备数据面完成。

目前，网络测量已经是一个热门话题，包含带外和 In-situ 带内测量两个技术分支。传统的 OWAMP[4]和 TWAMP[5]属于典型带外技术，其原理是生成测量报文插入业务流中进行测量；iFit 和 iOAM 属于 In-situ 带内测量，其原理是给业务报文中插入测量报文字段，俗称“染色”。

（1）带内技术的特点是轻便，新型的带内测量技术可要求设备支持 3.3 ms 发一次测量分组来提升测量精度，不要求时间同步仅做到频率同步即可。设备硬件在飞速发展，已有能力支撑该诉求。

（2）In-situ 带内测量可微观到具体业务流的质量，例如一个明确的 OTT 业务。但其特点是要测量业务流的每个报文。如果进行业务粒度的故障定位，该方式显然是最合适的。

所以不同的测量对象可以采用不同的测量方法。路径属于基本的传输通道，需要全网或者部分网络开启测量和测量评估，可以采用新型的带外测量技术；具体的业务流监控可采用 In-situ 测量方法，不必要全网开启以降低对设备和带宽资源的消耗，可按需部署，目的是提升业务故障的定界定位效率。

3.1.2　网络拥塞评估

收集到网络数据以后，温敏网络 nTouch 能力单元就可以进行网络拥塞度评估。

传统网络拥塞判断是通过检测链路利用率和分组丢失率两个指标来衡量的。例如链路利用率超过 80%，则认为链路拥塞；如果网络持续拥塞导致网络缓存不够则引发分组丢失，则链路拥塞可以被判断为变得更加严重。在传统方案中，设备上收集的网络性能数据会直接上送到一个外部系统，由外部系统做拥塞评估。如果发现某些链路拥塞，外部系统则通知运营商进行网络扩容。传统方案最大的问题是检测对象单一、缺乏 E2E 视图。

温敏网络 nTouch 的拥塞评估分为路径、链路等层级。路径最接近业务意图，链路次之，不仅限于本文所描述的对象。先讲路径和链路两个层级。

（1）路径拥塞

路径是一个 E2E 概念，以边缘头节点为始、边缘尾节点为终，中间可以跨域多跳网络设备。一对确定的边缘头节点为始和边缘尾节点之间可以有多组负载分担路径。

如果路径拥塞，其时延会增大，增大到途经网络节点缓冲（网络 buffer）不够用时就开始出现分组丢失。

（2）链路拥塞

网络链路属一跳的概念，属本节点和邻居节点之间。链路相比路径粒度更粗，如一条链路可以承载多条路径。

如果链路拥塞，则为带宽利用率超过阈值甚至分组丢失。

nTouch 能力应用在多条负载分担路径场景时，首先是要具备多条负载分担路径的视图，例如 Group*n*{LSP1, LSP2, LSP3, …, LSP*n*}。这一组路径要具备质量的可比性，则就需要在同一个时间窗发起和完成测量。

所以不管最终选择带外还是 In-situ 带内测量方法，都需要考虑多负载分担路径的场景，协议上要具备其扩展能力。

3.2　xRecognition

使能该能力单元有一个前提，就是网络是负载分担的。

nTouch 执行完毕以后，在评估出来网络有局部拥塞时，如果网络非负载分担模式，则 nTouch 直接通知运营商进行扩容；如果网络是负载分案模式，则 nTouch 可以联动 xRecognition 和 iX 能力单元进行智能负载分担。

那么本文继续讨论 xRecognition 能力单元。x 代表一个事物对象，它可能是路径，也可能是流，也可能是本文档未包含的其他内容。路径和流的具体情况如下。

（1）如果路径发生拥塞，业务识别 path Recognition（pRecognition）能力单元会从拥塞路径中挑选出合适的业务流，由 iX 进行流填充目的调整到其他轻载路径来解除拥塞。

（2）链路拥塞时，业务识别 flow Recognition（fRecognition）能力单元会从拥塞链路中挑选需要调整的路径，由 iX 进行重新算路绕行拥塞链路来解除拥塞。

xRecognition 相当于一个小型数据库，缓存了毫秒级精度的路径或者业务流统计信息。如果还需要在现有基础上继续提高精度，可基于 xRecognition 的数据进行路径或者流量的趋势预测。这个趋势预测不同于外部系统预测，主要体现在毫秒级的时间精度上。则 iX 可以根据算法和预测参数综合进行评估。

3.3　iX

iX（intelligent X）关键能力是智能算法。

传统方案已经具备设备级和路径级负载分担能力，但是负载分担算法是静态的，进一步地说，即无法根据流量的实际情况灵活地进行调整。流量变化越大，传统方案限制越大。所以，网络局部拥塞的解除依然是一个热门话题。

基于设备级和路径级负载分担能力，温敏网络 iX 提供了智能负载分担算法，即根据流量的变化调整负载分担算法，最终达到网络的负载均衡。

iX 包含 iFlow 和 iPath 两大组件。

（1）iFlow（intelligent Flow）

路径拥塞时，iFlow 从 fRecognition 中获取合适的流填充进负载分担轻载路径中。

理想状态，iFlow 启动调整之后不会引发轻载路径拥塞。轻载路径不拥塞有两个层面的衡量指标，第一指标是路径时延保持路径最小时延，第二个相对宽松的指标即分组不丢失。其次，轻载路径的剩余带宽[8]也是一个关键的研究点，后续将有专门的文章进行说明和论证。

关键问题是看被调整的流量属于什么流量。如果其要求比较苛刻，则路径时延需要一直保持路径最小时延；如果其质量要求比较宽松，则路径被要求不分组丢失即可。

（2）iPath（intelligent Path）

链路拥塞时，iPath 从 Recognition 获取合适的路径进行重新计算以绕行拥塞链路。链路拥塞基本的判定办法，已在网络中商用，本文不在此赘述。

4　温敏网络架构

温敏网络的架构设计主要考虑如何对其关键能力单元进行排兵布阵，以满足不同场景的需求。限定条件是在有限的软硬件资源条件下，如何最大程度地满足业务上的高质量诉求。

温敏网络架构分为集中式和分布式。

（1）分布式。设备从本地收集到网络质量信息可直接就地决策，特点是更快、更准。“快”体现在处理时效上，“准”体现在网络拥塞的有效机

会窗内解决问题，所以快和准是相应而生的。

（2）集中式。集中式控制器可以从网络每个节点收集到全网信息集中决策，擅长做全局资源调整。它的特点是慢但更全面。“慢”同样体现在处理时效上，“更全面”则是因为它掌握着全网信息。

要选择分布式和集中式，首先看各自拥有的完整业务视图范围。分布式设备的头节点具备多个负载分担路径的完整视图；集中式控制器拥有一个链路承载的多个不相关（例如不同源不同宿）的路径视图。

接着，要解决的具体问题如下。

（1）如果要解决网络微突发引发的网络拥塞问题，分布式架构的处理时效可以满足。而且路径视图更接近业务视图，对于业务流量的微突发更易感知和处理。

（2）如果要解决链路拥塞，那么拥有全网路径视图的集中式处理是最合适的。

4.1 基于温敏网络的分布式网络控制模型的建立

基于温敏网络的分布式网络控制模型的建立如图 2 所示。例如边缘节点 A 和边缘节点 B 之间存在多条负载分担路径，要更好地服务于业务，就必须更高程度地负载均衡。那么，该模型关注网络负载分担多路径的质量。

对于一条路径而言：

- 如果只承载一种业务，那么该网络路径的质量就代表所承载业务的质量；
- 如果承载多种优先级业务，那么该路径就拥有不同对应级别的质量。一般情况下，路径质量代表最低优先级的业务质量，因为高优先级业务可以抢占低优先级业务带宽，所以路径拥塞最直接的受害者就是低优先级业务。如果网络规划跟不上业务发展，则高优先业务直接被网络拥塞影响体验的情况也是存在的。理想状态就是，一个路径会存在针对不同优先级的业务质量。

对于一组负载分担路径而言，nTouch 在对等条件下对多路径的质量测量和评估，fRecognition 提供流信息识别方法，iFlow 从拥塞路径中挑选合适的流负载分担到其他轻载路径上。

如果要应对微突发的业务，这些能力都需要设备数据面完成。

该模型不仅限于第 4.1 节所述业务。

4.2 基于温敏网络的集中式网络控制模型的建立

基于温敏网络的集中式网络控制模型的建立如图 3 所示。例如节点 B 有条链路拥塞，实际上有多条路径经过该链路。在场景下，该模型需要关注网络链路质量。对于一条网络链路而言会承载多个路径的业务。多个路径的头节点和目的节

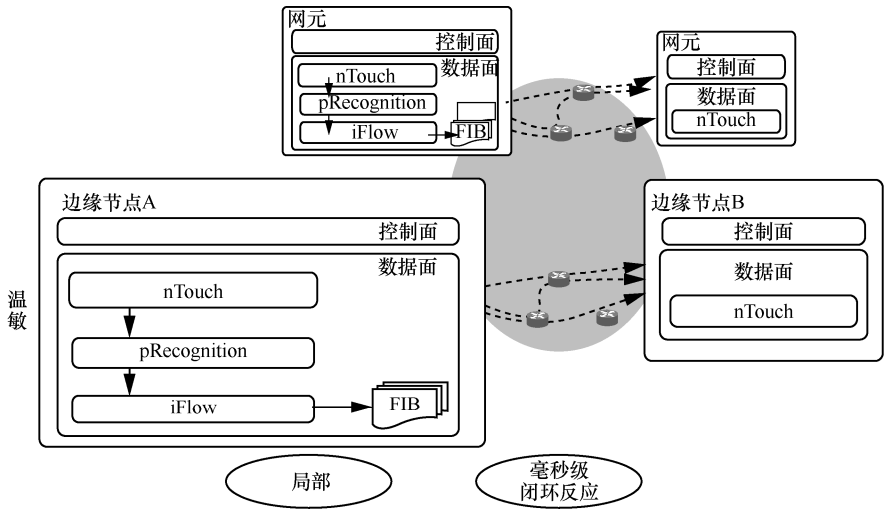


图 2 基于温敏网络的分布式网络控制模型的建立

点都是不一样的。所以要解除某条链路的拥塞，则需要有一个能容纳相关多条路径的视图的集中式能力单元去调整。

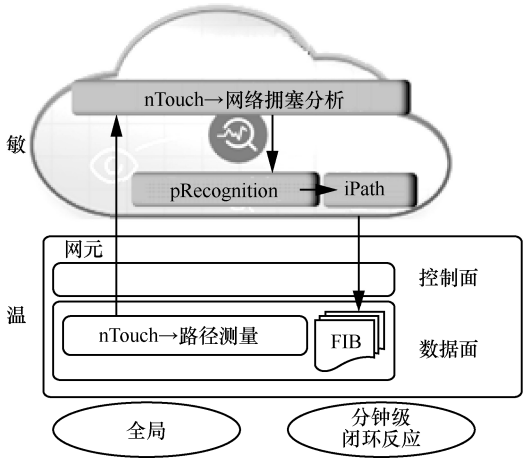


图 3　基于温敏网络的集中式网络控制模型的建立

sPath 负责识别出拥塞链路的多条路径，iPath 挑选合适的路径绕行拥塞链路。所以 pRecognition 和 iPath 要求部署在集中式能力单元 SDN 控制器。

该模型不仅限于第 4.2 节所述业务。

4.3　温敏网络的集中式和分布式混合部署

如果边缘节点设备在局部网络 E2E 负载分担调整时发现资源不够，那么它可以请求 SDN 控制器根据全网资源情况重新为某条路径计算路径。所以该场景中分布式和集中式混合部署效果更佳。

最终的方案效果需要在功能和效率上做取舍，以获得一个最佳平衡点。

5　结束语

本文提出的新型温敏网络架构，是一套完整的控制体系，增强了网络的弹性和稳健性。它可以让业务流根据网络的实时变化自适应以进入可用的路径中。其在传统 MPLS 场景分段路由、分段路由 IPv6 场景和未来新的隧道技术都具备广泛的通用性。笔者的后续计划是针对不同应用场景给出温敏网络的应用思考。

参考文献：

[1] LI S, XU L D, ZHAO S. 5G internet of things: a survey[J]. Journal of Industrial Information Integration, 2018(10): 1-9.

[2] 李子姝,谢人超,孙礼,等.移动边缘计算综述[J].电信科学,2018,34(1):87-101.
LI Z S, XIE R C, SUN L, et al. A survey of mobile edge computing[J]. Telecommunications Science, 2018,34(1):87-101.

[3] 3GPP. Study on LTE-based V2X services: TR36.885 [R]. 2016.

[4] 陈山枝,胡金玲,时岩,等.LTE-V2X 车联网技术、标准与应用[J].电信科学,2018,34(4):1-11.
CHEN S Z, HU J L, SHI Y, et al. Technologies, standards and applications of LTE-V2X for vehicular networks[J]. Telecommunications Science, 2018,34(4):1-11.

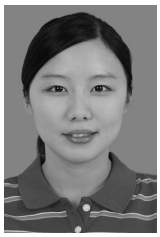
[5] JAIN S, KUMAR A, MANDAL S, et al. B4: Experience with a globally-deployed software defined WAN [C]//ACM SIGCOMM 2013 conference on SIGCOMM, August 12 - 16, 2013, Hong Kong, China. New York: ACM Press, 2013: 3-14.

[6] IETF. A one-way active measurement protocol: RFC4656[S]. 2006.

[7] IETF. A two-way active measurement protocol: RFC5357[S]. 2008.

[8] SALCEDO D, GUERRERO C D, MARTINEZ R, et al. Available bandwidth estimation tools metrics, approaches and performance[J]. International Journal of Communication Networks and Information Security (IJCNIS), 2018, 10(3).

[作者简介]



孙嘉琪（1987− ），女，中国电信股份有限公司智能网络与终端研究院工程师，主要从事 IP RAN/STN 术研究和相关工作。

杨广铭（1974− ），男，中国电信股份有限公司智能网络与终端研究院高级工程师，主要从事 IP RAN/STN 技术研究和相关工作。

党娟娜（1982− ），女，华为技术有限公司高级工程师，主要研究方向为下一代城域网络架构、分段路由 IPv6 的应用。

刘文杰（1989− ），男，华为技术有限公司高级工程师，主要研究方向为网络测量、网络优化、机器学习等。

研究与开发

面向多层次、多指标的光网络规划与评估方法

曹晓宏[1]，徐思雅[1]，周桂平[2]，王英杰[3]，于波涛[3]
（1. 北京邮电大学网络与交换技术国家重点实验室，北京 100876;
2. 国网辽宁省电力有限公司，辽宁 沈阳 110006;
3. 北京国电通网络技术有限公司，北京 100070）

摘　要：首先深入分析了影响业务可靠性的关键指标，并建立了面向可靠性和建设成本的网络规划模型，然后设计了自适应免疫算法以求解网络规划方案。进而，建立了面向多层次、多指标的网络规划方案评估模型，并使用神经网络算法获得指标权重，实现对网络规划方案的全面、有效评估。通过仿真可知，提出的面向多层次、多指标的光网络规划与评估方法能够提高光缆资源利用率和站点成环率、降低网络建设成本以及保障网络可靠性，有助于提高运维人员的工作效率和运维水平。

关键词：光网络；网络规划；评估模型；自适应免疫算法；神经网络算法
中图分类号：G304
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019206

Multi-level and multi-index planning and evaluation method for optical network

CAO Xiaohong[1], XU Siya[1], ZHOU Guiping[2], WANG Yingjie[3], YU Botao[3]
1. China State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
2. State Grid Liaoning Electric Power Co., Ltd., Shenyang 110006, China
3. Beijing Guodiantong Network Technology Co., Ltd., Beijing 100070, China

Abstract: The key indicators affecting business reliability were analyzed in depth, and a network planning model for reliability and construction cost was established. Then an adaptive immune algorithm was designed to solve the network planning model. Then, a multi-level and multi-index oriented evaluation model for network planning method was established, and the weights of the indexes were obtained by using neural network algorithm, so as to achieve a comprehensive and effective evaluation of the network planning method. The simulation results show that the multi-level and multi-index oriented optical network planning and evaluation method proposed can improve the utilization of optical cable resources and the site ring rate, reduce the network construction cost, ensure the network reliability, and help to improve the efficiency and quality of network operation and maintenance.

Key words: optical network, network planning, evaluation model, adaptive immune algorithm, neural network algorithm

收稿日期：2018−09−20；**修回日期**：2019−09−02
通信作者：徐思雅，1608764726@qq.com
基金项目：国家重点研发计划基金资助项目（No.2016YFB0901200）
Foundation Item: The National Key Research and Development Program of China (No.2016YFB0901200)

1　引言

光网络规划是指在网络建设初期对网络拓扑结构中站点、光缆和承载业务的传输路线进行合理部署。通过设计合理的网络规划方法以优化网络拓扑结构、保障网络的负载均衡是降低网络建设和改造成本以及获得高性能、可扩展网络的有效途径。然而，随着网络规模的不断增长、网络拓扑结构的日益复杂、业务种类和数量的不断增加，现有的面向单一优化目标的网络规划方法难以支撑网络的低成本、高质量的改造需求。并且，现有光网络规划方法大多针对物理层的站点、光缆，缺乏对业务层可靠性的考察，使运维人员无法全面地了解承载业务的分布情况，不能准确地评估网络所承载业务的运行风险。因此，设计一套高效、合理的网络规划方法，针对不同场景、不同需求生成针对性的网络改造方案，对优化网络拓扑结构、提高可靠性和降低网络建设成本具有重要作用。

网络规划方案评估是从网络架构的稳健性、网络可靠性和业务可靠性等方面对算法生成的网络规划方案进行性能评估。网络规划方案的评估结果可以直观地反映网络规划方案的合理性，并为网络规划方案的改进指明方向。然而，目前针对光网络规划方案评估的指标体系大多只面向物理层、网络层以及业务层中的两层，缺乏全面涵盖 3 个层次的评估方法，致使运维人员难以全面评估网络规划方案的有效性。因此，建立一套全面、科学的网络规划方案的评估方法，帮助运维人员准确、直观地评估网络规划方案，对验证网络规划方案的合理性、提高运维工作效率具有指导意义。

通过对国内外研究现状进行分析可知，各大运营商均在网络规划方案设计和网络规划方案评估分析方面进行了较为深入的研究[1-7]。针对网络规划方法，参考文献[2]提出了面向网络生存性的光网络规划方法，但缺乏对承载业务可靠性的考察；参考文献[3]从用户需求的角度出发，针对光网络的主干层、分配层、引入层进行规划设计，满足了用户对业务可靠性的要求，但网络建设成本较高；参考文献[4]从网络拓扑结构和网络可扩展性出发，针对佛山的地理环境和业务传输需求，设计了一种符合当前网络环境特点的网络规划方案；参考文献[5]给出了一套 EPON 终端通信接入网组网原则，并构建了以主干、汇聚和接入 3 层为标准的网络结构模型，阐述了网络结构、网络路径优化原则和方法。最后通过在电信息采集系统通道信道建设中的实践，证明了所提网络规划和优化方法的有效性。针对网络质量评估，参考文献[1]从网络性能角度，给出了光纤信息实时传输的具体指标；参考文献[6]提出了一种基于网络可靠性的光网络建模思想，对影响网络可靠性的关键性指标如信息流传输时延、分组丢失率等进行深入分析，并构建了一种在网络性能约束条件下的最大传输能力的网络模型；参考文献[7]针对网络环境的动态变化，提出了一种基于自治系统（autonomous system，AS）之间最优路径评估各个 AS 重要程度的方法，可以有效地发现 AS 中连接较少但重要程度较高的点，使得评估的重要性和实际重要性更吻合。基于以上参考文献的分析可知，针对光网络运行质量的研究主要集中在网络性能研究或业务质量的评估上，缺乏对二者结合的考虑，难以实现对网络规划的全面评估。

为解决以上问题，本文提出了一种面向多层次、多指标的光网络规划与评估方法，从成本、可靠性等角度对网络建设质量进行全面提升。首先，基于可靠性和建设成本构建网络规划模型，并从物理层、网络层、业务层 3 个维度构建网络规划方案评估模型。然后，设计自适应免疫算法以求解最优化网络规划方案，并采用神经网络算法获得评估模型权重，实现对网络规划方案的全面、有效评估。

2 网络规划与评估模型

本文所涉及的光网络规划方法是基于现有的网络拓扑结构，根据新增的业务需求和待选的光缆、站点，设计满足网络可靠性、经济性等约束要求的网络规划部署方案。网络规划与评估模型包括网络规划最优化问题模型和网络规划方案评估模型，本文将对两个数学模型进行分析。

2.1 网络规划最优化问题模型

由光网络的构成可知，光网络规划的对象为光纤和终端节点。光网络中的节点分布主要受用户分布影响，光纤分布主要由节点分布和节点间的互联关系决定。因此应针对用户需求，规划节点数量、位置和连接关系，并保障光缆纤芯余量率和光缆带宽冗余率等指标以满足网络可靠性的要求。

2.1.1 可靠性函数

本文从网络可靠性和业务可靠性两个方面建立可靠性函数，考察的参数包括光缆资源闲置率（光缆带宽冗余率和光缆纤芯余量率）、站点成环率和成环站点度加权值。

（1）站点成环率

网络中节点成环是指该拓扑结构中存在一个包含该节点的环，站点成环率为网络拓扑中成环的站点个数与站点总数的比值：

$$x=\frac{1}{s_{\text{sum}}}\sum \text{value} \tag{1}$$

其中，s_{sum} 表示网络拓扑中的站点总数。当站点处于环中时，value 的值为 1，否则为 0。

（2）成环站点度加权值

网络拓扑中站点 i 的度 k_i 为与该站点相连的站点数目。采用 min-max 标准化方法（min-max normalization，离差标准化）对网络拓扑中所有站点的度进行归一化处理：

$$\deg_j=\frac{d_j-d_{\min}}{d_{\max}-d_{\min}} \tag{2}$$

其中，$\deg_j$ 表示编号为 j 的站点度进行归一化处理后的值，d_j 表示编号为 j 的站点度数，$d_{\min}$ 表示网络拓扑中度数最小的站点度，$d_{\max}$ 表示网络拓扑中度数最大的站点度。

假设全网拓扑中站点集合为 S，$S=\{d_1,d_2,\cdots,d_n\},d_j\in S$，则成环站点度加权值可表示为：

$$Y=\sum_{j=1}^{j=n}e_j\deg_j \tag{3}$$

其中，$e_j\in\{0,1\}$，如果其值为 1，表示站点在环内；否则站点不在环内。

（3）光缆资源闲置率

光缆资源闲置率包括光缆带宽冗余率和光缆光纤余量率。从应急迂回和未来业务发展需求考虑，光缆带宽冗余率应不低于 30%；从网络可扩展性角度考虑，光缆光纤余量数应高于 4 根。

光缆带宽冗余率 A 为：

$$A=\frac{b_{\text{rem}}}{b_{\text{sum}}} \tag{4}$$

其中，b_{rem} 表示光缆中未使用的带宽值，b_{sum} 表示光缆中可用的最高带宽值。

光缆光纤余量率 B 为：

$$B=\frac{O_{\text{rem}}}{O_{\text{sum}}} \tag{5}$$

其中，O_{rem} 表示光缆中未使用的纤芯数，O_{sum} 表示光缆中的总纤芯数。

光缆资源闲置率 Z 为：

$$Z=\lambda\cdot A+\mu\cdot B \tag{6}$$

其中，λ 和 μ 是权重系数，且满足 $\lambda+\mu=1$。

根据式（1）、式（3）和式（6），可得到可靠性函数 R：

$$R=\alpha\cdot X+\beta\cdot Y+\gamma\cdot Z \tag{7}$$

其中，X 表示站点成环率，Y 表示成环站点度加权值，Z 表示光缆资源闲置率。而 α、β 和 γ 则为相应的权重系数，且 $\alpha+\beta+\gamma=1$。

2.1.2　网络建设成本函数

网络建设成本是指在网络规划方案的实施过程中，需要新增光缆、站点的建设成本：

$$C=\sum_{i=1}^{i=n}e_i m_i+\sum_{j=1}^{j=m}s_j m_j \tag{8}$$

其中，$i\in[1,n]$表示待选光缆条数为n，e_i的值为1时表示第i条光缆被选中，e_i值为0时表示第i条光缆未被选中；m_i表示建设第i条光缆的成本；s_j的值为1表示第j个站点被选中，s_j值为0表示第j个站点不被选中；m_j表示建设第j个站点的成本。

2.1.3　目标函数

在对网络规划方案进行设计时，通常以可靠性指标均达到最低阈值时的建设成本为基准，因此本文选择 min-max 算法对目标函数进行优化。为了以较低的建设成本获得较高性能的网络，本文将可靠性和建设成本的差作为优化变量F，将站点和链路的建设费用作为成本向量C，并根据网络规划方案的设计目标要求构成目标函数$S(F,C)$。目标函数的定义如下所示：

$$\begin{cases}F=\omega_1\cdot R-\omega_2\cdot\theta\cdot C\\ R=\alpha\cdot X+\beta\cdot Y+\gamma\cdot Z\\ C=\sum_{i=1}^{i=n}e_i\cdot m_i+\sum_{j=1}^{j=m}s_j\cdot m_j\\ S(F,C)=\left[\left(S_1(F,C_1)\right),\left(S_1(F,C_2)\right),\cdots,\left(S_1(F,C_n)\right)\right]^{\mathrm{T}}\\ \phi(F^*)=\min\max\left(S_1(F,C_i)\right),(i=1,2,3,\cdots,n)\end{cases} \tag{9}$$

其中，ω_1和ω_2表示建设成本函数C和可靠性函数R的权重系数，权重系数可根据不同使用场景对网络建设成本和可靠性的重视程度进行调整；θ为常数，用于调整网络建设成本C和可靠性R之间的数量级差异。

2.1.4　约束条件

首先，从可靠性函数分析，站点成环率应高于 60%；网络节点度尽量保持在 4 以下；根据部分业务的特定要求，比如为减少传输时延，需要在两个站点间直连光缆，从而导致某些待选光缆、站点成为必选。关于建设成本，可根据具体场景需求设置阈值。

2.2　网络规划方案评估模型

本文从物理层、网络层、业务层 3 个维度对影响网络规划方案质量的关键性指标进行深入分析，其中物理层从健康度、修复性、扩展性 3 个方面对站点和光缆进行评估；网络层从网络架构稳健度、网络性能对网络环境进行评估；业务层从故障保护、业务风险两个方面对传输业务质量进行评估。表 1 按评估维度给出了各级指标的定义、计算方法和正常指标值。

针对物理层，网络规划的对象主要为光缆和站点，因此本文将物理拓扑中光缆和站点可统计的属性信息作为本层的评估指标，评估指标包括健康度、修复性和扩展性 3 个方面。其中，健康度评估的是站点和光缆的完好率和故障历史，修复性评估的是光缆故障消除的及时率，扩展性评估的是站点和光缆的冗余情况。

针对网络层，其指标可分为网络架构稳健度和网络性能两大类。网络架构稳健度评估了网络拓扑的脆弱程度，网络性能表征了网络层的业务传输能力。针对网络架构稳健度，本文选取的评估指标包括网络边介数、网络关键点百分比、网络平均直径、网络成环率等；针对网络性能，本文选取的评估指标包括误码率、吞吐量、抖动时间、网络时延等。

针对业务层[11]，本文选取了通用的业务质量指标集合[6]中对业务可靠性影响程度较大的因素作为评估指标，并从故障保护和业务风险两个角度进行分析。故障保护可评价业务传输路线的冗余保护情况；业务风险评估了站点、链路的负载均衡情况。指标模型中各层次指标定义及计算方式见表 1[12-14]。

3　算法分析

网络规划是一个具有多目标性、多阶段性、

表 1　指标模型中各层次指标定义及计算方式

评估维度	二级指标	三级指标	计算方法	正常指标值
物理层	健康度	设备完好率	完好的台时数/日历工作台时数。 完好台时数等于日历工作台时减去故障及其修理的总台时数	≥95%
		平均故障间隔时间	设备的寿命单位总数/故障总次数	寿命为 10 年的设备，其故障次数应小于 100 次
		设备紧急故障次数	统计	≤28 次
		设备重大缺陷次数	统计	≤21 次
		光缆完好率	完好台时数/日历工作台时数	≥99.93%
		平均故障时间	光缆的寿命单位总数/故障总数	≥43 800 h
		光缆重大缺陷次数	统计	≤3
		光缆紧急缺陷次数	统计	≤5
	修复性	光电复合缆故障消除及时率	排除光电复合缆故障的时间/光缆故障排除标准时间	≤94.25%
		普通光缆故障消除及时率	排除普通类型光缆故障的时间/光缆故障排除标准时间	≤82.25%
		光缆段故障平均修复效率	总修复时间/总故障设备数目	≤98.6
		光电复合缆故障消除及时率	光电复合缆故障时间/修复故障时间	≥87.5%
	扩展性	核心板卡冗余率	空闲板卡数/设备总板卡数	≤4%
		业务槽位冗余率	可增加业务槽位数目/总槽位数目	≥20%
		光缆带宽余量率	剩余带宽/可用最高带宽	≥30%
		光缆光纤余量率	未用光纤数目/总光纤数目	光纤余量数≥4 根
网络层	网络架构稳健度	网络关键点百分比	关键点个数/总节点数	≤4%
		网络平均直径	两点间距离总和/节点数	[5.9,6.1]
		网络成环率	有效成环的设备数量/总设备数量	≥60%
		网络边介数	Math.sum(v 和 w 之间经过边 p 的最短路径数目/v 和 w 之间最短路径总数目)（v：站点 v（与 w 相连）；w：站点 w（与 v 相连）；p：链路 p）	≤0.1
		网络平均度	网络总度数/节点总数	［2.4，3.6］
		平均路径长度	网络路径总长度/路径条数	［2，6］
	网络性能	误码率	传输的误码/传输的总码数	$\leqslant 10^{-6}$
		抖动时间	最大时延−最小时延	<100 ms
		吞吐量	每秒请求次数	≥1 000
		过载分组丢失率	分组丢失数/发送分组数	≤1%
		网络时延/ms	发送时延+传播时延	(0,30)极快 (30,50)良好 (50,100)普通 (100,100+)差

（续表）

评估维度	二级指标	三级指标	计算方法	正常指标值
业务层	故障保护	故障保护率	故障保护成功次数/故障发生次数	≥99.3%
		波长冗余率	消耗的保护波长/总的占用波长总和	≥95%
		重光率	0.95（平均重光传输设备数量/光传输设备数量）+0.05 （平均重光缆数/光缆总数）	≥99.9%
		路径跳数变化率	Math.abs(故障前业务平均跳数-故障后业务平均跳数)/故障后业务平均跳数	≤24%
	业务风险	边风险值	链路承载业务数量	≤4
		节点风险值	站点承载业务数量	≤5
		网络风险值	链路风险值+节点风险值	–
		业务风险均衡度	业务发生风险性×风险发生的严重性	0.630 1

不确定性等特点的系统优化问题，免疫算法作为一种将生成和检测作为迭代过程的智能搜索算法，较适用于求解网络规划问题。但是，免疫算法的交叉、变异算子相对稳定，导致容易陷入局部最优的平衡态，并且进化后期常出现停滞不前的问题，从而很难获得所求问题的全局最优解。因此，本文设计了自适应免疫算法[9-10]，该算法始终采用最佳的交叉、变异算子，较好地平衡了收敛性和可行解多样性。同时本文采用神经网络算法获得评估模型权重，实现对网络规划方案的有效评估。

3.1　算子设计

（1）抗体编码

首先对光网络的网络规划过程中涉及的光缆、站点编号，编号从 1 开始，到 $N+M$ (待建设光缆、站点数目分别为 N 、 M)。抗体基因采用二进制编码方式，基因位数为 $N+M$ 。当第 i 号基因位为 1 时，表示第 i 条光缆需要被建设；当第 i 号基因位为 0 时，表示第 i 条光缆不需要被建设。网络规划中的站点标记采用同样方式。假设最终输出抗体上的基因序列为(101011100)，且低三位表示待建设站点信息，则网络规划方案中待建设光缆有 6 条，待建设站点有 3 个，且需要被建设的光缆为第 1 条、第 3 条、第 5 条和第 6 条，站点为第 1 个。

（2）计算抗体亲和度

免疫算法中，抗体亲和度表示优化问题可行解的质量，故本文中抗体亲和度表示光网络规划方案的质量，且由式（9）网络建设成本函数 C 和式（8）可靠性函数 R 共同决定。由于可行解质量的改进应朝着抗体亲和度增加的方向进行，故光网络规划问题中抗体亲和度可定义为：

$$\mathrm{Aff}_i = \frac{k_8 + s_i}{k_9} \tag{10}$$

其中，k_8 、k_9 为常数，s_i 为所求问题的目标函数。

（3）抗体克隆

本算法中克隆尺寸由抗体亲和度决定，并未采用等比例克隆方式。亲和度越高的抗体克隆子代个数越多。根据式（11），假设编号为 i 的抗体 $\mathrm{Aff}_i = 0.1$，且常数 $k_1 = 1$ ，则对抗体 i 进行克隆后，会生成包含 10 个子代的集合 Set_i ，$\mathrm{Set}_i = \{\mathrm{Set}_{i1}, \mathrm{Set}_{i2}, \cdots, \mathrm{Set}_{i10}\}$ 。

$$N_i = \mathrm{Floor}\left[k_1 / i\right] \tag{11}$$

其中，i 表示抗体标号，N_i 表示标号为 i 的抗体克隆后的个数，Floor 表示向下取整，k_1 为常数。

（4）抗体交叉

抗体间基因序列的交叉可以将父代的优良基因得到保存，获得具有更好基因序列的子代。

交叉概率为：

$$c_1=\begin{cases}k_4\cdot\sin\left(k_2\cdot\dfrac{a_{\max}-a_i}{a_{\max}-a_{\text{avg}}}\right),a_i>a_{\text{avg}}\\ k_3,a_i\leqslant a_{\text{avg}}\end{cases}\quad(12)$$

其中，a_i 表示第 i 号抗体的亲和度，a_{avg}、$a_{\min}$、$a_{\max}$ 分别表示种群抗体亲和度的平均值、最小值、最大值，k_2、k_3 和 k_4 为常数。

（5）抗体变异

抗体变异可以保证抗体种群的多样性。假设抗体 i 在繁殖后生成子代个数为 Child_i，且此类抗体的变异概率为 r_i，故此类抗体中需要参与变异操作的抗体个数为 $\text{Sum}_i=\text{Child}_i\cdot r_i$。

$$c_2=\begin{cases}k_7\cdot\cos\left(k_5\cdot\dfrac{a_{\max}-a_i}{a_{\max}-a_{\text{avg}}}\right),a_i>a_{\text{avg}}\\ k_6,a_i\leqslant a_{\text{avg}}\end{cases}\quad(13)$$

其中，a_i 表示第 i 号抗体的亲和度，a_{avg}、$a_{\min}$、$a_{\max}$ 分别表示种群抗体亲和度的平均值、最小值、最大值，k_5、k_6 和 k_7 为常数。

（6）计算选择概率

为保证算法具有良好的全局收敛性和高效的稳定性，需对良好抗体进行保存且保证种群抗体的多样性。因此在选择概率的定义中，本文选择正比抗体亲和度。抗体选择概率 Sec 低于平均选择概率 Sec_{avg} 的抗体将被舍弃。抗体 i 的选择概率为：

$$\begin{cases}\text{Sec}_i=\dfrac{\text{Aff}_i}{\text{Sum}_{\text{aff}}}\\ \text{Sum}_{\text{aff}}=\sum\limits_{j=1}^{j=n}\text{Aff}_i\end{cases}\quad(14)$$

其中，θ 为常数，Aff_i 表示抗体 i 的亲和度，Sum_{aff} 表示抗体亲和度总和。

3.2 算法流程

算法流程如图 1 所示。

算法的完整执行步骤如下。

步骤 1 系统初始化。

设定算法参数，导入光网络规划数据，包括站点、链路信息和相关常数等。

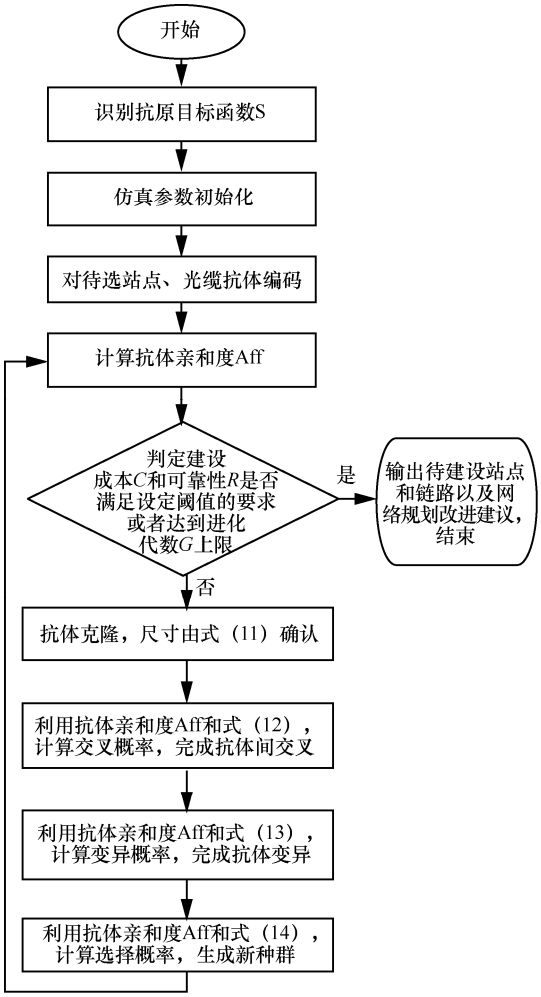


图 1 自适应免疫算法流程

步骤 2 抗原识别。

抗原即第 2.1 节中提及的目标函数和相关的约束条件。

步骤 3 产生初始抗体。

初始化抗体种群 Po_1，设定种群中抗体数为 N，对抗体进行编码并对其注射全局疫苗使得某些待选光缆和站点成为必选，设置进化代数 $G=1$。

步骤 4 计算抗体亲和度。

计算抗体亲和度 a_x，根据 a_x 值对所有抗体降序排序。检测此时获得的种群是否收敛，如果收敛则输出网络规划方案，停止循环；否则转步骤 5。

步骤 5 克隆操作。

抗体按亲和度降序排序，克隆尺寸定义如式（11），形成新的种群 Po_2。

步骤 6　交叉操作。

根据式（12）计算抗体间的交叉概率，形成新的种群 Po_3。

步骤 7　变异操作。

根据式（13）计算抗体的变异概率，形成新的种群 Po_4。

步骤 8　计算选择概率。

根据式（14）计算抗体的选择概率，形成新的种群 Po_5。进化代数 $G=G+1$，转至步骤 4。

自适应免疫算法（AI）的伪代码如下。

输入：站点和链路的抗体信息；站点数量 N；链路数量 M；默认参数 Params；最大迭代次数 G；建设成本函数 C；可靠性函数 R

输出：网络规划模型 Seq

初始化：站点和链路信息，包括位置、预制资金等；建设成本函数 C、可靠性函数 R；算法中所有默认的常量值

算法迭代（迭代次数≤G），针对每次迭代可做如下操作：

If 当前抗体群，做如下操作

如果 建设成本 C 和可靠性 R 和预置期望值一致 Then

End 算法并输出网络规划模型 Seq;

计算抗体亲和力 Aff_i；

根据亲和力对所有抗体进行降序排序；

单个抗体按照如下计算式完成克隆：

$$N_i=\mathrm{Floor}[k_i/i]$$

If $N_i \neq 1$ Then

获取克隆对象 a_i；

根据定义计算式，计算交叉、变异、选择概率；

根据交叉、变异概率更新所有抗体；

计算选择概率 Sec_i；

If $\mathrm{Sec}_i < \mathrm{Sec}_{\mathrm{avg}}$ Then

删除所有与当前抗体类型相同的抗体；

迭代次数 G=G+1；

If G=1 000 Then

End If 算法并输出网络规划评估模型 Seq。

4　网络规划方案仿真分析

4.1　参数设置

本文所使用的仿真数据来源于国内某区域光网络的部分网管数据。图 2 中，网络拓扑结构包含 21 个已建设站点，4 个待建设站点，26 条已建设光缆，18 条待建设光缆，且链路[2-4]和[2-6]的光缆资源闲置率略高于 30%，其余光缆的资源冗余率均高于 46%。

本节算法中涉及的通用参数见表 2，两种算法参数对比见表 3。

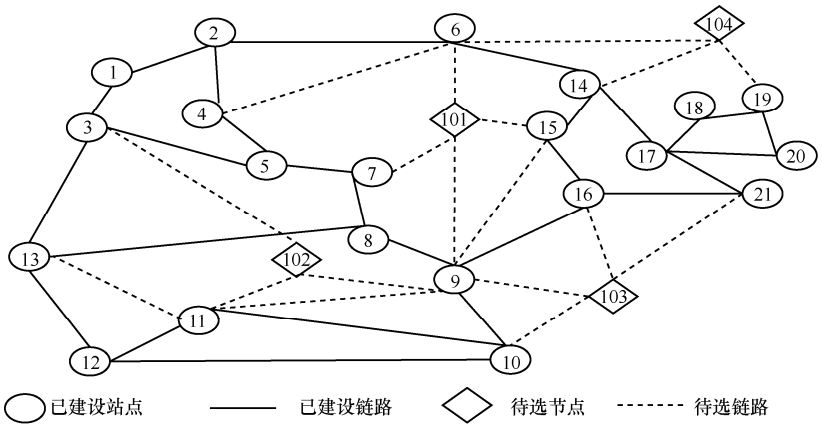


图 2　初始网络规划拓扑图

表 2 通用仿真参数设置

参数类型	参数值
种群规模	70
抗体长度	22
光缆带宽冗余率权重 λ	0.5
光缆光纤冗余率权重 μ	0.5
成环率权重 α	0.4
度加权值权重 β	0.3
光缆资源闲置率 γ	0.3
目标函数权重 ω_1	0.6（可调）
目标函数权重 ω_2	0.4（可调）
克隆公式中常数 k_1	30
交叉公式中常数 k_2	0.1
交叉公式中常数 k_3	0.8
交叉公式中常数 k_4	0.3
变异公式中常数 k_5	0.2
变异公式中常数 k_6	10
变异公式中常数 k_7	0.9
抗体亲和度中常数 k_8	1 000
抗体亲和度中常数 k_9	1

4.2 实验结果与分析

本文采用免疫算法作为对比算法，对本文提出的面向多层次、多指标的光网络规划与评估方法进行性能分析。由于不同应用场景对可靠性和网络建设成本的依赖程度不同，目标函数中的权重系数 ω_1 和 ω_2 被设为可调。现就权重系数 ω_1 和 ω_2 、网络建设成本、可靠性、站点成环率、成环站点度加权值、光缆资源闲置率之间的关系进行研究与仿真，仿真结果统计见表 4。

由表 4 可知，网络建设成本所占比重越小，可靠性越高，成环站点度数越低，光缆资源闲置率越高，站点成环率可达 60%以上，同时权重系数 $\omega_1:\omega_2$ 的取值在［1/9，7/3］之间均能满足光网络正常运行的最低标准。

图 3 是仿真后的网络拓扑结构。链路[4-6]成为待建设链路，原因是链路[2-4]和[2-6]的光缆资源闲置率接近 30%，需要通过新增链路的方式降低其上负载业务的运行风险。站点 103 未成为待建设站点，其原因是建设该站点将导致新增 4 条链路，建设成本明显增大，并且原有的拓扑结构能够满足现有业务和可靠性的需求。通过与图 2 对比可知，光缆资源闲置率提升了 9%，网络平均度上升了 0.43，但仍低于阈

表 3 两种仿真算法参数对比

算法名称	交叉概率 C_1	变异概率 C_2	克隆尺寸 N_i	进化代数 G_i	其他参数
自适应免疫算法	见表 2	见表 2	见表 2	1 000	见表 2
免疫算法	0.7	0.01	1:1	1 000	见表 2

表 4 某区域光网络仿真结果数据

(ω_1,ω_2)	网络建设成本 C	可靠性 R	站点成环率	成环站点度加权值	光缆资源闲置率
(0.1,0.9)	310	95.2%	100%	0.12	0.89
(0.2,0.8)	295	90.1%	100%	0.22	0.80
(0.3,0.7)	260	83.9%	100%	0.28	0.77
(0.4,0.6)	235	79.5%	98%	0.32	0.6
(0.5,0.5)	195	71%	91%	0.44	0.55
(0.6,0.4)	170	69%	88%	0.49	0.43
(0.7,0.3)	135	60.7%	74%	0.53	0.39
(0.8,0.2)	110	54%	66%	0.61	0.34
(0.9,0.1)	98	48%	52%	0.78	0.23

值 4，故本次仿真获得的最优化网络规划方案在保证建设成本较低的情况下提高了业务可靠性和网络可靠性。

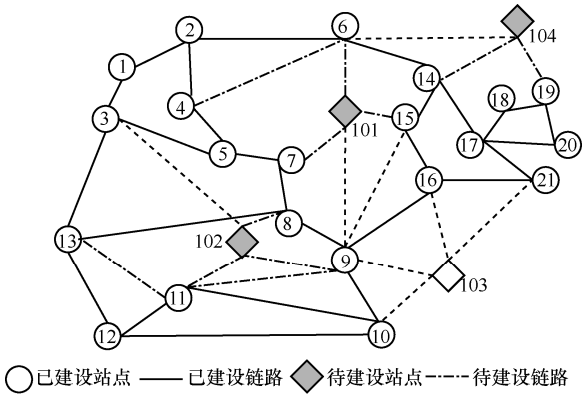


图 3　规划后网络拓扑结构

由图 4 可知，进化代数接近 200 时，两种算法获得的网络规划方案建设成本相同；低于 200 时，免疫算法获得的网络规划方案成本更低；超过 200 时自适应免疫算法获得的网络规划方案成本更低，且建设成本比免疫算法低 30% 左右。进化代数接近 400 时，两种算法获得的网络规划方案可靠性相同；超过 400 时自适应免疫算法获得的网络规划方案可靠性更高，并且可靠性比免疫算法高出接近 10%；低于 400 时则相反。造成上述现象的原因是本文采用的自适应免疫算法改进了免疫算法的交叉、变异算子，具有更好的全局收敛性，但收敛速度稍低于传统的免疫算法。

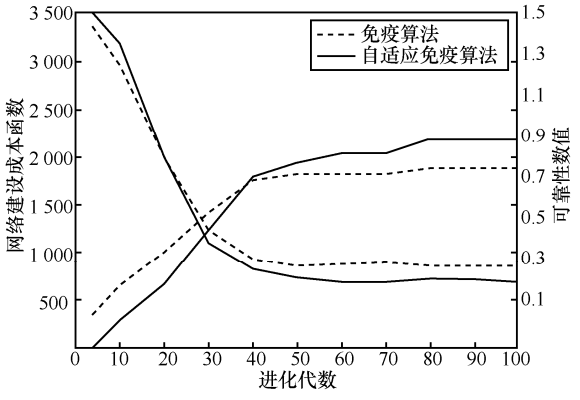


图 4　网络规划模型子函数与进化代数关系

通过对以上仿真结果的分析和对比可知，本文设计的网络规划与评估模型能够根据不同场景下对可靠性和网络建设成本的不同需求，设计出符合实际需要的网络规划方案。并且，本文采用的自适应免疫算法始终采用最佳的繁衍参数，能够在有限的时间内达到收敛且得到全局最优解，具有较高的准确性和稳定性。

5　结束语

为提高光网络可靠性并降低运维成本，提出了一种面向多层次、多指标的光网络规划与评估方法。首先，构建了以提高可靠性、降低建设成本为目标的网络规划模型，并利用自适应免疫算法求解最优化网络规划方案。然后，构建了基于物理层、网络层、业务层 3 个层次的网络规划方案评估模型，并通过神经网络算法获得指标模型权重，从而全面、客观地评估网络规划方案。通过仿真可知，本文提出的光网络规划与评估方法能够提高网络可靠性和业务可靠性、降低网络建设成本，是提高网络管理水平、改善工作效率的有效途径。

参考文献：

[1] 胥辉旗, 朱平云, 潘莉莉, 等. 基于高速光纤实时网的弹上信息传输技术研究[J]. 通信学报, 2010(S1): 161-165.
XU H Q, ZHU P Y, PAN L L, et al. Research on transmit technology for the missile system based on ultrahigh-speed real-time network[J]. Journal on Communications, 2010(S1): 161-165.

[2] 谭红君，解亚敏，刘业君. 可生存光纤—无线融合宽带接入网规划方法[J]. 电信科学，2016, 32(5): 138-145
TAN H J, JIE Y M, LIU Y J. Planning approach of survivable fiber-wireless broadband access network [J]. Telecommunications Science, 2016, 32(5): 138-145.

[3] 乔元志. 城市光纤接入网的规划与设计[J] . 电信科学, 2000, 16(4): 22-25.
QIAO Y Z. The research about optical access network planning in city[J]. Telecommunications Science, 2000, 16(4): 22-25.

[4] 李文权，赵奇禄，林位株. 接入网技术及佛山接入网规划建设[J]. 电信科学, 1998, 14(12):33-35.
LI W Q, ZHAO Q L, LIN W Z. Access network technology and

planning and construction of Foshan access network [J]. Telecommunications Science,1998, 14(12): 33-35.

[5] 江龙才. 终端通信接入网规划建设与优化管理[J]. 电力系统通信，2012(11): 21-26.
JIANG L C. Plan construction and optimization management of terminal communication access network[J]. Ower System Communication, 2012(11): 21-26.

[6] 赵娟, 郭平, 邓宏钟, 等. 基于信息流动力学的通信网络性能可靠性建模与分析[J]. 通信学报, 2011(8): 159-164.
ZHAO J, GUO P, DENG H Z, et al. Modeling and analysis of performance reliability for communication networks based on traffic dynamics[J]. Journal on Communications, 2011(8): 159-164.

[7] 刘红军, 胡晓峰, 邓文平, 等. 基于首选路由的 AS 重要性评估方法[J]. 软件学报, 2012(9): 2388-2400.
LIU H J, HU X F, DENG W P, et al. Technique of evaluating AS importance based on preferred route[J]. Journal of Software, 2012(9): 2388-2400.

[8] MEHDI M, BIJAN R, AHMAD A. Improving linear discriminat anylysis with artificial immune system-based evolutionary algorithms[J]. Information Science, 2011: 219-322.

[9] GOMEZ-GARDENES J, ECHENIQUE P, MORENO Y. Immunization of real complex commun- ication networks[J]. The European Physical Journal B, 2006(2): 259-264.

[10] 徐杨, 袁峰, 林琪, 等. 基于混合人工免疫算法的流程挖掘事件日志融合方法[J]. 软件学报, 2018(2): 396−416.
XU Y, YUAN F, LIN Q, et al. Merging event logs for process mining with a hybrid artificial immune algorithm[J]. Journal of Software, 2018 (2):396−416.

[11] ROSSIGNOLI F, LIONZO A. Network impact on business models for sustainability: case study in the energy sector[J]. Software & Systems Modeling, 2018, 2 (3): 694-704.

[12] 电力骨干光传输网络规划与运行质量评价系统用户手册[R]. 2016.
User manual of power backbone optical transmission network planning and operation quality evaluation system[R]. 2016.

[13] 终端接入网规划和运维技术需要分析和评价机制研究报告[R]. 2018.
Research report on the mechanism of analysis and evaluation of terminal access network planning and operation and maintenance technology needs[R]. 2018.

[14] 大容量光传输网络指标体系与质量评估研究报告[C]//中国电机工程学会电力通信专业委员会学术会议. 2013.
Research report on index system and quality assessment of large capacity optical transmission network[C]//Academic Meeting of the Electric Power Communication Professional Committee of China Electrical Engineering Society. 2013.

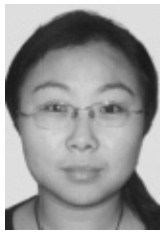
[作者简介]

曹晓宏（1991− ），女，北京邮电大学网络与交换技术国家重点实验室硕士生，主要研究方向为网络管理与通信软件。

徐思雅（1988− ），女，博士，北京邮电大学网络与交换技术国家重点实验室讲师，主要研究方向为智能电力通信网网络管理。

周桂平（1981− ），男，博士，国网辽宁省电力有限公司高级工程师，主要研究方向为设备状态监测及故障诊断。

王英杰（1964− ），男，北京国电通网络技术有限公司高级工程师，主要研究方向为电力信息通信技术。



于波涛（1977− ），女，北京国电通网络技术有限公司高级工程师，主要研究方向为电力通信信息管理。

研究与开发

差分混沌移位键控在水声通信中的应用

代红英[1]，陈梦蕾[2]，徐位凯[3]

（1. 重庆工程学院，重庆 400056;
2. 国家电网浙江省电力有限公司信息通信分公司，浙江 杭州 310000;
3. 厦门大学信息与通信工程系，福建 厦门 361005）

摘　要：水声信道作为目前已知最严酷、最复杂的无线通信信道之一，具有强多途干扰、时—频率双扩展、高噪、带宽窄等特征。水声信道的时变特性，使得估计与跟踪信道很困难，因此，研究无需信道估计与均衡的非相干稳健水声通信调制方法具有重要的意义。首先介绍了基于正交频分复用（OFDM）和扩频调制的水声调制技术的研究进展，然后，分析了差分混沌移位键控在水声信道下的可行性，提出了两种基于 OFDM 的多载波差分混沌移位键控方案，给出两个方案的调制解调器原理。在时—频双扩展信道和水声信道下，对两个方案进行了性能仿真和分析。性能结果表明所提出的方案在水声信道下具有良好的稳健性。

关键词：水声通信；多载波差分混沌移位键控；正交频分复用；稳健性

中图分类号：TN911

文献标识码：A

doi: 10.11959/j.issn.1000−0801.2019217

Underwater acoustic communications based on differential chaos shift keying

DAI Hongying[1], CHEN Menglei[2], XU Weikai[3]

1. Chongqing Institute of Engineering, Chongqing 400056, China
2. Information and Communication Branch of State Grid Zhejiang Electric Power Co., Ltd., Hangzhou 310000, China
3. Department of Information and Communication Engineering, Xiamen University, Xiameng 361005, China

Abstract: As the most stringent and complicated channel, underwater acoustic (UWA) channels are featured by high multipath interference, time-frequency doubly spread, high noise, and narrow bandwidth. Due to time varying of underwater acoustic, it is difficult to trance and estimate channel state information (CSI). The underwater acoustic communication techniques based on orthogonal frequency division multiplex (OFDM) and spread spectrum were reviewed. Then, the characteristic of differential chaos shift keying (DCSK) was analyzed, and two multicarrier differential chaos shift keying modulations were proposed. Finally, over time-frequency doubly spread channel and underwater acoustic channel, the bit errors rate (BER) performances for two proposed schemes were presented and analyzed. Numerical results show that the proposed schemes own good robustness over underwater acoustic channels.

Key words: underwater acoustic communication, multicarrier differential chaos shift keying, orthogonal frequency division multiplex, robustness

收稿日期：2019−06−28；修回日期：2019−08−27

1 引言

海洋占据 71%的地球面积，含有丰富的海洋资源，具有巨大的经济及军事战略价值。随着人类海洋活动的日益增加，海洋高新技术的研究已成为新科技革命的重要领域之一[1]。海上石油开采、海底地形勘测、地震海啸监测、水下机器人遥感、蛙人语音通信等海洋活动对水下信息通信提出了重大需求。数据、语音、图像的水下传输对通信的可靠性以及传输速率都提出了巨大的挑战。另一方面，随着社会对水下通信的需求逐渐增加，水下通信的科研与商业价值愈加凸显，使水下通信成为近年来极为活跃与发展迅速的领域之一[2]。随着水下通信技术的成熟，在未来的 5G 或者 6G 通信系统中[3-6]，水下通信网络将成为未来移动通信系统必不可少的组成部分。因此，开展水下无线通信相关的技术研究不仅具有重要的理论价值，而且还具有现实的必要性。

研究表明，水下信号能量的衰减与载波频率的平方成正比，因此无线电波无法在水下进行远距离传播，即使低频段电磁波也只能穿透 100 m 左右的海水。要实现水下中/远距离信号传输，声波是目前唯一的信号载体，水声通信也由此成为水下通信与组网的首选方式[7-8]。海面、海底对声波的反射散射，水中介质不均造成的折射，导致接收机不仅收到声波的直射分量，还收到了大量不同时延不同强度的折射、反射、散射分量，由此产生了严重的多途现象，当最大时延相对于符号周期不可忽略时，即发生频率选择性衰落；水体的流动、海平面的随机起伏使多普勒频偏现象不可避免，由此产生了信道的时变性。因此具有强多途干扰、长传输时延、大随机起伏、高噪等特性的水声信道是严重的时–频双扩展信道，也是目前已知的最复杂、最严酷的无线信道之一[9]。

水声通信主要面临四大问题。第一，时延扩展大。由于海平面反射，海底表面起伏产生漫射以及海水介质不均产生折射等原因，严重的符号间干扰（ISI）可以达到上百个传输符号长度。第二，可用传输带宽窄。由于信号在海水中信号衰减与频率的平方成正比，传播衰减较小的可用频段只有几十 kHz，因此水声通信的数据传输率也较低。第三，多普勒频偏严重。由于水下声速在 1 500 m/s 左右，海面波浪的随机起伏，收、发端的小幅移动，都会造成水声信道的时变，从而影响信号的传输。第四，接收信噪比低。由于较高的海洋环境噪声以及信号在传播过程中的衰减，信号接收端的信噪比较低，影响水下通信质量。因此寻找合适的调制技术克服水声信道对传输信号的影响，实现可靠、稳健、高速的通信是研究者共同追求的目标。

与地面通信技术发展趋势相同，水声通信经历了模拟到数字、单载波到多载波、单输入输出（single input single output, SISO）到多输入输出（multiple input multiple output, MIMO）的发展历程。数字通信技术主要包括振幅键控（amplitude shift keying, ASK）、相移键控（phase shift keying, PSK）、频移键控（frequency shift keying, FSK）。ASK 技术性能易受噪声的影响，在水声信道中并不被广泛采用。FSK 调制技术被认为是一种适用于水声信道的调制方式。非相干的多进制 FSK 调制技术具有良好的抗干扰抗多径能力，且接收端不需要相位同步易解调，在水声信道被广泛运用[10]。1989 年，Catipovic 等人[11]设计的 MFSK 调制水声系统在浅海传输距离水平方向 3 km，工作带宽 10 kHz，误码率（bit error rate, BER）可达 10^{-3} 量级。1998 年，美国 WHOI 与 Datasonics 公司采用 MFSK 调制方式设计了水声数据遥感系统[12]，在传输距离 4 km、输入信噪比 10~12 dB 的情况下，BER 可达 10^{-3}~10^{-2}。然而 FSK 有一个明显缺点，其频带利用率极低，对于带宽受限的水声信道有很大的制约。近期，正交频分复用（OFDM）的高频谱效率的特性，使得其在高速

水声通信中受到广泛的关注，而扩展频谱通信具有强的抗干扰能力同样在水声通信中获得了青睐。

本文首先简要综述了扩频与正交频分复用（OFDM）的水声通信研究进展情况，然后分析了差分混沌移位键控（DCSK）应用于水声信道的优缺点，最后提出了两个基于差分混沌移位键控调制的水声通信系统方案，并在水声信道下研究了所提方案的误比特率性能。

2　正交频分复用（OFDM）水声通信技术

OFDM 技术由于其较高的频谱效率以及良好的抗 ISI 能力，在 20 世纪 90 年代，逐步被研究者们引入水声通信领域。1994 年，Coatelan 和 Glavieux 提出了多载波水声通信系统[13]，该系统的每一条子载波采用的是 FSK 调制，采用非相干解调，实际上是 MFSK 系统。隔年，该团队将(133/171)卷积编码以及交织器加入多载波系统[14]，以降低传输系统误码率。在时变水声信道下，OFDM 技术最主要面对的问题是严重的多普勒扩展导致的 OFDM 子载波之间的正交性破坏，从而产生 ICI 问题。参考文献[15]提出利用多项式抑制编码（polynomial cancellation coding, PCC）从载波频偏角度来抑制子载波间干扰（inter carrier onterference, ICI）。然而时变信道的信道响应是不断变化的，每次频偏都不同，因此 PCC 的适用性非常有限。参考文献[16]提出利用 2×ID 最小均方误差（minimum mean squared error, MMSE）来完成信道估计与均衡。参考文献[17]提出一种线性 MMSE 估计器用于时变信道，该系统用两项泰勒级数展开来线性近似时域信道变化。然而 MMSE 均衡算法对于信道响应矩阵求逆的算法复杂度较高，其复杂度为 $O(K^3)$，K 为子载波个数。对于子载波数较大的系统，硬件实现难度较大。为了降低 OFDM 系统的 ICI 均衡复杂度，参考文献[18]提出一种复杂度较低的两级均衡器，该方案首先利用线性预处理将干扰符号压缩到子载波间隔中，其次利用迭代 MMSE 来估计频域符号，其本质将信道响应矩阵简化为带状对角矩阵。参考文献[19]利用基扩展模型对时变信道进行降维后再进行 ICI 均衡，以此降低计算复杂度。该类方法主要以增加接收端信号处理复杂度为代价来降低 ICI 的影响。参考文献[20]利用子载波之间的结构特点通过冗余传输的方法消除 ICI。与此同时，系统的传输效率却降低了。

3　扩频水声通信技术

扩频技术利用伪随机序列对传输信号进行扩频调制，接收端利用伪随机序列良好的自相关特性对传送信号进行解扩，其具备抗干扰、抗多径、隐蔽性好等优点。与此同时，由于扩频技术可以获得扩频增益，其可以在低信噪比环境下工作。因此，扩频技术被认为是一种有效的水声通信技术，我国“蛟龙号”载人潜艇就使用了该技术。扩频通信在水声通信中主要面临两大问题，首先信号传输效率较低，其次对多普勒扩展较为敏感。针对数据传输率低的问题，参考文献[21]提出了循环位移键控（cyclic shift keying, CSK）的扩频通信方式，利用扩频序列的循环位移特性对信息序列进行映射编码，成倍提供扩频系统的数据传输效率。但对于时变信道，多普勒扩展导致的载波相位跳变将会造成 CSK 系统的扩频增益严重下降。参考文献[22]提出了 M 元直接序列扩频方式，根据输入的信息比特进行二进制转换，转换后的数字从 M 个扩频序列中选出符合映射的扩频序列进行传输，该方案可以提高 lb$M$ 倍的传输效率。参考文献[23]提出将 CSK 技术与 M 元直接扩频技术及多载波技术进行结合，进一步提高系统的数据传输率。针对信道时变问题，研究者们考虑了自适应均衡器，参考文献[24]提出了假设反馈均衡算法，对假设切普（chip）序列进行 chip 速率更新代替实际判断反馈，实现对时变信道的跟踪与

补偿。但该方案中的假设反馈均衡器是非中心化的，对 ISI 问题并没有进行研究。参考文献[25]将 RAKE 接收机运用于直接扩频系统用于多径信号延时合并。但是在低信噪比环境下，RAKE 接收机难以达到预期的效果。

基于 OFDM 技术与扩频技术本身具有的优势，将扩频技术与 OFDM 技术结合运用于时变水声信道亦不失为一种好办法[26-27]。多载波直接序列扩频（multi-carrier direct sequence spread spectrum, MC-DSSS）技术最早是由 Kondo 与 Milstein 于 1996 年在参考文献[28]提出的，其基本设计方法是将经过 PN 码扩频后的信息码片并行加载到相互正交的子载波上，在接收端通过每个子载波的自相关器以及最大比合并（maximal ratio combining, MRC）进行解调。参考文献[29]提出了短扩频的 MC-DSSS 系统用于水声信道，该系统接收端利用自适应均衡器与载波相位估计器进行解调，仿真结果显示，特别是大时延扩展信道下，多载波扩频系统要比单载波扩频系统具有明显优势。

4 混沌水声通信技术

混沌扩频序列具有的类随机性以及初始敏感性，使其广泛运用于扩频通信中。参考文献[30]研究表明，基于混沌序列的扩频通信系统性能要优于基于 GOLD 码序列扩频系统。研究者们也将混沌扩频系统应用于时变水声信道，参考文献[31]将量化后的混沌序列作为扩频序列运用于水声通信系统，研究混沌扩频系统在浅海水声信道中的性能。但该方案接收端采用相干解调，由于混沌序列的初始敏感性，在现实中获得与发送端一样的混沌序列是较为困难的。为了克服伪随机序列周期性和二值性的不足，保证更好的保密性，参考文献[32]直接采用非量化的混沌序列作为扩频序列，研究了多通道混沌调相扩频系统在水声通信中的可行性。由于混沌信号对初值异常敏感，在实际中很难保证发射机和接收机能够获得相同的混沌序列。对于不需要混沌同步的差分混沌移位键控，参考文献[33-34]研究了差分混沌移位键控在水声传输环境下的性能表现。参考文献[35]将多载波技术与差分混沌移位键控进行结合，该方案在时−频双扩展信道下展现出良好的性能，该方案采用了非相干解调，不需要载波同步、信道估计，系统实现复杂度低，但系统数据传输率较低。为了克服参考文献[35]中的系统频率效率不高、缺乏灵活性的问题，参考文献[36]提出了另一种多比特并行传输的多载波差分混沌调制键控系统。该系统将多个 CS-DCSK 信号并行调制到不同子载波上，即一个差分混沌移位键控符号的码片仅占据一个子载波，由此增加系统灵活性。其在双扩展信道以及水声信道环境中展现出良好的性能。

4.1 差分混沌移位键控水声通信技术

已有研究表明，DCSK 在多径衰落信道下具有良好的性能表现，且由于采用非相干接收，在接收端不需要载波同步、信道估计与信道均衡，其具有简单的发送/接收机结构。然而在大时延扩展的水声信道下，直接采用 DCSK 并不能有效地对抗时延扩展带来的 ISI。因此，结合码复用的 DCSK 调制[37]，基于 OFDM 的多载波原理，提出了两种多载波码复用差分混沌移位键控调制方案，它们在水声信道下均表现出优良的性能。

4.1.1 多载波码复用差分混沌移位键控方案 1: MC- CS-DCSK-I

MC-CS-DCSK-I 系统发射端的原理如图 1 所示，该系统首先将输入的信息比特 $b\in\{0,1\}$ 通过 BPSK 调制为调制信息 $a\in\{-1,+1\}$。同时，混沌信号发生器通过 Logistic 映射 $c_{i+1}=1-2c_i^2, i\in N^+$ 生成长度为 β 的混沌码片 $\boldsymbol{c}=\left[c_1 \quad c_2 \quad \cdots \quad c_\beta\right]$。Walsh 码矩阵发生器生成一个 $P\times P$ 的 Walsh 码矩阵 $\boldsymbol{W}=\left[\boldsymbol{w}_1^{\mathrm{T}} \quad \boldsymbol{w}_2^{\mathrm{T}} \quad \cdots \quad \boldsymbol{w}_P^{\mathrm{T}}\right]$，其中，$\boldsymbol{w}_i=\left[w_{i,1} \quad w_{i,2} \quad \cdots \quad w_{i,P}\right], 1\leqslant i\leqslant P$。该矩阵由

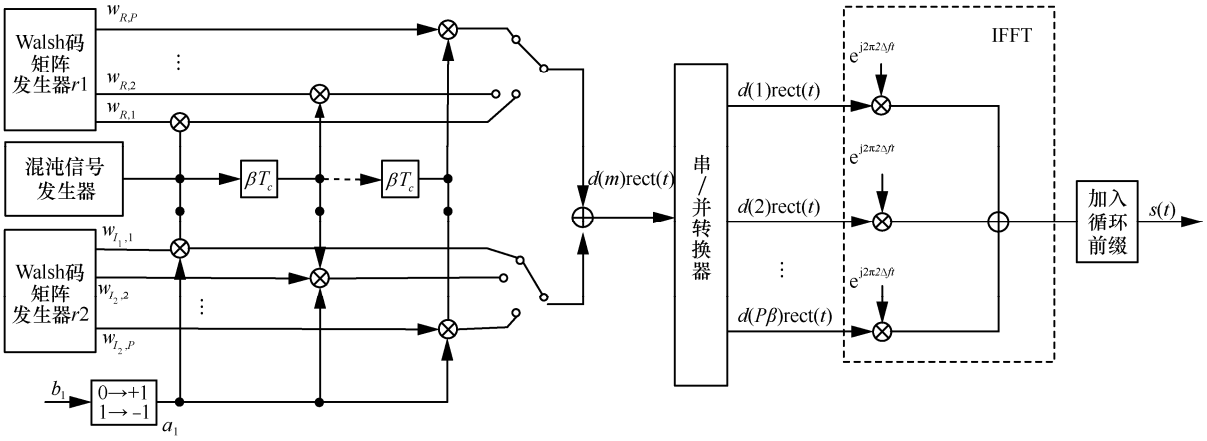


图 1　MC-CS-DCSK-I 发射端原理

Hadamard 构造方式生成，矩阵中不同行之间保证相互正交，不同列之间保证相互正交，满足 $\boldsymbol{w}_n \times \boldsymbol{w}_m^{\mathrm{T}} = \begin{cases} 0, & m \neq n \\ P, & m = n \end{cases}$。

接着从 Walsh 码矩阵中选取两行相互正交的 Walsh 码 $\boldsymbol{w}_R = \left[w_{R,1} \quad w_{R,2} \quad \cdots \quad w_{R,P} \right], R \in [1, P/2]$，$\boldsymbol{w}_I = \left[w_{I,1} \quad w_{I,2} \quad \cdots \quad w_{I,P} \right], I \in [P/2, P]$ 分别与 β 长度的混沌序列 $\boldsymbol{c} = \left[c_1 \quad c_2 \quad \cdots \quad c_\beta \right]$ 进行调制。其中，混沌序列 c 分别与 $\boldsymbol{w}_R$ 中的每一个元素相乘由此构成长度为 $P\beta$ 的扩频码片作为参考信号，混沌序列 c 与调制信息 a 以及 $\boldsymbol{w}_I$ 中的每一个元素相乘，由此构成的长度为 $P\beta$ 的扩频信号作为信息承载信号。然后将参考信号与信息承载信号对应码片相加，由此构成的信息符号 $\boldsymbol{d} = \left[d_1 \quad d_2 \quad \cdots \quad d_{P\beta} \right]$，实现参考信号与信息承载信号在时–频域重叠，但在码域正交的特性。之后，将信息码片 d 进行串行变换，调制到各个相互正交的子载波上，信号结构如图 2 所示，可表示为式（1）：

$$s(t) = \sum_{m=1}^{P\beta} \left[d_m \, \mathrm{rect}(t) \mathrm{e}^{\mathrm{j}2\pi m \Delta f t} \right] \tag{1}$$

其中，d_m 可表示为：

$$d_m = \boldsymbol{w}_R \otimes \boldsymbol{c} + a\boldsymbol{w}_I \otimes \boldsymbol{c}, m = 1, 2, \cdots, P\beta \tag{2}$$

其中，$\boldsymbol{c} = \left[c_1 \quad c_2 \quad \cdots \quad c_\beta \right]$ 表示长度为 β 的混沌序列，$a \in \{-1, +1\}$ 表示调制信息，$\boldsymbol{w}_R = \left[w_{R,1} \quad w_{R,2} \quad \cdots \quad w_{R,P} \right]$ 与 $\boldsymbol{w}_I = \left[w_{I,1} \quad w_{I,2} \quad \cdots \quad w_{I,P} \right]$ 表示长度为 P 的相互正交的 Walsh 码序列，$\otimes$ 表示克罗内克（Kronecker）运算符。$\Delta f = 1/(P\beta T_c)$ 表示子载波之间的频率间隔，rect(*t*)表示矩阵函数，其表达式为：

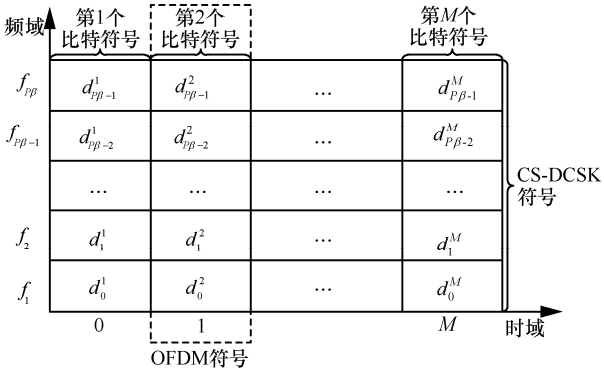


图 2　MC-CS-DCSK-I 发射信号结构

$$\mathrm{rect}(t) = \begin{cases} 1, & t \in [0, P\beta T_c] \\ 0, & \text{其他} \end{cases} \tag{3}$$

其中，$T_s = P\beta T_c$ 表示一个 OFDM 符号周期，T_c 表示符号码片周期。参考信号与信息承载信号的正交性，通过式(4)证明。

$$\Delta = \sum_{p=1}^{P} w_{R,p} \sum_{m=1}^{\beta} c_m \cdot a \sum_{p=1}^{P} w_{I,p} \sum_{m=1}^{\beta} c_m = a \sum_{p=1}^{P} w_{R,p} w_{I,p} \left(\sum_{m=1}^{\beta} c_m \right)^2 = a \frac{E_b}{2P} \sum_{p=1}^{P} w_{R,p} w_{I,p} = 0 \tag{4}$$

其中，Δ 表示内积符号，$E_{\mathrm{b}}=2P\sum_{m=1}^{\beta}c_m^2$ 表示 $2P\beta$ 个混沌码片能量，同时也表示为传输比特信号能量，$E_{\mathrm{b}}/2P=\sum_{m=1}^{\beta}c_m^2$ 表示 β 个混沌码片能量，$\sum_{p=1}^{P}w_{R,p}w_{I,p}=0$。

最后将加载到子载波后的信号进行并串转化，加入循环前缀，送入信道。

信号经历的信道为时–频双扩展信道。信道的冲激响应函数为：

$$h(t,\tau)=\sum_{l=0}^{L-1}A_l(t)\delta\left(\tau-\tau_l(t)\right) \tag{5}$$

其中，L 表示多径数，$A_l(t)$ 表示第 l 径信道系数，$\tau_l(t)=\tau_{l0}-a_l t$ 表述第 l 径时延。

MC-CS-DCSK-I 系统的接收机工作原理如图 3 所示。接收端收到的信号 $r(t)$ 首先进行时间间隔为 T_c 的采样，将模拟信号 $r(t)$ 转化为离散数字信号 r_n。收到的第 $m\left(m\in[1,P\beta]\right)$ 个离散数字信号 r_m 可表示为：

$$r_m=\sum_{i=0}^{L-1}h_i^{(m)}s\left(\left(n-i\right)\right)_{P\beta}+\eta_m \tag{6}$$

其中，$\left((\bullet)\right)_{P\beta}$ 表示以 $P\beta$ 为周期进行循环移位，η_m 表示均值为 0、方差为 N_0 的复高斯噪声。L 表示时–频双扩展信道中的多径数，$h_i^{(m)}$ 表示传输第 m 个码片持续时间 $t\in[(m-1)T_c,mT_c]$ 内的第 i 个抽头系数。这里假设信道在 $(m-1)T_c\leqslant t\leqslant mT_c$ 时间内信道冲激响应保持不变。

接着将经过采样后的离散信号移除循环前缀，利用串/并变换器，将串行信号变成并行信号，然后将并行信号送入长度为 $P\beta$ 的FFT变换器中，将信号从子载波中解调出来。第 k 个子载波上的信号 R_k 可表示为：

$$R_k=S_kH_{k,k}+\underbrace{\sum_{i=1,i\neq k}^{P\beta}H_{k,i}S_i}_{\text{ICI}}+N_k \tag{7}$$

其中，$N_k=\sum_{i=1}^{P\beta}\eta_i\mathrm{e}^{-\mathrm{j}2\pi ik/(P\beta)}$，$1\leqslant k\leqslant P\beta$ 表示复高斯白噪声的频域表达式，$S_k=1/\sqrt{P\beta}\sum_{i=1}^{P\beta}s_i\mathrm{e}^{-\mathrm{j}2\pi ki/(P\beta)}$ 表示第 k 个子载波上传输的信号。$H_{k,i}$ 表示时–频双扩展信道的频域冲激响应，其式为：

$$H_{k,i}=\frac{1}{P\beta}\sum_{n=0}^{L-1}F_n(i)\mathrm{e}^{-\frac{\mathrm{j}2\pi n(k-i)}{P\beta}},1\leqslant k,i\leqslant P\beta \tag{8}$$

其中，F_n 为时变信道第 n 个抽头系数的频域值，其计算式为：

$$F_n(k)=\sum_{u=1}^{P\beta}h_n^{(u)}\mathrm{e}^{-\mathrm{j}2\pi uk/(P\beta)},\ 0\leqslant n\leqslant L-1,1\leqslant k\leqslant P\beta \tag{9}$$

从式（7）可以看出，经过傅里叶变换后的信道频域响应矩阵 $\boldsymbol{H}_{P\beta\times P\beta}$ 不再是对角矩阵，这是因为在时变信道中，多普勒频移会造成多载波信号子载波间干扰（ICI）。

接着，将经过傅里叶变化后的信号 $\boldsymbol{R}=\begin{bmatrix}R_1 & R_2 & \cdots & R_{P\beta}\end{bmatrix}$ 分别与参考信号的 Walsh 码序列 $\boldsymbol{w}_R=\begin{bmatrix}w_{R,1} & w_{R,2} & \cdots & w_{R,P}\end{bmatrix}$ 以及信息承载

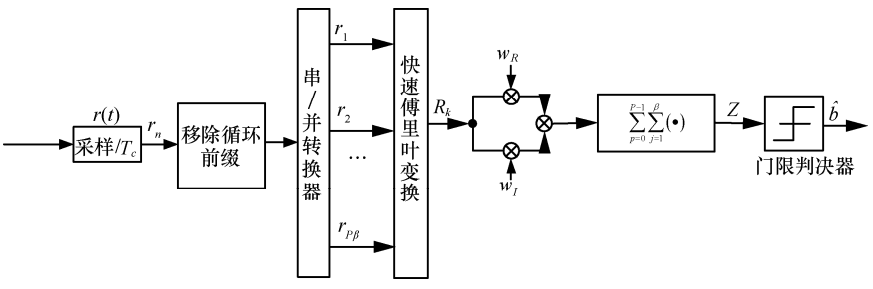


图 3　MC-CS-DCSK-I 接收端原理

信号的 Walsh 码序列 $\boldsymbol{w}_I=\begin{bmatrix}w_{I,1} & w_{I,2} & \cdots & w_{I,P}\end{bmatrix}$ 相乘，相乘后的两路信号再次相乘，最后将相乘后的对应码片累加，得到输出判决量 Z，其计算式为：

$$Z=\mathrm{Re}\left\{\sum_{p=0}^{P-1}\left(w_{R,p+1}w_{I,p+1}\sum_{j=1}^{\beta}R_{p\beta+j}R_{p\beta+j}^{*}\right)\right\} \tag{10}$$

其中，$\mathrm{Re}\{x\}$ 表示对变量 x 取实部，$(x)^*$ 表示对变量 x 取共轭。

最后，利用门限判决器对传输信息进行估计，获得估计比特 $\hat{b}$，其判决门限值为 0。

$$\hat{b}=\mathrm{sign}(Z) \tag{11}$$

其中，sign(x) 为符号函数。

4.1.2　多载波码复用差分混沌移位键控 2：MC-CS-DCSK-Ⅱ

MC-CS-DCSK-Ⅱ系统的发送端原理如图 4 所示。在该系统中，首先将 N 个并行的比特流 $\boldsymbol{b}=\begin{bmatrix}b^1 & b^2 & \cdots & b^N\end{bmatrix}$ 放入 N 个 CS-DCSK 调制器中进行调制，在 CS-DCSK 中，每个信息比特 b^i 将被混沌序列扩频成长度为 $P\beta$ 的信息符号流。由此，获得 N 个并行的长度为 $P\beta$ 的 CS-DCSK 符号。之后，每一列并行的 N 个码片通过交织器进行循环位移以此提高相邻码片之间的独立性。接着将经过交织器循环位移后的信息码片调制到相互正交的子载波上，最后经过并串变化，加入循环前缀，送入信道。

其中，CS-DCSK 调制器首先由混沌信号生成器生成 β 个码片 $\boldsymbol{c}=\begin{bmatrix}c_1 & c_2 & \cdots & c_\beta\end{bmatrix}$，其中，混沌码片的映射方式为 Logistic 映射，即 $c_{i+1}=1-2c_i^2, i\in N^+$。Walsh 码矩阵生成一个 $P\times P$ 的 Walsh 码矩阵 $\boldsymbol{W}=\begin{bmatrix}\boldsymbol{w}_1^{\mathrm{T}} & \boldsymbol{w}_2^{\mathrm{T}} & \cdots & \boldsymbol{w}_P^{\mathrm{T}}\end{bmatrix}$，其中，$\boldsymbol{w}_i=\begin{bmatrix}w_{i,1} & w_{i,2} & \cdots & w_{i,P}\end{bmatrix}, 1\leqslant i\leqslant P$。参考信号由 β 长的混沌码片与 $\boldsymbol{w}_R=\begin{bmatrix}w_{R,1} & w_{R,2} & \cdots & w_{R,P}\end{bmatrix}$，$R\in[1,P/2]$ 中每一个元素进行相乘得到长度为 $P\beta$ 的混沌码片。信息承载信号是由 β 长的混沌码片与 $\boldsymbol{w}_I=\begin{bmatrix}w_{I,1} & w_{I,2} & \cdots & w_{I,P}\end{bmatrix}, I\in[P/2,P]$ 中每一个元素以及与调制信息 $a_i\in\{-1,+1\}$ 相乘，由此构成 $P\beta$ 长的信息承载信号。最后将参考信号与信息承载信号在时域上叠加，由此得到 CS-DCSK 符号 $\boldsymbol{d}=\begin{bmatrix}d_1 & d_2 & \cdots & d_{P\beta}\end{bmatrix}$，第 l 个 CS-DCSK 符号可表示为：

$$\boldsymbol{d}^l=\boldsymbol{w}_R\otimes\boldsymbol{c}+a_l\boldsymbol{w}_I\otimes\boldsymbol{c} \tag{12}$$

其中，$a_l\in\{-1,+1\}$ 由 $b_l\in\{0,1\}$ 经过 BPSK 调制所得，$\otimes$ 表示克罗内克运算符。

由此，获得的 N 行并行的 CS-DCSK 信号表示为：

$$\boldsymbol{T}=\begin{bmatrix}d_1^1 & d_2^1 & \cdots & d_{P\beta}^1\\ d_1^2 & d_2^2 & \cdots & d_{P\beta}^2\\ \vdots & \vdots & \ddots & \vdots\\ d_1^N & d_2^N & \cdots & d_{P\beta}^N\end{bmatrix} \tag{13}$$

令 $\boldsymbol{Y}_k=\begin{bmatrix}d_k^1 & d_k^2 & \cdots & d_k^N\end{bmatrix}^{\mathrm{T}}$，$1\leqslant k\leqslant P\beta$ 表示第 k 个 OFDM 符号，由不同 CS-DCSK 信号的相同位置的码片构成。接着将矩阵 T 中的每一列，

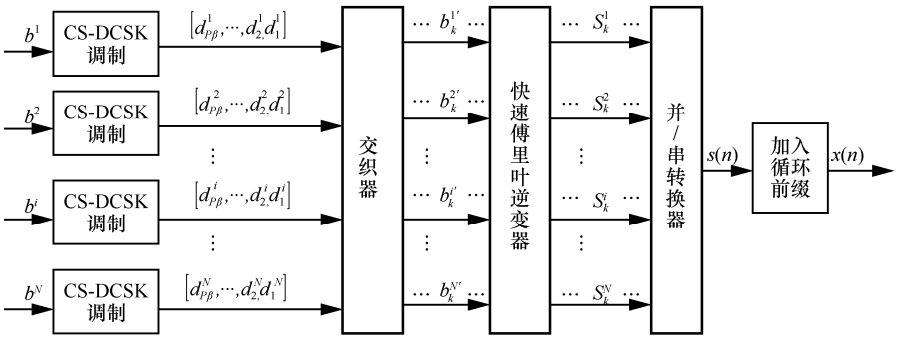


图 4　MC-CS-DCSK-Ⅱ系统发送端原理

即每一个 OFDM 符号送入交织器中进行循环位移，即对 Y_k 进行循环位移，循环位移 j 个位置用计算式可表示为：

$$\boldsymbol{Y}_k^j = \boldsymbol{Q}^j \boldsymbol{Y}_k \tag{14}$$

其中，Q^j 表示为矩阵 Q 的 j 次幂，矩阵 Q 被定义为：

$$\boldsymbol{Q} := \begin{bmatrix} 0_{1\times N-1} & 1 \\ \mathrm{I}_{N-1} & 0_{(N-1)\times 1} \end{bmatrix} \tag{15}$$

$\boldsymbol{I}_N$ 为 $N\times N$ 的单位矩阵。例如，N=4 时，矩阵 Q 可表示为：

$$\boldsymbol{Q} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{16}$$

当 j=3 时，矩阵 Q^j 可表示为：

$$\boldsymbol{Q}^j = \boldsymbol{Q}^3 = \boldsymbol{Q}\cdot\boldsymbol{Q}\cdot\boldsymbol{Q} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{17}$$

最后，经过交织器后的 $P\beta$ 个 OFDM 符号加载到子载波上，信号构成如图 5 所示。d'^i_k 表示经过循环位移后的第 k 个 OFDM 符号 $\boldsymbol{Y}_k^j$ 向量上的第 i 个码片，该码片将调制到第 i 个子载波上。第 $k\left(k\in[1,P\beta]\right)$ 个 OFDM 符号发送信号为：

$$s(t) = \sum_{k=1}^{P\beta}\sum_{i=1}^{N}\left\{d'^i_k \cos\left[2\pi\left(f_c + i\Delta f\right) + \varphi_i\right]\mathrm{rect}\left(t - kT_s\right)\right\} \tag{18}$$

其中，N 表示载波数量，f_c 表示载波频率，φ_i 表示第 i 个子载波的随机相位，Δf=1 / T_s 表示载波间频率间隔，T_s 表示 OFDM 符号周期，$P\beta$ 表示 OFDM 个数，即传输完整的 N 个 CS-DCSK 扩频信号需要的 OFDM 个数。

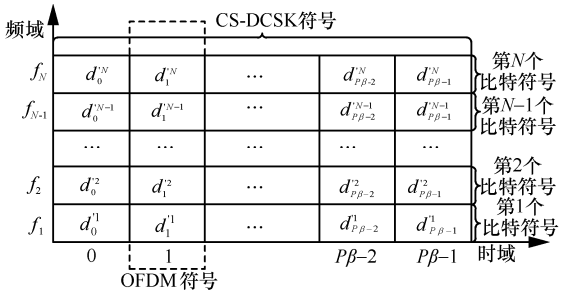


图 5　MC-CS-DCSK-Ⅱ发送信号结构

MC-CS-DCSK-Ⅱ系统的接收端工作原理如图 6 所示。在接收端，接收到的信号 $y(t)$首先进行时间间隔为 T_c 的采样，将模拟信号 $y(t)$转化为数字信号 y_n，其中，T_c=T_s/N。接着移除循环前缀，获得信号 r_n，随后，将 r_n 进行串行变换，r_k^i 表示第 $k,k\in[1,P\beta]$ 个 OFDM 上的第 $i,i\in[1,N]$ 个采样点，其表达式为：

$$r_k^i = \sum_{l=0}^{L-1} h_{i,l}^k s_k\left(\left(i-l\right)\right)_N + \eta_{k,i} \tag{19}$$

其中，$((\bullet))_N$ 表示以 N 为周期进行循环位移，$\eta_{k,i}$ 表示均值为 0、方差为 N_0 的复高斯噪声。L 表示时–频双扩展信道中的多径数，$h_{i,l}^k$ 表示传输第 k 个 OFDM 符号中第 i 个码片持续时间 $t\in\left[(i-1)T_c, iT_c\right]$ 内的第 l 个抽头系数。这里假设

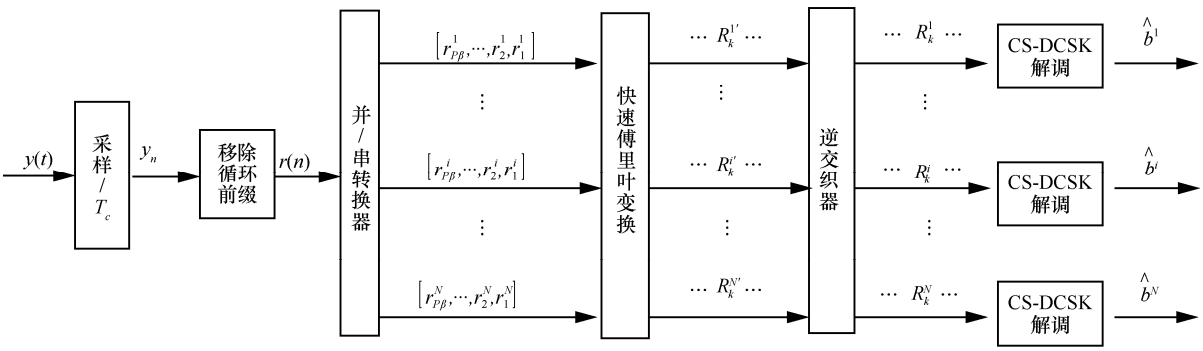


图 6　MC-CS-DCSK-Ⅱ系统接收端原理

信道在$(i-1)T_c \leqslant t \leqslant iT_c$时间内信道冲激响应保持不变。

接下来将收到的第 k 个 OFDM 符号的 N 个并行码片进行 FFT 变换，将信息码片从子载波中解调出来，$R_k^{'i}$ 表示第 k 个 OFDM 符号中第 i 个码片，其计算式为：

$$R_k^{'i} = S_{k,i}H_{i,i}^k + \underbrace{\sum_{j=1,j\neq i}^{N} H_{i,j}^k S_{k,j}}_{\text{ICI}} + N_{k,i}, 1 \leqslant k \leqslant P\beta, 1 \leqslant i \leqslant N \tag{20}$$

其中，$S_{k,i} = 1/\sqrt{N}\sum_{j=1}^{N} s_{k,j} e^{-\mathrm{j}2\pi ij/N}$ 表示第 k 个 OFDM 符号中第 i 个子载波上的信号，$N_{k,i} = 1/\sqrt{N}\sum_{j=1}^{N} \eta_{k,i} \mathrm{e}^{-\mathrm{j}2\pi ij/(N)}$ 表示复高斯白噪声的频域表达。$H_{i,j}^k$ 表示时−频双扩展信道的频域冲激响应，其计算式为：

$$H_{i,j}^k = \frac{1}{N}\sum_{l=0}^{L-1}\sum_{n=1}^{N} h_{n,l}^k \mathrm{e}^{\frac{-\mathrm{j}2\pi(jn+l(i-j))}{N}}, 1 \leqslant i,j \leqslant N, 1 \leqslant k \leqslant P\beta \tag{21}$$

然后将经过 FFT 解调后的信号送入解交织器中，进行反向循环移位，第 k 个 OFDM 符号 $\boldsymbol{R}_k^{'}$ 经过解交织器后的表达式为：

$$\boldsymbol{R}_k = \left(\boldsymbol{Q}^j\right)^{\mathrm{T}} \left(\boldsymbol{R}_k^{'}\right)^{\mathrm{T}}, 1 \leqslant k \leqslant P\beta \tag{22}$$

其中，$\boldsymbol{R}_k^{'} = \left[R_k^{'1} \quad R_k^{'2} \quad \cdots \quad R_k^{'N}\right]$。

将每次解交织后的信号存储于 $N \times P\beta$ 的矩阵存储器 $\boldsymbol{G}$ 中的一列，第 k 个解交织后的信号 $\boldsymbol{R}_k = \left[R_k^1 \quad R_k^2 \quad \cdots \quad R_k^N\right]$ 存储于矩阵 $\boldsymbol{G}$ 的第 k 列。经过 $P\beta$ 次解交织后，获得 $N \times P\beta$ 的信息矩阵，即：

$$\boldsymbol{G} = \begin{bmatrix} R_1^1 & R_2^1 & \cdots & R_{P\beta}^1 \\ R_1^2 & R_2^2 & \cdots & R_{P\beta}^2 \\ \vdots & \vdots & \cdots & \vdots \\ R_1^N & R_2^N & \cdots & R_{P\beta}^N \end{bmatrix} \tag{23}$$

接着将矩阵 $\boldsymbol{G}$ 的每一行送入每一个 CS-DCSK 解调器中，第 i 行信号 $\boldsymbol{D}^i = \left[R_1^i \quad R_2^i \quad \cdots \quad R_{P\beta}^i\right], 1 \leqslant i \leqslant N$ 送入第 i 个 CS-DCSK 解调器进行解调。将输入信号分别与参考信号的 Walsh 码序列 $\boldsymbol{w}_R = \left[w_{R,1} \quad w_{R,2} \quad \cdots \quad w_{R,P}\right]$ 以及信息承载信号的 Walsh 码序列 $\boldsymbol{w}_I = \left[w_{I,1} \quad w_{I,2} \quad \cdots \quad w_{I,P}\right]$ 相乘，相乘后的两路信号再次相乘，接着将相乘后的对应码片累加，得到输出判决量 Z^i，其表达式为：

$$Z^i = \mathrm{Re}\left\{\sum_{p=0}^{P-1}\left(w_{R,p+1}w_{I,p+1}\sum_{j=1}^{\beta} R_{p\beta+j}^i \left(R_{p\beta+j}^i\right)^*\right)\right\} \tag{24}$$

其中，$\mathrm{Re}\{x\}$ 表示对变量 x 取实部，$(x)^*$ 表示对变量 x 取共轭。

最后，利用门限判决器(11)即可恢复发送的比特。

4.2　仿真结果

4.2.1　时频双扩展衰落信道下的性能

本节给出MC-CS-DCSK-I和MC-CS-DCSK-Ⅱ在时−频双扩展信道下的性能，并与 CS-DCSK[37]和基于 OFDM 的多载波 DCSK[38]进行比较。在以下的仿真结果中，相关的参数如下：L 为信道多径数，SF 为扩频因子，CP 为循环前缀的长度，f_{norm} 为归一化多普勒频移（定义了多普勒频移与系统带宽的比值）。

图 7 展示了 MC-CS-DCSK-I 与 MC-DCSK 在不同多普勒频偏的频率平坦信道下的 BER 性能曲线，图 7 中参数设置 f_{norm}=0、0.001、0.01、0.1、L=1，SF=64，CP=T_s/8。此时，子载波带宽为 15.6 Hz，OFDM 符号周期为 64 ms，循环前缀长度为 8 ms，多普勒频移分别为 f_{d}=0、1 Hz、10 Hz、100 Hz。从图 7 中可以看出，在静态平坦信道下，即 f_{norm}=0，L=1 时，MC-CS-DCSK-I 系统与 MC-DCSK 基本一致，MC-CS-DCSK-I 稍优于 MC-DCSK。但随着归一化的多普勒频移 f_{norm} 增大，MC-CS-DCSK-I 的 BER 曲线斜率随着 f_{norm} 的增加而增加，即 f_{norm} 越大，曲线斜率越大，系统性能越好。当 f_{norm}=0.1，

SNR=24 dB 时，MC-CS-DCSK-I 的 BER 曲线已达到 10^{-6} 级别，显示出优越的性能。与之相反，MC-DCSK 性能随着 f_{norm} 增加而变差。在 $f_{\text{norm}} \geqslant 0.01$ 时，MC-DCSK 出现明显的“错误地板（error floor）”，随着 SNR 增加，BER 曲线仅达到 10^{-1} 级别。由此可认为相比于 MC-DCSK，MC-CS-DCSK-I 在时变信道中具有更好的稳健性。

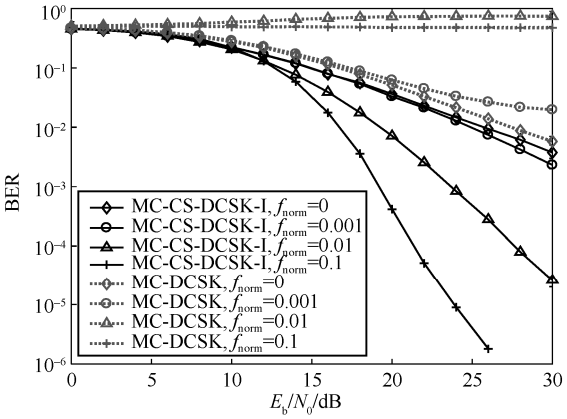


图 7　不同多普勒频移下，MC-CS-DCSK-I 与 MC-DCSK 性能比较

图 8 展示了 MC-CS-DCSK-I、MC-DCSK 以及 CS-DCS 在时−频双扩展信道中的 BER 性能曲线，图 8 中参数为 f_{norm}=0.001、0.01、0.1，L=7，SF=64，CP=T_s/8。从图 8 中可以看到，MC-CS-DCSK-I 性能优于 MC-DCSK 以及 CS-DCSK。对于 MC-DCSK，在 f_{norm} 较小时，即 f_{norm}=0.001 时，其性能曲线在 SNR=24 dB 时可达 10^{-6} 级别，但是随着 f_{norm} 增加，该系统出现明显的“错误地板”，在 f_{norm}=0.01、SNR=30 dB 时，系统 BER 曲线仅达 10^{-1} 级别。对于 CS-DCSK，在 L=7 的严重多径时延时变信道中，出现了明显的 ISI 问题。但 CS-DCSK 抗信道时变能力要优于 MC-DCSK，在 f_{norm} 从 0.01 增加到 0.1 时，BER 性能有所提高。对于 MC-CS-DCSK，其性能明显优于前两个系统，虽然在 f_{norm}=0.1 系统性能略有下降，但在 SNR=26 dB 时，其 BER 曲线已达到 10^{-6} 级别，展现良好的性能。由此可以认为 MC-CS-DCSK-I 在时−频双扩展信道中具有良好的稳健性。

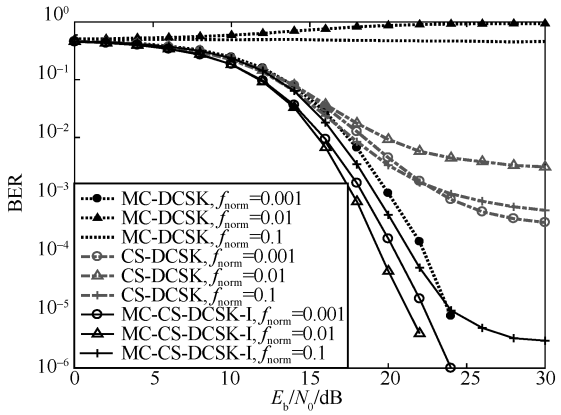


图 8　双扩展信道下，MC-CS-DCSK-I、MC-DCSK、CS-DCSK 的 BER 性能比较

图 9 研究了在双扩展信道下，循环前缀对 MC-CS-DCSK-I 的影响。其中参数设置为 f_{norm}=0.01，SF=64，L=1、3、7，CP=0、T_s/8。此时，信道多普勒频移为 f_{d}=10 Hz，信道最大时延分别为 0、4 ms、12 ms，OFDM 符号周期为 64 ms，循环前缀分别为 0、8 ms。从图 9 中可以看到，当 L 较小的时候，如 L=1、3 时，加不加循环前缀对系统性能没有很大影响，加循环前缀的曲线与不加循环前缀的曲线几乎重合，尤其是平坦信道(L=1)下，不加循环前缀的系统 BER 曲线反而要稍优于加循环前缀的系统曲线。但当 L 较大时，如 L=7 时，加循环前缀的曲线性能优于不加循环前缀的曲线。这是因为随着 L 的增大，系统符号间干扰（ISI）增加，加循环前缀有助于信号抵抗 ISI 提高系统性能。但是，从图 9 中可以看到，尽管在时延较大的时变信道中(L=7，f_{norm}=0.01)，

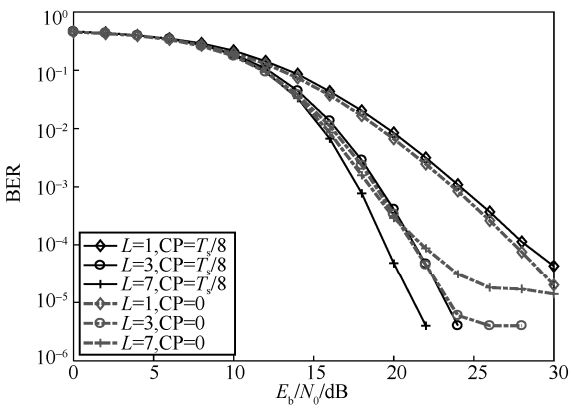


图 9　双扩展信道下，循环前缀对 MC-CS-DCSK-I 系统性能的影响

MC-CS-DCSK-I 系统在 CP=0 时，BER 曲线仍可展现出较好的性能，当 SNR=26 dB 时，BER 可达到 10^{-5} 级别。这是 MC-CS-DCSK-I 所具备的另一个良好特性。在 BER 要求不是很严格的通信系统中，可以牺牲部分传输性能，不加循环前缀以提高系统频谱效率，这是传输性能与频谱效率之间的一个权衡。

图 10 展示了 MC-CS-DCSK-Ⅱ与 MC-DSSS 在双扩展信道下的 BER 性能曲线，图中参数设置为 N=64，SF=32，L=7，f_{norm}=0、0.001、0.005、0.01，CP=T_s/8。此时，子载波带宽为 15.6 Hz，OFDM 符号周期为 64 ms，循环前缀长度为 8 ms，多径最大时延为 12 ms，多普勒频移分别为 f_d=0、1 Hz、5 Hz，10 Hz，分别占子载波带宽的 0、6.4%、32%、64%。从图 10 中可以看到，MC-CS-DCSK-Ⅱ系统性能明显优于 MC-DSSS 系统。在 f_{norm}=0 时，信道为静态频率扩展信道，MC-CS-DCSK-Ⅱ相比于 MC-DSSS 展现出更优的抗 ISI 能力，SNR=25 dB 时，MC-CS-DCSK-Ⅱ的 BER 达到了 10^{-7} 级别。MC-DSSS 的 BER 虽然随着 f_{norm} 的增加而降低，但是其最小 BER 仍远高于 MC-CS-DCSK-Ⅱ。对于 MC-CS-DCSK-Ⅱ，在 f_{norm}=0.005 时，系统性能虽相比于 f_{norm}=0.001 时有所下降，但是在 SNR=23 dB 时，系统性能仍可达到 10^{-5} 级别，完全满足实际通信要求。不过当 f_{norm}=0.01 时，系统出现明显“错误地板”，BER 曲线仅到达 10^{-2} 级别。由此表明 MC-CS- DCSK-Ⅱ在一定多普勒效应范围内具有良好的性能。

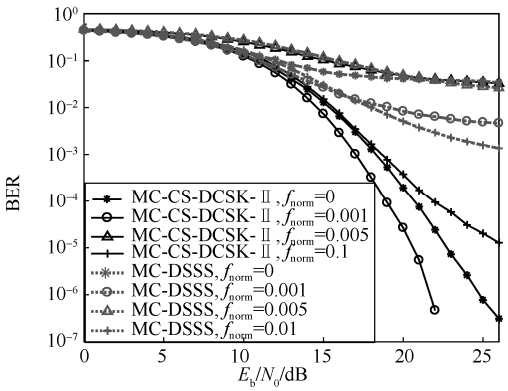


图 10　MC-CS-DCSK-Ⅱ系统与 MC-DSSS 系统在双扩展信道下的 BER 性能比较

为寻找 MC-CS-DCSK 的多普勒效应适应范围，图 11 展现了固定 E_b/N_0 情况下，不同多普勒频移、不同多径时延下的 MC-CS-DCSK-I 和 MC-CS-DCSK-Ⅱ的 BER 性能曲线。图 10 参数：SNR=20 dB，SF=32，CP=T_s/8，L=1、3、8，多径最大时延分别为 0、4 ms、14 ms、横坐标多普勒频移（Doppler shift）表示多普勒频移与子载波带宽的比值，当 Doppler shift=200%时，表示多普勒频移为子载波带宽的两倍。首先，从图 11 中看到，MC-CS-DCSK-I 与 MC-CS-DCSK-Ⅱ系都具有良好的抗 ISI 能力，随着多径数 L 的增加，性能都提高，两个系统都具有频率分集增益。其次，对于 MC-CS-DCSK-I，随着多普勒频移增加，系统的 BER 曲线一直保持在一个较好的水平上，系统具有良好的抗时变性能。对于 MC-CS-DCSK-Ⅱ，在 Doppler shift >50%时，系统出现明显“错误地板”，BER 曲线保持在 10^{-2}~10^{-1} 水平。然而在多普勒频移较小时，MC-CS-DCSK-Ⅱ系统则展现出优于 MC-CS-DCSK-I 的性能，并且在多普勒频移=10%左右时，展现出最优性能。因此，可以认为 MC-CS-DCSK-Ⅱ在 Doppler shift 为 0~30%，具有良好的抗时变性能。面对不同的时变信道，可通过调整系统的子载波带宽，使 MC-CS-DCSK-Ⅱ展现出最好的 BER 性能。

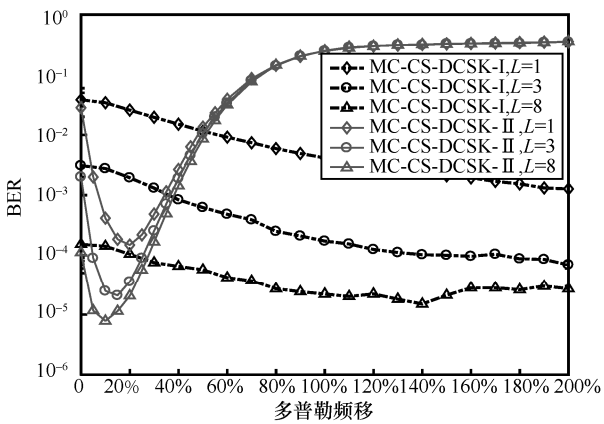


图 11　固定 SNR，MC-CS-DCSK-I 与 MC-CS-DCSK-Ⅱ在不同多普勒频移的双扩展信道下的 BER 性能

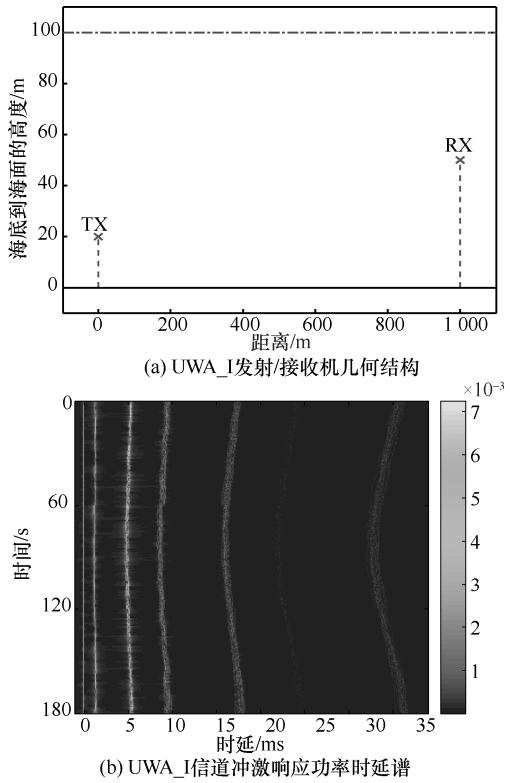
4.2.2 水声信道下的系统性能

本节基于 millitsa 水声信道模型[39]，对两个系统在水声信道下的性能进行了仿真。选用两种水声信道场景，其信道参数见表 1，设定系统带宽等于信道带宽。两个水声信道的几何结构以及信道冲激响应的功率时延谱如图 12 和图 13 所示。从图中可以看出，第一个水声信道（UWA_I）是一个大时延短距离水声信道，第二个水声信道（UWA_Ⅱ）为小时延中长距离水声信道。

表 1 水声信道仿真参数

仿真参数	取值
发射端距离底部距离	20 m
接收端距离底部距离	50 m
水中的声速	1 500 m/s
海底的声速	1 200 m/s
仿真信道时间长度	180 s
时间分辨率	0.05 s
频率分辨率	25 Hz
发射端与接收端水平距离(UWA_I)	1 000 m
中心频率(UWA_I)	15 kHz
带宽(UWA_I)	10 kHz
发射端与接收端水平距离(UWA_Ⅱ)	5 000 m
中心频率(UWA_Ⅱ)	1.5 kHz
带宽(UWA_Ⅱ)	400 Hz

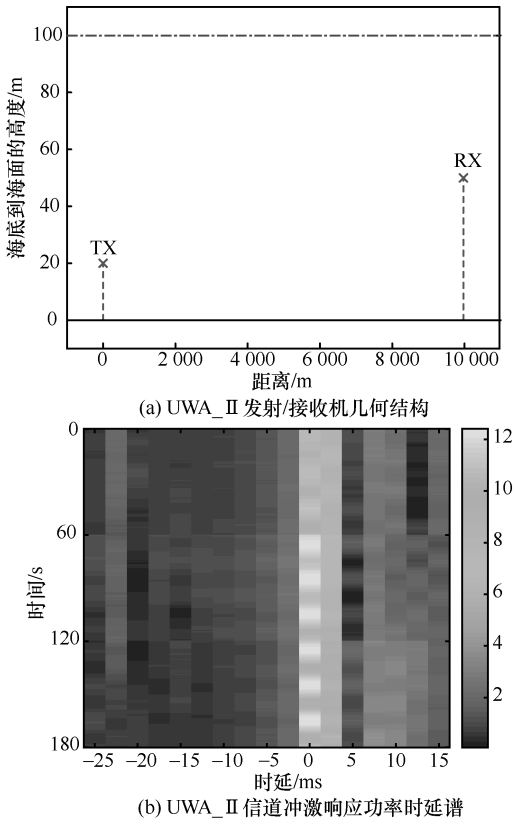
图 14 展示了 MC_CS_DCSK_I 在两种不同水声信道下的 BER 性能，并比较了在不同水声环境下循环前缀的对系统性能的影响。图 14(a)展示了 MC-CS-DCSK-I 在 UWA_I 信道下的 BER 性能，其中参数设置为 N=SF=128、256、512、1 024，CP=0、T_s/8。从图 14 中可以看到，系统 BER 性能首先随着子载波数的增加而提升，然后随着子载波数的增加而降低。这是因为当子载波数较小的时候，随着子载波的增加，OFDM 符号周期变长，系统抗符号间干扰能力增强，由此性能提升。而当子载波较多时，如 N=512，此时子载波带宽为 19.5 Hz，



(a) UWA_I发射/接收机几何结构

(b) UWA_I信道冲激响应功率时延谱

图 12 UWA_I 水声信道几何结构与冲激响应图



(a) UWA_Ⅱ发射/接收机几何结构

(b) UWA_Ⅱ信道冲激响应功率时延谱

图 13 UWA_Ⅱ水声信道几何结构与冲激响应图

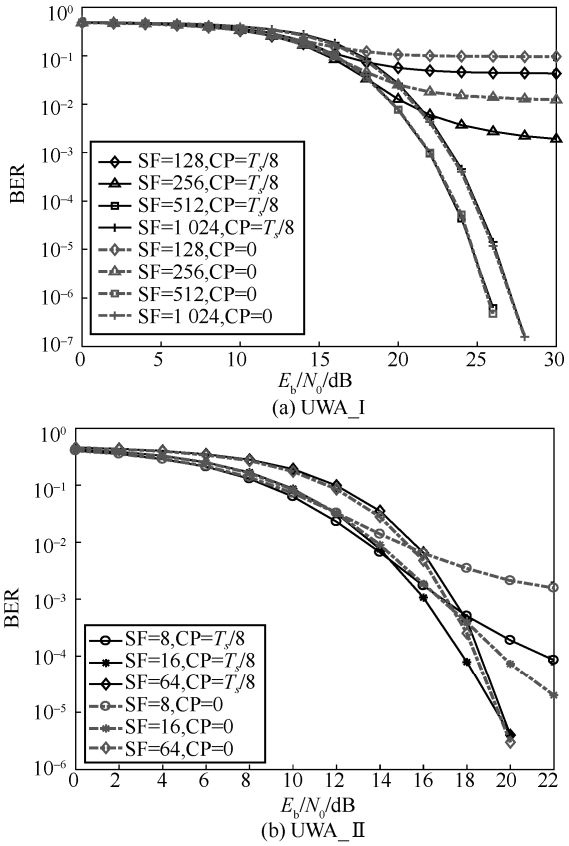


图 14　MC_CS_DCSK_I 在不同水声信道下的 BER 性能

而 UWA_I 信道的相干带宽在 28 Hz 左右，此时，系统已具备良好的抗 ISI 能力，再增大子载波数，反而导致接收端噪声的增加，降低系统性能。此外，从图中可以看到，当 N=512 时，系统展现出优秀的 BER 性能，当 SNR=26 dB 时，系统 BER 曲线达到 10^{-7} 级别。因此对于 MC-CS-DCSK-I，选取合适的子载波数有助于性能优化。从图 14 中还可以看到，当子载波数较小的时候，如 N=128、256，加循环前缀的系统 BER 性能要优于不加循环前缀的系统。但当子载波数较大时，如 N=512、1 024 时，加不加循环前缀对系统 BER 性能基本没有影响，因为此时系统已具备良好的抗 ISI 能力。

图 14（b）展现了 MC-CS-DCSK-I 在 UWA_Ⅱ水声信道中的 BER 性能，其中参数设置为 N=SF=8、16、64，CP=0、T_s/8。相比于图 14（a），图 14（b）在子载波数较小的时候即可获得较好的 BER 性能，这是由于 UWA_Ⅱ信道是时延较小的水声信道，即使子载波数较小时，ISI 的影响也不大。从图 14（b）中可以看到，系统性能也呈现先随着子载波数的增加而提高，然后随着子载波数的增加而降低，因此选择一个合适的子载波数有助于提升系统性能。此外，从图 14（b）中同样可以看到，当子载波数较多时，当 N=64 时，系统不加循环前缀的 BER 曲线与加循环前缀的 BER 曲线基本重合，当 N=16 时，不加循环前缀的系统性能虽低于加循环前缀系统的性能，但当 SNR=22 dB 时，不加循环前缀的系统 BER 曲线仍可达 10^{-5} 级别，展现出良好的性能。

图 15 展示了 UWA_I 水声信道下，MC-CS-DCSK-Ⅱ的 BER 性能曲线，并与 MC-DSSS 系统进行了比较。图 15 中参数为 N=128、256、512，SF=32，CP=T_s/8。从图 15 中可以看出，MC-CS-DCSK-Ⅱ与 MC-DSSS 的 BER 曲线都随着子载波数的增大而降低，这是由于随着子载波数的增大，OFDM 符号周期变长，信号抗 ISI 能力增加，系统性能提升。但同时从图 15 中可以看到，MC-CS-DCSK-Ⅱ优于 MC-DSSS，在 N=512、SNR=22 dB 时，系统 BER 曲线可到达 10^{-6}。

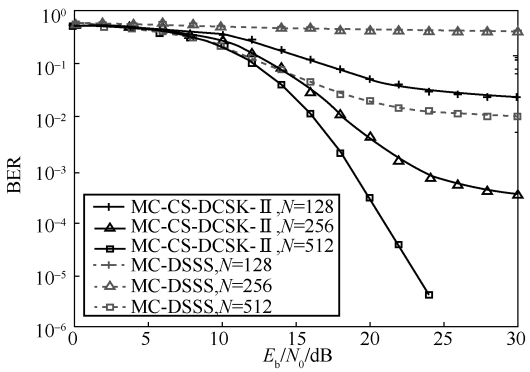


图 15　UWA_I 信道下，MC-CS-DCSK-Ⅱ与 MC-DSSS 的性能比较

图 16 讨论了在 UWA_I 信道下，循环前缀对 MC-CS-DCSK-Ⅱ的影响，图 16 中参数为 N=128、256、512，SF=32，CP=T_s/8、0。从图 16 中可以看到，加了循环前缀的系统 BER 性能优于没有加循环前缀的系统。但从图 15 中还可以看到，当子

载波数比较大时，不加循环前缀的系统 BER 性能仍可达到一个较优的范围。如 N=512、SNR=22 dB 时，不加循环前缀的系统的 BER 曲线可达到 10^{-5} 级别，展现出优秀的性能。因此，在某些传输场景的水声信道中，可以考虑不加循环前缀以提高系统传输效率与能量效率。

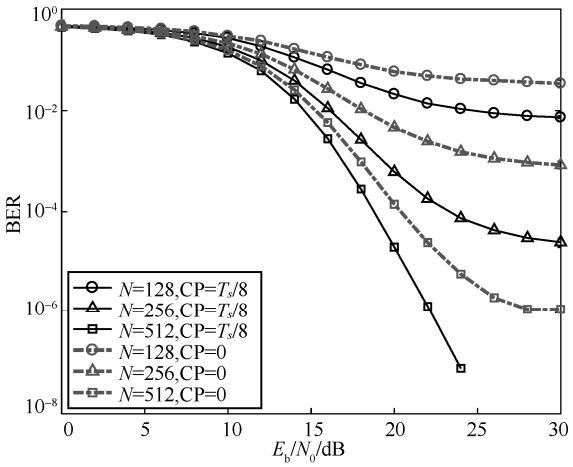


图 16　UWA_I 水声信道下，循环前缀对 MC-CS-DCSK-Ⅱ性能的影响

图 17 讨论了在 UWA_I 信道下，循环位移交织器对 MC-CS-DCSK-Ⅱ的影响。图 17 中参数为 N=128、256、512，SF=32，CP=T_s/8，P=0、1，其中，P=1 表示系统使用了循环位移交织器，P=0 表示系统没有使用循环位移交织器。从图 16 中可以看到，使用了循环位移交织器的系统性能要远优于不使用交织器的系统。由此可认为循环位移交织器在水声信道中仍具备其优秀的特性，增强系统在水声信道中的稳健性。

5　结束语

本文首先对水声信道下的调制解调技术进行了回顾，重点介绍了正交频分复用、扩展频谱技术以及混沌调制技术在水声通信中的研究进展，然后引出非相干调制对于稳健水声通信的重要性。作为一种非相干接收的混沌调制技术，差分混沌移位键控在多径衰落信道下具有优良的表现。但在严重时–频双扩展的水声信道下，直接采用差分混沌移位键控，并不能很好地适应新的信道环境，为此，本文设计了两种基于多载波的码复用差分混沌移位键控调制方案，并在时–频双扩展信道和水声信道下对这两种调制进行研究，仿真结果证明所提出的两种方案具有良好的抗信道双扩展的性能，并且在水声信道下表现出优良的稳健性。

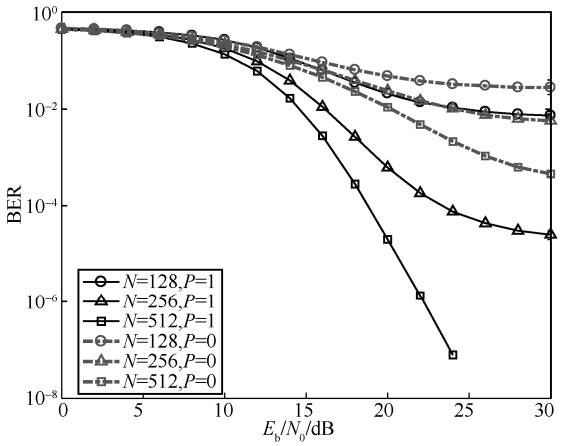


图 17　UWA_I 水声信道下，交织器对 MC-CS-DCSK-Ⅱ性能的影响

参考文献：

[1] Committee on Guidance for NSF on National Ocean Science Research Priorities, Decadal Survey of Ocean Sciences, Ocean Studies Board, et al. Sea change: 2015-2025 decadal survey of ocean sciences [R]. The National Academies Press, 2015.

[2] 徐文, 鄢社锋, 季飞, 等. 海洋信息获取、传输、处理及融合前沿研究评述[J]. 中国科学: 信息科学, 2016(46): 1053-1085.
XU W, YAN S F, JI F, et al. Overview of marine information acquisition, transmission, processing and fusion frontier research, SCIENTIA SINICA: Informationis, 2016 (46): 1053-1085.

[3] 程日涛, 张海涛, 王乐. 5G 无线网部署策略[J]. 电信科学, 2018, 34(S1): 1-8.
CHENG R T, ZHANG H T, WANG L. Deployment strategy of 5G wireless network[J]. Telecommunications Science, 2018, 34(S1): 1-8.

[4] 王祖阳, 杨传祥, 张进, 等. 5G 无线网技术特征及部署应对策略分析[J].电信科学, 2018, 34(S1): 9-16.
WANG Z Y, YANG C X, ZHANG J, et al. Analysis on technology characteristics and deployment strategies of 5G wireless network[J]. Telecommunications Science, 2018, 34(S1): 9-16.

[5] 王庆扬, 谢沛荣, 熊尚坤, 等. 5G 关键技术与标准综述[J].

电信科学, 2017, 33(11): 112-122.
WANG Q Y, XIE P R, XIONG S K, et al. Key technology and standardization progress for 5G[J]. Telecommunications Science, 2017, 33(11):112-122.

[6] 李渝舟, 江涛, 曹洋, 等. 5G 绿色超密集无线异构网络: 理念、技术及挑战[J]. 电信科学, 2017, 33(6): 34-40.
LI Y Z, JIANG T, CAO Y, et al. Green 5G ultra-dense wireless heterogeneous networks: guidelines，techniques, and challenges[J]. Telecommunications Science, 2017, 33(6): 34-40.

[7] STOJANOVIC M. Recent advances in high-speed underwater acoustic communications[J]. IEEE Journal Oceanic Engineering, 1996, 21(2): 125-136.

[8] CCHEN K, MA M, CHENG E, et al. A survey on mac protocols for underwater wireless sensor networks[J]. IEEE Communications Surveys & Tutorials, 2014, 16(3): 1433-1447.

[9] STOJANOVIC M, PREISIG J. Underwater acoustic communication channels: propagation models and statistical characterization[J]. IEEE Communications Magazine, 2009, 47(1): 84-89.

[10] HEARN P. Underwater acoustic telemetry[J]. IEEE Transaction on Communications Technology, 1966, 14(6): 839-843.

[11] CATIPOVIC J, DEFFENBAUGH M, FREITAG L, et al. An acoustic telemetry system for deep ocean mooring data acquisition and control[C]//IEEE Process Oceans, Sep 18-21, 1989, Seattle, USA. Piscataway: IEEE Press, 1989: 887-892.

[12] STOJNOVIC M, PROAKIS J G, RICE J A, et al. Spread spectrum underwater acoustic telemetry[C]// IEEE OCEANS' 98 Conference Process, Sep 28, 1998, Nice, France. Piscataway: IEEE Press, 1998: 650-654.

[13] COATELAN S, GLAVIEUX A. Design and test of a multicarrier transmission system on the shallow water acoustic channel[C]//IEEE OCEANS'94 Conference Process, Sep 13-16, 1994, Brest, France. Piscataway: IEEE Press, 1994: 472-477.

[14] COATELAN S, GLAVIEUX A. Design and test of a coding OFDM system on the shallow water acoustic channel[C]//IEEE OCEANS' 95 Conference Process, Oct 9-12, 1995, San Diego, USA. Piscataway: IEEE Press, 1995: 2065-2070.

[15] ARMSTRONG J, GRANT P M, POVEY G. Polynomial cancellation coding of OFDM to reduce intercarrier interference due to Doppler spread[C]//IEEE Global Telecommunications Conference, Nov 8-12, 1998, Sydney, Australia. Piscataway: IEEE Press, 1998: 2771-2776.

[16] KIM B C, LU I T. Parameter studies of OFDM underwater communications systems[C]//Process MTS/IEEE Oceans, Sep 11-14, 2000, Providence, USA. [S.l: s.n], 2000: 1251-1255.

[17] LINNARTZ J P M G, GOROKHOV A. New equalization approach for OFDM over dispersive and rapidly time varying channel[C]// IEEE International Symposium on Personal, Sep 18-21, 2000, London, UK. Piscataway: IEEE Press, 2000: 1375-1379.

[18] SCHNITER P. Low-complexity equalization of OFDM in doubly selective channels[J]. IEEE Transaction on Signal Processing, 2004, 52(4): 1002-1011.

[19] QU F, YANG L. Basis expansion model for underwater acoustic channels[C]// OCEANS 2008, Sep 15-18, 2008, Quebec City, Canada. Piscataway: IEEE Press, 2008: 1-7.

[20] WEN M, CHENG X, YANG L, et al. Index modulated OFDM for underwater acoustic communications[J]. IEEE Communications Magazine, 2016, 54(5): 132-137.

[21] HEC, ZHANG Q, HUANG J. Passive time reversal communication with cyclic shift keying over underwater acoustic channels[J]. Applied Acoustics, 2015(96): 132-138.

[22] 殷敬伟, 惠俊英, 王逸林, 等. *M* 元混沌扩频多通道 Pattern 时延差编码水声通信[J]. 物理学报, 2007(10): 5915-5921.
YIN J W, HUI J Y，WANG Y L, et al. *M*-ary chaotic spread spectrum Pattern time delay shift coding scheme for multichannel underwater acoustic communication[J]. Journal of Physics, 2007(10): 5915-5921.

[23] 尹艳玲, 周锋, 乔钢, 等. 正交多载波 *M* 元循环移位键控扩频水声通信. 物理学报[J], 2013(23): 224302-223302.
YIN Y L, ZHOU F, QIAO G, et al. Orthogonal multicarrier *M*-ary cycle shift keying spread spectrum underwater acoustic communication[J]. Journal of Physics, 2013(23): 224302-223302.

[24] STOJANOVIC M, FREITAG L. Hypothesis-feedback equalization for direct sequence spread spectrum underwater communications[C]//CEANS 2000 MTS/IEEE Conference Exhibition, Sep 11-14, 2000, Providence, USA. Piscataway: IEEE Press, 2000: 123-129.

[25] SOZER E M, PROAKIS J G, STOJANOVIC M, et al. Direct sequence spread spectrum based modem for underwater acoustic communication and channel measurements[C]//OCEANS 1999 MTS/IEEE Conference Exhibition, Sep 13-16, 1999, Seattle, USA. Piscataway: IEEE Press, 1999: 228-233.

[26] ZHANG H J, LEHNERT J S. Performance of multicarrier DS/SSMA over fast Rayleigh fading channels with Doppler diversity[J]. IEEE Transaction on Communications, 2006, 54(2): 273-283.

[27] KADOUS T A, SAYEED A M. An integrated framework for MC-CDMA reception in the presence of frequency offsets, phase noise, and fast fading[J]. IEEE Transactions on Wireless Communications, 2004, 3(4): 1224-1235.

[28] KONDO S, MILSTEIN L B. Performance of multicarrier DS CDMA systems[J]. IEEE Transactions on Communications, 1996, 44(2): 238-246.

[29] KONSTANTAKOS D P, ADAMS A E, SHARIF B S. Multicarrier code division multiple access (MC-CDMA) technique for underwater acoustic communication networks using short spreading sequences[J]. IEEE Proceedings - Radar, Sonar and Navigation, 2004, 151(4): 231-239..

[30] KUTIAN A P, PUTHUSSERYPADY S, HTUT S M. Performance enhancement of DS/CDMA system using chaotic com-

plex spreading sequence[J]. IEEE Transaction on Wireless Communications, 2005, 4(3): 984-989.

[31] LEI L, XU F, XU Y, et al. A chaotic direct sequence spread spectrum communication system in shallow water[C]//Process 2011 International Conference on Control, Automation and Systems Engineering (CASE), July 30-31, 2011, Singapore. Piscataway: IEEE Press, 2011: 1-4.

[32] 舒秀军, 王海斌, 汪俊, 等. 一种多通道混沌调相扩频方式及其在水声通信中的应用[J]. 声学学报，2017（2）:159-168 SHU X J, WANG H B, WANG J, et al. A method of multichannel chaotic phase modulation spread spectrum and its application in underwater acoustic communication[J]. ACTA ACUSTICA, 2017(2):159-168

[33] BAI C, REN H P, LI J. A differential chaos-shift keying scheme based on hybrid system for underwater acoustic communication [C]//Process 2016 IEEE/OES China Ocean Acoustics (COA), Jan 9-11, 2016, Harbin, China. Piscataway: IEEE Press, 2016: 1-5.

[34] BAI C, REN H P, GREBOGI C, BAPTISTA M. Chaos-based underwater communication with arbitrary transducers and bandwidth [J]. Applied Sciences, 2018(3): 1-11.

[35] CHEN M, XU W, WANG D, er al . Design of a multi-carrier different chaos shift keying communication system in doubly selective fading channels[C]//Process the 23rd Asia-Pacific Conference on Communication (APCC), Dec 11-13, 2017, Perth, Australia. Piscataway: IEEE Press, 2017: 1-6.

[36] CHEN M, XU W, WANG D, et al. A multi-carrier chaotic communication scheme for underwater acoustic communications [J]. IET Communications, 2019(14): 2097-2105.

[37] XU W, WANG L, KOLUMBAN G. A novel differential chaos shift keying modulation scheme[J]. International Journal of Bifurcations and Chaos, 2011, 21(3): 799-814.

[38] WANG S, ZHANG Z. Multicarrier chaotic communications in multipath fading channels without channel estimation[J]. Aip Advances, 2015, 5(1): 711-731.

[39] QARABAQ P, STOJANOVIC M. Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels[J]. IEEE Journal Oceanic Engineering, 2013, 38(4): 701-717.

[作者简介]

代红英（1977− ），女，重庆工程学院副教授，主要研究方向为通信技术。

陈梦蕾（1995− ），女，国家电网浙江省电力有限公司信息通信分公司工程师，主要研究方向为无线通信。

徐位凯（1976− ），男，博士，厦门大学信息与通信工程系副教授，主要研究方向为无线通信、水声通信。

研究与开发

GreyFan：一种 Wi-Fi 隐蔽信道攻击框架

马栋捷[1,2]，金成强[1,2]，陈园[1,2]，陈铁明[1,2,3]
（1. 浙江工业大学计算机科学与技术学院，浙江 杭州 310023；
2. 浙江省网络空间安全创新研究中心，浙江 杭州 310023；
3. 之江实验室工业互联网研究中心，浙江 杭州 311100）

摘　要：针对现实网络中诸如侧信道攻击、HID 攻击等传统的物理攻击，物理隔离被认为是一种较为彻底的抵御网络攻击的安全防护手段。2018 年，业界首次提出了一种物理隔离环境下的 Wi-Fi 隐蔽信道方法——Ghost Tunnel，即在 Wi-Fi 尚未连接的状态下，无线 AP 可成功将数据传给发起连接请求的计算机。提出了一种基于 Ghost Tunnel 方法的攻击框架—— GreyFan，利用该攻击框架攻击者可以对未连接 Wi-Fi 的用户实施无感知攻击，如文件隐蔽传输、任意代码执行等，并分析了相应的防御技术。

关键词：HID 攻击；网络隔离；隐蔽信道；GreyFan 攻击
中图分类号：TP309
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019179

GreyFan: a network attack framework using Wi-Fi covert channel

MA Dongjie[1,2], JIN Chengqiang[1,2], CHEN Yuan[1,2], CHEN Tieming[1,2,3]
1. College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China
2. Zhejiang Innovation Center of Cyberspace Security, Hangzhou 310023, China
3. Research Center of Industrial Internet, Zhejiang Lab, Hangzhou 311100, China

Abstract: For traditional physical attacks such as side channel attacks and HID attacks in real networks, physical isolation is considered to be a relatively complete security protection against network attacks. In 2018, a Wi-Fi hidden channel method in the physical isolation environment——Ghost Tunnel was firstly proposed, that is, in the state that Wi-Fi didn't connected, the wireless AP could successfully transmit data to the computer that initiates the connection request. An attack framework based on the Ghost Tunnel method——GreyFan was proposed. This attack framework enabled attackers to implement non-aware attacks on users who didn't connected to Wi-Fi, such as file concealed transmission and arbitrary code execution, etc. The corresponding defense technology was also analyzed.

Key words: HID attack, network isolation, covert channel, GreyFan attack

收稿日期：2019−05−24；修回日期：2019−07−03
基金项目：国家自然科学基金资助项目（No.61202282，No.61772026）；国家自然科学基金与浙江省政府联合项目（No.U1509214）
Foundation Items: The National Natural Science Foundation of China (No.61202282, No.61772026), The Joint Project of National Natural Science Foundation and Zhejiang Provincial Government (No.U1509214)

1 引言

隐蔽信道是一种违反安全策略的秘密传输信息的机制。它将数据报作为通信载体，将指定数据嵌入数据报文中，达到指定数据在网络中秘密输出而不被发现的目的。利用隐蔽信道，攻击者不仅可以将目标主机中的数据信息传送至攻击机获取情报，还可以传送控制命令，控制目标主机。它不仅会对网络数据的安全造成危害，甚至还会损害计算机硬件[1]。

基于隐蔽信道的网络攻击不需要复杂的计算机环境和操作条件，仅需要目标主机感染恶意客户端，即可将指定数据根据已经构建好的协议嵌入载体数据中，并与正常通信的流量混杂在一起，使其难以区分，从而隐藏通信通道。攻击机收到数据后进行响应，将嵌入的数据从伪装的数据报中分离出来，提取其中的关键数据。通过隐蔽信道进行攻击的方式往往让目标主机使用者难以察觉，隐蔽性极高[2]。

随着对现代隐蔽信道技术研究的不断深入和反隐蔽信道攻击技术的持续发展，使用仅由 ICMP、HTTP 等协议构建的隐蔽信道作为攻击框架核心已无法满足实际应用中对具有高隐蔽性、高安全性、高效率性的隐蔽信道攻击框架的迫切需求。因此，本文构建了一种隐蔽信道攻击框架——GreyFan 攻击框架。在架构上，该框架以 Ghost Tunnel 理论为基础，根据使用目的、使用对象、使用场景的不同，定义了全新的传输帧格式，设计了全新的传输协议，不仅保证了数据在传输过程中具有较好的稳定性和可靠性，同时还有效地将传输的数据进行了模块化划分，使各模块具有较好的规范性，使框架具有较好的扩展性；在功能上，利用该框架能够在不确定主机是否成功建立网络连接的情况下建立长连接的隐蔽信道，拥有远程控制、文件传输等功能，并可自主选择加密方式、发送分组初始量、发送分组间隔时间等，同时预留了异常检测逃避、系统实时监控等常用接口。

本文首先介绍了 GreyFan 攻击框架的研究背景，其次阐述了 GreyFan 攻击框架的基本结构、组成和功能，接着通过实验对该框架的可靠性、稳定性、通用性和时效性进行了验证，最后提出了对该种攻击框架的检测方法并进行总结。

2 相关工作

2.1 国内外对隐蔽信道的研究现状

隐蔽信道作为攻击框架的核心，它的构建使攻击框架的设计成为了可能。目前，国内外对于隐蔽信道的研究主要集中在从网络应用服务、网络协议和移动网络中寻找构建的可能性。

在网络应用服务方面，王娟等人[3]提出了一种基于浏览器帮助对象（browser helper object，BHO）构建网络隐蔽信道的方法，这种方法不仅能躲避杀毒软件和防火墙，还能为流量隐蔽和代理穿透提供有利条件。

在信道建立方法以及网络协议方面，Ahmadzadeh 等人[4]提出了利用差分通信信道导出公式来建立网络隐蔽通道的新方法；姬国珍等人[5]实现了基于 ICMP 数据分组时间间隔的时分型隐蔽信道；李卫等人[6]针对网络地址转换情况下信息传输行为和内容的强隐蔽需求，提出了一种适用于 NAT 环境的隐蔽信道构建方法；朱越凡等人[7]利用互联网不可或缺的 NTP（network time protocol，网络时间协议），设计了下行通道和上行通道分离的 NTP 隐蔽信道，提出了基于 NTP 的隐蔽信道构建机制。

在移动网络方面，Tan 等人[8]针对 LTE 视频流提出了一种基于 VoLTE 的构建隐蔽定时信道的方法，该方法通过故意丢弃视频分组来调制屏蔽信息，进而构建隐蔽定时信道；Yang 等人[9]提出现有移动系统可能会因受到恶意软件的攻击，而被迫改变无线网络接口的阻抗，从而迫使移动设备反射周围的 RF 信号，达到建立隐蔽信道传达信息的目的。Schulz 等人[10]在智能手机上的 Wi-Fi 中构建隐蔽通道，通过在传输之前预先过滤 Wi-Fi

帧，利用信道状态信息（channel state information，CSI）在接收器侧提取嵌入信息。

这些隐蔽信道虽然具有隐蔽性强、传输效果好、灵活性高等特点，但是在物理隔离的条件下，这些隐蔽信道均无法正常使用。而目前，对在物理隔离环境下的隐蔽信道研究还基本处于空白。

2.2　物理隔离环境下的攻击方式

物理隔离指的是计算机禁止直接或间接地接入公共网等非安全网络，它能够阻止重要计算机设备与非安全网络发生物理层面的接触。一般而言，在物理隔离环境下，计算机无法通过互联网被注入病毒或木马，避免了数据泄露或设备瘫痪等安全隐患，一定程度上提高了安全性。

然而，针对物理隔离环境同样也存在一些特别的攻击手段。U 盘攻击是一种将恶意代码存放于 U 盘的固件中，利用 U 盘在设备上的插拔使计算机设备被感染恶意代码的攻击方式；CD/DVD 攻击指的是利用 CD/DVD 等介质对 CD/DVD 的刻录软件的感染，从而完成对光盘类介质的感染，使光盘成为针对隔离网络的攻击工具和感染途径；水坑攻击是一种通过将恶意软件伪装成系统补丁、软件工具等常用软件，并利用社会工程学诱导计算机工程师下载、安装至物理隔离环境中的计算机，达到攻击目的的攻击方式；HID 攻击是一种将包含恶意脚本的 USB 设备模拟成人机交互设备，使计算机会由于错误地将其识别为人机交互设备而导致执行攻击脚本的攻击方式。

这些针对物理隔离环境下的攻击方式虽然隐蔽性好，但对恶意软件在目标主机上的维护则较为困难，灵活性较差。

2.3　Ghost Tunnel 隐蔽信道

在 2018 年年初，一种 Wi-Fi 物理隔离环境下的隐蔽通信技术——Ghost Tunnel 首次在 HITB 国际会议上被提出，在 Wi-Fi 尚未连接的状态下，无线接入点可将数据传给发起连接请求的计算机。

攻击者首先建立一个具有特殊 SSID（service set identifier，服务集标识）的恶意无线接入点，并向周边发送 Beacon 广播帧宣告本网络的存在；当被感染了恶意软件的目标主机打开无线网络时，原本用于扫描所在区域内是否存在 IEEE 802.11 网络的 probe request 帧，可携带具有特殊含义的数据进行发送；当无线接入点收到客户端的请求帧后，原本用于对客户端的 SSID 探测请求进行应答的 probe response 帧，也可携带指定数据进行发送，由此构成的 Wi-Fi 隐蔽传输通道就称作 Ghost Tunnel。

Ghost Tunnel 技术在理论上为物理隔离环境下构建隐蔽传输信道或隐蔽操控目标主机提供了一种新的途径。但在实际环境中，Ghost Tunnel 技术无法解决复杂的电磁环境干扰带来的分组丢失、出错等问题；此外，Ghost Tunnel 技术未对传输的内容进行细节化、模块化，导致无法在其基础上扩展出新的功能。

3　GreyFan 攻击框架

基于现有的隐蔽信道传输技术，结合 Ghost Tunnel 技术，本文构建了一种名为 GreyFan 的隐蔽信道攻击框架，而利用该框架进行的攻击被称为 GreyFan 攻击。该框架可分为三层结构，框架结构如图 1 所示。第一层为 GreyFan 隐蔽信道模型，用于在攻击者与目标主机之间建立稳定的长连接；第二层为传输帧结构，用于规范不同应用下的传输帧所携带的具体内容，方便应用层功能的添加与修改；第三层为应用层，用于提供具体的功能，该框架可自主选择传输加密方式、发送分组初始量、发送分组间隔时间等，并预留了异常检测逃避、系统实时监控等接口，方便后续开发。目前，该框架可实现远程控制目标主机、向目标主机传输文件等功能。

图 1　GreyFan 攻击框架结构

3.1 GreyFan 隐蔽信道模型

GreyFan 隐蔽信道模型为在 Wi-Fi 物理隔离的条件下建立稳定的隐蔽信道提供了基础。在 Wi-Fi 建立阶段，无线接入点与客户端均采用 IEEE 802.11 协议中的管理帧进行通信，管理帧主体为长度固定的字段和长度不定的信息元素。正是因为信息元素部分的内容可变，因此给隐蔽信道的建立提供了可能。

信息元素通常包含 3 个部分：元素 ID 字段、信息长度字段以及信息内容字段[11]，如图 2 所示。元素 ID 字段用于标识每一种信息元素的类型，信息长度字段标识第三部分的长度，信息内容字段携带具体的信息。GreyFan 攻击在不建立完整 Wi-Fi 连接的情况下，使 probe request 帧和 probe response 帧携带指定 ID 的冗余信息元素，从而建立隐蔽通信通道。

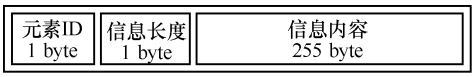


图 2 管理帧信息元素结构

攻击者与目标主机之间的基本通信流程如图 3 所示。目标主机在感染恶意程序后，无论其是否连接网络，均会主动发送携带含有特定元素 ID 的冗余信息元素的 probe request 帧；攻击者在接收到 probe request 帧后，解析其 ID 字段及具体信息内容，判定目标主机是否处于准备状态，并发送携带由恶意指令组成的冗余信息元素的 probe response 帧；目标主机在收到 probe response 帧后，解析其 ID 字段及具体信息内容，并根据内容执行后续操作。

然而在现实环境下，电磁环境复杂，容易导致数据分组的丢失，造成数据传输失败。因此，被控端在发送 probe request 帧时，帧内需要携带上一次收到 response 帧的时间的散列值；而控制端在发送 probe response 帧时，帧内需要携带本次帧发送时间的散列值。

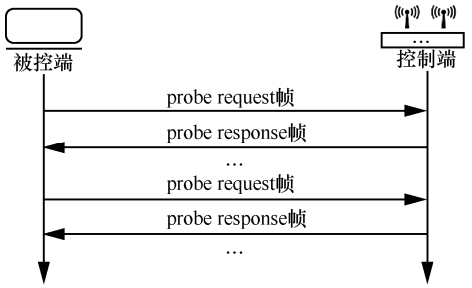


图 3 攻击者与目标主机之间的基本通信流程

此外，计算机存在缓存机制，会导致重复接收数据，因此被控端在解析信息内容后，将缓存该信息元素的散列值，当再次收到 probe response 帧时，将校验该帧内信息元素的散列值是否与缓存下的散列值相同，若相同，则不执行相关操作。

3.2 传输帧结构

在建立隐蔽信道通信的过程中，传输帧携带的内容不仅关系到传输效率，还会影响传输的可靠性。而在不同的应用场景下，传输帧需要携带的信息也不尽相同。因此，为了平衡传输的效率和可靠性，同时为应用层提供一种可复用的帧结构、简化应用层的扩展难度，根据帧工作性质的不同，将帧分为指令传输帧及文件传输帧。

指令传输帧的帧信息元素结构如图 4 所示。它将 255 byte 的冗余信息元素分为 4 个部分：第一部分为特殊字段，占 3 byte，用于标识该帧用于执行命令；第二部分为散列值字段，占 8 byte，用于可靠传输；第三部分为具体命令信息字段，占 240 byte，用于携带需要传输的具体命令；第四部分为保留字段，占 4 byte，用于未来扩展及优化。

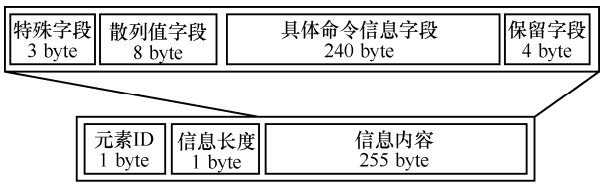


图 4 指令传输帧格式

文件传输帧的帧信息结构如图 5 所示，它将 255 byte 的冗余信息元素分为 7 个部分，各部分含义及用途见表 1。

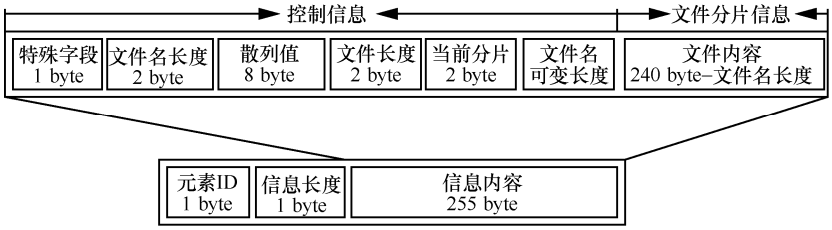


图 5　文件传输帧格式

表 1　文件传输帧功能解释

字节数/byte	含义	用途
1	特殊字段	标识该帧用于传输文件
2	文件名长度	标识文件名的长度，用 16 进制表示
8	散列值	用于可靠传输
2	文件总长度	标识文件的总分片数，用 16 进制表示
2	当前分片序号	标识当前分片所属的位置，用 16 进制表示
0～240	文件名	标识文件的保存名称，可带具体路径
0～240－文件名长度	文件的具体内容	发送具体的文件信息

3.3　远程控制及文件传输

远程控制和文件传输是 GreyFan 攻击框架应用层中的具体应用。远程控制指的是无论用户是否建立网络连接，攻击者均可无感知地操控目标主机，进行执行脚本、开关计算机、修改计算机配置文件等敏感操作。文件传输指的是无论用户是否建立网络连接，攻击者均可将本地文件无感知地传输至目标主机的指定目录下。

在远程控制应用中，被控端与控制端均发送指令传输帧，完成一次隐蔽远程控制，被控端系统流程如图 6 所示，控制端系统流程如图 7 所示。而在文件传输应用中，被控端发送指令传输帧，

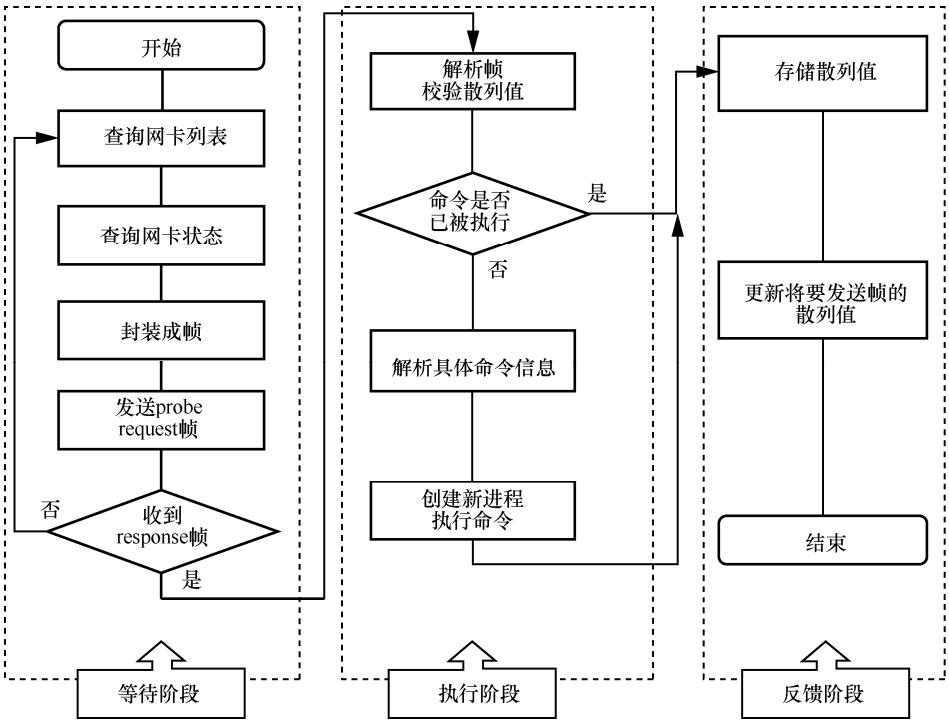


图 6　远程控制时被控端系统流程

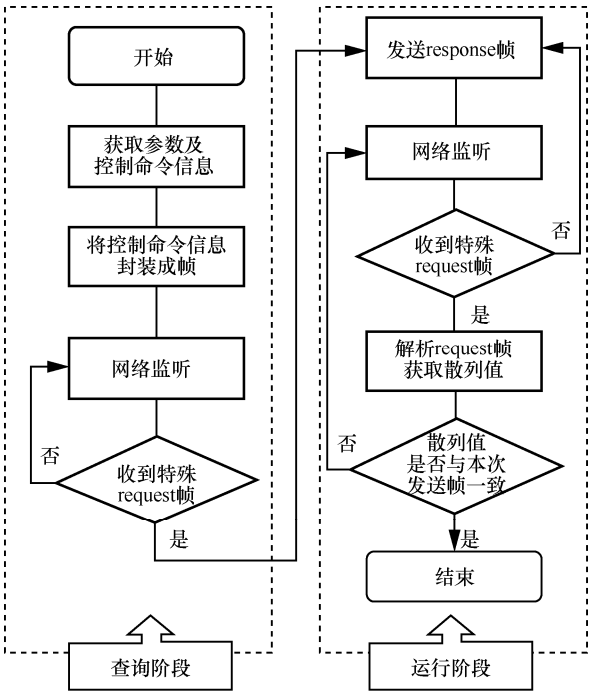


图 7　远程控制时控制端系统流程

控制端发送文件传输帧，完成一次隐蔽文件传输，被控端系统流程如图 8 所示，控制端系统流程如图 9 所示。

对于被控端而言，完成一次 GreyFan 攻击，需经历等待、执行和反馈 3 个阶段。在等待阶段，被控端调用系统 Wi-Fi 模块资源，持续发送包含特定信息头的 probe request 帧，并等待 probe response 帧，一旦收到，则进入执行阶段，否则依然处于等待阶段。在执行阶段，若进行远程控制，则解析帧内容，并创建进程，执行命令；若进行文件传输，则解析帧内容，并根据控制信息内的文件名字段按指定的保存路径保存文件信息。在反馈阶段，被控端需要将 probe request 帧内携带的时间散列字段更新为最新收到的 probe response 帧的时间散列值，若进行的操作为文件传输，还需将具体命令信息字段更新为被控端进行解析、接收和存储后的结果代码，之后持续发送，等待下一次命令或下一个文件。

对于控制端而言，完成一次 GreyFan 攻击，需经历查询和运行两个阶段。在查询阶段，控制端启动网络监听功能，持续监听是否收到含有指定信息头的 probe request 帧，一旦收到，则进入

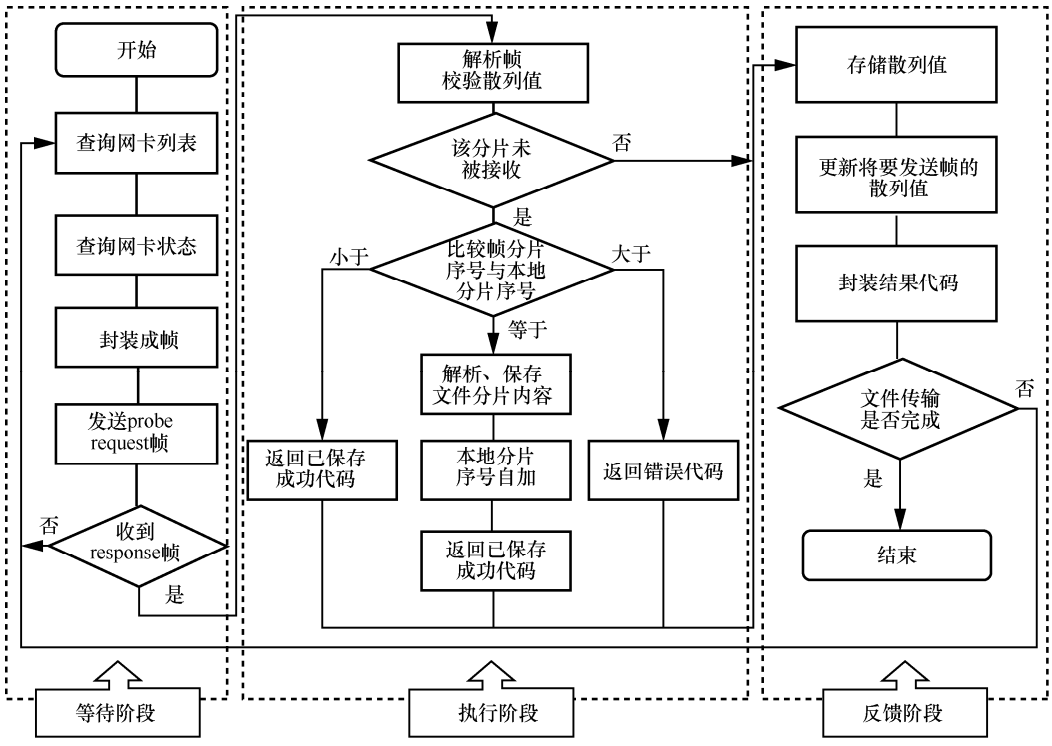


图 8　远程文件传输时被控端系统流程

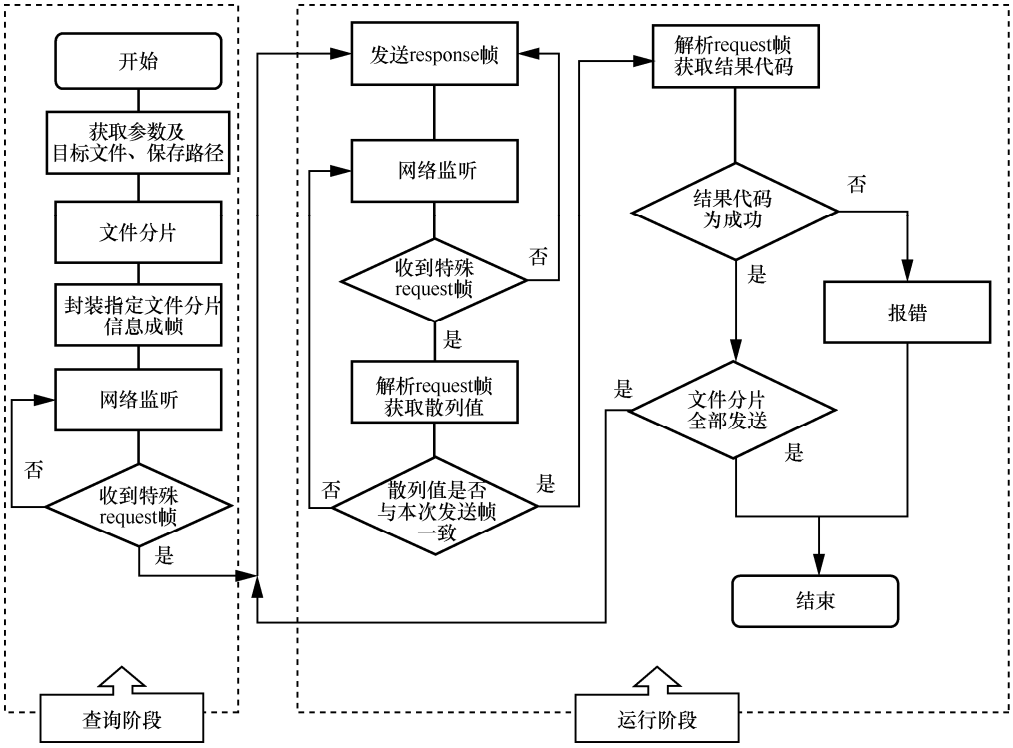


图 9　远程文件传输时控制端系统流程

运行阶段，否则依然处于查询阶段；同时，如果需要执行文件传输，则还需将用户指定发送的文件进行分片。在运行阶段，若进行远程控制，则发送携带当前时间散列值及用户远程控制命令的 probe response 帧；若进行文件传输，则发送携带控制信息及文件分片信息的 probe response 帧。之后等待被控端返回包含当前发送帧所携带的时间散列值的确认信息 probe request 帧。如果超过一定时间依旧未收到确认信息，则增加发送分组量重新发送。

3.4　扩展功能

由于应用 GreyFan 攻击框架的场景不同，攻击者需要一次性发送的分组量也会有所不同。为了提高框架应用的灵活性，GreyFan 攻击框架可自主选择发送分组的初始量以及超时时间。当 GreyFan 攻击框架在超时时间内未成功接收到来自于被控端的消息响应时，则判定被控端未正常接收指令或信息，此时，控制端会在其上一次发送分组量的基础上增加 100，并重新发送分组，以增加被控端成功接收的概率。

在文件传输中，由于外界干扰，传输的过程可能会被中断。因此，为了提高文件传输的可靠性及灵活性，GreyFan 攻击框架支持从断点续传功能，即当文件传输中断时，可从中断分片处或已完成传输分片处继续传输。

为了提高 GreyFan 攻击的隐蔽性，GreyFan 攻击框架支持自主选择是否加密及传输加密方式，目前 GreyFan 攻击框架支持的加密方式有 AES 加密和 RC4 加密。

为了使攻击者能够更好地使用 GreyFan 攻击，GreyFan 攻击框架预留了异常检测逃避、系统实时监控等接口，从而使攻击者能够更好地将该框架与其他系统对接。

4　实验与分析

本节针对远程控制和文件传输两种场景，分

别设计了多组对照实验，对 GreyFan 攻击框架的稳定性、时效性、可靠性和通用性进行了验证。

在硬件方面，控制端采用处理器为 Intel Core i5-3317U 的计算机，四核，主频为 1.7 GHz，内存 4 GB，华硕 UX32V 内置网卡，操作系统为 64 位 Kali 系统；被控端采用处理器为 Intel Core i7-6500U 的计算机，四核，主频为 2.5 GHz，内存 8 GB，Dell 灵越 15-5559 内置网卡，操作系统为 64 位的 Windows 7 系统。

在软件方面，控制端运行 GreyFan 攻击框架的控制应用程序，模拟攻击者；被控端运行更新至 2018 年 9 月最新病毒库的 360 杀毒软件、GreyFan 攻击框架的被控应用程序，模拟被恶意软件感染的目标主机。为了便于统计与分析，控制端与被控端详细的通信时间及过程分别打印至本地控制台。

为避免真实环境中其他网卡的干扰，实验环境为地下车库。

4.1 远程控制测试

在测试远程控制的稳定性、时效性和可靠性的实验中，将 Windows 环境下打开计算器的 cal 指令作为实验指令，将初始发送分组量设置为 500，将超时时间设置为 10 s，分别在控制端与被控端间隔为 10 m、15 m、20 m、25 m、30 m、40 m、45 m 的条件下进行实验，同一距离下测试 20 次。实验步骤如下。

在确定控制端和被控端的间距距离后，首先在被控端启动被控端应用程序，确定被控端已被激活。之后在控制端发送远程控制指令，并开始计时，直至被控端接收到指令并成功打开计算器后，记录下当前消耗时间，并将其作为被控端响应时间。由于 GreyFan 攻击框架要求被控端在执行完命令后需要返回确认信息给控制端，因此当控制端收到返回的信息后，停止计时，并将从被控端打开计算器到控制端接收到确认信息的时间作为控制端消耗时间。若控制端在传输过程中出现错误、被控端未成功打开计算器或杀毒软件提示异常，则视为控制失败。

图 10 展示了在相隔距离不同的条件下，去除最大最小值后，被控端响应及控制端消耗的平均时间及时间标准差。图 11 展示了在相隔距离不同的条件下，被控端响应时间及控制端消耗的时间离散图。

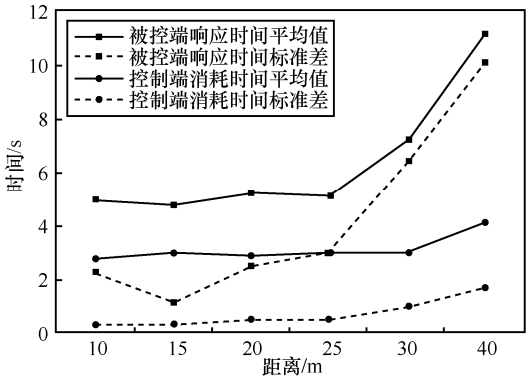
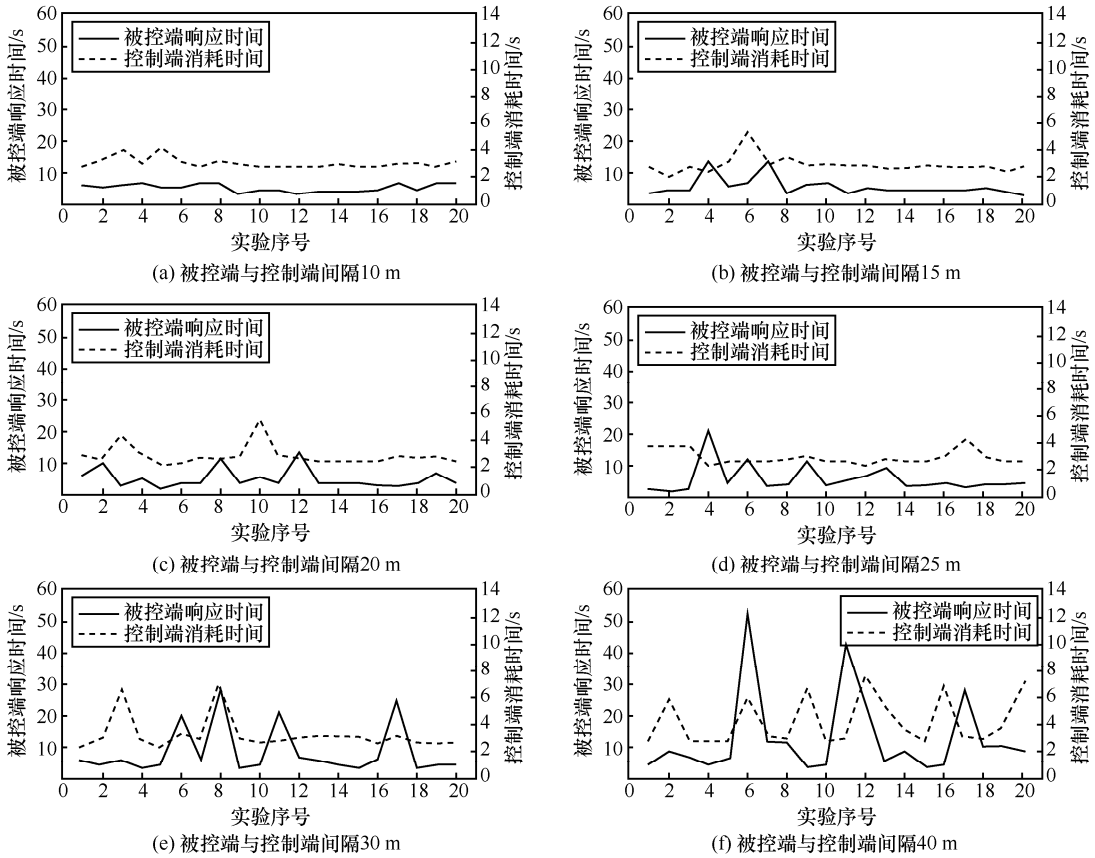


图 10 不同距离下被控端响应及控制端消耗时间平均值及标准差

从稳定性角度看，随着距离的逐渐增大，攻击的不稳定性逐渐上升。在实验中，当控制端与被控端之间的距离在 0~25 m 时，每次实验时的被控端响应时间和控制端消耗时间基本一致，说明此时 GreyFan 攻击框架具有较好的稳定性；在 25~45 m 时，被控端响应所消耗的时间和离散程度急剧上升，而控制端消耗的时间和离散程度也有所增加，说明此时攻击出现了不稳定性；当距离超过 45 m 时，攻击失效。

之所以随着距离的增大稳定性逐渐变差，是由于受到了实验双方内置网卡传输距离和计算机自身功率的限制：在距离较近的条件下，双方网卡均能够及时收到对方发送的 probe 帧，故每次测试消耗的时间基本一致，传输较为稳定。而当距离间隔较远时，实验双方可能会存在分组丢失的情况，然而由于 GreyFan 攻击框架具有可靠传输机制，当控制端发送控制命令后，一旦被控端丢失数据分组，则无法在规定时间内返回确认信息，控制端则会增加发送分组量，重新发送命令，因此被控端响应时间会急剧上升，其上升的幅度



(a) 被控端与控制端间隔10 m　(b) 被控端与控制端间隔15 m

(c) 被控端与控制端间隔20 m　(d) 被控端与控制端间隔25 m

(e) 被控端与控制端间隔30 m　(f) 被控端与控制端间隔40 m

图 11　不同距离下被控端响应时间及控制端消耗时间离散图

与设定的超时时间有关；而当被控端返回的确认信息丢失时，由于该信息封装在 request 帧中，是不断发送的，因此，控制端消耗时间只是略微增加。当双方距离超过网卡能够接受的距离后，无论控制端如何增加发送的数据分组，被控端均无法接收，故攻击失效。综上所述，在使用 GreyFan 攻击框架时，为了增加控制的稳定性，控制端应当尽可能地在距离被控端周围较近的位置进行控制。

从时效性角度看，若被控端能够及时响应控制端发送的控制命令，其平均时间为 5~6 s，若控制端能够及时接收确认信息，其平均时间为 2~3 s。因此在近距离情况下，完成一次远程控制工作需要 7~10 s。而在远距离情况下，被控端响应时间为 5~55 s，而控制端消耗时间为 2~8 s。

之所以在近距离情况下被控端响应时间会比控制端消耗时间长，是因为被控端在接收到控制端发送的 probe response 帧并解析后，还需要调用系统进程打开计算器，通过测算，该过程消耗的时间为 2~3 s。而控制端收到被控端发送的 probe request 帧确认信息并解析后，只需要进行一次判断运算，其消耗时间可忽略不计。因此，probe response 帧的实际传输时间与 probe resquest 帧的实际传输时间基本一致，为 2~3 s。

之所以在远距离情况下被控端响应时间远大于控制端消耗时间，是因为 GreyFan 攻击框架中的可靠传输机制。控制端重传的条件是当在设置的超时时间内未收到确认信息，则重传，而在本次实验中超时时间设置为 10 s，因此两次发送数据分组的间隔时间最小为 10 s，当多次在被控端响应阶段发生分组丢失情况时，被控端响应时间则成倍增加；而对于被控端发送的确认消息而言，

由于其确认消息被封装于 request 帧中，而该帧不断向外发送以表示本机已被激活，因此，由于两次发送帧的时间间隔短，当多次在控制端消耗阶段出现发送分组丢失时，控制端消耗时间仅略微增长。

之所以完成一次远程控制工作需要两倍的传输时间加执行命令时间，同样是因为 GreyFan 攻击框架中具有可靠传输机制，该机制为了提高传输的可靠性，要求控制端发送控制指令后，需要等待被控端返回的确认信息，否则将重新进行发送。因此牺牲了一定的传输效率。

从可靠性角度看，在间隔 10~40 m 的测试条件下，虽然执行的时间随着距离的不同不断波动，但是所有命令最终均可以成功执行，未出现控制失败的情况。而可靠性讨论的是被控端与控制端之间在能够正常通信的情况下完成预定功能的能力，故在 40 m 以外的条件下，被控端与控制端双方因受到计算机设备硬件的限制而无法攻击成功，其不属于可靠性的讨论范畴。因此，GreyFan 攻击框架虽然牺牲了传输效率，但是却拥有极高的攻击可靠性。

在测试远程控制的通用性实验中，被控端与控制端间隔 20 m，在控制端分别输入 Windows 环境下 15 大类共计 90 种命令，检测被控端能否正常执行以及杀毒软件是否能够检测出风险。若被控端未成功执行或杀毒软件检测出了病毒或进行了风险提示，则该命令记为不可执行。测试命令个数与可执行命令个数的统计情况如图 12 所示。

从通用性角度看，Windows 环境下所有的 DoS 指令均可以正常运行，其指令类别涵盖了复制操作、进程操作、目录操作、文件操作、时间操作、控制操作、环境变量操作、CMD 操作、网络操作、ACL 操作、计划操作、机器操作、驱动操作、硬盘操作和其他操作，其中部分命令甚至涉及一些计算机的敏感操作。但在执行这些命令时，计算机并未进行阻止，并且成功绕过了运行在被控端上的杀毒软件。因此，GreyFan 攻击框架具有通用性。

4.2 文件传输测试

在测试远程文件传输的实验中，在被控端与控制端间的距离间隔为 15 m 的条件下，将一个大小为 500 byte 的 txt 文件作为测试文件，初始发送分组量设置为 500，超时时间为 10 s，被控端存放文件的路径为“1.txt”，共测试 20 次，实验步骤如下。

首先在被控端启动被控端应用程序，确定被控端已被激活。之后在控制端发送远程文件传输指令，将本地的文件发送至被控端，并开始计时，直到控制端收到最后一次返回的确认信息并提示

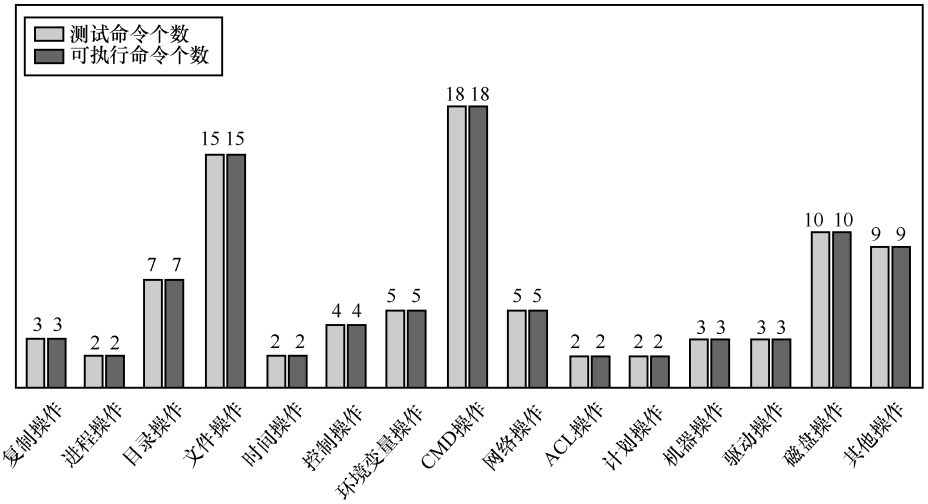


图 12 执行命令情况统计

完成传输后，停止计时，将这段时间记为远程传输文件的总消耗时间。若控制端在传输过程中出现错误、被控端接收的文件出现错误或杀毒软件提示异常，则视为传输失败。远程传输文件消耗时间及中断次数如图 13 所示。

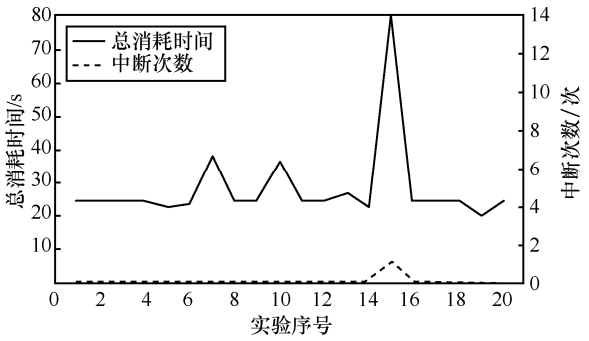


图 13　远程传输文件消耗时间时间及中断次数统计图

从时效性角度看，有 17 次传输测试花费的时间为 22~28 s，有 2 次传输测试花费了将近 40 s 的时间，有 1 次传输花费了将近 80 s 的时间。

之所以大部分测试的总耗时在 22~28 s，是因为在本实验中，一个文件传输帧可携带的具体内容字节数最高为 235 byte，因此 500 byte 的 txt 文件需要 3 个分片，即至少需要完成 3 次可靠传输，才能将文件发送至被控端。在控制端与被控端均能够良好地接收到数据分组的情况下，完成一次可靠传输需要 8 s 左右，因此，在预定的实验条件下，至少需要 24 s 左右的时间才能完成 500 byte 文件的传输。

之所以有少部分测试会有较高的总耗时，是因为在间隔为 15 m 的条件下，可能会出现被控端与控制端间通信数据分组的丢失，而由 GreyFan 攻击框架中的可靠传输机制可知，控制端判定一帧是否发送失败的条件是在超时时间内是否收到了被控端发送的确认信息，因此，在本实验中两帧的间隔时间至少为 10 s，这也导致了总消耗时间的增加。此外，由于 GreyFan 攻击框架支持断点续传，因此在第 15 次实验中发生中断故障后，需要人为再次启动命令程序，这势必会消耗大量的时间。

从可靠性角度看，除了第 15 次传输测试外，其余测试虽然可能发生了分组丢失情况，但在可靠机制的保护下，最终均完成了传输任务，保证了传输的可靠性。而在第 15 次传输的过程中发生了中断的情况，主要是因为在控制端在规定时间内未收到确认信息，下一次发送帧时需要增加发送分组量进行重新发送，然而发送 probe response 帧的过程需要占用较多的内存，故当发送分组量过大时，会导致内存不足，从而被迫中断。因此，虽然 GreyFan 攻击框架可以无限制地增大发送分组量以提高被控端接收的概率，但在实际情况下，由于受到缓存空间的限制，发送分组量是有上限的，而随着文件大小的增加，完成可靠传输的次数逐渐增加，因分组丢失而增加发送分组量的次数也随之增加，这也间接地增加了缓存的压力，提升了中断出错的概率。因此，文件传输在传输小型文件（如脚本文件）时，可靠性较高。

5　攻击检测与防御

GreyFan 攻击的关键在于对 IEEE 802.11 协议中管理帧的发送与接收，而对于操作系统系统和杀毒软件而言，它们并没有对其进行防御或检测，这也是被控应用程序能够在运行时成功绕过杀毒软件检测和系统防火墙防御的重要原因。因此，为了保护用户的个人财产和隐私不受到侵犯，在主机层面和网络层面提出了防御 GreyFan 攻击的方法。

5.1　主机层面

CPU 使用率指的是计算机中运行的程序所占用的 CPU 资源，表示计算机在某个时间点的程序运行情况。利用 GreyFan 攻击框架进行攻击的目的往往是窃取信息、破坏系统等，这势必会消耗系统的部分 CPU 资源。因此可检测 CPU 使用率，并重点关注使用率最高的前 3 个任务，当所属任务的分类属于报文接收和发送类任务时，可考虑采

取暂时关闭 Wi-Fi 模块等措施，达到防御的目的。

5.2 网络层面

WIPS（wireless intrusion prevention system，无线入侵防御系统）指的是能够监测无线范围内未授权的接入、并能自动采取应对措施的网络设备[13]。在网络、空间内部署该系统后，可利用该系统对区域内的无线网络环境进行实时监听。一旦识别出攻击行为，则通过多种渠道立即阻断对应的 2.4 GHz 和 5 GHz 的频段，同时记录下 Wi-Fi 设备的特征、攻击行为样本并报告管理员，达到拦截、阻断、取证、追踪的目的。

因此，可利用 WIPS 解析无线管理帧内容，验证其是否具有 GreyFan 攻击的特征信息[12]，如信息元素中的元素 ID 字段存在非特定内容等，从而达到识别 GreyFan 攻击行为的目的。

6 结束语

针对在物理隔离环境下建立隐蔽信道的需求，本文提出了一种 Wi-Fi 隐蔽信道攻击框架——GreyFan 攻击框架。实验测试表明，GreyFan 攻击框架能够执行的指令丰富，具有较强的通用性，且在半径为 25 m 的攻击范围内具有良好的可靠性和稳定性，适合在网络环境不佳的情况下传送较小的本地文件。对于 GreyFan 攻击，可以通过在主机层面监控计算机资源使用率或在网络层面利用网络嗅探解析帧内容的方式进行防御。

此外，GreyFan 攻击框架仍然可以在反检测方面进行改进。如在执行窃取、破坏等攻击任务时，可采用延时等待的方式进行攻击，即每隔一段时间攻击一次，以达到降低 CPU 使用率、逃避主机检测的目的；可以将有效载荷的内容通过 RSA 加密算法进行加密和混淆，以降低对有效载荷内容的识别。下一步将重点研究 GreyFan 攻击框架的新功能以及探究如何利用机器学习对 GreyFan 攻击进行智能识别，帮助人们降低网络空间的威胁。

参考文献：

[1] DAKHANE D M, DESHMUKH P R. Active warden for TCP sequence number base covert channel[C]//International Conference on Pervasive Computing, Jan 8-10, 2015, Pune, India. Piscataway: IEEE Press, 2015: 1-5.

[2] ZANDER S, ARMITAGE G, BRANCH P. Covert channels and countermeasures in computer network protocols[J]. IEEE Communications Surveys & Tutorials, 2007, 9(3): 44-57.

[3] 王娟, 郭永冲, 王强, 等. 基于 BHO 的网络隐蔽通道研究[J]. 计算机工程, 2009, 35(5): 159-161, 164.
WANG J, GUO Y C, WANG Q, et al. Research of network covert channel based on BHO[J]. Computer Engineering, 2009, 35(5): 159-161, 164.

[4] AHMADZADEH S A, AGNEW G. Turbo covert channel: an iterative framework for covert communication over data networks[C]//IEEE INFOCOM, April 14-19, 2013, Turin, Italy. Piscataway: IEEE Press, 2013: 2031-2039.

[5] 姬国珍, 谭全福. 基于数据包时间间隔的隐蔽通道实现及检测方法研究[J]. 通信技术, 2018, 51(1): 189-194.
JI G Z, TAN Q F. Covert channel implementation based on between-packet time intervals and detection method[J]. Communications Technology, 2018, 51(1): 189-194.

[6] 李卫, 嵩天. 适用于 NAT 环境的隐蔽通道构建方法[J]. 计算机工程与应用, 2018, 54(17): 103-109.
LI W, SONG T. Covert channel applying to NAT environment[J]. Computer Engineering and Applications, 2018, 54(17): 103-109.

[7] 朱越凡, 马迪, 王伟, 等. 一种 NTP 协议隐蔽通道[J]. 计算机系统应用, 2017, 26(5): 119-125.
ZHU Y F, MA D, WANG W, et al. Covert channel based on NTP protocol[J]. Computer Systems & Applications, 2017, 26(5): 119-125.

[8] TAN Y A, XU X, LIANG C, et al. An end-to-end covert channel via packet dropout for mobile networks[J]. International Journal of Distributed Sensor Networks, 2018, 14(5).

[9] YANG Z, HUANG Q, ZHANG Q, et al. NICScatter: backscatter as a covert channel in mobile devices[Z]. 2017.

[10] SCHULZ M, LINK J. Shadow Wi-Fi: teaching smartphones to transmit raw signals and to extract channel state information to implement practical covert channels over Wi-Fi[C]//The 16th Annual International Conference on Mobile Systems, Applications, and Services Table of Contents, June 10-15, 2018, San Francisco, California, USA. [S.l.:s.n.], 2018: 256-268.

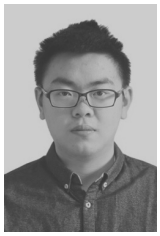
[11] WU K, WEI Z S, TEACHER W Z. A study on the application

of intrusion detection technology to WLAN[C]//IEEE International Conference on Communication Software and Networks, May 27-29, 2011, Xi'an, China. Piscataway: IEEE Press, 2011: 344-346.

[12] 谭彦, 厉萍, 卢洪涛, 等. Wi-Fi 无线钓鱼攻击分析及应对技术研究[J]. 电信科学, 2013, 29(Z2): 143-146, 151.
TAN Y, LI P, LU H T, et al. Wi-Fi wireless phishing attack analysis and coping technology research[J]. Telecommunications Science, 2013, 29(Z2): 143-146, 151.

[13] 高波, 潘毅明, 黄国瑾. 基于城域组网的运营级 WLAN 组网技术[J]. 电信科学, 2015, 31(10): 199-203.
GAO B, PAN Y M, HUANG G J. Technology of WLAN in operation level based on metropolitan area network[J]. Telecommunications Science, 2015, 31(10): 199-203.

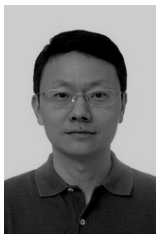
[作者简介]



马栋捷（1994− ），男，浙江工业大学计算机科学与技术学院硕士生，主要研究方向为信息安全。

金成强（1995− ），男，浙江工业大学计算机科学与技术学院硕士生，主要研究方向为信息安全。

陈园（1994− ），女，浙江工业大学计算机科学与技术学院硕士生，主要研究方向为物联网安全。



陈铁明（1978− ），男，博士，浙江工业大学计算机科学与技术学院教授，主要研究方向为网络空间安全与大数据智能分析。

研究与开发

基于区块链的结果可追溯的可搜索加密方案

翁昕耀，游林，蓝婷婷
（杭州电子科技大学，浙江 杭州 310018）

摘　要：在可搜索加密方案中，无论是云端服务器还是用户，都可能存在欺骗行为。为了解决这种安全问题，给出公平性安全的定义，提出基于区块链的可搜索加密方案。通过第三方可信机构（trusted authority，TA）验证数据传输过程中数据的一致性，区块链记录完整验证结果以防止篡改，使所涉及的实体达成一致的安全共识，从而实现公平性安全。安全性与复杂性分析表明，该方案是可行的。当该方案与搜索结果可验证的可搜索加密方案相结合时，可优化成基于区块链的搜索结果可验证的可搜索加密方案，其安全性将得到进一步的提高。

关键词：可搜索加密；公平性安全；区块链；可信机构；散列函数
中图分类号：TP309.7
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019183

Blockchain-based result-traceable searchable encryption scheme

WENG Xinyao, YOU Lin, LAN Tingting
Hangzhou Dianzi University, Hangzhou 310018, China

Abstract: In a searchable encryption scheme, whether it is a cloud server or a user, it may be deceptive. In order to solve the security problem, the definition of fairness security was given and a new searchable encryption scheme based on blockchain was proposed. Through the third-party trusted authority to verify the consistency of data during data transmission, the complete verification results were recorded on the blockchain to prevent tampering, and it could make all the related entities to reach the consistent security consensus and achieve the fairness security. The security and complexity analysis show that the proposed searchable encryption scheme was feasible. If it is combined with a searchable encryption scheme with its searched results verifiable, the proposed scheme can be effectively improved to be a blockchain-based searchable encryption scheme with its searched results verifiable and its security will be more strengthened.

Key words: searchable encryption, fairness security, blockchain, trusted authority, hash function

收稿日期：2019−03−22；修回日期：2019−06−28
通信作者：youlin@hdu.edu.cn
基金项目：浙江省自然科学基金重点项目（No. LZ17F020002）；国家自然科学基金资助项目（No. 61772166）
Foundation Items: The Key Project of Natural Science Foundation of Zhejiang Province (No. LZ17F020002), The National Natural Science Foundation of China (No. 61772166)

1　引言

随着人工智能、物联网等技术的快速发展，大数据的重要性越来越凸显。在云存储技术已经成熟的当下，用户和企业都倾向于将数据存储在云端服务器。但云端服务器是不可信赖的，虽然提供了诸多便利，但仍存在各种安全问题。用户的敏感数据（比如电子邮件、个人医疗信息或金融数据等）以明文形式上传，就会有被第三方恶意窃取的风险，导致用户隐私泄露。仅 2018 年，全球就发生多起重大信息安全泄露事件。2018 年 1 月，印度一个包含姓名、联系方式、指纹等敏感信息的 10 亿公民身份数据库 Aadhaar 遭泄露。2018 年 3 月，Facebook 被暴露超过 1.2 亿用户的个人信息经由第三方合作公司泄露，被用于非法政治用途。

为解决上述安全问题，可搜索加密技术应运而生，在加密数据上实现关键词检索，只获取感兴趣的目标数据。在保证数据安全的前提下，提高数据的使用效率。最早的可搜索加密方案是由 Song 等人[1]提出的 SWP 方案。Curtmola 等人[2]参考了 Goh 等人[3]的安全索引概念，提出了 SSE-1 和 SSE-2 方案。随着大数据时代的到来，可搜索加密方案的安全性受到越来越广泛的关注。Kamara 等人[4]提出的可搜索加密云端存储系统 CS2 以及 Kurosawa 和 Ohtaki[5]提出的 UC-security 可验证可搜索加密方案，是基于恶意服务器的假设构建的。此外，用户设备有受到攻击和密钥被攻击者窃取的可能性[6]。在多用户场景下，多个恶意用户也可能为了非法获取数据，联手欺骗云端服务器[7-9]。因此，可搜索加密系统中的任何实体都可能是恶意的，会存在欺骗行为。

区块链[10-11]是一种基于现代密码算法的分布式数据库，具有去中心化、难以篡改、可追溯、公开透明等优势。区块链的主要作用是存储信息，任何需要保存的信息，都可以写入区块链，也可以从中获取。区块链的结构如图 1 所示，由一个个串联的区块组成。每个区块包含两个部分：区块头（head）和区块数据（data）。在区块头上记录当前区块的特征值，如生成时间、区块数据的散列值、上一个区块的散列值等；在区块数据上记录实际的数据。

散列函数，是密码体制中能够将任意长度的消息映射成某一固定长度消息的公开函数。在密码学中，散列函数满足 3 个安全性方面的特性：单向性（one-way），如果对任意给定的 y，寻找使 $h(x)=y$ 成立的 x 在计算上是困难的；弱抗碰撞性（weakly collision resistance），已知 x，寻找 $x'\neq x$，使得 $h(x)=h(x')$ 在计算上是困难的；强抗碰撞性（strongly collision resistance），寻找两个不同的 x 和 x'，使得 $h(x)=h(x')$ 在计算上是困难的。根据上述安全特性可以推断，散列值在区块链中起着决定性作用。因为每个区块都拥有一个独一无二的散列值，如果区块的内容发生变化，则它的散列值也一定会发生改变。

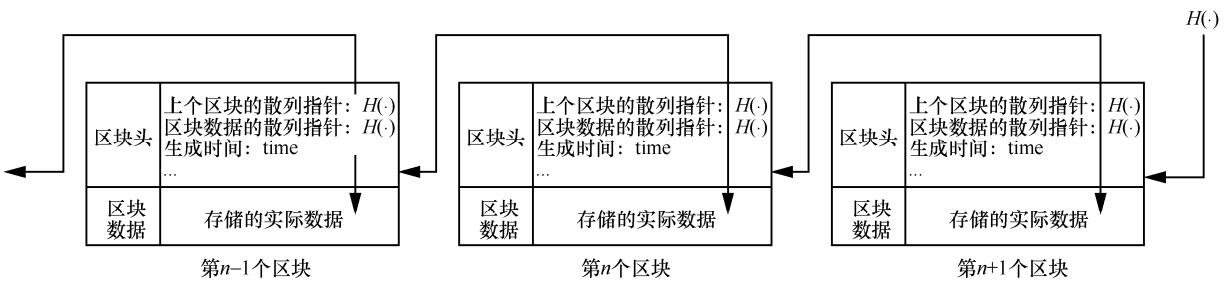


图 1　区块链的结构

为了进一步提高可搜索加密技术的安全性，本文提出基于区块链的可搜索加密方案，主要贡献如下。

- 给出可搜索加密研究领域中公平性安全的定义。完善可搜索加密的安全体系，能同时保障用户和云端服务器的权益。
- 提出基于区块链的可搜索加密方案。结合区块链与第三方可信机构（trusted authority，TA），实现公平性安全。其中，区块链起决定性作用。
- 本方案具有良好的拓展性，能与现有可搜索加密方案相结合，提高方案的安全性。
- 安全性分析验证了新方案的可行性与安全性，复杂性分析验证了新方案的有效性。

2 系统模型

可搜索加密技术的一般模型主要由 3 个实体组成：数据拥有者、数据使用者和云端服务器，如图 2 所示。

系统的设计目的在于，确保用户数据的安全性，同时提高数据的使用效率。如果下列性质保持不变，则可搜索加密方案是安全的：当云端服务器只获得密文时，它无法了解任何关于明文文档的信息；当云端服务器执行搜索算法时，它同样无法了解任何关于明文文档和关键词的任何信息，除了加密的检索结果。

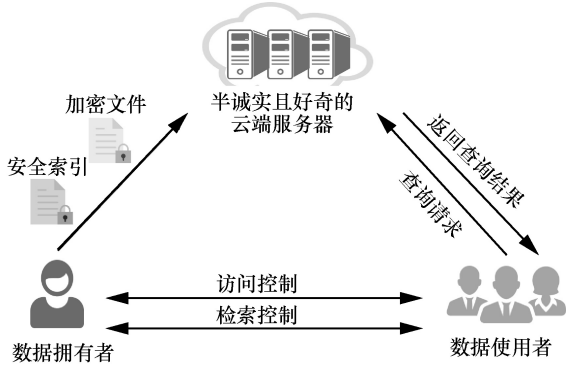


图 2 可搜索加密一般模型

可搜索加密的主要过程可分为 7 个，如图 3 所示。

（1）索引生成 $I \leftarrow \mathrm{Index}(F,W,K)$：数据拥有者在本地执行索引构建算法，扫描文档集合 $F=(f_1,f_2,\cdots,f_n)$，获得关键词集合 $W=(w_1,w_2,\cdots,w_n)$。然后将密钥 K、文档集合 F 和关键词集合 W 作为输入，得到作为输出的索引 I。

（2）数据加密 $C, S \leftarrow \mathrm{Enc}(F,I,K)$：数据拥有者在本地执行加密算法，将密钥 K、文档集合 F 和索引 I 作为输入，得到作为输出的加密文档集合 C 和加密索引 S，然后将其上传给云端服务器。

（3）权限授予 $\mathrm{Auz}(K,\delta)$：数据拥有者将密钥 K 和相关参数 δ 发送给数据使用者，授予合法

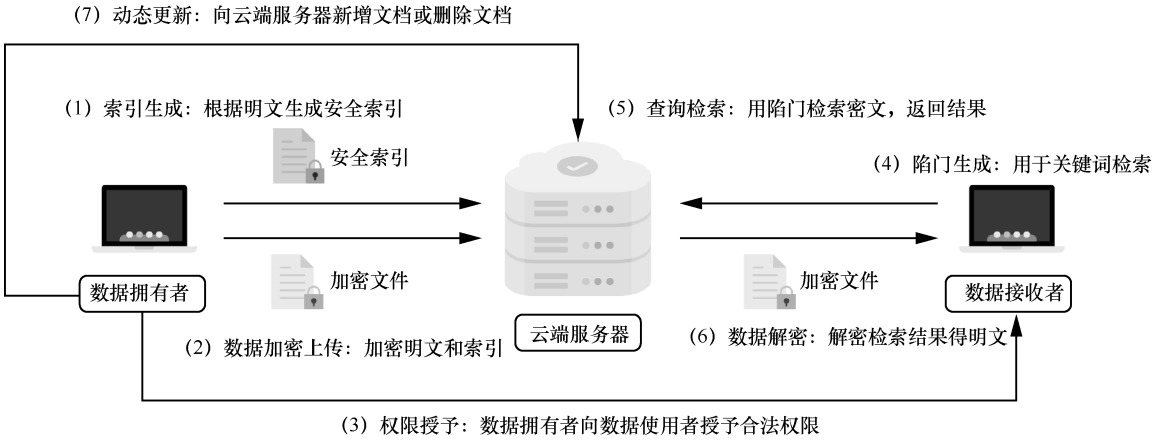


图 3 可搜索加密主要过程

权限。

（4）陷门生成 $\mathrm{TR}_w \leftarrow \mathrm{SrchToken}(w,K)$：数据使用者在本地执行关键词检索陷门生成算法，将密钥 K 和待检索关键词 w 作为输入，得到作为输出的检索陷门 TR_w，然后上传给云端服务器。

（5）查询检索 $c \leftarrow \mathrm{Search}(\mathrm{TR}_w, C, S)$：云端服务器执行检索算法，输入检索陷门 TR_w、加密文档集合 C 和加密索引 S，得到检索结果 c 返回给数据使用者。

（6）数据解密 $f \leftarrow \mathrm{Enc}(c,K)$：数据使用者在本地执行解密算法，将密钥 K 和加密文档集合 c 作为输入，得到作为输出的明文文档集合 f，即关键词检索结果。

（7）动态更新 $C', I' \leftarrow \mathrm{Update}(F', C, K, I, \tau)$：数据拥有者向云端服务器发起文档更新请求，新增文档或者删除文档。其中，待更新的文档为 F'，更新凭证为 τ；更新后的加密文档集合为 C'，加密索引为 I'。

3　公平性安全

伴随着可搜索加密方案的不断完善，越来越多新的安全问题也随之产生。从最早的密文安全，到后来的索引安全、关键词检索安全、动态更新安全等。经过研究后，注意到可搜索加密仍旧存在被忽略的安全问题。

定义 1　（数据传输）数据拥有者成功发送数据，数据接收者成功接收数据，才能称作一次数据传输。

定义 2　（公平性安全）在一次数据传输中，数据发送者与数据接收者对外声明一致，即数据发送者发送的数据与数据接收者接收的数据是相同的，双方都不存在任何欺骗行为，则认为方案实现了公平性安全。

在可搜索加密方案中，数据拥有者、数据使用者和云端服务器都有可能成为数据发送者或数据接收者。对数据发送者来说，应该如实告知发送了什么数据。对于数据接收者来说，应该如实告知接收了什么数据。

无论是数据发送者，还是数据接收者，只要有一方发生欺骗行为，就损害了方案的公平性安全。比如，用户上传完整密文给云端服务器，云端服务器收到完整密文，但却谎称收到的密文有缺失，恶意欺骗用户，降低用户的信用评级。已知目前存在的可搜索加密方案，都无法帮助用户验证云端服务器欺骗行为，用户的公平性受到损害。同样，用户也可能有欺骗行为。比如，用户收到来自云端服务器返回的正确关键词检索结果，但却谎称检索结果不正确，陷害云端服务器。所以，为了同时保障可搜索加密方案中不同实体的利益关系，实现公平性安全是有必要的。

公平性安全模型　在任意的过程 Process() 中，实体 U 向其他实体 Q 发送数据，实体 Q 接收来自实体 U 的数据。实体 Q 返回的接收结果可能是正确的（true），也可能是错误的（false）。当且仅当实体 U 和 Q 都能判断双方发送或接收的数据是否一致以及识别存在欺骗行为的实体时，可搜索加密方案是公平性安全的。

4　基于区块链的可搜索加密方案

4.1　基于区块链的可搜索加密方案系统模型

如果只依靠用户和服务器的自觉性，是无法完全保证方案公平性安全的，仍旧容易遭受不可信敌手的假冒攻击。而解决这个问题，就需要第三方可信机构验证公平性以及区块链记录结果以防篡改，从而实现方案的公平性安全。

在可搜索加密一般模型的基础上，增加了第三方可信机构，用于验证实体间传输数据的公平性。在每个实体处增加了区块链节点，形成整体的区块链，用于记录并同步验证结果，如图 4 所示。

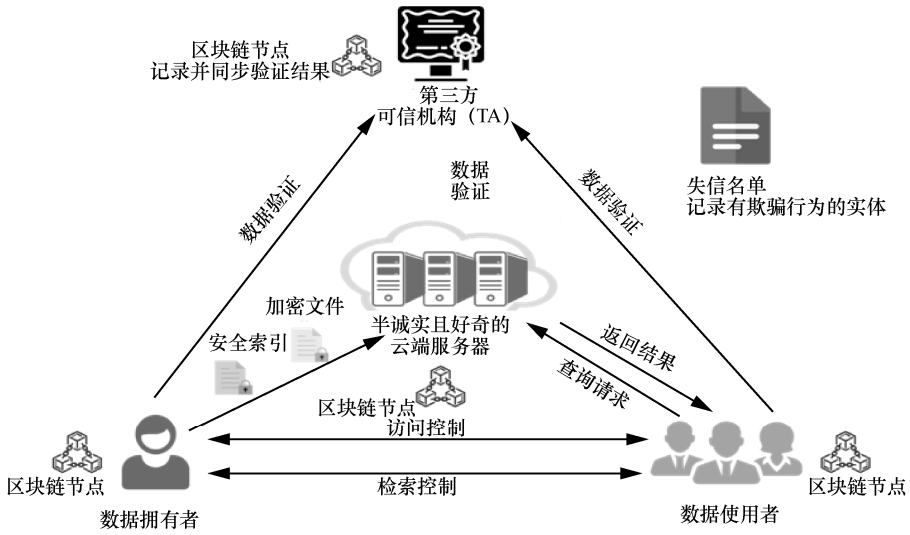


图 4　基于区块链的可搜索加密方案的模型

此外，失信名单用于记录有欺骗行为的实体，为可搜索加密领域构建一个完整的信任体系。

4.2　第三方可信机构（TA）的验证流程

在可搜索加密方案中，任何实体都可能成为数据发送者和数据接收者。需要保证方案的公平性安全，也意味着需要保证方案中每次数据传输的公平性安全。第三方可信机构的验证流程，如图 5 所示。

步骤 1　数据拥有者向数据接收者发送加密数据 data；

步骤 2　数据拥有者计算加密数据 data 的散列值 Hash(data)，上传给 TA；

步骤 3　数据接收者计算接收到的加密数据 data′的散列值 Hash(data′)，上传给 TA；

步骤 4　TA 对比验证散列值是否一致，并将验证结果记录到区块链上；

步骤 5　TA 将验证结果返回给数据发送者与数据接收者；如果第一次验证不成功，则重复步骤 2~5 继续验证，此时要求实体上传加密的数据，由 TA 来计算散列值 Hashagain(data) 和 Hashagain(data′) 并对比验证。Hashagain(.) 是与 Hash(.) 相同的散列函数，只是标记不一样。

（1）如果 Hashagain(data) 等于 Hash(data)，而 Hashagain(data′) 不等于 Hash(data′)，则证明云端服务器是欺骗行为；如果 Hashagain(data) 不等于 Hash(data)，而 Hashagain(data′) 等于 Hash(data′)，

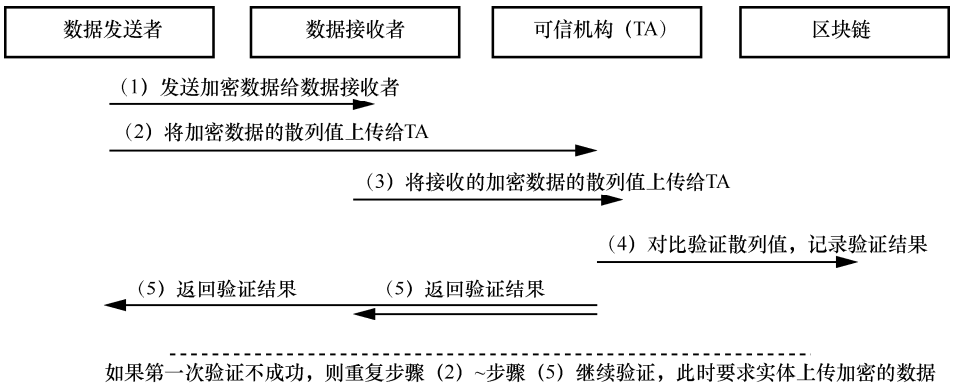


图 5　第三方可信机构的验证流程

则证明数据拥有者是欺骗行为；

（2）如果 Hashagain(data) 等于 Hash(data)，而 Hashagain(data′) 等于 Hashagain(data′)，则证明本次对比没有验证结果，需进行第三次验证。再次要求数据拥有者和云端服务器上传各自拥有的加密数据，由 TA 来计算对比散列值。在本次验证中，虽不能证明哪一方是欺骗行为，但能肯定两者中一定存在欺骗行为。

如果 TA 验证次数超过 5 次，仍旧验证不出结果，则将两者都添加到失信名单上，留下信用可疑的记录。此外，TA 需要将每次的验证结果完整记录到区块链上。

4.3　区块链记录验证结果

为了保证可搜索加密方案的公平性安全，也需要考虑第三方可信机构的安全性。因为其存在被恶意敌手攻击或串通其他实体来进行欺骗的可能性。

本文采用在区块链系统上记录验证结果，使方案中所有涉及的实体达成一致的安全共识。区块链中每个参与节点都可以记录数据且保存着整个区块链的数据。区块链上的数据是永久记录且难以篡改的，如果实体的欺骗行为被证实，则实体会被列入失信名单，造成巨大损失。此外，区块链可追溯、公开透明等特点，也使方案整体的安全性得到了一定完善。

TA 需要将每次对比认证的结果记录到区块链上。以一个区块为例，如图 6 所示，在区块头上记录上一个区块的散列值、区块数据的散列值、生成时间等其他相关参数。在区块数据中记录参与验证的实体、验证的次数、每次验证的内容、内容的散列值以及验证结果。每当区块链有更新时，处于区块链中的所有节点需要同步更新，例如数据拥有者和数据使用者的节点。

4.4　公平性安全验证过程

涉及公平性安全的，在可搜索加密方案中主要有 4 个过程：数据加密上传、权限授予、查询检索、动态更新。在不同过程中，都需要第三方可信机构验证结果、区块链记录结果，只是参与数据传输的实体不同，传输的数据不同，具体的区别如下。

区块头	上个区块的散列指针：$H(\cdot)$ 区块体的散列指针：$H(\cdot)$ 生成时间：time …
区块数据	验证实体：数据拥有者、云端服务器 验证次数：1次 验证内容：密文文档的标题集合 内容散列值：$H(C)$=$H(C')$ 验证结果：实现公平性安全 …

图 6　区块链中区块的结构

（1）数据上传过程

Upload()：数据发送者为数据拥有者，数据接收者为云端服务器，传输的数据为加密后的密文 C 和索引 I。

（2）权限授予过程

Authorization()：数据发送者为数据拥有者，数据接收者为数据使用者，传输的数据为密钥 K 和相关参数 δ。

（3）查询检索过程

Search()：此过程涉及两次数据传输，发起查询请求以及返回查询结果。其一，数据发送者为数据使用者，数据接收者为云端服务器，传输的数据为加密的搜索陷门 TR_w；其二，数据发送者为云端服务器，数据接收者为数据使用者，传输的数据为密文检索结果 R。

（4）动态更新过程

在动态更新过程中，具体可以分为两种情况：新增与删除。

- Add()：数据发送者为数据拥有者，数据接收者为云端服务器，传输的数据为加密后的新增文档 F 与新增凭证 τ_a。
- Delete()：数据发送者为数据拥有者，数据接收者为云端服务器，传输的数据为加密后的删除凭证 τ_d。

云端服务器根据删除凭证 τ_d 检索文档 D 并将其删除。此外，云端服务器删除文档 D 后，需要将删除结果返回给数据拥有者。此时，数据发送者变为云端服务器，数据接收者变为数据拥有者，传输的数据为删除的文档 D 。

5 安全性与复杂性

5.1 安全性

由于本文方案引入了区块链技术和第三方可信机构，可能带来新的安全风险。故下面将从两个方面证明本文方案的安全性。

定理 1 本文提出的方案是可行的。

证明如下。

（1）第三方可信机构 TA 安全性。作为可信的第三方，TA 是与通信双方实体互相独立的，并且保持绝对的公平公正。在初次验证中，只需要将待验证的加密数据散列值发送给 TA。散列函数具有单向性，因此只获得了散列值的 TA 无法计算得到加密数据。在二次验证及后续验证中，需要将待验证的加密数据发送给 TA。数据都是加密的，没有解密密钥的 TA 无法从中获得关于明文数据的任何信息。在数据传输过程中，如果加密数据被第三方恶意敌手截获，只要解密密钥不泄露给敌手，同样也可以保证明文数据的安全。在结果验证完后，TA 还需要将结果完整记录到区块链上。由于区块链上记录的数据是难以篡改且可追溯的，这在一定程度上也约束了 TA 的行为。综上所述，第三方可信机构 TA 是安全的。

（2）区块链安全性。在本方案中，记录到区块链上的相关数据有验证实体、验证次数、验证内容、内容散列值、验证结果，而这些内容都是可以公开的。其中，验证内容只是标题集合，用来指代本次验证是对什么数据进行验证的，不包括加密后的数据。根据散列函数的安全特性，仅通过内容散列值无法计算获得有关加密后数据的任何信息，从而确保区块链上记录数据的安全。区块链本身是去中心化的，每个平等的节点都存储着完整的数据库，这意味着区块链是不能随意篡改的，并且公开透明，接受监督。综上所述，区块链是安全的。

因此，本文提出的方案是可行的。

定理 2 本文提出的方案是公平性安全的。

证明：根据定义 1 可知，在一次数据传输中，数据发送者必定发送了数据，数据接收者必定收到了数据。根据定义 2 可知，数据发送者和数据接收者两者的数据完全一致，那么方案是属于公平性安全的。

在基于区块链的可搜索加密方案中，涉及数据传输的过程主要有：数据加密上传、权限授予、查询检索、动态更新。在这些过程中，通过第三方可信机构验证数据传输的公平性，当第一次验证成功时，则直接证明方案实现了公平性安全。若双方始终坚持自己的声明，第一次验证失败，则进行后续的验证，直到结果验证成功。若验证一直失败，在次数超过 5 次后，则证明数据传输的双方中必定存在欺骗者，是不公平的。每次验证的结果，将完整记录在区块链上，难以篡改但可追溯。综上所述，在可搜索加密方案的每次数据传输过程中，本方案可以保证各方实体享有公平对等的权益。

因此，本文提出的方案是公平性安全的。

5.2 复杂性分析

本文所提方案是基于一般可搜索加密模型，增加了区块链、第三方可信机构和失信名单，这 3 个实体是互相独立的，不影响可搜索加密方案主要操作。因此该方案能与现有的可搜索加密方案相结合，提高方案整体的安全性。比如，将本文方案与 Kurosawa 等人[5]所提方案结合，与其他方案的比较结果，见表 1。其中，n 表示文档的数量，r 表示检索某个关键词时检索到的文档数量，p 表示并行服务器的数量，m 表示交易事务的大小。Kamara 等人[12]所提方案采用多服

表 1　本文方案与其他方案的比较结果

方案	搜索复杂度	通信复杂度	可验证	敌手	公平性安全
Kurosawa 等人[5]	$O(D(w))$	$O(r)$	是	恶意服务器	否
Kamara 等人[12]	$O(r\log(n)/p)$	$O(1)$	是	诚实但好奇的服务器	否
Li 等人[13]	$O(D(w))$	$O(6m)$	是	恶意服务器和恶意用户	是
本方案	$O(D(w))$	$O(r)$	是	恶意服务器和恶意用户	是

务器并行计算，故其关键词检索效率最高，但无法实现公平性安全。Li 等人[13]所提方案虽然能实现公平性安全，但其需要“数字货币”系统的支持，且每次关键词检索至少需要 3 次数据通信和 6 次交易事务，降低了获取检索结果的效率。本文所提方案是基于恶意服务器和恶意用户提的，在关键词检索时，只需要验证并记录传输数据散列值的一致性，就能确保方案的公平性安全。

6　结束语

为解决可搜索加密中涉及的实体都可能存在的欺骗行为的问题，本文给出了可搜索加密研究领域中公平性安全的定义，并提出基于区块链的可搜索加密方案。该方案可实现公平性安全，使实体各方达成一致的安全共识，具有占用资源少、可拓展性好等优势。实现公平性安全的意义在于：同时保障实体各方的权益，及时发现不可信赖的实体，避免后续的损失。本文方案也存在不足，如果双方实体传输数据本身就是错误的（物理存储故障等），那么公平性安全检测会失效。这些问题将在今后的工作中研究与改进。

参考文献：

[1] SONG D X, WAGNER D, PERRIG A. Practical techniques for searches on encrypted data[C]//the 2000 IEEE Symposium on Security and Privacy (S&P 2000), May 14-17, 2000，Washington, USA. Piscataway: IEEE Press, 2000: 44-55.

[2] CURTMOLA R, GARAY J, KAMARA S, et al. Searchable symmetric encryption: improved definitions and efficient constructions[C]//the ACM Conference on Computer and Communications Security, Oct 30-Nov 3, 2006, Alexandria, USA. New York: ACM Press, 2006: 79-88.

[3] GOH E J. Secure indexes[J]. IACR Cryptology ePrint Archive, 2003: 216.

[4] KAMARA S, PAPAMANTHOU C, ROEDER T. CS2: A searchable cryptographic cloud storage system[R]. 2011.

[5] KUROSAWA K, OHTAKI Y. UC-secure searchable symmetric encryption[C]//the International Conference on Financial Cryptography and Data Security, Feb 27- Mar 3, 2012, Kralendijk, Bonaire. Heidelberg: Springer, 2012: 285-298.

[6] DAI S, LI H, ZHANG F. Memory leakage-resilient searchable symmetric encryption[J]. Future Generation Computer Systems, 2016(62): 76-84.

[7] JARECKI S, JUTLA C, KRAWCZYK H, et al. Outsourced symmetric private information retrieval[C]//The 2013 ACM SIGSAC Conference on Computer & Communications Security, Nov 4 - 8, 2013, Berlin, Germany. New York: ACM Press, 2013: 875-888.

[8] FISCH B A, VO B, KRELL F, et al. Malicious-client security in blind seer: a scalable private DBMS[C]//the 2015 IEEE Symposium on Security and Privacy, May 18-21, 2015, California, USA. Piscataway: IEEE Press, 2015: 395-410.

[9] ISHAI Y, KUSHILEVITZ E, LU S, et al. Private large-scale databases with distributed searchable symmetric encryption[C]// The Cryptographers' Track at the RSA Conference, Feb 29 - Mar 4, 2016, San Francisco, USA. Cham: Springer, 2016: 90-107.

[10] NAKAMOTO S. Bitcoin: a peer-to-peer electronic cash system[Z]. 2008.

[11] 李董, 魏进武. 区块链技术原理、应用领域及挑战[J]. 电信科学, 2016, 32(12): 20-25.
LI D, WEI J W. Theory, application fields and challenge of the blockchain technology[J]. Telecommunications Science, 2016, 32(12): 20-25.

[12] KAMARA S, PAPAMANTHOU C. Parallel and dynamic

searchable symmetric encryption[C]//the International Conference on Financial Cryptography and Data Security, Apr 1-5, 2013, Okinawa, Japan. Heidelberg：Springer, 2013: 258-274.

[13] LI H, TIAN H, ZHANG F. Block chain based searchable symmetric encryption[J]. IACR Cryptology ePrint Archive, 2017: 447.

[作者简介]

翁昕耀（1994− ），男，杭州电子科技大学硕士生，主要研究方向为信息安全与可搜索加密。

游林（1966− ），男，博士，杭州电子科技大学网络空间安全学院教授、博士生导师，主要研究方向为密码学、区块链技术、生物特征识别及其应用等。

蓝婷婷（1994− ），女，杭州电子科技大学硕士生，主要研究方向为信息安全与生物特征加密技术。

研究与开发

非视距环境下基于 RSS-TOA 的定位算法

卢志刚[1]，李有明[1]，贾向红[2]，常生明[1]，王晓丽[1]
（1. 宁波大学信息科学与工程学院，浙江 宁波 315211;
2. 中国联合网络通信有限公司，北京 100033）

摘　要：在无线传感器网络定位系统中，尤其是在室内定位中，非视距（NLOS）误差的存在使定位性能急剧下降。为克服非视距传播带来的定位误差，提出了一种针对非视距环境下联合接收信号强度（RSS）和到达时间（TOA）的定位算法。该方法首先通过 RSS 和 TOA 的测量结果建立关于目标位置的非凸优化问题，然后通过二阶锥松弛理论，将原始的非凸优化问题转换为一种凸优化问题，由此能够快速得到原问题的一个次优解。通过计算机模拟仿真验证，新方法的估计精度更高，性能更好。

关键词：接收信号强度；到达时间；二阶锥松弛；非视距
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019146

RSS-TOA based localization algorithm in non-line-of-sight environment

LU Zhigang[1], LI Youming[1], JIA Xianghong[2], CHANG Shengming[1], WANG Xiaoli[1]
1. Faculty of Electrical Engineering and Computer Science , Ningbo University, Ningbo 315211, China
2. China United Network Communications Co., Ltd., Beijing 100033, China

Abstract: The existence of non-line-of-sight (NLOS) error can degrade the positioning performance in wireless sensor network localization system, especially in indoor localization. To overcome the localization error caused by NLOS propagation, a localization algorithm was proposed based on received signal strength (RSS) and time-of-arrival (TOA). Firstly, a non-convex optimization problem was established based on RSS and TOA. Then, the original non-convex optimization problem was transformed into a convex optimization problem through the second-order cone relaxation technique, therefore a sub-optimal solution to the original problem could be obtained efficiently. Finally, computer simulation results show that the proposed method can provide higher estimation accuracy and better performance.

Key words: received signal strength, time-of-arrival, second-order cone relaxation, non-line-of-sight

收稿日期：2019−02−25；**修回日期**：2019−04−15
通信作者：李有明，liyouming@nbu.edu.cn
基金项目：国家自然科学基金资助项目（No.61571250）；浙江省自然科学基金资助项目（No.LY18F010010）；宁波市自然科学基金资助项目（No.2015A610121）
Foundation Items: The National Natural Science Foundation of China (No.61571250), The Natural Science Foundation of Zhejiang Province of China (No.LY18F010010), The Natural Science Foundation of Ningbo of China (No.2015A610121)

1 引言

无线传感器网络（wireless sensor network，WSN）是指由多个设备组成的无线通信网络，被分配到一个被监控区域，以测量某些局部感兴趣的信息[1]。近年来，无线传感器网络在目标跟踪、导航、应急服务、智能交通等领域都得到了广泛的应用[2-3]。在这些应用中，对目标位置定位至关重要，而有些特殊空间如室内、水下等，无法用 GPS/北斗等卫星定位，需要利用其他技术对目标定位。通常，通过人工或者其他手段部署位置已知的传感器节点称为锚节点。预先不知道自身位置，需要通过锚节点来定位的节点称为目标节点。传感器定位的主要思想是利用带有噪声的测量值来确定目标节点的位置。根据节点获取信号信息的方式不同，可以将目标定位方法分为：到达时间（time-of-arrival，TOA）[4]、到达角度（angle-of-arrival，AOA）[5]、到达时间差（time difference of arrival，TDOA）[6]、接收信号强度（received signal strength，RSS）[7]以及它们之间的联合方式。

早期的研究主要基于视距条件下的定位方式，这类方法只考虑了目标节点和锚节点之间是直线通信，没有任何阻碍物；但是在实际环境下，比如在室内，节点之间的通信会受到很多阻碍物的遮挡，这就会形成一种非视距（non-line-of-sight，NLOS）传播方式，从而导致对未知节点的估计存在较大的偏差。NLOS 误差作为无线传感器网络定位是最主要的误差来源之一[8]，如何有效地抑制 NLOS 误差并提高定位精度已经成为无线传感器网络定位技术的热点问题。

由于在复杂的环境下，各种测量结果存在较大的测量误差，一些传统的算法在非视距环境下定位误差较大，算法精度较低。因此，抑制非视距误差和提高定位精度引起了研究者的重视。参考文献[9]研究了视距条件下基于 RSS 的定位方法，作者分别分析了协作和非协作两种定位方式，但没有考虑非视距误差带来的影响。参考文献[10]研究了非视距环境下基于 TOA 的定位算法，分析了已知非视距状态和未知非视距状态两种条件下的定位性能，但是当噪声较大时，定位性能较差。在参考文献[11]中，首先将所有链路视为视距（line-of-sight，LOS），然后应用交替优化方法，并以迭代的方式改进位置估计和平均 NLOS 偏差估计，但该方法很难求得全局最优解，并且无法保证收敛性。参考文献[12]利用最大似然法估计位置坐标，但是需要完全已知噪声功率和 NLOS 偏差精确的先验信息，实际环境中应用受限。参考文献[13]提出了一种加权最小二乘方法，该方法不需要 NLOS 的误差统计信息，然而估计精度比较差。参考文献[14]研究了基于 RSS 和 AOA 的混合算法，提高了算法的精度，但是只考虑了视距条件下，无法直接应用到非视距环境。参考文献[15]利用最大似然准则建立测距优化问题，但是该类问题的目标函数是非线性且非凸的，求解十分困难。为了克服最大似然估计问题的缺陷，参考文献[16]将该问题转化为半正定规划（semidefinite programming，SDP）问题，然后求取次优的近似解。

非视距情况下，单独 RSS 方法会造成未知参数估计不准确，定位误差大。单独 TOA 方法需要高精度的时间同步，会造成定位精度不高。本文使用联合定位方法，锚节点可以从多渠道得到更多的可用信息，定位精度更高，所以本文建立了一类基于 RSS 和 TOA 的无线传感器网络联合定位问题。运用二阶锥松弛技术，将原始的非凸优化问题转化为一个凸的优化问题，由此可以精确求解全局最优解。仿真结果表明新算法的估计性能得到明显改善。

2 系统模型

考虑具有 N 个锚节点和 1 个目标节点的二维

无线传感网络定位场景，如图 1 所示，其中锚节点位置 $s_1,s_2,\cdots,s_N$ 已知，目标节点位置 x 未知。在集中处理方式下，假设目标节点向锚节点发出信号，且锚节点可以从接收的信号中提取 RSS 和 TOA 的测量信息，并且发送给中央处理器进行集中处理。在此期间，假设所有传感器节点的位置保持不变。由锚节点测量得到的两个度量信息，分别是接收功率 P_i 和测量距离 d_i，可以建模为[17]：

$$P_i = P_0 - b_i - 10\gamma \lg \frac{\| x - s_i \|}{d_0} + n_i \tag{1}$$

$$d_i = \| x - s_i \| + \beta_i + m_i \tag{2}$$

其中，P_0（dBm）是目标节点的发射功率，b_i（dB）和 β_i（m）是非视距误差，γ 是路径损失指数，x 表示真实的目标节点位置，s_i 表示第 i 个锚节点的位置坐标。d_0 是一个参考距离（$\|x-s_i\| \geqslant d_0$），d_i 为第 i 个锚节点与目标节点之间的测量距离，n_i 为第 i 个锚节点与目标节点之间的功率测量误差，m_i 为第 i 个锚节点与目标节点之间的距离测量误差，均为零均值高斯随机变量，分别表示为 $n_i \sim \mathcal{N}(0,\sigma_{n_i}^2)$，$m_i \sim \mathcal{N}(0,\sigma_{m_i}^2)$，为了简单起见，在本文的其余部分，假设所有噪声 $\sigma_{n_i}^2=\sigma_n^2$、$\sigma_{m_i}^2=\sigma_m^2$。根据参考文献[10-11]，假设 N 个不同的 NLOS 误差在某区间上是均匀分布的，即 $0 \leqslant b_i \leqslant b_{\max}$ 和 $0 \leqslant \beta_i \leqslant \beta_{\max}$。

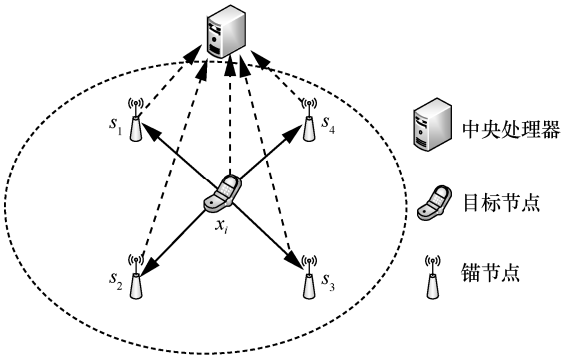


图 1　定位网络示意图

对于给定的观测向量 $\theta=\left[P^{\mathrm{T}},d^{\mathrm{T}}\right]^{\mathrm{T}}$（$\theta \in R^{2N}$），其中，$P=\left[P_1,P_2,\cdots,P_N\right]^{\mathrm{T}}$，$d=\left[d_1,d_2,\cdots,d_N\right]^{\mathrm{T}}$，则目标位置 x 的联合最大似然（maximum likelihood，ML）估计可表示为：

$$\min \sum_{i=1}^{2N} \frac{(\theta - f_i(x))^2}{\sigma_i^2} \tag{3}$$

其中，$\sigma_i=\left[\sigma_{n_i},\sigma_{m_i}\right]^{\mathrm{T}}$，$f_i(x)=[P_0-b_i-10\gamma \lg \frac{\| x-s_i \|}{d_0}, \| x-s_i \|+\beta_i]^{\mathrm{T}}$。可以看出，式（3）是非凸的，很难找到全局最小解。为了克服 ML 估计的非凸性和非线性，下文将构建如何应用凸优化的方法来求解目标位置的有效方法。

3　混合估计算法

本节设计一种求解式（3）的快速定位算法，具体过程如下。

首先将式（1）和式（2）分别近似为：

$$P_i = P_0 - b - 10\gamma \lg \frac{\| x - s_i \|}{d_0} + n_i \tag{4}$$

$$d_i = \| x - s_i \| + \beta + m_i \tag{5}$$

由于 N 个不同的 NLOS 误差是均匀分布的，所以将其近似为一个平均误差，故令 TOA 和 RSS 的非视距误差均值分别为 b 和 β。可以将式（4）写为：

$$10^{\frac{P_0-P_i-b+n_i}{10\gamma}} = \frac{\| x - s_i \|}{d_0} \tag{6}$$

进一步将式（6）等效表示为：

$$\mathrm{e}^{\ln(10^{\frac{P_0-P_i-b+n_i}{10\gamma}})} = \frac{\| x - s_i \|}{d_0} \tag{7}$$

将式（7）用一阶泰勒公式展开，可以近似表示为：

$$\rho + \varepsilon_i = \xi_i \| x - s_i \| \tag{8}$$

其中，$\rho=d_0 10^{\frac{P_0-b}{10\gamma}}$，$\xi_i=10^{\frac{P_i}{10\gamma}}$，$\varepsilon_i \sim \mathcal{N}(0,(\rho \frac{\ln 10}{10\gamma}\sigma_{n_i})^2)$。

对式（8）中 ε_i 移向等式右边且等式两边求平

方，得到：

$$\rho^2 \approx \xi_i^2 \| x-s_i \|^2 -2\varepsilon_i \xi_i \| x-s_i \| \tag{9}$$

由于 $2\varepsilon_i \xi_i \| x-s_i \|$ 远大于 ε_i^2，故可以忽略二阶噪声项，将式（5）中非视距误差项 β 移向等式左边，并将等式两边求平方，忽略二阶噪声项得到：

$$(d_i-\beta)^2 \approx \| x-s_i \|^2 +2m_i \| x-s_i \| \tag{10}$$

对式（9）、式（10）移项分离噪声项后，根据最小二乘准则，表示为如下的最小化问题：

$$\min_{x,\rho,\beta}\left(\sum_{i=1}^{N}\left(\frac{\xi_i^2 \| x-s_i \|^2 -\rho^2}{2\xi_i \| x-s_i \|}\right)^2 + \sum_{i=1}^{N}\left(\frac{(d_i-\beta)^2-\| x-s_i \|^2}{2\| x-s_i \|}\right)^2\right) \tag{11}$$

将式（11）中 $\|x-s_i\|^2$ 和 $(d_i-\beta)^2$ 展开，由于目标函数非凸，故引进辅助变量，令 $\|x\|^2=v$、$\|\beta\|^2=u$、$\|\rho\|^2=h$，则式（11）可进一步表示为：

$$\begin{aligned}&\min_{x,\rho,\beta,v,u,h}\left(\sum_{i=1}^{N}\frac{(\xi_i^2(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)-h)^2}{4\xi_i^2(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)} + \sum_{i=1}^{N}\frac{(u-2d_i\beta+d_i^2-v+2s_i^{\mathrm{T}}x-\| s_i \|^2)^2}{4(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)}\right)\\ &\text{s.t. } \|x\|^2=v,\ \|\beta\|^2=u,\ \|\rho\|^2=h\end{aligned} \tag{12}$$

由于式（12）中约束变量是非凸的，很难求得最优解，为此，分别将 $\|x\|^2=v$、$\|\beta\|^2=u$ 和 $\|\rho\|^2=h$ 松弛为 $\|x\|^2 \leqslant v$、$\|\beta\|^2 \leqslant u$ 和 $\|\rho\|^2 \leqslant h$，从而式（12）被松弛为一个优化问题：

$$\begin{aligned}&\min_{x,\rho,\beta,v,u,h}\left(\sum_{i=1}^{N}\frac{(\xi_i^2(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)-h)^2}{4\xi_i^2(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)} + \sum_{i=1}^{N}\frac{(u-2d_i\beta+d_i^2-v+2s_i^{\mathrm{T}}x-\| s_i \|^2)^2}{4(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)}\right)\\ &\text{s.t. } \|x\|^2 \leqslant v,\ \|\beta\|^2 \leqslant u,\ \|\rho\|^2 \leqslant h\end{aligned} \tag{13}$$

由于式（13）中目标函数是非凸的，故引入辅助变量 t_i 和 g_i，式（13）可以表示为：

$$\min_{x,\rho,\beta,v,u,h,t_i,g_i}\left(\sum_{i=1}^{N}t_i+\sum_{i=1}^{N}g_i\right) \tag{14}$$

$$\begin{aligned}&\text{s.t. } \frac{(\xi_i^2(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)-h)^2}{4\xi_i^2(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)} \leqslant t_i\\ &\frac{(u-2d_i\beta+d_i^2-v+2s_i^{\mathrm{T}}x-\| s_i \|^2)^2}{4(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)} \leqslant g_i\\ &\|x\|^2 \leqslant v,\ \|\beta\|^2 \leqslant u,\ \|\rho\|^2 \leqslant h\end{aligned}$$

综上所述，式（14）可以表示为二阶锥松弛问题，等价于：

$$\begin{aligned}&\min_{x,\rho,\beta,v,u,h,t_i,g_i}\left(\sum_{i=1}^{N}t_i+\sum_{i=1}^{N}g_i\right)\\ &\text{s.t. } \left\|\begin{bmatrix}2(\xi_i^2(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)-h\\ 4\xi_i^2(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)-t_i\end{bmatrix}\right\| \leqslant 4\xi_i^2(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)+t_i\\ &\left\|\begin{bmatrix}2(u-2d_i\beta+d_i^2-v+2s_i^{\mathrm{T}}x-\| s_i \|^2)\\ 4(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)-g_i\end{bmatrix}\right\| \leqslant 4(v-2s_i^{\mathrm{T}}x+\| s_i \|^2)+g_i\\ &\left\|\begin{bmatrix}2x\\ v-1\end{bmatrix}\right\| \leqslant v+1\\ &\left\|\begin{bmatrix}2\beta\\ u-1\end{bmatrix}\right\| \leqslant u+1\\ &\left\|\begin{bmatrix}2\rho\\ h-1\end{bmatrix}\right\| \leqslant h+1\end{aligned} \tag{15}$$

这是一个凸优化问题，可以使用通用 CVX 工具箱来求解。

4 计算机仿真分析

本节通过 MATLAB 蒙特卡洛实验来验证本文所提算法的性能，其中包括：参考文献[9]中的 RSS-SOCP 方法、参考文献[10]中的 TOA-SDP 方法以及参考文献[11]中的 GTRS 方法。仿真环境设置为：所有的传感器节点随机部署在 30 m×30 m 的正方形区域内。目标节点与锚节点之间的测量噪声 n_i 和 m_i 服从高斯分布，$n_i \sim \mathcal{N}(0,\sigma_{n_i}^2)$，$m_i \sim \mathcal{N}(0,\sigma_{m_i}^2)$，非视距误差 b_i 和 β_i 服从均匀分布，$b_i \sim U[0,b_{\max}]$，$\beta_i \sim U[0,\beta_{\max}]$。设置传感器

节点初始发射功率 $P_0 = 20$ dBm，路径损失指数 γ=3，参考距离 $d_0 = 1$ m，蒙特卡洛循环次数 $M_c = 10\,000$。本文以均方误差（root mean square error，RMSE）评估所有定位算法的性能，其定义为：

$$\mathrm{RMSE} = \sqrt{\frac{1}{M_c}\sum_{i=1}^{M_c} \| \hat{x} - x_i \|^2} \qquad (16)$$

其中，$\hat{x}_i$ 是第 i 次蒙特卡洛运行中真实目标位置 x_i 的估计值。

在本次实验中，设置锚节点个数为 8 个，非视距链路个数为 6 个，视距链路为 2 个，最大非视距误差 $b_{\max} = 6$ dB，$\beta_{\max} = 6$ m。图 2 比较了不同方法的 RMSE 随噪声标准差的变化曲线。从图 2 中可以看出，随着噪声标准差的变大，各种方法的 RMSE 均呈上升趋势，性能均变差。进一步，对于各种不同的噪声标准差，与其他两种单一方式定位方法比较，本文提出的联合方法和参考文献[11]中的联合方法均方误差更小，因此联合定位方法性能要明显优于单一方式定位方法。此外，在所有讨论方法中，本文方法均方误差最低，定位性能最好。

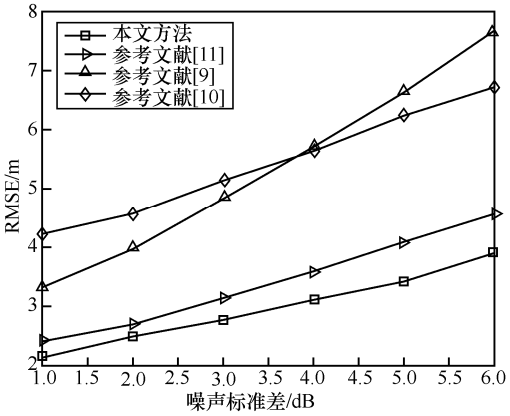


图 2　不同方法的 RMSE 随噪声标准差的变化

在本组实验中，设置锚节点个数为 8 个，非视距链路个数为 6 个，视距链路为 2 个，最大非视距误差 $b_{\max} = 6$ dB，$\beta_{\max} = 6$ m。图 3 比较了 4 种方法的累积分布函数（cumulative distribution function，CDF）随估计误差的变化曲线。从图 3 中可知，在估计误差范围内，本文所提算法都具有较好的性能。具体来说，当估计误差为 4 m 时，本文提出方法的累积分布函数可以达到 90%，而其他 3 种方法均未达到 90%。当估计误差为 6 m 时，本文方法的累积分布函数可以达到 99%，而参考文献[11]中达到 97%，参考文献[9-10]中的单一方式定位方法性能较差，累积分布函数分别为 81% 和 78%，因而在估计误差同等假设条件下，本文所提算法与其他方法比较，定位性能最好。

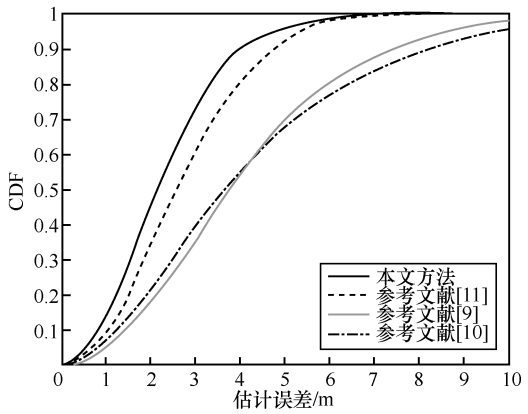


图 3　不同方法的 CDF 随估计误差的变化

在本次仿真中，设置锚节点个数为 8 个，噪声标准差 $\sigma_{n_i} = 3$ dB，$\sigma_{m_i} = 3$ m。最大非视距误差 $b_{\max} = 6$ dB，$\beta_{\max} = 6$ m。图 4 比较了 RMSE 随着非视距链路的个数的变化曲线。从图 4 中可以看出，随着非视距链路个数的增加，本文方法和参考文献[11]中的联合定位方法相对于单一方式定位方法的均方误差更小，都表现出很好的非视距偏差的抑制能力，而本文所提方法随着非视距链路数的增加，其均方误差性能变化幅度较小，并且相对于其他几种方法均方误差最小，可以看出其抑制非视距偏差能力最好。

在本组仿真中，设置锚节点个数为 8 个，非视距链路个数为 8 个，$\sigma_{n_i} = 3$ dB，$\sigma_{m_i} = 3$ m，最大非视距误差 $b_{\max} = 6$ dB，$\beta_{\max} = 6$ m。图 5 比较

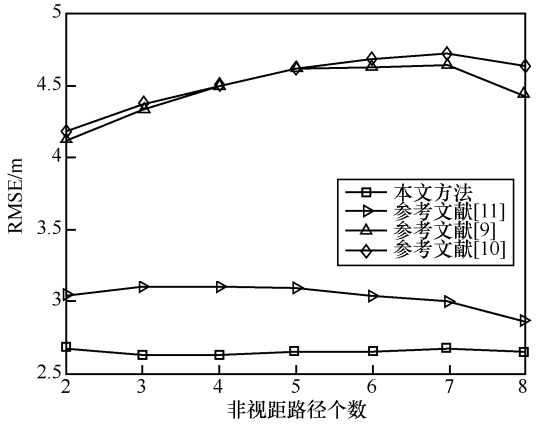


图 4　不同方法的 RMSE 随非视距路径个数的变化

了 RMSE 随着非视距误差大小的变化曲线。从图 5 中可以看出，随着非视距误差大小的增大，本文方法对非视距误差大的情况下敏感性较强，定位精度呈现出一种缓慢恶化的趋势。在非视距误差的较小的情况下，本文方法相对于参考文献[11]定位精度具有明显的优势。随着非视距误差的误差增大，本文方法虽然与参考文献[11]方法的 RMSE 趋于接近，但本文方法的 RMSE 性能总体上还是较低，也表现出较好的定位性能，此外，也优于其他单一定位方法。

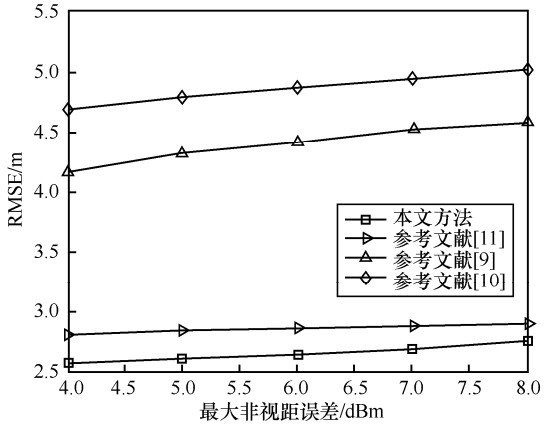


图 5　不同方法的 RMSE 随最大非视距误差的变化

5　结束语

本文针对非视距环境下联合 RSS 和 TOA 定位问题进行研究，提出了一种运用二阶锥松弛的技术，将定位目标函数松弛为一种凸优化问题的无线传感器网络联合定位方法。由于本文联合定位方法，可以从多渠道得到更多的可用信息，不仅可以测量 RSS 接收功率信息，也可以测量 TOA 距离信息。相对于传统单一定位方法测量得到的可用信息更多，能够有效地抑制非视距误差的影响，提高了定位的精度。此外，仿真结果表明本文的联合方法在非视距环境下的均方根误差和累积分布函数的性能上优于传统单一的 RSS 和 TOA 定位方法，也优于现有的联合方法，具有较高的定位精度，达到了较好的定位性能。

参考文献：

[1] YAQOOB I, AHMED E, HASHEM I A T, et al. Internet of things architecture: recent advances, taxonomy, requirements, and open challenges[J]. IEEE Wireless Communications, 2017, 24(3): 10-16.

[2] AKYILDIZ I F, SU W, SANKARASUBRAMANIAM Y, et al. Wireless sensor networks: a survey[C]//International Conference on Advanced Information Networking & Applications Workshops, May 26-29, 2009, Bradford, UK. Piscataway: IEEE Press, 2009: 393-422.

[3] PATWARI N, ASH J N, KYPEROUNTAS S, et al. Locating the nodes: cooperative localization in wireless sensor networks[J]. IEEE Signal Processing Magazine, 2005, 22(4): 54-69.

[4] 卢倩倩, 李有明, 常生明, 等. 基于到达时间的无线传感器网络协作定位算法[J]. 电信科学, 2019, 35(1): 62-66.
LU Q Q, LI Y M, CHANG S M, et al. Cooperative localization algorithm based on time-of arrival in wireless senor network[J]. Telecommunications Science, 2019, 35(1): 62-66.

[5] LIU C F, YANG J, WANG F S. Joint TDOA and AOA location algorithm[J]. Journal of Systems Engineering & Electronics, 2013, 24(2): 183-188.

[6] XING J J, FU C G. Passive localization based on short time TDOA sequence[J]. Acta Aeronautica Et Astronautica Sinica, 2011, 55(4): 25-31.

[7] BAHL P, PADMANABHAN V N, RADA R. An in-building RF-based user location and tracking system[C]//IEEE INFOCOM 2000, March 26-30, 2000, Tel Aviv, Israel. Piscataway: IEEE Press, 2000.

[8] GUVENC I, CHONG C C. A survey on TOA based wireless localization and NLOS mitigation techniques[J]. IEEE Communications Surveys and Tutorials, 2009, 11(3): 107-124.

[9] TOMIC S, BEKO M, DINIS R. RSS-based localization in wireless sensor networks using convex relaxation: noncooperative and cooperative schemes[J]. IEEE Transactions on Vehicular Technology, 2015, 64(5): 2037-2050.

[10] WANG G, CHEN H, LI Y, et al. NLOS error mitigation for TOA-based localization via convex relaxation[J]. IEEE Transactions on Wireless Communications, 2014, 13(8): 4119-4131.

[11] TOMIC S, BEKO M, TUBA M, et al. Target localization in NLOS environments using RSS and TOA measurements[J]. IEEE Wireless Communications Letters, 2018, 13(12): 224-229.

[12] CHAN Y T, HANG H, CHING P C. Exact and approximate maximum likelihood localization algorithms[J]. IEEE Transactions on Vehicular Technology, 2006, 55(1): 10-16.

[13] GUVENCU I, CHONG C C, WATANABE F. NLOS identification and mitigation for UWB localization systems[C]//2007 IEEE Wireless Communications and Networking Conference, Mar 11-15, 2007, Hong Kong, China. Piscataway: IEEE Press, 2007: 1571-1576.

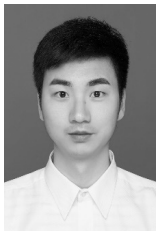
[14] TOMIC S, BEKO M, DINIS R. 3-D target localization in wireless sensor network using RSS and AoA measurements[J]. IEEE Transactions on Vehicular Technology, 2016, 13(2): 1242-1254.

[15] SHI X, MAO G, YANG Z, et al. Localization algorithm design and performance analysis in probabilistic LOS/NLOS environment[C]//2016 IEEE International Conference on Communications, May 22-27, 2016, Kuala Lumpur, Malaysia. Piscataway: IEEE Press, 2016: 1-6.

[16] SALARI S, SHAHBAZPANAHI S, OZDEMIR K. Mobility aided wireless sensor network localization via semidefinite programming[J]. IEEE Transactions on Wireless Communications, 2013, 12(12): 5966-5978.

[17] ZHANG J, DING L, WANG Y, et al. Measurement-based indoor NLOS TOA/RSS range error modelling[J]. Electronics Letters, 2016, 52(2): 165-167.

[作者简介]



卢志刚（1995− ），男，宁波大学信息科学与工程学院硕士生，主要研究方向为无线传感网络定位。

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

贾向红（1971− ），女，中国联合网络通信有限公司高级工程师，主要研究方向为通信及 IT 应用。

常生明（1982− ），男，宁波大学信息科学与工程学院博士生，主要研究方向为无线传感网络定位。

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

综述

区块链在蜂窝移动通信系统中的研究现状及发展趋势

左益平[1]，金石[1]，张胜利[2]
（1. 东南大学移动通信国家重点实验室，江苏 南京 210096;
2. 深圳大学区块链技术研究中心，广东 深圳 518060）

摘　要：区块链本质上是分布式数据库，无需第三方中介机构即可安全更新状态。将区块链技术引入 6G 蜂窝移动通信系统中以保障用户的隐私安全，减少资源分配和通信服务成本，支持不同分布式应用，从而实现移动通信和区块链技术的有机结合，被预测为 6G 蜂窝移动通信的关键技术之一。从区块链结合物联网（IoT）、边缘计算、频谱分配、干扰管理方面展开了详细的介绍，阐述了近年来国际学术界在该方向的最新研究进展，并在此基础上对 6G 蜂窝移动通信中区块链技术的发展趋势进行了进一步的展望。

关键词：区块链；物联网；边缘计算；频谱分配
中图分类号：TN929.5
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019192

Research status and development trend of blockchain in cellular mobile communication system

ZUO Yiping[1], JIN Shi[1], ZHANG Shengli[2]
1. National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
2. Blockchain Technology Research Center, Shenzhen University, Shenzhen 518060, China

Abstract: The blockchain is essentially a distributed database that can be safely updated without the need for third-party intermediaries. The blockchain technology is introduced into the 6G cellular mobile communication system to ensure the privacy of users, reduce the cost of resource allocation and communication services, and support different distributed applications. In order to realize the organic combination of mobile communication and blockchain technology, it is predicted to be one of the key technologies of 6G cellular mobile communication. From the blockchain combined with internet of things (IoT), edge computing, spectrum allocation, interference management, a detailed introduction was introduced, and the latest research progress of international academic circles in this direction was expounded in recent years, and on this basis, the development trend of blockchain technology in 6G cellular mobile communication was further prospected.

Key words: blockchain, internet of things, edge computing, spectrum allocation

收稿日期：2019−04−23；修回日期：2019−07−19
基金项目：国家自然科学基金杰出青年科学基金资助项目（No.61625106）；国家自然科学基金资助项目（No.61531011）
Foundation Items: NSFC for Distinguished Young Scholars of China (No.61625106), The National Natural Science Foundation of China (No.61531011)

1　引言

大约每一个十年，都会有新一代蜂窝移动通信系统被设计和部署，大大改变了人们工作和生活的方式。1G 移动通信网络仅用于语音呼叫；2G 移动通信技术是一种数字技术并支持文本消息传递；3G 移动技术提供更高的数据传输速率和容量，并支持交互式媒体业务；4G 移动通信技术将 3G 与固定互联网相结合，以帮助无线移动互联网，并且它超越了 3G 的局限，带宽增加，资源成本降低。5G[1]相比 4G 无线传输速率提升 10~100 倍，峰值传输速率达 10 Gbit/s，端到端时延减少至毫秒级，连接设备密度增加 10~100 倍，流量密度提高 1 000 倍，频谱效率提升 5~10 倍，能够在 500 km/h 的速度下保证用户体验。5G 主要从 3 个维度实现上述指标，即空口增强、更宽的频谱和网络密集化，代表性技术分别对应着大规模 MIMO、毫米波通信以及超密集组网。5G 尚未正式使用，但其部署已经开始，预计将在 2020 年出现。然而 5G 无法解决许多未来需求，例如极高的比特率、超低时延、高能效、抗阻塞性和无线充电等，所以现在有很多国家已经开始关注 6G[2-3]领域了。

6G 既要在 5G“大带宽、低时延、广联接”这 3 个应用场景上实现更好的通信基础服务，又要实现更多的应用场景的突破。6G 时代才是真正的万物万联时代，因为从目前来看，5G 在广联接也就是 IoT 的应用还不太理想，6G 可能会在这个场景上扩展，向着海陆空一体化融合网络的方向发展。目前，中国、美国、欧洲、韩国、日本、芬兰等国家和地区已陆续着手 B5G 或者 6G 技术的研发。不过现在还未得到统一的 6G 定义，各方都有自己的愿景。美国 FCC（Federal Communications Commission）认为 6G 将进入太赫兹（THz）频段，频段更高，基站的覆盖范围会变小，导致网络更加致密化；6G 将使用空间复用技术，6G 基站可同时接入数百个甚至数千个无线连接，其容量将会是 5G 的 1 000 倍；美国 FCC 还认为 6G 将会采用更智能、分布更强的基于区块链的动态频谱共享接入技术。欧盟认为 6G 就是 5G 结合卫星通信/卫星导航，可实现国际上几大主流卫星导航的互通漫游。韩国认为 6G 技术方案是“太赫兹+去蜂窝化结构+高空无线平台（如卫星等）”，其中，去蜂窝结构是指 6G 基站未必按照蜂窝状布置，终端未必只和一个基站通信，这种结构能够大大提高频谱效率。日本认为 6G 应该引入大规模 MIMO 技术与 OAM 技术的结合，有望在未来实现 40 个电波的叠加传输。最近在芬兰开的全球首届 6G 峰会的主办方、全球最早开始研究 6G 的科研机构——奥卢大学无线通信中心将即时通信与无限制连接、分布式计算和智能、极高频段材料和天线作为 6G 的主要研究方向。我国也启动了对 6G 的研发工作，科学技术部发布了关于国家重点研发计划“宽带通信和新型网络”重点专项，如“大规模无线通信物理层基础理论与技术”以及“太赫兹无线通信技术与系统”等，都是与 6G 相关的。可想而知，6G 蜂窝移动通信系统下会产生海量的无线大数据，资源分配和调度复杂度极高，用户个人信息的安全性和隐私性遭到更多方面的威胁。现有的技术手段还不能完全解决上述一系列问题，所以需要将创新性的技术加入 6G 中，以解决 6G 蜂窝移动通信系统的瓶颈问题。

另一方面，区块链作为一项新兴的前沿技术[4-5]，近年来受到了广泛的关注，是因为比特币使用了一个分布式的区块链来跟踪所有的交易，但实际上区块链技术并不局限于比特币一种应用。区块链技术天然地拥有去中心化、分布式、防篡改等优点，所以移动通信领域的研究者期望将其应用于系统的各个层面，期望大幅度提升蜂窝移动通信系统的性能，实现真正意义的万物万联。美国 FCC 在 MWCA2018 大会上也提出利用区块链进行动态频谱分配的构想，因此区块链技术被认为是 6G 的关键技术之一，其基本思想就是

利用区块链技术的去中心化、分布式、防篡改等特点，将区块链技术引入 6G 蜂窝移动通信系统中保障用户的隐私安全，减少资源分配和通信服务成本，从而实现移动通信和区块链技术的有机结合。学术界和工业界正在上述领域开展研究工作，使用区块链去中心化的分布式账本来记录各种无线接入信息的研究刚刚起步，未来可能会进一步激发新技术的创新，甚至改变未来使用无线频谱的方式。

尽管区块链技术融入蜂窝移动通信系统是趋势所在，但是研究都处于构想和初步探索阶段，机遇和挑战并存。追溯历史，蜂窝移动通信系统从 1G 到 5G 的发展，基本上解决了面向个人终端的陆地的网络覆盖问题，实现了信息的传播。未来移动通信向着海陆空一体化发展，加入区块链技术以实现信息价值的转移，IoT、资源的配置、隐私安全等方面一定会出现很多有意思的变化。目前这方面的研究面临诸多挑战，国内外研究者进行了初步构想和探索。本文主要介绍区块链技术应用于蜂窝移动通信技术的最新研究进展，主要包括 IoT、边缘计算、频谱分配、干扰管理。

2 区块链简介

2008 年 10 月 31 号，Nakamoto[6]第一次提出了区块链的概念，在随后几年中成为电子货币比特币的核心组成部分，作为所有交易的公共账本。通过利用点对点网络和分布式时间戳服务器，区块链数据库能够进行自主管理，无需第三方机构的辅助。比特币即“区块链 1.0”的设计思想已经成为其他应用程序的灵感来源。2013—2014 年间，程序员 Vitalik Buterin 受比特币启发后提出了以太坊即“区块链 2.0”，它是一个开源的有智能合约功能的公共区块链平台，通过其专用加密货币——以太币提供去中心化的以太虚拟机来处理点对点合约。下面详细介绍比特币、以太坊的基本原理以及区块链技术的发展现状。

2.1 比特币

2.1.1 区块

区块分为区块头和区块体两部分：区块头包含前一个区块的散列值、难度值、当前区块所有交易的 Merkle 根节点的散列值、时间戳和随机数，区块体包含当前区块的所有交易信息。区块链就是按创建的时间顺序进行排列的区块链条，它十分完美地实现了一个安全可靠、不断推进的比特币交易数据库，区块数据结构示意图如图 1 所示。比特币系统大约每 10 min 产生一个区块，该区块包含这 10 min 内未确认的交易以及前一个区块的散列值，因此从第一个区块问世至今就形成了一条完整的区块链。

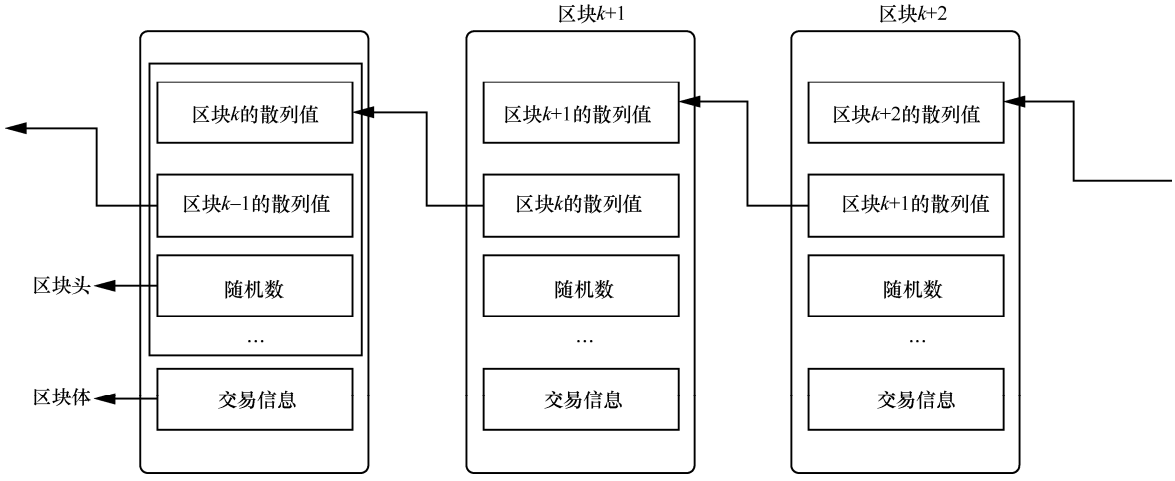


图 1 区块数据结构示意图

2.1.2　PoW

PoW 共识机制的本质是防止低算力的实体随意或恶意发布区块。散列函数的输入为区块头，输出是一个 256 bit 的散列值。比特币系统会把每个区块完成的时间控制在 10 min 左右。如果难度低于 10 min，系统就自动调高难度值，增加散列值开头 0 的位数；如果难度高于 10 min，就适当减少散列值开头 0 的位数，以调低难度值。这个 PoW 的过程被称为挖矿，挖矿流程如图 2 所示，挖矿前要构造好区块，通过不停改变随机数 Nonce 值使得区块头的散列值小于目标难度值。

挖矿的本质是争夺记账权，矿工收集、检验和确认过去一段时间内发生的交易。当找到一个符合 PoW 机制的散列值时，矿工就能够将自己封装的区块广播出去，让其他矿工验证该区块。如果有矿工接受该区块并以它为基础继续挖下一个区块，那么该区块中的所有交易单就获得一次确认。每延长一个区块就等价于该区块中的交易多了一次确认。若得到 6 次确认，那么该区块就获得全网的认可，封装到历史区块中。区块链可以记录任意两个节点的交易，且不需要第三方机构就可以在网络中更新节点的信息状态，本质上是分布式数据存储、点对点传输、共识机制、加密算法等技术融合的新型技术。

2.2　以太坊

以太坊就是一个开源的区块链平台，用户可以在该平台上按需创建应用或程序，以实现去中心化的自治。因此，以太坊是一个开源且去中心化的区块链平台或应用集合。以太坊的原理和比特币基本类似，以太坊最大的创新就是引入智能合约。智能合约就是写在区块链上的代码，一旦某个事件触发合约中的条款，代码即自动执行。智能合约由矿工执行，执行结果记录在区块链上。基于区块链技术的智能合约不仅可以发挥智能合约在成本效率方面的优势，而且可以避免恶意行为对合约正常执行的干扰。将智能合约以数字化的形式写入区块链中，由区块链技术的特性保障存储、读取、执行整个过程透明可跟踪、难以篡改。同时，由区块链自带的共识算法构建出一套状态机系统，使智能合约能够高效地运行。

2.3　区块链技术的发展现状

在过去 10 年中，区块链技术的发展十分迅猛，区块链不可变的数据组织框架和区块链网络的数据调度及维护方法是推进区块链技术发展的主要动力。从数据组织的角度看，区块链技术采用许多现成的加密技术[7]，加密地将用户的伪身份与标记的资产交易联系起来，所以区块链能证明资产的所有权。区块链通过“块”的形式将交易集合加密连接到前面的交易目录中来维持交易记录的次序。另一方面，对任意节点具有容忍度的开放网络中，区块链网络中的共识机制创造性地解决了交易记录重复的问题[8]。与传统的共识机制相比，在没有身份认证和更小消息传输的开销条

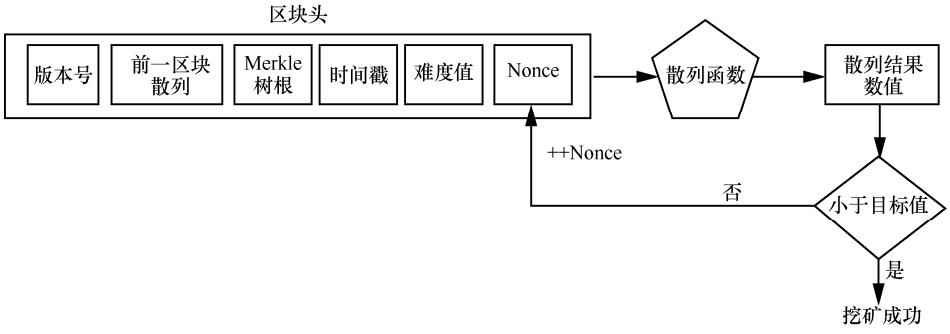


图 2　PoW 挖矿流程

件下，区块链共识机制能够在大量无信任节点之间提供关于全局分类账—数据状态的共识[9]。在这个意义上，区块链可以被看作网络的通用存储器，并且网络可以被看作包括每个节点的分布式虚拟机（virtual machine，VM）[10]。因此，区块链网络提供了一个通用平台，用于在分布式的应用程序中执行事务驱动的指令，例如自组织网络编排[11]和 P2P 资源交易[12]。

尽管随着区块链技术已经在快速发展，但有着更高服务质量需求的区块链应用，对区块链机制的设计提出了更严峻的挑战。特别地，区块链网络的性能极大地依赖于所采用的共识机制性能，与可扩展性非常有限的经典机制相比[13]，公有区块链网络（例如比特币）中的大多数共识机制以处理吞吐量为代价保证了更好的可扩展性。为了实现无信任节点之间分布式同步和确保交易完成，许多机制也对物理资源（例如计算能力）消耗产生了巨大需求[9]以及需要施加更长的等待时间来确认交易。参考文献[9]对于上述这些问题，进行了一些简短的调查研究，目的是在特定方面提高区块链共识机制的性能。

与此同时，在过去的 10 年中，区块链已经从不可变的、防篡改的分布式账本进行了进一步扩展。例如参考文献[14-15]讨论了关于比特币范围内的用户端应用（即钱包）、P2P 网络协议、共识机制和用户隐私的问题。参考文献[16]提供了从金融到物联网的新兴区块链应用的简要总结。此外，参考文献[17]提供了关于比特币网络中的安全问题的系统调查，包括对共识机制的攻击以及比特币客户端的隐私/匿名问题。参考文献[18-19]在以太坊网络范围内，提供了对智能合约的设计、应用和安全性等特殊问题的调查。

3 区块链在 6G 中的应用

6G 蜂窝移动通信系统将会带来超高数据量以及设备的大规模连接，通过其低时延、高速度和高容量的覆盖能力，可以使 IoT 设备获得更加广泛的应用。6G 移动通信网络边缘产生的数据量还在不断增加，直接在网络的边缘进行数据处理会更加高效，边缘计算毫无疑问是 6G 中还很有发展前景的技术。同时无线频谱一直是稀缺资源，其需求持续快速增长。另一方面，随着网络越加致密化，网络基础设施的投资成本正在不断加大，如何高效利用频谱资源将会是 6G 的核心挑战。6G 的网络致密化部署将会导致严重的干扰问题，所以干扰管理问题也是 6G 的关键研究方向。综上所述，6G 中的 IoT、边缘计算、频谱分配、干扰管理还是特别重要的研究课题。所以本节面向区块链在 6G 移动通信系统的应用，从 IoT、边缘计算、频谱分配、干扰管理 4 个方面展开详细的介绍，展示了近年来国际学术界在该方面的最新研究进展。区块链在蜂窝移动通信系统中的研究现状分类见表 1。

3.1 IoT

现在存在的中心化的网络，包括数据中心的高额维护成本等在内的诸多瓶颈，并不能适应万物万联的要求。未来 6G 的 IoT 需要一个去中心化的网络架构，即采用区块链技术。参考文献[20-23]都研究了基于区块链的 IoT 技术，参考文献[22]分析了在区块链网络中同步的 IoT 设备的通信流量，因为 IoT 设备需要连接到一个或多个验证器节点以观察或修改账户的状态，为了与最新的账户状态进行交互，需要将设备与验证器节点存储的区块链副本同步。介绍了具有不同通信成本和安全级别的同步协议，使得 IoT 端点与区块链同步，并且对同步协议产生的流量进行建模和分析。参考文献[22]提出 IoT 设备和区块链网络同步的两种协议，讨论了 IoT 设备的本地状态到区块链网络的全局状态的转移概率，受同步协议运行时间、设备睡眠时间、区块的大小等参数的影响。分析和仿真结果表明，保证同步过程可靠性的无线技术选择取决于区块链参数。而且如果协议执

表 1　区块链在蜂窝移动通信系统中的研究现状分类

研究方向	主要参考文献
物联网（IoT）	参考文献[22-23]主要讨论了 IoT 设备的局部状态与区块链网络的全局状态保持同步的过程
边缘计算	参考文献[24-27]提出将区块链 PoW 挖矿任务卸载到边缘计算，构建两阶段的 Stackelberg 博弈和拍卖机制来模拟边缘计算服务器和移动区块链用户的交互过程
频谱分配	参考文献[28]提出基于智能合约的频谱分配方案，移动网络运营商可以利用其他主要用户未使用的频谱来扩容；参考文献[29]利用区块链的工作方式模拟了主次要用户之间的频谱租用过程
干扰管理	参考文献[30]提出利用区块链的货币机制和协调协议的贪婪分布式算法实现以前采用中央控制器达到的最优信息分配，消除了用户间的干扰

行持续时间与平均块生成周期相当，则保持同步的概率迅速降低。最后发现区块链协议与主要产生上行链路流量的典型 IoT 应用不同，需要分配大量的下行链路资源。

参考文献[23]分析了具有轻量级 IoT 客户端的区块链系统的时延和通信权衡。区块链技术为 IoT 设备的协调提供了分布式架构，但是保留区块链分类账本的本地副本对于低功耗和内存受限的设备是不可行的。出于这个原因，提出用轻量级软件实现，仅下载有用的数据结构。提出了一种新颖的基于区块链网络的压缩方案，该方案在周期性更新中压缩区块链数据并进一步降低 IoT 设备的通信成本。参考文献[23]讨论了区块链将同步信息发送到 IoT 设备的下行链路帧长度的分布、帧的传输时间以及利用它们计算通信成本和时延。分析和仿真结果表明，如果已知账户更新和信道状态的统计信息，则轻量级客户端可以构建感兴趣的事件列表，提供可预测的平均通信成本，该方案可用于设计更高级的区块链轻量级协议。

3.2　边缘计算

现有的针对区块链和边缘计算的研究，大部分是因为 IoT 与区块链的结合虽然提高了系统的性能，但是区块链本身的挖矿过程需要很大的算力，不符合 IoT 设备计算能力相对较低、功耗较低以及零散低带宽的无线连接的特点，边缘计算技术可以解决这个问题。

参考文献[24]和参考文献[25]研究的问题相同，而且都提出了一种解决方案，具体是将区块链 PoW 挖矿任务卸载到附近的移动边缘计算服务器，允许移动用户访问和利用边缘的资源或计算服务。然而，边缘计算服务由边缘服务提供商（edge services provider，ESP）部署，首要目标是最大化自己的利益，因此使用两阶段 Stackelberg 博弈提供的定价方案来模拟 ESP 和移动用户之间的交互。参考文献[25]对两阶段 Stackelberg 博弈模型进行了更加详细的推导和分析。通过反向归纳得出了博弈的纳什均衡点，提出了 ESP 分别采用统一和歧视定价的最佳资源管理方案。此外，在两种定价方案中，Stackelberg 均衡的存在性和唯一性都已经得到了证明。

参考文献[26]采用基于拍卖理论的方案，其中 ESP 尝试制定实用的分配计算资源策略，首先，ESP 向移动用户宣布其资源服务，然后移动用户提交他们对计算资源的估值。最后 ESP 选择给赢得拍卖的用户提供服务，并通知所有用户当前的分配。对于需求不断的矿工，提出了一种可以实现最佳矿工收益的拍卖机制。对于有多种需求的矿工，将矿工收益最大化问题转化为非单调子模块最大化与背包约束问题。然后，设计了两种有效的机制，最大限度地提高矿工的收益。实验结果已经证明，所提出的拍卖机制是真实的、个体

理性的和高计算效率的，并且能够解决矿工收益最大化问题。

参考文献[27]对参考文献[26]提出的拍卖方案进行了增强，引入了深度学习的方法，具体来说，构建了一个基于最优拍卖解决方案的多层神经网络架构。神经网络首先对矿工的出价进行单调变换，然后计算矿工的分配和有条件付款规则。使用矿工的估值作为训练数据来调整神经网络的参数，以优化 ESP 的损失函数。实验结果证明了使用深度学习的好处，可获得高收益的移动区块链的最佳计算资源拍卖。

上述参考文献提出的区块链和边缘计算结合的算法，分别通过两阶段 Stackelberg 博弈定价和基于拍卖理论的方案来模拟 ESP 和区块链移动用户之间的交互过程，以用户的任务量和服务器收取的费用为技术指标，讨论了用户和服务器为了保证自身利益如何来选择最佳的任务和费用策略。可以看出，边缘计算作为辅助工具可以改进区块链本身的算力复杂度高的缺陷。未来的一个研究方向可以是区块链技术结合人工智能技术对边缘计算的安全隐私性和资源分配等方面进行改善。

3.3 频谱分配

从 1G 到 5G，为了提高速率、提升容量，移动通信永远向着更多的频谱、更高的频段扩展。无线频段越高，单位时间内所能传递的数据量就越大，覆盖范围越小，5G 的基站密度比 4G 高很多，为了提升覆盖范围和网络速率，将在基站端引入大规模 MIMO 和波束成形技术。5G 已经扩展到毫米波频段（3~6 GHz），6G 将使用太赫兹频段（100 GHz~10 THz），6G 的基站密集度将会更高。无线频谱是稀缺资源，其需求正在快速增长，另一方面，随着网络更加致密化，网络基础设施的投资成本正在不断加大。如何高效利用频谱资源将会是 6G 的关键话题之一。美国 FCC 提出 6G 可以采用更智能、分布更强的基于区块链的动态频谱共享接入技术，而不再通过集中式的数据库来支持频谱共享接入，不仅可以降低动态频谱接入系统的管理费用，提升频谱效率，还能进一步增加接入等级、接入用户数量等。6G 对原来被 3G/4G 占用的频谱的重新使用，使得整个移动网络的效率更高。并且借助区块链技术来实现分层次多中心的频谱使用共识和动态分配，即使是跨牌照协同也成为技术性的可能和商业性的必然。

从基于区块链技术的频谱分配出发，参考文献[26]提出了基于智能合约的频谱分配方案，移动网络运营商可以通过聚合其他主要用户未使用的许可频谱来扩展其容量。为了准确识别频谱机会，移动网络必须部署多个传感器，或者它可以将此任务卸载到具有感知能力的附近节点，即所谓的帮助者。但是对于帮助者执行频谱感知的激励是有限的，所以笔者引入了一种基于智能合约的频谱感知服务解决方案，促进网络运营商从其附近的帮助者购买频谱服务。参考文献[28]讨论了基于智能合约的主用户频谱感知的检测概率，恶意节点越多，频谱检测概率就越低。每秒次要用户的净收益等于总收益减去成本，净收益必须是正的，可以推导出最小的频谱效率表达式，分析了频谱感知帮助者的数目、写入智能合约的数据压缩系数对频谱效率的影响。参考文献[29]利用区块链的工作方式模拟了主次要用户之间租用频谱的过程，并且设计了相应的同步协议。参考文献[28]和参考文献[29]提出的基于区块链的频谱分配方法，都比传统拍卖更公平、更安全，并且节约了部分第三方中介的成本。但是现有文献对于区块链帮助频谱分配方案还都是一些表面的构想，具体实现和更深入的研究工作还有待后续的科研人员好好探索。

3.4 干扰管理

Gamal 等人[30]提出了贪婪分布式算法，利用区块链货币机制和协调协议，能够实现以前采用

中央控制器达到的最优信息分配，消除了用户之间的干扰。货币机制假设只有一种硬币类型，硬币能从一个发送机—接收机对节点到下一个节点，因为硬币从付款节点 1 到收款节点 2，付款人的发送机对收款人的接收机造成困扰，所以付款人就需要租用付款人的发送机来消除干扰。CM1（协调信息 1）是节点 i 愿意接受节点 i+1 的汇款，但要关闭接收机 i；CM2 是节点 i 愿意接受节点 i+1 的汇款，同时使用接收机 i；CM3 是节点 i 不愿意接受节点 i+1 的汇款，节点 i+1 是激活状态，这时从接收机 i+1 端可以看到用户间是无干扰的；CM4 是节点 i 不愿意接受节点 i+1 的汇款，节点 i+1 是非激活状态，这时从接收机 i+1 端可以看到用户间的干扰是无法消除的，因为整个回程负载约束失效了。基于上述区块链的货币机制和协调协议，提出了贪婪分布式算法，可以完美实现以前采用中央控制器得到的最优信息分配。

参考文献[30]具有一种硬币类型的货币机制和依赖于 2 bit 消息的协调协议实现，传输方案基于协作迫零和干扰避免。节点支付硬币是因为在收款人的接收器上造成干扰，所以要出租收款人的发射器消除干扰。在未来的工作中研究如何将所提出的方案扩展到一般网络拓扑，特别是硬币的数量、协调协议的开销以及具有任意拓扑的网络大小的时延标度。区块链技术在干扰管理方面的详细实施，也是一个未来的研究方向。

4　未来发展方向

由本文第 3 节的分析可知，区块链在蜂窝移动通信领域的应用，主要思想是将区块链技术引入 IoT、边缘计算、频谱分配、干扰管理领域。虽然区块链技术应用在蜂窝移动通信系统中的前景很广阔，但是目前还面临着诸多瓶颈问题。首先区块链技术本身存在一些问题，比如区块链技术的可扩展性差，挖矿需要的算力很大，确认交易的时间很长，节点规模、性能、容错性三者之间难以平衡，目前缺乏统一的区块链技术应用标准等。其次面临的问题是如何将区块链应用到蜂窝移动通信的各种场景中，采用什么样的技术架构，在性能指标、安全性、稳定性等方面缺乏指导性的思路和目标。毫无疑问，区块链技术应用到蜂窝移动通信系统中还有许多工作需要展开和完善。结合已有的研究工作，展望了区块链技术在蜂窝移动通信系统中的应用，可能会出现下述研究方向。

（1）区块链与 IoT 结合

当前绝大多数 IoT 设备基于中心化的分布式架构，边缘节点仍受到中心化的核心节点能力的制约。可以考虑利用区块链的去中心化思想，研究如何把物联网的核心节点的能力下放到各个边缘节点，让核心节点仅控制核心内容或做备份使用，各边缘节点为各自区域内设备服务。并且可以研究如何设计更加灵活的协作模式以及相关共识机制，完成原核心节点承担的认证、账务控制等功能，使得区块链在这种平台架构中扮演公开账本的角色，对 IoT 设备之间所有的信息交换进行安全、可信和稳定的记录。

（2）区块链与边缘计算结合

IoT 设备有限的计算能力和存储能力是制约区块链应用的重要瓶颈，但边缘计算可以解决这一问题。以移动边缘计算为例，未来可以研究移动边缘计算服务器如何替移动区块链网络下的 IoT 设备完成计算、存储等任务，并在区块链技术的帮助下保证数据的可靠性和安全性。边缘计算与区块链的融合能提高 IoT 的整体性能。同时还可以研究如何分配和协调区块链网络下的边缘计算、存储和无线通信等资源。

（3）区块链与频谱分配结合

如何高效利用频谱资源将会是 6G 的关键话题之一。美国 FCC 提出 6G 可以采用更智能、分布更强的基于区块链的动态频谱共享接入技术，而不再通过集中式的数据库来支持频谱共享接入，不仅可以降低动态频谱接入系统的管理费用，

提升频谱效率，还能进一步增加接入等级、接入用户数量等。区块链的发展可为频谱的分配管理提供新的思路。

（4）区块链与人工智能结合

到目前为止，6G 所描绘的愿景可以预见到有大量的数据交换，需要普遍使用人工智能方法，例如强化学习和深度学习，以智能有效的方式处理数据和管理无线通信资源[31-32]。显然，这在安全性、隐私性和信任方面构成了重大挑战。在真正庞大的物联网部署中，分布式身份验证是提供可伸缩性的一个关键点，区块链可以解决上述问题，同时人工智能方法也能够确定区块链的最佳参数。所以区块链结合人工智能显著改善 6G 网络的计算、通信和存储性能。这是未来区块链发展的方向之一。

总结已有研究和展望未来，可以发现区块链在 6G 中的应用很广泛，与物联网、边缘计算、资源管理结合，未来区块链在 6G 中的发展向物理层推进，还有与人工智能强强联手的发展趋势。目前这些研究都还刚刚起步，具有广阔的发展前景，面对诸多挑战。

5 结束语

本文首先介绍了区块链技术，包括比特币、以太坊的工作原理以及区块链技术的发展现状，然后详细阐述了区块链应用于蜂窝移动通信技术的最新研究成果，包括物联网、边缘计算、频谱分配、干扰管理。从最新研究进展可以看出，区块链作为 6G 发展的主流技术之一，可以利用自身网络的去中心化、安全性高等特点，对蜂窝移动通信中复杂的资源管理和隐私安全等方面进行颠覆性的改变，将会成为最具潜力的发展方向之一。

参考文献：

[1] 张静, 金石, 温朝凯, 等. 基于人工智能的无线传输技术最新研究进展[J]. 电信科学, 2018, 34(8): 46-55.
ZHANG J, JIN S, WEN C K, et al. The latest research progress of wireless transmission technology based on artificial intelligence[J]. Telecommunications Science, 2018, 34(8): 46-55.

[2] CASSARA G A. 6G: the next frontier[M]. Hoboken: Wiley, 2015.

[3] KLAUS D, HENDRIK B. 6G vision and requirements: is there any need for beyond 5G?[J]. IEEE Vehicular Technology Magazine, 2018,13(3): 72-80.

[4] 李董, 魏进武. 区块链技术原理、应用领域及挑战[J]. 电信科学, 2016, 32(12): 20-25.
LI D, WEI J W. Theory, application fields and challenge of the blockchain technology[J]. Telecommunications Science, 2016, 32(12): 20-25.

[5] 梅海涛, 刘洁. 区块链的产业现状、存在问题和政策建议[J]. 电信科学, 2016, 32(11): 134-138.
MEI H T, LIU J. Industry present situation, existing problems and strategy suggestion of blockchain[J]. Telecommunications Science, 2016, 32(11): 134-138.

[6] NAKAMOTO S. Bitcoin: a peer-to-peer electronic cash system[R]. 2008.

[7] MOHR A. A survey of zero-knowledge proofs with applications to cryptography[R]. 2007.

[8] RAYNAL M. Communication and agreement abstractions for fault-tolerant asynchronous distributed systems[J]. Synthesis Lectures on Distributed Computing Theory, 2010,1(1): 1-273.

[9] BANO S, SONNINO A, AL-BASSAM M, et al. Consensus in the age of blockchains[J]. arXiv: 1711.03936, 2017.

[10] KOSBA A, MILLER A. Hawk: the blockchain model of cryptography and privacy-preserving smart contracts[C]//2016 IEEE Symposium on Security and Privacy (SP), May 22-26, 2016, San Jose, CA, USA. Piscataway: IEEE Press, 2016: 839-858.

[11] BOZIC N, PUJOLLE G, SECCI S. Securing virtual machine orchestration with blockchains[C]//2017 Cyber Security in Networking Conference (CSNet), Oct 18-20, 2017, Rio de Janeiro, Brazil. [S.l.:s.n.], 2017: 1-8.

[12] CHATZOPOULOS D, AHMADI M. FlopCoin: a cryptocurrency for computation offloading[J]. IEEE Transactions on Mobile Computing, 2018, 17(5): 1.

[13] CASTRO M, LISKOV B. Practical byzantine fault tolerance and proactive recovery[J]. ACM Transactions on Computer Systems, 2001, 20(4): 398-461.

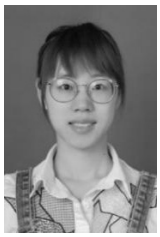
[14] TSCHORSCH F, SCHEUERMANN B. Bitcoin and beyond: a technical survey on decentralized digital currencies[J]. IEEE Communications Surveys Tutorials, 2016, 18(3): 2084-3123.

[15] BONNEAU J, MILLER A, CLARK J, et al. Sok: research perspectives and challenges for bitcoin and cryptocurrencies[C]//IEEE Symposium on Security and Privacy, May 17-21, 2015, San Jose, CA, USA. Piscataway: IEEE Press, 2015: 104-121.

[16] ZHENG Z, XIE S. Blockchain challenges and opportunities: a survey[R]. 2016.

[17] CONTI M, SANDEEP K E, LAL C, et al. A survey on security and privacy issues of bitcoin[J]. arXiv: 1706.00916, 2017.
[18] ATZEI N, BARTOLETTI M, CIMOLI T, et al. A survey of attacks on ethereum smart contracts[M]. Berlin: Springer, 2017.
[19] CHRISTIDIS K, DEVETSIKIOTIS M. Blockchains and smart contracts for the internet of things[J]. IEEE Access, 2016(4): 2292-2303.
[20] DORRI A, KANHERE S S, JURDAK R, et al. Blockchain in internet of things: challenges and solutions[J]. arXiv: 1608.05187, 2016.
[21] FENG S H, WANG W B, NIYATO D, et al. Competitive data trading in wireless-powered internet of things (IoT) crowdsensing systems with blockchain[J]. arXiv: 1808. 10217, 2018.
[22] DANZI P, KALØR A E, STEFANOVIĆ C, et al. Analysis of the communication traffic for blockchain synchronization of IoT devices[J]. arXiv: 1711.00540v1, 2018.
[23] DANZI P, KALØR A E, STEFANOVIĆ C, et al. Delay and communication tradeoffs for blockchain systems with lightweight IoT clients[J]. arXiv: 1807.07422, 2018.
[24] XIONG Z H, ZHANG Y, NIYATO D, et al. When mobile blockchain meets edge computing[J]. arXiv:1711.05938, 2018.
[25] XIONG Z H, FENG S H, WANG W B, et al. Cloud/fog computing resource management and pricing for blockchain networks[J]. arXiv: 1710.01567, 2018.
[26] JIAO Y T, WANG P, NIYATO D, et al. Auction mechanisms in cloud/fog computing resource allocation for public blockchain networks[J]. arXiv: 1804.09961, 2018.
[27] LUONG N C, XIONG Z H, WANG P, et al. Optimal auction for edge computing resource management in mobile blockchain networks: a deep learning approach[J]. arXiv: 1711.02844, 2017.
[28] BAYHAN S, ZUBOW A, WOLISZ A, et al. Spass: spectrum sensing as a service via smart contracts[C]//IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Oct. 22-25, 2018, Seoul, South Korea. Piscataway: IEEE Press, 2018: 1-10.
[29] KOTOBI K, BILEN S G. Blockchain-enabled spectrum access in cognitive radio networks[C]//2017 Wireless Telecommunications Symposium (WTS), Apr 26-28, 2017, Chicago, IL, USA. Piscataway: IEEE Press, 2017: 1-6.
[30] GAMAL A E, HESHAM E G. A blockchain example for cooperative interference management[J]. arXiv: 1808.01538, 2018.
[31] 王志宏, 杨震. 人工智能技术研究及未来智能化信息服务体系的思考[J]. 电信科学, 2017, 33(5): 1-11.
WANG Z H, YANG Z. Research on artificial intelligence technology and the future intelligent information service architecture[J]. Telecommunications Science, 2017, 33(5): 1-11.
[32] 王海坤, 潘嘉, 刘聪. 语音识别技术的研究进展与展望[J]. 电信科学, 2018, 34(2): 1-11.
WANG H K, PAN J, LIU C. Research development and forecast of automatic speech recognition technologies[J]. Telecommunications Science, 2018, 34(2): 1-11.

[作者简介]



左益平（1993− ），女，东南大学移动通信国家重点实验室博士生，主要研究方向为移动通信关键技术、区块链技术等。

金石（1974− ），男，东南大学移动通信国家重点实验室教授、博士生导师，主要研究方向为 5G/B5G 移动通信理论与关键技术、物联网理论与关键技术以及机器学习与大数据处理在无线通信中的应用等。

张胜利（1978− ），男，深圳大学区块链技术研究中心教授、博士生导师，主要研究方向为无线网络、区块链关键技术、物理层网络编码等。

全面实施携号转网对我国移动通信市场影响

胡文玉[1,2]，窦晓燕[2]

（1.首都经济贸易大学，北京 100070；2.北京中智博咨询有限公司，北京 102200）

摘　要：以我国移动用户为研究对象，采用随机抽样的调查方法，通过电话外呼的调查方式，对我国移动业务携号转网后移动通信市场格局及影响进行研究，运用“马尔可夫概率转移矩阵”对实施携号转网后全国及南北方移动电话市场份额进行了预测，并创新地提出“携号转网流向路径模型”。此外，重点围绕“区域特征、在网时长、ARPU 值和携号转网人数”等对 3 家通信运营商携号转网用户的关注点及转出和转入原因进行深入剖析。结果表明：网龄越长、ARPU 值越高，携号转网意愿越强；北方用户携号转网意愿高于南方。本文结论将为通信企业经营决策提供参考。

关键词：携号转网；马尔可夫概率转移矩阵；携号转网流向路径模型；市场份额预测；马尔可夫性检验

中图分类号：C331

文献标识码：A

doi: 10.11959/j.issn.1000−0801.2019214

Impact of the nationwide implementation of mobile number portabilityon mobile communications market

HU Wenyu[1,2]，DOU Xiaoyan[2]

1. Capital University of Economics and Business, Beijing 100070, China
2. Sciresearch Consulting Co., Ltd., Beijing 102200, China

Abstract: China’s mobile users were taken as a research object, using the random sampling survey method and the survey way of CATI (computer-assisted telephone interviewing) to conduct research on the mobile subscribers market’s structure and impact after implementation of MNP (mobile number portability). The Markov probability transfer matrix was used to predict the market share of mobile phones in China, North China and South China after the implementation of MNP. And the inflow and outflow path model was proposed innovatively. In addition, the focus was on regional characteristics, duration of use, ARPU value and number of MNP to further analyze the concerns and reasons of MNP. The result shows that the longer duration of use, the higher the ARPU value, and the stronger the willingness to choose other operators and the northern subscribers have more willing than the southern. The conclusions provide reference for the decision-making of communication operators.

Key words: MNP, Markov probability transfer matrix, inflow and outflow path model, market share forecast, Markov characteristic test

收稿日期：2019−05−16；修回日期：2019−09−02

1　引言

近年来，我国移动用户数增长速率逐渐减小，移动用户市场趋于饱和，具体变化情况如图 1 所示。截至 2018 年年底，我国移动用户累计达到 15.43 亿户，其中中国移动累计用户市场份额达到 59.9%[1]，中国联通为 20.5%[2]，中国电信为 19.6%[3]，我国移动市场格局依旧处于失衡状态。为给广大移动用户提供更加便捷的服务，打破移动电话市场一家独大的市场竞争格局，早在 2011 年工业和信息化部在天津、海南进行第一批携号转网试点，随后 2014 年在江西、湖北、云南进行第二批携号转网试点。2019 年政府工作报告中明确提出 2019 年年底前全面实施携号转网，至此我国步入携号转网推进和落实的“快车道”。携号转网对通信运营企业而言既是机遇也是挑战，将面临如何稳定在网用户的问题，确保用户转出最小化，同时面临如何制订有效策略吸引更多用户转入，以进一步扩大市场份额；对用户而言，最终去留关键是通信运营企业在“网络、业务、服务、技术、终端及内容”的整体市场竞争能力。

如今马尔可夫（Markov）链技术在预测市场份额方面的应用日趋成熟，因此本文借助马尔可夫链技术预测实施携号转网后我国 3 家运营企业的市场份额。在这方面的应用上，胡文玉等[4]运用马尔可夫模型分析了实施携号转网对我国移动通信市场份额的影响，结果表明携号转网后中国移动和中国联通受损，中国电信受益。秦书慧[5]应用马尔可夫链模型对电信市场份额进行预测。肖会敏等[6]借助马尔可夫理论中的市场占有率的预测模型，对我国 3 家运营商未来市场占有率的变化情况进行了预测。Oyatoye 等[7]通过问卷收集的原始数据，运用马尔可夫链建模，预测实施携号转网后的市场份额变化。在携号转网的研究上，张莅黎等[8]通过具体分析实施携号转网政策下我国电信运营商所受到的影响和面临的挑战，为提高我国电信运营商的综合竞争力提出相应的对策建议。Oloia 等[9]探讨了携号转网的优势及其在电信行业中的应用，研究表明随着服务质量的提高接受率将增加。Kim[10]研究了韩国实施非对称监管对移动电话市场的影响。Tyagi[11]研究用户对携号转网的意愿，结果表明人口因素对转换意愿有更大的影响。Park 等[12]研究了携号转网对韩国移动通信市场的影响，结果表明如果市场结构不对称且具有强大的主导地位，则需要监管机制来促进和落实携号转网以减少其副作用。

结合诸多国内外学者的研究成果[3-7]，本文运用“马尔可夫概率转移矩阵”对我国实施携号转网后的移动用户市场竞争格局及其影响进行分析，调查重点突出时间维度、不同区域、

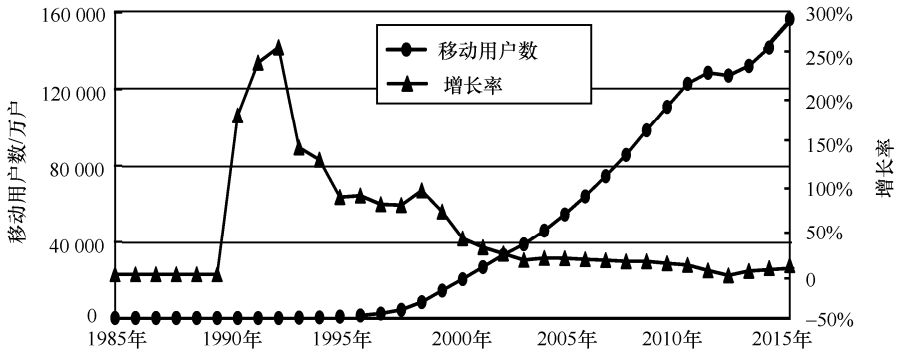


图 1　1985-2018 年我国移动用户数量及增长率

不同人群的携号转网意愿，挖掘不同用户群体转出和转入的原因，剖析影响用户携号转网的重点因素。通过本文的研究，帮助运营企业了解实施携号转网后市场份额的变化及用户在 3 家运营企业的流动情况，并分析不同用户群体的携号转网意愿，找出用户携号转网的主要原因，为通信运营企业提出有针对性应对策略提供参考，更好地把握机遇应对挑战。

2 相关理论及研究方法

2.1 样本设计及获取

2019 年 3 月 27 日至同年 4 月 2 日采用随机抽样的调查方法，通过 CATI 外呼电话调查方式对全国各省随机抽取的样本进行访问，共接触 138 632 个样本，获得 4 962 个成功样本和 815 个有携号转网意愿的样本，样本覆盖全国七大区的三大运营商的移动电话用户。具体见表 1。

2.2 应用理论及方法

马尔可夫链预测法是一种适用于随机过程的科学、有效的动态预测方法，其原理和方法多用于预测企业产品的市场份额。数学中具有马尔可夫性质的离散时间随机过程称为马尔可夫链。随机过程在当前信息已给定的情况下，如果想要预测将来只需知道当前的状态就可以，而过去（即当前以前的历史状态）对于预测将来（即当前以后的未来状态）是无关的[13]。

2.2.1 马尔可夫模型概述

马尔可夫链是由一个条件分布来表示的，$P(X_{n+1}\mid X_n)$ 被称为随机过程中的“马尔可夫概率转移矩阵”。

如果 X_{n+1} 对于过去状态的条件概率分布仅是 X_n 的一个函数，则：

$$P(X_{n+1}=x\mid X_0,X_1,X_2,\cdots,X_n)=P(X_{n+1}=x\mid X_n) \quad (1)$$

其中，X 为过程中的某个状态，式（1）是一个马尔可夫链。马尔可夫链一般具有以下要素：

- 时间集：$T=\{0,1,2,\cdots\}$；
- 状态集：$I=\{1,2,\cdots,m\}$，其中 m 有限；
- 初始状态：初始状态 s_0 既可以为服从某一分布的随机变量，也可以是一个固定值；
- Markov 性（即马氏性）：$P(s_{t+1}\mid s_t,s_{t-1},\cdots,s_0)=P(s_{t+1}\mid s_t)$；

要建立马尔可夫链模型需要一个转移概率矩阵，转移概率矩阵 P_{ij} 定义如下：

$$P_{ij}=\begin{bmatrix}P_{00} & P_{01} & \cdots\\ P_{10} & P_{11} & \cdots\\ \vdots & \vdots & \vdots\end{bmatrix} \quad (2)$$

表 1 携号转网成功样本及有携号转网意愿样本分布

区域运营商	成功样本/个				有携号转网意愿样本/个			
	中国联通	中国电信	中国移动	小计	中国联通	中国电信	中国移动	小计
东北	179	179	270	628	33	37	39	109
华北	273	254	284	811	35	44	43	122
华东	258	164	245	667	34	42	43	119
华南	255	155	241	651	32	47	40	119
华中	250	207	293	750	35	35	38	108
西北	244	199	220	663	33	44	42	119
西南	373	150	269	792	33	43	43	119
总计	1 832	1 308	1 822	4 962	235	292	288	815

其中，以中国电信携号转网为例，P_{00} 可认为是中国电信留在本网的市场份额，P_{01} 为中国电信转往中国联通的市场份额，P_{10} 为中国联通转往中国电信的市场份额，以此类推，这就是随机过程中的概率转移矩阵。

其计算式为：

$$AP_{ij} = B \tag{3}$$

具体计算式为：

$$(a_1 \ \ a_2 \ \ a_3) \times \begin{bmatrix} P_{00} & P_{01} & \cdots \\ P_{10} & P_{11} & \cdots \\ \vdots & \vdots & \vdots \end{bmatrix} = (b_1 \ \ b_2 \ \ b_3) \tag{4}$$

本文中矩阵 A 代表 2018 年年底三大运营商市场份额基数，矩阵 P_{ij} 代表携号转网后的概率转移矩阵，矩阵 B 代表实施携号转网后三大运营商的市场份额。

2.2.2　马尔可夫性检验

应用马尔可夫概率转移矩阵进行预测，随机变量序列必须通过马尔可夫性检验（简称“马氏性检验”）。马氏性检验通常用 x^2 统计量来检验[15]：当统计量 $x^2 = 2\sum_{i=1}^{m}\sum_{j=1}^{m} f_{ij}\left|\ln\frac{P_{ij}}{P_j}\right|$ 服从自由度为 $(m-1)^2$ 的 x^2 分布时，该序列即具有马氏性。其中，f_{ij} 为状态 i 到状态 j 的频数，m 为序列状态数，P_{ij} 为从状态 i 到状态 j 的一步转移概率，P_j 为状态 j 的边际概率。选定置信度 α，查表得 $x_\alpha^2((m-1)^2)$，如果 $x^2 > x_\alpha^2((m-1)^2)$，则可认为 x_n 符合马氏性，否则认为不是马尔可夫链。

2.3　研究前提假设

（1）携号转网政策假设

假设2019年年底前我国实施“双向携号转网”政策。

（2）移动电话市场假设

假设市场份额预测只考虑存量市场，由《中国统计年鉴》可知我国移动用户增长率逐渐减小，日趋饱和，因此，没有考虑 2019 年新增市场份额。

3　携号转网结果分析

3.1　携号转网意愿及动向

3.1.1　携号转网意愿

（1）全国

从全国来看，有携号转网意愿的用户占比较低，仅为 16.4%，较 2010 年增加了 2.4 个百分点，2019 年不考虑携号转网的用户占比为 57.4%，有 26.2%的用户还没有考虑好，见表 2。北方用户携号转网意愿高于南方，但差异不显著，见表 3。

表 2　携号转网意愿分析—全国

年份	是	否	没考虑好
2019 年	16.4%	57.4%	26.2%
2010 年	14.0%	52.0%	34.0%

表 3　携号转网意愿分析—南方/北方

区域	是	否	没考虑好
南方	16.1%	57.2%	26.8%
北方	16.7%	57.5%	25.8%

（2）七大区

从 3 家运营商来看，中国联通用户携号转网意愿较高（22.3%），中国移动次之（15.8%），中国电信携号转网意愿相对较低（12.8%）。从全国七大区域来看，华南区用户携号转网意愿较高，华中区用户携号转网意愿较低。其中，中国电信东北区携号转网意愿较高，中国联通华南区携号转网意愿较高，中国移动西北区携号转网意愿较高。具体见表 4。

3.1.2　携号转网动向

（1）全国

中国电信携号转网流向分析：中国联通流入中国电信的用户占比为 2.2%，中国电信流向中国联通的用户占比为 0.6%，中国联通净流入中国电信的用户占比为 1.6%；中国移动流入中国电信的用户占比为 5.0%，中国电信流向中国移动的用户占比

表 4　携号转网意愿分析—七大区/运营商

区域	总体	中国电信	中国联通	中国移动
东北	17.4%	18.4%	20.7%	14.4%
华北	15.0%	12.8%	17.3%	15.1%
华东	17.8%	13.2%	25.6%	17.6%
华南	18.3%	12.5%	30.3%	16.6%
华中	14.4%	14.0%	16.9%	13.0%
西北	17.9%	13.5%	22.1%	19.1%
西南	15.0%	8.8%	28.7%	16.0%
总计	16.4%	12.8%	22.3%	15.8%

为 1.7%，中国移动净流入中国电信的用户占比为 3.3%。实施携号转网后中国电信净流入用户为 4.9%，中国电信是携号转网的最大受益者，如图 2 所示。

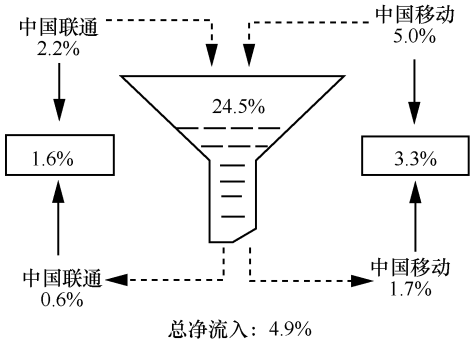


图 2　携号转网流向路径—中国电信

中国联通携号转网流向分析：中国电信流入中国联通的用户占比为 0.6%，中国联通流向中国电信的用户占比为 2.2%，中国电信净流入中国联通的用户占比为−1.6%；中国移动流入中国联通的用户占比为 3.8%，中国联通流向中国移动的用户占比为 2.1%，中国移动净流入中国联通的用户占比为 1.7%。实施携号转网后中国联通净流入用户为 0.1%，携号转网对中国联通影响不显著，如图 3 所示。

中国移动携号转网流向分析：中国电信流入中国移动的用户占比为 1.7%，中国移动流向中国电信的用户占比为 5.0%，中国电信净流入中国移动的用户占比为−3.3%；中国联通流入中国移动的用户占比为 2.1%，中国移动流向中国联通的用户占比为 3.8%，

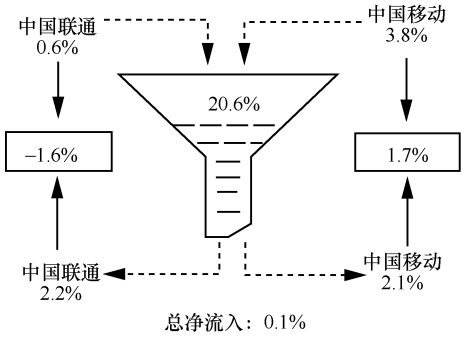


图 3　携号转网流向路径—中国联通

中国联通净流入中国移动的用户占比为−1.7%。实施携号转网后中国移动净流入用户为−5.0%，中国移动是携号转网的最大受损者，如图 4 所示。

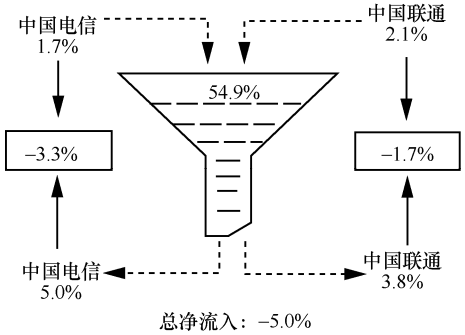


图 4　携号转网流向路径—中国移动

（2）南方/北方

从南方来看，中国电信是最大的受益者，中国联通与中国移动均有不同程度受损，如图 5~图 7 所示；从北方市场来看，中国电信是最大的受益者，中国联通也有所获益，中国移动受损严重，如图 8~图 10 所示。中国电信北方市场净流入要高于南方市场，中国移动北方市场受损较南方市场更严重。

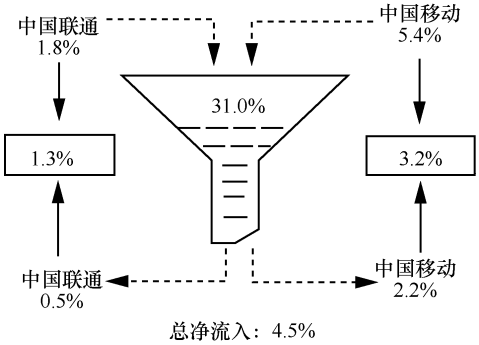


图 5　携号转网流向路径—中国电信（南方）

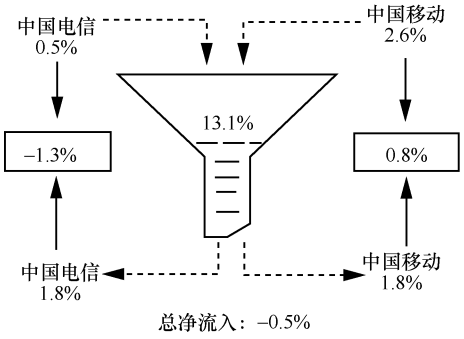


图 6　携号转网流向路径—中国联通（南方）

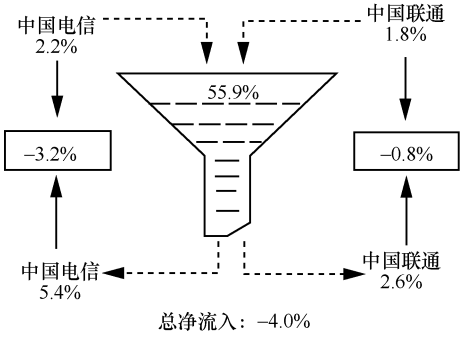


图 7　携号转网流向路径—中国移动（南方）

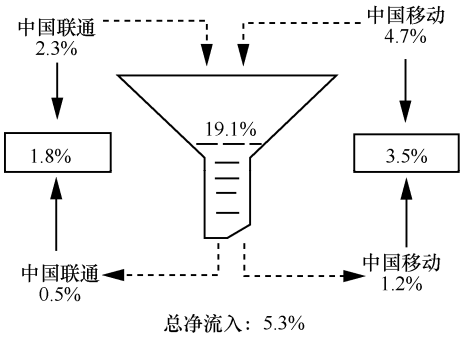


图 8　携号转网流向路径—中国电信（北方）

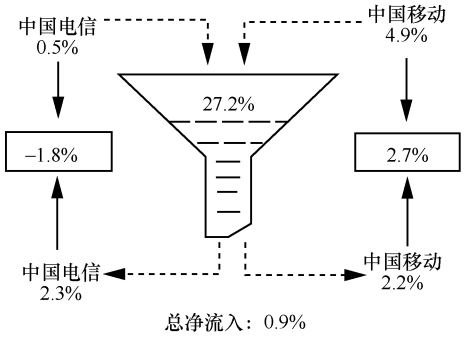


图 9　携号转网流向路径—中国联通（北方）

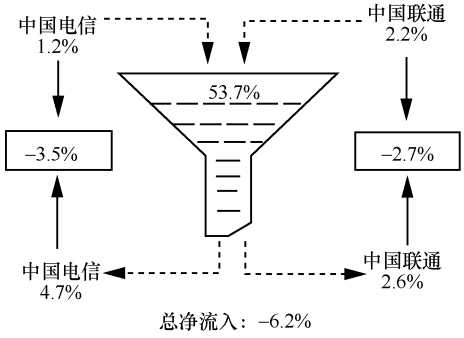


图 10　携号转网流向路径—中国移动（北方）

（3）七大区

中国电信东北区转出占比较高，用户更倾向于选择中国移动；中国联通华南区转出占比较高，用户更倾向于选择中国移动；中国移动西北区转出占比较高，用户更倾向于选择中国电信，见表 5。

3.1.3　市场份额预测

（1）市场份额预测—全国

通过对移动电话用户携号转网调查，得到携

表 5　携号转网流向—七大区

流向	中国电信			中国联通			中国移动		
	中国移动	中国联通	转出	中国移动	中国电信	转出	中国联通	中国电信	转出
东北	11.2%	5.6%	16.8%	8.9%	8.9%	17.9%	6.3%	5.9%	12.2%
华北	8.4%	2.9%	11.4%	9.1%	6.7%	15.7%	5.6%	8.8%	14.4%
华东	8.5%	3.1%	11.6%	11.6%	12.2%	23.8%	8.6%	8.6%	17.1%
华南	9.8%	1.6%	11.4%	15.5%	12.9%	28.4%	6.6%	9.5%	16.2%
华中	9.2%	4.0%	13.2%	8.7%	8.2%	16.9%	6.8%	5.5%	12.3%
西北	9.0%	3.7%	12.7%	7.0%	14.6%	21.6%	4.1%	14.1%	18.2%
西南	5.9%	1.6%	7.5%	12.7%	12.7%	25.3%	5.9%	7.8%	13.8%

号转网矩阵，如式（5），由携号转网矩阵计算携号转网概率转移矩阵，如式（6）。由式（6）可知，中国电信、中国联通和中国移动分别有88.43%、79.28%和85.29%的用户不考虑携号转网（注：没有考虑清楚的用户视为留在原运营商），中国电信用户稳定性较高，中国联通用户稳定性较差。

$$\begin{array}{c} \\ \text{中国电信} \\ \text{中国联通} \\ \text{中国移动} \end{array}\begin{bmatrix} \text{中国电信} & \text{中国联通} & \text{中国移动} \\ 1\,620 & 55 & 157 \\ 138 & 1\,037 & 133 \\ 153 & 115 & 1\,554 \end{bmatrix} \quad (5)$$

$$\begin{array}{c} \\ \text{中国电信} \\ \text{中国联通} \\ \text{中国移动} \end{array}\begin{bmatrix} \text{中国电信} & \text{中国联通} & \text{中国移动} \\ 88.43\% & 3.00\% & 8.57\% \\ 10.55\% & 79.28\% & 10.17\% \\ 8.40\% & 6.31\% & 85.29\% \end{bmatrix} \quad (6)$$

基于2018年年底3家运营商的市场份额和携号转网概率转移矩阵，实施双向携号转网后，中国电信、中国联通、中国移动的用户市场份额分别为24.5%、20.6%和54.9%，如式（7）的预测。中国移动净流出5.0%，中国联通和中国电信净流入分别为0.1%和4.9%。此外，与2010年携号转网调查结果相比中国电信受益增多，中国联通由受损转为受益，中国移动持续受损并增加，但近10年中国移动仅多流出1.6个百分点。具体分析见表6。

$$(19.6\%\quad 20.5\%\quad 59.9\%)\times \begin{bmatrix} 88.43\% & 3.00\% & 8.57\% \\ 10.55\% & 879.28\% & 10.17\% \\ 8.40\% & 6.31\% & 85.29\% \end{bmatrix} = (24.5\%\quad 20.6\%\quad 54.9\%) \quad (7)$$

表6 2019年与2010年实施携号转网净流入对比分析

运营商	2019年净流入	2010年净流入	较2010年变化
中国电信	4.9%	4.6%	0.3%
中国联通	0.1%	−1.1%	1.2%
中国移动	−5.0%	−3.4%	−1.6%

（2）马氏性检验

按第2.2.2节中提及的方法，计算得统计量 $x^2 = 2\sum_{i=1}^{3}\sum_{j=1}^{3} f_{ij}\left|\ln\frac{P_{ij}}{P_j}\right| = 9\,894.81$，令自由度为4，取置信度为 $\alpha = 0.05$，查 x^2 分布表得 $x^2_{0.05}(4) = 9.487\,7$，$x^2 \geqslant x^2_{0.05}(4)$，所以，此过程具有马氏性，可以利用马尔可夫链进行预测。

（3）市场份额预测—南方/北方

从南方市场来看，实施携号转网后，中国电信、中国联通、中国移动的用户市场份额分别为31.0%、13.1%和55.9%，中国电信净流入为4.5%，中国联通和中国移动净流出分别为0.5%和4.0%。从北方市场来看，实施携号转网后，中国电信、中国联通、中国移动的用户市场份额分别为19.1%、27.2%和53.7%，中国电信和中国联通净流入分别为5.3%和0.9%，中国移动净流出为6.2%。具体见表7。

3.2 携号转网影响要素

3.2.1 携号转网原因

从3家运营企业用户携号转网原因的对比分析来看，中国电信用户转入和转出更加关注移动网络质量和宽带网络质量，其次是资费套餐；中国联通用户转出原因主要表现为移动网络质量和

表7 实施携号转网后市场份额对比—南方/北方

区域运营企业	南方			北方		
	实施前	实施后	净流入	实施前	实施后	净流入
中国电信	26.5%	31.0%	4.5%	13.8%	19.1%	5.3%
中国联通	13.6%	13.1%	−0.5%	26.3%	27.2%	0.9%
中国移动	59.9%	55.9%	−4.0%	59.9%	53.7%	−6.2%

宽带网络质量，转入原因主要是资费套餐；中国移动转出主要表现为资费套餐，转入原因主要是移动网络质量，见表 8。具体分析如下。

（1）中国电信用户携号转网原因分析

用户转出原因网络提及率为 66.0%，主要表现为手机网络覆盖、手机上网速度和宽带上网速度，提及率分别为 41.0%、37.3%、15.6%；业务方面主要表现为资费套餐，提及率为 27.4%；服务方面主要表现为营业厅和客服热线服务及效率，提及率分别为 5.2%。中国联通和中国移动用户转入中国电信原因的网络提及率为 63.2%，主要表现为手机网络覆盖、上网速度和宽带上网速度，提及率分别为 36.8%、32.6%和 32.6%，宽带上网速度提及率超过 30%；业务方面表现为资费套餐，提及率为 37.5%。

（2）中国联通用户携号转网原因分析

中国联通用户转出原因的网络提及率为 76.8%，主要表现为手机网络覆盖、手机上网速度、宽带上网速度，提及率分别为 53.5%、36.2%、23.6%；业务和服务因素转出原因占比相对较小。中国电信、中国移动转入中国联通业务方面提及率为 52.4%，主要原因为资费套餐，提及率为 45.9%；网络和服务转入原因占比相对较小。

（3）中国移动用户携号转网原因分析

中国移动转出原因的网络提及率为 69.2%，主要表现为手机网络覆盖、上网速度和宽带上网速度，提及率分别为 26.9%、26.2%、22.7%；业

表 8　用户携号转网原因分析

携号转网原因	中国电信		中国联通		中国移动	
	转出	转入	转出	转入	转出	转入
网络	66.0%	63.2%	76.8%	50.0%	69.2%	59.2%
手机网络覆盖	41.0%	36.8%	53.5%	27.1%	26.9%	57.4%
手机上网速度	37.3%	32.6%	36.2%	28.8%	26.2%	39.7%
宽带上网速度	15.6%	32.6%	23.6%	14.7%	22.7%	15.4%
业务	31.6%	43.6%	29.9%	52.4%	42.3%	37.5%
资费套餐	27.4%	37.5%	24.4%	45.9%	37.1%	21.3%
捆绑政策	4.2%	4.8%	4.1%	6.5%	5.2%	3.7%
全家共享资费/话费，实惠	3.8%	5.2%	3.3%	7.1%	5.9%	2.6%
宣传促销力度	0.9%	1.7%	2.2%	1.2%	1.0%	1.5%
服务	9.0%	5.5%	5.9%	6.5%	5.6%	8.8%
营业厅服务效率	5.2%	2.4%	2.2%	5.3%	3.5%	4.0%
客服热线服务效率	5.2%	2.1%	2.6%	3.5%	2.4%	4.8%
互联网渠道办理业务方便	4.2%	4.1%	4.4%	3.5%	2.4%	3.7%
终端与技术	2.4%	0.3%	0.4%	0.6%	0.7%	1.5%
手机终端种类	0.9%	0.3%	0.4%	0.0%	0.3%	1.1%
5G 技术上网	0.9%	0.0%	0.0%	0.6%	0.0%	0.4%
电视设计及功能	0.9%	0.0%	0.0%	0.0%	0.0%	0.7%

务方面提及率为 42.3%，资费套餐提及率为37.1%，相对较高。中国联通、中国电信转入中国移动的原因网络提及率为 59.2%，主要表现为手机网络覆盖和手机上网速度，提及率分别为 57.4%和 39.7%；业务方面提及率相对较低，仅为 37.5%，其中资费套餐提及率为 21.3%。

3.2.2 携号转网影响

（1）在网时长

用户携号转网意愿随着在网时长增加而增强，在网时长在 3 年以上用户的转出意愿相对较高，见表 9。中国电信重点关注在网时长为 3~4 年、1~2 年和 4~5 年老用户（注：按转出百分比由大到小排序），这部分用户转网更倾向于中国移动；中国联通重点关注在网时长为 3 年以上的老用户，转出用户被中国移动和中国电信均等瓜分；中国移动重点关注在网时长为 3~4 年、5 年以上和 2~3 年的老用户（注：按转出百分比由大到小排序），这部分用户更倾向于转向中国电信。

（2）ARPU 值

3 家运营商高 ARPU 值用户的转网意愿更高，见表 10。中国电信高 ARPU 值用户转出占比相对较低，大部分用户更倾向于转向中国移动；中国联通高 ARPU 值用户转出占比最高，更倾向于转向中国移动；中国移动高 ARPU 值用户转出居中，ARPU 值在 200~300 元用户更倾向于转往中国电信，300 元以上用户更倾向于中国联通。

（3）携号转网人数

携号转网只 1 人携号转网占比为 54.9%，2 人以上一起携号转网超过 40%，4 人以上一起携号转网占比超过 15%。多人一起携号转网中国移动占比较高，中国联通相对较少。结果表明携号转网不仅是一个人基于移动业务的转网，更是一家人或全家人（注：一家人指三口之家，全家人指 4 人及以上的一家人）基于全业务综合考虑家庭转网，见表 11。

表 9 携号转网流向分析—在网时长

转向	中国电信			中国联通			中国移动		
	中国移动	中国联通	转出	中国移动	中国电信	转出	中国联通	中国电信	转出
1 年以内	5.7%	5.7%	11.4%	8.5%	6.4%	14.9%	7.6%	3.3%	10.9%
1（含）~2 年	12.6%	1.9%	14.5%	8.9%	8.9%	17.8%	5.1%	6.4%	11.5%
2（含）~3 年	4.4%	2.6%	7.0%	6.1%	7.6%	13.7%	8.5%	5.1%	13.6%
3（含）~4 年	14.2%	1.8%	16.0%	9.9%	16.0%	25.9%	9.3%	11.4%	20.7%
4（含）~5 年	12.7%	1.7%	14.4%	14.5%	9.9%	24.4%	5.9%	6.3%	12.2%
5（含）年以上	8.6%	3.7%	12.3%	10.9%	11.8%	22.7%	6.4%	10.5%	16.9%

表 10 携号转网流向分析-ARPU 值

转向	中国电信			中国联通			中国移动		
	中国移动	中国联通	转出	中国移动	中国电信	转出	中国联通	中国电信	转出
100 元以内	9.4%	2.9%	12.3%	8.6%	9.2%	17.8%	6.7%	8.6%	15.4%
100（含）~200	9.6%	4.5%	14.0%	10.1%	13.0%	23.1%	6.1%	9.8%	15.9%
200（含）~300	9.8%	3.9%	13.7%	16.3%	12.4%	28.7%	7.3%	10.0%	17.3%
300（含）元以上	10.4%	1.3%	11.7%	16.1%	12.9%	29.0%	9.6%	7.7%	17.3%

表 11　携号转网影响分析—携号转网人数

携号转网人数	总体	中国电信	中国联通	中国移动
1 人	54.9%	55.8%	59.2%	49.8%
2~3 人	28.6%	27.6%	27.2%	30.7%
4~5 人	10.6%	12.6%	9.6%	10.0%
5 人以上	5.90%	4.0%	3.9%	9.5%

4　结束语

本文应用“马尔可夫概率转移矩阵”再次对携号转网后的通信市场竞争格局进行研究和分析，较参考文献[4]创新之处主要表现如下。

一是创新地提出“携号转网流向路径模型”，并利用该模型对通信市场竞争格局进行分析，结果表明中国电信是携号转网的最大受益者，较 2010 年受益增加，不仅转入用户多，且转入用户大部分为中国联通和中国移动的在网老用户和高价值用户；中国联通携号转网前后转入和转出基本持平，较 2010 年受益增加，但高价值用户流出严重；中国移动携号转网后持续受损，且流出持续增加，但短期仍很难扭转中国移动在移动通信市场一家独大的局面。

二是加入“携号转网人数”的研究，结果表明 2 人以上一起携号转网占比超过 40%，5 人以上一起携号转网占比接近 10%，因此携号转网不仅是一个人只考虑移动业务的转网，而是一家人或全家人考虑家庭业务的转网，其影响面更大。

三是携号转入转出用户除考虑移动网络质量和资费套餐外，有超过 30%的用户会考虑宽带网络质量，因此，未来携号转网绝不简简单单表现为移动业务，更多要考虑融合业务[16]，尤其是宽带业务。未来携号转网用户关注的重点是基于家庭用户的全业务信息通信服务综合解决方案的供给能力，不仅要关注网络、业务、服务，还要关注技术、终端和内容，因此运营商要面向未来家庭用户需求，提供全方位、智能化信息通信服务综合解决方案。

参考文献：

[1] 中国移动通信集团有限公司. 2018 年年度业绩公告[R]. 2019.
China Mobile Communications Group Co., Ltd. 2018 annual results announcement[R]. 2019.

[2] 中国联合网络通信股份公司. 2018 年年度报告[R]. 2019.
China United Network Communications Limited. 2018 annual report[R]. 2019.

[3] 中国电信股份有限公司. 截至 2018 年 12 月 31 日止之年度业绩公布[R]. 2019.
China Telecom. Announcement of annual results for the year ended December 31, 2018[R]. 2019.

[4] 胡文玉, 李红霞. 我国实施移动业务携号转网政策的研究[J]. 电信科学, 2010(8): 133-138.
HU W Y, LI H X. Research on China's implementation of mobile service port number transfer policy[J]. Telecommunications Science, 2010(8): 133-138.

[5] 秦书慧. 马尔可夫链模型在电信市场份额中的应用[J]. 西部皮革, 2018(18): 33-34.
QIN S H. Application of Markov chain model in telecom market share[J]. Western leather, 2018(18): 33-34.

[6] 肖会敏, 葛敬云, 贾明泽, 等. 基于马尔科夫链的我国三大运营商市场占有率与预测分析[J]. 数学的实践与认知, 2019(4): 103-107.
XIAO H M, GE J Y, JIA M Z, et al. Market share and forecast analysis of China’s three major operators based on Markov chain[J]. Mathematics Practice and Cognition, 2019(4): 103-107.

[7] OYATOYE E O, ADEBIYI S O , AMOLE B B. Modeling the switching behavior of multiple-SIM GSM subscribers in nigeria using Markov chain analysis[J]. The IUP Journal of Operations Management, 2015, 14(1): 8-31.

[8] 张莅黎, 李美娟. 携号转网政策对我国电信运营商的影响分析[J]. 经济研究导刊, 2016, 299(18): 53-54.
ZHANG L L, LI M J. Analysis of the impact of port number transfer policy on China's telecom operators[J]. Economic Research Guide, 2016, 299(18): 53-54.

[9] OLOIA S V, KUBOYE B M. Interopertability of mobile number portability in South West Nigeria[J]. International Journal of Computer Applications, 2013,81(5).

[10] KIM J Y. The effect of regulations on the Korean mobile telecommunication industry[D]. Claremont: Graduate University, 2006.

[11] TYAGI V. Customer’s awareness and their switching intention towards mobile number portability with special reference to national capital region[J]. International Journal of Management Research and Reviews, 2013, 3(8): 3266-3273.

[12] PARK M C, KIM D D J, LEE S W. Demand for number porta-

bility in the Korean telecommunications market: Contingent valuation approach[J]. Journal of Global Information Management, 2007, 15(1): 43-67.

[13] 马成功. 基于马尔可夫链模型的软件可靠性测试方法的研究[D]. 成都: 电子科技大学, 2009.
MA C G. Research on software reliability test method based on Markov chain model[D]. Chengdu: University of Electronic Science and Technology, 2009.

[15] 金峻炎, 陈进. 加权马氏链在房地产投资回收期预测中的应用[J]. 湖南科技大学学报, 2010(2): 63-66.
JIN J Y, CHEN J. Application of weighted Markov chain in real estate investment recovery period prediction[J]. Journal of Hunan University of Science and Technology, 2010(2): 63-66.

[16] 周一青, 李国杰. 未来移动通信系统中的通信与计算融合[J]. 电信科学, 2018, 34(3): 1-7.
ZHOU Y Q, LI G J. Convergence of communication and computing in future mobile communication systems[J]. Telecommunications Science, 2018, 34(3): 1-7.

[作者简介]

胡文玉（1997− ），男，首都经济贸易大学博士生，北京中智博咨询有限公司运营总监，主要研究方向为信息通信技术（ICT）服务与消费洞察研究、数据分析与建模、空间计量分析。

窦晓燕（1996− ），女，现就职于北京中智博咨询有限公司，主要研究方向为通信企业客户感知服务咨询、数据分析与挖掘。

面向 5G 的核心网演进

马洪源，肖子玉，卜忠贵，赵远
（中国移动通信集团设计院有限公司，北京 100080）

摘　要：5G 致力于“信息随手至，万物触手及”的愿景。5G 网络采用场景化设计，以更加灵活的方式满足多样化的业务需求。基于面向 5G 的核心网目标架构，从网络设备形态、组网架构、网络和业务能力及网络管理编排等方面分析了网络演进中的问题。遵循“网络架构布局一步到位、网元功能平滑升级演进、容量流量逐步迁移调整”的原则，给出了不同时期 5G 演进策略建议。

关键词：5G；核心网；演进；架构
中图分类号：TN929.5
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019159

Evolution of core network oriented to 5G

MA Hongyuan, XIAO Ziyu, BU Zhonggui, ZHAO Yuan
China Mobile Group Design Institute Co.,Ltd., Beijing 100080, China

Abstract: The 5G network is dedicated to the vision of “internet of everythingand information is easy to reach”. 5G network adopts scene design to meet diverse business needs in a more flexible way. Based on the core network target architecture oriented to 5G, the problems in network evolution were analyzed from the aspects of network equipment form, networking architecture, network and service capabilities, and network management orchestration. Following the principle of “one-step network architecture layout, smooth evolution of network element functions, and gradual migration adjustment of capacity traffic”, the 5G evolution strategy recommendations for different periods were given.

Key words: 5G, core network, evolution, architecture

1　5G 网络概述

5G 网络提出了“万物互连”的目标及增强型移动宽带（eMBB）、海量物联网（mMTC）和高可靠低时延（uRLLC）三大应用场景。eMBB 相对于 4G 网络速率可以提供更高的速率、移动性以及频谱效率，可以满足 4K/8K 超高清视频、VR/AR 等大流量应用，为用户提供更好的使用体验。mMTC 和 uRLLC 是针对垂直行业推出的全新场景，分别在流量密度、连接密度和端到端时延、可靠性方面进行了网络设计，用以满足海量物联网连接、车联网、工业控制、智慧工厂等应用，推动 5G 由移动互联网时代向万物互联时代转变。

收稿日期：2019−04−01；修回日期：2019−05−28

2 5G 核心网及目标架构

2.1 5G 核心网特点及问题

为满足不同场景下多样化的业务需求，5G 需要提供灵活的、按需服务的核心网，如图 1 所示。5G 核心网充分借鉴各领域技术优势，打破传统局限，通过架构变革提供全新能力，但同时也给网路部署和实现带来了一些问题和挑战。

- 核心网从 4G 时期的“网元”解耦重构为 5G 时期的“网络功能”。基于统一的通用硬件设施，软件化的功能重构实现网络的低成本、灵活部署；但过多的网络功能与接口增加了核心网部署及异厂商组网的复杂度。
- 基于通用硬件设施，控制面与用户面功能分离，新型的会话管理和移动性管理为网络提供了更多可选的功能组合。但虚拟化转发设备性能与专用设备相比仍有较大改进空间，三层解耦不充分，虚拟化网络管理与编排仍在探索和完善。
- 采用互联网化协议（HTTP/2），提供标准开放的 API。基于面向服务的理念采用服务化架构及服务化接口实现 ICT 深度融合，共享互联网成熟技术；但原有通信协议与新的互联网协议如何实现良好兼容；面向服务的设计理念需要网络在设计态和运维态之间频繁切换，对运维管理又提出新的要求。
- 5G 通过边缘计算和网络切片功能，提供专用逻辑网络，提供按需配置的网络功能，满足基于服务等级协议的个性化需求。但网络切片属于端到端的概念，需要多专业协同和成熟的切片管理；行业用户千差万别，5G 对垂直行业的发展规划和商业模式也需加快制定和探索。

5G 网络多场景设计面临多样化的业务需求，不同的需求对网络能力提出了不同的诉求。在网络转型的关键节点首先应该考虑未来网络的目标架构，以终为始，减少不必要的重复建设与投资浪费。

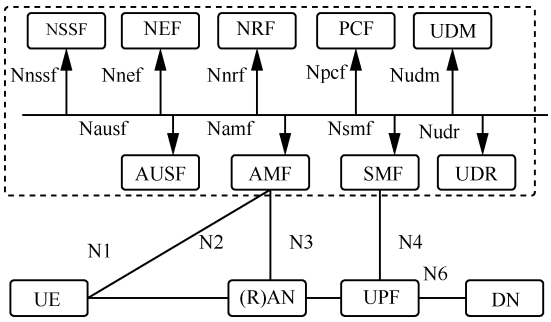


图 1 基于服务的 5G 网络架构

2.2 面向 5G 的核心网目标架构

如图 2 所示，未来的核心网是以 NFV、SDN 技术为基础的资源可全局调度、业务可快速部署、能力可全面开放、容量可弹性伸缩、架构可灵活调整的新一代网络；网络编排是整个网络的管理核心，涉及网络资源的调度编排和网络服务的运维管控。

未来的网络将从目前“以语音和数据为核心”转向“以内容和流量为核心”，核心网网络架构具有以下特点。

- 核心网从扁平的集中化架构调整为“中心+边缘”的分布式架构。传统机房重构为云数据中心，新建或改造边缘基础设施，能够满足新设备形态所需的空间、电源、散热及网络配套需求。
- 传统网元与虚拟化网元长期共存，多制式混合组网，传统网元容量逐步向虚拟化网络迁移；部署面向服务的虚拟化网络功能，按需生成基于 SLA 的网络切片，满足不同场景及用户的业务需求。
- 控制面与转发面进一步分离，转发面网元进一步向边缘下沉，通过 SDN，实现连接可编程及不同数据中心之间的流量集中调度。
- 通过引入新型编排管理系统实现业务、资源和网络的协调管理，通过业务设计、业务部署、配置上线、闭环控制等功能特性，实现 NFV/SDN 网络编排和全生命周期的自动化和智能化管理。

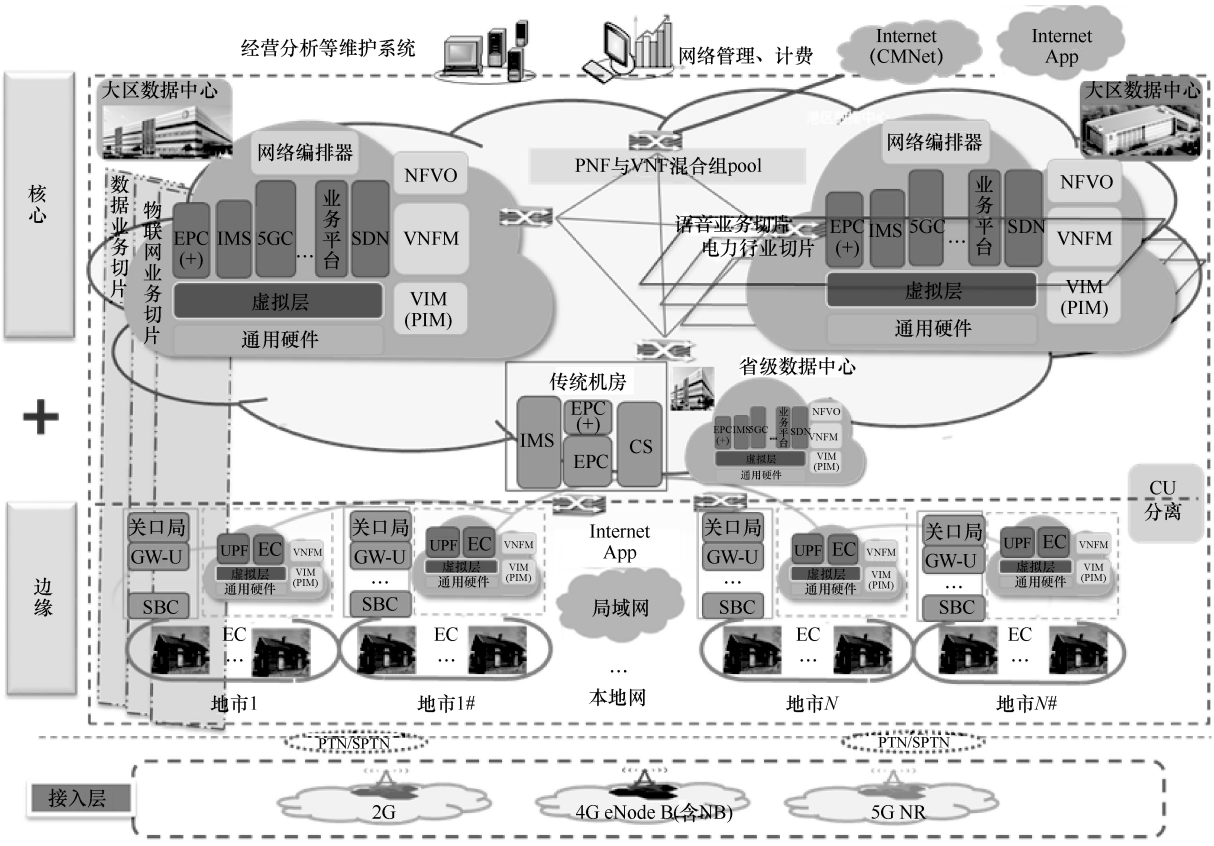


图 2　面向 5G 的核心网目标架构

3　演进中的几个问题

面向 5G 的核心网演进需要考虑设备形态、网络架构、网络和业务能力及网络管理编排等问题。

3.1　设备形态的转型

从目前行业发展来看，网络设备虚拟化、云化是核心网发展的趋势，而解耦后的重构才是网络发展的目标。虽然经历了多轮测试和优化，但虚拟化设备性能和解耦依然存在诸多问题，不同厂商部署相同容量虚拟网元对虚机资源的需求也存在较大差异。

抛开转发面设备不谈，以 500 万容量服务化的 5GC 产品为例，在获得灵活性、可扩展性、开放性的情况下，性能依然弱于传统物理设备，见表 1。

表 1　形态网元性能对比

产品形态	物理网元	虚拟化	服务化（5GC）
性能（资源相同，均采用服务器）	100%	85%	76%
功耗（资源相同，均采用服务器）	100%	80%	83%
占地	100%	120%	120%
新业务上线 TTM	以年为周期	以季度为周期	以周为周期

注：基于现网话务模型，同型号 CPU 配置。

类似地，NFV3 层解耦也需要长期的积累和迭代才能完善。硬件、虚拟层和上层应用充分解耦可以最大程度避免绑定，是目标架构；但目前存在接口标准成熟度低、集成难度大、部署周期长、故障定位复杂和运维管理难度大等问题。软硬件二层解耦技术相对简单、成熟，5G 初期采用基于

三层架构的二层解耦有利于快速部署。

另外，对于核心网部署容器技术目前尚没有特别明确的结论，后续是否引入以及如何引入都有待进一步商榷和研究。

3.2 网络架构的调整

面对多样的业务需求以及更为严苛的网络指标，“中心+边缘”的分布式网络架构既能实现网络控制集中协同，又能灵活地实现转发设备就近接入、本地分流。集中化建设策略和边缘基础设施规划是网络架构调整中的重要课题。

3.2.1 集中化建设思路

针对目前各省机房资源、维护支撑力量不均衡的情况，网络集中化部署有利于集中力量搞攻坚，实现业务快速部署、新技术快速迭代，实现真正意义上 Devops（研发运维一体化），同时提高运维效率，降低运维成本。集中化部署便于全网业务统一、网络切片划分及业务开展，有利于通过集群效应发挥网络规模效益。

集中化建设需要从网络时延、传输带宽、集中运维、集中化组网可靠性以及网络协同和演进等多个角度综合评估。

（1）时延角度

在传输距离增加 1 000 km 的情况下，5G 核心网集中化部署理论上可以满足大部分业务指标需求，用户体验基本不受影响。但集中化部署（基于 N26 的跨系统）造成切换时延增长可能影响切换成功率；交互型业务累积时延对用户体验的影响需要通过测试进一步验证。

（2）带宽角度

大带宽场景 UPF 设备集中部署对传输挑战巨大，以 N3 接口为例，假设大区覆盖 1 亿 5G 用户，按照 DOU 48 Gbit/s/日均 10 个忙时测算，传输网和承载网分别需要 423 个 100GE 端口，建设成本巨大。

（3）集中运维角度

集中运维需要对组织架构和维护分工界面进行适配调整；信令监测、日志留存、安全监控、网络管理等配套建设同步考虑；基于完备的组织架构，专业化的网络运维队伍才能更好地发挥网络集中化的优势。

（4）网络安全及可靠性角度

集中化+虚拟化为大区中心虚拟层/VNF 容灾带来了极大挑战；网元容量更加集中，DC 级故障及信令风暴影响面较大，需要更高级别的安全容灾方案。

（5）网络协同和演进角度

支撑网、信令网需要考虑与大区同步规划，集中化大区与现有基地之间的关系需进一步明确；针对后续 mMTC 及 uRLLC 场景，集中化建设的网元应考虑通过平滑演进满足业务需要。

总而言之，网络集中化建设初期建议多区域共享基础物理资源，不同区域分权分域管理。进行网络资源集中的同时，同步考虑管理系统、支撑系统和维护体系的建设。

3.2.2 边缘基础设施规划

5G 边缘基础设施主要满足大带宽业务本地分流以及边缘计算业务，通常部署 UPF、边缘计算平台、CDN、边缘应用服务器以及网络和管理支撑设备。5G 边缘机房包括地市核心机房、区县重要汇聚机房、普通汇聚机房以及站点接入机房等。

对于新建的地市核心机房，优先考虑按照数据中心标准建设，通过业务、管理、存储三域的物理隔离实现数据中心的分布式拓扑组网，通过 SDN 功能实现网络配置自动化，通过多样化硬件加速技术满足不同场景对网络的能力诉求。

区县及以下的边缘基础设施资源相对匮乏，如果采用云化架构需要考虑轻量级设计，可以采用定制化服务器，组网方面也可以考虑虚拟隔离方案。

边缘基础设施既要支持无人值守情况下的远程维护，也要能够满足异构资源共平台部署；在资源管理方面采用轻量化的 OpenStack 和 SDN，通过集中的网络编排器实现对边缘基础设施的统一管理。

3.3　网络及业务能力的演进

为满足不同运营商网络演进需要，5G 标准提供了两种网络架构：SA（standalone，独立组网）和 NSA（non-standalone，非独立组网）。网络组网及部署架构选项分析见表 2，SA 模式终端单连接，5G 接入网独立组网，通过与 4G 互操作完成系统间切换，5G SA 特指 option2。NSA 则采用 4G/5G 无线混合组网，终端同时与 4G 和 5G 接入网保持连接。

NSA 是向 SA 演进的过渡，option3 系列架构是大部分运营商部署的首选，核心网通过 EPC 升级实现，能够快速实现对 eMBB 场景的支持，而 option7 及 option4 系列是对 LTE 后续演进的一种补充，核心网技术跨度大。SA option2 是 5G 目标网络架构，采用全新的 5G 核心网，能够支持 5G 全部新特性，但标准及产业进展相对滞后。

5G 网络不同的架构需要不同的核心网与之匹配，而架构的选择涉及业务需求及网络定位、频谱资源分配、产业链成熟度、无线改造范围及建设难度、建设周期和投资成本等多个因素。

3.3.1　EPC 核心网升级

支持 option3 系列的 EPC+重用 4G 现有接口和协议，通过功能升级，在 EPC 基础上部署完整的 5G NSA 能力。升级的网络功能根据业务需要分为必选项、条件必选项和条件可选项，PC 核心网功能积极情况汇总见表 3。

基于网络功能需求，EPC 核心网可以确定网元改造范围，如 NSA 业务覆盖区内 MME 需全部升级，SAE-GW 可根据覆盖区内转发流量大小按需确定升级规模。若业务覆盖区 SAE-GW 全部升级，则无需对网关选择功能进行扩展增强；若不需要扩展 QoS、接入控制标识功能（ARD），则

表 2　5G 网络组网及部署架构选项分析

组网模式		NSA			SA
部署架构选项		option3/3a/3x	option7/7a/7x	option4/4a	option2
架构描述		LTE 作为控制面锚点提供连续覆盖，NR 提供容量补充	eLTE 作为控制面锚点提供连续覆盖，NR 提供容量补充	5G NR 作为控制面锚点提供基础覆盖，eLTE 提供容量补充	5G NR 独立组网、提供基础覆盖和容量
网络特性	终端	终端双连接	终端双连接	终端双连接	终端单连接
	无线	4G/5G 混合组网	4G/5G 混合组网	4G/5G 混合组网	5G 独立组网
	核心网	EPC+	5GC	5GC	5GC

表 3　PC 核心网功能升级情况汇总

	MME		SGW	PGW	PCRF	HSS	CG	备注
基本功能	承载迁移	必选	—	—	—	—	—	
	支持双连接	必选	—	—	—	—	—	
	安全	必选	—	—	—	—	—	
扩展 QoS	扩展 QoS	条件可选						（1）HSS 根据 MME NSA 能力下发新签约信息（扩展 QoS 和 ARD）；（2）目前低频最大带宽未超过 4 Gbit/s，不用扩展 QoS 也可支持双连接，当使用高频最大带宽超过 4 Gbit/s 时需要支持
网元选择	网元选择	条件必选			—	—	—	若仅部分 SAE-GW 或 SAE GW-C 升级，则需支持
区分计费	用量上报	条件必选			—	—	—	若需 4G/5G 区分计费，则需支持
	计费	—	条件必选		—	—	条件必选	
移动性管理	ARD 限制接入	条件必选	—	—	—	条件必选	-	若需限制用户对 5G 的接入，则需支持

HSS、PCRF 可不改造。

3.3.2 5GC 部署

支持 option2 系列需要新的 5G 核心网，即 5GC。5GC 既可以基于 vEPC 升级，也可以新建。鉴于 4G/5G 网络长期共存的基本现状，新建 5GC 需支持 EPC 功能，兼容现有 4G 业务，同时满足向 5G 平滑演进。从主要设备厂商产品路标反馈，5GC 控制面建议基于虚拟化部署，用户面可以采用物理设备也可以部署虚拟化设备。

5GC 与 EPC 之间的互操作可以基于标准的 N26 接口。通过 N26 控制面接口传递上下文等互操作信息，可提前在目标网络预留资源，提高切换效率。连接态下，互操作通过执行 PS 切换或重定向完成，PS 切换比重定向中断时延更短，基本无业务中断感知；空闲态下，5GC 通过执行 TAU 切换到 EPC 或通过执行注册流程完成 EPC 向 5GC 的系统间切换。支持 N26 接口需要对与 5G 网络有业务交集区域的 MME 进行改造。

3.3.3 用户数据的融合迁移

运营商基本确立了存量用户不换号，通过更换 5G 终端，接入 5G 网络实现 5G 商用的市场策略。从 4G 到 5G 实现用户不换号需要 HSS/UDM 融合部署，5G 用户共享 4G 鉴权/签约数据，但现网设备不具备向融合 HSS/UDM 演进的条件，因此存量用户数据迁移成为 5G 规模商用的重要前提。而用户是否需要换卡，则涉及是否启用 IMSI 隐私保护功能，即终端接入网络的第一条消息中 SUCI 是否加密。加密所用公钥证书存储在 USIM 卡，该功能为可选功能，取决于运营商。

鉴于用户数据是运营商的核心资产、规模庞大；存量用户数据迁移方案既需要具备通用性、与现网厂商解耦；又要尽量做到用户无感知，对业务支撑系统/信令网/现网 HSS 改造最小，节约投资，简化运维。

基于上述原则，建议采用虚拟化融合 HSS+UDM FE+BE 方案，用户分阶段按需迁移。融合设备满足用户不换号要求，虚拟化基础设施在架构层面一步到位，通过最短路径演进避免重复建设投资。用户迁移过程中新建虚拟化融合设备与传统设备共存，如图 3 所示，方案涉及以下

表 4 5GC 主要 NF 引入策略及部署建议

5GC NF	与 EPC 网元对比	融合网元	融合部署分析	引入部署策略	组网及容灾方式
UDM	HSS	UDM/HSS	统一签约管理（涉及存量）	新建（云化）、vHSS 升级	BE：1+1 FE：*N*+1
PCF	PCRF	PCF/PCRF	统一策略管理（涉及存量）	新建（云化）、vPCRF 升级	pool
SMF	MME 和 GW-C 会话管理功能	SMF/GW-C	统一会话锚点	新建（云化）、vGW-C 升级	pool
UPF	GW-U 用户平面功能	UPF/GW-U	统一用户面锚点	新建（云化或物理）、vGW-U 升级、传统 SAE-GW 升级	pool
AMF	MME 中 NAS 接入控制功能	AMF/MME	建议融合，统一信令接入锚点	新建（云化）、vMME 升级	pool
NEF	SCEF	NEF/SCEF	建议融合，统一能力开放	新建（云化）、SCEF 升级	1+1
NRF	5G 新引入，类似增强 DNS 功能	不涉及		新建（云化）	1+1
NSSF	5G 新引入	独立部署或与 AMF 合设		新建（云化）	1+1 或组 pool

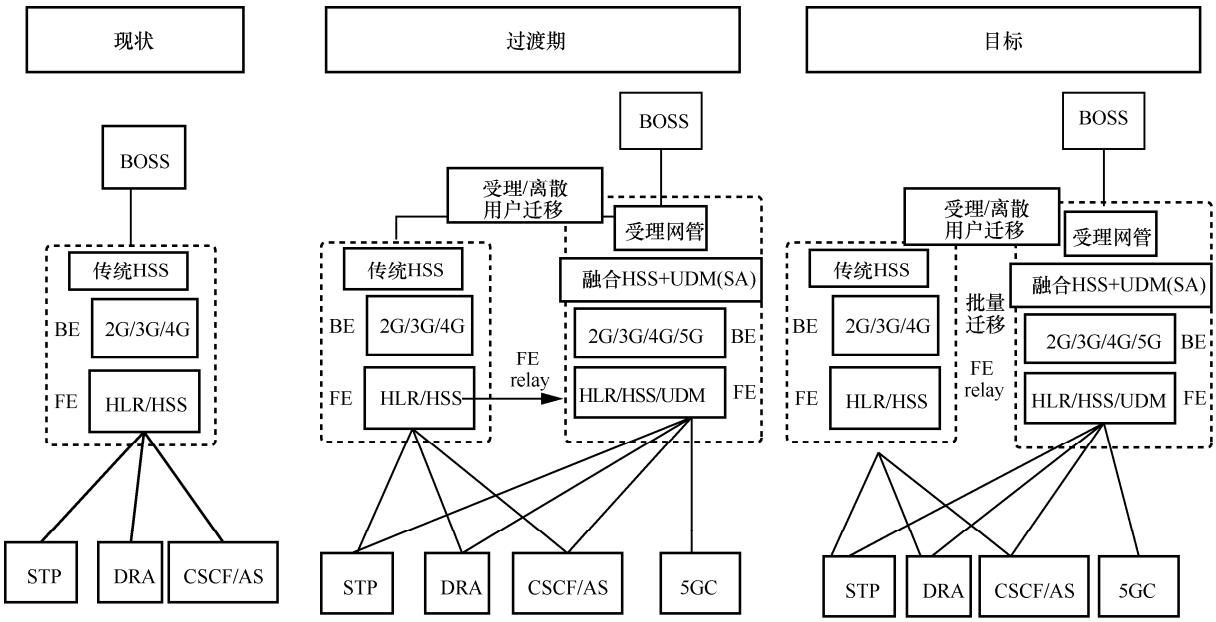


图 3　5G 的用户数据迁移方案

几个关键步骤。

（1）新建虚拟化融合设备需部署受理网关，主要完成离散用户自动迁移和业务支撑系统统一受理开通功能。

（2）5G 初期用户规模较小，采用标准的 FE relay 方案，支持离散用户完成从传统设备向新建融合设备的迁移。为保持信令路由不变，可优先在传统 HSS 支持 FE relay，传统 HSS FE relay 5G 用户到融合 HSS。该方案无需对传统 HSS 设备扩容，但需要评估对信令网的新增负荷。

（3）5G 中后期按号段批量迁移用户，快速完成传统 HSS 用户数据向融合设备的迁移，对于已经迁移过的离散用户无需进行二次数据规整；该方案可在完成用户数据迁移后，再统一修改信令路由。

3.3.4　基础业务的继承

基础业务的继承主要指 5G 语音和短信。3GPP R15 提出了 EPS fallback 和 VoNR 两种主要的语音解决方案，同时针对终端驻留在 eLTE 还提供了 VoeLTE 和 RAT fallback 解决方案。如图 4 所示，5G 网络针对不同场景提供了不同语音解决方案。

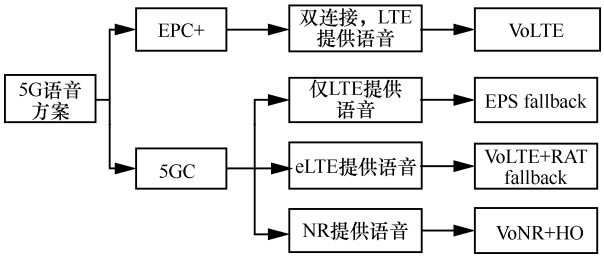


图 4　5G 语音承载方案选择与演进路径

根据网络部署方案和终端能力，5G 短信业务有 3 种可选方案：SMS over SG、SMS over IP 和 SMS over NAS。其中前两种方案在现网已经部署，基本可以满足 5G 初期短信业务需要。方案三主要引入 SMSF 为 5GC 与短信业务系统提供交互，短消息通过 UE 与 AMF 间的 NAS 信令传送。考虑到 mMTC 场景物联网对短信业务的大量需求，建议短信中心与信令网同步升级支持 SMS over NAS。

从无线覆盖范围、终端成熟度、业务需求场景等角度考虑，5G 语音初期优选 VoLTE 和 EPS fallback、逐步向 VoNR 目标方案过渡。初期 NSA 组网场景下，4G 网络覆盖优于 5G，建议 VoLTE 始终通过 4G 无线承载。对于不支持 IMS 协议栈的终端模组按需部署基于 SMSF 的短信解决方案。

3.3.5 5G 信令网的演进

4G 采用 Diameter 信令网旨在进一步简化网络配置，提供负载均衡、流量控制和会话绑定等功能。5G 也面临上述问题，由于 5G 核心网采用了互联网化协议，因此 5G 信令网引入 HTTP proxy。

基于以下考虑，5G 初期可暂不引入 HTTP proxy；待后续方案成熟，再考虑独立演进。

（1）码号寻址可通过 NF 配置或查询 NRF 实现，在初期 NRF 处理能力不受限情况下，可基于 NRF 服务发现实现信令转接。

（2）引入 HTTP proxy 进行链路汇聚可以降低 NF 的信令连接管理复杂度，但也带来一些额外的问题：例如汇聚后信令连接关闭、重建频繁；每次服务调用都需要经过 proxy 转接，引入时延；对 NF 存在路由功能要求等。

（3）5G 会话绑定方案本身与 proxy 功能没有关联，没有引入会话绑定 proxy 的需求。5G 语音方案中 4G 和 5G 共用 IMS 网络，AF 仅支持 Rx 接口接入 DRA，且 DRA 无 5G 用户的会话绑定信息。5G 引入 BSF，通过 DRA 查询实现语音业务会话绑定。BSF 可以独立设置，也可以与 SMF 合设。

（4）集中化建设，大区层面引入 proxy 进行链路汇聚和简化寻址可以降低 NF 的信令连接管理复杂度和 NRF 的处理负荷，但方案存在一定的复杂度且无标准支持。

由于 HTTP proxy 及 BSF 与 DRA 和 STP 存在功能差异且信令协议不同，可暂不考虑采用现有七号信令网和 Diameter 信令网演进支持。为解决 4G/5G 共 IMS 网络的 Rx 接口信令寻址问题，并避免 5G 对七号信令网的扩容需求，建议 DRA 升级支持 5G 服务化接口及短信 Diameter 接口。

3.4 网管编排的提升

通过编排器引入设计域、控制域以及资源管理与编排域，为编排器提供了核心能力。其中，设计域实现对业务、网络等的设计、测试与发布；控制域基于 EMS/OMC、VNFM、控制器等，实现对网络的直接控制；资源管理与编排域实现动态编排、资源管理与配置激活。

NFVO+是下一代网络的编排管理入口和核心，VNFM 是 NFV 网络的核心功能组件，涉及虚拟化网元 VNF 的整个生命周期管理。与网络能力相匹配的网管支撑系统能够充分发挥未来网络的技术优势；通过自研有利于快速满足各类网络管理需求，提升对虚拟化网元、虚拟层、硬件层的管理和支撑能力。

4 结束语

面向 5G 的核心网部署建议采用“网络架构布局一步到位、网元功能平滑升级演进、容量流量逐步迁移调整”的策略。

在 5G 规划期积极实践 5G 技术 4G 化、夯实 NFV 及 SDN 基础，坚持三层解耦，推动虚拟层收敛，扎实推进软硬解耦实现电信网络硬件通用化、标准化和池化。基于中心+边缘的核心网目标架构，提前启动机房规划和改造，提供与分布式核心网相匹配的网络环境。

在 5G 引入期基于运营策略选择不同建设方式，无论是先部署 NSA 逐步向 SA 演进，还是直接部署 SA，抑或是 NSA 和 SA 并举，都应当明确网络定位，避免大规模二次改造。若采用 EPC+，建议控制现网改造规模；同等条件下优选虚拟化建设方式，支持网络升级演进。若部署 5GC，则需考虑简化方案；5GC IoT 接口数量超过 30 个，其中服务化接口超过 20 个，操作/消息组数量超过 200 条，如果按照标准协议，将会对现网部署、尤其是异厂商联调带来巨大工作量。因此建议在真正实现网络自动化部署之前，5G 核心网采用简化部署策略：针对初期商用非必须的网元和功能进行裁剪，对网络功能、接口根据商用需求进行合并，减少虚拟层与应用层厂商配对数量，在不

改变标准协议真实意图的情况下简化部分特性、流程，对于明确的需求、固化部分配置，具体步骤如下。

（1）签约数据接口开放。5GC 控制面作为一个逻辑整体合一部署，仅实现基于 NRF 的签约数据服务化接口异厂商互通。

（2）控制面主要网元间服务化接口开放。AMF、SMF、PCF 之间采用服务化接口，实现异厂商互通；辅助性网元与主网元同厂商合一部署，接口不开放。

（3）全网元服务化接口开放。标准定义采用服务化接口的所有网元全解耦部署，支持端到端切片，支持包括 eMBB、uRLLC、mMTC 等全业务应用场景。

在 5G 发展期，不确定性因素逐步消除，5G 建设规模及投资力度逐步加大。5G 核心网能够充分发挥 5G 网络特性，通过网络切片形式对外提供服务化网络功能，不断推动网络从集中化向自动化及智能化演进。

参考文献：

[1] 3GPP. Technical specification group services and system aspects; system architecture for the 5G system [S]. 2018.

[2] 3GPP. Technical specification group services and system aspects; procedures for the 5G system [S]. 2018.

[3] 3GPP. Technical specification group services and system aspects; service requirements for the 5G system [S]. 2018.

[4] 3GPP. Technical specification group core network and terminals; organization of subscriber data[S]. 2018.

[5] 3GPP. Technical specification group core network and terminals; numbering, addressing and identification[S]. 2018.

[6] 马洪源, 肖子玉, 卜忠贵. 5G 标准及产业进展综述[J]. 电信工程技术与标准化, 2018(3).
MA H Y, XIAO Z Y, BU Z G. 5G standard and industry progress review [J]. Telecom Engineering Technics and Standardization, 2018(3).

[7] 马洪源, 肖子玉, 卜忠贵. 5G 网络语音及短信解决方案[J]. 移动通信, 2018, 42(9): 27-32.
MA H Y, XIAO Z Y, BU Z G. Voice and SMS solutions in 5G network [J]. Mobile Communications, 2018, 42(9): 27-32.

[作者简介]

马洪源（1984− ），男 ，中国移动通信集团设计院有限公司高级咨询设计师，主要研究方向为移动通信核心网。

肖子玉（1969− ），女，中国移动通信集团设计院总工程师、教授级高级工程师、院科学技术委员会委员、中国通信协会高级会员，主要研究方向为通信工程咨询、设计和研究。

卜忠贵（1976− ），男，中国移动通信集团设计院有限公司高级工程师、咨询设计总监，主要研究方向为移动通信核心网。

赵远（1986− ），男，中国移动通信集团设计院咨询设计工程师，主要研究方向为通信工程咨询、设计和研究。

电信运营商物联网发展模式比较分析

刘凯凯，张勋
（中国联合网络通信有限公司研究院，北京 100176）

摘　要：近年来，物联网发展如火如荼。对于国内运营商来说，它对目前的转型发展有着十分重要的意义。在此背景上，运营商如何实现在物联网产业“弯道超车”，商业模式至关重要。基于对国内运营商物联网运营现状的梳理，比较国内外物联网行业的发展策略，总结为 3 种不同的发展策略：单一模式、多环节覆盖模式和全产业模式。电信运营商在客户、渠道等方面具有优势，但受限于多方面因素。建议在构建发展模式时，遵循“循序渐进”的原则，充分根据自己的企业定位、结合国家的政策，注重自身在物联网产业链的定位，关注风险。最终实现较好的发展模式，改善电信行业目前“增量不增收”的现状。

关键词：电信运营商；物联网；发展模式
中图分类号：F626
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019144

A comparative analysis of the development modes of the IoT for telecom operators

LIU Kaikai, ZHANG Xun
Research Institute of China United Network Communication Co., Ltd., Beijing 100176, China

Abstract: In recent years, the development of IoT is in full swing. It is of great significance for the current transformation and development of domestic telecom operators. In this context, the business model is very important for operators to achieve “overtaking” in the IoT industry. Based on the analysis of the operation status of the IoT of operators, the development strategies of the IoT industry at home and abroad were compared, and three different development strategies was summarized. Telecom operators had advantages in customers, channels and other aspects, but limited by many factors. It was suggested that in constructing the development model, they should follow the principle of gradual progress, fully according to their own enterprise positioning, combined with national policies, pay attention to their own positioning in the industry chain of the IoT, pay attention to risks, and ultimately realize it. Better development model improve the current situation of increasing revenue in the telecommunications industry.

Key words: telecom operator, IoT, development mode

收稿日期：2019−02−03；修回日期：2019−05−01

1　引言

技术的变革，也是生产力的变革，物联网就被看作推动下一个经济增长的重要生产力，物联网市场蕴含着巨大的市场，在运营商整体发展趋缓的局面下，不管是国内还是国际电信运营商均将物联网作为业务增长点的重要发展方向。从国内的情况来看，政策驱动性给物联网的发展赋能，物联网+不断在不同行业落地。早在 2009 年，温家宝总理就曾在《让科技引领中国可持续发展》的讲话中提到“着力突破传感网、物联网关键技术，早部署后 IP 时代相关技术研发，使信息网络产业成为推动产业升级、迈向信息社会的发动机”。短短数年，围绕物联网的一条技术密集的产业链已经初具规模，物联网产业链如图 1 所示。产业链上游包括芯片和传感器等终端设备、网络设备、软件与应用的提供商，中游是系统集成到网络提供以及运营服务提供，通过链条传导，导入产业链下游的客户中。国内运营商在产业链扮演的角色也日益重要，但对标国际运营商，国内运营商的外部问题与内部问题逐渐暴露出来，核心技术的薄弱、自身定位的不清晰、发展模式的不成形，需要行业内不断地进行摸索与反思，才能在物联网这个新兴的行业寻求新的增长点，摆脱目前电信业发展的困境。随着互联网企业的壮大与生态化运营，电信行业商业模式、发展策略的定位与明确对未来发展至关重要，利用好自身的先天优势，才能在竞争中抢到位置，在增值服务领域分到更大份额的“蛋糕”。

2　发展历程

1978 年至今，改革开放已经 40 年了，如同国家发生的诸多变化一样，电信业从少数的通信产品，大哥大、BB 机，到如今人人都有的智能手机，产业内技术不断更新，产品不断迭代，从迅速发展到转型升级，从 1999 年拆分重组，中国电信业内

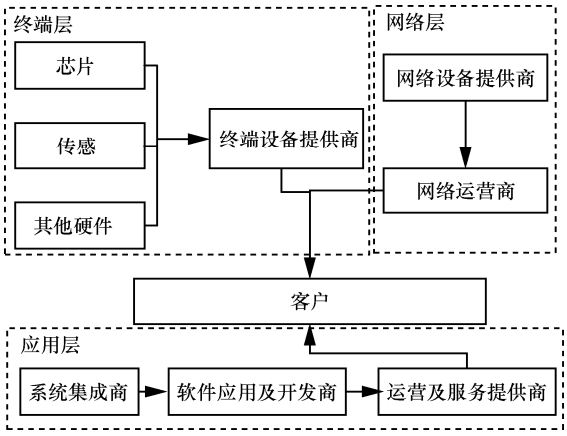


图 1　物联网产业链

的各运营商开始了分业务经营；2001 年，南北拆分中国电信，固话领域垄断被打破；2008 年，重组中国电信业实现中国移动、中国联通、中国电信三足鼎立。行业的变迁离不开大时代的要求，通信技术与计算机技术在行业变迁中起了非常大的作用，行业的发展趋势早已出现端倪。互联网的发展早已被认为不是终点，物联网早在 20 世纪末有了雏形，现如今，从中国信息通信研究院公布的数据来看，物联网的市场规模 2018 年已经预计达到了 1.2 万亿元，根据十三五规划，物联网行业的规模要达到 1.5 万亿元，这将远超过互联网的市场规模。对电信运营商来说，物联网技术的发展必会引领商业模式的变革与更迭，也为传统的电信运营商的服务带来新的增长机会。

2.1　国外运营商的物联网发展情况

物联网的发展开始由需求拉动，一方面来自于行业转型的需求，另一方面来自于消费升级的需求，不同行业与不同领域不断交叉波次推动物联网从刚开始的基础技术研究、基础设施建设推动到现在规模化行业需求的阶段。作为新一代的信息技术，主要面向工业应用，推动全球服务面向数字化、网络化、智能化发展。

工业和信息化部（以下简称工信部）的研究显示，2018 年上半年物联网连接增速达到 72%，从物联网的连接数来看，目前全球的总连接数已

经达到了 9 亿（数据来源：Counterpoint Research）。不管从增速还是总量上来说，电信行业已经将物联网作为了战略性业务，而这背后是各国对此投入巨额资金的结果，也希望能将物联网作为未来的的领先优势，全球移动物联网商用情况见表 1。

国外的运营商普遍以追求行业领先为目标，提供端到端的服务，希望在未来的电信领域竞争中能够抢到一席之地，国外代表性运营商物联网发展情况见表 2。连接市场是运营商发展物联网的首要入口，也是运营商考虑的主要盈利渠道，但

表 1　全球移动物联网网络商用情况（中国信息通信研究院，2018 年）

运营商	国家或地区	网络	运营商	国家或地区	网络
AIS	泰国	LTE-M	T-Mobile	斯洛文尼亚	NB-IoT
AIS	泰国	NB-IoT	T-Mobile	荷兰	NB-IoT
亚太电信	中国	LTE-M	T-Mobile	北美	NB-IoT
亚太电信	中国	NB-IoT	大哥大	中国	NB-IoT
AT&T	北美	LTE-M	Telefonica	西班牙	NB-IoT
AT&T	墨西哥	LTE-M	Telia	芬兰	NB-IoT
			Telia	挪威	NB-IoT
中国移动	中国	NB-IoT	Telia	丹麦	NB-IoT
中国电信	中国	NB-IoT	Telia	瑞典	NB-IoT
中国联通	中国	NB-IoT	Telecom Italia	意大利	NB-IoT
中华电信	中国	NB-IoT	TIM Brazil	巴西	NB-IoT
Dialog Ataxia	斯里兰卡	LTE-M	Telstra	澳大利亚	LTE-M
DNA	芬兰	NB-IoT	Telstra	澳大利亚	NB-IoT
Dialog Ataxia	斯里兰卡	NB-IoT	Telenor	挪威	NB-IoT
Elisa	芬兰	NB-IoT	TRUE Co.	泰国	NB-IoT
Etisalat	阿联酋	LTE-M	Turkcell	土耳其	LTE-M
Etisalat	阿联酋	NB-IoT	Turkcell	土耳其	NB-IoT
远传电信	中国	NB-IoT	Velcom	白俄罗斯	NB-IoT
KDDI	日本	LTE-M	Verizon	北美	LTE-M

表 2　国外代表性运营商物联网发展情况

国家或地区	运营商	主要定位	典型业务	运营支撑情况	运营机构
美国	AT&T	作为 M2M 服务提供商（解决方案主要承包商），提供包括网络连接及支持、计费、客户服务等	物流、智能电网、健康医疗、安全、车联网	网络：新建物联网专用 GGSN、HLR、SMSC；平台：自建平台+与 Jasper 合作	由 EDO 部门（负责消费类电子产品服务）和 AMS 部门（制定移动解决方案）组成，总计 200 人
欧洲	Vodafone	提供终端和网络管理服务，与合作伙伴共同拓展应用	自动抄表、安全监测、零售、设备跟踪、远程医疗、药品管理、车联网	网络：新建物联网专用 GGSN、HLR、SMSC；平台：自建平台	设立 M2M 事业部，垂直管理所有 M2M 资源，约 250 人，在 18 个国家组建下属机构
	Telefonica	作为端到端 M2M 服务提供商和价值链整合者，参与从模块制造到提供服务的工作	智能抄表、嵌入式 UICC 及其远程管理、远程医疗、健康服务、车联网	网络：新建物联网专用 GGSN、HLR、SMSC；平台：自建平台+与 Jasper 合作	物联网业务位于 Telefonica Dig-ital 公司下，负责物联网平台建设，产品支撑，业务推广
日本	NTT DoCoMo	聚焦于提供通信服务，通过定制嵌入模块推动规模发展	工程机械监控、自动售货机监控、智能电表、移动 POS、医疗监测、车联网	网络：新建物联网专用 GGSN、HLR、SMSC；平台：自建平台+与 Jasper 合作	总部设物联部门，100 多人，由 M2M 推进室负责开发物联网品牌，战略规划

这一市场竞争充分，某一家的运营商难以凭借该能力完成差异化的优势，现在的趋势是利用横向连接服务的规模，针对特定垂直领域的特定企业用例共同创建 IoT 解决方案。目前来看，运营商主要通过收购、合作关系和内部创新来支持这些垂直专业化的发展。

2.2　国内运营商物联网发展情况

我国的“十二五”规划中，将物联网的发展列入其中，国家发展和改革委员会、工业和信息化部、国务院都曾纷纷发出指导性文件。在下一个五年计划中，物联网的发展步入了一个新的阶段，跨界融合、集成创新与规模化发展成为该领域的重要特征，“十三五”规划完成的情况见表 3。

我国的三大运营商在基础网络建设方面的优势具有不可替代性，三大运营商也从 2012 年左右开始了布局物联网领域，纷纷搭建平台。在“十三五”规划的上半期仍处于网络建设阶段，下阶段将是应用阶段，从公众网络 M2M 连接数来看，三大运营商中国移动、中国联通、中国电信的连接数目前（截至 2018 年 6 月）已经达到 3.8 亿、0.84 亿、0.74 亿。预计到应用阶段，连接数还会呈暴发式增长。

中国移动 2012 年将自身的物联网基地转变组建成中移物联网有限公司，目前在行业内处于领先地位，深耕通道类、器件类及应用类三大业务，主要在物流、交通、市政方面物联网提供解决方案。

中国电信 2010 年开发 IoT 运营支撑平台，2014 年组建分公司主攻物联网，以“天翼物联”为主打品牌，以市场为导向，推进集约化发展，专注通道类业务，聚焦重点行业，平台具有较强的优势，但受制于自身资源，由省分公司推动，用户增长较慢。

在 4G 时代落后的中国联通，则以“做物联网领跑者”为目标，整体战略进行了业务聚焦，物联网领域聚焦制造业、农业、车联网等重点垂直行业。2010 年，在无锡迅速组建了物联网研究院，积极布局物联网领域，近几年，在自主知识产权方面，具有较强优势。推崇合作协同，与众多产业链上的商家有深度合作，推动技术方面的研究，目前仍是以“管道”服务为主，在生态、旅游、环保等领域的解决方案具有较强优势。

三大运营商在不同领域都有相关的业务展开，但都有自身的差异化优势，国内运营商物联网发展情况见表 4。

3　运营商发展模式分析

3.1　发展模式的内涵

发展模式是企业对自身的定位过程，通过资源整合，流程重构，研究自身的盈利路径，从而进行可持续的发展，目的是为企业创造更大的价值。对于发展模式可以从不同的角度进行诠释。面对物联网的发展，电信运营商只有使自己的洞察力更加敏锐，努力探求物联网的发展模式，才能实现产业链参与各方共赢，共同推动物联网产业的发展，部分运营商发展模式见表 5。

3.2　运营商在产业链中定位

（1）最基础的角色：网络运营商

目前，各行各业的数字化转型离不开物联网

表 3　中国信息通信研究院“十三五”规划完成情况（估计值）

	指标	“十三五”规划期末目标值	截至 2018 年 6 月	完成占比
1	物联网总体产业规模/万亿	1.5	1.2	80%
2	公众网络 M2M 连接数/亿	17	5.4	31.80%
3	特色产业集聚区基地/个	10	5	50%
4	产值超 10 亿元的骨干企/家	200	120	60%
5	制定国家和行业标准/项	200	81	40.50%

表 4　国内运营商物联网发展情况

运营商	主要定位	典型业务	运营支撑情况	运营机构
中国移动	物联网业务服务的支撑者；专用芯片和模组的提供者；物联网专用产品的推动者	通道类：物联网机器卡；器件类：通信模组、无线数传终端等；应用：车务通、行车卫士、亲情通、宜居通、电梯卫士等	网络：新建物联网专用 GGSN、HLR、SMSC；平台：自建平台，包括运营管理平台、业务网关和 PBOSS；码号：10648（1 亿个）	中国移动物联网有限公司（2012 年）
中国电信	物联网能力平台的运营者；物联网重点行业应用的集成者；物联网通信管道的提供者	标准通道产品：智机通；应用及解决方案：车管专家、无线 LED 发布、电梯无忧等标准产品及远程抄表、智能交通、智能环保等解决方案；能力服务：应用测试、发布及能力仓库等	平台：自建平台，包括全国统一运营平台和业务网关；码号：10649（1 亿个）	中国电信物联网分公司（2014 年）
中国联通	物联网管道服务商；物联网平台提供商	智能抄表、车辆、无线 POS、移动媒体、生产及环境监测等重点行业应用	网络：新建专用实验网元，包括 GGSN、HLR，尚未商用；平台：自建实验平台 M2MSP，尚无全国统一的支撑平台；码号：10646（1 亿个）	暂无独立机构

表 5　部分国内外电信运营商发展模式

运营商	运作模式
德国电信	以总集成商的角色在提供管道的同时，从产业链的上中下游提供完整的解决方案
西班牙电信	主要提供网络通道与管理，在业务上与诸多厂商合作，物联网收入占比较小
Verizon	以间接提供通道为主，但 2019 年来一开始通过收购车联网反面的公司，逐渐延伸自身的物联网能力
NTT	“+d”计划，B3B2x 的运作模式，招募众多的合作伙伴，逐渐形成自身的生态系统，挖掘客户的多样化需求

技术的应用，是实现业内信息化的必由之路。各行各业的应用种类多、技术跨度大、需求千差万别，存在自身传统的价值链和行业壁垒。电信运营商相比较其他的对手来说，覆盖面最广的无线通信网络无疑是最大的优势，而网络的价值是巨大的，因为进入物联网产业绕不开的就是通过系统集成商间接提供的网络数据通道。

（2）电信运营商另一种身份是系统集成商

它在这方面有一定的服务积累，身份集成的定位决定了它可以直接参与建设物联网以及应用的系统集成。但行业的千差万别，物联网的业务范围非常广阔，量身定制才能满足行业客户的具体需求，这也对电信运营商在解决方案上的灵活性提出了要求。

（3）电信运营商的另一个角色是运营服务提供商

目前拥有了上亿的用户规模，渠道分布很广，几乎遍布全国，品牌价值与业务经验在数十年内的积累也形成了一个优势壁垒，这些都是其他企业所不具备的，而且也对开展物联网的相关业务有很大的帮助。明确自己的身份，选择合适的合作方，提供最基本的网络连接，同时开展相应的支持，全渠道开展自己的业务。

电信运营商在整个通信产业链中的规模投资较大，产生的附加值却最低，处于“微笑曲线”的低端。这与运营商长期仅扮演“通道”的功能不无关系，尤其是近年来，全球经济下滑，电信行业整体低迷，竞争加剧，又受限于一系列的监管，2018 年的 500 强榜电信企业的整体排名下滑，营收下降，利润率下滑，而互联网企业诸如国外的亚马逊与国内的阿里巴巴则抓住趋势，大力发展“云大物智”等创新业务，业绩遥遥领先。

电信运营商也纷纷开始了自己的转型之路，德国电信不断延伸自己在创新业务的影响力，提供全产业的服务，凭借早期的戴姆勒 IT 部门的技术积累，T-system 专门的解决方案部门已经开始逐渐针对物联网领域提供一体化的解决方案，2013 年 6 月，德国电信就因为提供物联网产业链

下端的车联网解决方案完成了当时世界上最大规模的智能安全交通技术现场测试，此举有助于挽回德国常年因交通的堵塞造成的 170 亿欧元的经济损失。以智能家居为例，德国电信向设备生产商、服务提供商开放 QIVICON 平台，开发不同的盈利模式，收取平台使用费和业务收入分成；向智慧家庭运营商提供集平台、应用、设备、网关、品牌、客服等一体化的解决方案，截至 2018 年第 3 季度，德国电信智能家居 Smart Home 已经有 32.6 万个用户，同比增幅达到了 55%；2017 年德国电信物联网收入同比增幅达到了 4.9%。

AT&T 的物联网业务是 AT&T Communications 的一部分，在企业服务部门（AT&T business）中进行独立核算。在连接领域，发布的 Multi-Network Connect 平台使客户能够将这些服务与其他服务提供商的平台结合在一起，除了连接服务，还包括设备和软件管理、分析和 API 服务。目前已有针对特定垂直领域的不同团队，这些垂直领域包括工业物联网、资产管理和智慧城市等。2017 年已经实现了 3 850 万的连接，在物联网领域有 13 亿~14 亿美元的收入（数据来源：AT&T 官网）。AT&T 在 IoT 价值链上的定位如图 2 所示。

3.3　发展模式的类型

在梳理国内外物联网行业的发展过程后，本文可以总结抽象出电信运营商可能的 3 种物联网发展模式：一是单一模式；二是多环节覆盖模式；三是全产业模式。国内的运营商目前虽然全面发力，目标是发展成为全产业发展模式，但同业竞争的加剧，仍然无法做到击穿全产业，实现业务成熟规模化运营。

（1）单一模式

如图 3 所示，是指电信运营商将自身在物联网领域的业务角色进行集中，主要是网络运营商的角色。在产业发展中，凭借其所拥有的基础网络资源和能力建立核心竞争优势，实现价值创造，从而获取利润。

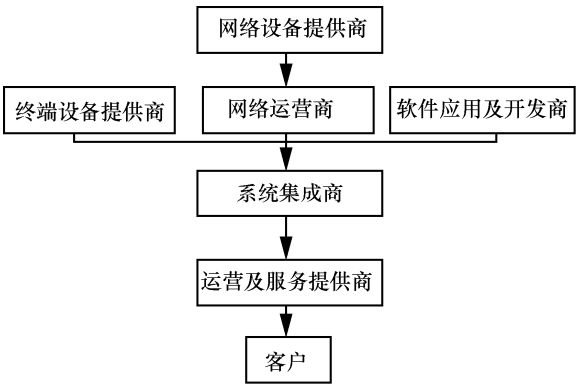
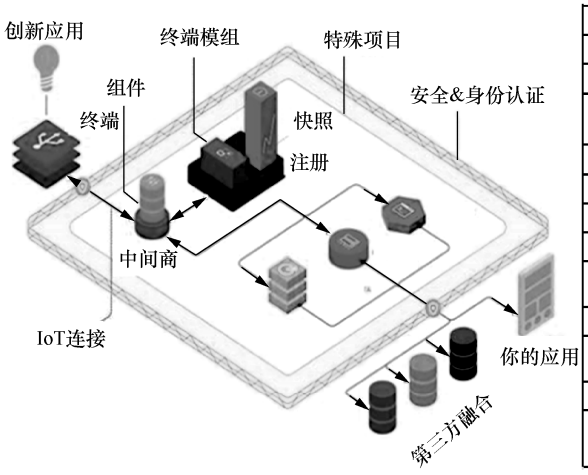


图 3　单一模式

这种模式是电信运营商所能提供的最基础的服务，通过其他物联网服务提供商占用通道消



硬件			
终端		网关/集线器/路由器	
软件			
数据存储	安全	网络运营/客户管理（CRM）	发展
平台			
关系管理	终端管理	应用管理	分析
定制费率计划	自我诊断	M2M/IoT编码库	边缘计算运用
加密	固件更新 (OTA)	预打包M2M/IoT功能	数据查询
压缩	订阅管理OTAP		商业智能
账单分拆			数据备份（云）
QoS和SLA			
连接			
WAN	public mobile	LAN	non-public mobile

图 2　AT&T 在 IoT 价值链上的定位（资料来源：OVUM 数据库，AT&T 官网）

耗的流量的费用而进行的盈利活动，扮演的角色单一，且由于多家运营商提供的服务差异化程度较小，竞争比较充分，利润率相对较低，比如早期的 Verizon。

（2）多环节覆盖模式

多环节覆盖模式相比较单一模式而言是指电信运营商从事的经营业务不仅局限于网络运营，并且参与到其他价值创造环节，比如，同时作为系统集成商存在。多环节覆盖模式如图 4 所示。

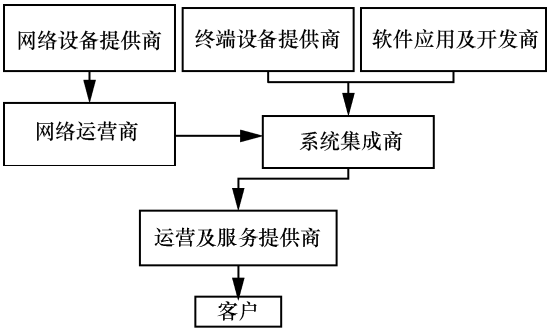


图 4　多环节覆盖模式

这一模式脱离了单纯的管道提供服务，也逐渐提供管道之外的解决方案，但由于技术的积累不够，或者转型的时间需要，没能达到全产业链条的运营能力。相比第一种模式的盈利能力有了进一步的提升，收入结构也能体现出转型的力度，比如法国电信等。目前的电信运营商也基本做到了多环节覆盖，但受限于各种因素，除了基础管道，其他环节的营收能力并没有明显的优势。

（3）全产业模式

全产业模式是指电信运营商在自身资源足够丰富的情况下，集成与服务能力都处于较强的优势，在充分了解行业需求的前提下，网络运营、系统集成、运营服务可以进行全类型提供，能够针对其他行业独立进行解决方案的定制，开发行业套件，实现用户个性化的开发，发挥规模效应，出现规模定制。全产业模式如图 5 所示。

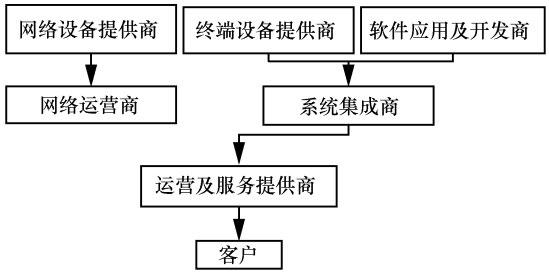


图 5　全产业模式

这一模式一般是电信运营商转型相对较为成功的一种运营模式，参与了全产业链的各个部分，由于垄断性较强，从而获得较强的盈利能力。比如德国电信专门的 T-system 由于原先的技术积累较好，而且启动时间早，在物联网领域的影响力相比其他运营商较强，对特殊领域已经有成熟的一体化解决方案，实现了全环节盈利。

3.4　优缺点比较

对以上的发展模式进行优劣势比较，可以从多维度进行分析，本文从盈利能力、产业价值链中的角色地位、投资风险与管理复杂性的角度进行分析。

（1）盈利能力

市场规则下，开拓业务提供服务的目的就是进行盈利。3 种模式里，营收渠道自然随着在产业链条中的占比而变广。单一模式中，收取流量费是电信运营商的获利渠道，但这种模式下竞争的模式单一，只能凭借价格优势获客，且对于整体的提速降费的影响，简单的流量经营无法给运营商带来明显的盈利点。多环节覆盖模式中，电信运营商有了其他的角色，便可以增加诸如系统集成商系统集成收入运营服务提供商的的用户服务费，可以通过不同的角色盈利，也可以通过不同的策略组合来形成利益最大化，除此之外，因计费、远程维护、业务管理等获得收益。一体化模式中，电信运营商可以获得以上模式中的所有收入。

（2）产业价值链中的角色地位

通过对 3 种模式的比较，电信运营商在物联网产业链条中的角色越来越多元，相应的话语权也会增加。单一模式中，电信运营商客户端为系统

集成商，与客户的连结性较弱，对市场的敏感性较低，增值幅度较低，且话语权不大。在其他两种发展模式中，电信运营商占据主导地位，尤其是在通道业务上，在客户的连结性较强，敏感度提高，在后期的价值链延伸时能够处于有利的主导地位，且由于竞争程度的下降，利润率会有相应的提高。

（3）投资风险

风险与获益与生共存，只要是以获利为基本目的的商业活动必然会存在风险，电信运营商具备无线网络及接入、落地的基础设施，在单一模式下，投资额度比较小，相应的风险小，活力少，一旦延伸自己的物联网产业链，边际效用得到提高，成本的投入带来了快速的增长，相应的风险等级也会上升。

（4）管理复杂度

不同的发展模式有一定的管理要求，而要求的多样性形成了管理的复杂度。对于相对应的 3 种模式来说，物联网在产业价值链扮演的角色越来越多，重要性不断增强，不管是技术管理还是架构管理上，需要全员投入更多，链条的长度与管理的复杂度成正相关，单一模式到一体化模式的管理复杂度依次增强。

4　结束语

目前国内运营商正处于行业转型升级的关键阶段，传统业务在“提速降费”的影响下，增量不增收已然成为一个常态。各大运营商都在展开一系列的创新业务，努力寻找下一个风口。虽然目前 3 家运营商都在行业标准、核心技术方面有了一定的突破，但通过对标国际运营商在物联网方面的战略动作，结合发展模式的分析，本文建议国内运营商以聚焦自身业务，在深耕自己擅长领域形成规模解决方案的同时，逐步击穿产业链上下游，在不同领域形成自身的护城河。

目前的大环境下，物联网行业的发展与政策的导向是密不可分的，国内的物联网公司行业集中度不高，主力都有自己的细分市场。物联网增量的两个浪潮如图 6 所示。

结合国外运营商发展的经验，国内运营商要着力抓手两个行业，比如技术标准相对比较成熟的行业—— 车联网，或者政策倾斜力度比较大的行业，比如涉及公共管理的交通、物流。首先聚焦于一个行业，做成熟的技术标准，成为行业标杆型的解决方案，再逐渐延伸到其他的行业应用上。在构建发展方面，需要注意以下几点。

- 以客户为核心，像民营经济做的相对较好的企业学习，要始终把客户作为核心的关注对象，注重行业里的客户群体的画像特征，只有摸透需求点，才有成为优质标杆的可能性。
- 差异化的产品与服务。对客户来说，获得创新的价值体验能带来更加积极的合作，在产品的生产上，要注入自身的价值主张，提升异质性，使客户形成路径依赖，注重后续持续性的服务。
- 注重强强联合，把握好具体的战略，适当的业务选择更优质，更具潜力的战略合作伙伴，整合核心资源，形成最佳的成本战略，形成一个拳头发力的局面，提升自身的核心竞争力，改善运营商面临的产业转

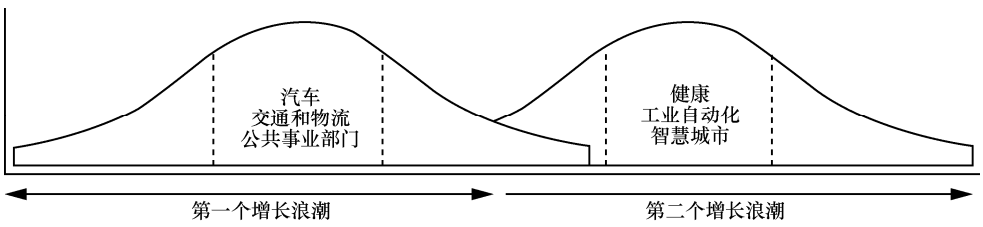


图 6　物联网增量的两个浪潮（资料来源：Gartner）

型压力。

国内的 3 家运营商，中国移动相对具有一定的实力，能够凭借母公司的雄厚资本通过收购合作等方式迅速布局全产业链，但需要注意管理上的复杂度，物联网时代面向的是万物互联，需要整体协同才能发挥作用，庞大的管理系统架构导致的极高的管理成本不适用于物联网需要的敏捷的发展环境；中国电信与中国联通在资本和客户资源上相对较弱，但能有效提高资源使用效率，尤其是中国联通，凭借混改的优势、合作方的技术、客户上的优势等，明确发展定位，完成多环节的覆盖，形成具有标杆性的垂直行业优秀应用，增强自身的能力。

参考文献：

[1] 温家宝. 让科技引领中国可持续发展[EB]. 2009.
WEN J B. Let science and technology lead China's sustainable development[EB]. 2009.

[2] 刘琛琛, 卢云, 庄性华. 物联网发展及运营商物联网应用研究[J]. 现代电信科技, 2016, 46(6): 59-63.
LIU C C, LU Y, ZHUANG X H. Internet of things development and operator internet of things applied research[J]. Modern Telecommunications Technology, 2016, 46(6): 59-63.

[3] 梁渭雄. 电信运营商构建“云管端”助力发展物联网浅谈[J]. 移动通信, 2016, 40(13): 47-52.
LIANG W X. Telecom operators build “cloud tube end” to help develop the internet of things[J]. Mobile Communications, 2016, 40(13): 47-52.

[4] 吕恒. 海外运营商物联网运营经验[J]. 通信企业管理, 2016(5): 70-73.
LV H. Overseas operator's internet of things operating experience[J]. Communications Enterprise Management, 2016 (5): 70-73.

[5] 宁玮. 电信运营商物联网平台型发展模式研究与评价[D]. 长沙: 湖南大学, 2016.
NING W. Research and evaluation on platform-based development mode of internet of things for telecom operators[D]. Changsha: Hunan University, 2016.

[6] 许梅, 梁娜, 王娜. 电信运营商物联网发展需找准发展模式[J]. 世界电信, 2015(10): 50-55.
XU M, LIANG N, WANG N. Telecom operators need to find a correct development model for the development of internet of things[J]. World Telecom, 2015(10): 50-55.

[7] 代成斌, 万功伟. 运营商逐步理清发展模式及策略抢占物联网产业竞争一席之地[J]. 世界电信, 2015(10): 45-49.
DAI C B, WAN G W. Operators gradually clarify the development model and strategy to occupy a competitive position in the internet of things industry[J]. World Telecom, 2015(10): 45-49.

[8] 孙亮, 高寅欣, 孙一平. 电信运营商物联网发展挑战及策略建议[J]. 邮电设计技术, 2015(2): 83-87.
SUN L, GAO Y X, SUN Y P. Challenges and strategic suggestions for the development of internet of things for telecom operators[J]. Posts and Telecommunications Design Technology, 2015(2): 83-87.

[9] 尹丽英, 魏明. 电信运营商物联网发展模式构建——基于奥斯瓦尔德框架[J]. 中国流通经济, 2013, 27(6): 92-96.
YIN L Y, WEI M. Construction of internet of things development model for telecom operators based on oswald framework[J]. China's Circulation Economy, 2013, 27(6): 92-96.

[10] 王凯, 范鹏飞, 黄卫东. 产业链视域下电信运营商发展物联网的发展模式研究[J]. 重庆邮电大学学报(社会科学版), 2013, 25(1): 96-101.
WANG K, FAN P F, HUANG W D. Research on the development model of telecom operators internet of things in the perspective of industry chain[J]. Journal of Chongqing University of Posts and Telecommunications (Social Science Edition), 2013, 25 (1): 96-101.

[11] 马刚, 杨大伟, 葛雯. 浅析电信运营商物联网发展策略[J]. 邮电设计技术, 2012(6): 8-11.
MA G, YANG D W, GE W. Brief analysis on the development strategy of Internet of Things for telecom operators[J]. Posts and Telecommunications Design Technology, 2012(6): 8-11.

[12] 王淑玲, 胡云, 从光磊, 等. 电信运营商物联网平台发展思考[J]. 邮电设计技术, 2017(8): 3, 7-10.
WANG S L, HU Y, CONG G L, et al. Telecom operator internet of things platform development[J]. Posts and Telecommunications Design Technology, 2017 (8): 3, 7-10.

[作者简介]

刘凯凯（1993− ），男，中国联合网络通信有限公司研究院管理研究中心财务管理研究员，主要研究方向为企业战略、财务、成本侧等。

张勋（1990− ），男，中国联合网络通信有限公司研究院大数据研究中心软件开发工程师，主要研究方向为开源技术、容器技术。

5G 网络共享共建方案

马涛
（中国通信建设集团设计院有限公司第四分公司，河南 郑州 450052）

摘　要：5G 是面向未来通信发展需求的移动通信网络，由于工作频率较高，在相同范围内需要建设更多的基站，投资成本巨大。网络共享将是运营商加快 5G 商用进程及“降本增效（降低成本，增加效益）”的有效途径。通过分析 SA/SA、SA/NSA 架构下的网络共享方案，探讨了 5G 网络共享的可能性，对于 5G 网络共享建设给出了系统化理论依据。

关键词：5G；SA；NSA；网络共享
中图分类号：TN929.53
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019182

5G RAN sharing &co-construction scheme

MA Tao
The Fourth Branch, China International Telecommunication Construction Group Design Institute Co., Ltd., Zhengzhou 450052, China

Abstract: 5G is a mobile communication network that meets the development needs of future communication. Due to the use of higher frequencies, more base BS are needed, results in huge CAPEX and OPEX. RAN sharing will be an effective way to speed up the 5G commercial process and reduce costs and increase efficiency. The possibility of 5G RAN sharing by analyzing the scheme under SA/SA and SA/NSA architecture was explored, then a systematic theoretical basis for 5G RAN sharing and co-construction was given.

Key words: 5G, SA, NSA, RAN sharing

1　引言

4G 时代，中国电信和中国联通突破了传统无线网络共享模式，由配套基础设施共建共享转变为网络共享，双方在基站共享、资源管理共享、无线参数策略、传输共享、核心网方案、异厂商插花等方面积极创新，并进行了规模共享建设，加快了 4G 网络建设进程，取得了巨大的经济和社会效益。5G 时代，由于工作频段较高（目前工业和信息化部发布了 5G 中频的试验许可，3 家运营商的频率为 2.6~4.9 GHz），针对中国电信股份有限公司河南分公司进行测算，预估 5G 目标网基站数量需达到现有 4G 基站的 1.2~2 倍，无论投资、基站数量，还是建站难度都远超 4G 网络，同时还要承担后期高额的基站维护费用以及资产折旧。在提速降费持续推进、4G 投资成本还未完全回

收稿日期：2019−03−17；修回日期：2019−06−09

收、5G 盈利模式尚不清晰等因素的影响下，为保障 5G 移动通信网络建设，基础设施以及网络共享或是运营商“降本增效（降低成本，增加效率）”的唯一途径。

2 网络共享共建基本方案

5G R15 标准仅支持 MOCN（multi-operator core network，一个 RAN 可以连接到多个运营商核心网节点)的共享共建方式，即 RAN 侧共享[1-2]。5G R15 小区最大 PLMN 个数扩充到 12 个（LTE 最大支持 6 个），共享的 NR 小区，每个 PLMN（public land mobile communication network）可以有自己的 TAC 和 cell-ID （LTE 协议 R14 版本引入）[3]。独立载波和共享载波各有优劣势：共享载波用户体验更好，但是由于异频组网，网络优化难度增加，尤其频率插花引入异频切换，降低网络性能；独立载波比共享载波更加具备独立性和 QoS（quality of service，服务质量）保障，保持原有同频组网，有利于优化。共建方案的研究，需关注互操作、交界区切换（交界区示意图如图 1 所示）、语音策略和网络规划及优化相关解决方案[4-6]。

3 SA/SA 共享共建方案分析

两家运营商均采用 SA 网络架构下共享共建的方式，可各自拥有核心网，共享 NR，NR 双连接至核心网。

3.1 互操作

共享区域内，互操作流程正常，与非共享场景一样；在共享区与非共享区交界处，非共享站与共享站异厂商时，由于 X2/Xn 接口对接困难，切换都需要走 N2 接口，如图 2、图 3 所示。

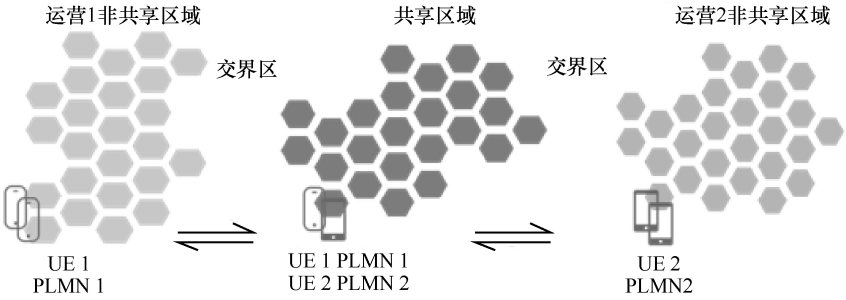
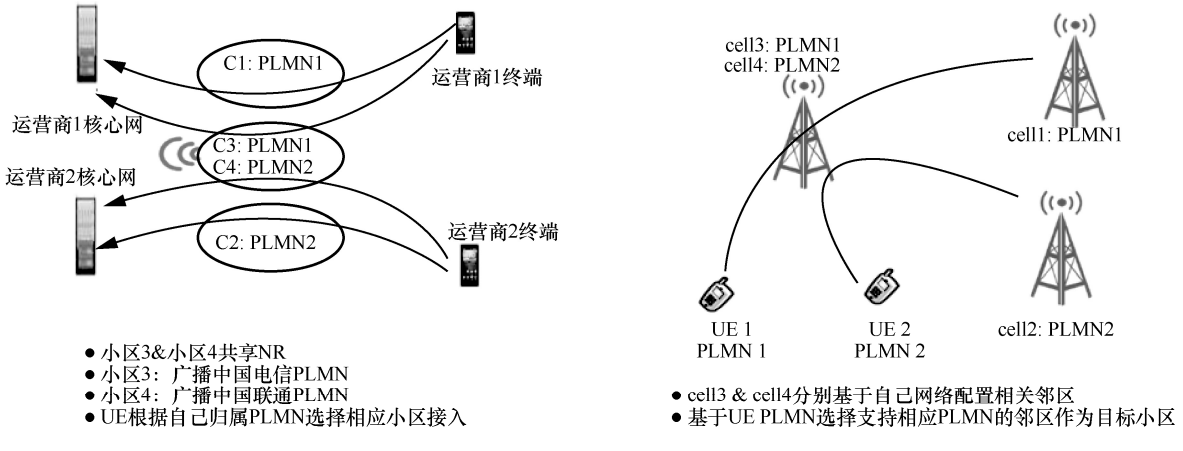
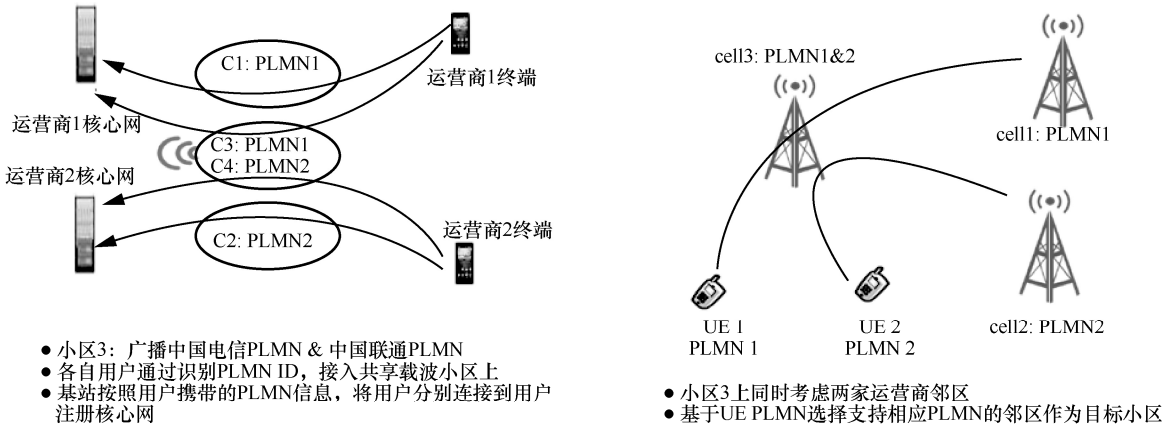


图 1 共享区域、非共享区域示意图



（a）独立载波：广播和注册　　（b）独立载波：切换和互操作

图 2 独立载波方式下互操作示意图



(a) 共享载波：广播和注册　　(b) 共享载波：切换和互操作

图 3　共享载波方式下互操作示意图

3.2　语音策略

建网初期建议双方采用 EPS 回落，由 LTE 网络承载语音；连续覆盖后双方可采用 VoNR。

3.3　网规网优

共享区域和非共享交界区域需要结合网络状况进行网络规划和优化，相对复杂。

3.4　NR 硬件要求

共建共享对硬件要求高，带宽和功率的需求成倍提升。

（1）宏基站 AAU

AAU 200 MHz/400 W 方案，3.5 GHz 频段设备预计 50 kg 以上（将超过铁塔公司 47 kg 要求）；从目前整机效率来看，整机功耗可能会接近 2 000 W，机房供电、备电需要改造，电源拉远也非常困难；AAU 200 MHz/320 W 方案，下行覆盖受限，吞吐量受限。

（2）小站 AAU

共建后设备指标 200 MHz，4×20 W，12 kg/12 L，已经和宏基站设备接近；同时，共享设备很难做到美化伪装需要的 10 kg/10 L 级别。

（3）室内覆盖

DAS 方式，现网的 DAS 无源器件不支持 3.5 Gbit/s 及以上。不能满足单独建设需求，更不能共建；3.5 Gbit/s 以上的馈线损耗比较大，如果和以前覆盖相当，需要提高信源功率或者增加信源，也比较困难。QCell 方式，按照 300 Mbit/s 250 mW 每通道，4T4R 的能力估算，功耗 100 W 左右会达到 POE 极限；光口需要 25 Gbit/s 代替 10 Gbit/s，体积接近 3 L，重量为 3 kg。

3.5　传输要求

共享 NR 的传输可以采用两种方式，共享方借用主建运营商承载和独立传输网。

（1）共享方借用主建运营商承载

• 方案一

借用主建运营商承载，跨运营商的互通在核心层。

应用场景：不依赖共享机房，不限定 CU/DU 合设方式

缺点：运营商 2 的非共享设备与共享设备的 Xn（SA 组网下）需要在核心层打通路径，当前现网不同运营商的设备 IP 地址都是独立分配，直接互通基本不可行，需要通过代理的方式进行互通。运营商 2 的业务需要通过核心层互通，Xn（SA 组网下）通过 2 个运营商的承载网，传输时延大幅增加。

• 方案二

借用主建运营商的部分承载，跨运营商互通在汇聚层。

应用场景：不依赖共享机房，不限定 CU/DU 合设方式，对非主运营商来说，接口时延较之方案一更小。

缺点：当前的承载网不同运营商都是独立建设，在承载网互通，节点多，波及范围大，跨运营商的承载网需要支持互操作。

（2）双方独立传输网络

无线设备同时连接多个运营商的承载，拥有更好的时延和解耦性。

3.6 核心网要求

5G 核心网对无线侧共享的要求包括以下 3 方面。

（1）4G/5G 互操作的实现方式

到 4G 互操作时，提供优选 PLMN 信息，要求 UE 从 4G 回到 5G 时，优先上次注册的 PLMN。AMF 支持多个 PLMN ID，主要是一个为 5G 网络 PLMN，一个是 4G 网络 PLMN。

（2）5G 网络下寻呼的区分方式

网络侧发现的寻呼，两个运营商要区分 5G-S-TMSI 的范围，以免相互冲突。

（3）切片的实现方式

共享 NG-RAN 的 PLMN 定义和支持共享 NG-RAN 的 PLMN 相关切片集。切片定义属于 PLMN 内部，无需特殊处理。切片提供的 QoS 能力以建成网络为基准，如网络部署不能满足 uRLLC 需求，则共享运营商无法提供 uRLLC 能力切片。共享载波情况下，如果资源完全共享，单方引入特殊切片（如 uRLLC），可能需要更多协商。

3.7 网管方案

主建方网管系统提供有限权限给共享方，共享方对共享网络只有查看和有限的操作权限。

（1）共享设置

UME 支持网络共享功能的设置，包括开/关网络共享功能、主/辅运营商设置（关联 PLMN）、辅运营商是否可见公共资源等。在北向接口开通时，通过设置 PLMN 信息，支持网络共享的配置、性能、告警北向接口。

（2）用户管理

UME 支持针对不同运营商，预定义不同的角色及操作集，并赋权（角色和操作对象）给不同的登录用户。

（3）数据权限

主建运营商能够操作和查看全部资源数据，包括：公共资源、共享资源、主运营商专用资源、辅运营商专用资源。共享方运营商能够操作和查看共享资源和共享方运营商专有资源。

4 SA/NSA 共享共建方案

两个运营商共享 NR，NR 一方面需连接到运营商 2 的 LTE eNB，通过运营商 2 的 eNB 对接运营商 2 的核心网，另一方面直接连接到运营商 1 的核心网。网络结构如图 4 所示。

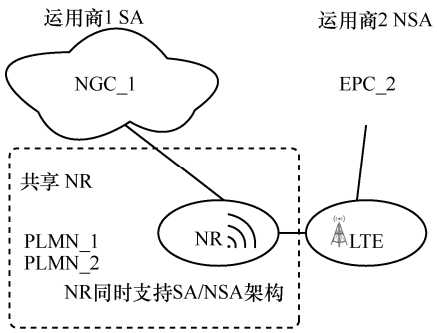


图 4 运营商 1 采用 SA、运营商 2 采用 NSA 结构示意图

运营商 1 采用 SA、运营商 2 采用 NSA，与 SA/SA 方案相比，在互操作、网规网优、核心网要求和网管要求方面均类似，但会面临以下问题。

- NR 硬件要求：需支持 SA/NSA 双架构，支持双方多种语音方案。对于主建方，其 NR 设备需保持与共享方设备相同的厂商，选择灵活性降低。
- 网络规划和优化：需要兼顾两种架构要求：SA/NSA 网络规划目标不同，存在覆盖性能差异，站间距也存在差异。

表 1　共享共建下不同网络架构优劣势对比

对比项	SA/SA	SA/NSA
语音	EPS 回落/ EPS 回落	EPS 回落/VoLTE
基站互操作	交界区异厂商时 N2 口切换	和运营商既有流程一致
异厂商	交界区边缘支持异厂商	共享区设备需和其中一方设备同厂商
网络规划和优化	共享区边缘结合双方需求进行	共享区需要同时考虑 SA/NSA，复杂
网管	统一架构网管	共享区网管需支持两种架构，复杂
终端	共享、非共享区要求一致	共享、非共享区要求一致
其他		（1）基站支持双架构连接，基站成本高； （2）当主建方为 NSA 组网时，与主建方需求冲突：主建方 NSA 组网的诉求在于 NSA 成熟度高，且建网初期投资较低，而共建共享方案建网初期成本较高，且设备成熟度低

- 传输方案：需采用独立传输网方案，共享承载 NSA 的 X2 口时延增大，难以接受。
- 语音方案：建网早期，运营商 1 需采用 EPS 回落，运营商 2 采用 VoLTE；建网中后期运营商 1 可选 VoNR 承载语音。
- 切换：对于运营商 1，需支持 5G 内和 5G 到 4G 间的切换；对于运营商 2，需支持 4G 下 SENB 切换，处理逻辑复杂。

5　结束语

通过表 1 中共享共建下不同网络架构优劣势对比，推荐采用 SA/SA 架构下的共享共建，相比 SA/NSA 下共享共建较为容易实现。5G 网络的共建共享将最大程度地节约成本，提高 5G 建设效率，大大地加快 5G 商用进程，进一步落实政府的“提速降费”政策。

参考文献：

[1] 3GPP. System architecture for the 5G system; stage 2 (release 15): 3GPP 23.501[S]. 2019

[2] 3GPP. NR; radio resource control (RRC) protocol specification (release 15): 3GPP 38.331[S]. 2019.

[3] 赵春华. 5G 承载网的架构演进及带宽分析[J]. 电信科学, 2019, 35(2): 79-83.
ZHAO C H. Architecture evolution and bandwidth analysis of 5G bearer networks[J]. Telecommunications Science, 2019, 35(2): 79-83.

[4] 由宗铭. 5G 无线接入网络架构设计探讨[J]. 网络安全技术与应用, 2019(2): 50-56.
YOU Z M. Discussion on the design of 5G wireless access network architecture[J]. Network Security Technology & Application, 2019(2): 50-56.

[5] 张晓江. 面向 5G 的中国铁塔配套改造分析[J]. 电信技术, 2019(2): 72-74.
ZHANG X J. Analysis of the transformation of Chinese iron towers for 5G[J]. Telecommunications Technology, 2019(2): 72-74.

[6] 张建强, 付道繁. 5G 技术演进对通信基础设施的影响及解决建议[J]. 电信快报, 2019(1): 6-8.
ZHANG J Q, FU D F. Impact of 5G technology evolution on communication infrastructure and suggestions for resolution[J]. Telecommunications Information, 2019(1): 6-8.

[作者简介]

马涛（1981− ），男，中国通信建设集团设计院有限公司第四分公司高级工程师，主要研究方向为移动通信新技术及无线网络规划设计。

230 MHz 电力无线通信技术的优化

赵训威，白杰，丁高泉，相里瑜
（国网信息通信产业集团有限公司，北京 102211）

摘　要：针对 230 MHz 频段电力无线专网连续分配载波的情况，给出了提高频谱效率的优化设计。通过取消载波间的预留保护带，采用连续子载波频域资源映射、自适应 FFT 与滤波处理，减少了运算量，提高了业务速率。针对 230 MHz 电力无线专网两种技术制式标准统一的问题，给出了必要性和重用公网成熟终端与系统设备产业链的初步分析。

关键词：LTE 230；连续子带；频域资源映射；中频滤波
中图分类号：TN915
文献标识码：A
doi: 10.11959/j.issn.1000−0801.2019195

Optimization of electric power wireless communication technology

ZHAO Xunwei, BAI Jie, DING Gaoquan, XIANG Liyu
State Grid Information & Telecommunication Group Co., Ltd., Beijing 102211, China

Abstract: An optimized design for improving spectral efficiency was given for the case where the power wireless private network in the 230 MHz band continuously allocates carriers. By canceling the reserved guard band between carriers, continuous subcarrier frequency domain resource mapping, adaptive FFT and filtering processing were used, which reduced the amount of computation and improves the service rate. Aiming at the problem of unification of two technical standards of 230 MHz power wireless private network, the necessity and the preliminary analysis of the mature terminal and system equipment industry chain of public network were given.

Key words: LTE 230, continuous subband, frequency resource mapping, intermediate frequency filtering

1　引言

230 MHz 是国家无线电管理委员会（以下简称无委）规定供遥测、遥控和数据传输使用的频段，目前主要被气象、电力、水利、地矿等行业使用。230 频段总带宽为 12 MHz，划分为 480 个频点[1]，每个频点对应的载波带宽为 25 kHz，载波又称为子带，其中主要有两段共 7 MHz 合计 278 个载波可供电力无线应用。在这 7 MHz 里，有 43 个零散分布频点已被其他行业占用，故电力行业总共可以使用 235 个离散分布的载波频点[2]，其中，能够连续分配的最大频点数为 41 个。

收稿日期：2019−04−23；修回日期：2019−07−25

2　230 MHz 电力无线通信技术

电力无线通信技术的候选技术包括 4G/5G 公网技术、无线专网技术[3]。230 MHz 频段电力行业分配的频点，其频谱特性较为离散[4]、连续带宽较窄，如图 1 所示。而公网技术设计为适用连续频谱、较大带宽，目前暂不考虑采用。

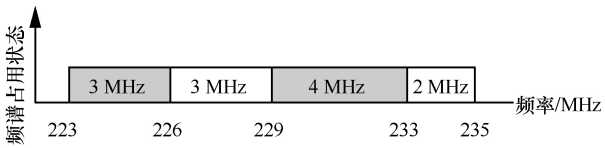


图 1　电力专用频谱分布图

230 MHz 电力无线专网技术[5]目前主要有两种制式：LTE-G 与 IoT-G，两者均基于 LTE/NB-IoT 技术，针对 230 MHz 频段离散频率资源进行设计，主要区别在于空口无线参数设计[6]，包括帧结构、时频资源结构、FFT 点数、子载波间隔及由此衍生出的调度控制机制等。

LTE-G 的一个无线帧长为 25 ms，包括 5 个时长 5 ms 的子帧[7]，分别为 1 个下行子帧、1 个特殊子帧和 3 个上行子帧，每个子帧包含 9 个 OFDM 符号，OFDM 符号采用 64 点 FFT，子载波间隔 2 kHz，每个子带包含 11 个子载波，基带采样速率为 128 ks/s，有效子载波共占用 22 kHz 带宽，每个 25 kHz 载波的两侧各留 1.5 kHz 的保护带，频谱占用率为 88%。

IoT-G 的帧结构与 LTE-G 类似，差别在于 DL 与 UL 子帧各占 2 个，第 3 个子帧是特殊子帧；其无线帧长为 10 ms[8]，包括 5 个时长 2 ms 的子帧，每个子帧包含 6 个 OFDM 符号，OFDM 符号采用 16 点 FFT，子载波间隔 3.75 kHz，每个子带包含 6 个子载波，基带采样速率为 60 ks/s，有效子载波共占用 22.5 kHz 带宽，每个 25 kHz 载波的两侧各留 1.25 kHz 的保护带，频谱占用率为 90%。

LTE-G 目前是按照各个载波单独处理的[9-10]，包括频域资源子载波映射、FIR 滤波以降低对邻带干扰、上采样与中频滤波等[11-12]，LTE-G 的处理流程如图 2 所示。IoT-G 的处理流程大致相同，区别仅在于 FFT 点数为 16，基带采样速率为 60 ks/s，中频采样速率为 15.36 Ms/s，是 256 倍上采样。

3　230 MHz 电力无线专网技术的优化设计

3.1　优化设计原理

每个载波单独处理，都有各自的保护带。但电力通信终端具有多种能力等级，已出现分别支持单子带、4 子带、16 子带、40 子带、280 子带等多种规格。

当多子带能力终端接入网络时，如果基站对

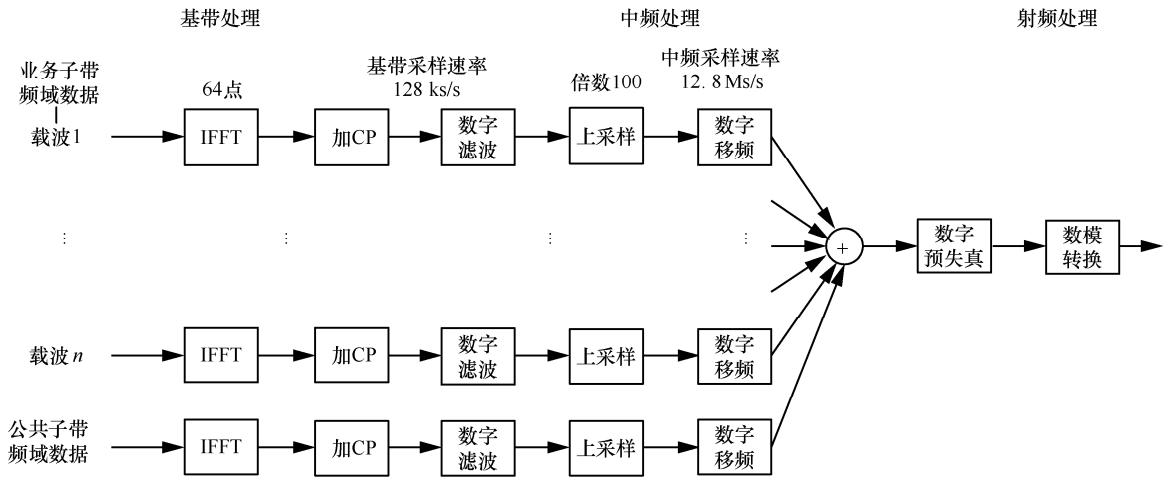


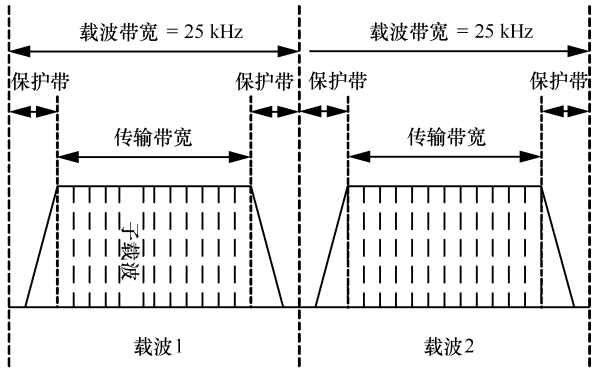
图 2　LTE-G 信号处理过程

终端分配多个连续载波资源，由于多个连续载波都分配给电力行业使用，而载波中间仍然保留保护带，这些保护带不能传输有效数据，会造成这部分频率资源浪费，频谱利用率不高，影响业务速率。

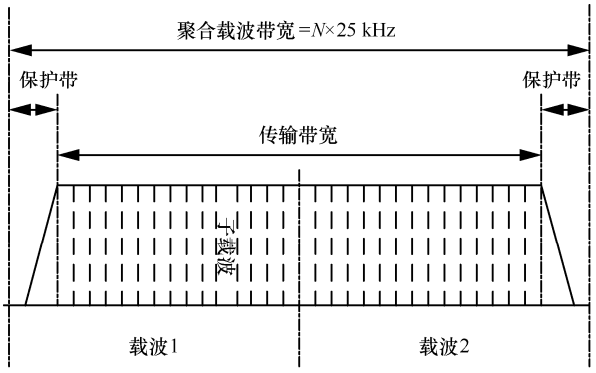
本文设计一种优化基带频域资源映射与中频的方法，在降低运算量复杂度的同时提高频谱利用率，其原理如下。

- 分配给终端连续占用的载波子带，不再单独进行单个载波的基带频域资源映射与滤波处理[13]，而是把连续占用载波组合作为整体带宽资源进行资源映射与滤波处理，仅在整体带宽的两侧预留保护带。
- 自适应 FFT 与滤波处理，针对不同连续带宽分配，对应采用不同 FFT 点数与 FIR 滤波器系数。

以两个连续载波为例，如图 3 所示，原来两个载波中间的保护带宽部分，采用优化设计方案后，可以用于有效的基带频域子载波映射，从而提高了频谱效率；另外，多个连续载波资源映射后，仅需要进行一次 FFT 运算与 FIR 滤波处理，相比现在每个载波都需要单独进行处理，优化方案的运算量大大降低。



(a) 现有载波占用特点



(b) 优化方案载波占用特征

图 3 优化的电力无线频谱资源占用

根据无委的频率规划使用表，电力占用的频率资源，最多可以分配连续的 41 个载波[1]，故存在连续分配 1~41 个载波/子带的情况，需要在此约束下分别设计针对 LTE-G 和 IoT-G 的优化方案。

3.2 LTE-G 优化设计方案

根据优化设计原理重新设计不同连续带宽所对应的参数[14]，考虑到提高频谱占用率，同时减小 FIR 滤波器阶数[13]，按照频谱占用率不超过 96%的原则，分别计算有效子载波个数 N_{sc}、保护带大小、FFT 点数、基带采样速率，并根据这些参数计算频谱利用率和业务速率可提高的百分比。

由此，可计算出 LTE-G 的优化设计参数如下。

- 分配两个连续子带时，子载波个数 N_{sc} 为 23，单侧保护带为 2 kHz，频谱占用率为 92%。
- 分配连续子带个数 N_{sb} 为 3~40 时，子载波个数 $N_{sc} = 12 \times N_{sb}$，单侧保护带为 $0.5 \times N_{sb}$，频谱占用率为 96%。
- 分配连续子带个数 $N_{sb}= 41$ 时，子载波个数 $N_{sc} = 12 \times N_{sb} = 492$，单侧保护带为 20 kHz，频谱占用率为 96%。
- 分配连续子带个数 N_{sb} 为 1~5 时，FFT 点数为 64，基带采样速率为 128 ks/s。
- 分配连续子带个数 N_{sb} 为 6~10 时，FFT 点数为 128，基带采样速率为 256 ks/s。
- 分配连续子带个数 N_{sb} 为 11~20 时，FFT 点数为 256，基带采样速率为 512 ks/s。
- 分配连续子带个数 N_{sb} 为 21~41 时，FFT 点数为 512，基带采样速率为 1 024 ks/s。

表 1 给出了示意几种分配连续带宽下计算的

参数配置，其他带宽配置同样可按照上面 7 条规则计算。

表 1　LTE-G 优化设计参数

连续分配子带个数	带宽/kHz	子载波个数	保护带/kHz	FFT 点数	基带采样速率/(ks·s^{-1})
1	25	11	1.5	64	128
2	50	23	2	64	128
3	75	36	1.5	64	128
6	150	72	3	128	256
11	275	132	5.5	256	512
21	525	252	10.5	512	1 024
40	1 000	480	20	512	1 024
41	1 025	492	20	512	1 024

3.3　IoT-G 优化设计方案

同样地，可以根据优化设计原理重新设计不同连续分配载波带宽对应的参数，考虑到提高频谱占用率[15]，同时减小 FIR 滤波器阶数，按照频谱占用率不超过 96%的原则，分别计算各种带宽配置下的优化参数。

由于 IoT-G 的子载波间隔是 3.75 kHz，不能被载波带宽 15 kHz 整除，故子载波个数 N_{sc} 和保护带大小并不随载波个数 N_{sb} 而线性变化，其优化设计参数计算如下。

- 按照不超过频谱占用率 96%为限制，计算子载波个数 N_{sc} 及对应的保护带。
- 分配连续子带个数 N_{sb} 为 1~2 时，FFT 点数为 16，基带采样速率为 60 ks/s。
- 分配连续子带个数 N_{sb} 为 3~4 时，FFT 点数为 32，基带采样速率为 120 ks/s。
- 分配连续子带个数 N_{sb} 为 5~9 时，FFT 点数为 64，基带采样速率为 240 ks/s。
- 分配连续子带个数 N_{sb} 为 10~19 时，FFT 点数为 128，基带采样速率为 480 ks/s。
- 分配连续子带个数 N_{sb} 为 20~38 时，FFT 点数为 256，基带采样速率为 960 ks/s。
- 分配连续子带个数 N_{sb} 为 39~41 时，FFT 点数为 512，基带采样速率为 1 920 ks/s。

表 2 给出了示意几种分配连续带宽下计算的参数配置，其他带宽配置同样可按照上面 7 条规则计算。

表 2　IoT-G 优化设计方案

连续分配子带个数	带宽/kHz	子载波个数	保护带/kHz	FFT 点数	基带采样速率/(ks·s^{-1})
1	25	6	1.25	16	60
2	50	12	2.5	16	60
3	75	19	1.875	32	120
5	125	32	2.5	64	240
10	250	64	5	128	480
20	500	128	10	256	960
39	975	249	20.625	512	1 920
40	1 000	256	20	512	1 920
41	1 025	262	21.25	512	1 920

3.4　复杂度与性能分析

采用连续子载波进行频域资源映射的优化设计后，不再需要每个子带分别进行 FFT 运算与 FIR 滤波处理，可以降低运算量，同时提高频谱效率。以 LTE-G 分配连续 30 子带为例，数字 FIR 滤波器性能如图 4 所示。

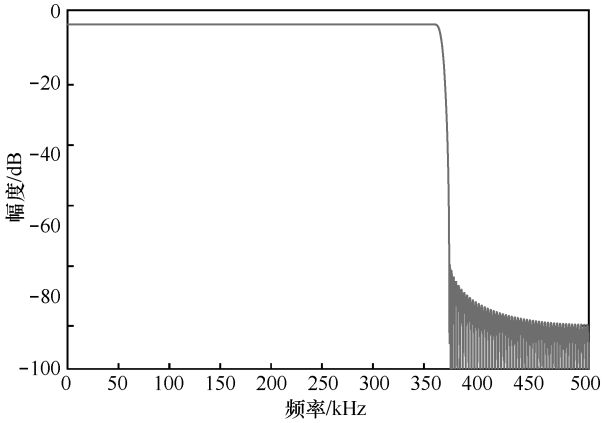


图 4　连续分配 30 子带的 FIR 滤波器性能

可以看出，优化设计增加了过渡带比例，从而降低了滤波器设计难度[16]，数字域可达到对邻带 ACLR 80 dB 的抑制比，加上模拟器件的非线

性底噪抬升，总体可以达到50~60 dB的抑制性能，而无委的指标要求为47 dB[1]；故经过滤波处理后的连续载波信号，同样可以满足无委的带内无用功率指标。

优化设计方案与现有各个载波分别进行中频处理的方案相比较，减少的运算复杂度与提升的性能对比见表3。

表3 复杂度与性能分析

	运算量减小	传输速率提升
LTE-G	77.46%	9.09%
IoT-G	16.18%	6.07%

以LTE-G为例，优化设计的信号处理流程如图5所示。IoT-G的优化处理流程类似。

4 230 MHz电力无线专网技术的标准统一

4.1 标准统一与融合

如前所述，230 MHz目前有两种技术制式：LTE-G与IoT-G，两种制式标准均基于4G/NB-IoT相关技术，结合电网业务特点进行了一些共性针对性设计，如控制信道和数据信道的多帧分配与多次重复以提升可靠性与扩展覆盖范围，简化广播信息和RRC信令以简化协议栈复杂度，降低RF收发指标以降低终端设备实现与成本等；但两种制式仍存在较大差异，如帧结构、载波间隔与时频资源、上下行配比、跳频与频谱感知、免调度传输与同帧调度等，这就造成两种技术制式的产业链力量较为分散，无法集中产业资源进行电力无线专网产业链的培育与促进，同时在实际网络部署时也需要考虑两种技术制式边界的交叉干扰问题，增加了工程实施的难度。

鉴于此，国家电网公司正在推动两种技术制式的标准统一，吸取两种技术制式的优点并进一步优化设计，同时考虑采用一些5G先进技术[15]来完善230 MHz电力无线专网标准，以促进产业链的集中培育与壮大，使无线通信感知层的功能能够满足泛在电力物联网的需求，同时也为电力无线专网标准结合5G的国际化应用做必要的技术铺垫。

4.2 重用4G公网产业链分析

230 MHz电力无线专网与4G/NB-IoT技术相比，最本质的差别在于采用了物理层的离散载波聚合技术，而非公网的MAC层载波聚合技术，此外系统设计基本参数存在差异，由此造成了电力无线专网的各设备网元复用4G/NB-IoT公网产业链的程度存在差异。

整体上，电力无线专网在系统设计时，已尽可能考虑最大化重用现有产业链[17-18]，充分利用

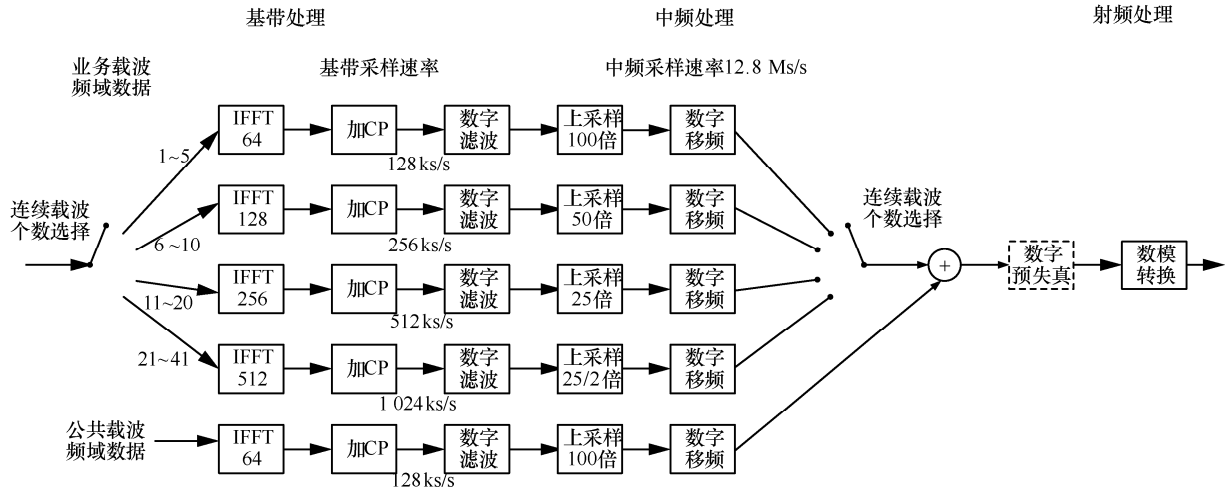


图5 LTE-G优化设计方案的信号处理过程

已成熟的各个环节，降低终端、系统设备研发与网络部署成本。

对于核心网、网管等，其功能与底层无线技术无关，可完全重用公网产业链。

对于基站，硬件部分，射频模拟部分、天线及 RRU 板卡由于 230 MHz 频段与公网不同，不能重用公网部分，需要重新设计板卡，其他如 BBU 板卡、时钟板、主控板等，可完全重用；软件部分，RRU 的 FPGA 数字中频处理[19]、基带 PHY 与 MAC 协议栈的复用程度较低，需较大改动，这些也是工作量最大的部分，RLC/PDCP/RRC/NAS/S1 等改动较小，复用程度较高。整体上在 RRU 适配 230 MHz 频段而重新设计板卡后，基于目前基站能力，仅需要软件升级。

对于终端，硬件部分，如基带芯片[20]、射频芯片、天线等，无法复用公网部分，需要重新设计，其他部分如载板、与电力业务终端的接口板等，可以重用；软件部分，与基站类似，主要修改数字中频、基带 PHY 与 MAC 协议栈，基于终端公网协议栈进行软件功能的修改升级即可。

对于测试仪表，频谱仪、信号源等不涉及协议分析功能，可以重用 R&S、Keysight 等公网常规仪表，涉及协议分析功能的仪表如协议一致性分析仪、综测仪等，需要重新设计硬件来适应 230 MHz 频段，同时进行软件部分的功能修改升级。

5　结束语

本文针对 230 MHz 频段电力无线通信技术，对专网技术方案进行了优化研究。

对于电力无线专网，根据电力行业频点占用的特性，给出了连续占用载波时的优化设计，通过取消连续载波间的预留保护带，可以有效降低运算复杂度、提高频谱效率、提升业务速率；并且适用于目前的两种电力专网技术制式，满足无委的邻带抑制指标要求，有较强的扩展性。

两种 230 MHz 电力无线专网技术制式——LTE-G 与 IoT-G，有必要进行标准统一与融合，进一步优化系统设计，满足泛在电力物联网业务需求，最大化利用现有的成熟产业链，同时适合向 5G 技术演进[21]，待后续国网公司确定统一标准后，可进行进一步的研究。

参考文献：

[1] 工业和信息化部. 工业和信息化部关于调整 223~235 MHz 频段无线数据传输系统频率使用规划的通知[EB]. 2018.
Ministry of Industry and Information Technology. Notice of the Ministry of Industry and Information Technology on adjusting the frequency use plan for wireless data transmission systems in the 223~235 MHz band[EB]. 2018.

[2] 国家电网有限公司. 230 MHz 离散多载波电力无线通信系统 第 1 部分: 总体技术要求: Q/GDW 11806.1[S]. 2018.
State Grid Co., Ltd. 230 MHz discrete multi-carrier power wireless communication system-part 1: general technical requirements: Q/GDW 11806.1[S]. 2018.

[3] WU Z C, JIANG C L, MIAO W W, et al. Research on frequency band selection for lte power wireless private network supporting IMS services for state grid[J]. DEStech Transactions on Materials Science and Engineering, 2017.

[4] 曹津平, 刘建明, 李祥珍. 基于 230 MHz 电力专用频谱的载波聚合技术[J]. 电力系统自动化, 2013, 37(12): 63-68.
CAO J P, LIU J M, LI X Z. Carrier aggregation technology on 230 MHz dedicated spectrum of power systems[J]. Automation of Electric Power Systems, 2013, 37(12): 63-68.

[5] 周建勇, 田志峰, 李艳, 等. 广覆盖 LTE 230 系统在电力配用电应用中的研究与实践[J]. 电信科学, 2014, 30(3): 168-171.
ZHOU J Y, TIAN Z F, LI Y, et al. Research and practice of LTE 230 system with wide coverage characteristics in the power distribution and utilization application[J]. Telecommunications Science, 2014, 30(3): 168-171.

[6] JIANG C L, JIANG S, GUO B, et al. An overview of LTE 230 system in smart grid[C]//International Conference on Information Sciences, Machinery, Materials and Energy (ICISMME), April 11-13, 2015, Chongqing, China. [S.l.:s.n.], 2015: 1002-1005.

[7] 国家电网有限公司. 230 MHz 离散多载波电力无线通信系统 第 2 部分: LTE-G 230 MHz 技术规范: Q/GDW 11806.1[S]. 2018.
State Grid Co., Ltd. 230 MHz discrete multi-carrier power wireless communication system-part 2: LTE-G 230 MHz specification: Q/GDW 11806.1[S]. 2018.

[8] 国家电网有限公司. 230 MHz 离散多载波电力无线通信系统 第 4 部分: IoT-G 230 MHz 技术规范: Q/GDW 11806.1[S]. 2018.

State Grid Co., Ltd. 230 MHz discrete multi-carrier power wireless communication system-part 4: IoT-G 230 MHz specification: Q/GDW 11806.1[S]. 2018.

[9] 周春良, 周芝梅, 王连成, 等. LTE230 数字中频发送机的设计[J]. 电子设计工程, 2018(6): 114-119.
ZHOU C L, ZHOU Z M, WANG L C, et al. Design of a digital IF transmitter for LTE230[J]. Electronic Design Engineering, 2018(6): 114-119.

[10] 周春良, 周芝梅, 王连成, 等. LTE230 数字中频接收机的设计[J]. 电子技术应用, 2017(9): 46-49.
ZHOU C L, ZHOU Z M, WANG L C, et al. Design of a digital IF receiver for LTE230[J]. Application of Electronic Technique, 2017(9): 46-49.

[11] 铁奎, 张慷, 凌云志. 通信系统中数字上变频技术的研究与设计[J]. 电子设计工程, 2012, 20(15): 190-192.
TIE K, ZHANG K, LING Y Z. Design and research of digital up conversion in communication system[J]. Electronic Design Engineering, 2012, 20(15): 190-192.

[12] 田增山, 李路. TD-LTE 多带宽数字下变频设计与 FPGA 实现[J]. 电讯技术, 2016, 5(7): 808-814.
TIAN Z S, LI L. Design and FPGA implementation of multi-bandwidth digital down converter in TD-LTE system[J]. Telecommunication Engineering, 2016, 5(7): 808-814.

[13] 北京智芯微电子科技有限公司. 一种基于 LTE230 的中频信号处理装置和方法: CN 104768190 B[P]. 2018-08-17.
Beijing Smartchip Microelectronics Technology Co., Ltd. A LTE230 based IF signal processing device and method: CN 104768190 B[P]. 2018-08-17.

[14] GAO Y Q, YIN Y, JIA W, et al. A reconfigurable digital intermediate frequency module for software defined radio transmitters[C]//The 12th IEEE International Conference on Solid-State and Integrated Circuit Technology, Oct 28-31, 2014, Guilin, China. Piscataway: IEEE Press, 2014: 1-3.

[15] DAHLMAN E, PARKVALL S, SKOLD J. 5G NR: the next generation wireless access technology[M]. Pittsburgh: Academic Press, 2018.

[16] HAYKIN S. Adaptive filter theory[M]. New York: ACM Press, 2014.

[17]YANG L X. The application of TD-LTE 230M wireless power broadband network technology in smart power-stealing prevention system[J]. Applied Mechanics and Materials, 2015(738-739): 1217-1220.

[18] ZHONG YQ, WANG H, CHEN B R. Research on application principles of communication technologies in smart distribution network[J]. Electric Power Information Technology, 2013, 11(5): 43-47.

[19] ZHANG Y, YUAN X M, QIN J, et al. Research and implementation of the digital intermediate frequency in LTE superheterodyne transmitter[C]//The 8th IEEE International Conference on Communication Software and Networks, June 4-6, 2016, Beijing, China. Piscataway: IEEE Press, 2016: 287-292.

[20] 周春良, 张峰, 程伦, 等. LTE230 无线通信基带芯片的设计与应用[J]. 电子技术应用, 2015(12): 48-50.
ZHOU C L, ZHANG F, CHENG L, et al. Design and application of baseband chip for LTE230 wireless communication[J]. Application of Electronic Technique, 2015(12): 48-50.

[21] HASSEBO A, MOHAMED A A, DORSINVILLE R, et al. 5G-based converged electric power grid and ICT infrastructure[Z]. 2018.

[作者简介]

赵训威（1974− ），男，博士，国网信息通信产业集团有限公司高级工程师，主要研究方向为无线通信技术和物联网技术。

白杰（1981− ），男，国网信息通信产业集团有限公司工程师，主要研究方向为电力无线通信技术。

丁高泉（1984− ），男，国网信息通信产业集团有限公司工程师，主要研究方向为电力无线通信技术。

相里瑜（1978− ），男，国网信息通信产业集团有限公司工程师，主要研究方向为电力无线通信技术。