

田生文 龚军 著

覆盖网络

# 覆盖网络

## 弹性路由与跨层优化

清华大学出版社



# 覆盖网络弹性路由 与跨层优化

田生文 龚 军 著

清华大学出版社  
北 京

## 内 容 简 介

覆盖网络技术是下一代互联网、云计算数据中心网领域研究的热门技术。本书紧紧围绕互联网发展的前沿课题,介绍下一代互联网的热点问题及其解决思路。全书从覆盖网络的产生背景、基本原理及应用出发,通过分析现有互联网存在的问题,以及覆盖网络的不足之处,研究了覆盖网络拓扑构建、多路径负载均衡路由、覆盖网无状态多播,以及覆盖路由与底层物理网络流量工程之间目标不一致而导致的路由冲突问题。

本书是作者在密切跟踪该领域研究成果的基础上深入研究的结果,是一本全面论述覆盖网络弹性路由与跨层优化的著作。全书图文并茂、深入浅出,理论与实践相结合,可读性强。可作为计算机科学与技术、网络工程、物联网、信息工程等相关专业的高年级本科生和研究生的教学参考用书,也可以作为关注新一代互联网理论与实践的研究者的参考用书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

### 图书在版编目(CIP)数据

覆盖网络弹性路由与跨层优化/田生文,龚军著. —北京:清华大学出版社,2017  
ISBN 978-7-302-46469-3

I. ①覆… II. ①田… ②龚… III. ①覆盖网(广播、电视)—研究 IV. ①G221

中国版本图书馆 CIP 数据核字(2017)第 024659 号

责任编辑:白立军

封面设计:常雪影

责任校对:梁 毅

责任印制:杨 艳

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, [c-service@tup.tsinghua.edu.cn](mailto:c-service@tup.tsinghua.edu.cn)

质量反馈:010-62772015, [zhiliang@tup.tsinghua.edu.cn](mailto:zhiliang@tup.tsinghua.edu.cn)

课件下载: <http://www.tup.com.cn>, 010-62795954

印 装 者:三河市中晟雅豪印务有限公司

经 销:全国新华书店

开 本:185mm×230mm

印 张:11

字 数:192千字

版 次:2017年5月第1版

印 次:2017年5月第1次印刷

印 数:1~800

定 价:49.00元

---

产品编号:069378-01

# 前 言

随着互联网的进一步普及,基于网络的新型应用不断涌现,如视频点播、视频会议、远程教育、网络多媒体交互协作平台、在线网络游戏等,这些新型应用要求通信网络提供高可靠性、高带宽、低时延的服务。然而,在互联网规模日益扩大的同时,Internet 本身的僵化现象越来越明显。例如,链路故障恢复时间长以及路由膨胀等问题,严重影响端到端的传输性能;IP 多播的部署难和可扩展性差等问题,限制了日益增长的一对多和多对多数数据传输;缺乏对 QoS 的有效保障,无法满足某些业务的需求。此外,互联网的基础服务提供商(ISP)之间存在复杂的商业利益关系,这使得对现有技术的重大调整变得异常困难。

覆盖网络(Overlay Network)的出现为现有互联网的改造和升级提供了新的思路。覆盖网络是建立在已有网络上的一种逻辑网络,利用隧道或封装机制将覆盖节点(Overlay Nodes)互连起来,形成覆盖网络拓扑来完成数据传输,而无须改变原有的互联网基础设施。覆盖网络除了能完成类似 P2P(Peer-to-Peer)、CDN(Content Delivery Network)这样的内容分发与共享服务之外,随着终端节点性能(CPU 带宽和存储能力)的不断提升,也可以提供路由和组播这些原本只能由路由器完成的基础性服务。另一方面,随着服务器虚拟化和存储虚拟化技术的日臻完善,网络虚拟化技术成为国内外科学界研究的热点。覆盖网络技术作为网络虚拟化的一种有效的解决方案,为下一代互联网、云计算和数据中心网的规划设计提供了一种可行的思路。

虽然覆盖网络在提高路由质量、保障 QoS、提供组播服务等方面能够对现有的互联网基础设施起到很好的补充作用,但在构建覆盖网络时如何感知基础设施结构,达到上下层优势互补,还有许多关键问题亟待解决。例如,如何考虑物理网络中部分关键节点对覆盖网络拓扑构造、路由和数据分发的影响;如何解决两条或多条覆盖链路共享物理链路,造成性能下降的问题;如何建立具有节点邻近意识的覆盖网络,减少端到端的时延;在多播通信中如何降低节点状态的维护代价等。其次,覆盖路由是通过覆盖网络进行的路由模

式,是根据用户的特定需求在应用层上计算路由,是覆盖网络研究的关键技术,受到了国内外学者的广泛关注。在现实中,为了提升服务的性能,服务提供商(SP)在互联网上部署了大量支持各种类型服务的覆盖网络。然而,覆盖路由的本质是自私路由,它根据具体服务的需求进行路由,与物理网络流量工程的路由目标通常并不一致,因此常常会发生冲突;同时,共存的覆盖网络之间也可能会因为竞争网络资源而出现冲突。这种冲突所引起的路由交互问题,严重影响了网络的效率和稳定性。本书针对上述问题,对覆盖网络拓扑构建、多路径路由、应用层多播、上下层路由冲突等覆盖网络关键问题进行了深入细致的研究。在研究过程中,遵循IP网络的运行规律,通过问题抽象、模型构建、算法设计与求解、理论证明、实验验证等一系列严谨的过程,其结论对于下一代互联网的规划与设计具有一定的借鉴意义。

全书共分为8章,其中1~6章由田生文编写,第7章和第8章由龚军编写。在撰写过程中,参考且引用了国内外有关覆盖网络、Internet网络等方面的大量文献,在此,向相关作者表示衷心的感谢。

本书的出版得到国家自然科学基金(No. 61170161)的资助。另外,本书的编写还得到鲁东大学邹海林教授、王刚教授、杨洪勇教授;北京邮电大学廖建新教授、王晶教授、王敬宇老师、戚琦老师的大力支持,在此对他们的支持和帮助表示衷心的感谢。

特别感谢清华大学出版社,感谢责任编辑及其他参与此书编辑和出版的各位老师,为本书的顺利出版付出的辛勤劳动。

由于作者水平有限,书中不足之处在所难免,恳请广大读者和同行批评指正。

作者邮箱:sw\_tian@yeah.net。

编 者

2017年1月

# 目 录

|       |                      |    |
|-------|----------------------|----|
| 第 1 章 | 绪论 .....             | 1  |
| 1.1   | 覆盖网络产生的背景 .....      | 1  |
| 1.1.1 | 互联网的设计缺陷 .....       | 1  |
| 1.1.2 | 下一代互联网设计思路 .....     | 2  |
| 1.2   | 覆盖网络的本质 .....        | 7  |
| 1.2.1 | 覆盖网络体系结构 .....       | 7  |
| 1.2.2 | 覆盖网络的缺陷 .....        | 7  |
| 1.3   | 本书的研究内容及结构安排 .....   | 11 |
| 第 2 章 | 覆盖网络概述 .....         | 12 |
| 2.1   | 覆盖网络基本概念 .....       | 12 |
| 2.2   | 覆盖网络的分类 .....        | 13 |
| 2.3   | 覆盖网络的应用 .....        | 15 |
| 2.3.1 | P2P 网络 .....         | 15 |
| 2.3.2 | 内容分发网 .....          | 16 |
| 2.3.3 | 应用层多播 .....          | 17 |
| 2.3.4 | 增强服务质量(QoS) .....    | 19 |
| 2.3.5 | 云计算数据中心网 .....       | 20 |
| 2.3.6 | 覆盖网络在 SDN 中的应用 ..... | 22 |
| 2.4   | 覆盖网络研究现状 .....       | 23 |

|            |                                 |           |
|------------|---------------------------------|-----------|
| 2.4.1      | 覆盖网拓扑构建 .....                   | 24        |
| 2.4.2      | 覆盖网路由 .....                     | 31        |
| 2.4.3      | 覆盖网多播 .....                     | 34        |
| 2.4.4      | 覆盖网感知方式 .....                   | 41        |
| 2.5        | 本章小结 .....                      | 42        |
| <b>第3章</b> | <b>基于超节点的覆盖网络拓扑构建算法 .....</b>   | <b>43</b> |
| 3.1        | 研究背景 .....                      | 43        |
| 3.2        | 基于超节点的覆盖网拓扑构建 .....             | 46        |
| 3.2.1      | 问题的描述 .....                     | 46        |
| 3.2.2      | 超节点的选取 .....                    | 47        |
| 3.2.3      | $K$ 最小生成树覆盖网拓扑 .....            | 48        |
| 3.3        | 覆盖网一跳路由快速恢复机制 .....             | 51        |
| 3.4        | 算法性能评价 .....                    | 53        |
| 3.4.1      | 仿真实验 .....                      | 53        |
| 3.4.2      | 结果分析 .....                      | 55        |
| 3.5        | 本章小结 .....                      | 62        |
| <b>第4章</b> | <b>负载均衡的覆盖网络多路径路由机制研究 .....</b> | <b>63</b> |
| 4.1        | 研究背景 .....                      | 63        |
| 4.2        | 负载均衡的一跳覆盖网多路径路由 .....           | 64        |
| 4.2.1      | 网络模型 .....                      | 64        |
| 4.2.2      | 线性规划模型 .....                    | 66        |
| 4.3        | 启发式算法 .....                     | 68        |
| 4.3.1      | 中继节点的选择 .....                   | 68        |
| 4.3.2      | 多路径流量切割率 .....                  | 71        |
| 4.4        | 多路径覆盖网络的部署 .....                | 72        |

|            |                                  |           |
|------------|----------------------------------|-----------|
| 4.5        | 性能评价 .....                       | 72        |
| 4.5.1      | 仿真实验 .....                       | 72        |
| 4.5.2      | 结果分析 .....                       | 74        |
| 4.6        | 本章小结 .....                       | 78        |
| <b>第5章</b> | <b>基于包内布隆过滤器的无状态覆盖网络多播 .....</b> | <b>80</b> |
| 5.1        | 研究背景 .....                       | 80        |
| 5.2        | 具有节点邻近意识的覆盖网拓扑 .....             | 82        |
| 5.2.1      | 节点的网络坐标 .....                    | 82        |
| 5.2.2      | Transit-Stub 覆盖网层次拓扑 .....       | 83        |
| 5.3        | 基于包内布隆过滤器的覆盖网多播 .....            | 87        |
| 5.3.1      | 包内布隆过滤器 .....                    | 88        |
| 5.3.2      | 无状态多播机制 .....                    | 89        |
| 5.3.3      | 数据转发中的误判率 .....                  | 91        |
| 5.4        | 性能评价 .....                       | 92        |
| 5.4.1      | 仿真实验 .....                       | 92        |
| 5.4.2      | 结果分析 .....                       | 93        |
| 5.5        | 本章小结 .....                       | 96        |
| <b>第6章</b> | <b>覆盖网络路由冲突问题 .....</b>          | <b>98</b> |
| 6.1        | 研究背景 .....                       | 98        |
| 6.2        | 流量工程概述 .....                     | 98        |
| 6.2.1      | 流量工程概念 .....                     | 98        |
| 6.2.2      | 流量工程分类 .....                     | 99        |
| 6.3        | 路由冲突问题模型 .....                   | 103       |
| 6.3.1      | 冲突问题模型 .....                     | 103       |
| 6.3.2      | 冲突问题研究现状 .....                   | 105       |

|              |                                |            |
|--------------|--------------------------------|------------|
| 6.4          | 覆盖路由之间的交互 .....                | 107        |
| 6.5          | 本章小结 .....                     | 109        |
| <b>第 7 章</b> | <b>多覆盖网络共存环境下的混合交互 .....</b>   | <b>110</b> |
| 7.1          | 研究背景 .....                     | 110        |
| 7.2          | 网络模型和问题描述 .....                | 112        |
| 7.2.1        | 网络模型 .....                     | 112        |
| 7.2.2        | 交互目标 .....                     | 114        |
| 7.3          | $n+1$ 参与者非合作博弈 .....           | 117        |
| 7.3.1        | 纳什均衡 .....                     | 117        |
| 7.3.2        | 静态最优反应算法 .....                 | 118        |
| 7.3.3        | 动态最优反应算法 .....                 | 119        |
| 7.4          | 1 领导者 $n$ 跟随者斯塔克尔伯格-纳什博弈 ..... | 123        |
| 7.4.1        | 斯塔克尔伯格-纳什均衡 .....              | 124        |
| 7.4.2        | 斯塔克尔伯格-纳什博弈求解算法 .....          | 125        |
| 7.5          | 性能评价 .....                     | 126        |
| 7.5.1        | 仿真实验 .....                     | 126        |
| 7.5.2        | 结果分析 .....                     | 128        |
| 7.6          | 本章小结 .....                     | 135        |
| <b>第 8 章</b> | <b>混合交互的联盟合作机制 .....</b>       | <b>136</b> |
| 8.1          | 研究背景 .....                     | 136        |
| 8.2          | 联盟博弈 .....                     | 136        |
| 8.2.1        | 联盟目标函数 .....                   | 136        |
| 8.2.2        | 联盟分配方案 .....                   | 137        |
| 8.3          | 联盟形式 .....                     | 139        |
| 8.4          | 性能评价 .....                     | 141        |

|                  |     |
|------------------|-----|
| 8.4.1 仿真环境 ..... | 141 |
| 8.4.2 仿真结果 ..... | 141 |
| 8.5 本章小结 .....   | 146 |
| 参考文献 .....       | 147 |

# 第1章 绪论

## 1.1 覆盖网络产生的背景

### 1.1.1 互联网的设计缺陷

互联网的普及和发展改变了人们的生活和生产方式。随着宽带、无线、移动通信等通信技术的发展,互联网应用类型和应用领域进一步拓展,网络规模 and 用户数量呈膨胀趋势。根据《2016 中国互联网发展状况最新统计报告》,截至 2015 年 12 月,中国网民规模达 6.88 亿,互联网普及率为 50.3%;预计 2020 年网民规模将达到 11 亿,互联网普及率将达到 85%。互联网规模 and 用户数量持续增加的同时,互联网应用更加丰富,逐渐被应用到金融、商贸、公共服务、社会管理、新闻出版、广播影视等经济社会生活的各个领域,为经济、政治和社会文化等方面都带来了极大的变化,也直接影响了整个社会的信息化过程。此外,互联网应用形式层出不穷:社交网络(Facebook 和新浪微博)、搜索引擎(Google 和百度)、电商平台(阿里巴巴和京东商城);新型应用不断涌现,如物联网、云计算、大数据分析等。2015 年 3 月,在第十二届全国人民代表大会第三次会议上,李克强总理提出了“互联网+”行动计划,这更为国内互联网行业注入新的发展动力。

随着互联网规模及应用的增加,原先主要面向科学研究设计的 TCP/IP 体系架构已经难以满足社会经济发展的需求,暴露出各种各样的弊端,无论服务质量、路由故障恢复、可扩展性和可管理性,还是商业模式都存在问题,严重影响了互联网的进一步发展,具体表现如下。

#### 1. 服务质量难以保障

互联网服务质量(Quality of Service, QoS)问题由来已久,新型应用对 QoS(如时延、带宽的要求)的要求越来越高,这与互联网采用“端到端”的层次化体系结构设计思想存在

一定的矛盾。现有互联网从功能上划分为通信子网和资源子网。通信子网采用无连接的数据传输方式,提供“尽力而为”的服务。资源子网(即端系统本身)对服务质量的影响较大。这导致在传输过程中,端系统无法对传输节点进行合理的管控,服务质量难以保障。虽然在 IP 层设计了 IntServ(Integrated Services)<sup>[1]</sup>、DiffServ(Different Services)<sup>[2]</sup> 和 MPLS(Multiple Protocols Label Switching)<sup>[3]</sup> 机制,但这需要传输路径上的大量网络元素(如路由器)协同配合,不仅需要通信子网的节点支持这一服务,同时增加了互联网的复杂性,影响了传输的效率。

## 2. 路由优化与故障恢复效率低下

现有互联网体系结构采用“漏斗”模型,通过 IP 协议将业务与承载分离,达到快速传输数据的能力。然而,根据 OSPF 和 BGP 协议原理,当物理链路出现故障时,OSPF 和 BGP 需要重新计算路由,由于其收敛速度慢,导致故障链路恢复时间长,需要花几十秒,甚至数分钟时间<sup>[4][5]</sup>。在此期间,可能存在数据包的丢失问题<sup>[6]</sup>。另一方面,由于各 ISP 之间的商业化运营,AS(Autonomous System)间的路由协议 BGP 的路由策略在很大程度上反映了 ISP 之间的商业利益关系,AS 间数据包的传输并非完全遵循最短路径协议<sup>[7]</sup>,这使得路由优化变得复杂。

## 3. 可扩展性差

这里的可扩展性指的是互联网传送能力的可扩展性和路由器容量的可扩展性。随着网络规模的不断增加和域间路由的日益复杂,互联网主干部分 BGP 路由表急剧膨胀,这增加了路由计算和维护的代价。虽然 IPv6 以 128 位的地址空间弥补 IPv4 空间不足的问题,但 IPv6 仍采用现行的路由机制,路由表极度膨胀导致路由更新代价大幅增加的问题严重影响互联网的可扩展性。另一方面,虽然无类域间路由 CIDR 和网络地址转换 NAT 技术的使用扩展了网络地址空间,但这些措施导致子网数目增多以及带宽资源利用率不平衡等问题,造成路由过于集中,并进一步引发网络拥塞,使得数据传输效率下降。

### 1.1.2 下一代互联网设计思路

针对上述问题,目前对互联网的改进主要有“改良式”“革命式”和“整合式”三种思路。

#### 1. “改良式”思路

“改良式”即演进式(Incremental)思想,其特点可以形象地比喻成“打补丁”,具体来说

是针对 IP 在路由效率、QoS 保障、移动性支持、地址空间等方面存在的问题,分别进行优化改进。主要研究计划包括美国的 NGI、Internet2, 欧盟的 Ambient 和 GTRN, 日本的 APAN, 中国的 CNGI 等。总的来说,演进式方案通过“打补丁”的方法,对现有互联网体系结构以及网络运行体制进行相应的修改与增补,如图 1-1 所示。例如,为高效路由而设计的多协议标签交换协议(Multi-Protocol Label Switching, MPLS),为 IP 地址不足而设计的 IPv6,为解决用户移动性而提出的 Mobile IP,为解决 QoS 问题设计的 DiffServ 和 RSVP,为安全性设计的 IPsec 等。上述方法都是对现有的 IP 协议进行相应的修补与增改,属于烟囱式、拼盘式设计和改进,既进一步加剧了网络本身的复杂性,也使得网络的全局优化更加艰难,不足以支持互联网的进一步发展,没有改变互联网的协议、组网结构缺乏反馈和自适应的本质。虽然十多年前 IPv6 被选择作为下一代互联网(NGI)协议,但目前的 IPv6 继续沿用了 IPv4 的体系架构,难以给互联网的发展带来革命性的影响。因此,这种演进式和向后兼容的设计理念只能对 TCP/IP 进行“头痛医头、脚痛医脚”的局部优化,难以在互联网上得到全面有效的利用,更无法从根本上解决互联网面临的问题,且增加了 IP 层的复杂性,违背 TCP/IP 协议设计初衷。

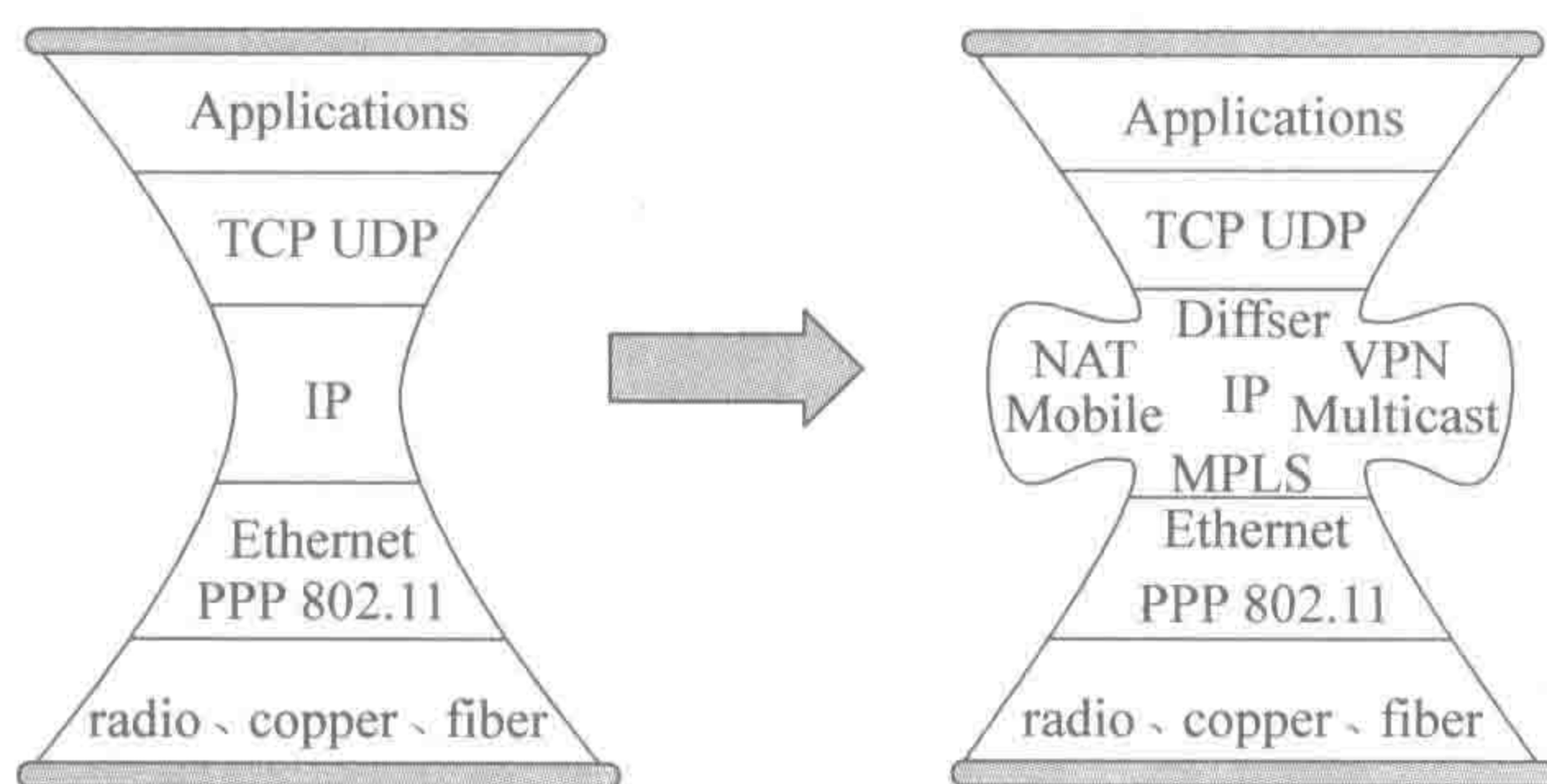


图 1-1 演进式“改良”路线

## 2. “革命式”思路

“革命式”思想,即“从零开始”(Clean-Slate)<sup>[8]</sup>研究并设计下一代互联网,也称为“后 IP 时代”互联网。该路线就是要彻底摒弃原有 TCP/IP 的束缚,采取推倒重来的方法重新设计未来互联网。通过对体系架构及相关网络运行、管理机制的重新设计,把互联网打造为集“计算”“通信”以及“存储”为一体的未来信息服务平台,彻底解决可扩展性、安全性等

问题。摆脱传统设计理念的束缚,以跨学科的思想及科学研究与工程创新相结合的方法开展下一代互联网(即未来互联网,Future Internet)的研究,可以满足社会经济等多维度的需求,构建全新的具有强鲁棒性、高自适应性、支持移动性、安全和可控可管的下一代互联网体系,推动互联网经济的健康可持续发展。

对于“革命式”路线的研究,国外已启动多个项目,主要包括美国的全球网络创新环境(Global Environment for Network Innovations,GENI)、未来互联网设计(Future Internet Design,FIND)、未来互联网科学基础(Scientific Foundations for Internet's Next Generation,SING)等;欧盟的未来互联网研究与试验(Future Internet Research and Experimentation,FIRE);日本的未来网络架构设计(Architecture Design Project for New Generation Network,ADPNGN)。

在这方面,我国政府及相关科研机构也进行了大量研究,并得到了相关政策和资金的支持,如国家自然科学基金、国家重点基础研究发展(“973”)计划、国家高技术研究发展(“863”)计划、国家科技支撑计划、中国下一代互联网(CNGI)项目等。近年来,“973”计划支持的有关未来网络体系结构的项目有一体化可信网络与普适服务体系基础研究(2007CB307100)、可测可控可管的IP网络的基础研究(2007CB310700)、新一代互联网体系结构和协议基础研究(2009CB320500)、面向服务的未来互联网体系结构与机制研究(2012CB315800)、可重构信息通信基础网络体系研究(2012CB315900)、智慧协同网络理论基础研究(2013CB329100)等。其中,由中国工程院刘韵洁院士担任首席科学家的“面向服务的未来互联网体系结构与机制研究”项目,提出全新的面向服务的未来互联网体系结构SOFIA,以服务标识为核心进行路由,是一种支持“革命式”思路的未来互联网体系结构;由北京交通大学张宏科教授负责的“智慧协同网络理论基础研究”创建了全新的智慧协同网络“三层、两域”体系架构的理论与机制,研究解决资源的动态适配、智慧映射机理和网络复杂行为的博弈决策理论等科学问题。

经过多年的研究,“革命式”研究已取得了一些研究成果,其中最具有代表性的是施乐公司的帕洛阿托研究中心(PARC)的Van Jacobson等人提出的内容中心网络CCN(Content Centric Networking)以及由美国GENI项目资助的斯坦福大学Clean Slate研究组<sup>[8]</sup>提出的软件定义网络SDN(Software Defined Network)<sup>[9][10]</sup>。

CCN来源于2010年美国NSF资助的4个未来互联网体系结构(Future Internet

Architecture, FIA) 研究项目<sup>[11]</sup>之一的“命名数据网络”(Named Data Networking, NDN)<sup>[12]</sup>。CCN 也属于信息中心网络 (Information Centric Networking, ICN) 范畴<sup>[13][14]</sup>, 即以内容为中心, 而不是以传统互联网的 IP 为中心, 如图 1-2 所示。CCN 保持沙漏模型, 通过对内容进行命名标识, 以命名标识识别内容, 取代传统互联网以 IP 地址识别主机的方式。CCN 网络弱化位置信息, 用户请求的内容与位置无关, 采用名字路由, 通过路由器来缓存内容, 从而使数据传输更快, 并能提高内容的检索效率。

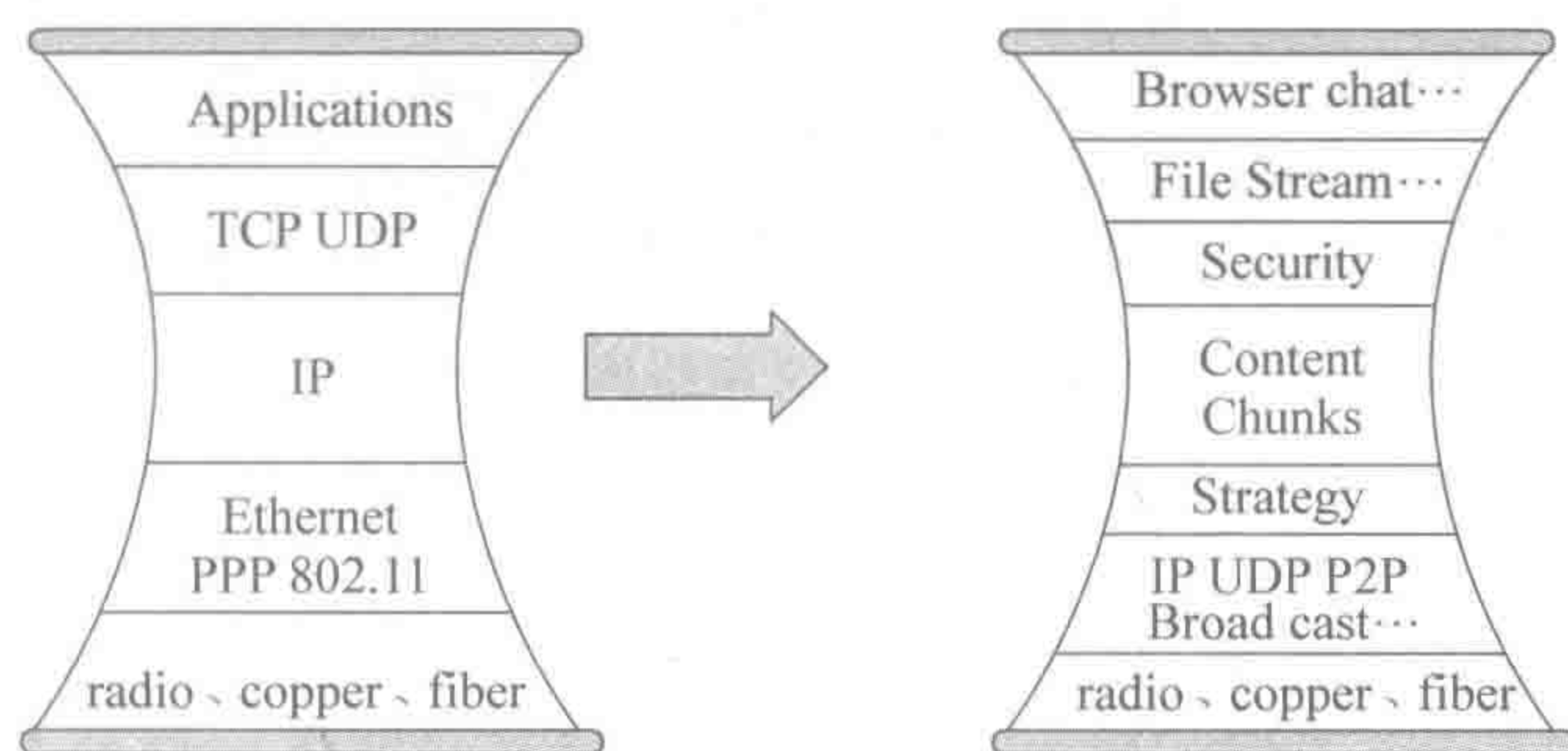


图 1-2 内容中心网络 (CCN) 模型

SDN 的设计理念是将网络的控制平面与数据转发平面进行分离, 并实现可编程化控制。SDN 的典型架构共分三层, 如图 1-3 所示, 最上层为应用层, 包括各种不同的业务和应用; 中间为控制层, 属于控制平面, 主要负责处理数据平面资源的编排, 维护网络拓扑、状态信息等; 最底层为基础设施层, 属于数据平面, 负责基于流表的数据处理、转发和状态收集。其中控制平面与数据平面之间的接口, 称为控制数据平面接口, 即南向接口, 主要功能是流表的下发以及拓扑数据收集等, 其核心技术是 OpenFlow<sup>[15][16]</sup>。而应用层与控制层之间的接口服务于用户, 是北向接口, 目前没有统一的标准。

现有网络中, 用于决策和管理的控制平面与负责数据转发的数据平面紧耦合在每个网络设备中 (如路由器、交换机)。而在 SDN 网络中, 网络设备只负责单纯的数据转发, 可以采用通用的硬件; 而原来负责控制的操作系统将提炼为独立的网络操作系统, 负责对不同业务特性进行适配。网络操作系统一方面抽象了底层网络设备的具体细节, 同时还为上层应用提供了统一的管理视图和编程接口。总结起来, SDN 本质上具有“控制和转发分离”“资源的集中管理与控制”和“通用硬件及软件可编程”三大特点。

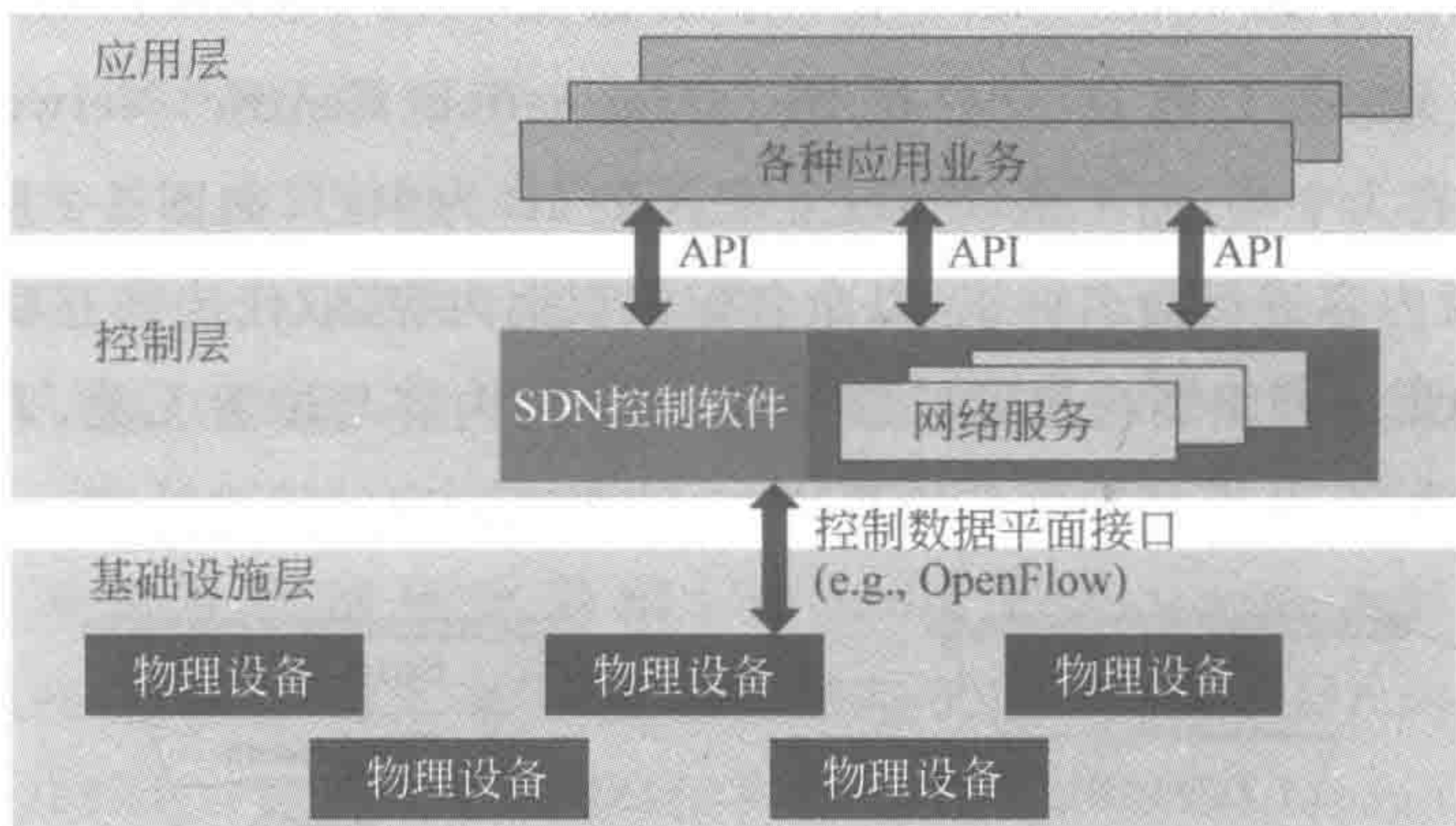


图 1-3 软件定义网络体系架构

虽然“革命式”思想超前，提出了对互联网现有的体系结构进行重新设计，但是，由于现在基于 IP 互联网的广泛部署，采用“革命式”的方案对互联网进行彻底革新还需要一个漫长的过程。

3. “整合式”思路

“整合式”思路是一种介于“改良式”和“革命式”之间的折中方案，即对于现有互联网需要迫切解决的各种问题，寻求一个系统性、大范围、整体性的修补。“整合式”线路采用覆盖网改造现有互联网。覆盖网也称为层叠网，或叠加网络，其本意是指建立在另一个网络上的网络。在互联网领域，覆盖网是指叠加在互联网 IP 层之上，应用层之下增加一个“中间层”，为上层业务和应用提供针对性的服务，如图 1-4 所示。覆盖网络不需要大规模改变现有网络架构就能提供更为可靠、容错性更好的服务；也不抵制用户的创新，只对已有的应用和业务进行适当管理和控制，是对现有互联网体系架构的“系统性”修补。使用覆盖网络技术，即使网络层出错误，如链路断裂或拥塞，应用系统也可以凭借覆盖网络快速找到替代路由，并且

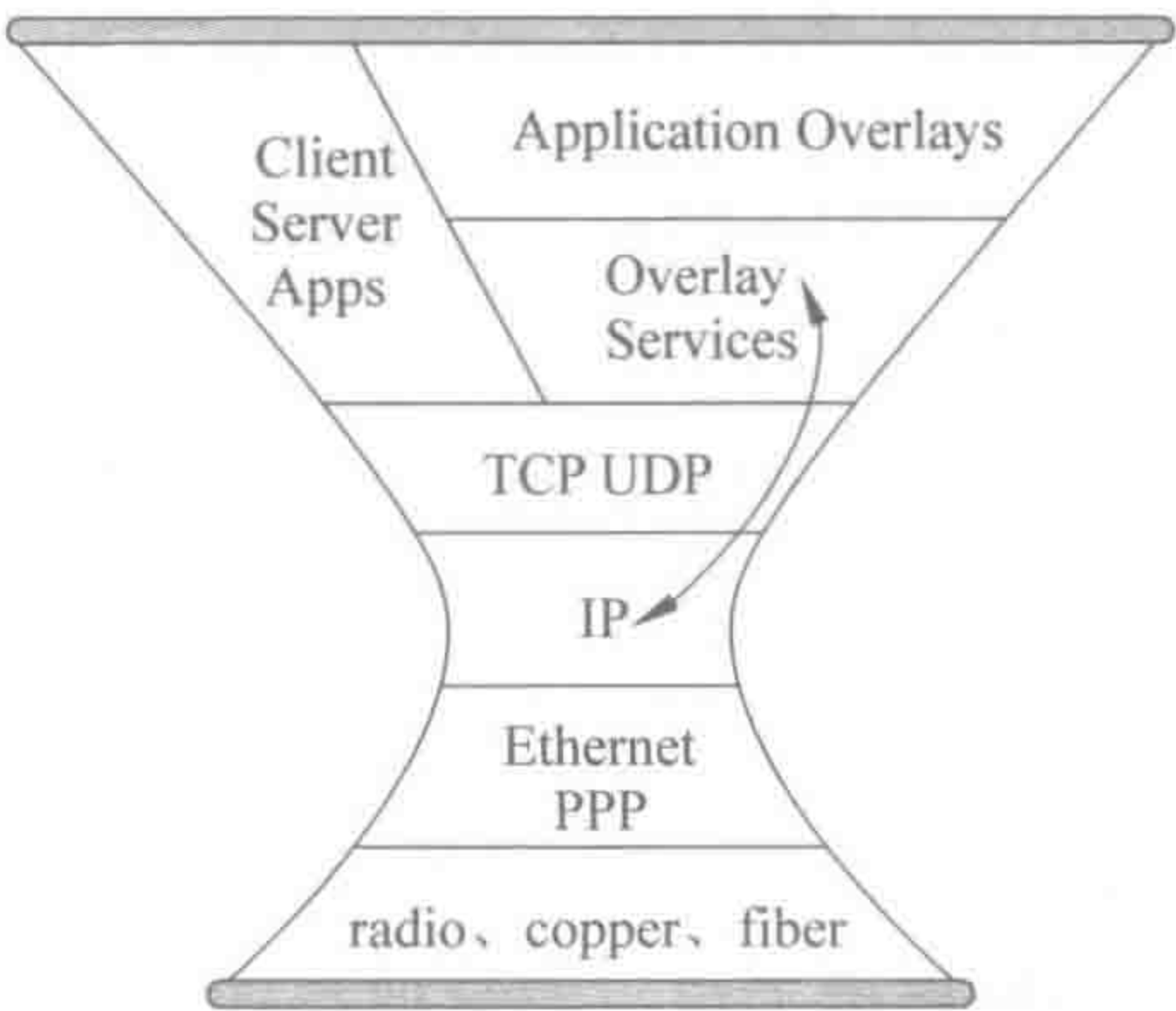


图 1-4 覆盖网络体系结构

可以根据应用服务对 QoS 的不同要求寻找相应的最优路径。

## 1.2 覆盖网络的本质

### 1.2.1 覆盖网络体系结构

覆盖网是构筑在已有的互联网基础之上,通过选择并合理连接节点构建一层新的网络,提供类似基础设施所提供的基础性服务,如路由、组播、内容分发等。覆盖网是一种逻辑网络,其节点由连接在互联网的部分终端节点或应用类服务器组成,具有数据转发、处理和存储等功能,节点间通过虚拟逻辑链路连接起来,一条覆盖网链路对应一条或多条物理链路。覆盖网络探测底层物理网络的链路状态信息,并根据自己的策略为覆盖网络中的数据流计算路由,然后将数据流的转发路径发送给底层物理网络,由底层物理网络按照指定的路径进行传输。覆盖网络上的路由是应用层的路由机制,独立于网络层的路由机制,比网络层路由反应快。例如,当底层物理网络路径出现故障时,可以通过覆盖网络快速找到替代路径,极大地减少了故障恢复的时间。

覆盖网络是一种面向服务的网络,由服务提供商们(Service Providers)根据用户的需求,在一个物理网络上创建多个独立的覆盖网络,定义各自的虚拟拓扑图,来完成一些特定应用或用户群体的特殊需求,如图 1-5 所示。覆盖网络克服了传统网络路由的功能局限性,帮助改进了现有互联网的路由性能,满足了新型业务的特定需求。例如,提供满足用户 QoS 需求的网络传输服务;实现快速故障恢复,帮助提高网络可靠性;以及提供网络内容分发和多播服务等。

### 1.2.2 覆盖网络的缺陷

虽然覆盖网络可以帮助改进现有互联网的网络性能,但是也会带来一些新的问题,概括如下。

#### 1. 过度依赖终端节点

首先,虽然终端主机节点的性能和存储容量在不断提升,但它对数据包的处理和转发的能力低于路由器。已有的覆盖网研究成果在构建覆盖网拓扑时,其节点依赖于参与服