未来互联网体系结构研究综述

谢高岗 张玉军 李振宇 孙 毅 谢应科 李忠诚 刘韵洁

(中国科学院计算技术研究所网络技术研究中心 北京 100190)

摘 要 互联网逐渐成为社会基础设施,现有TCP/IP体系结构面临诸多挑战,未来互联网体系结构成为研究热点.文中分析了现行互联网在可扩展性、动态性、安全可控性等方面面临的根本性问题,综合比较了面向可扩展性、面向动态性及可信未来互联网体系结构研究,讨论了相关体系结构存在的问题.实验验证是未来互联网研究的重要手段,论文进一步分析了支持互联网体系结构持续创新所需的可编程虚拟化路由器及其试验床的研究进展.论文最后讨论了未来互联网体系结构有待重点研究的相关问题.

关键词 未来互联网;体系结构;试验床;可编程虚拟化路由器;下一代网络;下一代互联网

中图法分类号 TP393 **DOI号**: 10.3724/SP.J.1016.2012.01109

A Survey on Future Internet Architecture

XIE Gao-Gang ZHANG Yu-Jun LI Zhen-Yu SUN Yi XIE Ying-Ke LI Zhong-Cheng LIU Yun-Jie

(*Network Technology Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190)

Abstract Internet has become as a social infrastructure. The current Internet architecture based on TCP/IP is faced with many challenges. This fact makes the clean-slate design of future Internet architecture be a hot research topic. This paper analyzes the essential problems that are inherent in the current Internet architecture, including scalability, mobility, security and controllability, and comprehensively introduces the existing research works on the future Internet architecture such as the scalability-oriented architecture, the mobility-oriented architecture etc. Considering experiment-based validation is an important means for the research of future Internet, this paper further presents the development of programmable virtual routers and the corresponding testbeds. Finally, open issues on the future Internet architecture design are discussed.

Keywords future Internet; architecture; testbed; programmable virtual router; next generation network; next generation Internet

1 引 言

经过40多年的发展,互联网(Internet)已经成为集信息采集、传输、存储与处理于一体的信息社会的重要基础设施.Internet设计之初主要用于解决大型机等资源的时分复用问题,尽管计算技术、通信技术与应用模式不断进步,计算模式经过单机模式、

收稿日期:2011-01-31;最终修改稿收到日期:2012-03-01.本课题得到国家"九七三"重点基础研究发展规划项目基金(2012CB315801)、国家自然科学基金(61133015)资助.**谢高岗**,男,1974年生,博士,研究员,博士生导师,当前研究兴趣为未来互联网体系结构、可编程虚拟化路由器、网络测量分析与模型化.E-mail: xie@ict.ac.cn.**张玉军**,男,1976年生,博士,副研究员,博士生导师,当前研究兴趣为未来互联网体系结构、可信互联网.**李振宇**,男,1980年生,博士,副研究员,当前研究兴趣为未来互联网体系结构、P2P计算.**孙 毅**,男,1979年生,博士,副研究员,当前研究兴趣为未来互联网体系结构、移动互联网.**谢应科**,男,1972年生,博士,副研究员,当前研究兴趣为可编程虚拟化路由器.**李忠诚**,男,1962年生,博士,研究员,博士生导师,研究领域为计算机网络.**刘韵洁**,男,1943年生,博士生导师,中国工程院院士,研究领域为计算机网络.

客户/服务器模式、点对点模式发展到目前的云计算模式，但作为互联网最重要技术基础的 TCP/IP 体系结构却基本保持不变[1]. 与此同时，互联网接入方式和网络角色定位发生了根本性变化，移动接入、物联网以及延迟容忍网络(Delay Tolerant Network，DTN)和固定接入一样，成为互联网的主要接入方式；云计算成为重要的互联网计算模式，多媒体内容分享应用和社会网络(Social Network)成为互联网的新型应用，网络已经成为集信息采集、传输、存储与处理于一体的信息平台，而不仅仅是数据传输通道. 接入方式的变化和网络角色定位转换导致以 IP 地址为核心、以传输为目的、按照端到端原理设计的 TCP/IP 体系结构在路由扩展性、动态性、安全、可管理、可靠性、QoS 以及能耗等方面的问题日益突出[2].

为解决 TCP/IP 体系结构存在的问题，人们进行了大量研究，并取得了丰硕成果，如 IP Multicast、InterServ 与 DiffServ 等，但真正被全网部署的协议屈指可数[3]. 这既有网络规模太大不易统一部署的原因，也有互联网经济学方面的原因. 互联网目前更多地体现为社会经济基础设施，而非单纯的科研设施，除非出现现有体系结构及其改进方法(如 Mobile IP、Classless Inter-Domain Routing、Network Address Translator、IP Security 等)无法承载的新型应用，或者在目前体系结构下互联网即将崩溃，或者进一步网络投资大于预期收益，否则在目前这种状况下，即使对互联网协议进行微小的修改都缺乏部署动力，如 IPv6 产生十余年以来，一直缺乏大规模部署，对互联网体系结构进行革命性全新设计更是无从谈起. 毕竟当前的互联网与以前用 DNS(Domain Name System)代替 hosts.txt 以及部署 CIDR(Classless Inter-Domain Routing)时的互联网所处的环境已经完全不同.

随着云计算、物联网、移动通信技术的发展，TCP/IP 体系结构是否到了需要革命性重新设计(Clean-Slate Redesign)的时候[4]？回答这个问题首先要分析当前体系结构是否面临无法克服的根本性障碍以及运营商是否有采用革命性体系结构的动机[5]，这些也是未来互联网体系结构(Future Internet Architecture，FIA)研究的出发点与目标.

本文分析了 TCP/IP 体系结构面临的根本性问题，总结并分析对比了近年来 FIA 相关的重要研究工作，并指出 FIA 领域需要重点关注的问题. 本文第 2 节分析 TCP/IP 体系结构面临的根本性问题；第 3 节综合分析比较了目前 FIA 的研究进展，包括面向可扩展性的体系结构、面向动态性的体系结构以及解决其它问题的体系结构；第 4 节总结未来互联网系统实现及试验床方面的研究进展；第 5 节总结 FIA 领域需要重点关注的问题.

2 TCP/IP 体系结构面临的根本性问题

未来互联网体系结构研究必须首先明确当前体系结构面临的根本性问题. 本节分析当前 TCP/IP 体系结构面临的根本性问题，包括可扩展性问题、动态性问题及安全可控性问题等.

2.1 可扩展性问题

互联网的基本模型是基于全局地址的点到点通信，通信链路因路由器转发数据包而被统计复用. 新型应用和计算模式的出现导致互联网流量日益增多：数据密集型计算与文件分发业务(例如视频)的发展以及接入带宽的增加(例如 LTE)使得网络流量持续增加；云计算模式由于规模经济效应的原因，逐渐得到广泛应用，数据被提交到云计算平台存储处理，云计算模式进一步增加了网络的流量负载；物联网延伸了互联网感知物理世界的触角，并将逐渐从互联网的边缘网络转变为互联网的共生网络，成为网络数据资源的重要来源. TCP/IP 体系结构的点到点通信模型导致这些新增的网络流量最终都被汇聚到骨干网络与数据中心接入链路上，导致网络流量增加的速度远远超过摩尔定律与路由器性能提升的速度①，网络规模不断地被动扩张. 另外，路由器通过维护路由表实现 IP 地址路由，IP 地址分配不均衡、多宿主技术、地址碎片、流量工程等使得路由表难以实现高效率聚合，导致骨干路由表急剧膨胀. 目前，活跃路由表条目已经达到 35 万条②，而且以每两年 1.3 倍的速度增长[6]. 路由条目的快速增长将极大降低路由查找性能，增加路由器实现开销.

流量激增和路由表急剧膨胀相互作用已经严重影响到互联网的可扩展性. 而互联网可扩展性问题的根源在于 TCP/IP 体系结构基于 IP 地址的点到点通信模式，这一基本特征导致所有通信流量都被汇聚到骨干网络上，并且所有骨干路由器都需要维护到达任意节点(子网)的路由. CDN(Content

① http://www.huawei.com/broadband/iptime_backbone_solution/era/100g_transport_era.do

② The CIDR Report. Available at http://www.cidr-report.org

Delivery Network)[7]通过在互联网之上部署新的大规模基础设施来缓存数据，从而缓减流量激增问题.但是CDN不仅昂贵而且仅对签约用户的特定应用数据进行优化，不能从根本上解决流量激增带来的互联网可扩展问题.

从上述分析可以看出：可扩展性问题是目前互联网面临的根本性问题之一.在现有互联网体系结构中，只能通过不断地增加硬件设备投资来缓解这个问题，无法找到根本性的解决方案.设计不以IP为中心的未来互联网体系结构，采用数据与带宽结合统计复用的方法，构建网络友好的业务系统与业务友好的网络系统，将为未来互联网中可扩展性问题提供根本解决方案.

2.2　动态性问题

早期互联网主要为以计算机为代表具有一定处理能力的固定终端提供数据交换服务.经过几十年的发展，互联网终端形态发生了很大变化，突出表现在网络的动态性显著增加.造成网络动态性增加的主要原因有两个：(1)便携式移动终端日益普及，移动终端上的业务流量占网络流量的比重日益增加.对日本最大社会网络Mixi访问方式的分析结果表明，2010年84%的访问是通过移动终端进行的，而这一数字在2006年时仅为14%.摩根斯坦利公司的调查报告预计，到2012年全球智能手机的业务流量将首次超过PC①.终端的移动性显著加剧了网络的动态性及其上承载服务的不稳定性[8].(2)物联网的出现将使得低智能终端数量爆炸式增长.同PC等固定终端不同，以传感器、RFID为代表的物联网终端能量和处理能力都受到很大限制，休眠模式、自组织、按需路由等机制的引入使得这些节点上连接的动态性较之固定终端明显加强[9].

网络节点动态性的增强导致数据传输路径频繁变换，严重破坏了上层应用服务的连续性，影响了互联网用户的服务质量.如前所述，TCP/IP体系架构最初就是为具有一定处理能力的固定终端设计的：(1)IP地址的二重表达特性(既表达身份又表征位置)不能很好地支持终端移动，移动IP的思想使得协议栈冗余，处理效率低；(2)TCP/IP端到端的通信模式将服务连接的维护管理工作交由终端完成，这对于低智、低能的物联网终端是个很大的挑战.

从上述分析可以看出：动态性是目前互联网面临的另一个根本性问题.增加网络的智能性、减少终端的负担，设计身份和位置分离的新型体系结构，将移动管理和路由设计统一考虑，使得节点在移动过程中具有唯一、稳定的标识，将为未来互联网中动态性问题提供根本的解决方案.

2.3　安全可控性问题

TCP/IP互联网设计之初，网络规模限制在一个相对封闭、可控的范围内，基本不存在安全隐患.网络技术和应用的发展最终使互联网发展成为一个开放、不可控的复杂系统，不可避免地面临多种安全威胁.特别地，目前出现的在线社会网络(Online Social Network，OSN)应用对安全和隐私保护提出了更高的要求.针对互联网安全问题，现有解决方案大体可分为两种手段：扩展网络协议(如IPSec、SSL/TLS等)，采用加密/认证技术手段保证通信数据的安全，这种方式导致协议栈臃肿不堪，通信效率下降；在网络中增加多种安全设备(如防火墙、入侵检测设备等)，检测和抵御攻击行为，安全设备的多样化导致其难以协调发挥综合防控效果.可以看出，目前互联网的安全手段基本处于被动应对状态，不是一个系统的解决方案.基于IP地址的点到点通信模式注定现有安全手段只能够提供端到端安全通道，无法实现针对服务及内容的个性化安全服务[10-11].

应用性能由网络传输性能，典型的指标如IETF IPPM(IP Performance Metrics)，与服务提供者的服务性能综合决定.尽管业界提出了大量的资源保障机制用于增强尽力而为的传输机制，但这些机制仅能保证单跳链路的传输性能，无法保证最终用户需要获得的服务质量.应用性能的可控不仅取决于路径的传输性能，还取决于传输与数据资源的全局优化：Verizon网络测量发现，P2P数据块平均传输距离为1000英里以上，平均经过5.5个Metro-hops[12].但在以IP路由为核心的互联网体系结构中，服务资源与物理网络缺乏互感知机制，要提供满足应用性能要求或者可预期性能要求的资源优化非常困难.例如P2P覆盖网络与网络物理拓扑的失配问题，导致在BitTorrent系统中，尽管50%～90%的数据块已经存储在本地在线用户上，但是用户依然选择在外部节点下载[13].现有互联网基于IP地址的通信模式导致用户在获取服务时需要以IP地址加端口的组合方式唯一指定服务提供者，只能采取在服务请求者和服务提供者之间的路径上预留资源的方式来保障服务质量.

---

① Meeker M, Devitt S, Wu L. Internet trends. CM Summit, New York, 2010. http://www.morganstanley.com/institutional/techresearch

从上述分析可以看出，缺乏安全可控的服务性能是互联网面临的另一个根本性问题.设计面向服务的未来互联网体系结构，支持服务资源和物理网络的相互感知，构建面向服务和数据的安全与可信架构，从源头上限制网络攻击行为的发生，构建服务性能与网络服务能力、用户需求的定量模型，以服务性能最大化为目标决定网络提供服务的方式，将为未来互联网中安全可控性问题提供根本的解决方案.

综上所述，现有 TCP/IP 体系结构无法解决互联网面临的可扩展性、动态性和安全可控性等根本性问题，需要设计新的互联网体系结构并研究相关关键机理.对未来互联网的研究思路可以分为两种：演进式或增量式(Incremental)和变革式(clean-slate).演进式路线通过“打补丁”的方式对 TCP/IP 互联网进行修改和补充，其核心仍然是 TCP/IP 体系结构，并未从根本上解决 TCP/IP 体系结构面临的问题，不能或者不能很好地适应互联网接入方式和数据交换的巨大变化.最近几年，国际上开始以变革式路线研究未来互联网体系结构，即“从零开始”，不受现有互联网体系结构的约束，研究新的体系结构，这种变革式思路已经成为设计未来互联网体系结构的共识.尽管如此，为推动新的互联网体系结构的部署和实施，在设计过程中也需要考虑与现有互联网的互通，支持现有互联网向新型互联网的演进式部署.

3 未来互联网体系结构(FIA)研究进展

针对现有 TCP/IP 体系结构存在的可扩展性、动态性、安全可控性等问题，本节综合分析了现有解决上述问题的研究进展.

3.1 面向可扩展性的体系结构

互联网可扩展性问题包括两个方面：流量可扩展和路由可扩展.德国 Ipoque 的统计数据表明，目前网络带宽“消费大户”是 P2P 文件共享，在北美和东欧地区分别占据了 43%和 70%的网络带宽①.造成上述问题的原因在于 P2P 的数据传输机制，P2P 过于强调“对等”，每个节点之间的数据交换完全是无序的，为解决这一问题，美国耶鲁大学的研究人员提出了 P4P 方案[13].与 P2P 随机挑选伙伴节点不同，P4P 方案充分利用网络运营商掌握的网络拓扑、收费策略等控制信息，通过智能运算选择地域临近的网络来进行数据交换，最大程度降低数据传输的跳数，提高数据传输容量.

P4P 方案有助于消除 P2P 覆盖网络与实际网络拓扑之间的失配，它的核心原则是在现有网络体系结构的基础上增加新的设备，基于用户数据请求特征，实现数据感知与资源调度，并且依赖于运营商与服务提供商之间的协助.正是由于上述问题，人们期望从体系结构本身来解决流量增加问题，CCN[10]/NDN[11](Content Centric Network/Named Data Networking)是目前解决流量扩展性问题的革命性体系结构.

CCN/NDN 旨在通过新体系结构的设计，实现以地址、位置为中心的网络向以数据为中心的网络的转变，拟解决地址空间爆炸、动态性、地址可扩展等问题，同时借助存储优化流量，避免拥塞，提升网络吞吐量.如图 1(a)所示，CCN/NDN 在体系结构中引入命名数据作为协议栈的细腰，不再采用基于 IP 地址的方式，允许应用按照需要独立选择命名方案；基于命名数据的设计可直接满足应用需求，无需中间件实现从 IP 地址到所需内容的映射，简化网络架构设计；通过在中间路径上路由器中缓存命名数据，便于后续请求就近获取数据，达到避免拥塞、优化流量的目的.如图 1(b)所示，CCN/NDN 采用基于名字的路由和转发，通过定义标准接口将路由和转发分离，有助于两者平行研究，特别是可在转发不变的前提下研究路由优化.CCN/NDN 使得内容分发变得更加自主，用户可自由选择最优链路.CCN/NDN 模型与现有网络模型具有较好的兼容性，可以运行于任何层次之上，同样任何层也可以运行于 CCN/NDN 之上，对内容分发具有优势，并能很好地支持流量的可扩展性.

CCN/NDN 以数据为中心的思想能够很好地减少网络中传统数据密集型应用(如在线视频、文件下载等)的流量.然而在未来互联网中，业务类型将更加多样化，许多新型应用如实时应用、交互式应用等将日益流行，为了保持这些动态数据的连接，CCN/NDN 要求周期性地广播探测包从而带来大量的信令开销.此外，在未来互联网云计算模式下，CCN/NDN 仅仅缓存数据，而对数据的处理仍然需要到远端的云计算平台上进行，这也一定程度上限制了该架构对网络流量优化的作用.从上述分析可以看出，未来互联网仅仅以数据为中心是不够的，数

① Schulze H, Mochalski K. Internet study 2008/2009. http://www.ipoque.com/resources/internet-studies/internet-study-2008_2009, 2011

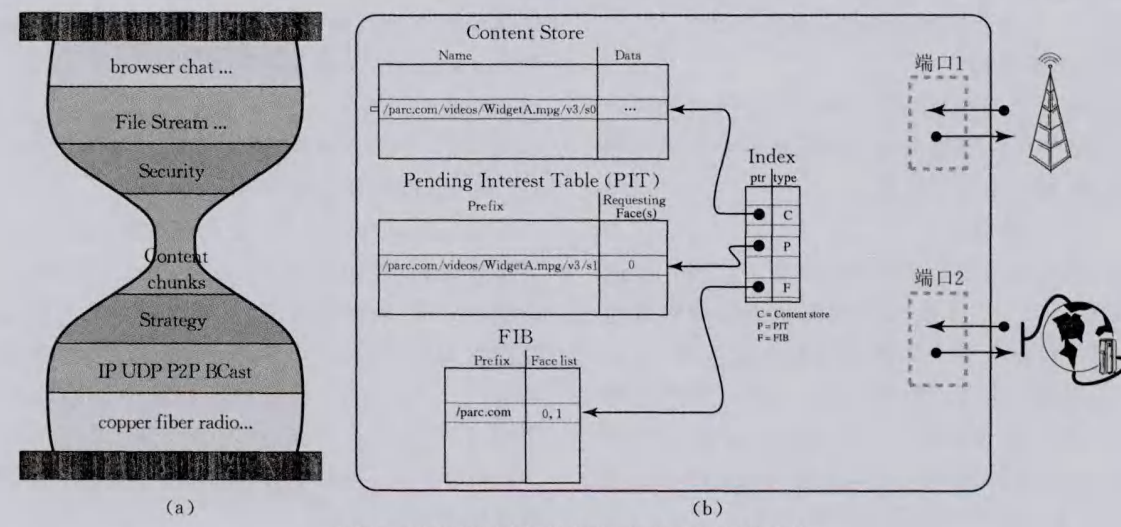

图 1　CCN/NDN 协议栈和节点转发模型[10-11]

据是服务的载体，要想全面支持流量的可扩展性，必须统筹考虑数据和数据处理的边界与融合，构建以互联网服务为中心的网络体系架构.

针对路由可扩展问题，Cisco 公司提出了位置/标识分离协议 LISP(Locator/Identifier Separation Protocol)[14]，其目的是通过采用聚集度更高的机制解决路由可扩展问题，目前 Cisco 已经发布了多个 LISP 的实现版本，其基本思想是将基于 IP 地址的单一编址空间划分为两个编址空间：EID(端点标识)和 RLOC(路由位置符)，其中 EID 仅被用于站内通信，RLOC 则可用于全局通信. LISP 可采用渐进式方式部署在现行网络上，具体通信模型则采用 Map-and-encap 方案，即首先将不可路由的 EID 映射为可路由的 RLOC，再进行隧道封装后在现有网络上传输. LISP 是在 TCP/IP 体系结构框架内针对路由可扩展性的一种改进，在一定程度上解决移动性问题，仍然无法解决流量增长问题.

3.2　面向动态性的体系结构

动态性支持是未来网络的基本特征之一[15]，位置和标识分离是解决动态性问题的主要解决方案之一. 在面向动态性的 FIA 研究方面，当前主要有两方面的工作：斯坦福大学的 Clean-Slate 项目[16]、美国自然科学基金的 MobilityFirst 项目[17]等.

斯坦福大学的 Clean-Slate 项目将重新设计互联网基础设施和服务，为设备互联、计算和存储创造新的创新平台，并使得他们能够为研究者和用户所用，而新的架构将主要考虑移动计算的需求. 考虑到目前大量的 WIFI 和蜂窝接入网络，Clean-Slate 项目的目标是为用户提供最优的接入选择，其基本思想是将服务提供者和网络运营商分离，实现安全的用户和设备接入控制，由网络跟踪用户的移动过程和位置信息，从而提供移动过程中的高效切换和服务质量. 按照 Clean-Slate 的规划，需要在路由、切换、目录服务、安全和接入控制等方面进行重大创新.

MobilityFirst 项目是美国自然科学基金会启动的 4 个面向 FIA 的研究项目之一，由美国罗格斯大学联合其他 8 家大学共同实施，其目标是针对目前移动接入设备激增的现实，设计面向移动/无线世界的 FIA 及协议. MobilityFirst 的两条基本设计原则是：将无线/移动终端作为主流接入设备设计网络架构；设计更为强大的安全和信任机制. MobilityFirst 体系架构的基本技术特征包括：支持快速的全局名字解析；采用公钥基础设施实现网络设备的验证；核心网络采用扁平地址结构；支持存储-转发的路由方式；支持逐跳的分段数据传输；支持可编程的移动计算模式等. MobilityFirst 可以很好地支持移动计算，但同样无法解决流量激增带来的带宽消耗问题.

在上述方案中，Clean-Slate 主要考虑接入优化的问题，并强调在网络侧实现移动管理；MobilityFirst 的目标是设计一种健壮、可信的以服务为中心的网络架构，通过全局名字解析和转发存储实现路由，其基本思想仍然与现有网络架构一致，即先获取节点位置再获取服务；LISP 在现有网络结构中，通过定义二维节点标识在现有网络体系结构下实现位置和标识相分离. FIA 支持动态性的基本思路应该是：位置和标识分离，使得节点在移动过程中具有唯一、稳定的标识；移动管理和路由设计统一考虑，支持主机和网络移动；采取缓存机制，提升节点移动过程中服务和数据的获取效率；设计良好的安全机制，支持用

户自认证，支持动态信任传递，提升接入效率.

3.3 其它体系结构

除面向可扩展性及动态性的 FIA 外，当前还有许多体系结构关注网络的可控可管、网络可信、以服务为中心的网络设计等问题.

3.3.1 可管可控结构

网络的可管可控是未来互联网研究的重点问题之一. 现有 TCP/IP 体系结构把控制逻辑和数据包处理紧耦合在分布于全局的路由器和交换机上. 由于数据网络通常由多家机构运维且部署在各种各样的环境中，因此需要复杂的网络层面的管理. 这些需求引发了对路由控制协议以及管理平面的改进和修补，导致了复杂的控制和管理机制.

美国卡内基梅隆大学的研究者重新设计了互联网控制和管理结构，提出了 4D(Decision, Dissemination, Discovery, Data)[18]体系结构. 其基本思想是把网络决策逻辑从管理网络设备交互的协议中分离出来，即决策逻辑从网络设备中分离出来. 该体系结构遵循 3 个设计原则，即网络层面的目标、全局网络视图和直接控制. 从功能上来说，决策平面实现在集中的服务器上，借助全局的网络视图来集中地实现管理目标，并直接把决策下发到路由器；分发平面提供到路由器的可靠通信通道；而路由器上的发现平面收集自己的资源和局部环境的信息；数据平面则负责数据转发. 这种分离的结构有助于实现简单、可靠和灵活的网络控制和管理.

3.3.2 可信网络结构

互联网在设计之初假设网络节点是可信的，因此只需要提供尽力而为的数据包传输就足够了. 而事实上，随着互联网的迅速发展，网络节点不再可信，导致了大量的安全问题，如 ARP 欺骗、DoS 攻击等. 北京交通大学的张宏科教授等提出了两层可信网络体系结构[19]，即“交换路由层”和“普适服务层”. 前者提供网络和终端的接入，而后者负责各种业务的会话、控制和管理. 为了解决 IP 欺骗问题、支持移动性和提高网络管理和控制能力，交换路由层把接入标识和路由标识区分开来，即把接入终端的标识和位置分离开来，而不再集中在 IP 上. 普适服务层统一处理和描述网络服务和资源，并采用多连接多路径传输机制提高传输效率

互联网面临的另一个问题是源地址欺骗，这一问题产生的根源是网络设备仅检查数据包的目的地址，而对源地址不进行验证，增加了安全、管理以及计费的难度. 清华大学的吴建平教授等提出了源地址认证体系结构 SAVA(Source Address Validation Architecture). 这是一个层次化体系结构，包括接口 ID 层、IP 前缀层以及 AS 层，该结构已经在 CERNET 上部署和测试[20-21].

3.3.3 面向服务的结构

对互联网用户来说，访问互联网的目的在于获得某种服务，比如文件下载服务、语音通话服务或者是软件服务. 云计算将应用、软件以及平台作为服务提供给用户，是未来计算的一种模式. 美国自然科学基金项目 Nebula① 研究以云计算为核心的体系结构，其目标是构建可靠、可信的核心网络连接各个云计算数据中心. 数据中心可由多个提供商提供，互相通过副本备份等技术使得用户可连接到最近的数据中心. 核心网络通过冗余高性能链路和高可靠的路由控制软件实现高可用性. 用户通过安全和可信的数据链路访问数据中心. Nebula 天然地支持云计算这一面向服务的未来计算模式. 该结构的研究重点在于数据中心构成的核心网络，对边缘用户关注较少，缺乏对移动计算问题的有效解决方案.

美国自然科学基金的 XIA(eXperessive Internet Architecture)项目②不再局限于对某一种特定通信场景的支持，而强调演进的网络通信实体和场景. XIA 对每一种通信规则(principle)，如主机、内容和服务等，定义一个细腰，用户通过指定适当的通信实体来表达通信的意愿，从而支持演进的网络实体[22].

SOFIA[23]是一种面向服务的互联网体系结构. 互联网不再仅仅作为传输通道，而是被看作是服务池. SOFIA 以标识的服务作为协议栈的细腰，通过服务的迁移等技术可实现服务的本地化，有效解决互联网流量激增带来的问题；通过标识和地址的分离有效支持泛在移动计算. 和云计算类似，该结构借助 Pay-as-you-go 支付模式发挥统计复用的规模效益. 面向服务的体系结构研究与协议栈实现还包括 Serval[24].

3.4 FIA 研究进展的综合比较

上面给出了可扩展性的体系结构、面向动态性的体系结构及其它体系结构的研究进展，介绍和分析了每一类体系结构的典型代表及其主要技术思路. 表 1 综合比较了现有 FIA 的主要研究进展.

① http://nebula.cis.upenn.edu/

② XIA. http://www.cs.cmu.edu/~xia/, 2011

表 1　现有 FIA 研究的综合比较

分类	具体方案	设计目标	主要思想	与 TCP/IP 的关系	当前进展	进一步分析
面向可扩展性的体系架构	CCN/NDN	提高扩展性与安全性	以数据为中心	采用非 IP 架构	进行中	适合静态大数据的共享
面向动态性的体系架构	Clean-slate	无线接入与移动性能	通过网络侧的支持实现移动管理	支持 TCP/IP 架构下的移动	进行中	主要考虑 WIFI 和蜂窝的接入优化问题
	MobilityFirst	以移动为中心的网络结构	全局名字解析、存储转发	标识和地址分离	进行中	先获取节点位置再获取服务
其它体系架构	可管可控结构(4D)	重新设计互联网控制和管理结构	把网络决策逻辑从管理网络设备交互的协议中分离出来	针对 TCP/IP 网络设计	已经结束	网络状态的感知与资源调度控制
	可信网络结构	增强互联网的可信性	源 IP 地址认证、传输和服务分层的两层结构	对 TCP/IP 互联网的改进	正在进行	身份与资源安全认证
	以云计算为中心的结构(Nebula)	构建可靠、可信的核心网络连接各个云计算数据中心	核心网络通过冗余高性能链路和高可靠路由控制软件实现高可用性	全新的面向云计算的体系结构	正在进行	云计算资源的高效可靠互联与访问
	SOFIA/Serval	构建面向服务的未来互联网体系结构	定义服务标识，实现可扩展服务路由	非 IP 架构，支持 IPv6 环境中的演进式部署	正在进行	解决互联网面临的可扩展性、动态性和安全可控性问题

4　系统实现与验证

未来互联网体系结构的研究需要大规模网络试验床的实际测试与验证，可编程虚拟化路由器是构建未来互联网试验床的核心设备. 路由器需要在控制平面和数据平面向用户提供灵活的编程接口，允许用户根据不同的协议需求，修改路由器的功能模块，这是未来互联网新协议和新架构研究的基础. 路由器平台需要同时虚拟运行多个不同体系结构的路由器实例，各个实例之间需要独立无干扰. 路由器的这种特性可以使未来多样化的网络体系结构与现有 TCP/IP 体系结构长期并存，并支持未来网络体系结构的渐进式试验与部署. 可编程虚拟化路由器对未来互联网体系结构的研究具有重要意义，成为近几年网络研究的热点问题之一[25-30]. 本节分析了可编程虚拟化路由器和试验床的研究进展.

4.1　可编程虚拟化路由器

从系统实现来看，可编程虚拟化路由器的研究与实现可以分为两类：基于通用平台的路由器和基于专用硬件的路由器.

基于通用平台的可编程虚拟化路由器用普通网卡实现数据的采集和转发，通过在 CPU 上运行虚拟化软件（Xen[31]、OpenVZ① 等）和路由器软件（Click[32]、XORP[33]、Quagga② 等）实现可编程虚拟化路由器，其特点是具有非常高的灵活性，但转发性能低. 近几年来，随着硬件技术的发展，特别是高速网卡及多核 CPU 技术的发展，通用平台给可编程虚拟化路由器带来了新的契机. 研究者们在通用平台上，设计了高性能的可编程虚拟化路由器，如 RouteBricks[25]、PacketShader[26] 等. RouteBricks 利用 Intel 公司高性能的多队列网卡，设计了一个 35 Gbps 的路由器原型系统. PacketShader 利用 GPU 强大的并行计算能力进行 IP 查找，有效地卸载了 CPU 的负担；通过简化 Linux 内核中数据包缓存区的数据结构、静态分配大的连续缓冲区、采用批处理方式等，优化数据包 I/O 引擎. PacketShader 在单台服务器上，64 字节小包的转发性能接近 40 Gbps.

基于专用硬件平台的可编程虚拟化路由器，利用专用硬件（如 FPGA 等）的高速、并行特性，实现高性能的虚拟化路由器数据平面. Lockwood 等人[34] 首次在 NetFPGA 中实现了 8 个相同的虚拟数据平面，控制平面在软件中实现[27]. 在硬件中实现虚拟数据平面，一方面可以提供更好的隔离性，另一方面，很容易获得线速的数据包转发性能. 由于 FPGA 的资源相对有限，基于 FPGA 的可编程虚拟化路由器的扩展性成为最大的问题. 为了克服这一问题，Unnikrishnan 等人[28] 使用 FPGA 的动态重构技术以及软硬件数据平面迁移技术，解决可编程虚拟化路由器的可扩展性问题；Yin 等人[29] 提出使用 FPGA 局部动态可重构技术解决硬件虚拟数据平面的可扩展性问题，进一步降低了软硬件数据平

① OpenVZ. Available: http://wiki.openvz.org/Main_Page, 2011

② Quagga, 2011. Available: http://www.quagga.net/

面的迁移时间.借鉴 Click 模块化的思想,Anwer 等人[30]提出了模块化的硬件虚拟数据平面架构 SwitchBlade,在硬件数据平面中支持 IPv4、IPv6、OpenFlow 以及其它非 IP 协议,这种架构提高了模块的可复用性,支持新协议快速的原型设计和部署.

PEARL 是中国科学院计算技术研究所研发的高性能、可编程的虚拟路由器平台[35]①,整个系统结合了硬件的高性能和软件的高灵活性优势:采用 TCAM 实现了高性能的路由转发表,使用 FPGA 对数据包进行快速的预处理,提供硬件虚拟化支持,实现高性能数据平面;在软件上,使用操作系统级的虚拟化技术 LXC,对软件虚拟路由器进行隔离,保证各个虚拟路由器之间互不干扰,通过软硬件的灵活控制,每个虚拟机可以实现一台可编程的路由器,在系统上研发了灵活的控制接口对整个系统进行定制和管理,满足当前各种新协议对转发数据通路的不同需求.

表 2 综合比较了可编程虚拟化路由器的研究进展.从近几年的研究趋势来看,可编程虚拟化路由器的研究将会朝着两个不同的方向继续发展:一方面,随着商用高性能多队列网卡的出现以及 SR-IOV 等 I/O 虚拟化技术的发展,通用服务器平台的 I/O 性能有了质的飞跃,如何充分利用这些新兴的 I/O 加速技术和多核 CPU 的并行优势,进一步提高可编程软件路由器的性能,将是一个重要的研究方向.另一方面,FPGA、GPU、众核处理器的专用硬件平台的发展,给可编程虚拟路由器的研究与设计提供了新的思路,如何更好地发挥专用的硬件平台的特点,对可编程虚拟化路由器进行加速,也是当前的研究热点.

表 2 可编程虚拟化路由器的研究进展

具体方案	设计目标	硬件平台	主要特点	试验结果
RouteBricks	采用集群服务器实现高性能的路由器	CPU+NIC	采用商用器件,纯软件路由器,具有比较好的灵活性和可扩展性,通过对集群系统的服务器间和服务器内部的并行优化来提高性能	原型系统达到 35 Gbps 转发性能
PacketShader	采用商用硬件实现高性能路由器	CPU+GPU	NUMA 系统结构;多核 CPU、高性能网卡、GPU;采用 GPU 实现数据包处理的加速	40 Gbps 的转发性能
SwitchBlade	在可编程平台上实现新协议的试验	CPU+NetFPGA	虚拟数据平面,用户可定制硬件模块,灵活的匹配和转发引擎,采用软件方法进行例外处理,满足新协议应用需求	重点验证虚拟的数据平面,具有很好的隔离性和可扩展性
PEARL	采用可编程硬件实现虚拟化路由器	CPU+FPGA+TCAM	采用 FPGA 和 TCAM 实现高性能的路由转发,通过虚拟化硬件技术实现高性能数据平面,具有良好隔离特性	20 Gbps 线速转发能力,体现高性能和良好隔离性

4.2 未来互联网试验床

未来互联网体系结构的研究需要大规模网络试验床的支撑.一方面,试验床为 FIA 的研究人员提供了一个试验平台,各种新的架构可以在试验床上进行验证;另一方面,虽然 FIA 的设计可以从零开始,但是其验证和部署必然是演进式的.大规模试验床给未来互联网架构的演进式验证和部署提供了一条有效的途径.因此,世界上各个国家和地区在研究未来互联网架构的同时,都在积极研究和部署大规模网络试验床,如美国的 PlanetLab[36]、GENI[37]等,欧盟 FIRE[38] 项目下的 OneLab②,日本的 JGN2plus③ 和韩国的 FIRST[39]等.

PlanetLab 是一个由美国国家科学基金资助的全球范围内的覆盖网络.其目的是提供一个开放式的平台,用于开发、部署和访问新的网络服务.到目前为止,PlanetLab 已经包含 517 个地区的 1126 个节点,为全世界各地顶级学术和工业研究机构的 1000 多名研究人员提供服务,用于支持分布式存储、网络映射和 P2P 系统等方面的新技术研究.PlanetLab 的短期目标是提供一个服务验证平台,支持大量新应用的开发和部署;长期目标是通过支持各种创新的服务研究,成为未来互联网的一个缩影,最终实现一个面向服务的未来互联网体系结构.GENI 是美国自然科学基金于 2005 年启动的一个促进未来互联网革命性创新的计划,其目的是建设一个可编程的、支持虚拟化的、可联合的、基于 Slice 的大规模试验基础设施,支持互联网前沿科学与工程问题的研究.

FIRE 计划是欧盟关于未来互联网研究的一个重要计划,包括两个主要的研究内容:试验床部署和试验驱动.OneLab 是 FIRE 计划下的试验床研究项目,提供了 4 个核心试验床的访问服务,包括

① ProgrammablE virtuAl Router pLatform, PEARL, 2011. Available at http://fi.ict.ac.cn

② OneLab, 2011. Available at http://www.onelab.eu/

③ JGN2plus, 2011. Available at http://www.jgn.nict.go.jp/english/index.html

PlanetLab Europe、ETOMIC、NITOS 和 DIMES. PlanetLab Europe 是 PlanetLab 的一部分，用于开发新的网络服务，目前已为 500 多个研究人员提供服务；ETOMIC 是一个高精度的测量基础设施，分布于欧洲，能够提供高精度的主动测量服务；NITOS 是一个无线网络试验床，用于测试和评估未来的无线协议和应用；DIMES 测试床用于对现有 Internet 结构和拓扑的测试.

JGN2plus 是和日本新一代互联网研究 AKIRA 项目对应的新一代互联网试验平台项目，它是在 JGN2 试验床基础上的拓展实现，也是日本最大的网络试验床项目. 与 JGN2 相比，JGN2plus 集成了虚拟化的服务，能够提供各种的服务用于网络技术的研发和网络应用的试验. FIRST 是韩国于 2009 年 3 月启动的一个设计和部署未来互联网试验床的项目，该项目的主要内容包括：研究未来互联网平台实现的核心技术，包括可编程、虚拟化和控制架构等；开发未来互联网试验平台并部署试验床，试验平台包括小规模的基于 PC 的试验平台和中大规模的基于 ATCA 的试验平台. FIRST 的最终目标是在韩国的 KOREN① 网络上部署一个面向服务的未来互联网试验床，并与 GENI 兼容，开展广泛的国际合作.

从试验床研究的发展来看，目前已有相当多的国家和地区正在部署未来网络试验床，而且越来越多的国家和地区将会开始启动未来互联网研究计划，着手部署未来互联网研究的试验床. 在试验床的研究和部署当中，普遍采用两个关键技术[40]：虚拟化和联邦互联. 虚拟化主要解决资源共享问题，允许多个不同的试验在同一个物理网络上同时进行. 联邦互联主要为了实现大规模的、统一的测试平台，通过将不同国家和地区的多样化的网络试验床进行互连，提供统一的控制框架，为研究者提供更大规模和更加多样化的测试环境.

5 总　结

信息传递仅仅是互联网的基本功能之一，服务才是互联网价值的根本体现. 互联网承载的服务呈现出爆炸式增长态势，未来互联网实质是一种具有超大规模、虚拟化、高可靠性、通用性、高可扩展性的服务系统，支持用户在任意位置、使用任意终端方便地获取各种服务. 未来网络体系结构必须适应这些服务模式和服务需求，在服务驱动的设计理念下，未来互联网体系结构设计需要重点研究和解决以下问题：

(1) 网络模型. 网络是网络体系结构、网络设备、业务与用户相互作用的复杂系统，任何一个因素的改变都将影响其他因素的状态. 未来网络体系结构需要在动态获取多维指标与建立精确的数学模型的基础上进行探索与验证.

(2) 可扩展路由协议. 未来互联网设计中，位置和标识分离已经成为发展趋势，非 IP 寻址也已经成为共识，需要设计可扩展路由协议，解决不定长、不规则标识的路由问题，解决节点移动过程中位置和标识的映射问题.

(3) 服务本地化管理策略. 服务本地化是未来互联网中解决流量扩展问题的基本思想，需要设计合理、高效的服务本地化管理策略，解决存储决策、存储管理、服务获取、存储命中率低等问题.

(4) 安全和可靠性策略. 在未来互联网中，安全设计理念将从被动防御转向主动防御，其基本思想包括：用户和服务验证、安全传输、动态信任管理等，服务安全将作为一个整体的网络属性为最终用户感知.

(5) 未来互联网实现及试验. 在设备层面，研究网络核心设备的可编程和虚拟化关键技术. 在试验床层面，探讨复杂网络行为的纯化、简化方法以及加速、延缓和再现网络现象的实验技术，并设计互联网科学实验的标准过程，探索互联网科学实验理论，构建一个动态的、可持续发展的、大规模的未来互联网实验床. 通过未来互联网测试环境与验证技术的研究，满足新一代服务友好的未来互联网关键技术的测试与验证需求.

参 考 文 献

[1] Leiner B M, Cerf V G et al. A brief history of the Internet. ACM SIGCOMM Computer Communication Review, 2009, 39(5): 22-31

[2] Feldmann A. Internet clean-slate design: What and why?. ACM SIGCOMM Computer Communication Review, 2007, 37(3): 59-64

[3] Handley M. Why the Internet only just works. BT Technology Journal, 2006, 24(3): 119-129

[4] Rexford J, Dovrolis C. Future Internet architecture: Clean-slate versus evolutionary research. Communications of the ACM, 2010, 53(9): 36-40

① KOREN: Korea Advanced Research Network. Available at http://koren2.kr/koren/eng/

[5] Ratnasamy S, Shenker S, McCanne S. Towards an evolvable Internet architecture. ACM SIGCOMM Computer Communication Review, 2005, 35(4): 313-324

[6] Meyer D, Zhang L, Fall K. Report from the IAB workshop on routing and addressing. RFC 4984, 2007

[7] Pallis G, Vakali A. Insight and perspectives for content delivery networks. Communications of the ACM, 2006, 49(1): 101-106

[8] Gao W, Cao G. Fine-grained mobility characterization: Steady and transient state behaviors//Proceedings of the 11th ACM International Symposium on Mobile Ad Hoc Networking and Computing. Chicago, USA, 2010: 61-70

[9] Azimi N, Gupta H, Hou X, Gao J. Data preservation under spatial failures in sensor networks//Proceedings of the 11th ACM International Symposium on Mobile Ad Hoc Networking and Computing. Chicago, USA, 2010: 171-180

[10] Jacobson V. Special plenary invited short course: (CCN) Content-centric networking. Future Internet Summer School, Bremen, Germany, 2009

[11] Zhang L, Estrin D et al. Named data networking (NDN) project. PARC Technical Report NDN-0001, 2010

[12] Xie H, Yang R et al. P4P: Provider portal for applications//Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication. Seattle, USA, 2008: 351-362

[13] Karagiannis T, Papagiannaki K, Faloutsos M. BLINC: Multilevel traffic classification in the dark//Proceedings of the 2005 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. Philadelphia, USA, 2005: 229-240

[14] Farinacci D, Fuller V et al. Locator/ID separation protocol (LISP). IETF draft-ietf-lisp-09, 2010

[15] Zhu Z et al. A Survey of Mobility Support in the Internet. RFC 6301, 2011

[16] Parulkar G. Reinventing Internet infrastructure and services. Technical Report, USA: Standford University, 2011

[17] Raychaudhuri D. MobilityFirst: A robust and trustworthy mobility-centric Architecture for the Future Internet. NSF FIA Project Summary. Technical Report of MobilityFirst Project Team: NDN-0001, 2010

[18] Greenberg A, Hjalmtysson G et al. A clean slate 4D approach to network control and management. ACM SIGCOMM Computer Communication Review, 2005, 35(5): 41-54

[19] Zhang H K et al. Theory and key technologies of new generation Internet. ZTE Communications, 2008, 14(1): 17-20

[20] Wu J, Cui Y et al. 4over6 transit solution using IP encapsulation and MP-BGP extensions. RFC 5747, 2010

[21] Wu J, Bi J et al. A source address validation architecture (SAVA) testbed and deployment experience. RFC 5210, 2008

[22] Han D et al. XIA: Efficient Support for Evolvable Internetworking//Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI'12). San Jose, USA, 2012

[23] Xie G, Sun Y, Zhang Y et al. Service-oriented future Internet architecture. INFOCOM-Poster, Shanghai, China, 2011

[24] Nordstrom E et al. Serval: An End-host Stack for Service-Centric Networking//Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI'12). San Jose, USA, 2012

[25] Dobrescu M et al. RouteBricks: Exploiting parallelism to scale software routers//Proceedings of the ACM SIGOPS 22nd Symposium on Operating Systems Principles. Big Sky, USA, 2009: 15-28

[26] Han S et al. PacketShader: A GPU-accelerated software router. ACM SIGCOMM Computer Communication Review, 2010, 40(4): 195-206

[27] Anwer M B, Feamster N. Building a fast, virtualized data plane with programmable hardware. ACM SIGCOMM Computer Communication Review, 2010, 40(1): 75-82

[28] Unnikrishnan D et al. Scalable network virtualization using FPGAs//Proceedings of the 18th Annual ACM/SIGDA International Symposium on Field Programmable Gate Arrays. Monterey, USA, 2010: 219-228

[29] Yin D et al. Customizing virtual networks with partial FPGA reconfiguration//Proceedings of the 2nd ACM SIGCOMM Workshop on Virtualized Infrastructure Systems and Architectures. New Delhi, India, 2010: 57-64

[30] Anwer M B et al. SwitchBlade: A platform for rapid deployment of network protocols on programmable hardware. ACM SIGCOMM Computer Communication Review, 2010, 40(4): 183-194

[31] Barham P et al. Xen and the art of virtualization//Proceedings of the 19th ACM Symposium on Operating Systems Principles. Bolton Landing, USA, 2003: 164-177

[32] Kohler E et al. The Click modular router. ACM Transactions on Computer Systems, 2000, 18(3): 263-297

[33] Handley M et al. XORP: An open platform for network research. ACM SIGCOMM Computer Communication Review, 2003, 33(1): 53-57

[34] Lockwood J W et al. NetFPGA—An open platform for gigabit-rate network switching and routing//Proceedings of the IEEE International Conference on Microelectronic Systems Education. San Diego, USA, 2007: 160-161

[35] Xie G G, He P et al. PEARL: A programmable virtual router platform. IEEE Communication Magazine, Special Issue on Future Internet Architectures: Design and Deployment Perspectives. Communications Magazine, 2011, 49(7): 71-77

[36] Chun B et al. PlanetLab: An overlay testbed for broad-coverage services. ACM SIGCOMM Computer Communication Review, 2003, 33(3): 3-12

[37] Elliott C. GENI: Opening up new classes of experiments in global networking. IEEE Internet Computing Magazine, 2010, 14: 39-42

[38] Gavras A et al. Future internet research and experimentation: The FIRE initiative. ACM SIGCOMM Computer Communication Review, 2007, 37(3): 89-92

[39] Park M K et al. Implementation of a future Internet testbed on KOREN based on NetFPGA/OpenFlow switches//Proceedings of the NetFPGA Developers Workshop. Stanford, USA, 2009

[40] Paul S et al. Architectures for the future networks and the next generation internet: A survey. Computer Communications, 2011, 34(1): 2-42

XIE Gao-Gang, born in 1974, Ph.D., professor, Ph. D. supervisor. His research interests include future Internet architecture, programmable virtual router and network measure and modeling.

ZHANG Yu-Jun, born in 1976, Ph. D., associate professor, Ph. D. supervisor. His research interests include future Internet architecture and trusted Internet.

LI Zhen-Yu, born in 1980, Ph. D., associate professor. His research interests include future Internet architecture and P2P computing.

SUN Yi, born in 1979, Ph. D., associate professor. His research interests include future Internet architecture and mobile Internet.

XIE Ying-Ke, born in 1972, Ph. D., associate professor. His research interest is programmable virtual router.

LI Zhong-Cheng, born in 1962, Ph. D., professor, Ph. D. supervisor. His research interest is computer networks.

LIU Yun-Jie, born in 1943, Ph. D. supervisor, member of Chinese Academy of Engineering. His research interest is computer networks.

Background

The development of emerging Internet applications (e. g. Internet of Things, Mobile Computing, Cloud computing and social networks) bring great challenges to the IP address-oriented TCP/IP Internet architecture, which only aims at transferring packets. The challenges require the research on future Internet architectures and their mechanisms. This paper analyzes the current future Internet architectures, the programmable virtual router and the future Internet testbeds. ICT started the research on future Internet in 2009 and has achieved some result, including the service-oriented future Internet architecture (SOFIA) and the programmable virtual router (PEARL).