摘 要

过去的三十年中，Internet 已经从一个小型的实验性研究性的网络发展壮大为一个以路由器，交换机和主机组成的复杂网络。如今维护一个准确的网络拓扑关系对所有网络管理系统都是最基本的要求,是得到良好网络设计构架的关键。研究人员已经不仅注重 Internet 性能测量，也越来越注重对 Internet 拓扑进行探测，使得拓扑探测发展成为一个非常有挑战性的研究方向。

多年来，很多组织和研究人员关注于 Internet 自治系统级和路由器级的拓扑测量和分析;网络管理系统及其研发和使用单位往往更关注于管理域内路由器级拓扑发现技术，并要求部署的拓扑发现系统具有一定的网络拓扑监控能力。

如今已经存在很多针对 IPv4 网络的拓扑发现技术，其中基于 ICMP 的探测方式对路由器级公共网络拓扑发现最为有效,并且较小的地址空间为 IPv4 拓扑发现系统提供了良好的性能基础。IPv6 巨大地址空间给以上方法带来性能上的挑战，严重影响拓扑发现系统性能和对目标网络的覆盖，并且由于 IP 和 ICMP 协议上的变化，已有的 IPv4 网络拓扑发现技术无法直接移植到 IPv6 网络环境中。

本文介绍了一种 IPv6 拓扑发现系统的设计与实现,扩展了基于 ICMP 探测方式的适用范围。系统解决了 IPv6 路由器级拓扑探测中的实际网络环境问题，其中包括中间路由报文限制、别名解析、匿名端口、路由循环、不稳定路由等。IPv6 拓扑发现系统性能的提高和对目标网络的覆盖也是本文的关注重要点，系统提供了种子节点，构造地址和 IPv6 地址管理模块作为目标地址集合来源。Web Services 技术的应用提高了 IPv6 拓扑发现系统部署、配置和管理的灵活性。

本文系统对中国移动 CNGI 骨干网和 CERNET2 进行了实际拓扑，在此基础上对获取的 IPv6 拓扑数据进行了分析和总结。

关键词：网络探测，拓扑发现,ICMPv6,IPv6，网络测量

Design and Implementation of Topology Discovery System for IPv6 Networks

Chen Hanlin

Directed By Zhang Guoqing

In just three decades, the Internet has grown from a small experimental research network into a complex network of routers, switches, and hosts. Now maintaining an accurate map of the network topology is one of the basic requirements of any management solution and is essential to the procurement of good architectural design decisions. Researchers have paid much attention not only to Internet performance measurement but also to Internet topology measurement that has hence been growing into a novel and challenging research area.

The topology measurement and analysis of Internet AS-Level and Router-Level are concerned by many organizations and researchers in these years. But network management systems and developing-deploying departments always pay more attentions on the technologies of intre-domain topology discovery and require some scout function of the deployed system.

A number of techniques for IPv4 network topology already exist. Of these ICMP-based probing has shown to be most useful in determining router-level topologies of public networks and IPv4's limited address space offers a good base for high performance of IPv4 topology discovery systems. Because of the IPv6's huge address space, the system confronts challenges of the performance and covering on the target networks. Due to the changes in protocols of IP and ICMP, the existed techniques for IPv4 network topology discovery can't be directly ported to IPv6 networks.

Design and implementation of a topology discovery System for IPv6 networks is introduced, which enlarges the applicability of ICMP-based probing. This system overcomes the key issues in router-level topology discovery, including intermittent router limitation, alias resolution, anonymous interfaces, routing loop, instable routing and so on. The performance and networks' covering of topology discovery system are also our key issues. The system supports seeds, IP constructing and IPv6 address space management as the resource of probing IP and Using Web Services promotes flexibilities on deployment, configuration and management.

IPv6 networks such as Cernet2 and China Mobile's backbone of CNGI is probed. At the end the topology date collented from them will be shown and analysed.

Keywords: Network probing, topology discovery, ICMPv6, IPv6, network measurement.

图目录

表目录

声　　明

我声明本论文是我本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，本论文中不包含其他人已经发表或撰写过的研究成果。

作者签名：陈韩林　　日期：2007-4-15

论文版权使用授权书

本人授权中国科学院计算技术研究所可以保留并向国家有关部门或机构送交本论文的复印件和电子文档，允许本论文被查阅和借阅，可以将本论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编本论文。

（保密论文在解密后适用本授权书。）

作者签名：陈韩林　导师签名：张国清　日期：2007-4-15

第一章 引言

随着 IPv6 技术的日益成熟，许多国家都在下一代互联网的建设方面投入大量资金，构筑本国的 IPv6 试验网。根据以往网络的经验，随着 IPv6 网络的快速扩张、接入的随机性和业务的多样性使网络速度、容量以及底层拓扑结构都发生了巨大变化。随着 IPv6 网络重要性的日益提高和网络结构的日益复杂，了解其结构和特性对于 IPv6 网络的应用、扩展、优化、增强网络安全性等方面能够发挥重要作用。

互联网的高速发展早已超越了互联互通的基本要求，加速了高端网络应用的不断提出与普及。当前互联网正面临来自本身技术，管理运营方面的挑战，诸如网络的可扩展性（IPv4 网络的地址空间问题，路由表爆炸问题等），网络系统的维护和安全，端到端的高性能通信和服务质量等问题。在现有框架中，对上述问题的解决主要依赖于在现有协议基础上的不断修补，无法在统一环境中得到完善的解决。籍此 IPv6 实验网正在全球范围扩展，而获得一个精确的网络拓扑关系图是任何一个网络管理解决方案的基本要求。拓扑信息对网络管理的其他领域（故障管理，配置管理，用户管理，性能与安全等等）有着相当重要的作用。但由于 IPv6 协议体系与 IPv4 的巨大差别，给 IPv6 网络的自动拓扑发现带来了新的困难与挑战。

1.1 网络管理介绍

广义上的网络管理是对资源的管理，是指调度和协调资源，以便在所有时间都能使计划、营运、管理、分析、评估、设计和扩充网络以合理的成本和最佳的能力满足服务等级的目标。网络管理包括运行、管理、维护和供给功能，这些功能提供了管理网络资源的有效方法。

而从狭义上入手，网络管理是指对网络状态进行监控，当网络出现故障时能及时做出报告和处理，调整网络资源，使网络能正常、高效地运行。一般而言，网络管理有五大功能：配置管理、性能管理、故障管理、安全管理、计费管理。这五大功能是保证一个网络系统能够正常运行的基本功能集合。

配置管理可细分为拓扑管理和参数管理。拓扑管理是指自动发现网络内的所有设备以及设备之间的连接情况，对发现的结果能够自动更新以保证和实际网络的一致性。能够以图形化的方式表现网络拓扑结构、网络状态、设备信息。参数管理是保证设备配置信息的完整性，对配置信息的正确性检查，配置数据的查询与统计。

性能管理主要提供性能监测功能、性能分析功能。性能监测功能是对网络流量、设

备资源使用情况（如 CPU 占有率、内存占有率等）等性能数据进行连续地采集。性能分析功能主要是根据监测到的性能数据进行统计和计算，获得网络及其主要成分的性能指标，定期或在必要时生成性能报表和图表。

故障管理的目的是迅速发现和纠正网络故障，动态维护网络的有效性。故障管理的主要功能包括告警监视、故障定位、告警过滤。告警监视用来监视网络设备出现的故障。故障定位用来确定故障产生的位置或者产生故障的设备。告警过滤用来过滤大量告警中不重要的信息从而突出网络的故障所在；当网络中的某个地方出现故障以后，往往会引发很多告警信息，这就需要通过告警过滤从这些告警信息中找出根本问题。

安全管理的目的是提供信息的隐私、认证和完整性保护机制，使网络中的服务、数据以及系统免受侵扰和破坏。计费管理主要是正确的计算和收取用户使用网络服务的费用，同时进行网络资源利用率的统计和网络的成本效益核算。

在以上管理功能中，配置管理是整个网络管理的基础，因为它提供了网络管理所需的基础数据，也是网络管理的基本对象的集合。而其中，网络拓扑信息以及拓扑节点信息是配置数据中的最基本数据。因此网络拓扑以及节点的数据获取和更新成为网络配置管理的基础功能，也是整个网络管理的基础功能，而这一过程即网络拓扑发现和拓扑更新过程。

面对一个大规模的网络，采用人工方式进行网络拓扑管理工作量太大，同时准确性无法保证，这就需要网管系统能够自动发现网络的拓扑结构，并以图形化的方式呈现。网络拓扑管理是网管系统的基础，能够为性能、告警、配置数据的处理提供支持，这些数据的监测和处理需要在已知网络拓扑的基础上进行，从而全面、动态地反映网络的运行状况，为用户监视整个网络提供强有力手段。

网络拓扑自动发现系统就是要识别网络中各种类型的设备、设备之间的连接关系，并获取各类型设备的配置信息。目前最常用的方法包括利用 SNMP(Simple Network Management Protocol，简单网络管理协议)、IP 协议族命令和系统探测技术等手段发现给定范围内的网络设备及它们之间的互连关系。拓扑自动发现的结果通常存放于数据库中供网络管理系统的其它功能模块使用。

1.2 Internet 拓扑测量相关工作

现实的需要推动了 Internet 拓扑测量研究领域的形成和发展。对 Internet 拓扑结构进行动态描述具体有以下几个方面的应用：

（1）Internet 拓扑测量可以辅助宏观网络发展布局和网络管理。例如确定在哪里增加新路由器、网络扩容等。通过了解网络地理分布状况，既可以从宏观层次管理规划

大范围内的网络发展布局，也可以生产隐含地理模型的综合拓扑【33】，为仿真模拟、网络管理等服务。了解网络拓扑结构等信息是进行有效网络管理的基础之一。虽然测量得到的 Internet 拓扑不是物理拓扑，但它反映了实际的路由拓扑情况，这对于网络可达性研究、网络存活性分析、优化网络配置等具有重要的参考意义。

（2）Internet 拓扑测量能够为仿真模拟 Internet 环境【32】、协议设计与评价提供研究基础【34】【35】。只有在类似于 Internet 的拓扑上进行仿真模拟研究，其结果才具有现实可用性。只有通过对实际 Internet 拓扑结构进行分析，得到相应的特征参数，指导构造更符合实际网络的拓扑生成器【36】【37】，才有可能进一步对网络进行准确建模。

（3）拓扑测量选路拓扑结构，可以分析研究 Internet 选路的动态性质（例如发现迂回路由、转发环路、路由黑洞、中间的连通性发生变化的路由、无规则动态路由等病态路由）和路由配置策略，研究路由的收敛性质【38】【39】【40】或用于域间路由错误管理【41】，为进一步选路、升级改进设计利用拓扑性质的更有效的路由协议、提高网络路由性能提供可行性。

（4）Internet 拓扑测量可以为与拓扑机构相关的协议和算法的性能改进提供依据。此外，拓扑结构信息还有助于帮助选择多镜像服务器的位置、帮助 ISP 确定与那个 AS 相连能够具有更好的 Internet 连通性【32】【42】【43】等

（5）以拓扑结构信息为基础，结合性能测量，有助于准确定位并实施故障隔离（fault isolation）【44】以及为遏制蠕虫病毒和防范大规模网络攻击提供研究平台和预警手段，从而能够对整个网络更具宏观控制力。并且实际上，控制蠕虫扩散范围和阻断 DDoS 攻击本身是一种故障隔离，也是拓扑测量在网络安全研究中的应用。

（6）Internet 拓扑测量技术可用来观察灾难发生对网络连通性的影响【30】。例如 Cheswick 等人所进行的 Internet Mapping 项目【31】采用单点测量，曾成功描述了科索沃战争期间，由于网络设施或电力供应设施被破坏而使南斯拉夫地区的网络拓扑结构发生显著变化的情况（1999 年 5 月）。这从一个侧面反映了轰炸效果，也是网络拓扑测量在军事领域中应用的典范。路由器拓扑，是网络抗攻击能力【45】和可生存性【46】研究的基础之一。

（7）Internet 拓扑测量可以为 Internet 流量工程和网络行为学研究提供基础辅助依据。

随着 IPv6 网络近年来的广泛部属，对于 IPv6 网络进行拓扑发现的研究也开始兴起。IPv6 协议体系的改变使得传统的 IPv4 网络拓扑发现方法不能直接移植到 IPv6 网络中来，同时 IPv6 协议的许多新特征，如地址结构的变化以及 IPv4-IPv6 隧道的存在，也对 IPv6 网络的拓扑发现提出了全新的挑战。在国际上，CAIDA 组织主要关注全球 IPv6 网络信息的采集和拓扑发现【1】。它的工作重点在于对全球范围的 IPv6 网络发展状况进行检测和分析。在相关文献中， 贝尔实验室的 Daniel G 等人提出了一种基于源路由的拓扑发现思想，对目前的 6Bone 网络进行了大规模的探测，并对 IPv6 匿名路由器的问题提

出了解决方案【2】。法国 LORIA 实验室的 Astic I 等人提出了一种基于分级结构的拓扑发现思想【3】。文献【18】给出了 IPv6 和 IPv4 混合网络中隧道的发现方法。

1.3 拓扑发现工具及其分析

1. Ping

Ping 命令是 IP 网上最古老的一种工具，用来监测网络节点是否活着，或用于监测到网络节点间的往返时延（RTT）。通常 Ping 只涉及网络上的源和目的两节点，而忽略网络细节。另外我们可以使用广播 Ping，其 Ping 的地址不是一个单一的地址，而是子网的广播地址，所有位于该子网的主机均对此 Ping 包进行响应，从而一次就可得到子网内的全部活动主机。

2. Traceroute

Traceroute 命令是 TCP/IP 家族内另一个比较早的工具，它可用来发现测试点和目标主机之间的路由器。路由器在转发包之前总是将其 TTL 值减 1，如果 TTL 降为 0，则路由器向源地址发送 TTL-Expired ICMP 消息。Traceroute 实现的原理就是应用路由器的这个特性，通过发送 TTL 逐渐增大的探测包，由测试点到目标间这条路经上所有的路由器依次向测试点发送 TTL-Expired ICMP 包，从而发现所有路由器。因为几乎所有的路由器设计时都实现了发送 TTL-Expired ICMP 消息的功能，所以大多数情况下 Traceroute 的结果是准确可信的。由于采用逐渐增大 TTL 值的方法，每探测一个目标需要依次发送不同 TTL 值的多个包，因此用 Traceroute 获取结果比 Ping 要慢的多。可以设计一种并发式的 Traceroute 命令，一次发送不同 TTL 值的多个包，从而加速路由器的发现速度。

3. DNS

IP 地址是为网络上的路由器或主机等机器设计的，它不符合人类的记忆习惯，DNS（Domain Name System）就是为了解决这个问题而开发的。DNS 系统主要用于网络设备 IP 地址到名字的映射，同时也维护一些其他信息如设备的硬件平台及操作系统等。

4. SNMP

SNMP（简单网络管理协议）的基本思想是所有的网络设备维护一个 MIB（管理信息库）保存其所有运行进程的相关信息，并对管理工作站的查询进行响应。SNMP 协议描述了一种从 MIB 库中获取信息的方法，对设备唯一的要求是支持 SNMP 并且 MIB 中的信息足够丰富。

5. 其它工具或技术

除了上面介绍的几种常用工具外，我们还可利用节点的 ARP 表查询它直连的设备，利用路由协议（如 OSPF、BGP）发现所有子网或网络，发现所有的路 由器，在 BGP 下还可发现一条路径经过的自治域（Autonomous Systems）。对于非 IP 网络，

可利用专门的技术（对 IPX 网络可采用 SAP）发现网络拓扑信息。此外，一些厂家专有的技术如 Cisco 的 CDP（思 科发现协议，仅用于 Cisco 设备）、Netflow 技术等也可用于拓扑发现。

互联网拓扑发现的研究由来已久，按照被发现实体的粒度不同，可以分为三类：旨在发现自治系统间互连关系和商业关系的 AS 层次的拓扑发现，旨在发现路由器间连接关系的路由器级拓扑发现和旨在发现局域网内物理设备（包括路由器，交换机，Hub 和主机）之间互连关系的物理网络拓扑发现【7，50】。本文主要研究的是路由器级的拓扑发现，下面将对以上相关拓扑发现工具进行讨论。

1. Ping 工具分析

使用 Ping 的最大问题是，当 Ping 一个活着的主机时，其往返时延往往在几十毫秒左右，但 Ping 一个不存在的或宕着的主机，一般比较常用的超时通常为 20 秒，再加上为了减少丢包对测量结果的影响而采取发 2～3 个 Ping 包，这样对这类主机的监测代价就非常大。这个问题最直接的解决方案是减少超时值，但是必须注意不要小于网络实际的往返时延。通过精心设计超时和重发策略（随着跳数的增多，超时相应增大），可以有效减少等待时间同时又减少误判。而使用广播 Ping 的问题是，现在实际网络中广播 Ping 很少得到完全支持，部分网络由路由器代替子网内的主机响应。在另外一些网络中主机根本就不对广 播 Ping 进行响应，甚至路由器根本不转发能引起广播的包。这是基于网络安全的考虑，因为可以利用这个特性进行拒绝服务攻击，例如向几个大的子网进行广播 Ping，并把源地址设置为受害者的地址，这样受害者就会淹没于大量 ICMP Ping 的响应包，从而拒绝提供任何服务。对该问题的一个解决方案是设计一个专门的 Broadcast Ping 程序，其内部实现是直接将子网的广播地址转变为多个主机地址，然后启动多个线程或进程来分别向主机发送 Ping 包，从而获取子网内的全部主机地址。

IPv6 拥有巨大网址空间，协议地址空间由 IPv4 的 32 位扩大到 128 位，2 的 128 次方形成了一个巨大的地址空间。即使 64 位前缀子网的地址空间理论上允许地址相当与现有 IPv4 地址空间的总和。使用 Ping 命令来监测网络节点是否存活显得不可行。使用 IPv6 组播技术可以解决以上问题，组播是否被支持是个不确定因素，而且即使可行此方法只能得到网络节点的状态，无法进一步提供网络节点间的链接信息。另一方面组播技术的应用限制了拓扑发现的应用的范围，无法适用公共网络的拓扑发现。因此我们在 IPv6 拓扑发现系统放弃 Ping 工具的使用。

2. DNS 工具分析

使用 DNS 服务器提供的区域传输功能可以一次获取域内许多主机和路由器，快捷方便，这是它的优点。但如果主机的地址通过 DHCP 获得，则 DNS 对此就无能为力，此外，DNS 服务器提供的信息可能与实际情况不一致，甚至有

些 DNS 服务器没有提供区域传输功能。尽管有诸多缺点，DNS 在拓扑发现中还是很重要的，我们可以把 DNS 返回的信息作为其他算法的起点；我们还可以在不知道网络具体结构的情况下，使用不同时间返回来的信息直接用来估算网络的增长速度。

因为 DNS 工具较低的准确性和有限的使用范围，因此也未被采用。

3. SNMP 工具

使用 SNMP 的最大优点是信息自动随网络的状况更新，这样通过 SNMP 获取的拓扑信息总是反映网络最新的状况。其缺点是并不是所有设备都支持 SNMP 协议，而且除了标准的 MIB 信息外，各厂家都为自己的设备开发了专门的 MIB，如果在拓扑自动发现程序中使用了这些 MIB，其处理上可能不得不随厂家的不 同而作特殊的处理。拓扑发现中用到的 MIB 组有 System 组、Interfaces 组、IP 组，它们均为当前 MIB- II 下的标准组。

如今 IPv6 网络下 SNMP 的标准化工作尚未完成，相对于 IPv4，目前支持 IPv 的 MIB 库和库中被管理对象都还很少，特别是在最重要的 RFC2465 中规定的很多字段访问还有困难，以上问题给管理域内拓扑发现系统造成较大困难【14】。

如果拓扑发现的执行者是整个待发现网络的管理者并拥有相应路由器的管理权限，则可以应用 SNMP 来进行拓扑发现，这种方法简单易行且准确。但当该前提无法保证时，将导致拓扑信息的不完整。SNMP 协议本身的特点限制了其使用范围，使其无法作为一种通用的广范围的拓扑发现系统。因此本系统没有采用 SNMP 作为主要的探测手段。

4. 路由协议工具分析

网络拓扑发现可以使用域间的路由协议 BGP 和域内路由协议 OSFP 关注于不同的网络拓扑发现层次，但是这就要求对不同网络拓扑发现层次的关注需要在不同路由协议间进行切换。路由信息的获取需要事先获取相应路由器的地址和管理权限，单个路由器只能描述部分网络，因此不完备的路由器列表也将导致拓扑信息的不完整。另一方面对于不支持以上路由协议的网络拓扑发现系统将遇到极大的挑战。

虽然本系统没有采用路由协议作为探测手段，但我们对此将进一步关注，因为路由协议为特定范围网络提供了准确的拓扑关系描述，并有很好的实时性为网络拓扑监控提供了良好的基础。

5. Traceroute 工具分析

Traceroute 应用程序是 V. Jacobson 于 1988 年开发的【27】，其初衷是用于观察端到端的路由连接状况和故障位置。目前，基于 Traceroute 机制采集选路信息是路由器级拓扑测量的主要手段。Traceroute 已被广泛用于检测和诊断路由问题（例如路由循环、不稳定路由、匿名接口等）、刻画端到端的路由行为和特征、网络拓扑发现等。虽然 Traceroute 存在一些众所周知的不足之处，可能会影

响其探测的效率和准确性，例如第三方地址问题（匿名端口）、探测过程中可能发生的路由变化使测量不准确（不稳定路由）、探测过程产生的额外负载可能会影响网络性能、中间路由器可能使用 ICMP 报文转发限制等，但它是目前唯一的、不需要从每个管理域获取专有路由信息的、可有效地观察报文如何在网络中流动的方法【25】。

于前四种拓扑发现手段比较，Traceroute 工具是唯一通用、准确（其准确性依赖对实际网络环境问题的解决）和输入要求最少的拓扑发现手段。本文介绍的 IPv6 拓扑发现系统将以 Traceroute 作为主要的拓扑发现工具。

在传统的 IPv4 网络中，已经有许多路由器级拓扑测量的研究【1、4、6、16、17】。研究的重点主要在于测量目标地址的选择和测量点的覆盖对于测量完整性的影响，路由器的别名解析问题、路由循环问题和路由不稳定性对于测量正确性的影响。1996 年，J. Richard 提出了利用 Traceroute 来反映网络拓扑结构的设想【28】。此后，人们从不同角度开展了 Internet 级拓扑测量的研究。Internet 路由器级拓扑测量也经历了从单点测量到多点测量的发展过程。

拓扑发现技术的关键点在于保证最终拓扑信息的正确性和完整性，并提高拓扑探测的性能。拓扑信息的正确性是指最终的拓扑关系数据反映的节点和链路状况在对应的实际网络中真实存在；拓扑信息的完整性是指实际网络中的节点和链路状况在最终的拓扑关系数据中得到表达；拓扑发现性能关注于探测过程产生的额外负载所影响的网络性能和拓扑发现过程所需要的时间。

因为 Traceroute 工具自身的不足，如今其应用主要停留在对公共网络和管理域间的拓扑发现和测量。本文系统将关注以上拓扑发现技术关键点，解决最终拓扑信息的正确性、完整性，提高拓扑探测性能，以此扩展基于 ICMP 拓扑发现系统的适用范围。

1.4 Web Servicess 介绍

使用 Web Servicess 技术，应用程序可以通过与平台和编程语言无关的方式相互通信。Web Servicess 是一个软件接口，它描述了一组可以在网络上通过标准化的 XML 消息传递访问的操作。它使用基于 XML 语言的协议来描述要执行的操作或者要与另一个 Web Servicess 交换的数据。在面向服务的体系结构（Service-Oriented Architecture，SOA）中，一组以这种方式交互的 Web Servicess 定义了特定的 Web Servicess 应用程序。

软件业最终会接受这样的事实：跨多个操作系统、编程语言和硬件平台集成软件应用程序不可能由任何一种专门的环境来解决。传统上，这个问题一直是一个紧耦合问题，调用远程网络的应用程序通过自己发出的函数调用和请求的参数与远程网络紧密地联系在一起。在 Web Servicess 出现之前，在大多数系统上，采用的是固定的接口，但对于不断变化的环境或需求，这样做缺乏灵活性或适用性

Web Servicess 所使用的 XML 可以用真正与平台无关的方式来描述任何（所有）数据，以跨系统交换数据，因此转向了松耦合应用程序。而且，Web Servicess 可以在较抽象的层面上工作，较抽象层面可以按照需要动态地重新评估、修改或处理数据类型。所以，从技术层面上讲，Web Servicess 可以更方便地处理数据，并且允许软件更自由地进行通信。

从更高的概念层面上讲，我们可以将 Web Servicess 视为一些工作单元，每个单元处理特定的功能任务。再往上一步，可以将这些任务组合成面向业务的任务，以处理特定的业务操作任务，从而使非技术人员可以考虑一些应用程序，这些应用程序能够在 Web Servicess 应用程序工作流中一起处理业务问题。因此，一旦由技术人员设计并构建好 Web Servicess 之后，业务流程架构师就可以聚集这些 Web Servicess 来解决业务层面上的问题。这里借用汽车引擎来作类比，业务流程架构师考虑将整个汽车引擎与汽车框架、车身、变速器和其他系统组合在一起，而不是研究每个引擎内的各个部件。而且，动态平台意味着引擎可以与其他汽车制造商的变速器或部件一起工作。

最后一个方面是，Web Servicess 有助于在组织内的业务人员和技术人员之间架起一座桥梁。 Web Servicess 使业务人员更容易理解一些技术上的操作。业务人员可以描述一些事件和活动，然后技术人员可以将这些事件和活动与相应的服务相关联。

有了通用定义的接口和设计良好的任务，重用这些任务就变得更容易了，因而重用这些任务所代表的应用程序也就变得容易了。应用程序软件的可重用性意味着在软件上的投资有了更好的回报，因为可以从同一资源产生更多收益。可重用性使业务人员可以考虑以一种新的方式来使用现有的应用程序，或者以一种新的方式将应用程序提供给合作伙伴，因此可能增加合作伙伴间的业务交易。

Web Services 是在 Internet 上进行分布式计算的基本构造块。开放的标准以及对用户和应用程序之间的通信和协作的关注产生了这样一种环境，在这种环境下，Web Services 成为应用程序集成的平台。应用程序是通过使用多个不同来源 Web Services 构造而成的，这些服务相互协同工作，而不管它们位于何处或者如何实现。以 Web Services 方式提供现有应用程序，可以构建新的、更强大的应用程序，并利用 Web Services 作为构造块。

1.5 本文的贡献

IPv6 巨大地址空间给以上方法带来性能上的挑战，严重影响拓扑发现系统性能和对目标网络的覆盖，并且由于 IP 和 ICMP 协议上的变化，已有的 IPv4 网络拓扑发现技术无法直接移植到 IPv6 网络环境中。本文介绍了一种 IPv6 拓扑发现系统的设计与实现，扩展了基于 ICMP 探测方式的适用范围。系统解决了 IPv6 路由器级拓扑探测中的实际网络环境问题，其中包括中间路由报文限制、别名解析、匿名端口、路由循环、不稳定路由等。IPv6 拓扑发现系统性能的提高和对目标网络的覆盖也是本文的关注重要点，系统提供了

种子节点，构造地址和IPv6地址管理模块作为目标地址集合来源。Web Servicess技术的应用提高了IPv6拓扑发现系统部署、配置和管理的灵活性。

本文主要的贡献包括：

1. 实际IPv6网络环境问题的解决：
   a) 散列地给出目标地址，避免短时间内对同一路由器发送过多的探测报文，并结合小型的自适应过程，避免中间路由报文限制问题。
   b) 利用匿名端口合并方法，避免匿名端口膨胀导致最终拓扑结果的不准确性
   c) 利用UDP端口不可达报文并结合源路由技术解决路由器多址问题
   d) 解决复杂网络问题同时发生时（比如路由循环和中间路由报文限制）的路由循环问题
   e) 在探测和确定路径过程中加入已探明链路，结合IP层源路由选项迫使探测报文经过指定已知节点。利用此方法收集不稳定路由路径信息，丰富最终拓扑关系数据。

   本文的拓扑发现系统对主要网络现象的处理方法有效得减少了误差，保证了最终拓扑数据的准确性。
2. 扩大探测目标地址来源，扩展了IPv6拓扑发现系统的应用

   本文系统支持多种探测目标地址来源。不仅支持典型种子节点列表（seeds list）和根据地址空间前缀构造探测地址，而且结合IPv6地址管理模块作为目标地址集合来源。将IPv6地址空间分配模块作为IPv6拓扑发现系统可靠探测目标地址获取的途径：从中得到已分配的地址空间前缀进行目标地址构造，并对未使用的地址空间根据IPv6地址聚会和现有子网地址前缀长度采取一定随机性构造目标地址来确定所探测的空白地址空间是否在实际网络中被使用。IPv6地址空间分配作为探测目标地址获取途径，为基于ICMP拓扑发现系统应用于管理域内提供了基础，提高了对目标网络的探测效率和起到监测管理域网络拓扑的作用。
3. 充分利用了IPv6源路由选项：

   系统将源路由选项应用在路径探测和别名处理过程中。在实际探测过程中对符合特定前缀的IPv6地址利用源路由机制进行路径探测，比如针对移动CNGI网络我们制定的是2001:e80:ffff:,针对Cernet2制定的为2001:da8:1::；在别名处理过程中，用带有源路由选项的ICMPv6探测报文确定路由设备报文处理模型，并用带有源路由选项的UDP探测报文确认路由别名现象等。

   充分利用IPv6源路由选项为我们在IPv6网络环境下解决路由别名和不稳定路由提供了基础，提高了最终拓扑数据的准确性和对目标网络的覆盖度。
4. 基于Web Services构建了IPv6自动拓扑发现系统

   将Web Services作为通讯的基础，解析了拓扑发现系统的三个主要模块，降低了探测、数据获取和管理三者间的耦合度。

拓扑发现系统平台可使用系统内每个基本单位的探测节点，利用本单位和次级单位的探测节点构成对本层网络的多点或并行探测，通过合理的布置并行探测点，可以大大提高探测效率和准确性； Web Services 技术的应用，为拓扑数据获取提供了统一的接口并解决拓扑关系层次属性的表达；整个系统有较强的适应性，满足了不同网络环境对管理的需求。例如，对拓扑关系的表示完全可以脱离探测和拓扑管理平台，而只采用数据获取模块；或脱离数据获取或拓扑管理平台，而只为其他平台提供探测点等等。

本文系统针对中国移动 CNGI 和 Cernet2 进行实际的拓扑探测。实验数据表明：系统很好得解决了主要网络问题，确保了拓扑发现的准确性，特别是基于 UDP 端口不可达报文的方法和源路由机制在解决 IPv6 路由别名的问题中得到了很好的效果；地址确认步骤的必要性：其剔除了网络中的伪地址，保证了节点相对探测点距离的准确性和为 IPv6 路由多址的确认提供了方便；对符合（骨干网）前缀的已探明地址使用源路由进行链路探测，可以避免不必要数据的获得，提高链路探测效率；路由器级拓扑发现探测并不能准确反映实际端口之间真实的链接对应关系，只能反响网络中路由设备的 IP 地址及路由设备之间的链接关系。

1.6 论文的组织

文章组织结构如下：

第一章主要介绍拓扑发现相关工作。从网络管理和拓扑管理的作用开始，在拓扑测量相关工作中，介绍了 Internet 拓扑结构动态描述的应用以及目前拓扑发现领域相关组织和个人的研究成果；接着介绍了常用的拓扑发现手段：ping、Traceroute、DNS、SNMP、OSFP 和 BGP 等等，并分别对其进行分析。最后对 Web Serviec 技术进行了介绍。在本章第五小节中列举了本文的相关贡献。

第二章主要探讨路由器级网络和Traceroute工具。首先，对系统主要使用的Traceroute工具的探测原理进行具体的描述；第二节分析现有主要的路由器级拓扑测量方式；第三节中，提出了提高 Traceroute 探测性能的方法；接着描述了拓扑发现中单探测点及其“cross link”问题；最后介绍了种子列表并说明了种子列表探测过程。

第三章我们详细阐述了 IPv6 网络拓扑发现的关键技术，对影响拓扑发现结果正确性的诸多因素进行了讨论，其中第一节为中间路由报文限制，第二节为路由循环，第三节为别名解析，第四节为匿名端口，第五节为不稳定路由。相应的章节中都给出了系统对具体问题的解决方法。

第四章，本章首先介绍了典型的几种目标地址获取方式；接着，我们介绍了将 IPv6 地址空间分配模块和 IPv6 拓扑发现系统相结合，IPv6 拓扑发现系统从 IPv6 地址空间分配模块中得到已记录的地址空间前缀进行目标地址构造，并对未使用的地址空间，根据 IPv6 地址聚会和现有子网地址前缀长度采取一定随机性构造目标地址来确定所探测的空白地

址空间是否在实际网络中被使用；第三节我们介绍了 IPv6 拓扑发现系统的设计：系统将 Web Services 作为模块间通讯的基础，解析了拓扑发现系统的三个主要模块，降低了探测、数据获取和管理三者间的耦合度。第四节结合现有 IPv6 网络的特点提出了多点并行探测的方式：系统将整个探测任务发送给不同探测节点同时进行，因此可以成倍数得提高探测并行数，并对可能遇到的疑问做了分析；最后介绍了利用源路由机制的路径探测并特别说明了地址确认的必要性。

第五章主要分析了具体的实验数据，介绍本系统对中国移动 CNGI 和 Cernet2 网络的实际测量结果并作相应的分析；最后对实验数据进行了总结。

最后一章进行总结和介绍进一步的工作。

第二章 路由器级拓扑和 Traceroute 工具

目前，基于 Traceroute 机制采集选路信息是路由器级拓扑测量的主要手段。Traceroute 已被广泛用于检测和诊断路由问题（例如路由循环、不稳定路由、匿名接口等）、刻画端到端的路由行为和特征、网络拓扑发现等。虽然 Traceroute 存在一些众所周知的不足之处，可能会影响其探测的效率和准确性，例如第三方地址问题（匿名端口）、探测过程中可能发生的路由变化使测量不准确（不稳定路由）、探测过程产生的额外负载可能会影响网络性能、中间路由器可能使用 ICMP 报文转发限制、探测报文容易被路由器或防火墙过滤而失效等等【24】，但它是目前唯一的、不需要从每个管理域获取专有路由信息的、可有效地观察报文如何在网络中流动的方法【25】。

本章以下部分组织如下：第 1 节首先描述基于 Traceroute 的探测原理，详细解释了基于 Traceroute 机制采集选路信息；第 2 节介绍现有主要的路由器级拓扑探测方式，探测目标的构造等问题；第 3 节介绍 Traceroute 工具性能提高的方法；第 4 节单点探测和“cross link”问题；第 5 节介绍种子列表并说明了种子列表探测过程。

2.1 Traceroute 探测原理

我们知道每个 IPv6 报文头都有一个 Hop 域（IPv4 为 TTL），报文每经过一台路由器时，该路由器都要将 Hop 的值减 1，并检查 Hop 值是否有效，如果 Hop 为 1，路由器就丢弃该报文，如果该报文不是 ICMPv6 出错报告报文，则向该报文的源地址发送一个表明“超时”的 ICMPv6 报文。

假设目标 D 与源点 S 间隔 10 跳（hop），目标 D 运行 TCP/IP 协议，以采用 UDP 高端口探测报文为例，Traceroute 测量从源点 S 到目标 D 的转发路径的过程是：从 Hop=1 开始，依次向目标 D 发送 Hop 值渐增（每次加 1）的探测报文。到 Hop<=9 时，对每个 Hop 值，探测报文没有到达目标时中间路由器就发现其 Hop 无效，并向源点 S 发送一个 ICMPv6 超时报文。当 Hop=10 时，探测报文将到达目标 D，由于探测报文设置是实际并不存在 UDP 服务的端口，因此目标 D 将发出一个 ICMPv6 端口不可达报文。从这些依次返回的“超时”报文可以得到探测报文在网络经过的路径，最后的“端口不可达”可以判断是否到达目标 D。这么就得到了从源点 S 到目标 D 的一个转发路径。这里特别说明，如果到达目的地但目标 D 没有运行相应的协议，则可能发回一个表明“协议不可达”的 ICMP 报文。如果发送的探测报文是 ICMP 回响请求（echo request）报文，则当 Hop<10 时仍将得到超时报文，目标 D 将得到回响应答（echo reply）报文。基于 Traceroute 的探测机制如图一所示：

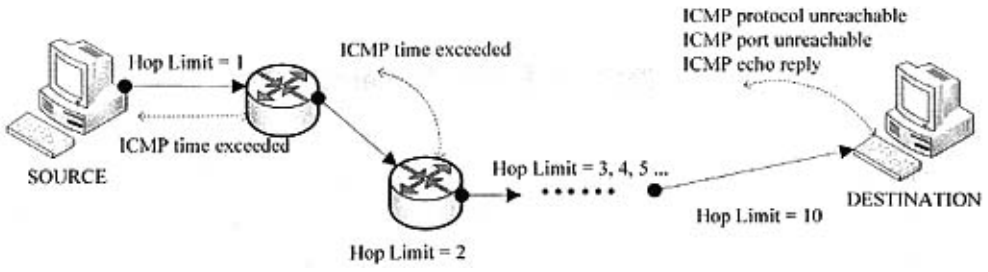


图 1　基于 Traceroute 的拓扑探测机制

实际上，无论发送什么类型的 hop（TTL）渐增探测报文，除了最后到达目的地时收到的响应报文类型可能有所区别外，从中间的路由器可能收到的都只是 ICMP 的超时报文。路由器级拓扑探测有两个显著特点。一方面，由于探测目标地址的选取具有随机性，实际到达目标的探测路径比例可能较低，例如对下一代中国教育和科研网（CERNET2， the China Education and Research Network 2）的拓扑测量实验中到达目标的探测路径比例就很低（参见第六章实验数据分析）。另一方面，路由器级拓扑探测的目的不是探测所有网络的选路拓扑结构，而主要是探测骨干网及其延伸部分，对于用户网络一般不在探测范围之内。因此，即使某些类型的探测报文到达目标而没有任何响应时，只要收到中间的路由器的超时 ICMP 报文，一般不影响对所测量范围的拓扑结构的观察。

2.2　路由器级拓扑测量

在 Internet 路由器级拓扑图中，一个节点代表一台路由器，节点之间额度边表示路由器之间的逻辑连接关系，一般表示路由器之间在网络层一跳（hop）可达【26】。通常采用无向图表示路由器拓扑图。目前，Internet 路由器级拓扑测量采集选路信息主要基于 Traceroute 机制。

Traceroute 应用程序是 V. Jacobson 于 1988 年开发的【27】，其初衷是用于观察端到端的路由连接状况和故障位置。1996 年，J. Richard 提出了利用 Traceroute 来反映网络拓扑结构的设想【28】。此后，人们从不同角度开展了 Internet 级拓扑测量的研究。

Internet 路由器级拓扑测量经历了从单点测量到多点测量的发展过程。以往的 Internet 路由器拓扑测量研究按照探测引擎的构成方式可分为以下三类：

1. 辅之以源路由机制的单点测量

主要包括 Pansiot 等人【5】的研究、Cheswick 等人【29】【30】进行的绘制 Internet 地图项目 IMP（Internet Mapping Project）【31】、Siamwalla 等人【32】的研究、Govindan 等人【6】开发的 Mercator 等。此后的研究基本都采用多点测量。

但在广域范围内采用 Traceroute 源路由选项进行拓扑测量。一方面源站选路速度较慢，自动确定采用哪些 IP 进行源路由也不是一件容易的事情；另一方面，Internet 上支持源路由选项的路由器非常少，不到 8%【6】（2000 年），并且随着安全措施的加强

将进一步减少，无法在大规模拓扑测量中使用。

单点测量的主要问题是不完全性，只能得到树状拓扑（未必严格的树）。

2. 采用公共 Traceroute 服务器（PTrS， Public Traceroute Server）进行多点测量

典型研究为 N. Spring 等人采用的 Rocketfuel、主要针对美国的大型 ISP 从 ISP 外部进行的拓扑测量。由于采用 PTrS，因此不需要布置探测引擎测量，这是其优点。但是，采用 PTrS 进行拓扑测量存在一些不利因素，最突出的一点是，目前，全世界内只有少数 ISP 的 PTrS（5.8%）支持在该 ISP 的多点探测，而单点测量无法得到较完全的拓扑结构。

3. 自主开发探测引擎进行多点测量

典型研究为 CAIDA【1】的 skitter 计划。1998 年 7 月 skitter 开始运行，其主要目的是进行 Internet 性能测量，其测量源点和目标的选择是为了满足性能测量的要求。目前，全世界已部署了 20 多个 skitter 监控器（测量源点），分布于亚洲，欧洲及北美洲，这些监控器探测 5 个不同地址列表中的目标地址，每个列表都针对某一特定用途。

国内关于 IP 网络拓扑发现算法主要是基于 SNMP 协议，通过访问 MIB，结合 ICMP 等协议，在局域网或一个管理域等小范围内构造网络拓扑。

采用 Traceroute 机制测量得到的选路路径是 IP 地址级的路径，除最后一跳之外，其它地址各对应一个路由器接口地址。直接从 IP 级路径生成的拓扑图称为 IP 级拓扑图（图中节点对应 IP 地址，即路由器接口）。为了得到路由器级拓扑图（图中节点对应路由器），需要进行别名解析（等同地址），也就是说，首先要识别出哪些 IP 地址在同一路由器上，然后将属于同一个路由器的接口地址合并起来成为一个节点。这个问题在 IPv6 网络中的解决和 IPv4 略有不同，此点将在一下章节中进行讨论。

在确定了测量的地理范围之后，一个重要的问题是如何选择探测的目标地址。总的原则应该是使过程具有覆盖完全、低负载、高效。

一般网络都采用路由聚合及面向目标地址的选路原则，也就是说，在路由稳定、无负载平衡的情况下，从子网络之外的某点到同一子网络地址范围内任意两个不同 IP 地址的路由是相同的。因此，当探测目标集合为完备目标集时，再增加探测目标也无法获得不同的转发路径，这时，给定探测范围内的拓扑测量的完全性就只由测量源点的数量和位置决定了。

路由器级拓扑探测中探测目标的选择主要有以下几种方式：

（1） 先筛选、编辑一个目标地址列表。这种方法会影响探测的完全性

（2） 根据网络地址空间前缀构造必要地址进行启发式目标选择及探测。在 IPv4 网络环境中，Govindan【6】等人指出，通过将启发式确定的前缀列表同一个骨干路由表中所包含的前缀相比较，发现路由表中有 8%的前缀没有被覆盖到，也就是说，有些子网根本就没有探测到。但这并不奇怪，因为即使是一个 Internet 骨干路由器，也只是包含全部地址空间中的以部分网络前缀。

（3） 从 Web 服务器列表收集大量地址，或从不同来源收集大量地址范围，然后从

每个可路由的，前缀长度为 24 的网络选取一个有响应的 IP 地址进行探测。从 Web 服务器列表随机抽样的方法存在的不足是：一方面，Web 服务器地址只是 Internet 整个地址空间的一部分，不可能保证每个用户子网络都存在 Web 服务器；另一方面，Web 服务器在 Internet 上的分布不均匀使得测量目标的选择可能是不完备的。后一种方法其测量目标集合的完备性一方面取决于搜集的地址列表的完全性，另一方面测量范围内的最小用户地址分配单位前缀长度不能大于 24，否则目标集合可能是不完备的。

（4） 根据 RouteViews 的 BGP 路由表中目的网络地址及 AS_PATH，选择 IP 地址进行探测【20-22】，称为有导向的探测。这种方法从面向 ISP 的拓扑测量来说，测量目标集合的完备性取决于 BGP 路由表中路由信息的完全性及其覆盖范围的完全性。关于 AS 级拓扑测量的研究已经表明，Internet 上任何一个路由器都只反映了到达一部分 Internet 地址空间的路由

特别需要说明的是，在 IPv4 网络中，在给定的探测范围地址空间（总体）内，一个选择探测目标的方法是按照一定的抽样率均匀随机地抽取目标地址。这时探测目标集的完备性将由抽样率来决定。如果抽样率较低，有可能是探测目标集不完备。有时为了增加对目标网络的覆盖往往对给定的地址空间进行遍历性的探测。遍历性的探测对较大的网络空间会带来性能上的瓶颈，而对抽样方法，抽样粒度往往决定了对网络的覆盖性，比如抽样率为 1%的探测过程往往探测不到子网络拓扑。并且现实网络情况相对复杂，很难判断出抽样比例增加到多少才能满足要求。对目标地址的均匀随机抽样无法保证探测目标集的完备性，而拓扑探测的目标之一是尽可能多地测量出通向子网络的转发路径，然后再生成相应的拓扑图。因此理想情况下，从每个子网络地址范围内选取一个可达的目标地址就足够了。现实的 IPv4 环境中，一方面在测量范围内从每个子网络确定一个可达的目标地址是不容易的，另一方面，路由可能不稳定。

2.3 tracroute 性能提高

通过 Traceroute 可以知道信息从你的计算机到互联网另一端的主机是走的什么路径。当然每次数据包由某一出发点（source）到达某一目的地(destination)走的路径可能会不一样，但基本上来说大部分时候所走的路由是相同的。UNIX 系统中，我们称之为 Traceroute,MS Windows 中为 Tracert。Traceroute 通过发送小的数据包到目的设备直到其返回，来测量其需要多长时间。一条路径上的每个设备 Traceroute 要测 3 次。输出结果中包括每次测试的时间(ms)和设备的名称（如有的话）及其 IP 地址（见图 3）。

```
Tracing route to www.yahoo.com [204.71.200.75]
over a maximum of 30 hops:

1 161 ms 150 ms 160 ms 202.99.38.67
2 151 ms 160 ms 160 ms 202.99.38.65
3 151 ms 160 ms 150 ms 202.97.16.170
4 151 ms 150 ms 150 ms 202.97.17.90
5 151 ms 150 ms 150 ms 202.97.10.5
6 151 ms 150 ms 150 ms 202.97.9.9
7 761 ms 761 ms 752 ms border7-serial3-0-0.Sacramento.cw.net [204.70.122.69]
8 751 ms 751 ms * core2-fddi-0.Sacramento.cw.net [204.70.164.49]
9 762 ms 771 ms 751 ms border8-fddi-0.Sacramento.cw.net [204.70.164.67]
10 721 ms * 741 ms globalcenter.Sacramento.cw.net [204.70.123.6]
11 * 761 ms 751 ms pos4-2-155M.cr2.SNV.globalcenter.net [206.132.150.237]
12 771 ms * 771 ms pos1-0-2488M.hr8.SNV.globalcenter.net [206.132.254.41]
13 731 ms 741 ms 751 ms bas1r-ge3-0-hr8.snv.yahoo.com [208.178.103.62]
14 781 ms 771 ms 781 ms www10.yahoo.com [204.71.200.75]

Trace complete.
```

图 2　Traceroute 示例

针对 tracroute 探测工具性能提高方法：

（1）　一次简单的 Traceroute 探测，对于每一跳即使得到了其回送报文，都会发生三个相同的探测报文。这对发现网络节点和网络路径其实没有太多实际的帮助。因此对发回回送报文的节点只发送一次报文，对无回应的节点（通过设定）将在较长时间间隔内发送一次或一次以上的报文，以确认其无回应并避免中间路由报文限制问题；

（2）　典型的 Traceroute 探测缺省的最大探测条数为 30，因此无论连续出现多少等待超时都将完整的做完 30 次的探测，这也造成了性能上的瓶颈。实际情况下，当同时连续 3 跳出现等待超时就可认为节点不可达，如此的处理完全不会影响最后拓扑数据的准确性。

（3）　在实际探测环境中，有必须经过的路由器，这些路由器往往是最近的网关，桩节点出口路由器等等，因此跳过对这些路由器的重复探测能很大地提高探测效率。

（4）　改进单点探测中类 Traceroute 过程的并行机制，提高探测线程的并行数。并行线程数方面，较好的网络环境中并行线程数可以达到 60 以上，一般情况下保持在 30 左右比较合适。

2.4　单点探测点

单节点发现的拓扑图是发散性的，辐射状的，往往会忽略所发现节点间的路径，造成“cross link”问题。多点探测的拓扑发现系统，能部分解决“cross link”问题。见图 2，只通过探测点 A 无法发现节点 2、3 之间的链路，导致拓扑链接状况的缺失；此时如果有探测点 B 的加入，就能自然地发现节点 2、3 之间的链路。

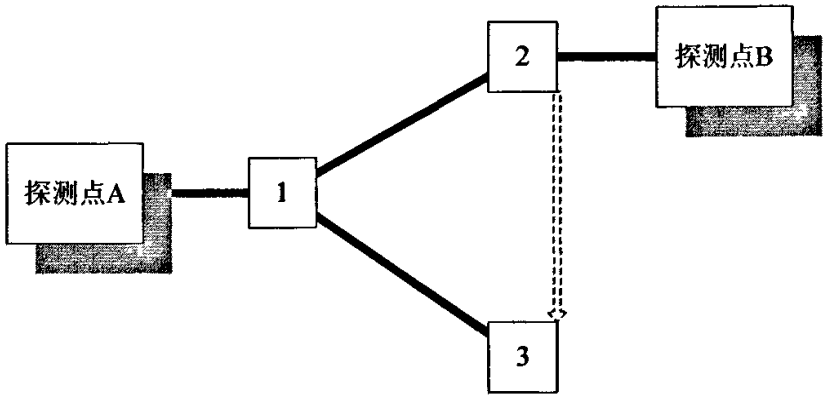


图 3　"Cross Link" 问题

一般单节点的探测方法都试图采用源路由（指定报文通过某些中间节点到达目的地）的手段提高链路发现的覆盖性。IPv4 和 IPv6 协议都提供源路由机制。但是在 IPv4 中，出于对安全的考虑，源路由机制基本上不可用（据【6】统计，可用率只有 8%）。源路由机制是 IPv6 强制要求的特性，而且在现有的 IPv6 试验环境中，大部分路由器也将源路由功能设置为打开的。这就为我们在新一代网络环境中研究和实现高效、准确、灵活的拓扑发现工具提供了良好的实验条件。一个高灵活性的拓扑发现系统，需要能自然融合单点探测和多点探测的优点，并尽可能提供不同的探测属性和方式。

2.5　种子节点列表

种子节点（seeds）是为网络拓扑发现而给出的路由器 IP 地址列表，它表示已知的路由器地址。对种子节点的探测是建立在网管人员对已有网络知识基础上的，所以无论准确度还是效率都有一定的保障。但这不仅要求用户要掌握整个网络的拓扑关系，而且要求种子节点尽可能得完备和尽可能选择靠近目标网络的边界。才能保证最终结果会有很好的覆盖性，能最大程度得接近实际网络。

种子节点能在已知边界节点集合的基础上，提供高效可靠的拓扑发现手段. 但是基于种子节点的发现手段，在边界节点不完全的前提下没有进一步提高准确性和发现更多节点的空间，这就需要其它的探测手段来弥补。种子节点列表的形成主要分为两个方面：一是将以往拓扑数据作为种子节点来源之一，二是用户对种子节点的维护（包括从 DNS 服务器得到 IP 地址列表，增加以往列表地址等）。种子节点地址列表进一步丰富了探测地址来源，提高对探测目标网络的覆盖和最终拓扑信息的准确性。

图 4 为基于种子节点的拓扑探测目标网络；图 5 为基于种子列表探测得到的拓扑关系图，其中白色节点和虚线为未发现节点和路径；图 6 添加种子列表记录，其中灰色节点为添加节点；最后图 7 为根据修改后的种子列表得到的最终拓扑关系。

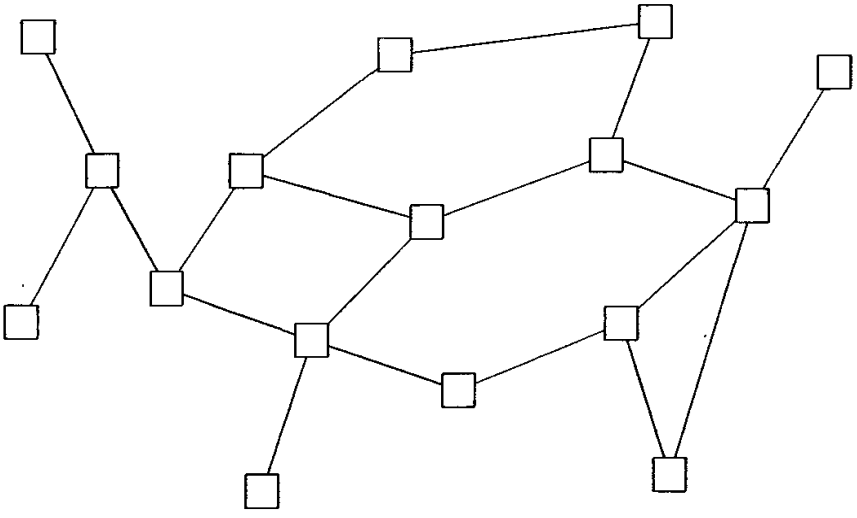


图 4　种子列表探测目标网络

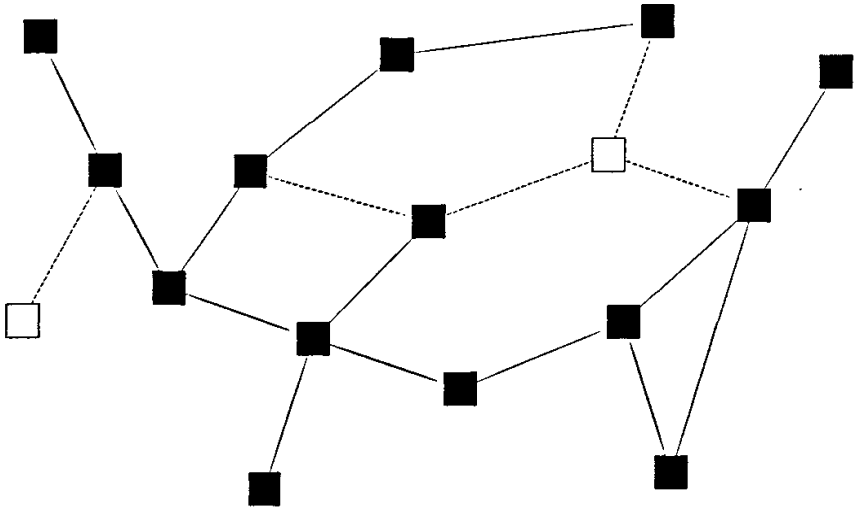


图 5　基于种子列表探测得到的拓扑关系图

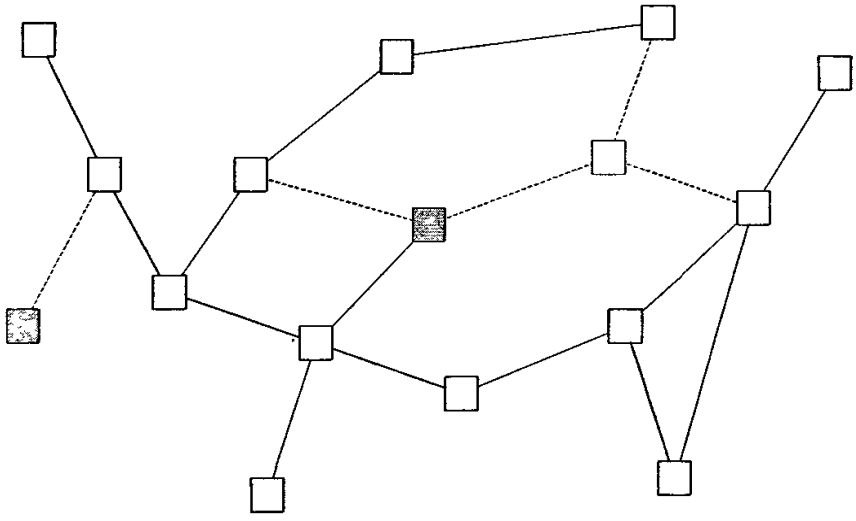


图 6　添加种子列表记录

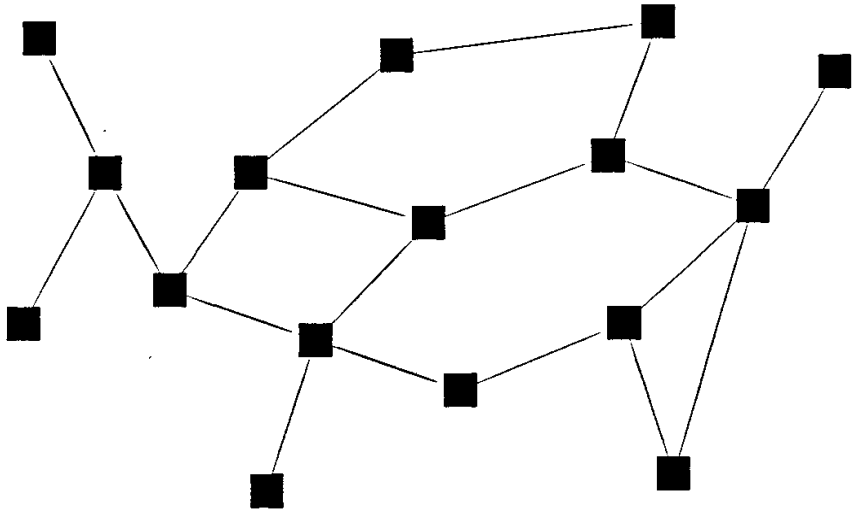


图 7　基于种子列表探测得到的最终拓扑关系图

2.6　本章小结

本章首先介绍了基于 Traceroute 的探测原理，详细解释了基于 Traceroute 机制采集选路信息；接着介绍现有主要的路由器级拓扑探测方式，描述了 Internet 路由器级拓扑测量从单点测量到多点测量的发展过程，列举了对 IPv4 网络测量中各种主要探测目标地址的来源；在第三节中讨论了 Traceroute 工具性能提高的方法：在实际探测中设定 Traceroute 的起始跳数，跳过对必须经过的路由器的探测；对发回回送报文的节点，系统只进行一次探测；对无回应的节点间隔一定时间进行一次以上的探测，以确认其无回应并避免中间路由报文限制问题；对多次连续无回应情况系统会在适当的时刻将探测终止而不等整个 Traceroute 过程结束；并行方面，较好的网络环境并行探测数可以达到

60 以上，一般情况下保持在 30 左右比较合适为提高探测性能，系统改进了探测过程并采用了并行机制；第四节介绍了网络拓扑中单点探测因为其拓扑结果是发散性的，辐射状而遇到的“cross link”问题，进一步说明 IPv6 源路由选项在解决“cross link”问题时作用。最后我们介绍种子列表并说明了种子列表探测过程。

第三章 实际网络环境问题

拓扑发现技术的关键点在于保证最终拓扑信息的正确性和完整性，并提高拓扑探测的性能。拓扑信息的正确性是指最终拓扑关系数据反映的节点和链路状况在对应的实际网络中真实存在；拓扑信息的完整性是指实际网络中的节点和链路在最终的拓扑关系数据中得到表达。由于众多实际因素的影响，对目标网络的探测很难保证最终拓扑数据 100%的正确和完整。

实际网络环境问题的有效解决很大程度上决定了最终拓扑信息的准确性，能为有效拓扑关系的获取提供基础。以下 1 到 5 小节分别对 ICMP 报文限制，匿名端口，路由循环，路由多址和不稳定路由等问题进行分析并提出实际解决方法。

3.1 中间路由报文限制

由于带宽限制，特别是防止 ICMP 的 DoS 攻击问题，网络节点会配置一定的 ICMP 报文限制策略，因此对同一路由器短时间内高强度地拓扑探测会引发报文转发限制的问题。这些限制主要分基于间隔时间和基于带宽的限制。不同的路由器往往采用不同的限制策略，甚至有些路由器不转发任何的 ICMP 报文直接将其丢弃或不回送 ICMP 报文。有一种避免报文堵塞的方法就是先对节点进行报文限制的探测。但是这种方法不仅耗费时间，不够准确，而且在一定程度上不可行和不必要。

本文开发的系统对中间路由报文限制问题主要采取避免的方法：首先，探测时间窗口内，系统散列地给出目标地址，因此避免了短时间内对同一路由器发送过多的探测报文；其次在探测过程中具有小型的自适应过程，当无回应报文时系统相应地增加探测报文的发送间隔。最后在极端情况下，由于报文阻塞的发生，已知的路由器会在探测过程中无法被唯一标识（匿名端口）。匿名现象的泛滥会影响最终拓扑信息的准确性，系统通过对匿名端口膨胀的处理保证即使这种极端情况的发生在最终拓扑关系数据上出现有且仅有一次。

3.2 匿名端口

在对网络进行探测时经常会遇到以下一种或几种情况：

（1）路由器不回送探测响应报文；

（2）路由器采用探测报文的目标地址而不是其接口地址，作为 ICMP 响应报文的源地址；

（3）路由器采用非全局地址，例如私有地址（private address）作为 ICMP 响应报文的源地址。

这些情况依次对应探测过程中的超时，回送报文源地址与探测报文目的地址相等以及回送报文源地址非全局地址等情况。

在一条探测路径中，这几种情况既可能单独出现，也可能连续出现。无论以上任何一种情况的发生，我们能确定存在这么的节点，但却无法唯一地表示它，因此我们称其为匿名端口，其表明存在一个路由器（接口），但无法得到其全局 IP 地址。

一般情况下，路由器在响应报文里使用自己端口中一个全球单播地址作为源地址，这是 Traceroute 作为探测工具的前提。所以匿名端口产生了新的问题：我们可以将拥有全球单播地址的网络节点区别，但却无法对匿名端口进行唯一标识和分配，导致匿名端口数量的膨胀，见图 8。

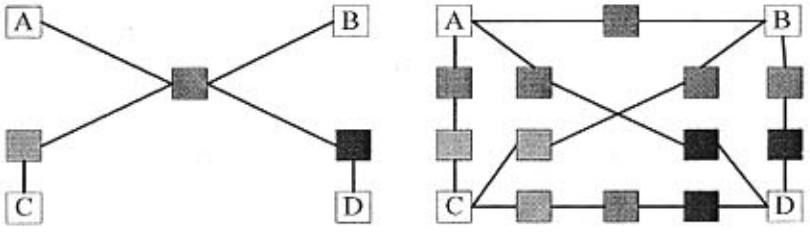


图 8　匿名端口膨胀示例

由于匿名端口情况的发生相对比较稀少，在实际网络中并没有很复杂的匿名端口情况，导致较大地影响最后拓扑关系数据。因此对匿名端口的处理本系统主要采用了合并的方法，即将可能相同的匿名端口合并。本文在解决匿名端口问题时，是在拓扑探测的同时解决匿名端口现象，将匿名端口带来的拓扑膨胀最小化。匿名端口的问题是在每个探测进程中解决的，而不是在拓扑探测完成后对拓扑数据修改而得出的。特别需要说明的是，如果在同一探测进程中发现不同的匿名端口，不是将其合并，而是应该当成不同的匿名端口进行处理。在现实网络环境下匿名端口的发送毕竟是少数，特别是使用了避免的措施后，对于其的进一步的讨论可见【10】。

系统在拓扑发现过程中，特别是对指定前缀的基于源路由的路径探测中往往遇到匿名端口膨胀的情况，本系统通过适当的方法很好地抑制了膨胀的发生，并保证最终数据的完整性。

3.3　路由循环

不当的网络配置会导致路由循环现象，虽然其不会产生错误的路径，但降低了探测效率并产生不必要的拓扑数据。不正确的路由表导致路由循环的产生，其往往发生在静态路由协议中或由隧道内路由器随机转发报文引发。比如探测路径会反映为 U-V-X-Y-X-Y…，表示路由器 X 和 Y 间有路由循环。路由循环会发送在两个或两个以上的路由器之间，但两个路由器间的循环比较常见。路由循环，匿名接口和中间路由报文限制等问题的同时发生，会给拓扑发现的正确性带来更多的挑战。这时就需要识别特殊的路由循环状况，避免对最后的拓扑发现结果产生不必要的影响。在路由循环和中间路由

报文限制问题同时发生会导致最终拓扑数据的偏差，如U-V-Y-X-Y1-X-Y2…（这时Y1，Y2为匿名端口，应被识别为Y）。本文系统能识别路由循环和中间路由报文限制问题同时发生产生的偏差，如将上例探测数据识别为U-V-Y-X.

3.4 别名解析

所谓别名解析（alias resolution），就是识别出那些IP地址属于同一个路由器，然后将别名接口地址合并的过程。别名解析无论在IPv4还是IPv6路由器级拓扑探测中都是一个重要的问题。IP层协议的改变给在IPv6网络环境下更好地解决别名解析问题带来了机遇和挑战。

一个路由器通常具有多个IP地址与不同的网络相连。不同的探测序列可能从不同的端口经过同一路由器，从而返回属于同一路由器的不同IP地址。如何确定多个IP地址是否属于同一台路由设备在路由器拓扑发现中称作路由多址问题或路由器别名解析问题。IPv4拓扑发现中主要有两种方案来解决别名解析问题。第一种是通过对IP报头序号的判断来解决路由多址问题，即向两个待确认地址同时发送两个ICMP报文，如果返回的两个IP报文的序号接近，则认为这两个IP地址属于同一台路由设备；第二种方法是基于UDP端口不可达报文，这种传统的别名解析方法是基于相同源地址的方法，最初由Pansiot【5】等人采用，之后在IMP【29】【30】、Mercator【6】、iffinder【47】以及Ally【16】中采用。其基本原理是：一般的路由器都会指定一个特定端口的地址作为发送UDP端口不可达报文的源地址，因此，对于待确认地址A和B，同时向其发送两个UDP报文，端口为一个未使用的端口（一般为大于1024的端口），则如果A与B属于同一台路由设备，则在回应的UDP端口不可达报文中，将会使用相同的源地址。一个别名集合是一个等价类。针对不同目的地址的别名探测收到具有相同源地址的“端口不可达”响应报文，则它们是别名。

基于UDP端口不可达报文的优点是一般不会产生错误的判断。但该技术也存在一些局限性，使别名解析不完全，从而导致随时间推移不完全的路由器级拓扑图的不可比性。首先，如果路由器被配置为“采用收到报文的目的地址作为响应报文的源地址”，那么通过响应报文将无法发现别名。其次，大约10%（也许更高）的核心路由器从来不响应未知端口的UDP报文，对这些路由器来说，该技术无法解析别名【48】【16】。

N. Spring等人【16】【49】提出了另外3中种不同的别名解析方法，但都有一定的程度的误报：

第一种别名解析方法基于报文标识符IPID的递增性和邻近性。这种方法不仅没有进一步解决解析的不完全性，另一方面大大提高了错误判断的危险性。网络中大量报文的发送和选择合理的阀值都是很难确定的实际因素。重要的是IPv6协议中取消了报文序号，彻底是这种方法在下一代网络中失去了生命力。

第二种是基于域名的别名解析。其是通过域名中可能包含的接口信息进行推断别名。

这种方法依赖于域名的命名方法及准确度，而域名中不包含接口信息以及拥有域名的接口较少，都会使这种方法失效。

第三种方法是基于图的别名解析方法，其主要的考虑为：一方面，在不存在路由循环情况下，同一条路径上的接口不可能是别名；另一方面，在路由器对点连接以及采用入口地址作为“超时”报文的源地址时，具有相同直接后续（即下一条）的两个地址是别名。这种方法对于采用多路访问媒介或交换网络连接路由器的情况就失效了。

通过以上分析，我们仍然采用传统的基于 UDP 不可达报文的别名解析方式，虽然可能存在解析不完全，但不会存在错误的解析。并且下面我们利用 IPv6 源路由选项增强了对解析不完全问题的解决。

对于 IPv6 网络，由于很大部分路由器沿用 IPv4 的 UDP 报文处理模型，因此基于 UDP 端口不可达报文的方法【5、6】仍然适用；此外根据 RFC2463 和具体的报文处理过程，部分路由器被配置为“采用收到报文的目的地址作为响应报文的源地址”，并且 IPv4 网络基本禁用源路由选项，因此造成在 IPv4 网络下部分路由器不适用已有别名解析方法，但因为在 IPv6 网络环境下源路由选项的基本开通，所以存在额外解决方法【3】。

本文系统实际别名解析方法描述如下：

1. 首先测试离探测点距离为 n 的地址 A（路由器）是否可用 UDP 端口不可达方法（即发送跳数为 n 的高端口 UDP 报文）。
2. 可用则结束，否则发送经 A 到可能为等同地址 Ax（其跳数 n）的 ICMPv6 报文。
3. 返回报文源地址为 A 则结束，否则发送跳数为 n+1 经 A 到 Ax 的 UDP 报文。
4. 产生的端口不可达 ICMPv6 报文源地址为 Ax 则 A-Ax 为等同地址；否则 A-Ax 不等同。

下面将对以上步骤进行说明：

- 步骤 1，我们的目的是尽可能识别出利用 UDP 端口不可达方法可识别地址等价类，减少别名解析处理的地址集合
- 步骤 2，判断地址 A（路由器）报文处理模型是否先处理源路由选项然后判断跳数值（见图 9），如果报文处理模型不符合要求，则将无法利用 ICMPv6 或 UDP 报文识别路由器别名
- 步骤 3、4，我们将发送带有路由选项的高端口 UDP 报文，其跳数值为 n+1；因为地址 A（路由器）采用了符合条件的报文处理模型，所以如果猜测的地址 Ax 和 A 等同，那回送的 ICMPv6 报文源地址将为 Ax
- 特别需要指出，Ax 地址的选择可以通过枚举，等同前缀或拓扑关系图的分析给出

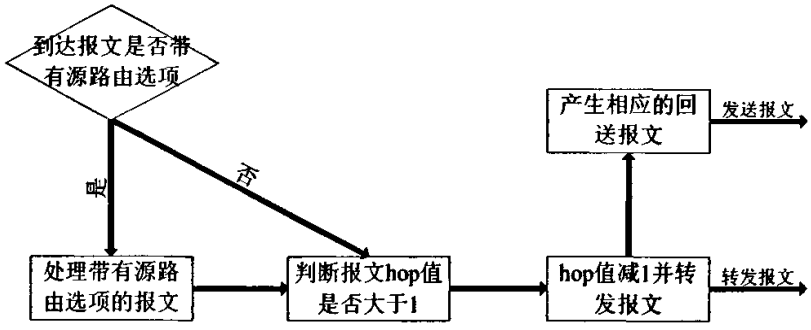


图 9　路由器报文处理模型

别名解析的不完全性使最后拓扑关系数据中的节点粒度不统一，有些节点代表的是路由器，有些代表的路由器中的一个接口。因此，目前通过探测得到的路由器级拓扑图实际上是介于接口级拓扑和路由器级拓扑之间的一种混合结构。我们研究的目标就是促使最终拓扑关系数据向实际路由器级拓扑靠近。

3.5　不稳定路由

对于基于 Traceroute 探测技术的一个主要假设是：在给定方向上，任意两个节点发送的报文将总是经过相同的路由器，也就说会有一样的 Traceroute 路径信息；这就是说其路径信息是稳定的。但是这个假设在现实网络环境中就不存在了。虽然很多时候路由是稳定的，但是部分路由器还是采用了动态路由来平衡负载。主要的路由协议（比如 OSPF 和 IS-IS）都支持对给定的目标地址采用等同的负载平衡，并且动态路由策略往往采用不同的实现，比如有基于源和目标地址的 hash 算法，轮询算法或一些负载相关的算法（参见 RFC 2328）。

不稳定路由主要影响 Traceroute 收集的链路信息，图 10 的例子中说明了这个问题：图中 A 路由器采用了动态路由策略。此时从 P 点到 D 的 Traceroute 探测因为 A 点动态路由策略的存在导致回送的路径信息为 P-A-B-D，但现实情况 B-D 的链路并不存在。

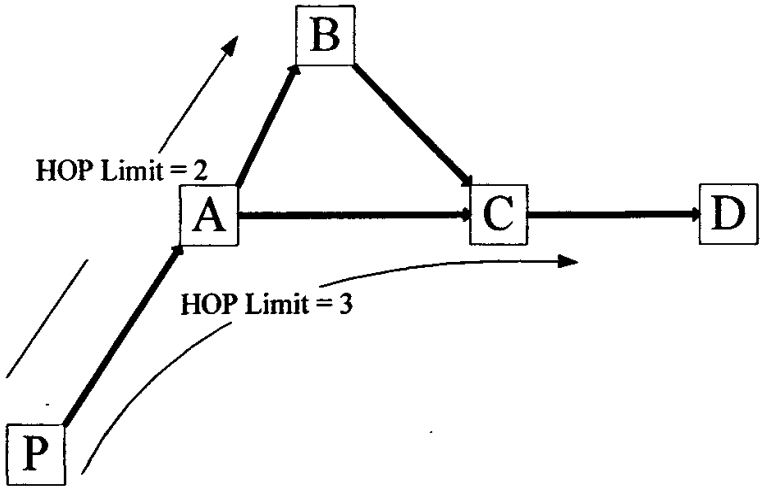


图 10　不稳定路由示例

在 IPv4 环境中，因为源路由选项的不充分设置，导致只能通过大量发生探测报文来判断动态路由的存在，此方法在实际使用中往往不可行，带来过多的错误信息。在 IPv6 环境中，我们在探测和确定路径过程中加入已探明链路，结合 IP 层源路由选项迫使探测报文通过指定的已知路径。此方法能收集绝大部分的不稳定路由路径信息，丰富了最后的拓扑数据。特别地，系统执行时往往针对特定的地址前缀进行探测，因为执行动态路由的设备主要分布在骨干网。

3.6　本章小结

本章主要介绍了 IPv4 和 IPv6 网络拓扑发现过程中遇到的主要网络现象并分别介绍了本文系统的解决方法：

1. 提出具体的解决中间路由报文限制问题的方法。
2. 利用简单可靠的合并方法解决了匿名端口膨胀问题，抑制了利用源路由探测路径时匿名端口膨胀的现象。
3. 不仅解决一般路由循环问题，而且能识别较复杂路由循环的发生
4. 对于别名解析问题，本文系统利用 IPv6 网络环境提供的优势提出了实际的解决方案：在采用传统基于 UDP 端口不可达方法的基础上，利用 IPv6 源路由选项进一步完善了对此问题的解决。
5. 充分利用源路由选项，解决 IPv4 网络环境下无法解决的不稳定路由问题，保证了最后路由信息的准确性。

第四章 IPv6 拓扑发现系统

探测目标地址集合的完备性保证了最终拓扑信息的完整性。探测目标地址集合意味着对目标网络的覆盖，覆盖应促使尽可能发现所管理范围（或探测范围）内的网络节点及其链接关系，同时满足拓扑发现系统对效率的要求，即以最小的地址空间覆盖目标网络。保证最终拓扑信息的完整性必将联系到探测效率问题。

IPv6 拓扑发现系统的输入是代表实际网络的探测目标地址，输出是最终的拓扑关系数据，以下主要介绍本文 IPv6 拓扑发现系统所需目标地址的来源和系统构架。

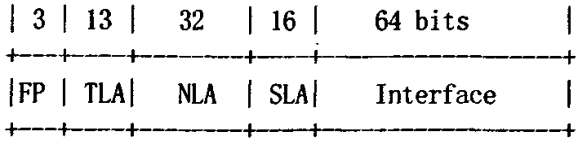
4.1 典型目标地址获取

路由器级拓扑探测中目标地址选择主要几种方式：目标地址列表（种子节点），地址空间范围内构造探测地址，DNS 服务器地址获取和 BGP Routing Table。以上方式构造的目标地址主要正对 AS 级网络拓扑测量，最终目的就是为拓扑发现或测量系统提供系统输入。

从 Web 服务器获取探测地址和 BGP Routing Table 获取地址做为拓扑系统输入合适于大规模公共网络测量。从 BGP Routing Table 中得到相连 AS 域的地址前缀信息，构造相应 IPv6 地址作为地址列表的来源进行 Traceroute 探测【20-22】，需要 AS 域边界路由器的存在列表和信息读取权限，并且其关注于域间的商业链接关系；从 Web 服务器获取探测地址，只是整个地址空间的一部分，不可能保证每个用户子网络都存在 Web 服务器，Web 服务器在探测网络上的分布不均匀使得测量目标的选择可能是不完备的。AS 级拓扑发现地址来源手段无法适应 AS 域内部的路由器级拓扑发现的细粒度要求，因此我们需要寻找更细粒度的目标地址集构造方案

目标地址列表和探测范围内地址空间构造做为拓扑系统输入比较通用，没有探测范围的限制。本文系统目前支持目标地址列表和探测范围内地址空间构造做为系统目标地址输入。

在现有 IPv6 网络中，由于 IP 层协议特点，特别强调了网络地址的聚会特性，这就根本上改善了探测范围内地址空间构造方法对目标网络的覆盖程度，特别是对单一管理域内的网络拓扑发现过程，能收到理想的拓扑关系。



按照现在的 IPv6 地址结构规定（RFC 2471），典型 IPv6 全球地址应该由固定前缀 001，13 位的顶级聚合 ID（TLA），8 位保留，24 位的下一级聚合 ID（NLA），16 为的场

点聚合 ID（SLA）和 64 位的接口 ID 组成。*APNIC, ARIN, RIPE NCC*制定了 IPv6 地址的分配政策，ISP 级的地址前缀一般为 32，在国内 Cernet2，中国移动，中国网通等 IPv6 的试验网所分配到的地址前缀就是如此。根据 RFC3177 建议 ISP 范围内所分配的地址大小分为两种，一种为前缀为 48 的地址块，另一种为前缀为 64 的单子网地址。实际中无论是 48 位还是 64 位前缀的地址块分配后，下一级的网络管理员也往往按照实际需要对地址块进一步规划。在目前情况下还很少发生地址空间不够，需要前缀合并的情况。

首先 IPv6 地址分配原则强调了聚会的特性，使得针对子网前缀构造地址进行探测能保证到达子网络，并且源路由选项在网络探测上的应用避免了路由不稳定对链路发现的影响并且能很大程度增加骨干网节点的发现程度。唯一需要解决的问题了，被留到了对子网络的覆盖，因为 IPv6 网络还处于发展阶段，大规模的应用还没有开始，所以对有些叫小的网络（地址空间前缀并不小）对子网络的探测就存在抽样或遍历的矛盾。比如对于地址前缀为32为的目标网络空间，如果桩地址前缀为48，对网络的完全覆盖需要65536次探测（虽然这种情况较少发生，32 位的地址空间一般会将桩地址前缀设定在 48 位以上）。这是所有可能桩地址前缀进行完全覆盖存在性能上的瓶颈，可以通过一下几种方法解决。

（1）随机抽取桩地址空间，构造地址进行探测。一般均匀的抽样在 10%以上，可以保证较好理想的覆盖度

（2）提高所构造地址的前缀，比如适当得将 48 位前缀改为 42 位。此方法对非均匀分配地址空间的网络效果不好

（3）为探测过程提供一个以上的并行探测点，并且每个探测点同时并行进行网络探测（具体参见下文中系统构架和多点并行探测）。

目标地址列表做为系统目标地址来源，其可以来自以往拓扑数据和用户手动输入等等，进一步丰富了地址来源，提高对目标网络的覆盖度和最终拓扑信息的准确性（具体见第二章种子节点列表）。

通过地址前缀构造域内地址仍然存在探测性能问题。例如，针对 32 位的地址前缀，如果运营商配置子网采用 48 位地址前缀，保证对目标网络的完全覆盖则需要类似 Traceroute 探测数为 65536 次，如果子网前缀改为 64 位那探测数将达到现有互联网所有地址的总和。通过在构造地址集合中随机抽取地址进行探测（比如 10%、20%等比例）或增加所构造地址的前缀长度（比如适当得将 48 位前缀改为 42 位）的方法以提高了提高探测效率，但仍然没有根本性解决探测效率问题，特别是对目标网络的覆盖和针对性的探测，特别是仍无法满足对目标网络进行监控的功能要求。

以上介绍的典型目标地址来源比较机械，对实际应用缺乏足够的针对性，没有彻底解决探测性能和覆盖度的问题。特别是没有满足基于 ICMP 报文探测的拓扑发现系统监控目标网络的实际需求。监控目标网络要求目标地址的给出具有针对性和动态性，并且要求一定是实时性。下节介绍的目标地址获取方法将一定程度上解决基于 ICMP 报文探测的拓扑发现系统监控目标网络方面的问题。

4.2 IPv6 地址空间分配和目标地址获取

一般而言，网络管理有五大功能：配置管理、性能管理、故障管理、安全管理、计费管理。这五大功能是保证一个网络系统能够正常运行的基本功能集合。地址空间分配属于配置管理，在网络规划中，地址分配方案的设计至关重要，好的 IP 地址分配方案不仅可以减少网络负荷，还能为以后的网络扩展打下良好的基础。

不同于 IPv4 网络协议，IPv6 网络协议提供了巨大的地址空间。IPv6 的地址结构和地址分配采用严格的层次结构，以便于进行地址聚合，从而达到减小路由器中路由表的规模。IETF（互联网工程任务组）对 IPv6 协议和地址类型及其分配做出了相关规定。IP 地址规划主要涉及到网络资源的利用的方便有效的管理网络的问题，IPv6 地址有 128 位，其中可供分配为网络前缀的空间有 64 位，按照最新的 IPv6 RFC3513，IPv6 地址分为全球可路由前缀和子网 ID 两部分，协议并没有明确的规定全球可路由前缀和子网 ID 各自占的 bit 数，目前 APNIC 能够 申请到的 IPv6 地址空间为/32 的地址。因此对于分配了 32 位地址前缀的 LIR 或较大 ISP，其 IPv6 地址空间内需要对 2^16 至 2^64 的子网地址进行规划和分配（见图 11），即使前缀为 64 为的子网内也存在地址段规划和分配的问题。由于 IPv6 的地址空间巨大，因此尽可能减少对路由表容量是非常重要的。合理划分地址空间是下一代互联网中地址有效聚合的关键。

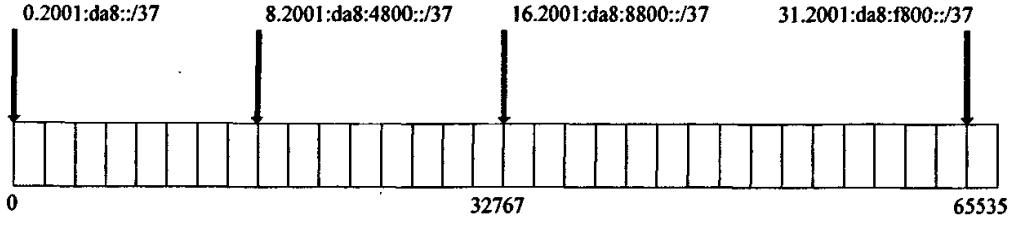


图 11 48 位子网前缀地址空间示例

运营商网络具有规模庞大、结构复杂、用户不断增加、需要不断扩容升级等特点，这就要求在进行地址空间分配时，需要进行仔细规划，遵循一定的地址分配原则，才能使得网络稳定、高效地运行，走上可持续发展之路。

运营商网络 IP 地址分配的原则：

1. 自治原则

公共互联网络被划分成几个大的自治区域，每个大自治区域中有被划分成几个小的自治区域。运营商的网络也可以这样进行管理。虽然运营商的网络比较庞大，但是 如果将整个网络分切成小块的网络，每个小块的网络管理与网络其

他部分相对独立，这样管理起来就比较方便了。

2. 顺序原则

按照自治原则将网络进行逻辑划分后，就可以根据地域、设备分布及区域内用户数量来进行子网规划。这样就充分考虑了网络层次和路由协议的规划，通过聚合网络 减少网络中路由的数目和地址维护的数量，充分体现了分层管理的思想。同时，IP 地址规划要和网络层次规划、路由协议规划、流量规划等结合起来考虑。在进行 地址分配时，为了提高地址分配效率和地址利用率，最好按照一定的顺序进行。选择的顺序可以是自上而下的顺序，也可以是自下而上的顺序，还可以是二者结合使用。

3. 可持续发展原则

由于网络用户数持续高速增长，全国性大集团性用户不断增加，再加上政府上网工程和电子商务的全面启动，电信网络所要承载的业务量和业务种类越来越多，这使 得电信网络需要频频进行技术升级、改造和扩容。所以，在进行地址分配时必须要从分考虑到这些因素，为网络的每个部分留有部分地址冗余，这样才能保证网络的可持续发展

4. 可聚合原则

互联网日新月异的发展和日益庞大的规模令当初设计互联网络的专家始料不及。在路由表的急剧膨胀情况下，可聚合原则是网络地址分配时所必须遵守的最高原则。 原因有以下两个方面。由于 IPv6 地址空间非常庞大，如果规划不好，其路 由条目也可能会非常庞大，而且还会以较高的速度急剧增长。可聚合原则要求我们在进行地址规划时，应提供足够的路由冗余功能

5. 整体和层次原则

IP 地址规划应该是网络整体规划的一部分，即 IP 地址规划要和网络层次规划、路由协议规划、流量规划等结合起来考虑。IP 地址的规划应尽可能和网络层次相对应，应该是自顶向下的一种规划。充分合理利用已申请的地址空间，提高地址的利用效率。

根据以上原则，克服现有人工 IPv6 网络地址分配方法的缺点，我们设计了一种基于关键字和优先级的自动 IPv6 地址空间分配方法。利用此方法能在庞大的 IPv6 地址空间中根据实际需要自动得到所需地址空间，使所分配地址空间具有合理的可持续发展空间，可以合理地规范 IPv6 网络地址分配并显著地提高 IPv6 网络的地址聚合度。

主要步骤：

1. 目标地址空间初始化：

根据运营商所分配的 IPv6 地址前缀（例如 32）和子网地址前缀（例如 48 位、64 等），给所有子网地址编号；根据用户设定的层次数量和其他参数初始化整个地址空间，并提供具体修改功能（见图 12）：

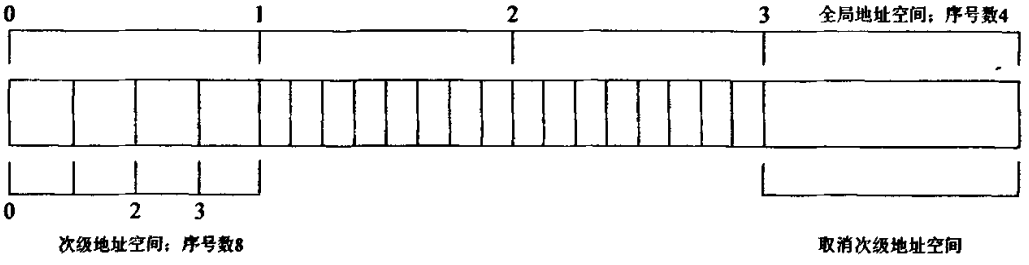


图 12　地址空间初始化的层次

2. 关键字分配：

在地址空间内，根据关键字序列分配相应地址空间。对已分配关键字序列则在相应的地址空间的进行优先级分配。

图 13 中展示了在运营商地址前缀内关键字空间的选择顺序，其选择原则为：确保最先分配关键字序列的扩展空间。图 14 为具体关键字分配示例。

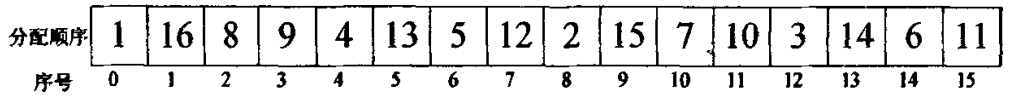


图 13　关键字空间选择顺序示例

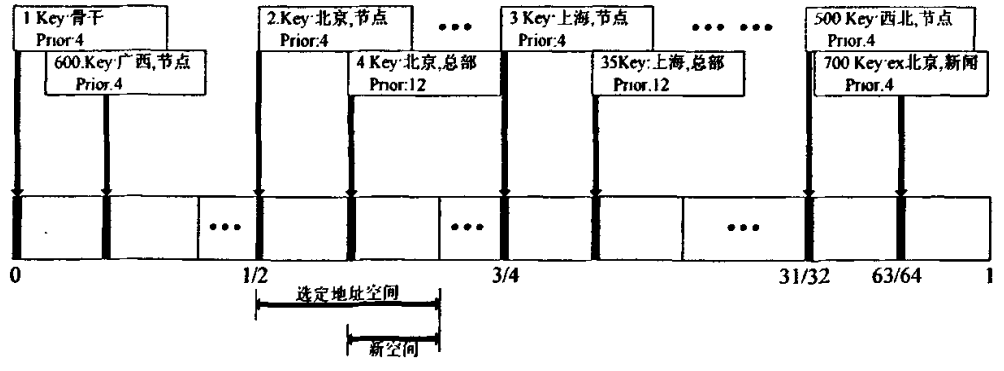


图 14　关键字分配示例

3. 优先级分配：

关键字分配过程后，找到关键字序列对应的地址空间，在此地址空间内进行基于优先级的地址空间分配：查找地址空间优先级最低的地址空间，如果优先级相同则按照用户设置偏好进行选择；对选定的地址空间进行分割，完成分配过程（见图 15）。

优先级的确定方法：

地址空间优先级= Prior + f(MaxGlobal, MaxLocal) + p(NumUsed)；

MaxGlobal 指定关键字序列对应空间中最大能分配到的新地址空间；

MaxLocal 优先级分配中所选择地址块内最大能分配到的新地址空间；

NumUsed 优先级分配中所选择地址块内已分配的地址空间数量；

f(MaxGlobal, MaxLocal)指关于 MaxGlobal 和 MaxLocal 的函数；其值于 MaxGlobal 成正比，MaxLocal 成反比；

p(NumUsed)指关于 NumUsed 的函数；于 NumUsed 成正比；

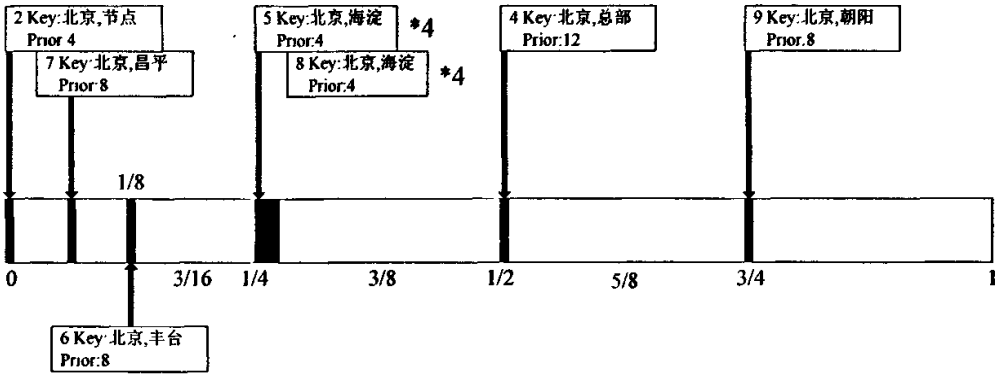


图 15　优先级分配示例

特别需要说明的是：当地址分配过程中遇到地址空间冲突时，比如在基于优先级的地址分配中发生地址耗尽等问题，将需要以下的冲突处理流程来处理。

4. 冲突处理：

对应关键子序列的地址空间后是否存在相邻序号未分配；

有则，扩充地址空间（见图 16）；

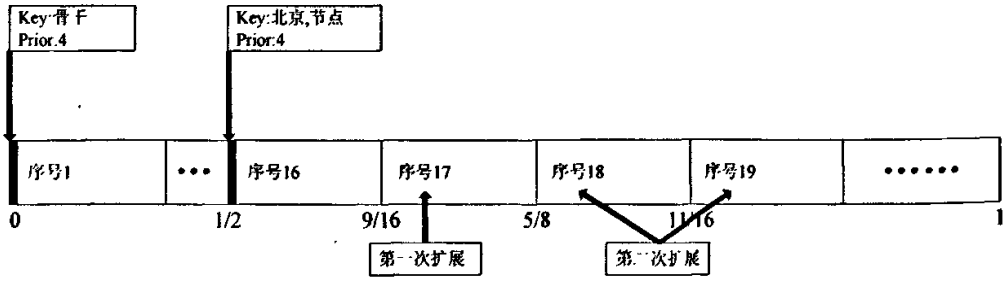


图 16　地址空间合并

否则，修改关键字序列，在关键字序列前加上“ex”，重新尝试分配地址空间（见图 14 中 700.ex 北京，新闻）；

以上操作都无效，则告知用户无法进行地址空间分配。

5. 其他手工操作：

完备的自动地址空间配方方案也离不开补助的手动设置，使其有更多的灵活性和适应能力。本系统就包括对关键字层次的修改，对特定地址空间合并等操作等等。

此 IPv6 自动地址空间分配方法与现有人工 IPv6 地址分配方法相比较有以下优点：

第一，根据用户设定预先分配较大地址空间，并规定了关键字地址空间选择原则。其有益效果是，避免在开始阶段对先到者分配过大的地址空间并充分利用先分配序号有更大地址空间增长需求这一实际规律，保证了对应关键字序号未用完情况下未来地址空间的增长并保证序号全部用完的情况下分配到最大可用地址空间，提高了地址空间分配的合理性，提高了 IPv6 地址聚合；

第二，在存在相符关键字的情况下，分配操作在关键字对应大地址空间内完成，可以根据实际需要将分配地址空间时需要考量的匹配因素考虑进来，比如地理信息、路由信息等等。

第三，在关键字空间相应序号用尽情况下，从能分配最大地址空间的已分配地址块中得到新的地址空间；其有益效果是，抢夺能分配最大地址空间的已分配地址块本质上是抢夺利用率不高地址空间，从而保证了地址空间分配的合理性，提高了 IPv6 地址聚合。

第四，可以根据用户需要修改关键字层次合并地址空间等，充分满足了不同的地址空间分配需求，提高了实际应用的灵活性。

第五，与现有基于增长空间的 IPv6 地址分配方法相比较，其有益效果是，在地址空间优先级的计算上。以动态优先级作为分配依据，其考量包括了人为经验判定成分，可增长的地址空间和其未来的增长趋势。可以使对应地址空间根据当前系统分配状况拥有相应的动态优先级，确保过多或不当的地址分配情况，从而进一步保证了地址空间分配的合理性，提高了 IPv6 地址聚合。

根据上节对典型目标地址获取的介绍，我们可以了解典型目标地址获取方式主要针对公共网络路由器级拓扑测量。其地址给出粒度较大，关注点集中在 AS 域间，拓扑测量周期长，效率要求不高等。

我们认为目标地址获取限制了基于 ICMP（Traceroute）的拓扑探测对更小粒度和范围的网络管理系统或域内网络拓扑发现系统的应用。

域内 IPv6 拓扑发现系统存在获取合理目标网络地址集合的问题。特别是没有满足基于 ICMPv6 报文探测的拓扑发现系统监控目标网络的实际需求。监控目标网络要求目标地址的给出具有针对性和动态性，并且要求一定的实时性。在此我们将 IPv6 地址空间分配模块作为 IPv6 拓扑发现系统可靠探测目标地址集合获取途径以解决基于 ICMP 报文探测的 IPv6 拓扑发现系统以上方面的问题，见下图 17：

首先拓扑发现系统根据 IPv6 地址空间分配模块中已记录的地址空间前缀（已分配的地址空间）构造目标探测地址（见下图上方的地址块）。系统根据这些地址空间前缀构

造 IPv6 拓扑探测所需目标地址集合的一部分。这部分目标地址具有很好的针对性，能代表被管理网络。这部分目标地址代表已注册网络，是可靠的地址信息来源。但注册信息落后与实际网络的发展，无法完成对实际网络完全的覆盖，特别是要求对实际网络的监控。因此要求更丰富的目标地址来源。

为了完成以上目标，我们在未使用的地址空间中，根据 IPv6 地址聚合和现有子网地址前缀长度构造目标地址集合的另一部分。这部分地址的给出将在空白地址空间中执行一定的随机性，来确定所探测的空白地址空间是否在实际网络中被使用，从而起到监测网络拓扑的作用（见下图下方）。

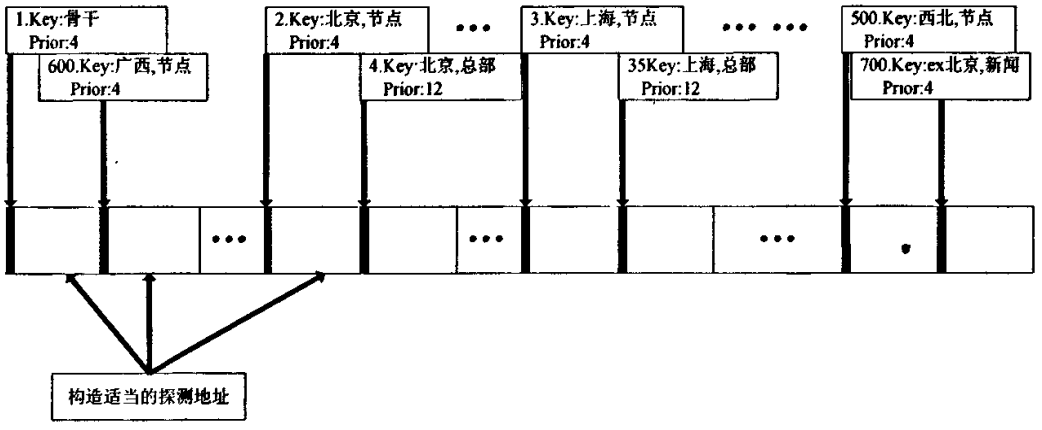


图 17　从 IPv6 地址空间分配中获取目标地址

具体对目标网络进行监控是个动态并持续的过程，对于不同类别的地址应采用不同处理方式。首先对于骨干网节点和已注册节点应进行维护，维护这个集合的准确性，增加其中地址或者确认其存活状态；其次对于未使用的地址空间的探测将较低频率地持续地发起报文探测，当发现未注册地址时进行记录，并在以后的探测过程中增加对其探测的频率，直到其注册状态改变。

域内拓扑发现和地址空间管理都属于网络管理中配置管理的范畴，网络管理系统需要针对不同的管理需要进行不同的调整。我们希望地址空间和拓扑发现系统的结合能在实际使用中得到进一步的改进和修正。

4.3　IPv6 拓扑发现系统设计

系统基本单位由探测节点、拓扑发现平台和本网拓扑数据模块构成。探测节点负责报文的发送和接收；拓扑发现平台负责拓扑数据的收集、验证和修改；本网拓扑数据模块提供拓扑数据获取接口。系统基本单位完成本层网络拓扑信息的收集、处理和发布。内部各模块通过 Web Services 交互。模块相互独立，各自部署，在拓扑发现平台内配置相应 WSDL 即可。

完备的自动拓扑发现系统由若干个基本单位构成，其中的一个拓扑发现平台作为主平台负责所有拓扑信息的合成。特别需要指出，拓扑发现系统平台可使用系统内每个基

本单位的探测节点，见图 18：主拓扑发现平台利用本单位和次级单位的探测节点构成对本层网络的多点或并行探测，同时使用本层拓扑数据和下层子网拓扑数据合成分层的拓扑数据关系。

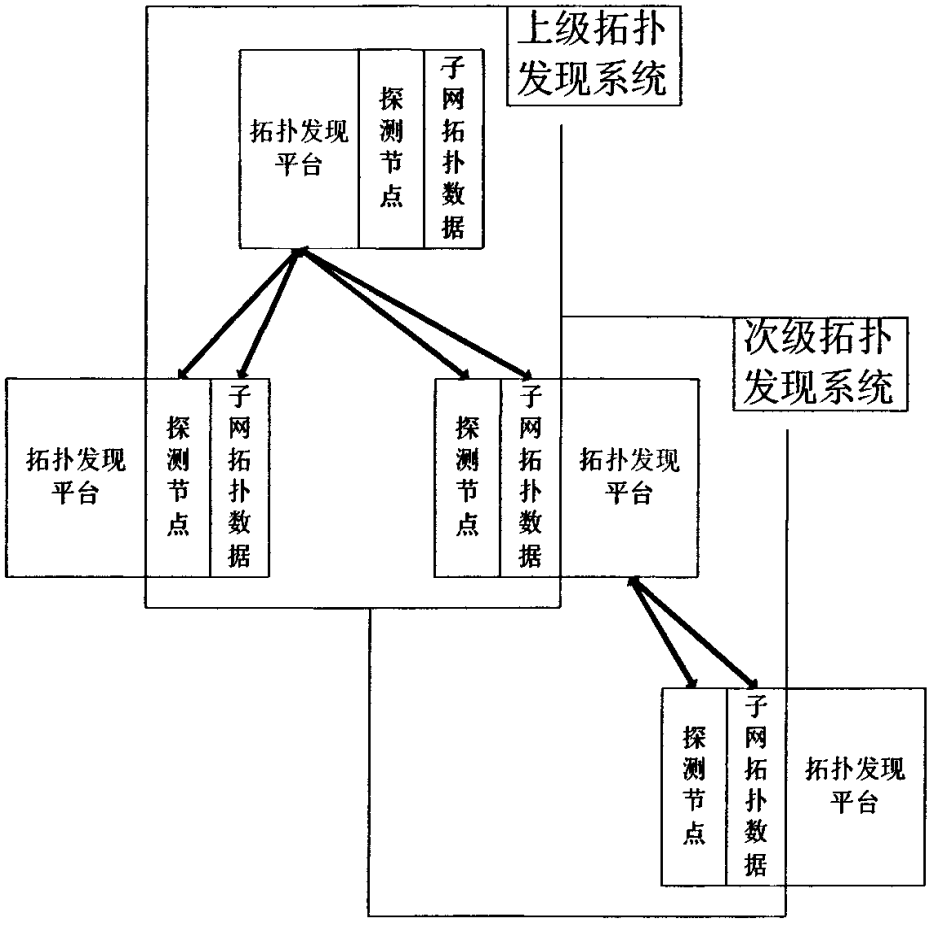


图 18　系统架构

本系统将 Web Services 作为模块间通讯的基础，解析了拓扑发现系统的三个主要模块，降低了探测、数据获取和管理三者间的耦合度。在探测方面可以重复利用不同的探测点从不同位置对目标网络进行单点或多点的探测作业，提高了拓扑发现的灵活性和准确性。在数据获取方面，因为 Web Servics 的应用，为其提供了统一的接口，解决拓扑关系层次属性的表达。Web Services 技术的应用提高了拓扑发现系统整体的适应性，满足了不同网络环境对管理的需求。例如，对某个网络拓扑关系的表示完全可以脱离探测和拓扑管理平台，而只采用数据获取模块；或脱离数据获取或拓扑管理平台，而只为其他平台提供探测点；等等。

4.4　多点并行探测

IPv6 拓扑发现系统地址前缀构造地址会需要探测效率的提高，利用 IPv6 地址空间

分配系统做为网络监控手段需要更大范围尽可能收集节点及其链路信息。因此系统探测点选择和设计成为了主要关注点之一。

现有拓扑发现系统，往往将探测模块和系统紧耦合，以致单点探测遇到性能的瓶颈。本系统将 Web Services 作为模块间通讯的基础，将探测模块和拓扑控制系统分开，并将前者作为并行探测的一部分。拓扑发现系统平台可使用系统内每个基本单位的探测节点，见图 19：主拓扑发现平台利用本单位和次级单位的探测节点构成对本层网络的多点或并行探测，能提高拓扑探测的效率，并将进一步提高拓扑发现的准确性。

将构造地址集合中地址随机抽取均匀发送到各个探测节点进行托探测。见下图 19，系统将拓扑任务均匀得分配给 A、B、C 节点，并使得 A、B、C 节点分别都实行一定的并行线程数。系统保持对 A、B、C 节点控制。拓扑发现的性能因此提高了三倍，如果合理得加入其它节点，拓扑发现性能因此能进一步得到提高。

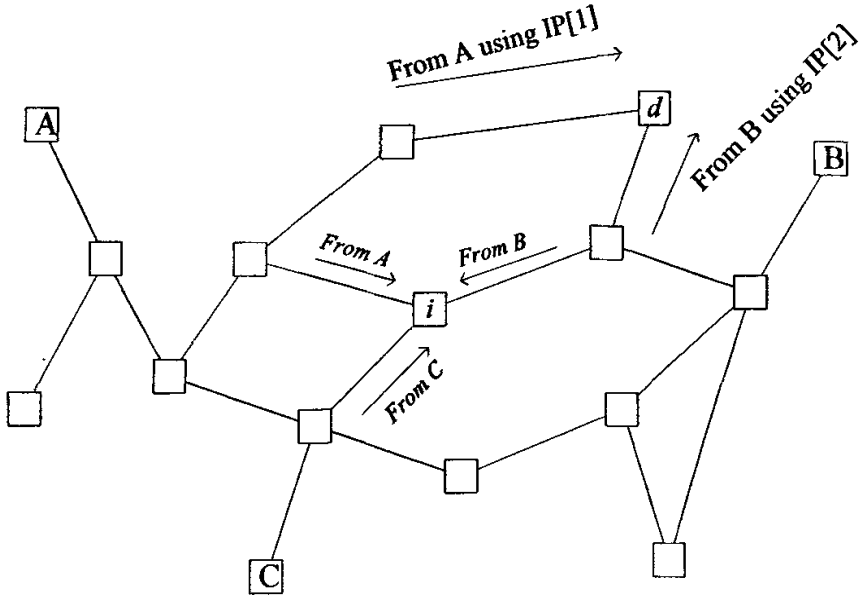


图 19　多点并行探测示例

这里就有拓扑发现准确度的疑问：

（1）　节点发现度是否下降？

因为多点探测所用构造的地址集合和单点探测地址集合相同，并且均匀分配给了各个探测点，节点发现度没有任何下降。例如图 5 中的点 d，因为均为分配节点的测量，在整个探测周期中，可能分别接受来自节点 A、B、C 的探测报文，同理对于网络中任何的节点，整个探测过程都将和单点探测一样将其覆盖。因此可以很确定地得到节点发现度没有任何下降的结论。

（2）　链路发现度是否下降？

IPv6 网络的地址利用率可见的将来都不会超过 10%，因此均匀得在节点间分配地址进行拓扑发现，不仅完全没有降低链路发现的程度，反而构成了近似

的多点探测模式，并很大程度上减少了重复拓扑发现，降低了最后拓扑数据合并的难度。再参考图 5 中的中间节点 i，其在探测过程中分别很有可能接收不同探测节点从不同路径发过来的探测报文，这种情况类似于多点探测模式，因此链路发现度得到了一定程度的提高。

特别需要说明的是：多点并行链路发现度的保证是建立在 IPv6 网络地址利用率相对较低的基础上的，均为的地址分配策略为探测提供了多方向探测的便利，并在此基础上不损失探测的覆盖度，提高了准确性和探测效率。其次，在对目标网络进行监控时，对于未使用的地址空间的探测将较低频率地持续地发起报文探测，并将其分配给不同的探测点，不仅能收集到更多的路径信息，而且有助于提高对目标网络的监控粒度和时效性，有效降低因为探测频率提高而对网络负载产生的影响。

4.5 后续路径探测和确认

经过一定规模基于 Traceroute 的拓扑发现探测后，被探测的目标网络基本已经查明。因此单点探测使得最后拓扑探测关系呈现树状拓扑，即使采用多点探测解决了部分“cross link”问题，但是由于无法进行充分的多点探测在骨干网络中仍然存在部分的“cross link”现象，如下图 3 中虚线链路。

在 IPv4 环境中，网络上支持源路由选项的路由器非常少，不到 8%【6】（2000 年），并且随着安全措施的加强将进一步减少，无法在大规模拓扑测量中使用。因为源路由选择的不充分设置，造成 IPv4 网络中的不稳定路由现现象。因此，导致只能通过大量发生探测报文来判断动态路由才存在，而此方法在实际使用中往往不可行，带来过多的错误信息。

在 IPv6 环境中，我们结合 IP 层源路由选项迫使探测报文通过特定的已知路径。如下图 20 中，规定探测路径经过 v 到达 D，此时未知节点 i 和虚线代表的未知路径都将被发现。通过这种方法，基本上解决了“cross link”和动态路由的问题。特别需要说明的是，为了减少不必要的数据，特别是匿名端口的问题，我们只针对符合特定前缀的节点进行探测。比如在移动 CNGI 网络使用符合 2001:e80:ffff::/48 的骨干网前缀，Cernet2 使用 2001:da8:1::/48 骨干网前缀利用源路由机制进行链路探测。

在对目标网络进行初步拓扑发现后，需要对获取地址进行确认。这是因为构造的目标地址在探测过程中绝大部分不可达且由于高密度探测和动态路由机制的存在导致其探测路径的偏差；其次由于小部分地址可能为匿名路由设备自动生成其不具备可达性。因此有必要对获取地址进行必要的确认，为别名解析等步骤提供方便。特别需要说明的是，地址确认应该在发起别名解析等需要确切路径信息（跳数）的节点进行。

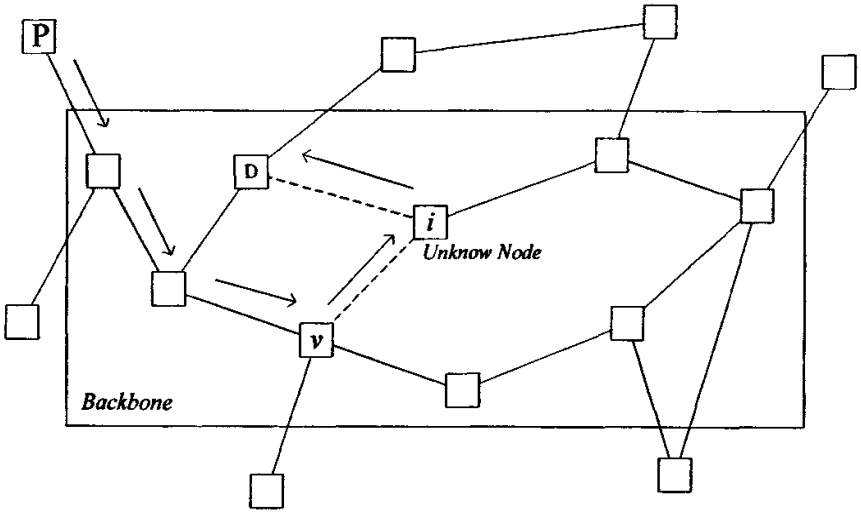


图 20　利用源路由选项进行路径探测示例

4.6　本章小结

本章首先介绍了典型的几种目标地址获取方式：从 BGP Routing Table 中得到 AS 级地址前缀信息，构造相应 IPv6 地址作为地址列表的来源进行 Traceroute 探测；从黄页网站或 6bone 等得到已注册数据作为地址列表的来源；根据 IPv6 地址前缀信息，利用协议聚合的特性，在一定地址前缀内构造用于拓扑探测 IPv6 地址列表；种子列表；等等。

接着，我们介绍了将 IPv6 地址空间分配模块和 IPv6 拓扑发现系统相结合。IPv6 拓扑发现系统从 IPv6 地址空间分配模块中得到已记录的地址空间前缀进行目标地址构造；并对未使用的地址空间，根据 IPv6 地址聚会和现有子网地址前缀长度采取一定随机性构造目标地址来确定所探测的空白地址空间是否在实际网络中被使用；具体对目标网络进行监控是个动态并持续的过程，对于不同类别的地址应采用不同处理方式。首先对于骨干网节点和已注册节点应进行维护，维护这个集合的准确性，增加其中地址或者确认其存活状态；其次对于未使用的地址空间的探测将较低频率地持续地发起报文探测，当发现未注册地址时进行记录，并在以后的探测过程中增加对其探测的频率，直到其注册状态改变；从而大大提高对目标网络的针对性和探测效率，起到监测网络拓扑的作用。

第三节我们介绍了 IPv6 拓扑发现系统的设计：系统基本单位由探测节点、拓扑发现平台和本网拓扑数据模块构成。完备的自动拓扑发现系统由若干个基本单位构成，其中的一个拓扑发现平台作为主平台负责所有拓扑信息的合成。系统将 Web Services 作为模块间通讯的基础，解析了拓扑发现系统的三个主要模块，降低了探测、数据获取和管理三者间的耦合度。

第四节结合现有 IPv6 网络的特点提出了多点并行探测的方式：系统将整个探测任务发送给不同探测节点同时进行（为探测过程提供一个以上的并行探测点，并且每个探测点同时并行进行网络探测），因此可以成倍数得提高探测并行数。系统将探测模块和拓扑

控制分开，将进一步提高拓扑发现的准确性和效率；并对可能遇到的疑问做了分析，比如节点发现度是否下降和链路发现度是否下降等。

最后介绍了利用源路由机制的路径探测并特别说明了地址确认的必要性：在对目标网络进行初步拓扑发现后，需要对获取地址进行确认，为别名解析等步骤提供方便。其应该在发起别名解析等需要确切路径信息（跳数）的节点进行。

第五章 实验数据分析

本系统针对中国移动 CNGI 和 Cernet2 进行实际探测实验。

5.1 相关环境及其参数

至今本 IPv6 拓扑发现系统的开发语言主要为 Java 和 C++；C++语言主要用于报文的发送与接收，然后使用 JNI 使得 Java 代码和 C++代码进行交互。系统其它部分开发都使用 Java 语言。

由于现今各个操作系统对 IPv6 网络的支持情况，特别是套接字编程中源路由选项的支持，本系统探测节点主要使用 Free BSD 系统，Free BSD 系统遵循 POSIX 规范，开源并且稳定；在 IPv6 套接字编程上提供了充足的接口，而 Linux 系统至今没有加入支持 IPv6 源路由选项的库函数，对其的支持需要额外的 patch。

本 IPv6 拓扑发现系统使用 apache 下面的 axis 项目实现对 Web Services 的支持。利用 MySQL 及其提供的库函数实现数据的表达和存储。此外没有使用其它第三方库函数，因此很好地保证了拓扑发现系统的跨平台特性。

中国移动 CNGI 网络地址空间前缀为 2001:e80::/32，分配给各地的子网地址前缀为 41 位；Cernet2 网络地址空间前缀为 2001:da8::/32，子网前缀一般为 48 位；中国移动 CNGI 骨干网地址前缀为 2001:e80:ffff::/48，Cernet2 骨干网前缀为 2001:da8:1::/48。

5.2 实验数据及其分析

对中国移动 CNGI 骨干网进行实际探测时，系统在 2001:e80::/32 空间内使用 41 位前缀长度构造目标地址进行探测，并对其骨干网(2001:e80:ffff::/48)使用源路由机制进行了链路探测，实验结果如下：

试验步骤	节点	链路	路由多址（对）
目标地址探测	38	42	0
地址确认和路由多址判断	38	58	17
源路由路径探测	38	98	17
第二次地址确认和路由多址判断	38	98	17

表 1 CNGI 拓扑发现结果

由于中国移动 CNGI 网络还处于试验阶段其上没有大规模开展应用，并且其少数几个接入子网都采取较严格管理不回送 ICMPv6 报文所以现在只能探测其骨干网。

和实际路由器端口地址及其链接情况得到如下结果：地址发现率 100%，链路发现率 100%，路由器多址发现率误差 1/19。误差设备为南京节点设备二（见表格 16），根据

相关测试显示此设备根据 RFC2463，路由器被配置报文处理过程为“采用收到报文的目的地址作为响应报文的源地址”，并且报文处理模型是先判断跳数值后处理源路由选项因此不能使用带有源路由报头高端口 UDP 探测报文，所以造成误差。其次关于源路由路径探测后链路数量增加的问题（表 1 中从 58 增加到 98），这是由于使用源路由选项进行链路探测，使部分路由器回送的 ICMPv6 源地址多样化，多样化的源地址由于是同一路由器不同端口地址所以并不影响最后路由器级设备之间的链接关系及设备地址的标识。

以上结果完全能符合既定目标，能很好得满足网管要求。

以下是中国移动实际拓扑图：

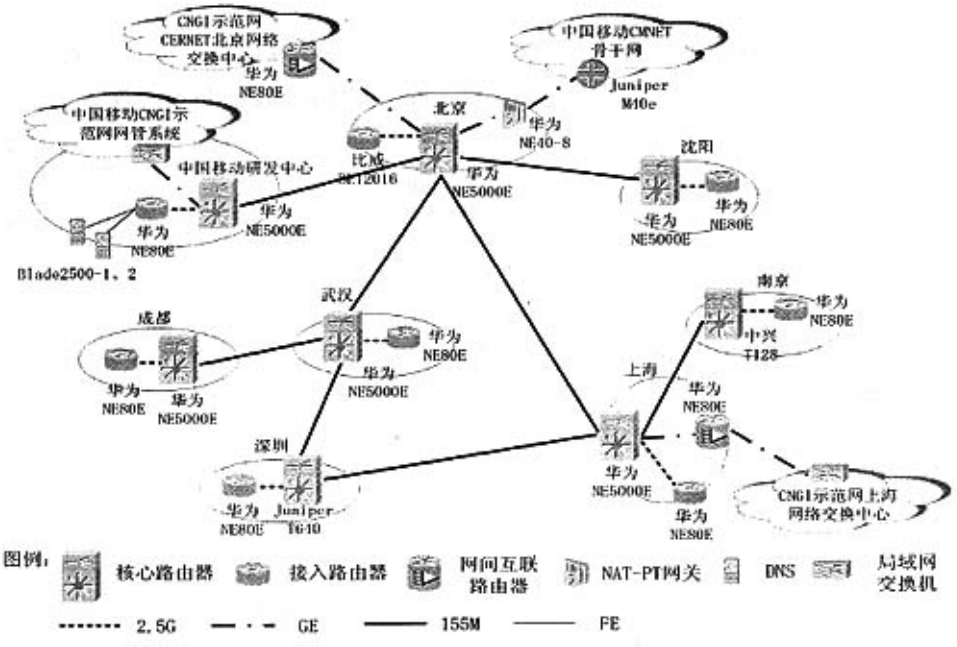


图 21　中国移动 CNGI 骨干网拓扑图

以下给出移动骨干网实际网络配置：

北京节点设备端口配置：

设备一：

本端设备地址	对端设备地址
2001:0E80:FFFF::1/126	2001:0E80:FFFF::2/126
2001:0E80:FFFF::5/126	2001:0E80:FFFF::6/126
2001:0E80:FFFF::9/126	2001:0E80:FFFF::A/126
2001:0E80:FFFF::D/126	2001:0E80:FFFF::E/126
2001:0E80:FFFF::11/126	2001:0E80:FFFF::12/126
2001:0E80:FFFF::15/126	2001:0E80:FFFF::16/126
2001:0E80:FFFF::19/126	2001:0E80:FFFF::1A/126

表 2　北京节点设备一配置

设备二：

本端设备地址	对端设备地址
2001:0E80:FFFF::1A/126	2001:0E80:FFFF::19/126

表 3 北京节点设备二配置

设备三：

本端设备地址	对端设备地址
2001:0E80:FFFF::6/126	2001:0E80:FFFF::5/126

表 4 北京节点设备三配置

设备四：

本端设备地址	对端设备地址
2001:0E80:FFFF::2/126	2001:0E80:FFFF::1/126

表 5 北京节点设备四配置

北京研发中心设备端口配置：

设备一：

本端设备地址	对端设备地址
2001:0E80:FFFF::A/126	2001:0E80:FFFF::9/126
2001:0E80:FFFF::1D/126	2001:0E80:FFFF::1E/126

表 6 北京研发中心设备一配置

设备二：

本端设备地址	对端设备地址
2001:0E80:FFFF::21/126	2001:0E80:FFFF::22/126
2001:0E80:FFFF::25/126	2001:0E80:FFFF::26/126
2001:0E80:FFFF::1E/126	2001:0E80:FFFF::1D/126

表 7 北京研发中心设备二配置

上海节点设备端口配置：

设备一：

本端设备地址	对端设备地址
2001:0E80:FFFF::2D/126	2001:0E80:FFFF::2E/126
2001:0E80:FFFF::E/126	2001:0E80:FFFF::D/126
2001:0E80:FFFF::31/126	2001:0E80:FFFF::32/126
2001:0E80:FFFF::35/126	2001:0E80:FFFF::36/126
2001:0E80:FFFF::39/126	2001:0E80:FFFF::3A/126

表 8 上海节点设备一配置

设备二：

本端设备地址	对端设备地址
2001:0E80:FFFF::3A/126	2001:0E80:FFFF::39/126

表 9 上海节点设备二配置

设备三：

本端设备地址	对端设备地址
2001:0E80:FFFF::2E/126	2001:0E80:FFFF::2D/126

表 10　上海节点设备三配置

武汉节点设备端口配置：

设备一：

本端设备地址	对端设备地址
2001:0E80:FFFF::12/126	2001:0E80:FFFF::11/126
2001:0E80:FFFF::41/126	2001:0E80:FFFF::42/126
2001:0E80:FFFF::45/126	2001:0E80:FFFF::46/126
2001:0E80:FFFF::49/126	2001:0E80:FFFF::4A/126

表 11　武汉节点设备一配置

设备二：

本端设备地址	对端设备地址
2001:0E80:FFFF::4A/126	2001:0E80:FFFF::49/126

表 12　武汉节点设备二配置

沈阳节点设备端口配置：

设备一：

本端设备地址	对端设备地址
2001:0E80:FFFF::16/126	2001:0E80:FFFF::15/126
2001:0E80:FFFF::29/126	2001:0E80:FFFF::2A/126

表 13　沈阳节点设备一配置

设备二：

本端设备地址	对端设备地址
2001:0E80:FFFF::2A/126	2001:0E80:FFFF::29/126

表 14　沈阳节点设备二配置

南京节点设备端口配置：

设备一：

本端设备地址	对端设备地址
2001:0E80:FFFF::3D/126	2001:0E80:FFFF::3E/126
2001:0E80:FFFF::36/126	2001:0E80:FFFF::35/126

表 15　南京节点设备一配置

设备二：

本端设备地址	对端设备地址
2001:0E80:FFFF::3E/126	2001:0E80:FFFF::3D/126

表 16　南京节点设备二配置

成都节点设备端口配置：

设备一：

本端设备地址	对端设备地址
2001:0E80:FFFF::46/126	2001:0E80:FFFF::50/126
2001:0E80:FFFF::4D/126	2001:0E80:FFFF::5E/126

表 17　成都节点设备一配置

设备二：

本端设备地址	对端设备地址
2001:0E80:FFFF::4E/126	2001:0E80:FFFF::4D/126

表 18 成都节点设备二配置

深圳节点设备端口配置：

设备一：

本端设备地址	对端设备地址
2001:0E80:FFFF::51/126	2001:0E80:FFFF::52/126
2001:0E80:FFFF::42/126	2001:0E80:FFFF::41/126
2001:0E80:FFFF::32/126	2001:0E80:FFFF::31/126

表 19 深圳节点设备一配置

设备二：

本端设备地址	对端设备地址
2001:0E80:FFFF::52/126	2001:0E80:FFFF::51/126

表 20 深圳节点设备二配置

针对对 Cernet2，系统在 2001:da8::/32 空间内使用 48 位前缀长度构造目标地址进行探测，并对其骨干网(2001:da8:1::/48)使用源路由机制进行了链路发现。

Cernet2 按照核心节点的接入能力以及在网络互联中发挥的作用不同，将核心节点分为两类，即一级节点和普通节点。在 CERNET2 分布的 20 个城市的 25 个核心节点中，北京－清华、上海－交大、广州、南京和武汉为一级节点，其他 20 个为普通节点。

由于北京和上海的高校比较集中，同时，北京又是 CERNET2 网络中心和 CNGI－6IX 所在地，因此，在北京和上海分别采用了分布式的设计方案。北京节点 分布在清华大学、北京大学、北京邮电大学和北京航空航天大学，基于已有的光纤传输基础设施，采用 2.5Gbps/10Gbps 组网技术，连接成环型拓扑结 构，构成分布式接入层，并以 2.5Gbps/10Gbps 接入核心层。北京－清华节点为一级节点，其余为普通节点。上海节点包括上海交通大学、复旦大学和 同济大学共同构成分布式的核心节点。其中，上海交大为一级节点，其余为普通节点。

以下是 Cernet2 公布的网络拓扑图：

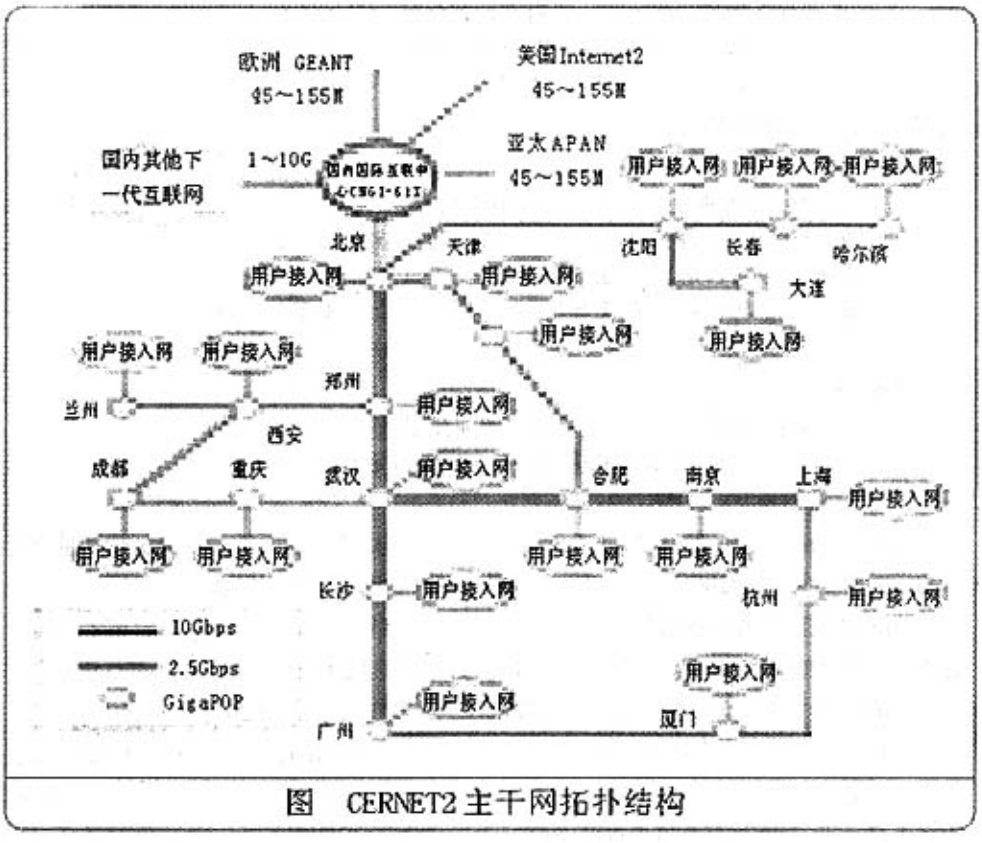


图 CERNET2 主干网拓扑结构

图 22 中国教育网 Cernet2 主干网拓扑结构

实验结果如下：

试验步骤	节点	链路	路由多址(对)
目标地址探测	139	158	0
地址确认和路由多址判断	146	188	21
源路由路径探测	159	254	21
第二次地址确认和路由多址判断	159	261	29

表 21 Cernet2 拓扑发现结果

在目标地址探测中，共计 65536 次类 Traceroute 过程。结果中到达了目标地址对应前缀的数量为 95，基本符合 Cernet2 目前高校接入数。经过源路由探测后骨干网地址从 44 增加到了 64。在地址确认和路由多址判断后，一共得到 35 个骨干路由器节点，经过 DNS 解析发现基本和实际情况相符。

5.3 实验小结

实验数据表明：

1. 系统很好得解决了主要网络问题，确保了拓扑发现的准确性，特别是基于 UDP 端口不可达报文的方法和源路由机制在解决IPv6路由别名的问题中得到了很好的效果
2. 地址确认步骤的必要性：其剔除了网络中的伪地址，保证了节点相对探测点距离的准确性，为解决别名问题提供了方便
3. 对符合（骨干网）前缀的已探明地址使用源路由进行链路探测，可以避免不必要数据的获得，提高链路探测效率，还可以解决不稳定路由问题
4. 路由器级拓扑发现探测并不能反映实际端口之间链接关系，只能反响网络中路由设备的 IP 地址及路由设备之间的链接关系

第六章 结束语

随着 IPv6 网络的快速发展，在逐渐成熟的协议体系上的应用将越来越丰富，必将带动新一代网络的普及和发展。本文通过深入分析 IPv6 网络及其拓扑发现技术的关键问题，保证最终拓扑信息的正确性和完整性，并提高拓扑发现的效率。本文提出网络环境关键问题解决方案，在此基础上利用 Web Services 技术设计和实现了 IPv6 网络拓扑自动发现系统。

6.1 本文工作总结

本文的工组成果主要分为如下几个部分：

1. 实际 IPv6 网络环境问题的解决：

   本系统充分利用传统 IPv4 网络拓扑发现方法结合 IPv6 网络特点解决了一系列网络环境问题：中间路由报文限制，路由循环，别名解析，匿名端口，不稳定路由。本拓扑发现系统对主要网络现象的处理方法有效得减少了发现误差，提高了拓扑发现的准确性。

2. 扩大探测目标地址来源，扩展了 IPv6 拓扑发现系统的应用

   本文系统支持多种探测目标地址来源。不仅支持典型种子节点列表（seeds list）和根据地址空间前缀构造探测地址，而且结合 IPv6 地址管理模块作为目标地址集合来源。将 IPv6 地址空间分配模块作为 IPv6 拓扑发现系统可靠探测目标地址获取的途径：从中得到已分配的地址空间前缀进行目标地址构造，并对未使用的地址空间根据 IPv6 地址聚会和现有子网地址前缀长度采取一定随机性构造目标地址来确定所探测的空白地址空间是否在实际网络中被使用。

   本文基于 ICMP 协议的拓扑发现系统相对于管理域内基于 SNMP 协议的系统在准确度和时效性方面仍有一定的差距，但其探测性能的提高，特别是无需设备管理权限的优点，使其能成为管理域内拓扑发现及管理系统的有益补充，并能在一定程度上替代基于 SNMP 协议的拓扑发现系统。系统将 IPv6 地址空间分配作为探测目标地址获取途径，为基于 ICMP 拓扑发现系统应用于管理域内提供了基础，提高了对目标网络的探测效率和起到监测管理域网络拓扑的作用。

3. 充分利用源路由选项

   系统将源路由选项应用在路径探测和别名处理过程中。在实际探测过程中对符合特定前缀的 IPv6 地址利用源路由机制进行路径探测，比如针对移动 CNGI 网络我们制定的是 2001:e80:ffff:，针对 Cernet2 制定的为 2001:da8:1:；在别名处理过程中，用带有源路由选项的 ICMPv6 探测报文确定路由设备报文处理模

型，并用带有源路由选项的 UDP 探测报文确认路由别名现象等。充分利用 IPv6 源路由选项为我们在 IPv6 网络环境下解决路由别名和不稳定路由提供了基础，提高了最终拓扑数据的准确性和对目标网络的覆盖度。

4. 基于 Web Services 构建拓扑发现系统，为拓扑效率的提高奠定了基础

将 Web Services 作为通讯的基础，解析了拓扑发现系统的三个主要模块，降低了探测、数据获取和管理三者间的耦合度。拓扑发现系统平台可使用系统内每个基本单位的探测节点，利用本单位和次级单位的探测节点构成对本层网络的多点或并行探测，通过合理的布置并行探测点，可以大大提高探测效率和准确性； Web Services 技术的应用，为拓扑数据获取提供了统一的接口并解决拓扑关系层次属性的表达；整个系统有较强的适应性，满足了不同网络环境对管理的需求。例如，对拓扑关系的表示完全可以脱离探测和拓扑管理平台，而只采用数据获取模块；或脱离数据获取或拓扑管理平台，而只为其他平台提供探测点等等。

6.2 下一步研究方向

以下是可以进一步开展研究工作的几个方向：

1. 进一步完善和改进 IPv6 拓扑发现系统

本文开发的系统由于较少的实验环境和缺乏真实数据的对比，存在进一步改进和完善的余地。因此下一步工作应该进一步和网络管理单位合作，用实践来检验和完善本系统。

2. 准确识别骨干网中的 Tunnel

本系统主要侧重点为发现 IPv6 网络的拓扑结构，目前还没有加入识别 IPv4-to-IPv6 和 IPv6-to-IPv4 Tunnel 的功能，无法标识骨干网中两个路由器之间的连接是通过 Tunnel 连接还是直接连接。从 IPv4 向 IPv6 的过渡将通过很长一段时间，在过渡期间，通过 Tunnel 将各 IPv6 站点相连是主要的过渡方法，因此能够发现骨干网内哪些路由器之间的连接是 Tunnel 将有益于了解整个网络的拓扑结构，能够体现出从 IPv4 向 IPv6 过渡期间网络结构的变化情况。

3. 实现更具实时性的网络拓扑监控工具

SNMP 和路由协议（如 OSPF、BGP）的最大优点是信息自动随网络的状况更新，这样通过这些手段获取的拓扑信息总是反映网络最新的状况，能保证很好的实时性，其优点对网络拓扑监控特别具有吸引力。但其缺点是并不是所有设备都支持这些协议，特别是通过这些手段获取拓扑信息都需要相应路由器管理和访问权限。针对不同的网络拓扑发现或监控要求，可以利用以上工具开发相应的拓扑发现和监控工具来反映实际网络状况。

4. 结合动态拓扑呈现手段

拓扑呈现和展示是必要的组件之一，本系统由于有上层系统的呈现接口所以现在只实现了提供拓扑数据的功能。现在有很多关于拓扑呈现和展示的项目和方案，特别有些项目具备动态生成拓扑关系功能。引入优秀的拓扑呈现手段将进一步提高拓扑发现系统对网络拓扑关系的表达，提高对网络监控的能力，因此将这些项目和方案引入本系统也是下一步的工作。

参考文献

[1] CAIDA Organization. http://www.caida.org

[2] Daniel G W, Fangzhe C, Ramesh V et al. Topology Discovery for Public IPv6 Networks. ACM SIGCOMM Computer Communications Review, 2003; 33(3): 59-68.

[3] Astic I, Festor O. A hierarchical topology Discovery Service for IPv6 network. In: Network Operations and Management Symposium (NOMS), 2002: 497-510.

[4] B. Cheswick, H. Burch, and S. Branigan. Mapping and Visualizing the Internet. In: *Proc. 2000 USENIX Annual Technical Conf.*, June 2000: 1-12.

[5] J. Pansiot and D. Grad, On routes and multicast trees in the Internet. In: *ACM Computer Communications Review*, vol. 28, no. 1, 1998.

[6] R. Govindan and H. Tangmunarunkit. Heuristics for Internet map discovery. In: *Proc. INFOCOM 2000*, March 2000. 1371-1380.

[7] Y. Breitbart, M. Garofalakis, C. Martin, R. Rastogi, S. Seshadri, and A. Silberschatz. Topology discovery in heterogeneous IP networks. In: *Proc INFOCOM 2000*, March 2000. 265-274.

[8] R. Hinden, S. Deering. IPv6 addressing architecture. RFC 2373, July 1998.

[9] APNIC, ARIN, RIPE NCC. http://www.ripe.net/docs/IPv6policy.html.

[10] B. Yao, R. Viswanathan, F. Chang, and D. Waddington. Topology inference in the presence of anonymous routers. In: INFOCOM 2003, San Francisco, Mar 2003.

[11] V. Paxon. End-to-end routing behavior in the Internet. In: IEEE/ACM Trans.on Networking, 1997, 5 (5): 601-615.

[12] G. Mansfield, M. Ouchi, K. Jayanthi, Y. Kimura, K. Ohta, and Y. Nemoto. Techniques for automated network map generation using SNMP. In: *Proc. INFOCOM 1996*, April 1996: 473-480.

[13] Gelnn K.SNMP in IPv6 Context. In:proc of Symposium on Applications and the Internet Workshops, 2003: 254-257

[14] 李云琪, 杨家海. 一个面向 IPv6 的网络拓扑管理系统的实现. 计算机工程与应用, 2004, 40(29).

[15] P. Francis, S. Jamin, C. Jin, Y. Jin, D. Raz, Y. Shavitt, and L. Zhang. IDMaps: A global internet host distance estimation service. In: *IEEE/ACM Transactions on Networking*,vol. 9, no. 5, 2001.

[16] N.Spring, R.Mahajan and D.Wetherall. Measuring isp topologies with rocketfuel. In: Proceeding of INFOCOM 2002, Oct, 2003.17.

[17] 姜誉, Internet 路由器级拓扑测量与分析技术研究, 哈尔滨工业大学博士论文, 2005

[18] Lorenzo Colitti, Giusepps Di Battista and Maurizio Patrignani. IPv6-in-IPv4 tunnel discovery: methods and experimental results. In: IEEE Transcations on Network and Service Management, 1(1), April, 2004.

[19] 6bone. www.6bone.net

[20] A.Broido and kc claffy. Internet Topology: Connectivity of IP Graphs. In: Proc. 2001 SPIE International Symp. on Convergence of IT and Communication (SPIE ITcom) Workshop on Scalability and Traffic Control in IP Networks, Denver, Colorado, Aug. 19-24, 2001, 4526: 172–187

[21] Z. Q. Mao, J. Rexford, J. Wang, and R. H. Katz. Towards and Accurate AS-Level Tracerourte Tool. In: Proc. ACM SIGCOMM 2003, Karsruhe, Germany, Aug. 2003, New York: ACM Press, 365-378

[22] Z. Q. Mao, D. Johnson, J. Rexford, J. Wang, and R. H. Katz. Scalable and Accurate Identification of AS-Level Forwarding Path. In: Proc. IEEE INFOCOM 2004, Hong Kong, Mar. 7-11, 2004, 3: 1605-1615

[23] IAB, IESG. Recommendations on IPv6 Address Allocations to Sites. RFC 2373, September 2001.
[24] A. Lakhina, J. W. Byers, M. Crovella, and P. Xie. Sampling Biases in IP Topology Measurement. In: Proc. IEEE INFOCOM 2003, San Francisco, California, 30 Mar. -3 Apr. 2003, 1:332-341
[25] Z. Q. Mao, J. Rexford, J. Wang, and R. H. Kartz. Towards an Accurate AS-Level Traceroute Tool. In: Proc. ACM SIGCOMM 2003, Karlsruhe, Germany, Aug. 2003, New York: ACM Press, 365-378
[26] D. Alderson, J. Doyle, R. Govindan, and W. Willinger. Toward an Optinimztion-Driven Framework for Designing and Generating Realistic Internet Topologies. ACM SIGCOMM CCR. 2003, 33(1): 41-46
[27] V. Jacobson. Traceroute(1988), Available: ftp://ftp.ee.lbl.gov/Traceroute.tar.Z
[28] J. Rickard. Mapping the Internet with Traceroute. BoardWatch Magazine 1996, (12)
[29] H. Burch and B. Cheswick. Mapping the Internet. IEEE Computer. 1999, 32(4): 97-98, 102
[30] B. Cheswick, H. Burch, and S. Branigan. Mapping and Visualizing the Internet. In: Proc. 2000 USENIX Annual Technical Conf., San Diego, California, June 2000, 1-12
[31] Internet Mapping Project. Available URL: http://research.lumeta.com/ches/db
[32] R. Siamwalla, R. Sharma and S.Keshav. Discovering Internet Topology. Technical Report, CS Dept., Cornell University, July 1998. Available URL: http://www.cs.cornell.edu/skeshav/paper/discovery.pdf
[33] A. Lakhina, J. W. Byers, M. Crovella, I. Matta. On the Geographic Location of Internet Resources. InL Proc. 2nd ACM SIGCOMM Workshop on Internet Measurement(IMW 2002), Marseille, France, Nov. 6-8, 2002, 249-250
[34] S. Floyd and E. Kohler. Internet Research Needs Better Models. ACM SIGCOMM Computer Communication Review (CCR). 2003, 33(1): 29-34
[35] E. W. Zegura, K. L. Calvert, and M. J. Donahoo. A Quantitative Comparison of Graph-Based Models for Internet Topology. IEEE/ACM Transactions on Networking. 1997, 5(6): 770-783
[36] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Network Topology Cenerators: Degree-based vs. Structural. ACM SIGCOMM Computer Communication Review (CCR). 2002, 32(4): 147-159
[37] 张宇，张宏莉，方滨兴. Inernet 拓扑建模综述. 软件学报. 2004, 15(8): 1220-1226
[38] C. Labovitz, A. Ahuja, and A. Bose. Delayed Internet Routing Convergence. ACM SIGGCOMM CCR. 2000, 30(4): 175-187
[39] C. Labovitz, A. Ahuja, R. Wattenhofer, and S. Venkatachary. The Impact of Internet Policy and Topology on Delayed Routing Convergence. In:Proc. IEEE INFOCOMM 2001, Anchorage, Alaska, Apr. 22-26, 2001, 1: 537-546
[40] L. X Gao and J. Rexford. Stable Internet Routing Without Global Coordination. ACM SIGMETRICS Performance Evaluation Review. 2000, 28(1): 307-317
[41] 赵邑新，尹霞，吴建平，于滨. 域间路由错误管理. 清华大学学报（自然科学版）. 2002, 42(1): 60-63
[42] P. Radoslavov, R. Govindan, and D. Estrin. Topology-Informed Internet Replica Placement. Computer Communications. 2002, 25(4): 384-392
[43] L. Subramanian, S. Agarwal, J. Rexford, and R. H. Katz. Characterizing the Internet Hierarchy from Multiple Vantage Points. In: Proc. IEEE INFOCM 2002, Hilton, New York, June 2002, 2: 618-627
[44] A. Reddy, D. Estrin, and R. Govindan. Large-Scale Fault Isolation. IEEE Journal on Selected Areas in Communication. 2000, 18(5): 733-743
[45] R. Albert, H. Jeong, and A. L. Barabasi. Error and Attack Tolerance of Complex Networks. Nature. 2000, 406(6794): 378-382

[46] H.F. Lipson and D. A. Fisher. Survivability – A New Technical and Business Perspective on Security. In: Proc. 1999 ACM Workshop on New Security Paradigms, Ontario, Canada, Sept. 21-24, 1999, 33-39

[47] B. Huffaker, D. Plummer, D. Moore, and k claffy. Topology Discovery by Active Probing. In: Proc. 2002 IEEE Symposium on Applications and the Internet Workshops(SAINT'02w), Nara, Japan, 28 Jan. -1 Feb. 2002, 90-96

[48] P. Barford, A. Bestavros, J. Byers, and M. Crovella. On the Marginal Utility of Network Topology Measurements. In: Proc. 1st ACM SIGCOMM Workshop on Internet Measurement(IMW 2001), San Francisco, California, Nov. 2001, 5-17

[49] N. Spring, M. Dontcheva, M. Rodrig, and D. Wecherall. How to Resolve IP Aliases. Technical Report #04-05-04, Dept. of Computer Science and Engineering, the University of Washington, May 2004, 1-13

[50] 郑海，张国清. 物理网络拓扑发现算法的研究. 计算机研究与发展. 2002,39(3): 264-268

致 谢

本文的工作是在张国清副研究员的悉心指导下完成的，他渊博的学识、敏捷的思路、严谨的治学作风和创新精神深深地影响了我。他的严格要求、谆谆教诲、热情关怀使我受益匪浅，在此谨向尊敬的导师表示衷心的感谢。

在整个课题工作与论文撰写的过程中，得到了张国强师兄的热心指导和帮助。他兢兢业业的工作态度、丰富的实际工程经验给我留下了深刻的印象，在此深表谢意。

课题组的阙伟科、范晶在课题研究期间给予我很多宝贵建议和帮助，在此表示真诚的感谢。她们的支持和合作使我的论文工作得以顺利进行。

感谢网络管理室的所有工作人员（傅川，李彦军，曹重英，屈春河，苏爱华，何斌，袁斌，陆彦斌，王迪，魏郑浩，周志勇，秦卓琼，杨清峰，覃涛），中科院计算所和研究生院学生处各位老师提供的良好的研究环境及诸多方便。

最后，我将最深挚的感激献给我的家人，我今天的成果也凝聚着他们的辛勤和汗水。

作者简历

姓名：陈韩林　　性别：男　　出生日期：1981.9.18　　籍贯：浙江诸暨

2004.9 -- 2007.7　　中国科学院计算技术研究所，硕士研究生
2000.9 -- 2004.7　　浙江大学计算机科学与技术系，本科，学士学位

【攻读硕士学位期间发表的论文】

[1] 陈韩林，张国清，张国强，基于 Web Services 的 IPv6 拓扑发现系统，计算机工程与设计（已录用)

[2] 陈韩林，一种基于关键字和优先级的层次性 IPv6 网络地址分配方法，专利申请号为 200610171650.8

[3] Guoqing Zhang,Guoqiang Zhang,Hanlin Chen 'Border Router Level Graph:Another View of the Internet Topology' submitted to IEEE Globecom 2007

[4] Guoqiang Zhang,Guoqing Zhang,Hanlin Chen 'The Weighted AS Graph' submitted to IEEE Globecom 2007

【攻读硕士学位期间参加的科研项目】

[1] 下一代互连网 IPv6 中日合作子课题《面向业务和逻辑的网络服务管理》
2005 年 03 月至 2005 年 07 月

[2] 下一代互联网示范工程 2005 年研究开发、产业化及应用试验项目《CNGI 网络监控系统》（CNGI-04-7-1D）　　2005 年 03 月至 2006 年 07 月

[3] 国家自然科学基金项目《针对 Scale free 网络的紧凑路由研究》批准号 60673168