虚拟分布式 IPv6 路由器级拓扑探测模型*)

刘振山　王清贤　罗军勇

(国家数字交换系统工程技术研究中心　郑州 450002)

摘　要　提出了一种虚拟分布式 IPv6 路由器级拓扑探测模型——VDPM(Virtual Distributed Probing Model)。VDPM 探测方式既达到了分布式拓扑探测效果,又避免了高昂的部署费用和繁琐的通讯维护工作。本文详细论述了 VDPM 实现的两个关键问题:虚拟探测源的选取和探测目标点集合的构建。通过对比 VDPM 方式和纯 IPv6 单源探测方式以 Cernet2 为目标网络进行拓扑发现的结果,体现了 VDPM 作为大规模 IPv6 路由器级拓扑发现原型系统设计依据的合理性。

关键词　虚拟探测源,拓扑覆盖率,探测冗余

Virtual Distributed Probing Model for IPv6 Router-level Topology Discovery

LIU Zhen-shan　WANG Qing-xian　LUO Jun-yong

(National Digital Switching System Engineering & Technology R&D Center, Zhengzhou 450002, China)

Abstract　This paper presents VDPM for IPv6 router-lever topology discovery. VDPM discovers the topology of IPv6 networks in a distributed model and avoids costly fees for deploying and tedious works for communications. This paper also discusses two key issues of VDPM: probing-target sets construction method and virtual probing souce selection method. With the contrasts of the two results which are got from discovery for Cernet-2 with VDPM and native IPv6 access probing model, it can draw the conclusion that VDPM can guide the design of large scale of IPv6 networks topology discovery system.

Keywords　Virtual probing souce, Topology coverage, Probing redundancy

1　引言

当前用于网络管理的 IPv6 网络拓扑发现技术[1-3]要求在网络运营者配合下,设置多个探测代理来发现网络拓扑结构,此类技术无法用于广域网拓扑探测。贝尔实验室开发的 Atlas 系统采用选路控制的单源拓扑探测技术[4],从一个网络节点探测一大片 IPv6 地址空间内的网络拓扑结构,其前提条件是事先获知这片地址空间中数量充分的有效目标 IPv6 地址,显然这个前提是非实验性质的大规模网络拓扑探测无法满足的,且单源探测方式执行效率较低。现有的分布式 IPv6 拓扑发现工具,如 scamper[5]要求在全球不同的位置部署几十乃至上百台探测引擎,可以获得这些探测引擎之间的 IPv6 路由器级拓扑结构。此类技术的实现建立在高度协作的前提下,不但部署困难、代价高昂、设备之间的通讯维护任务繁重,并且在多数情况下网络运营者出于安全考虑不会予以配合。本文结合过渡时期 IPv6 网络的发展现状,提出了一种可适用于非协作前提下的 IPv6 路由器级拓扑探测模型——VDPM,并针对 VDPM 实现的两个关键问题——探测目标点集合的构建和虚拟探测源的选取进行了详细的描述。最后通过对比 VDPM 方式和纯 IPv6 单源探测方式对 Cerner-2 上均匀分布的 120 个可达性目标点构成的探测目标点集合进行拓扑发现的结果,体现了 VDPM 作为大规模 IPv6 网络拓扑发现系统设计依据的合理性。

2　VDPM 的组成及原理

VDPM 由探测源(执行探测功能的主机)、虚拟探测源(国内外科研或网络运营机构设置对外提供服务的 ISATAP 和 6to4 中继路由器)、路径数据库(存储所有已探测路径信息的数据库)三个主要部分组成,如图 1 所示。

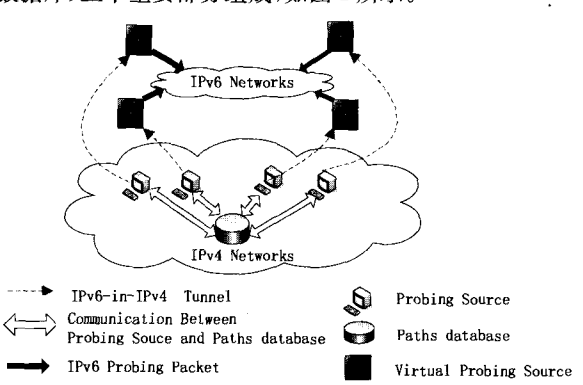

图 1　VDPM 的组成及原理

VDPM 将所有的探测源集中部署在一个 IPv4 的 C 类子网中,每一个探测源和一个虚拟探测源建立固定的 IPv6-in-IPv4 隧道。拓扑探测执行时,探测源对目标 IPv6 网络的探测报文经过 IPv4 网络的转发,投递至虚拟探测源,虚拟探测

*)基金资助:本研究得到国家 863 高技术研究发展计划资助,基金编号:2006AA01Z409。**刘振山**　博士研究生,主要研究方向为 IPv6 网络拓扑发现与网络特征分析、网络复杂度理论、非协作前提下大规模网络拓扑探测模型;**王清贤**　教授,博士生导师,主要研究方向为网络安全、计算复杂性理论;**罗军勇**　教授,硕士生导师,主要研究方向为网络安全协议分析、分布式计算。

源剥离其 IPv4 头部后再转发至探测目标点；探测目标点的应答报文经过 IPv6 网络的转发，投递至虚拟探测源，而后经虚拟探测源封装其 IPv4 头部，再转发至探测源。所有探测源在每次执行路径探测之前先要访问路径数据库，判断即将探测的路径是否已经存在。如果存在，结束此次路径探测，否则继续执行，同时将获得的路径信息实时存储进路径存储数据库中。通过这种方式可以有效地避免探测冗余。VDPM 探测方式既达到了分布式拓扑探测效果，又避免了高昂的部署费用和繁琐的通讯维护工作。

3 VDPM 实现的两个关键问题

一直以来 Internet 路由器级拓扑发现技术大都采用具有普遍适用性的 Traceroute 路径探测机制。在这种机制下探测目标点(Traceroute 的目的地址)的数量及探测源相对于目标网络的分布位置直接影响拓扑发现的覆盖率和探测冗余程度[6-9]，因而成为该领域研究的两个核心问题。

3.1 现有探测目标点集合构建方法的局限性分析

目前用于 IPv4 网络拓扑探测的目标点集合构建方法可以归纳为以下四种：

(1)启发式扫描：根据地址类别，选用不同的网络前缀长度，进行启发式扫描，将应答地址整理成探测目标点集合[10]；

(2)抽样法：在给定区间范围内，按照一定的抽样法则抽取探测地址，形成探测目标点集合[11]；

(3)基于 Web 站点采集：搜集 Web 站点信息，将 Web 站点对应的域名解析成 IP 地址，从而形成探测目标点集合[12,13]；

(4)基于 RouteView 信息：根据 RouteView 提供的 BGP 表中的目的网络和 AS_PATH 选择 IP 地址进行探测，称为有导向探测[14-16]。

IPv6 海量地址空间的特性，使得用启发式扫描来获取探测目标点的计算空间复杂度巨大，也使得基于抽样法的抽样率与覆盖率之间的矛盾无法找到折衷方案。

我们选取 RouteView 上 2007 年 7 月 1 日 00：00 至 2007 年 7 月 15 日 12：00 的所有 MPBTDs(360 个文件)和 MPBUDs(2880 个文件)数据源，并整理成表 1 格式。

表 1 IPv6 地址前缀与 AS 号的对应表

IPv6 前缀	前缀长度	AS 号
2001:1200:0:0:0:0:0:0	32	16531
2001:1210:0:0:0:0:0:0	32	2549
2001:12e8:0:0:0:0:0:0	32	16397
……	……	……

通过分析表 1 中的 IPv6 前缀信息，我们发现长度超过 48 位的 IPv6 前缀只占全部 IPv6 前缀的 5.3%；利用表 1 中的 IPv6 前缀信息进行基于前缀停止集的路径探测，得到的实验结论是：探测报文无法到达目标 IPv6 网络纵深，只能得到非常有限的拓扑覆盖率。因此基于 RouteView IPv6 前缀停止集的方法同样不适用于 IPv6 探测目标点集合的构建。

3.2 VDPM 中探测目标点集合的构建方法

由 3.1 节的分析可知要获得一定数量的有效 IPv6 探测目标点，目前唯一可行的方法是搜集足够数量 IPv6 Web 站点。我们从 Cernet2 中心(清华大学)的资源网站下载了提供的 IPv6 Web 站点信息，共获得 243 个 IPv6 Web 站点，经验证其中有 177 个 IPv6 Web 站点可以解析出对应 IPv6 地址，我们将其存储为表 2 格式。

表 2 Domian_IPv6Address Table

Domain_INFO	IPv6_ADDRESS
www.tsinghua.edu.cn	2001:da8:200:200::4:100
www.nrc.edu.cn	2001:da8:200:101::150
redweb.tsinghua.edu.cn	2001:da8:200:e288:8888:8888:8888:8888
news.tsinghua.edu.cn	2001:da8:200:200::4:28
sus6.sns.tsinghua.edu.cn	2001:da8:200:e288::89
www.ipv6.scut.edu.cn	2001:250:1800:17::57
ipv6.scnu.edu.cn	2001:251:4002:8::6
ftp6.sjtu.edu.cn	2001:da8:8000:1::80
bbs.neu6.edu.cn	2001:da8:9000:b255:200:e8ff:feb0:5c5e

显然这种通过搜集和解析 IPv6 Web 站点的方法得到的探测目标点集合是不完备的。因此在已得到的初始探测目标点集合的基础上，我们提出了一种基于前缀跨度模型(PSM，Prefix Span Model)的探测目标点集合扩建方法，该方法主要思想描述如下：

从初始的探测目标点集合中任意提取出一个探测目标点，得到该探测目标点所属的 64 位长度的桩网络前缀(绝大多数情况下分配给终端 IPv6 设备的前缀是 64 位)。通过查询公开的 IPv6 前缀分配信息，可以获知该桩网络前缀所属的顶级 IPv6 前缀长度，从而可以得出桩网络与顶级前缀之间的前缀跨度。比如 2001:da8:a5:e::/64 的所属的顶级 IPv6 前缀为 2001:da8::/32，那么前缀跨度为 32。

假定某探测目标点桩网络前缀描述为 $P_1, P_2, P_3, \cdots, P_m, P_{m+1}, P_{m+2}, \cdots, P_{64}$，$m$ 表示该前缀所属的顶级前缀长度。从第 64 位到 $m+1$ 位按照逐 1 比特递减的方法，构造出一个新的 IPv6 前缀集合：$F=\{P_1, P_2, P_3, \cdots, P_{m+1}, P_{m+2}, \cdots, Pn/n \mid m<n<64\}$，然后将 F 集合中的每个元素按照末位比特填充零的原则构造出伪地址集合 $F'=\{P_1, P_2, P_3, \cdots, P_{m+1}, P_{m+2}, \cdots, Pn, 0\cdots\cdots0 \mid m<n<64\}$，最后将 F' 中的每个元素作为 ICMPv6 请求报文的目标地址，进行遍历探测。如果 ICMPv6 应答结果返回了网络不可达(Destination net unreachable)信息，则说明该网络前缀不存在；如果某个节点返回主机不可达 (Destination host unreachable)信息，则将该节点添加至初始的目标点集合中。

说明：这种方法首先构造出桩网络到顶级前缀之间可能存在的 IPv6 前缀集合，然后利用 ICMPv6 请求和应答报文判断出这些前缀的可达性，从而将返回主机不可达节点扩充至初始的目标点集合中。通过这种方法对每一个初始探测目标点的扩建只需要 $64-m$(前缀跨度)次猜测即可。

3.3 VDPM 中虚拟探测源的选取策略

VDPM 中虚拟探测源的选取基于这样一个事实：虚拟探测源相对目标网络的分布位置决定了拓扑探测的执行效率和覆盖率。如图 2 所示，所有的探测源对目标点 H 的探测报文都要经过同一个入口点 E，因此针对 T 区域采用分布式探测和单源探测获得的拓扑覆盖率是完全一样的。

为了说明 VDPM 探测源选取策略，首先做如下定义：

定义 1 用 $T(h)$ 表示探测源 h 选取的虚拟探测源，用 $E(T(h))$ 表示 $T(h)$ 投递出的探测报文进入目标网络 T 的入口节点(在图 2 中为路由器 E)。h 发出的探测报文到达 $E(T(h))$ 的实际跳数用 $Hop(h, E(T(h)))$ 表示，h 发出的探测报文到达 $T(h)$ 的 IPv4 路径跳数用 $Hopv4(h, T(h))$ 表示，$T(h)$ 投递的探测报文到达 $E(T(h))$ 的跳数用 $Hopv6(h(S), E(T(h)))$ 表示。

定义 2 用于拓扑发现的探测源的集合用 P 表示，$P=$

(下转第 76 页)

参 考 文 献

[1] Recio R J. Server I/O networks past, present, and future// Proc. of the ACM SIGCOMM Workshop on Network-I/O Convergence: Experience, Lessons, Implications. New York: ACM Press, 2003: 163-178

[2] Intel in Communications. 10 gigabit Ethernet technology overview. White Paper, 2003. http://www.intel.com/network/connectivity/resources/doc_library/white_papers/pro10gbe_lr_sa_wp.pdf

[3] Chase J S, Gallatin A J, Yocum K G. End-System optimizations for high-speed TCP. IEEE Communications Magazine, 2001, 39(4): 68-74

[4] Wright G R, Stevens W R. TCP/IP Illustrated, Volume 2: The Implementation. Addison Wesley, 1995

[5] 章淼, 吴建平, 林闯. 互联网端到端拥塞控制研究综述[J]. 软件学报, 2002, 13(3): 354-363

[6] Yeh E, Chao H, Mannem V, et al. Introduction to TCP/IP offload engine. 2002. http://www.10gea.org/ SP0502 IntroTo TOE_F.pdf

[7] Adaptec Corporation. Advantages of a TCP/IP offload ASIC. 2004. http://graphics.adaptec.com/pdfs/ tcpio_adv_wp.pdf

[8] Minturn D, Regnier G, Krueger J, et al. Addressing TCP/IP processing challenges using the IA and IXP processors. Intel Technology Journal, 2003, 7(4): 39-50

(上接第 47 页)

$\{P_i | i=1,2,3,\cdots,n\}$, $COV(P_i, T)$表示 P_i 投递的探测报文到可以到达的进入 T 的入口点的集合。

探测源选取规则 1 如果 h 分别选择了探测源 $T(h)1$ 进行探测，满足 $E(T(h)1) \notin COV(P,T)$，则将 $T(h)1$ 添加到 P 中。

说明：$E(T(h)1) \notin COV(P,T)$，显然通过 $T(h)1$ 可以到达目标网络新的入口点，选择 $T(h)1$ 加入 P 可以提高拓扑发现的覆盖率。当 $E(T(h)1) \in COV(P,T)$时，则按照规则 2 和 3 继续选择。

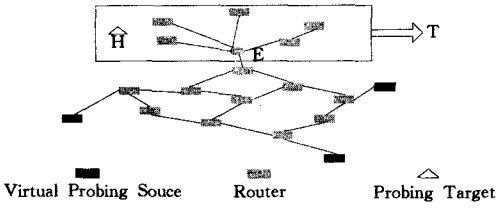

图 2 分布式拓扑探测构想图

探测源选取规则 2 如果 h 分别选择了探测跳板 $T(h)1$ 和 $T(h)2$ 进行路径探测，满足 $E(T(h)1) = E(T(h)2) \notin COV(P,T)$，且 $Hop(h, E(T(h)1)) < Hop(h, T(h)2)$则将 $T(h)1$ 添加到 P 中；

说明：当探测源选择不同的虚拟探测源进行路径探测都到达目标网络的同一个入口点时，则将其中路径最短的情况下选择的探测源加入拓扑发现的探测源集合。

探测源选取规则 3 如果 $Hop(h, E(T(h)1)) = Hop(h, T(h)2)$ 且 $Hopv4(h, T(h)1) < Hopv4(h, T(h)2)$，则将 $T(h)1$ 添加到 P 中。

说明：总路径跳数一样的情况下优先选择 IPv4 路径短的虚拟探测源添加到虚拟探测集合主要考虑到 IPv6 路由表为分级结构，而 IPv4 是扁平结构，所以在跳数相同的前提下报文在 IPv6 网络中的转发更为迅速。

4 实验及结论

为了验证 VDPM 相对单源 IPv6 拓扑探测在提高拓扑覆盖率方面的优势，我们进行了实际拓扑探测，整个过程分为两个步骤：

步骤 1 在 202.196.63.0 子网内设置了两个探测源，分别与上海交通大学的 ISATAP 路由器(2001:da8:8000:3:0:5efe:202.112.26.254)和清华大学的 ISATAP 路由器(2001:da8:900e:0:5efe:59.66.4.50)建立了 IPv6-in-IPv4 隧道；选出 Cernet-2 上分布相对均匀的 120 个可达性探测目标点作为探测目标点集合，进行拓扑探测，在未进行别名归并的前提下，发现路由器接口数为 373，发现链路数为 625。

步骤 2 从河南 Cernet-2 接入中心，前缀为 2001:da8:a5:e::/64 的链路内，设置纯 IPv6 探测源对相同的 120 个探测目标点进行拓扑探测，在未进行别名归并的前提下，发现路由器接口数为 209，发现链路数为 386。

通过对比看出 VDPM 相对单源 IPv6 拓扑探测在提高拓扑覆盖率方面有显著优势，唯一缺点是报文的应答时间过长，平均应答时间都在 500ms 以上。通过对测试数据的进一步分析，我们发现步骤 2 中只有 35.3%路由器接口地址与步骤 1 得到的相关数据重合，因此下一步考虑将纯 IPv6 接入环境下的单源探测方式补充为 VDPM 的一部份，这样可以进一步提高拓扑发现的覆盖率。在后续的工作中，我们将着重分布式探测冗余避免策略的完善，并依据 VDPM 设计出面向国家级的大规模 IPv6 网络拓扑发现原型系统。

参 考 文 献

[1] Astic I, Festor O. A hierarchical topology discovery service for IPv6 networks// IEEE/IFIP Network Operations and Management Symposium (NOMS). Apr. 2002: 497-510

[2] 李云琪, 杨家海. 一个面向 IPv6 的网络拓扑管理系统的实现[J]. 计算机工程与应用, 2004, 40(29): 73-75, 181

[3] 沈曾伟, 周刚. IPv6 接入网拓扑结构自动发现方法研究[J]. 计算机工程, 2006, 19: 136-138

[4] Waddington D G, Chang Fangzhe, Ramesh V, et al. Topology Discoveryfor Public IPv6 Networks[J]. ACM SIGCOMM Computer Communications Review, 2003, 33(3): 59-68

[5] IPv6 Scamper. http://www.wand.net.nz/~mjl12/ipv6-scamper/

[6] Mao Z Q, Rexford J, Wang J. Towards an Accurate AS-Level TracerouteTool // Proc. ACM SIGCOMM 2003. Karlsruhe, Germany, New York: ACM Press, Aug. 2003: 365-378

[7] Vukadinovic D, Huang P, Erlebach T. On the Spectrum and Structureof Internet Topology Graphs// H. Unger, T. Böhme, A. Mikler, eds. Lecture Notes in Computer Science (LNCS) 2346. Berlin: Springer-Verlag, 2002: 83-95

[8] Spring N, Mahajan R, Wetherall D. Measuring ISP Topologies w-ithRocketfuel. ACM SIGCOMM CCR, 2002, 32(4): 133-145

[9] Donnet B, Raoult P, Friedman T, et al. Deployment of an Algorithm for Large-Scale Topology Discovery. IEEE Journal on Selected Areas in Communications (JSAC), Special Issue on Internet, Special Issue on Internet Sampling, 2006, 24(12): 2210-2220

[10] Pansiot J J, Grad D. On Routes and Multicast Trees in the Internet. ACM SIGCOMM CCR, 1998, 28(1): 41-50

[11] Govindan R, Tangmunarunkit H. Heuristics for Internet Ma P-Discovery// Proc. IEEE INFOCOM 2000. Tel-Aviv, Israel, Mar. 2000, 3: 1371-1380

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

[13] Barford P, Bestavros A, Byers J, et al. On the Marginal Utility of N-etworkTopology Measurement// Proc. 1st ACM SIGCOMM Workshop on Internet Measurement (IMW 2001). San Francisco, California, Nov. 2001: 5-17

[14] Meyer D. Route Views Project. University of Oregon Available URL: http://www.routeviews.org/

[15] Spring N, Mahajan R, Wetherall D. Measuring IS PTopologies with Rocketfuel. ACM SIGCOMM CCR, 2002, 32(4): 133-145

[16] Mao Z Q, Rexford J, Wang J. Towards an Accurate AS-Level Traceroute Tool// Proc. ACM SIGCOMM 2003. Karlsruhe, Germany, New York: ACM Press, Aug. 2003: 365-378