

IP 网络拓扑发现研究^{*}

Research on Topology Discovery in IP Network

刘玉华^{1,2} 肖德宝¹ 余一姣¹ 高 丽¹

(华中师范大学计算机科学系 武汉 430079)¹ (华中科技大学计算机学院 武汉 430070)²

Abstract Auto topology discovering in large and heterogeneous network is becoming more and more critical to network management. With our experience in network management, we propose the main troubles at first, which existing algorithms are facing. Then, key technologies in discovery algorithms and important criterions to judge them are analyzed in detail. At last, we example some good discovery methods in popular network management tools step by step.

Keywords IP Network, Topology structure, Discovery algorithm

1. 引言

网络拓扑发现是网络管理功能的基本组成部分之一,也是衡量商业网络管理系统成败的重要尺度。拓扑发现在搜索网络瓶颈、解决交通拥塞、排查潜在致命点、增强网络连接性和网络升级方面都是至关重要的。拓扑发现还可用于实时网络仿真和帮助用户决定从何处切入网络(安装服务器)以获得最大带宽。拓扑发现对于开发新型网络协议和算法,如 QoS 路由策略等来改进网络性能也是很有利的。

目前国外一些网管产品中已实现了网络自动拓扑发现功能,其中最具代表性的有:HP 的 Openview、IBM 的 Netview、Cisco 的 CDP 等。国内网管软件开发也取得了一定进展,对局域网、小型网络拓扑发现较好,但大型网络成功的实例不多。在拓扑发现研究方面,美国 Cornell 大学的 CNRG 网络研究组、南加州大学的 SCAN 研究组和 Bell 实验室都走在前列,他们提出了启发式发现算法^[1,2]、可搜索交换机端口连接的链路层发现算法等^[3]。在国内开展拓扑发现研究的有上海交大^[4]、华中科技大学^[5]、西安交大^[6]、华中师大等,他们都设计了有自己版权的发现算法,并已用于商业网管软件中。但是迄今为止,对于现代大型异构 IP 网络,自动拓扑发现仍是一项挑战。因为:

- IP 网上运行的子网、主机、网络互连设备、网络协议、操作系统等都出自于不同供应商,型号各异、版本有别、种类繁多,这给拓扑信息获取带来了极大困难;

- 由于 IP 网的开放性,这些网络由不同的组织机构控制和管理,有各自的管理规则与安全策略,有的过滤甚至屏蔽一些发现工具和算法的搜索;

- IP 网络是一个实时动态网络,瞬息万变,可能费时费力生成的拓扑图到手时已经过时。

鉴于以上情况,研究设计好的高效拓扑发现算法尤为重要,而且需将多种发现算法结合使用,才能确保自动拓扑发现的速度快、开销小、准确率和完整性高。

2. 网络拓扑结构与图的遍历

网络拓扑发现是依靠计算机软件和程序自动确定网络中各元素相互之间的动态关系,其中网络元素是指网络中的路由器(低层网关)、交换机、网关(高层网关),还有子网、网桥、

Hub、主机和信道。网络拓扑直接呈现网络的物理连接与逻辑连接,这些物理连接按拓扑方式可分为总线型、星型、环型、树型、星环型、混合型等。IP 子网大都采用简单的拓扑结构,而 IP 主干网则是一种不规则的网状网。在分析网络拓扑结构时,我们常把网络中的路由器、子网、主机、网桥等设备抽象为一个点,把连结这些设备的信道抽象成两点之间的边,这样网络拓扑结构就被抽象成了一个图,图就是网络拓扑的数学模型^[7],那么一切研究图论的工具都可以用来研究网络拓扑结构^[8]。图论中的深度优先遍历与广度优先遍历在网络拓扑搜索中广为应用,即从图的任一点出发,顺着某些边,可以访问图中所有节点,而且使每一个节点仅被访问一次,这就是图的遍历。本文后面章节中的很多搜索算法都是基于图的遍历算法的。

3. 网络拓扑发现关键技术

研究 IP 网络拓扑发现要明确几个概念及其关键技术。

3.1 拓扑发现的规模

IP 网中有主干网、域网、子网等,其使用的互连设备不同。例如主干网通常由边界路由器连接不同的自治系统(AS)组成,而域网内多是由路由器-路由器连接或路由器-子网连接(也有主机充当路由器的),而子网一般是不同结构的 LAN 网,子网内多是主机。所以针对不同规模的网络、所需拓扑信息的详略,发现速度的快慢,搜索开销的大小,而采用不同机理的设计算法。一般有主拓扑发现算法、域拓扑发现算法、子网拓扑发现算法等。为了提高速度,可在所搜索域网内再细划分子域,缩小搜索粒度,在各个子域里同时运行多个进程并行搜索。

3.2 拓扑发现的层次

从协议层次角度来看,IP 网络有网络接口层、网络层、传输层和应用层。通常网络层(IP 层)是拓扑发现的首选位置,几乎所有网络都支持 IP 协议,而且基于 SNMP 协议的发现算法可实现性较强,所以传统的发现算法都是在单一管理域中的网络层搜索逻辑网络的拓扑结构。但在大型异构 IP 网中,此技术仅能发现小部分拓扑信息,所以有必要深入到网络接口层(OSI 的链路层、物理层)和虚网中进行探索。为了提供更大的带宽,交换机越来越多地分布在 IP 网中,以形成微子网段,这些架构对网络层是不可见的,所以为了真正掌握网络

^{*}湖北省自然科学基金资助项目(编号:2001ABB013)。刘玉华 博士生。

的拓扑连接,多层拓扑发现是必不可少的。

3.3 拓扑发现的信息采集

由于 IP 网络协议丰富,拓扑发现算法可利用协议特性来析取相关信息组成拓扑视图。我们总结了目前常用发现算法所涉及到的有关协议,如图 1 所示。

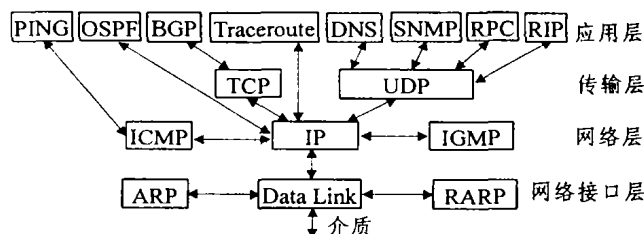


图 1 拓扑发现算法涉及到的相关协议

其中最常用的发现工具有 Ping、Traceroute、DNS,还有基于这些工具和协议的组合算法,如:广播 Ping+DNS Zone 转换算法;Traceroute+DNS Zone 转换算法;Ping+Traceroute 算法;基于 SNMP+ICMP 算法;基于 SNMP+ARP 算法;基于 IP+TTL 算法;基于 BGP(OSPF、RIP)动态路由协议算法;基于 IGMP 组播拓扑发现算法等等。这些算法利用协议性质采集设备中存放的数据库、链表或变量拓扑信息来构成拓扑图。这些协议分别适合于不同规模、不同层次、不同方式下的发现算法,以求达到其性能和开销的最佳比值。

3.4 拓扑发现的方式

拓扑发现有主动方式与被动方式两种方法^[6]。主动式一般由网管主机或运行发现程序的主机向需要搜索的网络发送探索指针或数据包,然后采集返回的信息,析取其相关拓扑成分形成拓扑视图。现在大多数算法都采用此类方式,它的优点是快捷,缺点是增加网络额外流量。被动监测技术是在所需探测网络内安置探测器或探索源,由它监视当前网络运行情况,得到信息后发送给网管主机,以形成拓扑视图。此技术的优点是不增加额外流量开销,缺点是速度相对慢,而且安装探索源对一个大型网络来说并不现实。

一个好的发现算法是将主动探测与被动监视技术结合起来,例如用 ICMP 包探查子网上活动主机,用 SNMP 报文采集监测到的有关路由器信息。

3.5 发现算法的目标和性能参数

拓扑发现算法追求高效率、高速度、高完整性、高准确率四个目标,即尽可能小的网络开销,尽可能快的运行速度,尽量完整的拓扑区域,尽量准确的拓扑信息。也有 4 个参数^[1]评价一个发现算法,即速度、开销、完整性和准确率。速度指拓扑发现算法运行的时间,通常以分钟计算;开销以测量发送 ping 包和 traceroute 追踪所经历的路由器跳数决定;完整性用发现网上主机、路由器、子网数的百分比表示;准确率即发现的拓扑元素是否正确无误,也是用百分比表示。

4. 拓扑发现工具的机理

IP 网络中有些协议和命令是可以直接作为拓扑发现工具的,它们是:

1) Ping 与广播 Ping^[9] Ping 操作主要用来检测目的主机是否存活在网络中,它采用异步方式发送和接收 ICMP 应答包。Ping 通一个存活的主机大约只需几十毫秒,一个 C 类网可在几秒钟内搜索完毕,但如果该主机关机或在 IP 层上不可达,则将耗费较长时间。广播 Ping 是将 ICMP 包发向整

个子网,例如发向子网段 192.168.16.0 或 192.168.16.255,它对发现子网中的全部主机是有用的,但这种并发送会增加网络流量,甚至会引起网络拥塞,而且许多子网并不支持此种操作。

2) Traceroute^[9] 可以探索 IP 数据报从一台主机发送到另一台目的主机所经过的全部路径。IP 包首部有一个 TTL 字段(生存周期),开始信源发送一个 TTL 值为 1 的 UDP 报文,第一个路由器收到此报文后,将 TTL 值减 1 为 0,便丢弃此报文,同时回发一个 ICMP 超时错报文给信源,这样信源得到了第一个路由器信息。然后信源第二次发 IP 包时,增加 TTL 值为 2,这样每重发一次时,TTL 值增 1。接收报文的路由器先将 TTL 值减 1,若 TTL 值不为 0,则继续向路由器的下一跳转发;若值为 0 则丢弃该报文,回发 ICMP 超时错信息给信源,信源便记下该路由器信息,在到达最终目的主机时,返回信源一个 ICMP 端口不可达报文。以此逐步探索可得到域内全部路由器拓扑信息。

3) DNS 的 Zone transfer^[10] IP 网络中每个节点有唯一的 IP 地址,DNS 中包含有每个主机名字与 IP 地址转换表,Zone transfer 命令可以很快地发现域内主机和其它网络设备的名字,通常它与广播 Ping 或 Traceroute 工具结合使用,搜索子网内的主机和设备或作为其它发现算法的起点。这种技术简单快捷,但是出于安全考虑,有些 DNS 服务器并不提供这种工具。而且 DNS 数据库存储数据是静态的,它不能随时更新,所以信息有时并不准确。表 1 给出了这三种工具性能的定性比较^[1]。

表 1 三种发现工具性能比较

	Ping	Traceroute	DNS zone transfer
速度	活动主机快、其它慢	慢	快
开销	低	高	低
完整性	全部域	全部域	大多数域
准确率	准确	通常准确	通常准确

5. 几类发现算法的研究

5.1 主拓扑发现

在主干网上由于受限条件多,利用 IP 网络最原始工具 Ping 和 Traceroute 为核心设计算法最为实用。在实现细节上有所不同,有的算法初始采用随机地址探索来构建 IP 地址集,这样可避免额外的数据报文开销。有的采用 DNS Zone 转换来形成初始 IP 集。一个好的算法是基于 BGP 协议的^[1],BGP 是一种不同自治系统(AS)路由器之间进行通信的网关协议。对边界路由器遍历搜索、析取其路由信息,可得到不同的 AS 域和大量的域 IP 地址,发送探索包到一个随机选中的 IP 地址上,可以获取主干网上路由器连接信息。为了确保不漏掉交叉链的节点,可采用多重搜索技术。该算法下一步再用 Traceroute 工具来追踪,逐步生成所需拓扑图。

算法描述如下:

1. 取 BGP 路由表信息
2. 从信息里提取全部域地址
3. 在每一个域内选取一个 IP 地址
4. 对每一个 IP 地址做:
 - a. traceroute 这个 IP 地址
 - b. 存储搜索到的全部节点和边
 - c. traceroute 返回
5. 获取相关结果

此算法的优点是实现简单,大多数网络可行;缺点是发现周期长,对主干网交叉点识别不够。

5.2 域拓扑发现

传统的域拓扑发现算法大都是基于 SNMP 协议的^[11]。SNMP 实现分为 NMS(管理者)和 Agent(代理)两部分,Agent 存在于各网络设备中,用以响应管理者的请求,对 MIB 库进行访问和设置。NMS 与 Agent 间有 5 种通信原语,拓扑发现算法使用其中 Getrequest 和 GetNextrequest 等原语,采用图论中的深度优先或广度优先遍历法搜索路由器 MIB 库中路由表变量 ipRouteNextHop、ipRouteDest 的信息,生成域内路由器-路由器、路由器-子网间拓扑结构图。

由于网络 IP 化趋势和 SNMP 的广泛应用,该算法可适用于大多数环境;另外由于路由表信息是动态刷新的,因此可实时获取当前网络的拓扑信息;第三个优点是该算法实现简单速度快。对于那些不支持 IP 的网络,可以在网络边界采用代理机制,转化为 Over IP 网,再实现此种技术。此算法的主要缺点是它只能搜索域内 IP 层路由信息,而不能析取更为复杂的链路层、物理层拓扑信息,并且对于那些拒绝访问其 SNMP 的网管域是无能为力的。

5.3 子网的拓扑发现

一种子网拓扑发现是根据 MIB 变量 iftype 值确定相应子网的拓扑结构;如总线型、星型、环型等。它也是对 5.2 节域拓扑算法的补充,是基于 ARP 和 SNMP 结合起来的^[5]。ARP 属于网络接口层,它提供 IP 层到物理层网卡 MAC 地址之间的动态映射,利用路由器上 ARP 表中的 MIB 变量 ipToMediaNet Address,可以逐一取得所有与它相连终端设备的 IP 地址。将这些 IP 地址与给定子网地址进行模式匹配,如果成功则证明该 IP 地址对应的主机在子网内。主机的 MAC 地址可由 MIB 变量 ipNetToMediaPhyAddress 获得。

此算法的优点是迅速简捷、网络开销小;缺点是更新 ARP cache 间隔时间是静态的,所以实时性较差,可用程序通过 Telnet 仿真修改轮询时间。

另一种子网拓扑算法主要是利用 TCP/IP 中的 ICMP 协议^[12],算法描述如下:

1. 确定该子网的网关、子网地址和子网掩码
2. 通过网关地址获得子网类型
3. 通过子网地址和子网掩码获得该子网的主机数

4. 在该网中,对 IP 地址进行 ping 操作,析取有 ICMP 报文回应的 IP 地址

5. 根据 Socket 的请求,可以获取存活主机的其他信息

此算法简单易行,但 ping 网络会增加开销,另外发现速度较慢。

结束语 近年来我们在开发园区网络拓扑发现系统中积累了一定的经验,但大型 IP 异构网络与园区网的拓扑发现有很大区别。本文在指出 IP 网络特性基础上广泛探讨了其拓扑发现理论与关键技术,构造并分析了一些发现算法,指出了其性能利弊。下一步应深入研究 IP 网络物理层/链路层和虚网的拓扑发现、探索组播拓扑发现技术,进一步从理论上定量分析算法的性能,设计出好的高效拓扑发现算法来。

参考文献

- 1 Siamwalla R, et al. Discovering Internet Topology. IEEE INFOCOM 1999
- 2 Govindan R, et al. Heuristics for Internet Map Discovery. IEEE INFOCOM 2000
- 3 Breitbhart Y, et al. Toplogy Discovery in Heterogeneous IP Networks. IEEE INFOCOM 2000
- 4 白英彩,等. 计算机网络管理系统设计与应用. 北京:清华大学出版社,1998
- 5 李佳,石冰心,等. 网络管理系统中的自动拓扑发现算法. 华中理工大学报,1998(1)
- 6 张勇,张德运,等. 网络拓扑发现的主动探测技术的研究和实现. 小型微型计算机系统, 2000. 8
- 7 Calvert K, et al. Modeling Internet Topology. <http://www.cc.gatech.edu/projects/gtutm>
- 8 Xu Junming. Topological Structure and Analysis of Interconnection Networks. New York. Kluwer Academic Publishers. October 2001
- 9 Stevens W R. Tcp/Ip Illustrated, Volume 1: The Protocols. Boston, MA. Addison Wesley. 1994
- 10 Qasem A, Melendez E A. Internet Topology Discovery. <http://www.cs.fsu.edu/>
- 11 Case J. A Simple Network Management Protocol (SNMP). RFC 1157. 1990
- 12 Deering S. ICMP Router Discovery Messages. RFC1256. Sept. 1991
- 5 Chen W, Lin E, Wei H. Dynamic location control for mobile nodes. [Technical Report 97-CSE-10]. Department of Computer Science and Engineering, Southern Methodist University, 1997
- 6 Yavatkar R, Pendarakis D, Guerin R. A Framework for Policy-based Admission Control. RFC2753, Jan. 2000
- 7 Seshan S, Balakrishnan H, Katz R H. Handoffs in cellular wireless networks: The daedalus implementation and experience. Kluwer International Journal on Wireless Personal Communications, Jan. 1997. 8
- 8 Mysore J, Bharghavan V. A new multicasting-based architecture for internet host mobility. In: Proc. of ACM/IEEE MobiCom, Budapest, Hungary, Sept. 1997. 161~172
- 9 Cheng Lin Tan, et al. A Fast Handoff Scheme for Wireless Networks. The Second ACM International Workshop on Wireless Mobile Multimedia (WoWMoM'99), Seattle, Washington. 1999
- 10 Williams C. Localized Mobility Management Requirements. Internet Engineering Task Force. draft-williams-mobileip-lmm-requirements-00.txt

(上接第 32 页)

动管理方案^[10],更好的移动管理方案必须具有网络开销更低、切换时延更小、数据丢失更少、安全程度更高、实现更容易以及应用范围更广等特点。

参考文献

- 1 Perkins C. IP Mobility Support. Network Working Group, Oct. 1996. RFC 2002
- 2 Perkins C. Mobile-IP Local Registration with Hierarchical Foreign Agents. Mobile-IP Working Group, Internet Engineering Task Force, Feb. 1996
- 3 Cho G, Marshall L F. An efficient location and routing scheme for mobilw computing environments. IEEE Journal on Selected Areas in Communications, 1995, 13(5): 868~879
- 4 Caceres R, Padmanabhan V N. Fast and Scalable Handoffs for Wireless Internetworks. In: Proc. of ACM MobiCom'96, Nov. 1996