

基于生成树算法的链路层拓扑发现研究

刘海华, 王萍萍

(西南交通大学 信息科学与技术学院, 四川 成都 610031)

摘要:随着大规模交换网络的发展,网络拓扑发现的研究由网络层拓展到数据链路层。链路层的拓扑发现能够发现网络层拓扑发现无法发现的局域网内部的详细的物理连接情况,对网络配置管理具有重要意义。研究了目前基于地址转发表(AFT)的方法,针对现有算法的不足作了一定分析,提出了一种基于生成树算法(STA)的链路层网络拓扑发现算法,利用SNMP获得网桥MIB中的生成树信息,通过分析这些信息计算出链路层的网络拓扑。该算法相比其它算法更简单、高效,有应用价值。

关键词:拓扑发现;地址转发表;生成树算法;简单网络管理协议;MIB

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2008)05-0101-04

Research of Datalink Layer Topology Discovery Based on Spanning Tree Algorithm

LIU Hai-hua, WANG Ping-ping

(School of Information Science & Technology, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: With the development of large scale switching network, study is transferred from network layer to data link layer. The link layer topology discovery can discover the physical topology internal a LAN, which the network layer topology discovery can not discover. Analyze the current link layer topology discovery algorithm: method base on address forwarding table, a link layer topology discovery algorithm based on spanning tree algorithm is proposed. The algorithm gets the STP information from the bridge MIB with tool of SNMP. It can calculate the link layer topology on the basis of information. The algorithm is simple and with high performance in application.

Key words: topology discovery; AFT; STA; SNMP; MIB

0 引言

随着网络对带宽的要求越来越高,大量的集线器被交换机所替代,这使得链路层的拓扑信息对于网络管理越来越重要,因为当某链路发生故障或拓扑连接关系发生改变时,拓扑信息可以很直观地显示出来。数据链路层拓扑发现算法是和网络层完全不同的,它实现起来很复杂,这主要是由于交换机硬件的透明性。网络中连接的设备并不能意识到交换机的存在,交换机本身仅仅在生成树协议中与相邻交换机进行通信。交换机所维护的唯一状态信息就是它们的地址转发数据库,它存储了转发与过滤数据帧的信息,这些信息被保存在SNMP Bridge MIB^[1]中。只有通过通过对网络各个交换机地址转发数据库的深入分析,才可能发现网络中交换机到路由器、交换机到交换机、交换机到终端设备之间的连接关系。现在,国外已经提出了几种基于

交换机AFT(地址转发表)的算法^[2,3]。

1 当前基于地址转发表的链路层发现算法

在交换域中,交换机的连接关系确定了整个交换域的拓扑结构,因此,发现交换域的拓扑首先要发现交换机之间的连接关系。在讨论算法之前,先了解几个定义和定理:

定义1:令拓扑图中第*i*台交换机的第*j*个端口为*S_{ij}*。

定义2:令对于任意端口*S_{ij}*,*A_{ij}*表示交换机*S_i*的地址转发表中通过端口*S_{ij}*收到的数据帧的源地址集合。

定义3:若一台交换机*S_i*端口*S_{ij}*的*A_{ij}*中未出现其它交换机的MAC地址,则称端口*S_{ij}*为叶端口。

定理1:如果 $A_{ij} \cup A_{kl} = U$ (*U*指交换域内所有交换机的集合)且 $A_{ij} \cap A_{kl} = \emptyset$,则端口*S_{ij}*与端口*S_{kl}*直连。

定理2:若路由器*R*与交换机*S_i*的*S_{ij}*端口直接

收稿日期:2007-08-27

作者简介:刘海华(1981-),男,硕士研究生,研究方向为软件、网络管理;导师:倪少权,教授,博导,研究方向为网络管理。

相连当且仅当 S_{ij} 是叶端口,且 A_{ij} 中包含路由器 R 的 MAC 地址。

目前的链路层拓扑发现算法中,基本上都用到了定理 1 来证明一个交换机与另一个交换机是否直接相连,用定理 2 来证明交换机与路由器是否直接相连,从而得出交换域中的各个交换机之间、交换机与路由器之间的连接关系。

虽然根据定理 1 可以确定交换机端口之间的连接关系。但存在以下不足:

(1)在整个过程中,需要将一个交换机上的端口和其它所有交换机上的端口进行比较判断。如果交换机很多,端口数量为 P (很大)时,基本上要进行 P^2 次运算,使得运算量很大。

(2)若交换机地址转发表完整,从定理 1 就可得出交换机的连接关系,但是很多时候地址转发表并不完整(因为交换机会定期扫描地址转发表,发现在一定时间内没有出现的 MAC 地址就把它从表中删除)。

因此引入了基于生成树的发现算法来对交换机连接关系进行修正。

2 基于生成树算法的链路层拓扑发现

目前在网络设计中,为了保证网络的可靠性,一般在网络设计时设置备份线路,这样在实际拓扑图中会出现环,但由于生成树协议的限制,交换机将自动封闭这些备份线路以保证系统中不出现拓扑环的情况。本算法捕获的是在算法运行时段内,子网内部所自动生成的当前“生成树”的快照,本算法拓扑发现的结果将包含以太网内的交换机集合、交换机及其端口之间的连接关系以及交换机与路由器之间和交换机与终端设备之间的连接关系等。

2.1 生成树协议的基本概念

使用生成树协议可以让交换机动态地探测一条无环路的数据通路,并当拓扑结构改变时,自动重配。使用生成树协议,交换机之间传输一种特殊的消息,称为网桥协议数据单元(BPDU),此信息用于计算一颗生成树的结构^[4]。这些消息有如下用途:

- (1)选择一个交换机为根交换机(Root Bridge)。
- (2)确定交换机到根交换机的最短路径。
- (3)在每个网段上选出一个指定交换机(designated bridge),该网段只有通过该交换机才能向根交换机方向发送数据。
- (4)在每个交换机内选定一个根端口(Root Port),只有通过该端口才能向根交换机方向发送数据。
- (5)确定交换机各端口的状态。
- (6)每个网段的指定交换机(designated bridge)必有

一个端口与该网段相连,称作该网段的指定端口(designated port)。

交换机之间不断交流这些信息,确定自己在生成树中的位置。该生成树中的节点可以理解成两类:一类是交换机,一类是网段。对于树中的每一个节点,若它的父母和子女都已确定,那么该树已经形成,并且每一个节点都明确知道自己在树中的位置。因此,我们的目的就是让每个节点明确自己的父母和子女。

对于树中的交换机节点,它的根端口所接的网段是它的父母节点,和它的指定端口(designated port)相连的网段是它的子女结点。对于网段节点,它的指定交换机(designated bridge)是它的父母节点,以根端口(root port)和本网段相连的交换机则是它的子女节点。这样,通过相互交换 BPDU,各交换机之间形成了一颗生成树。另外,在工作过程中,可能由于拓扑结构或路径长度的改变,引起生成树的重构,这种重构也是通过交换机之间的 BPDU 的交流动态实现的。

2.2 生成树算法的实现过程

一开始,网络上的各个交换机都假定自己是根交换机,假定各个端口都是指定端口,于是该交换机定时向外发送配置 BPDU,当某个交换机接收到相邻交换机数据帧中的信息后,与自己的特征进行判断,如果发送该帧的交换机的优先级比自己高,则该交换机修改自己的相应参数,使自己变成非根交换机,否则,接收交换机向发送交换机发送自己的配置 BPDU,这样,经过一段时间后,整个网络就可以选择出该网络的根交换机,指定交换机和交换机的指定端口和根端口,如果一个交换机的端口既不是根端口,也不是指定端口,则该端口被阻塞。于是整个生成树形成。作为网络的根交换机,要定时向外发送配置 BPDU,任何一个非根交换机都经过根端口从上一级接收根交换机发送的配置 BPDU,又通过指定端口向下一级发送该配置 BPDU,同时修改自己的 message 定时器。这样,当网络中的某一交换机出现故障时,它相邻的交换机就会很长时间接收不到根交换机发送的配置 BPDU,这样,它的 message 定时器就会超时,于是,该交换机向外发送拓扑改变 BPDU,接收到该拓扑改变 BPDU 的交换机又会进一步向外发送拓扑改变 BPDU,直到根交换机知道网络的拓扑结构发生了变化,或者根交换机出现故障。不管哪种情况,整个网络中的各个交换机又会像最开始时一样重新配置,最后形成一颗新的生成树。

2.3 相关的 MIB 变量

使用生成树的前提是交换机都支持 802.1D 生成树协议版本,同时实际网络配置中启用了生成树。文中所讨论的物理连接拓扑发现的过程,也是通过使用

SNMP 协议,查询专门的网桥 MIB 信息,即所谓的 Bridge MIB(RFC1286),并对信息进行分析。基于 SNMP 协议的发现方法的优点在于发现速度快,发现信息全,但它的局限性在于要被发现的设备必须支持 SNMP 协议^[5]。对在链路层设备的发现过程中,用到下面几个 MIB 对象:

(1)dot1dBase 组。

dot1dBaseBridgeAddress 表示交换机的 MAC 地址,可以用此值唯一地标识交换机。

(2)dot1dTp 组。

dot1dTpFdbTable 表中存储的是交换机地址转发表的信息。主要包括的 MIB 对象有:

* dot1dTpFdbPort:交换机各个端口号。

* dot1dTpFdbAddress:交换机对此 MAC 地址代表终端有转发信息。

(3)dot1dStp 组。

该组存储的是与生成树协议相关信息,主要用到的 MIB 对象有:

* dot1dStpDesignatedRoot:交换机所在的生成树中的根交换机的 BID(Bridge Identifier)。

* dot1dStpRootPort:该交换机根端口号。

* dot1dStpPortTable 表:主要含有端口相连信息。表中主要使用的 MIB 变量有:

·dot1dStpPort:该交换机的一个端口。

·dot1dStpPortDesignatedBridge:该端口的指定交换机的 BID。

·dot1dStpPortDesignatedPort:该端口的指定端口的 PID(Port Identifier)。

·dot1dStpPortStatus:通过运行生成树协议确定的端口的当前状态。1 代表 disabled;2 代表 blocking;3 代表 listening;4 代表 learning;5 代表 forwarding;6 代表 broken,出现了故障。只有 forwarding 状态才表示该端口属于生成树中的非阻塞端口。

由生成树协议可知,只要知道生成树中的根交换机、每个交换机的根端口、指定交换机、指定端口、端口状态就可以构建唯一的交换域拓扑结构。dot1dStp 组中的 MIB 变量完全提供了这些信息。

2.4 拓扑发现方法

根据生成树协议,要构建交换域中的拓扑结构,需要找出根交换机,并搜索各个交换机的根端口、指定交换机、指定端口、端口状态。

算法一开始首先找出交换域中的所有交换机,确定根交换机,找出以根交换机为指定交换机的二级交换机,然后找出以这些二级交换机作为指定交换机的三级交换机。依次类推,直到找出所有叶子交换机为

止。叶子交换机即没有交换机以它作为指定交换机的节点。这样,交换机之间的连接关系就确定了。

然后对于所有的叶子端口,查找它们的地址转发表,找出所有与之连接的终端设备和路由器,若有两个或以上的终端设备与一个端口相连,则表明它们通过一个 HUB 与交换机相连。再根据定理 2,找出交换机与路由器的连接关系。算法流程图如图 1 所示。

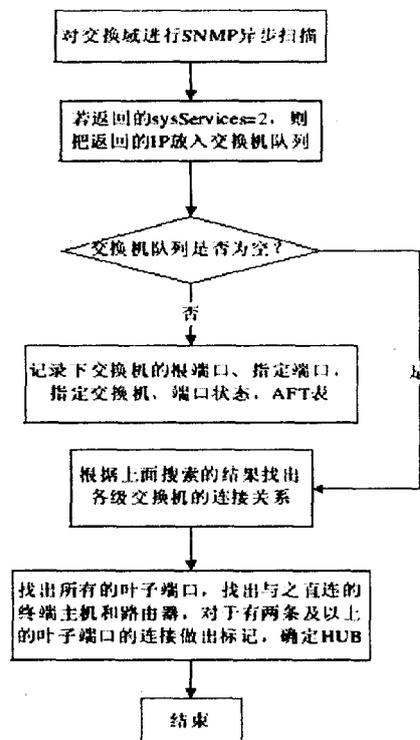


图 1 算法流程图

在算法发现拓扑结构过程中,找出交换域中的所有交换机是很关键的一步,因为这一步会增加很大的网络流量,并且方法使用不当时耗费时间很长,下面介绍一种异步方法提高发现速度。

2.5 使用异步方式找出所有交换机集合

通过读取 ipNetToMediaTable 中的物理地址和 IP 地址对 < IP, MAC >, 然后根据 ipNetToMediaIfIndex 值合并,可以得到一个交换域中的所有设备。为了确定这些设备是否交换机,需要向所有设备 161 端口试探性发送 SNMP 报文,看是否有响应(这里请求 sysServices 的值,若为 2,则该设备是交换机)。没有响应,则是普通计算机。这一过程有两种方式可以实现:

(1)同步方式:向所有要检测的 IP 地址依次发送请求报文,在上一次所发报文的应答或超时报文到达后才执行下一个报文的发送。

(2)异步方式:向所有要检测的 IP 地址,一次性发送所有请求报文,然后接收所有的应答报文。在一个交换域中,交换机的数量只占很少一部分,大部分是对

SNMP 无响应的计算机,若使用同步方式,很多时间浪费在等待超时报文上。所以这里使用异步方式实现。

为了实现这一过程,首先建立一个队列存储所有交换域中的设备 IP,然后对队列中 IP 异步发送 SNMP 请求报文。对于每个应答报文,查看返回的值,若为 2,则把该 IP 放入交换机队列(在这里还有 sysServices 的值大于 2 的应答报文,这是路由器发出的应答报文,通过查找网络层拓扑发现的结构替换序列对队列,找到该 IP 地址对应的 MaxIP,从而找到了相应的路由器标识^[6])。收到第一个超时报文后,等待 1s 并继续接收看是否还有应答报文,1s 后退出等待。

2.6 拓扑更新

上述算法是基于生成树协议来确定交换域内拓扑连接关系的,所以当生成树发生变化时,比如根交换机发生改变或是其它交换机发生故障,拓扑发现都要运行一次,这样才能反映出拓扑的实时性。

3 结束语

实验结果表明,以上的拓扑发现算法在准确度和效率上均有很好的效果。

目前,很多网络中通过三层协议交换机连接不同的 VLAN(虚拟局域网),每个 VLAN 都需要运行独立

的生成树协议,但是 802.1D 中的生成树协议并不支持 VLAN。很多交换机厂商开发自己的私有 MIB 变量来支持 VLAN,这就需要对各厂商的私有 VLAN MIB 变量有全面了解才能发现一个 VLAN 的链路层拓扑。目前的工作中,没有对 VLAN 进行研究。这将是下一步需要研究的内容。

参考文献:

- [1] McCloghrie K. Definitions of Managed Objects for Bridges [S]. Internet RFC1286. 1991.
- [2] Breitbart Y, Garofalakis M, Martin C. Topology Discovery in Heterogeneous IP Networks[J]. IEEE INFOCOM, 2000, 26 (1):265-274.
- [3] Lowekamp B, O'Hallaron D R, Gross T R. Topology Discovery for Large Ethernet Networks[J]. ACM SIGCOMM, 2001 (8):237-248.
- [4] Rose M, McCloghrie K. Structure and Identification of Management Information for TCP/IP - based Internets [S]. RFC1155. 1990.
- [5] 康华,周明全,房鼎益.基于 SNMP 协议的网络拓扑发现和显示[J].微机发展,2003,13(7):60-65.
- [6] 熊英,楚惟善,王春枝.基于 TCP/IP 的网络拓扑发现方法的研究[J].湖北工业大学学报,2005,20(4):53-55.

(上接第 100 页)

建设、成为实现以 XML/SOAP/WSDL/UDDI^[3-6]为主干的信息服务的 key 问题。

4 结束语

数字城市从概念的提出发展到现在仍然处于探索和研究阶段。虽然实现一体化的按需服务的架构的关键技术和标准均以提出和实现,但是在走向工程化的道路上尚有差距。基于 XML 的元数据描述空间和非空间数据成为整合信息港服务实现语义应用的必然,基于 Web 标准的服务体系成为实现 Internet 开放、共享的保障。

针对我国数字城市建设的状况,构建符合 Web 标准的应用平台、元数据标准化、服务标准化建设是加快实施的关键。

参考文献:

- [1] Klopfer M. Interoperability & Open Architectures: An Analysis of Existing Standardisation Processes & Procedures[EB/OL]. 2005-05. <http://www.opengeospatial.org/press-room/papers/>.
- [2] 王洪,李琦,承继成.数字城市元数据服务体系的研究和实践[J].北京大学学报:自然科学版,2004,40(1):107

- 115.

- [3] Web Service 体系[EB/OL]. 2002. <http://www.techng.com/content.asp?titleid=3797>.
- [4] UDDI 组织. UDDI Technical White Paper[EB/OL]. 2004-10. <http://uddi.org/pubs/uddi-tech-wp.pdf>.
- [5] Boubez T, Hondo M, Kurt C, et al. UDDI 程序员 API 规范 V1.0[EB/OL]. 2001-03. <http://www.uddi-china.org/pubs/ProgrammersAPI-V1.01-Open-20010327-CN.pdf>.
- [6] W3C. SOAP Version 1.2, W3C Working Draft[EB/OL]. 2001-07-09. <http://www.w3c.org/tr/2001/WD-soap12-20010709/>.
- [7] W3C Web Home[EB/OL]. 2001. <http://www.w3.org/Consortium/technology>.
- [8] XML project Web Home[EB/OL]. 2004. <https://www.seegrid.csiro.au/wiki/bin/view/Xrmm/WebHome>.
- [9] Sen M, Duffy T. GeoSciML: Development of a generic Geo-Science Markup Language[J]. Computers & Geoscience, 2005 (31):1095-1103.
- [10] 李德仁,朱欣焰,龚健雅.从数字地图到空间信息网格——空间信息多级网格理论思考[J].武汉大学学报:信息科学版,2003,28(6):642-650.
- [11] 王家耀,祝玉华,吴明光.论网格与网格地理信息系统[J].测绘科学技术学报,2006(2):1-7.