桥接局域网第2层拓扑结构的自动发现

李 涛 石志强 吴志美

(中科院软件所多媒体通信和网络工程研究中心 北京 100080)

Discovering Layer-2 Topology in Bridged Local Area Networks

LI Tao SHI Zhi-Qiang WU Zhi-Mei

(Institute of Multimedia Communication and Network Engineering, Institute of Software, Chinese Academy of Sciences, Beijing 100080)

Abstract With the rapid development and the wide use of Bridged Local Area Networks, the importance of below-OSI-Layer-3 topology comes to be realized. Recent research work has demonstrated that active layer-2 topology can be determined using Address Forwarding Information. The large quantities of address forwarding entries in a typical BLAN, however, make their algorithms both time-consuming and bandwidth-consuming. In view of the above and the fact that transparent bridges are used preponderantly in real networks, we base our algorithm on information supported by Spanning Tree Protocol MIB group, which reduces the cost of acquiring and processing information, and improves the accuracy. Our algorithm relies on standard MIB information and can be used in heterogeneous environments.

Keywords Topology discovery, SNMP, MIB, Spanning tree, Directed bipartite graph, Network management, LAN, MAC bridge

1. 引言

早期关于拓扑发现的研究主要集中于第3层,也就是网络层的拓扑结构,但是随着网桥被广泛用于连接多个局域网段构成桥接局域网(BLAN, Bridge Local Area Network)[1]以便提高网络的可靠性、获得更大的带宽或是提供不同局域网协议之间的互通[7],忽略第2层设备(网桥)之间连接关系的拓扑结构信息显得"粒度"太粗了,桥接局域网的网络管理需求使得网络拓扑结构的自动发现必须向下延伸。

在文中,我们说第2层拓扑结构是指网络在OSI协议栈第2层上的拓扑模型,由网络中所有网桥以及它们之间的连接组成。网桥之间的连接可以是通过一对端口直接相连接,也可以是由于连接到同一局域网段的相连接。在下文中,为简洁起见,除非特别指明,我们说拓扑是指第2层拓扑。

文[4,5]讨论了拓扑发现的问题,所给算法均基于网桥端口的地址转发信息,其中文[5]中算法是对文[4]的改进。文[4]中算法的核心是:分属两个网桥的一对端口是相连的,当且仅当这两个端口的地址转发条目集合的交集中没有关于其他任何网桥的地址条目,并且其并集中包含了关于该局域网中的所有网桥的地址条目。根据该性质能够得到完整无误的网络拓扑结构的前提条件是每个网桥上每个端口的地址转发信息都是完整无遗的。然而在实际网络中,由于网桥地址转发信息老化机制的存在,这一点通常难以保证。

虽然文[4]中给出了两个方法解决这个问题,然而这两个方法在处理实际网络时,还是遇到了困难[5]。文[5]中从另一个角度出发通过地址转发表推导端口间的连接关系。为了表述上方便,引入"通过集(through sets)"的概念。网桥Bi上端口x的通过集,是指该网桥上除去x以外的其它端口地址转发条目集合的并集。文[5]中算法所基于的性质是:如果两个端口的通过集相交不为空,则这两个端口必然不相连。否则就等于宣告交集中设备同时出现于两个不同的网段,而在第2层拓扑结构无回路的局域网中这显然是不可能的。反过来,如果仅仅知道一对端口的通过集相交为空,却不能判定这一对端口相连(具体例子参见文[5])。文[5]中指出如果这两个端口满足"Minimum Knowledge Requirement",那么它们之间否相连则是可以判定的。文[5]中说明了在通常情况下,该条件是容易满足的。

上述两个算法均称适用于任何种类的网桥组成的无回路桥接局域网。但我们也看到,其连接关系的推导基于集合上的并和交,运算很繁琐。事实上,对于一个中等规模的桥接局域网,地址转发条目的数量就可以达到几百条,挨个取得每个网桥每个端口的地址转发表所消耗的网络带宽将会很可观。而且,实际网络中由于地址转发信息老化机制的存在,网桥地址转发表不完整的情况将直接影响这两个基于地址转发信息算法的准确性。

事实上,实际网络中交换机等透明网桥的使用远远超出了源路由等其他种类的网桥。如果我们把算法的适用范围稍作牺牲,将问题域确定为透明网桥组成的桥接局域网,那么就可以完全抛开地址转发表,而仅利用采集和处理开销小得多的生成树协议组(dot1dStp,文[3]中定义的一个标准SNMP MIB)变量来构造网络的第2层拓扑结构。

本文第2部分介绍了与生成树相关的概念以及网桥MIB[3]中生成树协议组,接着在第3部分提出系统模型,并在此基础上给出基于生成树的第2层拓扑发现算法,并对算法作了比较和分析,第4部分中讨论了该算法实现的问题,最后总结了全文。

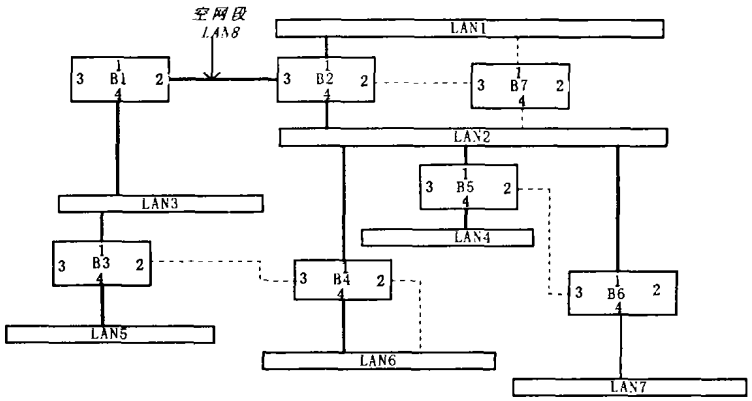
2. 生成树

2.1 生成树的概念

透明网桥组成的局域网的正常工作要求网络拓扑是树型的。然而有时网络管理员会在局域网段之间设置并行的两个或多个网桥以提高网络的可靠性[7]。如果不采取任何措施,由

于拓扑结构中产生了回路，这种配置将引发广播包风暴造成网络瘫痪。因此，网桥之间相互通信（例如使用文[1]中生成树协议）产生一个能够到达每个局域网段的生成树覆盖实际的拓扑结构来完成数据帧的转发。生成树构成了该局域网第2层的拓扑结构。图1中给出了一个有冗余配置的局域网，粗实线连接的网桥和局域网构成了该局域网的一个生成树，虚线表示了被协议阻塞的连接（连接两端的网桥端口也一起被阻塞）。

在透明网桥组成的局域网中，网桥之间将通过协议协商出一个网桥作为该局域网生成树的根，这个网桥称为根网桥。例如文[1]中生成树协议根据优先级和MAC地址选出一个网桥作为根网桥。



（粗实线表示的是被协议激活的连接，细虚线表示的是被协议禁止的连接）

图1　一个配置有冗余网桥的局域网及其生成树

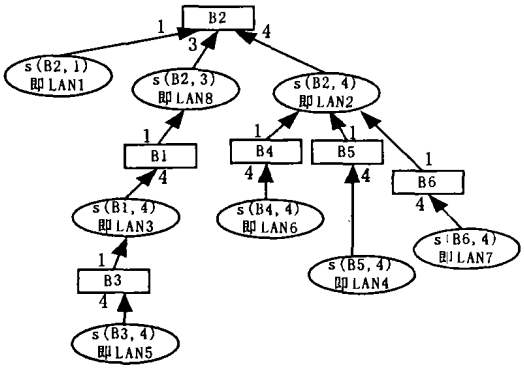


图2　拓扑结构的有向二部图表示

对于每一个网段，在所有连接到它的网桥端口中，选择一个到根网桥最近的端口负责将该网段上的数据帧传递到根网桥以及把从根网桥传递来的数据放到该网段上，这个端口就称作该网段的指派端口，该端口所属于的网桥称作该网段的指派网桥。如果网桥的端口是其所连接到的网段的指派端口，我们也称该端口是该网桥的指派端口。我们称参与生成树的端口为活动端口，显然根网桥上的活动端口一定是其所在网段的指派端口，根网桥是其所连接到的网段的指派网桥。网桥的根端口是该网桥上所有端口中到根网桥最近的端口，负责在本网桥和根网桥之间传输数据帧。即使某个网桥上有两个端口到根网桥一样远近，协议将保证只选择其中一个，如文[1]生成树协议选择端口号较小的端口。除去指派端口和根端口，其它的端口都被协议阻塞。网桥上根端口所连接到的网段称作该网桥的根网段。局域网中所有网段，所有网段的指派网桥，所有网桥上的根端口和指派端口组成了该局域网的生成树。

当两个网桥的一对端口直接相连时，我们也称这一对端口是通过一个空网段相连。如我们把图1中网桥B1上的端口2和网桥B2上的端口3看作是通过一个空网段LAN8相连。表1给出了图1所示局域网中每个网段的指派网桥和指派端口。

表1　网段及其指派网桥、指派端口

网段	指派网桥	指派端口
LAN1	B2	1
LAN2	B2	4
LAN3	B1	4
LAN4	B5	4
LAN5	B3	4
LAN6	B4	4
LAN7	B6	4
LAN8	B2	3

2.2　生成树协议组

这是IETF定义于文[3]中的一个标准SNMP MIB组，定义了与生成树相关的一些MIB变量，与本算法密切相关的有：1）根网桥（dot1dStpDesignatedRoot），即该网桥所在生成树的根；2）根端口（dot1dStpRootPort），即该网桥的根端口；3）指派端口和指派网桥（dot1dStpPortDesignatedBridge和dot1dStpPortDesignatedPort），对应于网桥上的每个端口给出了所连接到的网段的指派端口和指派网桥；4）enabled/disabled状态（dot1dStpPortEnable），对应于网桥上的每个端口给出了端口当前状态是enabled还是disabled。5）端口状态（dot1dStpPortState），对应于端口的状态，可以是转发（forwarding）、阻塞（blocking）、侦听（listening）等。

3. 基于生成树的拓扑发现算法

3.1　系统模型

我们将局域网抽象为一个有向二部图D[6]。生成树中所有的网桥组成了二部图D的一个顶点集Nb，所有网段组成了D的另一个顶点集Ns。我们将网桥和网段之间的连接分为两种：一种是网段到其指派网桥的连接，在D中，这是一条从Ns中顶点出发到Nb中顶点的有向边；另一种是网桥到其根网段的连接，在D中，这是一条从Nb中顶点出发到Ns中顶

点的有向边。

我们将 D 表述为一个四元组(Nb,Ns,E1,E2),Nb 和 Ns 如前所述分别是网桥和网段顶点集,E1 和 E2 是有向边集,其中 E1={⟨s,B⟩|(s∈Ns)∧(B∈Nb)∧(B 是 s 的指派网桥)},E2={⟨B,s⟩|(s∈Ns)∧(B∈Nb)∧(s 是 B 的根网段)}。图 2 以有向二部图 D 的形式表示了图 1 中的局域网。

引理 1 网段与其指派端口是一一对应的。

证明:由指派端口的定义可得每一个网段有且仅有一个指派端口;而任一端口只能连接到唯一段上,所以本引理成立。

我们用 p(Bi,x)表示网桥 Bi 上的端口 x,根据引理 1,我们可以用 s(Bi,x)标识指派端口是 p(Bi,x)的网段而不会引起任何歧义。

3.2 算法

对于一个给定的局域网,首先我们确定 Nb。例如可以使用 Ping 找出每个活动的设备,然后通过标准 SNMP MIB 系统组(system)[8]确定该设备是否是一个 2 层设备,找出局域网中所有网桥,然后对每一个网桥用 dot1dStpPortEnable 和 dot1dStpPortState[3]判断出它是否在 Nb 中。

然后我们找到根网桥,记为 R。

接下来我们给出两个推论用于发现网段以及 E1 和 E2 中的边。

推论 1 ⟨s(Bj,x),Bi⟩∈E1⇔j=i 且端口 x 是 Bi 的指派端口。

证明:"⇒" ⟨s(Bj,x),Bi⟩∈E1 即网桥 Bi 是网段 s(Bj,x)的指派网桥,而从网段 s(Bj,x)的表示我们知道 Bj 也是 s(Bj,x)的指派网桥,由引理 1 得到 j=i 且端口 x 是 Bi 的指派端口。

"⇐" 由 E1 定义显然可得。

推论 2 ⟨Bi,s(Bj,x)⟩∈E2⇔Bi 的根端口 r 所在的网段的指派端口是 p(Bj,x)。

证明:⟨Bi,s(Bj,x)⟩∈E2⇔s(Bj,x)是 Bi 的根网段⇔Bi 的根端口 r 所在的网段的指派端口是 p(Bj,x)。

对于一个给定的网桥 Bi∈Nb,对于它的每一个指派端口 x,根据推论 1,我们得到 s(Bi,x)∈Ns 且有向边⟨s(Bi,x),Bi⟩∈E1。

对于一个给定网段 s(Bj,x),我们查找所有根端口所在网段的指派网桥是 Bj 并且指派端口是 Bj 的第 x 个端口的网桥 Bi,根据推论 2,有向边⟨Bi,s(Bj,x)⟩∈E2。

我们从根网桥 R 出发,对于它的每个指派端口 x,构造网段 s(R,x),构造有向边⟨s(R,x),R⟩。对于每一个网段 s(R,x)找到满足根端口所在网段的指派端口是 p(R,x)的所有网桥 B,构造有向边⟨B,s(R,x)⟩。然后假设 B 就是根网桥,重复上述操作。如此循环最终我们得到桥接局域网的第 2 层拓扑结构 D(Nb,Ns,E1,E2)。

3.3 算法分析

本算法在推导网桥之间连接关系时,用到的 MIB 对象每个网桥的每个端口只有 10 个字节(dot1dStpPortDesignatedPort[3]的 2 字节+dot1dStpPortDesignagedBridge[3]的 8 个字节)。而文[4,5]中算法对于每个网桥的每个端口读取了地址转发表,通常这个表远远大于 10 个字节。因此本算法无论是网络传输时间还是内存空间上都要较之文[4,5]中算法更为节省。

本算法的时间复杂度由确定有向边集 E2 的时间花费决定。该值关于网络中网桥数目 n 的 2 次幂呈线性增长趋势,即 O(n*n)。虽然这与文[4,5]中算法的时间花费相当,然而由于本算法的基础运算因子是网桥 ID 和端口号的比较运算也就是整数、字符串上的比较运算,而文[4,5]中算法的基础运算因子是地址转发表、通过集这些集合上的并交运算,因此本算法时间效率更高。

接下来我们考察网络正常稳定的运行情况下算法的准确性。由于地址转发信息的老化机制,网桥端口的地址转发表并非总是完整的,因此,文[4,5]中基于地址转发信息的算法都可能存在端口连接情况无从判定的情况。但是这对于本算法不产生任何影响。网络稳定运行时,其生成树已经确定,通过本算法得到的网络拓扑结构是准确的。

4 算法的实现

我们在一个 vc++ 6.0 程序 TopoViewer 中实现了本算法,这将作为一个模块集成到我们的网管系统中去。TopoViewer 读取保存有 MIB 信息的数据文件,计算出网络的拓扑结构,如图 3 所示。

实际网络中支持生成树协议的透明网桥使用很广泛,然而众多的厂商对于如何读取 MIB 信息并没有一致的接口,甚至有些网桥并不能很好地支持网桥 MIB。因此在网管系统中

(下转第 15 页)

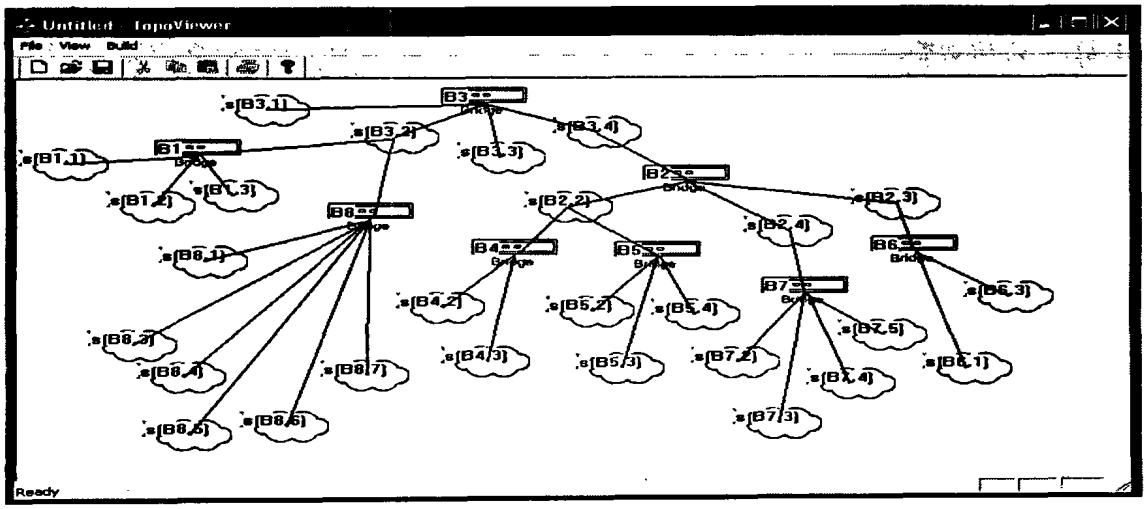


图 3 TopoViewer 基于表 3 中数据绘制的第 2 层拓扑图

格数据及资源服务，以及网格安全服务这三个方面对网格所使用的技术进行了分析。从分析中可以看出，网格体系结构实际上解决的问题是如何将网格中分布的各类资源提供者有效地组织起来，对外形成一个虚拟的整体组织，提供具有统一接口的各类服务；网格数据及资源服务则为网格内部各资源提供者之间，以及资源提供者与用户之间的资源和信息传输提供了一整套机制；网格安全服务的功能对于多数实用的网格系统是不可缺少的，它的目的在于协调网格中资源的提供者和资源的请求者对于资源所要求的控制权限，对于通用的网格安全服务而言，更进一步的要求就是提供尽可能灵活的控制策略。同时，这三类技术又是有着紧密联系的，网格数据及资源服务和安全服务都必然地要在网格体系结构的框架中来实现，而在网格数据及资源服务中也自然会涉及到网格安全的问题。

下面将对这三类技术可能的进一步发展进行讨论。实际上，其中任何一类技术的发展都可能会影响到其它技术的实现方式的改变。

以网格安全服务为例，对于网格安全结构的构建思想的讨论有助于思考对于网格整体的结构可能的扩展，甚至可能导出全新的结构，下面我们从另外一个角度对整个安全结构做一些讨论。

如果从资源提供者的角度来看，4.2.2节中所描述的结构的作法是由资源代理对外部的资源请求进行处理，也就是说，资源是在请求到来之后再动态分配的，这样保证了足够的灵活性，却使得资源分配的效率有所下降，这样在发生某一用户对某些资源的不断重复请求的情形时，灵活性的存在反而成了缺点。基于这种考虑，可以在资源代理对外部请求的处理中对以往的请求加以记录分析，对某一用户常用的资源，可以预先进行分配，并在资源代理处注册，这样当下一次该用户对该资源的请求到来之时，就可以直接对资源进行相应的操作，效率也将有所提高。但需要考虑的是，对以往资源请求的记录处理这一操作对于整体效率也将会有一定影响，如果可能的话可以在进行资源分配的时候进行。这里的讨论实际上已经与资源分配有关(但是对于用户信息的记录处理，则应该在网格安全服务的框架内进行)，这就是属于网格数据及资源服务的范畴了。

关于网格体系结构，现在提出的主要发展方向是开放式网格服务体系，即将构建网格所围绕的中心从网格资源转移到网格服务。这一改变同样对于网格的数据及资源服务和安全服务的实现方式会有相应的影响。

从以上这些论述可以看出，网格体系结构，网格数据及资源服务，和网格安全服务这三类技术对于网格构架的成功与否起着关键性的作用，对于这几类技术的具体实现方式的选择除了需要考虑其各自独立性能，还需要对各类技术之间的相互影响加以分析。

参 考 文 献

1 Beck M, Moore T, Plank J S. Exposed versus Encapsulated Approaches to Grid Service Architecture. GRID 2001, LNCS 2242,2001. 124～132

2 Grimshaw A S, Lewis M J, Ferrari A J, Karpovich J F. Architectural Support for Extensibility and Autonomy in Wide-Area Distributed Object Systems: [UVa CS Technical Report]. 1998(http://legion. virginia. edu/papers/CS-98-12. pdf)

3 Grimshaw A S, et al. Legion: The Next Logical Step Toward a Nationwide Virtual Computer: [UVa CS Technical Report]. 1994, (http://legion. virginia. edu/papers/CS-94-21. pdf)

4 Middleware Overview: Globus/Grids, Ian Foster, (http://www-fp. mcs. anl. gov/middleware98/glogr. ppt)

5 Globus: A Meta Computing Infrastructure Toolkit, Ian Foster, Carl Kesselman. (http://www. globus. org/)

6 Bent J, et al. Flexibility, Manageability, and Performance in a Grid Storage Appliance. HPDC'02, July 2002

7 Thain D, et al. The Kangaroo Approach to Data Movement on the Grid. In: Proc. of the Tenth IEEE Symposium on High Performance Distributed Computing. San Francisco, California, Aug. 2001

8 Iamnitchi A, Foster I. On Fully Decentralized Resource Discovery in Grid Environments. C. A. Lee(Ed.): Grid Computing - GRID 2001 Second International Workshop, Denver, CO, USA, November 12, Proceedings, Lecture Notes in Computer Science, LNCS2242, p. 51 ff. (http://link. springer-ny. com/link/service/series/0558/papers/2242/22420051. pdf)

9 Dongarra H C J. Netsolve's Network enabled Server: examples and applications. IEEE Computational Science & Engineering, 1998, 5(3): 57～67

10 Frey J, et al. Condor-G: A Computation Management Agent for Multi-Institutional Grids. Cluster Computing, 2002, 5(3): 237～246

11 GSI Roadmap Introduction talk from Grid Forum 5(Oct 2000). http://www. gridforum. org

12 Foster I, et al. A Security Architecture for Computational Grids. In: Proc. 5th ACM Conf. on Computer and Communications Security Conf. Describes techniques for authentication in wide area computing environments. 1998. 83～92

(上接第8页)

如何搜集到所需的MIB信息是一个值得关注的问题。例如为不同厂商的设备提供不同的MIB采集程序；对于不支持MIB信息读取的网桥，在其网段上设置代理侦听所有的生成树协议数据包提取相应MIB信息等。

结束语 准确的拓扑结构信息无论是对于网络管理还是一些网络相关的应用都是十分必要的。在本文中我们提出了一个基于标准生成树协议MIB组[3]的第2层拓扑结构发现算法。与基于地址转发信息的拓扑发现算法[4,5]相比，本算法信息采集和处理的开销小。在网络正常运行的情况下，生成树已经确定，运用本算法所得拓扑结构是准确无误的。

IETF于2000年推出物理拓扑MIB[2]，试图解决物理层拓扑结构的发现问题。但是由于没有确定如何获取这些MIB对象的机制，关于网络第2层、第1层拓扑结构的自动发现还有待更多的研究。

参 考 文 献

1 Media Access Control(MAC) Bridges. IEEE 8021. D-1990, May 1990

2 Bierman A, Jones K. Physical Topology MIB. Internet RFC-2922, Sept. 2000

3 Decker E, Langillle P, Rijsinghani A, McCloghrie K. Definitions of Managed Objects for Bridges. Internet RFC-1493, July 1993

4 Breitbart Y, et al. Topology Discovery in Heterogeneous IP Networks. In: Proc. of INFOCOM 2000, March 2000

5 Lowekamp B, O'Hallaron D R, Gross T R. Topology Disovery for Large Ethernet Networks. ACM SIGCOMM '2001

6 耿素云. 集合论与图论. 北京大学出版社, 1998

7 Tanenbaum A S. Computer Networks. Prentice Hall PTR, 1996

8 McCloghrie K, Rose M. Management Information Base for Network Management of TCP/IP-based internets: MIB-II Internet RFC-1213, March 1991