采用单测量源的拓扑推断算法

赵洪华[1]，丁 科[1]，陈 鸣[1]，张婷婷[2]，金凤林[1]，贺 汛[1]

(1. 解放军理工大学指挥自动化学院 南京 210007; 2. 解放军理工大学理学院 南京 210007)

【摘要】为了减少拓扑推断中需要时钟同步和节点间合作的限制，提出了一种仅需要单个测量源并且不需要时钟同步的“运输车”测量方法，该方法可以测量目的地址共享链路的排队时延。设计了根据排队时延推断拓扑结构的算法。理论证明了基于“运输车”测量方法和排队时延推断网络拓扑的可行性和正确性，并通过NS2进行了仿真，仿真结果表明基于“运输车”测量方法和排队时延能够准确的推断网络拓扑结构。

关 键 词 相关性; 网络层析成像; 拓扑推断; 运输车

中图分类号 TP393 **文献标识码** A doi:10.3969/j.issn.1001-0548.2010.02.026

Topology Inference Algorithm by Using One Measuring Node

ZHAO Hong-hua[1], DUNG Ke[1], CHEN Ming[1], ZHANG Ting-ting[2], JIN Feng-lin[1], and HE Xun[1]

(1. Institute of Command Automation, PLA University of Science and Technology Nanjing 210007;

2. Institute of Science, PLA University of Science and Technology Nanjing 210007)

Abstract In order to reduce the limit of time synchronization and cooperation between nodes, a measurement method named “transport train” is proposed. The measurement method could measure the queue delay of share links between nodes. by using only one measuring node and without need of time synchronization, A topology inference algorithm is put forward based on queue delay. The feasibility and correctness of topology inference algorithm based on queue delay and “transport train” measurement method are analyzed theoretically. The algorithm is simulated by NS2, the results validate that topology inference algorithm based on queue delay and “transport train” measurement method could infer network topology correctly.

Key words correlation; network tomography; topology inference; transport train

网络拓扑推断是网络层析成像技术(network tomography)[1-2]的最新应用之一，该技术根据网络中节点性能特性的相关性来推断网络的拓扑结构，因为研究表明网络中节点的共享链路越多，节点的性能越相近，即相关性越大[3]。

当前基于层析成像技术的拓扑推断在测量节点性能的过程中受到较多的限制，例如需要测量源节点和测量目的节点的配合、节点间的时钟同步等，实用性较低。为了提高拓扑推断技术的实用性，本文中提出了一种仅需要单个测量源的测量方法，并设计了相应的拓扑推断算法。

1 相关研究

基于网络层析成像技术的拓扑推断主要分为两个步骤：(1) 通过端到端的测量获得测量源节点到目的节点的端到端性能参数，根据端到端性能参数计算出目的节点间的相关性；(2) 根据目的节点间的相关性推断网络的拓扑结构。

层析成像技术中的端到端性能测量不同于普通的性能测量，需要采用特殊的测量方法，当前比较常用的测量方法是紧接(back to back)分组对[4-8]测量方法。紧接分组对由两个大小相同的分组组成，两个分组分别到达不同的目的地址，分组之间具有非常小的时间间隔。该方法正是利用分组之间较小的时间间隔使到达不同节点的测量分组在节点共享链路部分经历的网络状况相同。紧接分组对测量方法的实例如图1所示。

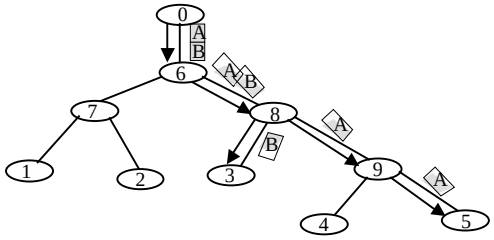

图1 紧接分组对测量方法

通过紧接分组对测量方法可以测量端到端单向时延和端到端单向丢包率，根据单向时延或单向丢

收稿日期：2008 – 09 – 10; 修回日期：2009 – 03 – 12

作者简介：赵洪华(1979 –)，男，博士，主要从事网络拓扑发现等方面的研究.

包率可以计算节点间的相关性，计算的相关性包括时延协方差或成功传输率。

当前的推断算法包括分层算法[6-7]、LBT算法[9]、DBT算法[10]和MLT[11]算法，这4种拓扑推断算法都基于节点间的相关性来推断网络的拓扑结构，例如可以通过时延协方差或成功传输率推断网络拓扑结构。

紧接分组对测量方法可以较准确地测量目的节点共享链路部分的性能特性，但该方法需要节点间的时钟同步及测量节点与目的节点的配合，因此在实际应用中受到较多的限制。

为了提高拓扑推断的实用性，提出了“运输车”测量方法和相应的拓扑推断算法。

2 “运输车”测量方法

“运输车”由3个分组组成，其中两个Ping分组和一个长分组，长分组位于两个Ping分组之间，两个Ping分组具有相同的目的地址，长分组具有另一个目的地址并且分组长度较长，例如可以设为1 000 B。为了减少测量分组对网络的影响，在“运输车”中Ping分组的长度较短，如可以设为50 B。

“运输车”测量方法中两个Ping分组到达目的地址后，目的地址返回ICMP的ECHO Reply分组到测量源节点，长分组直接到达目的地址，在两个目的地址的共享链路部分，长分组在中间节点排队，导致Ping分组之间的间隔增大。

在测量源节点根据返回的ECHO Reply分组的时间间隔计算目的地址对共享链路部分的排队时延，由于只计算时间间隔，因此“运输车”测量方法不需要节点间的时钟同步，并且不需要节点间的配合。采用“运输车”测量方法的实例如图2所示。图中两个Ping分组的目的地址为节点3，长分组的目的地址为节点4，在节点0计算返回分组的时间间隔。

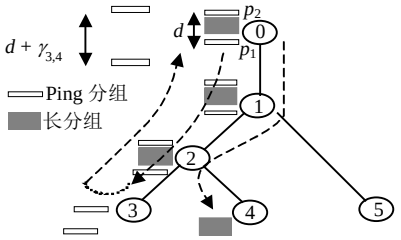

图2 “运输车”测量方法

如图2所示，“运输车”中的长分组到达节点4后丢弃，两个Ping分组到达节点3后返回ECHO Reply分组到测量源节点0，在测量过程中，长分组在节点对(3,4)的共享链路部分(0→1→2)对Ping分组之间的时间间隔产生影响，导致Ping分组之间的间隔增大，在往返路径的其余部分，分组之间的间隔不再增大。

在“运输车”测量方法中，为了准确测量目的地址共享链路的排队时延，分组的长度需要满足下列条件。

在“运输车”测量方法中，设长分组的长度为$s(q)$，Ping分组的长度为$s(p)$，设分组经过的共享链路为$(L,L_2,\cdots,L_m)$，链路的带宽为$(b_1,b_2,\cdots,b_m)$，则在共享链路部分长分组影响第二个Ping分组排队时延的条件为：

$$\frac{s(q)}{s(p)} \geqslant \mathrm{Max}\left[10, \frac{b_{i+1}}{b_i}(i=1,2,\cdots,m)\right] \tag{1}$$

设“运输车”测量方法中两个Ping分组的时间间隔为d，返回分组的时间间隔为$d+\Delta d$，设在目的地址的非共享链路部分，Ping分组经过的物理链路为$L_{m+1},L_{m+2},\cdots,L_n$，链路带宽为$b_{m+1},b_{m+2},\cdots,b_n$。在目的地址对非共享链路部分背景流量对Ping分组的时间间隔不产生影响的条件为：

$$\frac{s(p)}{d+\Delta d} \leqslant \min(b_i) \qquad m+1 \leqslant i \leqslant n \tag{2}$$

3 拓扑推断算法

通过“运输车”测量方法可以测量目的地址对共享链路部分的排队时延，在树状结构的拓扑中，目的地址对的共享链路越多，则排队时延越大，因此可以把排队时延作为拓扑推断的节点间相关性值，为了推断准确，测量时通常采用多次测量的方式。

在树状拓扑结构中，设$T=(V,L)$表示拓扑树，其中V为节点集，L为链路集，0表示根节点，R表示所有的叶节点，$R=\{1,2,\cdots,i,\cdots,j\}$，$a(i,j)$为节点$i$和节点$j$的父节点，$n_i<n_j$ 表示n_i是n_j的子孙。如果$a(k,j)<a(i,j)$，则节点k和j的共享链路比节点i和j的共享链路多。

3.1 拓扑推断参数计算

在“运输车”测量方法中，对于每个目的地址对通常发送50组“运输车”测量序列，每组测量序列中都有20个“运输车”，则测量过程中产生的流量与紧接分组对测量方法相当[6-7]。

设“运输车”测量分组的目的地址对为(i,j)，Ping分组的目的地址为i，长分组的目的地址为j，“运输车”中Ping分组的时间间隔为d，在测量源节点共收到50组ICMP ECHO Reply分组，每组测量序列中包括20对ICMP ECHO Reply分组。

设在测量源节点收到返回分组的时间集合为T_1，$T_1=\{T_{1,1},T_{1,2},T_{2,1},T_{2,2},\cdots,T_{i,1},T_{i,2},\cdots,T_{N,1},T_{N,2}\}$，通过第一个“运输车”获得的排队时延为$(T_{1,2}-T_{1,1})-d$，通过第$N$个“运输车”获得的排队时延为$(T_{N,2}-T_{N,1})-d$。

设m组“运输车”测量序列获得的测量源到目的地址对(i,j)的排队时延为$(\hat{d}_{1,1},\hat{d}_{1,2},\cdots,\hat{d}_{1,n},\cdots,\hat{d}_{1,20})$，$\cdots,(\hat{d}_{m,1},\hat{d}_{m,2},\cdots,\hat{d}_{m,n},\cdots,\hat{d}_{m,20})$，其中$d_{1,1}$表示在第1个“运输车”测量中返回分组之间的时间间隔。对目的地址对(i,j)所有的时间间隔计算均值，计算公式为：

$$\overline{D}_{i,j}=\sum_{m=1}^{50}\sum_{n=1}^{20}\hat{d}_{m,n}/20\times 50 \tag{3}$$

为了方便比较和分析，采用标准化的排队时延作为拓扑推断参数。设拓扑树T中有M个叶节点，则共有$\binom{M}{2}$个样本排队时延值。令所有叶节点对的排队时延集合为D，则$D=\{\overline{d}_{1,2},\overline{d}_{1,3},\cdots,\overline{d}_{i,j},\cdots,\overline{d}_{m-1,m}\}$。设$D(D)$为$D$的方差，Min$(D)$为$D$最小值，对集合$D$中的元素作标准化计算，生成新的集合$T$，其中每个元素的标准化公式为：

$$T_{i,j}=\frac{\overline{D}_{i,j}-\mathrm{Min}(D)}{\sqrt{D(D)}} \tag{4}$$

3.2　拓扑推断算法

通过上面计算的标准化排队时延作为节点间的相关性，可以推断网络的拓扑结构，推断算法如下。

输入：根节点{0}、叶节点集V_r={{1},{2},⋯,{M}}、边集$L=\varnothing$、$T=\{T_{1,2},T_{1,3},\cdots,T_{i,j},\cdots,T_{m-1,m}\}$

输出：Tree=(V,L)

While $|V_r|>1$ do

Begin

根据T对节点集进行划分，划分后的结果为$V_r=A_1\cup,A_2\cup,\cdots,\cup A_n$，对于每一个集合$A_i$，$\forall i\forall j(i\in A_i,j\in A_i)\rightarrow T_{i,j}>T_{i,m}(\forall_m m\notin A_i)$；

For each ($A_i\subset V_r$) do

Begin

If A_i 内节点数>1 then

Begin

合并A_i内的节点，生成新的节点m；

更新T，在T中减去A_i中节点的相关性值，并加入新值$T_{m,k}(k\notin A_i)=\mathrm{Max}(T_{i,k}(i\in A_i))$；

更新节点集V，加入新节点m，$V=A_i\cup\{m\}$；

更新边集L，$L=L\cup\{(m,x),x\in A_i\}$；

更新叶节点集V_r，$V_r=\{V_r-A_i\}$　　加入{m}；

End；

End；

End；

拓扑推断算法中有节点集的划分和添加新节点后相关性值的计算两个重要部分。节点集的划分是把所有的兄弟节点划分在同一个集合中，划分时采用Min($T_{i,j}$)作为参考值。兄弟节点合并后计算新的相关性值，计算时采用兄弟节点中与其他节点相关性最大的节点代表新节点，有：

$$T_{m,k}(k\notin A_i)=\mathrm{Max}(T_{i,k}(i\in A_i)) \tag{5}$$

4　仿真实验及结果

为了验证“运输车”测量方法和基于排队时延的拓扑推断算法的正确性，在NS2[12]环境下进行了仿真实验，在不同网络负载的情况下验证了算法的正确性。

仿真实验采用较为普遍的树状结构的网络拓扑，由于基于单个发送节点，多个接收节点的拓扑都为树状结构，因此采用树状结构的拓扑具有普遍意义。

树状拓扑如图3所示，图中与叶节点相连的链路带宽为1 Mb/s，链路时延为10 ms。内部链路带宽为2 Mb/s，内部节点采用RED的丢包策略。

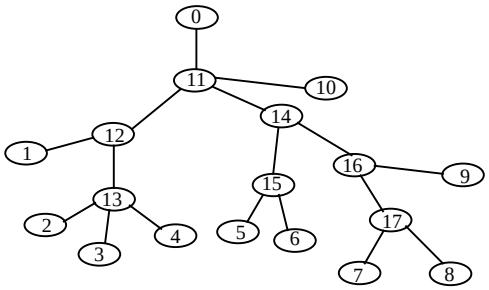

图3　NS2仿真拓扑结构

采用“运输车”测量方法在根节点0向所有叶节点对发送“运输车”测量分组，背景流量加入自相似流和泊松流。

在网络负载较轻和网络负载适中的环境下进行了多次实验，不同网络的负载情况验证了测量方法的有效性。当“运输车”测量分组数较多时，根据排队时延计算的节点相关性均能较准确反映节点共享链路的情况，图4为根据“运输车”测量方法测量的性能参数推断的拓扑结构，图中内部节点按推断的顺序编号。

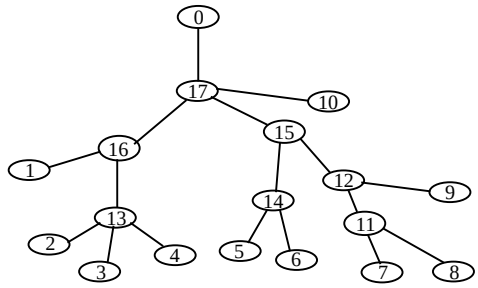

图4 推断的拓扑

在相同的仿真环境下，采用紧接分组对测量方法测量的单向时延和丢包率各进行了50次拓扑推断的仿真实验，表1为在相同的环境下不同测量方法正确推断拓扑结构情况的对比(其中*X*/*Y*，*X*为正确推断次数，*Y*为仿真次数)。

表1 3种性能参数推断拓扑结构的对比

	轻载/次数	中载/次数	正确率/(%)
单向时延	36/50	46/50	82
丢包率	21/50	39/50	62
排队时延	42/50	35/50	77

通过表1可知，采用紧接分组对测量的单向时延推断拓扑结构的效果最好，但单向时延的测量需要节点间的时钟同步和节点间的合作。采用“运输车”测量的排队时延推断拓扑的效果与单向时延推断拓扑的效果相差不大。采用紧接分组对测量的丢包率推断拓扑结构的效果最差，并且需要节点间的合作。

图5和图6分别为网络负载较轻和网络负载适中时，根据“运输车”测量方法测量的排队时延计算的一次具体节点对的相关性值。采用三维图形显示节点对的相关性值，在三维图形中*X*轴表示节点*i*，*Y*轴表示节点*j*，*Z*轴表示节点对(*i*,*j*)的相关性值。

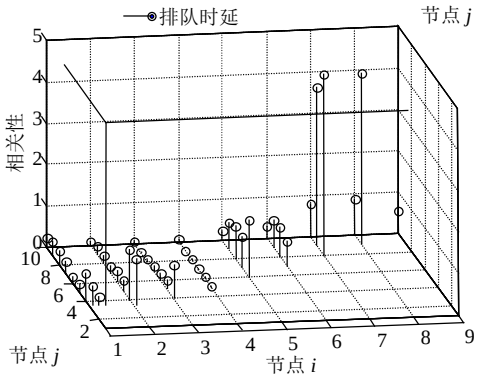

图5 网络负载较轻时节点对相关性

对比图5和图6，在网络负载较轻和网络负载适中时计算的节点相关性聚类效果较好，在树状结构的拓扑中具有相同共享链路的节点对的相关性值相等，节点间的相关性能够反映节点共享链路的情况。通过多次仿真实验表明，在网络负载较轻和网络负载适中的情况下，通过“运输车”测量方法测量的排队时延计算节点相关性能够较准确地推断网络拓扑结构。

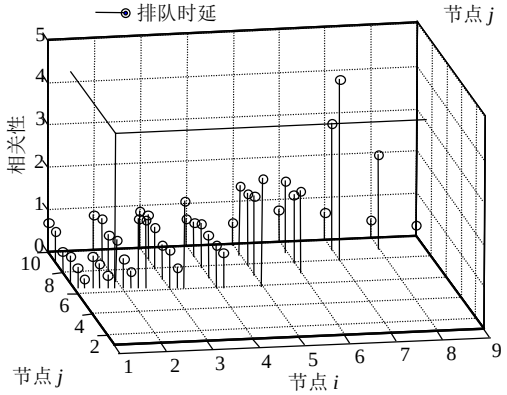

图6 网络负载适中时节点对相关性

5 总 结

为了提高拓扑推断技术的实用性，本文设计了可以测量目的节点对共享链路排队时延的“运输车”测量方法，该方法仅需一个测量源。此外，还设计了相应的拓扑推断算法，并通过NS2仿真验证了算法的正确性。

参 考 文 献

[1] CASTRO R, COATES M, NOWAK R, et a1.Internet tomography[J]. IEEE Signal Process Mag, 2002,19(3): 47-65.

[2] CASTRO R, COATES M, LIANG G, et al. Network tomography: Recent developments[J]. Statistical Science, 2004, 19(3): 499-517.

[3] RATNASAMY S, MCCANNE S. Inference of multicast routing trees and bottleneck bandwidths using end-to-end measurements[C]//IEEE INFOCOM. New York: [s,n.], 1999: 353-360.

[4] DUFFIELD N, PRESTI L, PAXSON V, et al, Network losstomography using striped unicast probes[J]. IEEE/ACM Transactions on Networking, 2006, 14(4): 697-710.

[5] MENG S, ALFRED O. Unicast-based inference of network link delay distributions with finite mixture models[J]. IEEE Trans on Signal Processing, 2003, 51: 2219-2228.

[6] MENG S, ALFRED O. Topology discovery on unicast networks: A hierarchical approach based on end-to-end measurements[EB/OL]. [2008-05-18]. Http://www.eecs.umich.edu/msim/Publications/cspl-357.ps.pdf, 2005.

[7] MENG S, ALFRED O. Hierarchical inference of unicast network topologies based on end-to-end measurements[J]. IEEE Transactions on Signal Processing, 2007, 55(5): 1708-1718.

(下转第310页)