



基于 H3 的拓扑可视化研究及实现

辛富国¹, 李荣芳¹, 王中振², 权东晓²

(1. 陕西邮电职业技术学院 咸阳 712000; 2. 西安电子科技大学 西安 710071)

摘要:给出了一种快速且对网络产生负载较小的网络拓扑发现算法,特点是利用数据库存储数据,可以方便地对数据进行检索和处理。基于 H3 拓扑可视化原理对数据进行预处理,得到需要的输入文件。最后利用双曲空间对网络拓扑进行显示。测试结果表明,该方法可以很好地对网络拓扑进行可视化,显示效果美观。

关键词:网络拓扑发现;拓扑可视化;H3 可视化;双曲空间

doi: 10.3969/j.issn.1000-0801.2014.05.018

Research and Implementation of Network Topology Visualization Based on H3

Xin Fuguo¹, Li Rongfang¹, Wang Zhongzhen², Quan Dongxiao²

(1. Shaanxi Post and Telecommunication College, Xianyang 712000, China;

2. XiDian University, Xi'an 710071, China)

Abstract: A network topology discovery algorithm was given which had the characteristic of rapid and less extra traffic. The path data were stored by database to facilitate the retrieval and processing of data. Then based on the principle of H3 visualization, the source data was processed to get the needed input file. Finally the network topology was displayed in the hyperbolic space. The test results show that the method is suitable for the network topology visualization, and it can display the network topology beautifully.

Key words: network topology discovery, topology visualization, H3 visualization, hyperbolic space

1 引言

随着社会的不断发展,网络在人们日常生活和工作中的作用越来越重要。网络的功能越来越强大,网络的规模也在不断扩大,流量也不断上升。即使是网络管理员也可能并不清楚网络的拓扑结构,因此拓扑发现的作用变得越来越重要。拓扑发现包括合作探测和非合作探测^[1]。合作探测时,管理员可以通过访问路由表、MBI 库^[2]等来辅助探测拓扑,这能够帮助管理员了解网络的连接关系,同时根据流量信息对网络进行扩容;非合作探测主要用作攻击,指的是在没有访问权限的情况下,通过向网络发送数据分

组,根据其返回的信息来推测网络拓扑。传统的算法是基于 ICMP(internet control messages protocol,互联网控制报文协议)来进行拓扑发现^[3],由于一些路由器不响应,使得发现的拓扑不完整,较好的解决方案是结合流量、时延^[4]、地址转发表^[5]等信息进行网络断层扫描^[3-5],推测拓扑信息。

通过拓扑发现可以得到网络地址的连接关系,但是如何将得到的拓扑信息很好地显示给用户,也是一个比较困难的问题。拓扑可视化就是通过已知的拓扑数据将目标网络的节点和连接状况完整、清晰地呈现在人们眼前,为人们分析、了解目标网络的状况提供依据。单点拓扑发现的结果是树型结构,如果以平面的形式来显示^[6],由于树的节

点数量随着树的深度的增加呈现指数增长,那么能够显示的节点数量就是有限的,并且层次越深,节点的符号就越小。但是由于双曲空间中空间大小呈现指数增长,正好与树的节点增长的趋势一致,所以可以借助双曲空间来进行显示^[7]。基于 H3 的拓扑可视化^[8]就是基于这样的原理实现的。

本文首先给出一个高效的网络拓扑发现算法,然后详细介绍了基于 H3 的拓扑可视化的基本原理,最后基于 MFC(microsoft foundation class,微软基础类)进行了编程实现并以其对校园网进行了测试与分析。

2 算法描述

2.1 拓扑发现算法

在这里采用常用的基于 ICMP 的拓扑发现算法,首先利用 ping 程序探测主机是否在线,如果在线则逐步改变 TTL (time to live, 生存时间)值进行路径探测。为了提高效率,对算法做了以下改进。

一是采用多线程编程来进行探测。由于利用 ping 程序探测时,如果主机在线则会立即返回信息,否则必须默认等待超时后才能确定主机不在线,需要的时间比较长。若主机在线,需要逐步改变 TTL 值进行路径探测。因此当探测的地址空间较大时,单线程执行需要的时间会很长,所以采用多线程编程来提高探测速度。

二是采用由远及近的探测方式。在图 1 所示的拓扑结构中,假设主机 A 探测到主机 H 在线后,依次设定 TTL 值为 2 和 1,则分别会收到节点 G 和 F 返回的超时信息 ICMP_timeout。A 到 H 的路径探测结束,然后探测节点 J,首先设置 TTL 值为 3,则会接收到节点 I 返回的信息,然后设置 TTL 值为 2,接收到节点 G 返回的信息。由于在前面已经探测过 A 到 G 的路径了,因此后面就没有必要设置

TTL 值为 1 继续探测,提高了探测速度。在真实网络中,特别是探测距离较远的外网时,能大大提高探测速度,减少发送分组的数量,对网络造成的影响也较小。

2.2 数据预处理

拓扑探测得到的路径信息,用数据库进行存储,包括目的节点以及经过的每一个路由器的 IP 地址。由于有的路由器对 ICMP 数据分组没有响应,如果没有收到回复,在数据库中则将路由器的地址记为 255.255.255.255,方便以后将 IP 地址转换成整数进行压缩存储。

如果对图 1 所示的拓扑结构进行探测,则探测结果见表 1。

表 1 图 1 的拓扑探测结果

目的节点	第 1 跳	第 2 跳	第 3 跳	第 4 跳	第 5 跳
C	B	C			
E	E				
D	B	D			
H	F	G	H		
K	K				
J	F	G	I	J	

而 H3 可视化软件的输入文件中每一行代表一个节点,其内容格式为:depth identifier [...]。

depth 是指当前节点在骨干树中的深度,如果当前行 A 的前面数行中某行 B 的深度比行 A 的深度小 1,则说明在树中 A 是 B 的孩子节点,有一个连线从 B 到 A。identifier 用来描述节点,一般为 IP 地址等信息,[]中的信息为附加信息,用来描述节点是主机、路由器、显示标志等。图 1 所示的网络拓扑结构,其对应的文件格式为:

```
0 A []
1 B []
2 C []
2 D []
1 E []
1 F []
2 G []
3 H []
3 I []
4 J []
1 K []
```

观察该文件和图 1 可知,该文件的输入顺序类似于图的深度优先遍历序列。要实现拓扑的可视化,关键是如何

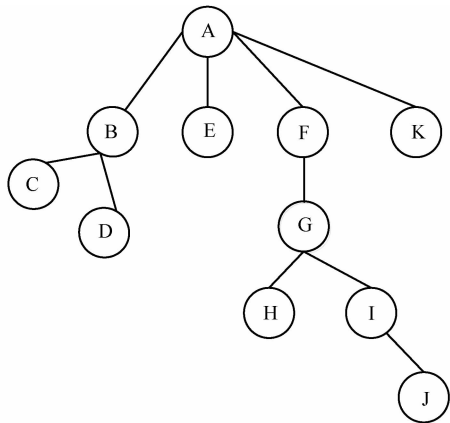


图 1 网络拓扑结构示意图



从表 1 得到 H3 要求的输入文件, 最终将一条条独立记录的路径信息归并成一棵树。并且在实现时必须考虑算法的高效性, 因为当拓扑记录达到数千上万条的时候, 时间复杂度超过 $O(n^3)$ 的算法都将会变得非常缓慢。

在处理时, 可以对数据库的内容先进行 hop_1 (第 1 跳) 的排序, 再对 hop_2 进行排序……在对 hop_30 进行排序后 (假设最多为 30 跳), 表 1 将会变成表 2 的结果。

表 2 排序后的探测结果

目的节点	第 1 跳	第 2 跳	第 3 跳	第 4 跳	第 5 跳
C	B	C			
D	B	D			
E	E				
H	F	G	H		
J	F	G	I	J	
K	K				

从表 2 中的数据可以看出, 相似度越高的记录越邻近, 那么在遍历数据库生成 H3 需要的文件的时候, 只需要比较当前记录和上一条记录的差异, 就可以依据这些差异添加相应的节点, 从而减少了不必要的比较。同时, 该过程中对每一条记录的每一跳的遍历刚好是一个深度遍历的序列, 这样就可以生成 H3 需要的文件。该算法的时间复杂度是 $O(n)$, 速度比较快。

2.3 拓扑可视化

得到文件以后, 就要根据文件对拓扑结构进行可视化。由于基于单点探测的网络拓扑是一个树型结构, 需要布局的节点数量随着树的深度的增加呈现指数增长。而在欧几里得空间中可用的布局空间则呈现多项式增长, 这样为了显示结果就要将离根节点较远的节点占用的空间减小, 从而导致能够看到离根节点较近的拓扑连接关系, 而较远的节点无法看清。由于双曲几何空间中可用的布局空间随着半径的增长呈现指数形式的增长, 所以 H3 借助双曲几何空间进行显示。通过投射模型将双曲空间投影到欧几里得空间进行显示, 因为直线的显示速度要快很多, 在用户拖动查看拓扑结构时速度较快。

在树的布局上, H3 可视化算法是由施乐帕克研究中心开发的圆锥树^[9]系统扩展来的。圆锥树将叶子节点布局在从双亲节点发散出来的圆锥的圆周上, 而 H3 算法则将叶子节点布局在一个覆盖在圆锥之上的半球面上。这个算法要进行两趟遍历: 自底向上执行一趟来估算每个半球容

纳所有的孩子半球所需要的半径大小, 自顶向下执行一趟来将每个孩子半球布局到双亲半球的表面。这两步是不能合并的, 因为在布局孩子半球的时候需要先确定双亲半球的半径^[8]。

(1) 自底向上的估算

设双亲半球的半径为 r_p , 在 $p+1$ 层有 n 个孩子半球, 每个孩子半球的地面圆的面积为 D_k , 则 p 层的半球的表面积为 $p+1$ 层的所有孩子半球的底面圆的面积和, 即:

$$H_p = \sum_{k=1}^n D_k = \sum_{k=1}^n \pi r_k^2 \quad (1)$$

从而得到双亲半球的半径为:

$$r_p = \sqrt{H_p / 2\pi} \quad (2)$$

则对应到双曲空间分别为:

$$H_p = \sum_{k=1}^n D_k = \sum_{k=1}^n 2\pi (\cosh r_k - 1) \quad (3)$$

$$r_p = \sinh^{-1} \sqrt{H_p / 2\pi} \quad (4)$$

通过自底向上的估计, 可以得到所需要的球的半径, 下面通过自顶向下的方式来布局各个节点。

(2) 自顶向下的布局

在布局的时候, 所有的孩子半球都依据其子孙的数量进行排序, 依据孩子半球的大小自内向外地在环状带上填充, 子孙数量最多的填充到极点位置。图 2(a) 为填充方式的侧视图, 图 2(b) 为填充方式的俯视图, 从图 2(c) 可以看出如果不对节点进行排序, 则会浪费很多空间。

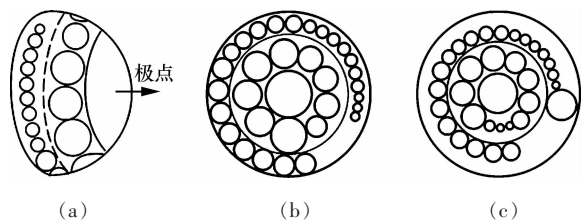


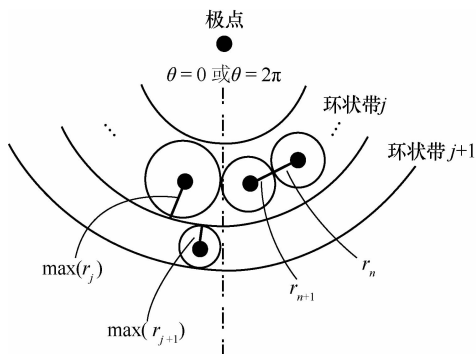
图 2 填充解决方案示意

设每个孩子半球的坐标为: (r_p, ϕ, θ) , 通过第一步, 已经得到了半径 r_p , 下面求出 ϕ 和 θ 。

由于图 3 可以得到, 两个相邻的孩子节点的 θ 的增量取决于两个孩子半球的半径, 通过推导可以得到:

$$\delta\phi_{\text{incr}} = \arctan\left(\frac{\tanh r_n}{\sinh r_p \sin \phi}\right) + \arctan\left(\frac{\tanh r_{n+1}}{\sinh r_p \sin \phi}\right) \quad (5)$$

在一个环状带内, 顺时针不断增大 θ 值来布局孩子节

图3 孩子半球的布局点 θ 的变化

点,当环状带 j 填充至 $\theta=2\pi$ 时,则移动至下一个离极点更远的环状带 $j+1$ 。由于对节点进行了排序,所以每个环状带内的第一个节点半径最大,可以通过第一个孩子半球的半径求得 ϕ 的增量:

$$\delta\phi_{\max} = \arctan\left(\frac{\tanh r_j}{\sinh r_p}\right) + \arctan\left(\frac{\tanh r_{j+1}}{\sinh r_p}\right) \quad (6)$$

(3)测试与分析

基于提出的拓扑发现算法和可视化方法,在VC6.0环境下借助MFC编写了拓扑发现软件,数据库采用MySQL,对校园网进行了测试,测试结果如图4所示。

左侧以列表的形式显示发现的主机,通过点击“+”号可以将到达目标节点的路由信息显示到下方的编辑框里。通过多次测量,得到了254个主机,并绘制了拓扑结构,初始拓扑中,根节点的 ϕ 和 θ 均为0,显示在球表面的中心位置。通过拖拽可以实现旋转效果,方便观察感兴趣的节点,如图5所示。

3 结束语

本文给出了一种快速且对网络产生负载较小的拓扑发现算法,利用数据库来存储信息,方便程序对其进行快速的检索。并基于H3的拓扑可视化原理对数据进行

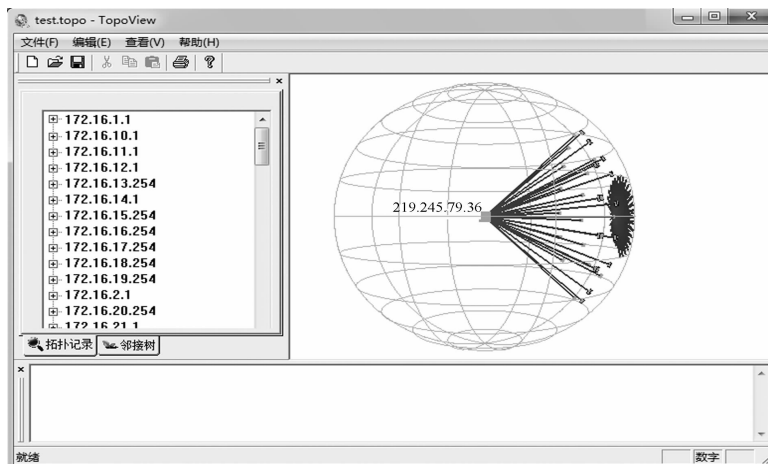


图4 拓扑结构初始状态

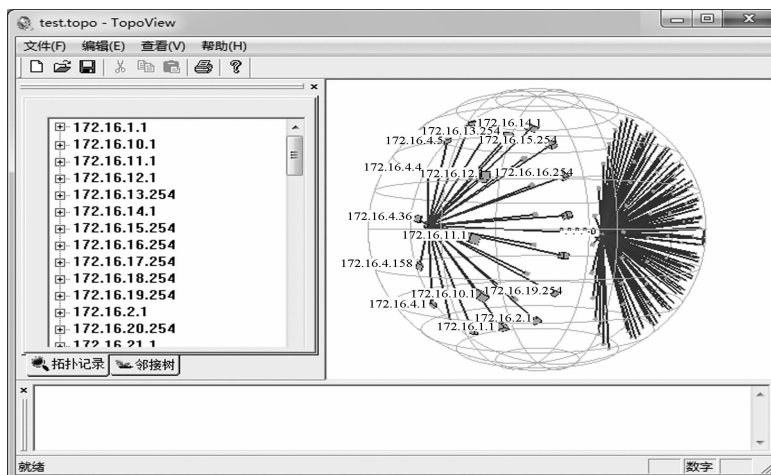


图5 旋转后的拓扑结构



预处理,最终实现了拓扑可视化。测试结果表明,该方法可以很好地对网络拓扑进行可视化,但是发现的拓扑不是很完整,需要通过多种手段识别匿名路由器,借助网络断层扫描技术推测拓扑结构,也可以进行分布式拓扑发现。

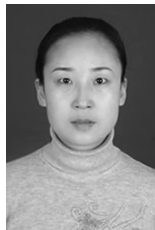
参考文献

- 1 闫兴纂,殷建平,蔡志平. 网络拓扑发现算法综述. 计算机工程与应用, 2007, 43(14): 131~135
- 2 卢红梅. 一种网络拓扑算法的研究和分析. 科技信息, 2012(31): 144~145
- 3 刘香丽,吴辰文,茹俊年等. 基于 Tarceroute 和邻接分组对的网络拓扑推测方法. 兰州交通大学学报, 2013, 32(1): 88~91
- 4 张耀方,孔德弟,石佳玉. 改进的网络拓扑推断算法. 电子测试, 2014(1): 36~41
- 5 李丹程,马东琳,韩春燕等. 面向 Trunk 技术的网络拓扑发现算法研究, 2012, 33(11): 2435~2441
- 6 徐阔海. 一种基于 Rapha I 图形库的网络拓扑生成系统. 软件导刊, 2014, 13(1): 149~151
- 7 Lamping J, Rao R. Laying out and visualizing large trees using a hyperbolic space. Proceedings of UIST'94, Marina del Rey, California, USA, 1994: 13~14
- 8 Munzner T. Interactive visualization of large graphs and networks. Doctoral Dissertation of Stanford University, 2000
- 9 Robertson G, Mackinlay J, Card S. Cone trees: animated 3D visualizations of hierarchical information. Proceedings of the Conference on Human Factors in Computing Systems (CHI '91), New Orleans, Louisiana, USA, 1991: 189~194

[作者简介]



辛富国,男,陕西邮电职业技术学院讲师、教务处副处长,主要研究方向为 GNSS(全球卫星导航系统)、网络抗干扰、光通信技术。



李荣芳,女,陕西邮电职业技术学院讲师,主要研究方向为数据挖掘、算法与数据结构、多媒体技术。



王中振,男,西安电子科技大学硕士生,主要研究方向为网络拓扑发现、拓扑可视化。



权东晓,女,博士,西安电子科技大学副教授、硕士生导师,主要研究方向为量子通信、图形图像处理。

(收稿日期:2014-03-16)

(上接第 111 页)

TD-LTE 已成为主流移动通信标准之一,初步具备同发达国家展开下一代通信技术标准竞争的實力,应权衡我国国情和风险承担能力,从国家安全角度出发,合理制定新通信标准的开发策略和管制措施,积极参与国际合作,在 5G 等新一代通信标准开发中争取更多的主动权,提高国家核心竞争力。

参考文献

- 1 葛亚力. 技术标准战略的构建策略研究. 中国工业经济, 2003 (6): 91~96
- 2 骆品亮,殷华祥. 标准竞争的主导性预期与联盟及福利效应分

析. 管理科学学报, 2009, 12(6): 1~10

- 3 胡武婕. 中国信息通信产业技术标准竞争与策略研究. 北京邮电大学博士学位论文, 2010

[作者简介]

徐朝锋,男,北京邮电大学经济管理学院博士生,主要研究方向为技术标准竞争、通信管制等。

忻展红,男,北京邮电大学经济管理学院教授、博士生导师,主要研究方向为运筹学应用、通信管理、服务科学与工程等。

(收稿日期:2014-04-29)