非协作方式下 IPv6 接入网络拓扑发现*)

刘振山　王清贤　罗军勇

(国家数字交换系统工程技术研究中心　郑州 450002)

摘　要　本文致力于提高非协作方式下 IPv6 接入网络拓扑发现的覆盖率，首先建立了描述 IPv6 接入网络的前缀聚合树模型(PATM，Prefix Aggregation Tree Model)，然后提出了基于 PATM 模型的网络拓扑发现方法。最后给出了对华东某高校 IPv6 校园网的拓扑发现测试结果。通过对测试数据进行的分析和对比，验证了基于 PATM 模型的拓扑发现方法的有效性。

关键词　前缀聚合树模型，非协作，兄弟节点，前缀跨度

Non-cooperative Topology Discovery Method for IPv6 Access Networks

LIU Zhen-shan　WANG Qing-xian　LUO Jun-yong

(National Digital Switching System Engineering & Technology R&D Center, Zhengzhou 450002, China)

Abstract　This paper is mainly applied to some methodologies on improving the coverage of non-cooperative topology discovery for IPv6 access networks. This paper proposed the PATM (Prefix Aggregation Tree Model) to depict the IPv6 access networks, and it proposed the topology discovery methodology on the base of the PATM. In the end, the paper proposed the results of topology discovery for an IPv6 access network of a university in the east of china and proved the efficiency of topology discovery methodology with the analysis of the test results.

Keywords　Prefix aggregation tree model, Non-cooperative, Brother nodes, Prefix span

1　引言

根据被测网络的运营者协作与否，网络拓扑发现可以分为协作式(cooperative)网络拓扑发现和非协作(non-cooperative)式网络拓扑发现。在网络运营者配合下，协作式网络拓扑发现一般能够准确、有效地得到拓扑发现结果，促进网络检测和管理。非协作式网络拓扑发现不需要网络运营者的参与。非协作方式下的 IPv6 网络拓扑发现总的要求是：在已知目标点数量非常有限的前提下，探测源从远程进行拓扑发现来获得目标网络内尽量完整的网络拓扑结构，同时确保拓扑发现过程遵循低负载、高效和弱感知性的原则。Cheswick 等人进行的 Internet Mapping 项目[1]采用单点非协作式探测方式，曾成功地描述了科索沃战争期间由于网络设施或电力供应设施被破坏而使南斯拉夫地区的网络拓扑结构发生显著变化的情况(1999 年 5 月)。

根据被探测目标网络的范围，IPv6 网络拓扑发现可以简单地分为 IPv6 接入网络拓扑发现和 IPv6 公共网络拓扑发现。现有的 IPv6 接入网络拓扑发现技术[2-4]基本上都采用协作式，要求在网络运营者配合下，设置多个探测代理来发现网络拓扑结构。这种技术易于实现，但只限于网络管理方面的应用。现有的公共 IPv6 网络拓扑发现技术虽然不需要网络运营者的配合，但实现的前提条件仍然比较苛刻。例如：贝尔实验室开发的 Atlas 系统采用探测引擎与源路由选择(source routing)功能相结合的技术[5]，可以从一个网络节点探测一大片 IPv6 地址空间内的网络拓扑结构，其前提条件是已知这片地址空间中足够数量的目标 IPv6 地址；现有的多点 IPv6 拓扑发现工具，如 scamper[6]在全球不同的位置部署 20 多台探测引擎，能够发现这些探测引擎之间的 IPv6 路由器级拓扑结构。但这种方式部署困难、代价高昂、设备之间的通讯维护任务繁重，其应用难以推广，并且拓扑发现的网络范围有限。

为了调和 IPv6 接入网络拓扑发现覆盖率和探测效率之间的矛盾，在已知目标点信息不充分的前提下提高网络拓扑发现的覆盖率，我们建立了 PATM 模型作为 IPv6 接入网络拓扑发现实现的理论依据，并且提出了基于 PATM 模型的根节点发现策略、叶节点发现策略和兄弟节点发现策略，并对每一种发现策略进行了说明。最后通过对实际 IPv6 接入网络的拓扑发现测试及分析，验证了 PATM 的合理性及相关拓扑发现方法的有效性。

2　IPv6 接入网络建模

只有通过对实际测量得到的网络拓扑结构进行分析，得到相应的特征参数，指导更符合实际的拓扑生成器，才有可能进一步对网络结构进行准确建模[7-9]。但是在进行网络拓扑发现之前，需要一个参考性的网络模型来指导网络拓扑发现方法的形成。

Internet 建模的研究按时间顺序大体上分为三代：第一代为 20 世纪 80 年代的随机图产生器，其中最为典型的是 Waxman 产生器[10]；第二代为 20 世纪 90 年代的结构产生器，Tiers[11]和 Transit-stub[12]就是明显的基于层次结构设计思想的两类拓扑产生器；第三代为 2000 年以来的基于网络节点度的产生器，如 BRITLE[13]和 Inet[14]等。IPv6 全球单播地址采用分级结构，如图 1 所示。在 IPv6 中同样也是使用 IPv4

*)本研究得到国家 863 高技术研究发展计划资助(基金编号：2006AA01Z409)。刘振山　博士研究生，主要研究方向为 IPv6 网络拓扑发现与网络特征分析、网络复杂度理论、大规模网络探测的无踪迹性；王清贤　教授，博士生导师，主要研究方向为网络安全、计算复杂性理论；罗军勇　教授，硕士生导师，主要研究方向为网络安全协议分析、分布式计算。

CIDR 中的最长前缀匹配法则。由于 IPv6 采用可聚合地址和分级结构，IPv6 网络更凸显出层次化的结构特征。目前尚没有专门描述 Internet 路由器级拓扑结构的层次化模型，因此我们借鉴描述 AS 级拓扑结构的 Transit-stub[9] 模型来描述现有的 IPv6 网络路由器级拓扑结构，如图 2 所示。

```
|          n bits          |  m bits  |        128-n-m bits        |
+--------------------------+----------+----------------------------+
global routing prefix      | subnet ID|        interface           |
```

图 1　RFC3513 定义的 IPv6 可聚合全球单播地址结构

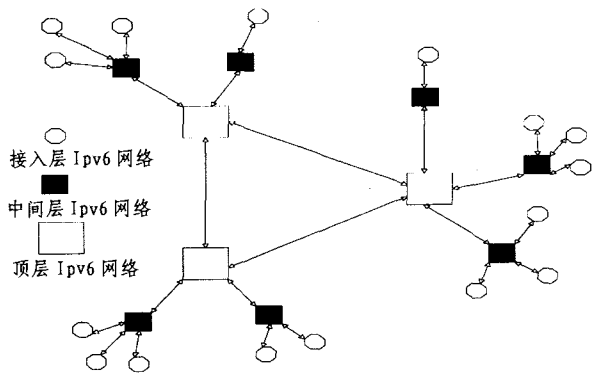

图 2　层次化的 IPv6 网络模型

现有的网络特征刻画参数主要有：平均路径长度(average path length)、聚类系数(clustering coefficient)、度与度分布(degree and degree distribution)等。顶层 IPv6 网络平均聚类系数较高，符合富人俱乐部特征[15]，采用源路由机制下的 Traceroute 机制进行探测，能显著提高拓扑覆盖率；接入层 IPv6 网络聚类系数较低，用源路由机制进行探测显然会产生过多的探测冗余，但在已知目标点有限的前提下，如果仅采用简单的非源路由探测机制进行拓扑发现，又会显著地引起拓扑覆盖率的缺失。如图 3 所示的 IPv6 接入网络中，在已知两个 IPv6 桩网络的前提下采用简单的非源路由探测机制进行拓扑发现，其拓扑发现节点覆盖率不足 50%，链路覆盖率甚至更低。

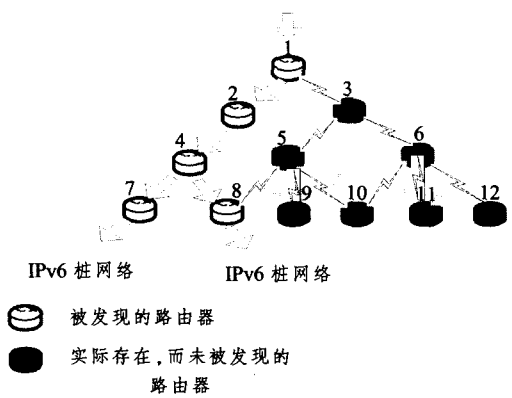

图 3　已知桩网络数量不充分的前提下
采用简单 Traceroute 的探测效果

我们在层次化的 IPv6 模型基础上引入新的特征参数—前缀聚合的单向性特征来刻画 IPv6 接入网络。根据前缀聚合的单向性特征，可以将 IPv6 接入网络描述成一棵以前缀长度为等级的树型结构。图 4 是实际的网络拓扑结构，图 5 是按照前缀聚合单向性特征刻画得到的前缀聚合树结构。

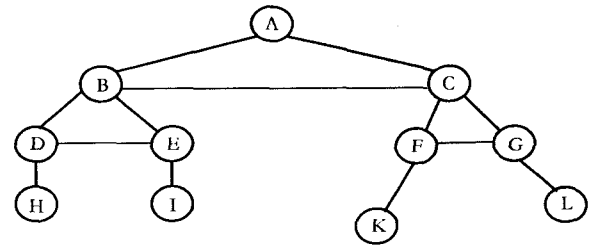

图 4　实际网络拓扑结构

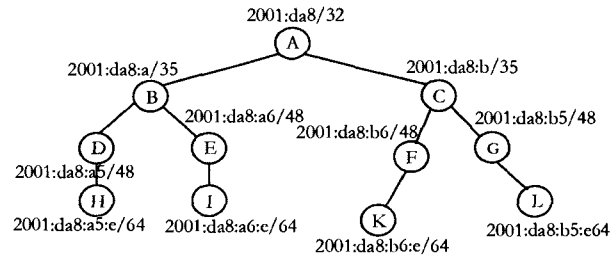

图 5　前缀聚合树结构

下面给出关于前缀聚合树模型中各元素的定义。

根节点：假设某 IPv6 接入网络的区间范围表示为前缀 P，该网络内某路由器向外公告的路由前缀也为 P，则该路由器为前缀聚合树的根节点，如图 5 中的 A 节点。

叶节点：前缀聚合树根节点以外的其它节点称为叶节点。

父子节点：如果 B 节点的前缀被 A 节点聚合，则称 A 节点为 B 节点的父节点，B 节点为 A 节点的子节点，如图 5 中的 A 节点成为 B 节点的父节点。

树枝：从根节点到桩网路之间的路径构成了一棵树枝，如图 5 中的 A-B-D-H 为一棵树枝。

兄弟节点：拥有同一个父节点的节点之间互称兄弟节点，如图 5 中的 D 节点和 E 节点互称兄弟节点。

前缀跨度：父节点聚合后的路由前缀长度与子节点聚合后的路由前缀长度之间的差值称为前缀跨度，如图 5 中的 A 节点和 B 节点之间的前缀跨度为 3。

桩网络：前缀聚合树的末梢，一般前缀为 64 位，如图 5 中的 H 和 I 点。

3　基于 PATM 模型的拓扑发现方法

假定基于 PATM 模型的网络拓扑发现有如下两个前提：

(1)已知目标网络地址区间前缀，这个前提是肯定具备的；

(2)已知目标网络内一定数量的桩网络前缀。

按照前提条件的利用方法，前缀聚合树的拓扑发现主要包括三方面内容：

(1)利用区间前缀进行根节点发现；

(2)利用已知的桩网络前缀发现桩网络所属的树枝上的叶节点；

(3)根据已发现的叶节点和前缀跨度来发现叶节点的兄弟节点。

3.1　根节点发现

根节点发现策略：假设目标网络区间范围表示为 $P_1.P_2.P_3\cdots P_m/m$，m 表示前缀长度，P_i 表示比特位，在 $P_1.P_2.P_3\cdots P_m/m$ 的 m 位后补充 128-m 比特的 0，得到 $P_1.P_2.P_3\cdots.P_m.\underbrace{0\cdots\cdots0}_{128-m}$，将 $P_1.P_2.P_3\cdots P_m.\underbrace{0\cdots\cdots0}_{128-m}$ 作为 Traceroute 探测的目标地址，则该 Traceroute 执行中返回主机不可达信息(Destination

host unreachable)的节点是前缀聚合树的根节点。

策略说明:由于探测报文的目标地址为 $P_1.P_2.P_...P_m.\underbrace{0\cdots\cdots0}_{128-m}$,探测报文被转发节点按照最大前缀匹配选路原则,或者默认路由选路原则,转发到与 $P_1.P_2.P_3\cdots P_m/m$ 完全匹配的一段链路(相当于 IPv4 中的子网)中,而此链路的一端是根节点,由于 $P_1.P_2.P_3...P_m.\underbrace{0.....0}_{128-m}$ 为无效的 IPv6 地址,因此根节点返回主机不可达信息。

3.2 叶节点发现

叶节点发现策略:假设已知的某个桩网络前缀为 $P_1.P_2.P_3...P_m.f_{m+1}.f_{m+2}\cdots f_{64}$,从第 64 位到 $m+1$ 位按照逐 1 比特递减的方法,构造出一个 IPv6 前缀集合:$F=\{P_1.P_2.P_3\cdots P_m.f_{m+1}.f_{m+2}\cdots f_n/n \mid m<n<64\}$,然后将 F 集合中的每个元素按照末位置零的原则构造出伪地址集合 $F'=\{P_1.P_2.P_3...P_m.f_{m+1}.f_{m+2}\cdots f_n.\underbrace{0\cdots\cdots0}_{128-m} \mid m<n<64\}$,将 F' 中的每个元素作为 Traceroute 探测的目标地址,进行依次探测。如果 Traceroute 执行的最终结果返回了网络不可达(Destination net unreachable)信息,则说明该网络前缀不存在;如果某个节点返回主机不可达 (Destination host unreachable)信息,则该节点是叶节点。

策略说明:桩网络处于前缀聚合树的末梢,并且一个桩网络只存在于唯一的树枝上,利用一个已知的桩网络前缀采用这种前缀长度递减猜测的方式就可以发现该桩网络所属树枝上的所有叶节点,而且猜测的最大次数为 $64-m$ 次。

3.3 兄弟节点发现

第 3.2 节利用桩网络前缀发现一部分树枝,及树枝上每一级的前缀跨度。如图 6 所示,假设同一树枝上的两个叶节点 R 和 R′之间的前缀跨度为 N,则 R′最多可能有 2^N 个兄弟节点。

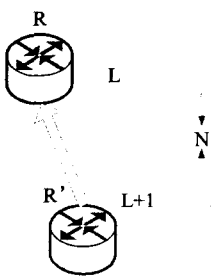

图 6 前缀跨度描述

兄弟节点发现策略:假设叶节点 R 的父节点 R'聚合后的前缀为 $P_1.P_2.P_3\cdots P_n$,R 聚合后的前缀为 $P_1.P_2.P_3\cdots P_n.P_{n+1}.P_{n+2}\cdots P_m/m$,则可以构造出从 $P_1.P_2.P_...P_m$ 到 $P_1.P_2.P_3\cdots P_n.P_{n+1}.P_{n+2}...P_m/m$ 之间 2^{m-n} 个可能的前缀,并按照和上面一样的构造方法产生 2^{m-n} 个伪地址,将这些地址分别作为 Traceroute 探测的目标地址,则返回主机不可达信息的节点是 R 的兄弟节点。

策略分析:兄弟节点发现策略的效率和前缀跨度的大小密切相关。如果前缀跨度过大,就不适合用这种方法来猜测兄弟节点。为了避免海量扫描,做到无踪迹探测,我们在系统实现的时候设定的前缀跨度域值为 8。

4 验证与分析

目前网络拓扑发现验证有两种方法:一种是对 NS-2 模拟器的功能进行拓展,使得支持拓扑发现算法,然后验证拓扑发现算法对 NS-2 模拟器随机产生的网络拓扑结构测试的覆盖率和效率;另一种方法是通过实际网络拓扑发现测试来验证拓扑发现的效能。这两种方法各有优势,但考虑到 NS-2 模拟器无法模拟实际的路由策略、网络延时、已知桩网络的分布等直接影响拓扑发现效果的关键因素,因此我们通过对实际网络的测试来初步评估拓扑发现方法的有效性。

我们的试验环境位于 Cernet-2 郑州大学网络中心,如图 7 所示。

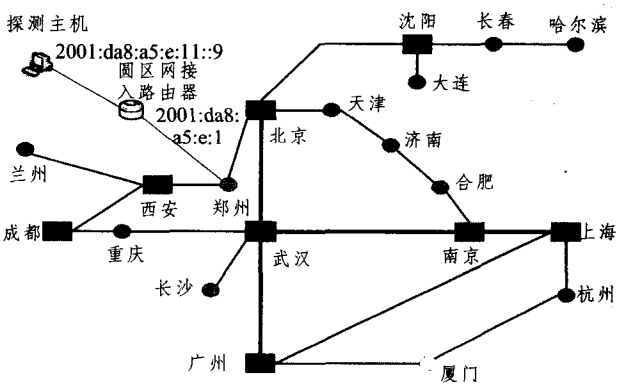

图 7 测试环境

测试目标网络:华东某高校 48 位前缀长度的 IPv6 校园网。

已知的前提条件:目标网络的范围为 2001:250:XXXX/48,测试目标网络内公开的 Web 域名有 16 个,经过解析获得 11 个有效桩网络前缀。

测试过程:因为目标网络实际的拓扑结构未公开,因此我们通过对比的方法来分析本文提出的方法对提高网络拓扑发现覆盖率的帮助。

表 1 和表 2 得到的数据分别是通过简单的 Traceroute 和本文提出的基于前缀聚合树模型的拓扑发现方法得到的数据。为了便于比较,表 1 和表 2 只给出未对路由器接口进行别名归并前的数据信息。

表 1

测试最大线程数	5
超时应答间隔	1500ms
测试总时间	约 2 分 20 秒
测试报文总数(除去别名归并需要的报文)	约 915
探测出路由器接口数(别名归并前)	39
探测出链路数	65

表 2

测试最大线程数	5
测试总时间	约 5 分 50 秒
超时应答间隔	1500ms
测试报文总数(除去别名归并需要的报文)	约 2740
探测出路由器接口数(别名归并前)	57,提高了 46.1%
探测出链路数	92,提高了 41.5%

通过测试结果可以看出,路由器接口覆盖率和链路覆盖率都有了显著提高,基于 PATM 模型的拓扑发现过程基本遵循了低负载、高效和弱感知性的原则。通过对测试现象分析和目标网络的实际求证发现了一些现象与本文提出的前缀聚合树模型有少许出入:

(1) IPv6 地址块的分配不连续现象

在测试过程发现目标网络除了 2001:250:XXXX/48 之外还有其它前缀,比如 2001:da8:XXXX/48。由于 2001:da8:XXXX/48 和 2001:250:XXXX/48 无法聚合,在这种情况下用前缀聚合树模型来描述目标网络得到的不是一棵完整的树型结构,而会出现孤立节点。

(2)路由前缀不聚合现象

虽然普遍情况下路由前缀按照从桩网络向接入点路由器的方向聚合,但有些情况下网管人员对某些路由器并没有设置前缀聚合功能,这样可能出现路由前缀短的路由器节点更接近桩网络,而路由前缀长的节点反而更接近根节点的情况。这种情况和前缀聚合树模型不吻合,但这种情况不具备普遍性。

结束语 非协作方式下 IPv6 接入网络拓扑发现的最大问题在于 IPv6 网络海量地址空间的特性使得探测效率和覆盖率的之间的矛盾不可调和。本文就此问题将 IPv6 接入网络按照其前缀聚合的单向性特征刻画成 PATM 模型,而后提出了基于 PATM 模型的网络拓扑发现方法。通过实际测试及分析,可以得出结论:本文提出的方法能够满足非协作方式下的 IPv6 网络拓扑发现总的要求,且对提高拓扑发现覆盖率有显著的帮助。

参 考 文 献

[1] Internet Mapping Project. Available URL: http://research.lumeta.com/ches/db/

[2] Astic I, Festor O. A hierarchical topology discovery service for IPv6 networks [C] //IEEE/IFIP Network Operations and Management Symposium (NOMS), Apr. 2002: 497-510

[3] 李云琪,杨家海. 一个面向 IPv6 的网络拓扑管理系统的实现[J]. 计算机工程与应用,2004, 40(29): 73-75,181

[4] 沈曾伟. 周刚. IPv6 接入网拓扑结构自动发现方法研究 [J]. 计算机工程,2006,19:136-138

[5] Waddington D G, Chang Fangzhe, Ramesh V, et al. Topology Discovery for Public IPv6 Networks [J]. ACM SIGCOMM Computer Communications Review, 2003, 33(3): 59-68

[6] IPv6 Scamper. http://www.wand.net.nz/~mjl12/ipv6-scamper/

[7] Floyd S, Kohler E. Internet Research Needs Better Models. - ACM SIGCOMM Computer Communication Review (CCR), 2003, 33 (1): 29-34

[8] 张宇,张宏莉,方滨兴. Internet 拓扑建模综述. 软件学报,2004, 15(8): 1220-1226

[9] Alderson D, Doyle J, Govindan R, et al. Toward an Optimization-Dreven Framework for Designing and Generating Realistic Internet Topologies. ACM SIGCOMM CCR, 2003,33(1): 41-46

[10] Waxman B M. Routing of Multipoint Connections. IEEE Journal on Selected Areas in Communications (JSAC), 1988, 6 (9): 1617-1622

[11] Doar M B. A Better Model for Generarting Test Networks// Proc. 1996 IEEE Global Telecommunications Conf. (GLOBECOM 1996). London, Nov. 1996: 86-93

[12] Calvert K L, Doar M B, Zegura E W. Modeling Internet topology. IEEE Communications, 1997, 35 (6): 160-163

[13] Media A, Lakhina A, Matta I. An approach to universal topology generation//Proceedings of MASCOTS. Washington, 2001: 346-353

[14] Winick J, Jamin S. Inet-3.0: Internet topology generator. Technical Report CSE-TR-456-02. Department of EECS, University of Michigan, 2002

[15] Zhou S, Mondragon R J. The rich-club phenomenon in the Internet topology. IEEE Communication Letters, 2004, 8(3): 180-182

(上接第 65 页)

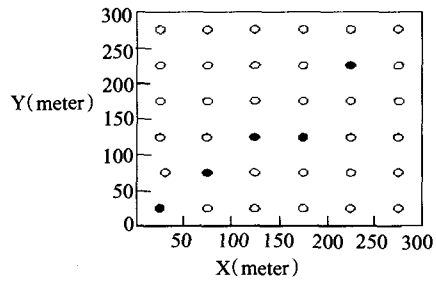

图 3 预设传感器网络节点分布图

参照 2.2 节的失效节点识别算法,设置 $e=0.1$,即可根据公式(1)对失效节点进行识别,识别结果如图 4。

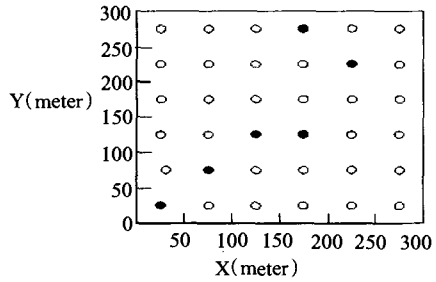

图 4 $e=0.01$ 失效节点识别示意

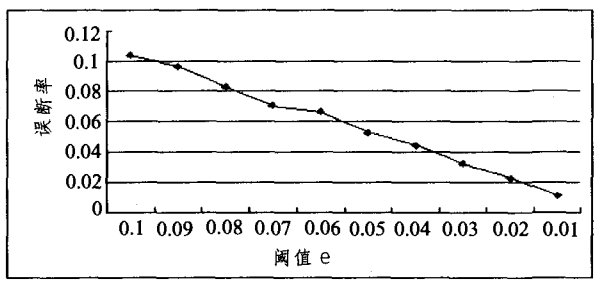

图 5 误判率与阈值 e 关系图

从图中可以看出,随机设置的失效传感器节点可以被准确地识别出来,但是有一个距离传感器汇聚节点距离较远的节点被误判为失效节点。由于该节点距离传感器汇聚节点较远,数据发送到汇聚节点可能需要经过更多的跳数,在设置的阈值 e 的范围内,产生了报文的连续丢失,因此需要提高阈值的精度去做出更准确的判断,将阈值 e 设置为 0.01,就可以对失效的传感器节点做出准确的判断而不会产生误判。

下面对误判率和阈值 e 的关系进行分析,进行 10000 次数据收集,并对距离汇聚节点跳数为 10 的节点,研究其随着阈值 e 的变化,节点失效的误判率变化情况,从图 4 中可以看出当 e 取值在 0.01 时,误判率可以控制在 1% 左右,对其他节点进行同样的抽样,也得出类似的结果,由此可以看出,节点失效误判的情况可以由阈值 e 很好地控制。

结束语 本文利用 Bloom filter 技术,提出了一种传感器网络失效节点的识别方法,并通过 NS2 仿真模型进行了验证。但是本文提出的算法是针对周期性数据采集的传感器网络,具有一定的局限性,如何在事件驱动和非周期性传感数据收集的传感器网络利用传感器节点采集数据的时空相关性进行失效节点的判断是未来研究的目标。

参 考 文 献

[1] Zhao J, Ramesh G, Deborah E. Sensor network tomography: Monitoring wireless sensor networks. Computer Communication Review, 2002, 32(1): 64

[2] Zhao J. Measurement and Monitoring In Wireless Sensor Networks. Ph. D. Thesis. Department of Computer Science, University of Southern California, Dec. 2003

[3] Zhao J, Govindan R, Estrin D. Residual energy scan for monitoring sensor networks// IEEE Wireless Communications and Networking Conference. vol. 1, IEEE USA, 2002: 356-362

[4] Hartl G, Li Baochun. Loss inference in wireless sensor networks based on data aggregation//Third International Symposium on Information Processing in Sensor Networks. ACM USA, 2004:396-404

[5] Li Yongjun, Cai Wandong, Tian Guangli, et al. Loss Tomog-raphy in Wireless Sensor Network Using Gibbs Sampling// Proceeding of EWSN 2007. LNCS 4373, 2007:150-162

[6] 肖明忠,代亚非. Bloom Filter 及其应用综述. 计算机科学, 2004,31(4): 180-183

[7] He T, Huang C, Blum B M, et al. Range-free localization schemes for large scale sensor networks// Proceedings of the 9th Annual International Conference on Mobile Computing and Networking. 2003: 81-95

[8] The Network Simulator 2, www.isi.edu/nsnam/ns2, 2005