

无线传感器网络启发式失效链路推断算法

赵 佐,蔡皖东

ZHAO Zuo, CAI Wan-dong

西北工业大学 计算机学院,西安 710072

School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China

E-mail: zhaozuo@nwpu.edu.cn

ZHAO Zuo, CAI Wan-dong. Heuristic algorithm to identifying lossy links in sensor networks. *Computer Engineering and Applications*, 2009, 45(14): 93-95.

Abstract: The technology of wireless sensor network has matured and has been built actual applications such as environment monitoring, surveillance etc. The experiences have demonstrated the obvious need for network failure managing tools. Lossy links use in a sensor network affect network performance, and hence need to be detected and repaired. Sensor nodes are restrained by limited resources, so inference technique based on network tomography is introduced by passive end-to-end measurement. Through the problem of lossy links inference is mapped to minimal set-cover problem, the algorithm based on heuristic strategy is proposed. The performance of inference algorithm is evaluated by simulation, and the simulation results indicate feasibility and efficiency of the method.

Key words: lossy link inference; network tomography; set-cover problem; heuristic strategy

摘 要: 无线传感器网络的实际应用产生了对网络故障管理的迫切需求。严重报文丢失的失效链路影响无线传感器网络的性能,需加以发现并修复。受有限资源的约束,采用被动端到端测量的方法,利用网络断层扫描技术推断失效链路。通过将失效链路推断问题映射为最小集合覆盖问题,提出了一种基于启发策略的失效链路推断算法。仿真实验结果表明该算法具有可行性和有效性。**关键词:** 失效链路推断;网络断层扫描;最小集合覆盖问题;启发式策略

DOI: 10.3778/j.issn.1002-8331.2009.14.028 文章编号: 1002-8331(2009)14-0093-03 文献标识码: A 中图分类号: TP393

1 引言

失效链路(lossy link)是指出现严重报文丢失的链路。随着无线传感器网络技术的日益成熟,无线传感器网络的实际应用系统逐步出现在环境监控、数据收集等多个领域。无线传感器网络中网络拥塞的发生、节点能量过低或无线通信干扰都将导致链路产生严重报文丢失,失效链路严重影响无线传感器网络的性能,必须加以发现和修复。

失效链路发现技术分为基于内部节点协同测量和基于端到端测量两大类。基于内部节点协同测量技术需要每个节点监控其邻接链路的报文丢失率,并报告给汇聚节点。这个方法虽直接明了,但导致了巨大的通信流量负担,不适合在资源有限的无线传感器网络中使用。每一个部署的传感器网络都面向某个或一些确定的应用,其节点有规律地发送应用数据到汇聚(sink)节点。利用端到端测量数据推断链路性能的技术称为网络断层扫描(Network Tomography, NT)技术^[1]。而利用端到端测量的方法具有不产生额外的监控流量的优点,且被动测量方式不需要额外消耗节点能量。因此,采用被动测量感知节点到汇聚节点的应用业务数据报文丢失,利用 NT 技术

推断失效链路。

现有基于 NT 技术推断失效链路的方法分为两类:测量数据包相关性法^[2-3]和简单断层扫描法^[4-6]。探测数据报文相关性法需要保证探测数据包间存在严格的相关性,文献[2]提出了基于数据融合的数据收集框架来保证数据报文之间的相关性。该方法可以推断出每条链路的具体丢包率,精确性高,但部署困难,限制了该方法的应用范围。简单断层扫描法主要用于推断具有布尔性质的网络性能(如连通性、失效链路等)。其假设网络中报文丢失的发生是少数链路失效所导致的^[6],不需要数据数据包之间具有相关性,测量方法简单。文献[4]提出了线性规划(LP)算法和基于 Gibbs 抽样的贝叶斯推断算法。文献[5-6]提出基于贪心策略的 SCFS 推断算法。Gibbs 抽样算法的精确度最高,但计算开销大,难以应用于大规模网络推断计算。SCFS 算法简单,计算速度快,但倾向于优先选取距离根节点近的链路作为失效链路,存在较大误差,且只能用于树型拓扑。本文在上述研究基础上,通过将失效链路推断问题映射为最小集合覆盖问题,提出了一种启发式失效链路推断算法。

作者简介: 赵佐(1974-),男,在职博士研究生,CCF 会员,主要研究方向为网络测量与性能评价、网络信息安全;蔡皖东(1955-),男,博士,教授,博士生导师,主要研究方向为计算机网络、分布式计算、网络信息安全与对抗等。

收稿日期:2008-03-17 修回日期:2008-06-10

2 网络模型

2.1 拓扑模型

无线传感器网络数据收集过程中形成的网络逻辑拓扑可建模为一棵逆向树 $T=(V, L)$, 如图 1 所示。其中 V 为节点集合, L 表示连接节点之间的链路集合。 T 的根节点 s 表示 sink 节点。 $n_v=|V|$ 表示节点数, $n_l=|L|$ 表示逻辑链路数。节点 i 到节点 j 的链路用有序节点对 $(i, j) \in V \times V$ 表示, 即节点 j 是节点 i 的下一跳节点, 在拓扑 T 中 i 是 j 的子节点。链路 $(i, j) \in L$ 简记为 l_{ij} 。除根节点 s 外, 任一节点 i 都有唯一的父节点 $f(i)$ 。即 $(i, j) \in L$, 则记 $j=f(i)$ 。设存在正整数 n 使得 $k=f^n(i)$ 成立, 则称节点 k 为 i 的祖先节点, 节点 i 为 k 的子孙节点。用集合 $d(i)=\{k \in V | \exists n > 0, i=f^n(k)\}$ 表示节点 i 的子孙节点集。在 $T=(V, L)$ 中, 从节点 i 到节点 s 的路径记为 p_i , 集合 P 表示所有到 sink 节点的路径, $n_p=|P|$ 为路径数量。 M_i 表示组成路径 p_i 的所有链路集合。

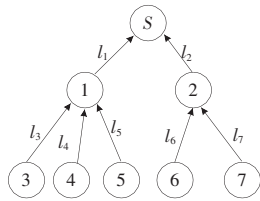


图 1 逆向树网络拓扑示意图

给定 $T=(V, L)$ 和 P , 可计算路由矩阵 $A=(a_{ij})_{n_p \times n_l}$: A 中的第 i 行对应路径 p_i , 第 j 列对应链路 l_j , 其中 $a_{ij}=1$ 表示路径 p_i 中包含链路 l_j , 即 $l_j \in M_i$ 。本文假设任一条链路至少被一条路径覆盖。

如果无线传感器网络内部节点 i 到汇聚节点 s 的路由在数据采集周期内没有改变, 就认为从感知节点 i 到汇聚节点 s 的拓扑关系是稳定的。在失效链路推断时, 假设网络拓扑是已知的。当拓扑未知时, 可以利用端到端测量数据推断出网络拓扑^[7]。

2.2 性能模型

本文使用的性能模型如下所述。假设传感器网络中的逻辑拓扑在数据收集周期 T_R 内保持相对稳定, 使得能够收集到足够多的数据。网络中的每个感知节点都向 sink 节点发送或转发感知数据, 在 sink 节点就可知那些感知节点的数据到达 sink 节点。根据感知节点的采样频率, 可知在 T_R 应该发送数据报文及在传输路径上丢失数据报文的个数。假设报文在各个链路之间发生丢失相互独立, 即服从 Bernoulli 分布, 则流经 T 的数据流可用随机过程 $Z=(z_{ij}), i \in d(j), j \in V$ 来描述, 其中 $z_{ij} \in \{0, 1\}$ 。 $z_{ij}=1$ 表示从节点 i 发出的数据成功到达 j , 反之数据在链路上发生丢失。设 ϕ_k 表示报文在链路 l_k 上的平均到达率, 则报文丢失率为 $1-\phi_k$ 。若路径 p_i 由 m 条链路组成, 即 $M_i=\{l_1, \dots, l_m\}$, 设路径报文平均到达率用 φ_i 表示, 则有 $\varphi_i=\prod_{k=1}^m \phi_k$ 。

文献[6]阐述了 ϕ_k 不具有统计可辨识性的原因, 即单纯用路径的报文丢失率无法计算出每条链路的报文丢失率。因此, 采用链路报文到达率阈值 t_l 对链路进行分类: 当 $\phi_k < t_l$ 时链路为失效 (lossy) 或坏链路; 否则为好 (not lossy) 链路。 t_l 的值可以依据具体应用需求或历史数据确定。文献[8]的研究表明无线传感器网络中链路或为好或为坏链路, 且好与坏链路是明确可区分的。可区分性定义为好路径全部由好链路组成, 而坏路径中至少包含了一条失效链路^[6]。在失效链路相对稀少的假设下, 利用端到端测量数据来推断失效链路的最可能解应当是具有最少

失效链路数目的解^[6]。

3 推断算法

用 $D=\{D_1, \dots, D_{n_p}\}$ 表示端到端的测量数据集, 其中 $D_i=(r_i, f_i)$, r_i 为路径 p_i 在测量期间到达的报文数, f_i 为丢失报文数。路径 p_i 报文到达率 $\varphi_i=r_i/(r_i+f_i)$ 。设定路径报文到达率阈值 t_p , 用其来划分好坏路径: 当 $\varphi_i \geq t_p$ 时, 路径 p_i 为好路径; 反之则为坏路径。用 t_p 将路径集 P 划分为好路径集 P_C 和坏路径集 P_B 。

因为 t_p 是人为指定, 所以不可避免引入两类判断误差: 假阳性错误, 好的路径判断为坏路径; 假阴性错误, 实际是坏路径但判断为好路径。当选择 $t_p=t_l$ 时, 若 p_i 中存在一条坏链路 $l_k (\forall l_k \in$

$M_i, \phi_k < t_l)$, 则路径传输到达率低于 $t_p (\varphi_i = \prod_{l_k \in M_i} \phi_k \leq t_p)$, 此时不存在假阳性错误; 当选择 $t_p=t_l^m$ ($m=|M_i|$ 为 p_i 中包含的链路数),

因为如果路径的传输成功率低于 t_l^m , 说明至少存在一条传输成功率小于 t_l 链路, 因而不存在假阴性错误。但在不知链路丢包率具体分布情况下, t_p 的最优选择是无法通过分析的方法得出。本文选取 $t_p=(t_l+t_l^m)/2$ 。

3.1 问题描述

定义 1 链路控制域是指包含指定链路的集合。即:

$$\text{Domain}(l_k)=\{p_i | p_i \in P \text{ and } l_k \in M_i\}$$

设 $\chi \subseteq L$ 为最可能的失效链路解, x 为长度为 n_l 标志向量, 即 $l_k \in \chi$ 时, $x_k=1$, 否则 $x_k=0$ 。失效链路推断问题可描述为:

$$\begin{cases} \min \min z & \mathbf{1}_n \mathbf{x}^T \\ \text{s.t.} & P_B = \bigcup_{x=1}^T \text{Domain}(l_i) \end{cases} \quad (1)$$

式(1)中 $\mathbf{1}_n=[1, 1, \dots, 1]$ 为 n_l 维行向量。由此, 失效链路推断问题映射为最小集合覆盖问题, 可以通过求解式(1)推断出失效链路。

3.2 算法描述

最小集合覆盖问题是典型 NP 困难问题^[9], 文献[6]的 SCFS 算法可看作集合覆盖问题的贪心法求解, 本文采用启发式算法求解。

定义 2 路径频度 $k(i)$ 为路径 p_i 在 P_B 中出现的所有链路构成的链路控制域集合中出现的次数。设 $p_i \in P_B$, 当且仅当 p_i 出现在 k 个链路控制域时其路径频度为 $k(i)$ 。

定义 3 链路覆盖度 $C(i)$ 为链路控制域 $\text{Domain}(l_i)$ 中包含的所有路径频度的最小值。即 $C(i)=\min\{k(j) | p_j \in \text{Domain}(l_i)\}$ 。

定义 4 必选链路是链路覆盖度为 1 的链路。

定义 5 链路必选度是指在必选链路的集合中, 如果不选择链路 l_i , 则出现 R 个必选链路, 称链路 l_i 的必选度为 R , 记为 $R(i)$ 。

利用上述定义, 给出如下启发式策略:

策略 1 若存在链路 l_i , 使得 $\text{Domain}(l_i)=P_B$, 则 l_i 被选择作为唯一的失效链路。

策略 2 必选链路一定被选择作为失效链路。

策略 3 如果 $\text{Domain}(l_i) \subseteq \text{Domain}(l_j)$, 则排除链路 l_i 。

策略 4 如果 $R(i) > R(j)$, 链路 l_i 被选择的优先级高于链路 l_j 。

因为不选 l_i 后, 在下一步必须选择的链路数比排除 l_j 的多, 必选链路越多越不利于优化, 所以选择 l_i 比较合理。这是防止“过分贪心”的策略。

策略 5 如果两条链路的必选度相同,则选择控制域基数大的链路。这是朴素贪心思想的推广。

基于启发式策略失效链路推断算法(Heuristic Lossy Link Inference, HLLI)具体描述如下。

Input:网络拓扑 $T(V, L)$, 测量数据集合 D , 链路报文到达率阈值 t_{i0}

(1)初始化:设 χ 为 lossy 链路集合; $\chi=\emptyset, P_c=\emptyset, P_b=\emptyset$, 标识向量初始全部为零,即 $x=[0, 0, \dots, 0]_{n_v}$ 。

(2)对网络中每一条路径,计算 t_p 和路径报文到达率, $\forall p_i \in P, \varphi_i = r_i / (r_i + f_i)$ 。如 $\varphi_i > t_p, P_c \leftarrow P_c \cup \{p_i\}$; 否则 $P_b \leftarrow P_b \cup \{p_i\}$ 。

(3)对在 P_b 中的每一条路径中的组成链路,置对应标志位为 1, $x_i = 1$, if $l_i \in M_j, p_j \in P_b$ 。

(4)计算路径频度、链路覆盖度和必选度。

(5)while $P_b \neq \emptyset$

①如存在链路 l_i , 使得 $\text{Domain}(l_i) = P_b$, 则链路 l_i 加入 $\chi, \chi = \chi \cup \{l_i\}$; break;

②如存在在必选链路, 加入必选链路到 χ 中: if $\exists C(j) = 1$ then $\chi = \chi \cup \{l_j\}, P_b \leftarrow P_b - \text{Domain}\{l_j\}$, continue; 否则, 转③;

③ $\exists l_i, l_j$, 满足 $\text{Domain}(l_i) \subseteq \text{Domain}(l_j)$, 则 $P_b \leftarrow P_b - \text{Domain}\{l_i\}$, continue; 否则, 转④;

④选择必选度最大的链路加入 χ : if $\forall x_j = 1, \exists l_i, \max R(i)$, then $\chi = \chi \cup \{l_i\}, P_b \leftarrow P_b - \text{Domain}\{l_i\}$, continue; 否则, 转⑤;

⑤选择链路控制域基数最大的链路加入 χ 。

(6)output χ 。

4 仿真与评价

算法的性能评价采用两个指标:覆盖率(Detection Rate, DR)和误判率(False Positive Rate, FPR)。设 F 表示网络中实际的失效链路集, X 表示用算法推断出的失效链路集, 则 DR 和 FDR 定义如下:

$$DR = \frac{|F \cap X|}{|X|}; FPR = \frac{|X \setminus F|}{|X|}$$

仿真实验采用 NS2 网络仿真器,通过扩展 NS2 使其能够仿真无线传感器网络中数据收集算法。在每个轮次的数据收集过程中,一个节点是否成功收到其子节点发送来的感知数据是随机决定的。仿真时每条链路都设定一个报文丢失率,当数据收集轮次逐渐增大时,该链路上实际报文丢失率趋近于预设定的报文丢失率。在仿真过程中,每条链路实际的报文丢失数也被统计,用于计算实际的链路报文丢失率,通过与推断结果比较,来评价算法的准确性和有效性。

仿真过程采用 GT-ITM 拓扑生成器的 Transit-Stub 图形模块生成树形网络,节点数 v 在范围 100~1 000 间变化。设 f 表示网络中失效链路所占比例, f 在范围 0.05~0.25 之间变化。丢包模式 LM 定义为:好链路的丢包率服从区间 $[0, 0.01]$ 上均匀分布,失效链路的丢包率服从区间 $[0.05, 1]$ 上均匀分布。一旦每条链路被指定一个丢包率,该链路的实际丢包过程采用 Bernoulli 过程。每个数据报文在传输过程中被丢失的概率由该链路的报文丢失率所确定。每次实验进行 200 轮数据收集,通过测量数据推断失效链路,并计算每次实验 DR 和 FPR。每种配置条件下共进行 100 次实验,计算 DR 和 FPR 的平均值,用其评价算法性能的优劣。

(1)坏链路比例的因素。图 2 显示当网络拓扑固定,节点数 $v=500$, 坏链路比例 f 在 $[0.05 \sim 0.25]$ 变化时算法性能的变化趋势。随着 f 的增加,DR 减少而 FPR 增加,算法的性能发生下降。出现这种趋势是因为推断算法是建立在假设失效链路的数量

是相对稀少基础之上,当坏链路的比例增加时,弱化了假设条件导致算法性能下降。

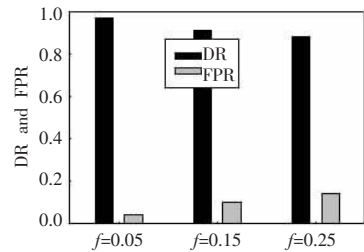


图 2 节点数 $v=500$ 时 f 变化时性能比较

(2)网络拓扑的影响因素。图 3 显示了网络中坏链路比例 $f=0.2$, 改变网络的规模,节点数 v 在 $[100 \sim 1 000]$ 内变化时算法性能的变化趋势。随着网络规模的增大,性能下降,但下降的幅度并不大,说明算法对网络规模的变化有较强的健壮性。

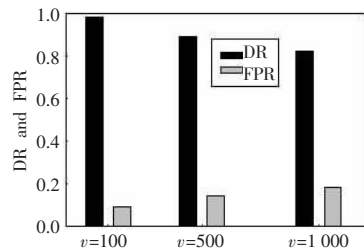


图 3 $f=0.2$ 时网络节点数变化时性能比较

(3)与 SCFS 算法比较

算法 HLLI 与 SCFS^[5]比较结果见表 1。可以看出 HLLI 算法在覆盖度上比 SCFS 算法略高,但在误判率上 HLLI 算法明显优于 SCFS 算法。这是因为 SCFS 算法采取基于链路控制域基数的贪心策略,优先选取距离根节点最近的链路作为失效链路。而 HLLI 算法利用启发式策略能够更加准确选择失效的链路。

表 1 算法 HLLI 与 SCFS 性能比较

性能	$f=0.05$		$f=0.15$		$f=0.25$	
	HLLI	SCFS	HLLI	SCFS	HLLI	SCFS
DR/(%)	98	98	93	91	89	88
FPR/(%)	0.7	2.0	1.4	4.0	1.9	6.9

5 结束语

失效链路发现是无线传感器网络管理的重要组成部分。采用被动测量无线传感器网络中的应用数据报文,通过网络断层扫描技术推断出失效链路的方法。在失效链路相对稀少假设下,将失效链路推断问题映射为最小集合覆盖问题,提出了一种启发式失效链路推断算法。通过仿真实验证明了该算法的有效性。

参考文献:

- [1] Coates M, Hero A, Nowak R, et al. Internet tomography[J]. IEEE Signal Processing Magazine, 2002, 19.
- [2] Hartl G, Li B. Loss inference in wireless sensor networks based on data aggregation[C]//Proceedings of the Third IEEE/ACM International Symposium on Information Processing in Sensor Networks (IPSN 2004), April 26-27 2004.

(下转 144 页)