

基于简单网络断层扫描的失效链路定位研究

赵 佐 蔡晓东

(西北工业大学计算机学院 西安 710129)

摘 要 失效链路是无线传感器网络中一种典型的网络故障现象,严重影响了无线传感器网络的运行与服务质量,必须加以发现并修复。主要研究了基于简单网络断层扫描方法定位失效链路的技术。引入二元分离模型描述链路状态,在已知链路状态先验分布条件下,失效链路定位问题描述为最大后验估计问题。通过将失效链路定位问题映射为加权最小集合覆盖问题,提出了一种基于启发策略的失效链路定位算法。仿真实验结果表明,该算法具有可行性和有效性。

关键词 失效链路定位,简单网络断层扫描,加权最小集合覆盖问题,启发式策略

中图分类号 TP393 **文献标识码** A

Localization Problem of Faulty Links Based on Simple Network Tomography

ZHAO Zuo CAI Wan-dong

(School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China)

Abstract Faulty links are a typical network fault symptom in wireless sensor network. It affected the operation of network and the quality of service seriously, therefore faulty links need to be detected and repaired. Localization technique of faulty links based on simple network tomography was introduced. This paper presented a binary disjunctive model for describing link faulty states; the localization problem of faulty links is referred to as the Maximum A-posteriori Probability problem. By making the Localization problem of faulty links mapped to weighting minimal set-cover problem, the algorithm based on heuristic strategy was proposed. The performance of inference algorithm was evaluated by simulation, and the simulation results indicated the feasibility and efficiency of the method.

Keywords Faulty link location, Simple network tomography, Weighting set-cover problem, Heuristic strategy

1 引言

无线传感器网络(Wireless Sensor Network, WSN)综合了传感器、嵌入式计算、网络及通信、分布式信息处理等多项技术,是涉及多学科高度交叉、知识高度集成的前沿热点研究领域。它利用大量的微型传感计算节点,通过自组织网络以协作方式进行实时监测、感知和采集各类环境或监测对象的信息,成为连接物理世界、数字虚拟世界和人类社会的桥梁。但相对于传统网络,无线环境内在的通信介质不可靠性、节点能量受限及带宽等资源有限,使得无线传感器网络更易受周围环境、传输距离、传输速率等因素的影响,故障发生的频率更高。因此,为提供较高的服务质量,需要设计有效的故障管理机制并提供高效的故障诊断工具。

失效链路是指出现不连通、严重报文丢失或延迟超长等现象的链路,是一类重要的网络故障形式。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失效。因此,失效链路是无线传感器网络一种典型的故障形式,它严重影响无线传感器网络的运行、性能及服务质量。传感器网络的故障诊断包括故障的发现、定位与隔离、

恢复 3 个阶段^[1]。本文重点针对失效链路发现与定位问题展开研究,以丢失链路(lossy link,即发生严重报文丢失的链路)为例,论述基于端到端测量的失效链路发现与定位技术。

2 相关研究工作

现有的传感器网络失效链路定位技术分为基于内部节点协同发现和端到端测量两大类。基于内部节点协同发现技术采取每个节点监控其邻接链路的报文状态,将结果报告给汇聚节点(sink)。该技术概念上直接明了,且有较高发现定位精度,但产生了很大的通信流量负载,导致节点能量损耗过大,降低了传感器网络的寿命。

利用端到端测量数据推断故障发生位置的定位技术称为网络断层扫描(Network Tomography, NT)技术^[4]。传感器网络的节点周期地发送感知数据到汇聚节点,如以应用业务数据报文作为测量工具,在汇聚节点被动监测应用业务数据报文则可获取报文延迟或丢失的测量值。因此,端到端被动测量的方法具有不产生额外的监控流量,且不需要额外消耗节点能量的优点。

依据探测报文间是否具有相关性,NT 技术推断失效链

到稿日期:2009-02-20 返修日期:2009-05-08 本文受教育部博士点基金(200806990030),西北工业大学科技创新基金(2008KJ02028)资助。

赵 佐(1974-),男,博士研究生,讲师,主要研究方向为网络测量与性能评价、网络信息安全,E-mail:zhaozuo@nwpu.edu.cn;蔡晓东(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。

路的方法分为两类:报文相关性法^[2,3]和简单断层扫描法^[5-8]。报文相关性法需要保证探测报文间存在严格的相关性,文献[2]提出了基于数据汇聚保证报文间相关性的数据收集方案,采用EM算法推断每条链路的具体丢包率,以此确定丢失链路的具体位置。该方法定位精度高,但计算量大且数据汇聚涉及的技术复杂、部署困难,从而限制了其应用范围。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、高丢包链路、延迟尖峰等),其假设网络中失效路径的出现是由少数链路失效导致的^[7],测量时不需要探测报文之间具有相关性,测量方法简单、易于部署。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。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。文献[5,6]提出了SCFS算法,该算法简单,计算速度快,采用的贪婪策略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,且只能应用于树型拓扑。文献[8]基于最小集合覆盖技术提出了LLIS算法,该算法增加了对不精确测量数据的处理过程,但其假设每条链路发生失效的概率相同,且链路失效发生是小概率事件,这样的假设违背了网络运行的实际情况。

3 网络和性能模型

3.1 网络拓扑模型

$T=(V,L)$ 表示一棵倒立的逆向树,用来描述传感器网络数据收集过程中形成的网络逻辑拓扑(如图1所示),其中 V 表示传感器网络节点集合, 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, $n_e=|L|$ 表示逻辑链路数。序偶 $(i,j) \in V \times V$ 表示节点 i 到节点 j 的链路,即节点 j 是 i 的下一跳节点,在 T 中节点 i 是 j 的子节点。链路 $(i,j) \in L$ 简记为 l_i 。除根节点 s 外,任一节点 i 都有唯一的父节点 $f(i)$,即 $(i,j) \in L$,则有 $j=f(i)$ 。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。集合 $d(i)=\{k \in V \mid \exists n > 0, i=f^n(k)\}$ 表示节点 i 的子孙节点集。在 $T=(V,L)$ 中,节点 i 到节点 s 的路径记为 $path_i$,集合 $PATH$ 表示所有感知节点到sink节点的路径集合, $n_p=|PATH|$ 为路径数。 M_i 表示组成路径 $path_i$ 的所有链路集合。

由 $T=(V,L)$ 和 $PATH$ 可计算路由矩阵 $A=(a_{ij})_{n_p \times n_e}$ 。 A 中的第 i 行对应路径 $path_i$,第 j 列对应链路 l_j ,其中 $a_{ij}=1$ 表示路径 $path_i$ 中包含链路 l_j ,即 $l_j \in M_i$ 。假设拓扑中任一条链路至少被一条路径覆盖。

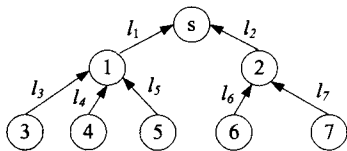


图1 逆向树网络拓扑示意图

3.2 性能模型

假设传感器网络中的逻辑拓扑在数据收集周期TR内保持相对稳定,从而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,则sink节点就可知哪些感知节点的数据已到达。以丢失链路为例,根据感知节点的采样频率,可知在TR内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,则流经 T 的数据流可用随机过程 $Z=(z_{i,j}), i \in d(j), j \in V$ 来描述,其中 $z_{i,j} \in \{0,1\}$ 。 $z_{i,j}=1$ 表示从节点 i 发出的报文成功到达 j ,反之报文在链路上发生丢失。

已知网络拓扑和端到端路径成功传输率(简称传输率),则路径与链路传输率之间存在如下线形关系^[4]:

$$\log(\varphi_i) = \sum_{k=1}^{n_p} \log(\phi_k) D_{ik}, 1 \leq i \leq n_p \quad (1)$$

式中, φ_i 为路径 $path_i$ 的成功传输率(简称传输率), ϕ_k 为链路 l_k 的传输率,链路 l_k 的丢包率为 $1-\phi_k$ 。为了求得链路丢包率,必须求解上述线性关系构成的方程组。但由于矩阵 D 通常是非满秩的,因此不存在唯一解。

我们的目的是定位失效发生链路的位置,因此并不一定需要计算出每一链路的实际丢包率,只需要找出丢包率超过阈值的链路即可。可以采用二元可分离模型建立故障模型,利用阈值对链路与路径状态进行划分。二元可分离模型适用于很多测量指标,如连通性、延迟尖峰等故障建模。设 t_l 是所指定的链路传输率阈值,当 $\phi_k \geq t_l$ 时,则链路 l_k 为好(non lossy),否则链路为坏(或失效,lossy)。设定路径报文传输率阈值 t_p ,用来划分好坏路径:当 $\varphi_i \geq t_p$ 时,则路径 $path_i$ 为好路径;反之为好路径。阈值 t_l 的选择依赖于具体应用的性能需求。 t_p 的选择为 t_p^d , d 为路径的长度。

在二元分离模型下,采用布尔变量 y_i 表示路径 $path_i$ 的状态: $path_i$ 为好路径,则 $y_i=0$;否则 $y_i=1$ 。用 x_k 表示链路状态:链路 l_k 为好时 $x_k=0$ 。二元可分离模型下路径与链路状态关系表示如下:

$$y_i = \bigvee_{k=1}^{n_p} x_k \cdot a_{ik} \quad (2)$$

式中,“ \bigvee ”表示二元逻辑加法运算,“ \cdot ”表示逻辑乘法运算。式(2)的向量表达形式为:

$$y = \bigvee_{k=1}^{n_p} x_k \cdot a_k \quad (3)$$

式中, a_k 表示 A 的第 k 列。

通过求解方程组(3)来获取失效链路具体精确位置。方程组(3)有唯一解当且仅当 D 的列在布尔代数中是线性独立,而这种情况是很少出现的。当方程组(3)存在多个解时,需要选取一个最可能解释测量值的解。因此,必须利用额外的链路状态信息来辅助发现这个最可能的解。对丢失链路而言,这个信息就是链路 l_k 成为丢失链路的概率 θ_k 。设 $\Theta=[\theta_1, \theta_2, \dots, \theta_{n_e}]^T$ 表示链路状态概率的向量,其中 T 表示转置。先验概率 Θ 的值可以基于历史数据获得。

4 基于先验概率的失效链路定位算法

4.1 问题描述

基于上述分析,故障链路定位问题可描述为:已知路由矩阵 A 、链路状态先验概率 Θ 、一次端到端测量值 $y=[y_1, y_2, \dots, y_{n_p}]^T$,通过求解方程组(3)获得 $x=[x_1, x_2, \dots, x_{n_e}]^T$ 的值。由于式(3)可能存在多个解,需要寻找一个最可能的解,即在给定测量值 y 条件下求解 x 使得条件概率 $P_\Theta(X=x|Y=y)$ 最大化。即:

$$\operatorname{argmax}_x P_\Theta(X=x|Y=y) \quad (4)$$

式中, $P_\Theta(\cdot)$ 表示在给定先验概率 Θ 下的条件概率。求解最

可能失效链路问题是一个最大后验估计(MAP)问题,式(3)的解给出了一个最大后验估计。由 Bayes 公式:

$$P(X=x|Y=y)=\frac{P(Y=y|X=x)P(X=x)}{P(Y=y)} \quad (5)$$

式(5)中的 $P(Y=y)$ 仅依赖于网络条件和测量值,并不影响 x 的选择。因此式(4)等价于:

$$\operatorname{argmax}_x P(Y=y|X=x)P(X=x) \quad (6)$$

由于链路状态 X_k 是相互独立的随机变量,即:

$$P(X=x)=\prod_{k=1}^{n_c} \theta_k^{x_k} (1-\theta_k)^{(1-x_k)} \quad (7)$$

对于任意路径 P_i ,在给定 $X=x$ 计算方程(2)时, $Y_i=0$ 当且仅当对所有 $k(1 \leq k \leq n_c)$, $A_{ik}x_k=0$ 或等价于 $(1-A_{ik})^{x_k}=1$ 。

因此,条件概率 $P(Y_i=0|X=x)=\prod_{k=1}^{n_c} (1-A_{ik})^{x_k}$ 和 $P(Y_i=1|X=x)=1-P(Y_i=0|X=x)$ 的值为 1 或 0。

定义 1 P_C 作为测量中好路径的集合,即当 $y_i=0$ 时 $P_i \in P_C$; P_B 作为发现故障的路径集合,当 $y_i=1$ 时 $P_i \in P_B$ 。

$$P(Y=y|X=x)=\prod_{P_i \in P_C} P(Y_i=0|X=x) \prod_{P_i \in P_B} P(Y_i=1|X=x) \quad (8)$$

则式(8)的概率计算结果为 1 或为 0。

因此,要保证式(6)的值不为 0,则有 $P(Y=y|X=x)=1$,故 x 必须满足如下条件:

(1)对每一条故障路径 $P_i \in P_C$,至少包含了一条坏链路(即 $\exists k, x_k=1$,满足 $A_{ik}=1$);

(2)对每一条好路径 $P_i \in P_C$,不包含任何一条坏链路(即 $\forall k, x_k=0$,满足 $A_{ik}=1$)。

简而言之,好路径上的所有链路都是好的,故障路径上必须至少存在一条故障链路。

利用上述结论,将故障链路定位问题进行如下简化:设简化生成矩阵 R 是从 A 中移除所有对应好路径的行和在好路径行上的链路所在的列后的矩阵,则 R 的每一行表示一条坏路径,每一列表示链路属于至少一条故障路径。设 N_R 表示矩阵 R 中列所代表的备选故障链路集合,在排除好路径和好链路后, R 的维数为 $|P_B| \times |N_R|$ 。故障链路定位问题转化为在集合 N_R 中寻找一个子集 x ,使得所有的故障路径至少包含 x 中的一条链路。

$$\operatorname{argmax}_{x \in N_R} P(x)=\operatorname{argmax}_{x \in N_R} \prod_{k=1}^{|N_R|} \theta_k^{x_k} (1-\theta_k)^{(1-x_k)} \quad (9)$$

采用对数形式表示式(9),消除其中不依赖于 x 的项,获得下述等价的优化问题描述:

$$\begin{aligned} \operatorname{argmax}_{x \in N_R} P(x) &= \operatorname{argmax}_{x \in N_R} \prod_{k=1}^{|N_R|} x_k \log \frac{\theta_k}{1-\theta_k} \\ &= \operatorname{argmin}_{x \in N_R} \prod_{k=1}^{|N_R|} x_k \log \frac{1-\theta_k}{\theta_k} \\ \text{subject to } & \sum_{k=1}^{|N_R|} R_{ik} x_k \geq 1, 1 \leq i \leq |P_C| \end{aligned} \quad (10)$$

4.2 失效链路定位算法

式(10)中的优化问题可以映射为一加权集合覆盖(WSCP)问题,而 WSCP 问题是典型的 NP-hard 问题^[9]。现已提出很多优化算法来求解 WSCP 问题,大多数算法都需要很高的时间复杂度。本文采用了基于贪心策略的启发式求解算法,首先定义链路域 $Domain(l_k)$ 为包含链路 l_k 的路径集

合。算法使用贪心策略构造可行的解集合,每一步按照 $\log((1-p_k)/p_k)/|Domain(n_k)|$ 最小的标准选取一条链路 l_k 。

算法描述如下:

WFLL 算法

Input: 简化矩阵 R , 坏路径集合 P_B

Step1: 初始化 X 为空集合,令 $Q_B := P_B$

Step2: while $Q_B \neq \emptyset$

1) 发现 $n_k \in N_R$ 满足 $\log((1-\theta_k)/\theta_k)/|Domain(n_k)|$ 最小

2) 把 n_k 加入到 X 中, $X := X \cup \{n_k\}$ 同时 $x_k := 1$

5 算法仿真与评价

为了评价算法性能,本文采用了两个评价指标:覆盖率(DR, detection rate)和误判率(FPR, false positive rate)。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:

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

仿真实验通过扩展 NS2 仿真工具,使其能够仿真无线传感器网络中数据收集算法,用 Matlab 实现定位算法并分析收集到的数据。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。在仿真时,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,统计每条链路实际的报文丢失数,用于计算实际的链路报文丢失率。通过与推断结果的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 图形模块生成网络拓扑,节点数 V 在范围 100~1000 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。仿真时的丢包模式 LM 采用:好链路的丢包率服从区间 $[0, 0.01]$ 上的均匀分布,失效链路的丢包率服从区间 $[0.05, 1]$ 上的均匀分布。一旦每条链路被指定一个丢包率^[10],该链路的实际丢包过程用 Bernoulli 过程描述。每个报文在该链路传输过程中发生丢失的概率由链路报文丢失率所确定。路径丢包率基于对 1000 个报文传输的情况进行计算。设定 l_k 为 0.95, l_p 为 l_k^d ,在此 d 为路径的长度。

每次实验进行 200 轮数据收集,通过测量数据推断失效链路,并计算每次实验的 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。

图 2 和图 3 显示当网络拓扑固定、节点数 $V=500$ 、坏链路比例 f 在 $[0.05 \sim 0.35]$ 变化时算法性能的变化趋势。随着 f 的增加,WFLL 算法的 DR 出现下降趋势, FPR 略微增加,但性能总体变化不大。通过与 LLIS 算法对比,可以看出 LLIS 算法的 DR 下降幅度较大。出现这种趋势,是因为 LLIS 算法是建立在假设失效链路的数量相对稀少的基础之上,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。从 FDR 性能上看,LLIS 算法比 WFLL 算法的 FDR 高,两者的 FDR 变化幅度均不大。WFLL 算法通过对历史数据的学习获得了链路状态的先验概率,所以推断过程中对 FDR 影响不大。

(下转第 117 页)

3)节点密度。为了能够使得网络容易部署,系统采用了两种通信频率,因此只需要对通信频率进行调节(就是调节通信距离)就可以使得网络节点密度达到可调节的效果,可以针对不同的应用场合进行网络部署。

结束语 MFALM 定位方法是一种适用于水下部署的动态传感器网络定位算法,具有能耗低、网络寿命长、定位精确度高、节点密度可调节等优点,可以应用于诸如生物科学、潜艇探测等多种水下定位场景。需要注意的是,Max Hops 的取值与节点密度和 relay 节点的数量有关,较好地设置 Max Hops 的取值,可以有效地提高网络寿命。同时,水下环境的反射、多径、噪声、多普勒效应等现象会对定位精度产生较大影响,因此需要考虑对这些不良情况的克服方法。

参考文献

[1] Akyildiz I F, Pompili D, Melodia T. Challenges for efficient communication in underwater acoustic sensor networks[J]. ACM SIGBED Review, 2004; 3-8

[2] Cui J-H, Kong J, Gerla M, et al. Challenges: building scalable mobile underwater wireless sensor networks for aquatic applications[J]. IEEE Network, Special Issue on Wireless Sensor Networking, 2006; 12-18

[3] Chandrasekhar V, Seah W K G, Choo Y S, et al. Localization in Underwater Sensor Networks- Survey and Challenges [C] //

Proc. ACM WUWNet. Los Angeles, CA, USA, Sep. 2006; 33-40

[4] Heidemann J, Ye W, Wills J, et al. Research challenges and applications for underwater sensor networking [C] // Proc. IEEE WCNC. Las Vegas, NY, USA, 2006, 1; 228-235

[5] Niculescu D, Nath B. DV Based Positioning in Ad Hoc Networks [J]. Telecommunication Systems, 2003, 22(1-4); 67-80

[6] Akyildiz I F, Pompili D, Melodia T. Underwater acoustic sensor networks; research challenges [J]. Ad Hoc Networks (Elsevier), 2005, 3(3); 257-279

[7] He T, Huang C, Blum B M, et al. Range - Free Localization Schemes for Large Scale Sensor Networks [C] // Proc. ACM MOBICOM. San Diego, CA, USA, Sep. 2003; 81-95

[8] 白菊蓉. 第五站到空间四站 TDOA 定位中的模糊影响 [J]. 西安邮电学院学报, 2007, 12(1); 12-15

[9] Bulusu N, Heidemann J, Estrin D. GPS-less low cost outdoor localization for very small devices [J]. IEEE Personal Communications Magazine, March 2000; 28-34

[10] Niculescu D, Nath B. DV based positioning in ad hoc networks [J]. Journal of Telecommunication Systems, 2003, 22(1-4); 267-280

[11] Partan J, Kurose J, Levine B N. A survey of practical issues in underwater networks [C] // WUWNet '06; Proceedings of the 1st ACM International Workshop on Underwater Networks. 2006; 17-24

(上接第 110 页)

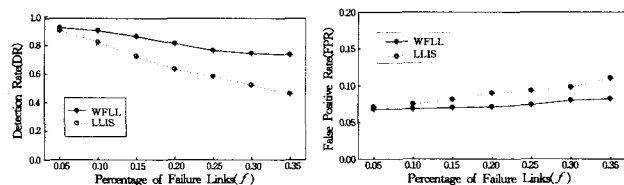


图 2 f 变化时 DR 性能变化趋势 图 3 f 变化时 FPR 性能变化趋势

结束语 故障管理的机制和方法是无线传感器网络管理中的重要组成部分。失效链路是无线传感器网络中故障的主要形式,研究其发现与定位的机制和方法有着重要意义。本文采用二元分离模型描述链路失效现象,基于简单网络断层扫描技术将失效链路定位问题归类为最大后验估计问题,通过分析、推导将失效链路定位问题映射为加权最小集合覆盖问题,提出了一种基于启发式策略的失效链路定位算法。通过仿真实验证明了算法的有效性。

参考文献

[1] Paradis L, Han Qi. A survey of fault management in wireless sensor networks [J]. Journal of Network and Systems Management, 2007, 15(2); 171-190

[2] Hartl G, Li Baochun. 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 2004

[3] Mao Y, Kschischang F R, Li Baochun, et al. A factor graph approach to link loss monitoring in wireless sensor networks [C] // IEEE JSAC, Special Issue on Self-Organizing Distributed Collaborative Sensor Networks. April 2005

[4] Coates M, Hero A, Nowa K R, et al. Internet tomography [J]. IEEE Signal Processing Magazine, 2002, 19

[5] Padmanabhan V N, Qiu L, Wang H J. Server-based inference of internet performance [C] // Proceedings of the IEEE INFOCOM'03. San Francisco, CA, April 2003

[6] Nick D. Simple network performance tomography [C] // Proceedings of the IMC'03. Miami Beach, Florida, October 2003

[7] Duffield N G. Network tomography of binary network performance characteristics [J]. IEEE Transactions on Information Theory, 2006, 52(12); 5373-5388

[8] Nguyen H X, Thiran P. Using end-to-end data to infer lossy links in sensor networks [C] // Proc. of IEEE INFOCOM. 2006

[9] Karp R M. Reducibility among combinatorial problems [M]. Complexity of Computer Computations. New York: Plenum Press, 1972

[10] Zhao Jerry, Govindan R. Understanding packet delivery performance in dense wireless sensor networks [C] // SenSys'03; Proceedings of the First International Conference on Embedded Networked Sensor Systems. 2003