

基于随机网络的故障节点定位算法设计

高建明

(浙江越秀外国语学院国际商学院 绍兴 312000)

摘 要 随着网络规模的增加,路由节点之间有着较大的随机分布性,节点之间难以建立精确的定位模型,基于特征匹配的传统方法在故障定位节点中难以体现其随机性的特点,从而造成故障的位置不准确。论述了随机网络的定义与特性、随机网络的度分布和特征谱;通过定义随机网络的直径和平均距离,计算随机网络的集聚系数,提取随机网络故障特征的有效参数,建立故障数据的特征模块,把随机行走网络节点故障检测算法的故障特征点作为登记点,对一个节点发生故障时产生的随机种子进行初始检测,再用故障特征和随机变化检测相结合的方法,提高检测的效率。仿真结果表明,改进的方法可以提高故障节点定位的精确度。

关键词 随机行走,搜索算法,节点,分布,集聚

中图法分类号 TP393.06 文献标识码 A

Design on Fault Localization Algorithm Based on Stochastic Network

GAO Jian-ming

(College of Internation Business, Zhejiang Yuexiu University of Foreign Languages, Shaoxing 312000, China)

Abstract With the increasing scale of the network, a large random distribution between routing nodes occurs. It is difficult to establish the accurate localization model, so the traditional methods based on feature matching is difficult to describe the random characteristics in fault location of nodes, resulting in inaccuracy of fault location. This paper discussed the overview and characteristics of stochastic networks, and degree distribution and characteristic spectrum of stochastic networks. Through defining the diameter and average distance for stochastic network, calculated the clustering coefficient of stochastic network, extracted effective parameters of stochastic networks fault feature, constructed a feature module of fault data. Using the fault feature node of random walk network fault detection algorithm as a random seed registration point and a node failure detection method for the initial, the methods prove the detection efficiency using the combination of random variation of fault characteristics and registration detection. Simulation results show that the improved method can improve the accuracy of fault localization.

Keywords Random walk, Search algorithm, Node, Distribution, Agglomeration

随着应用网络规模的增加,网络路由节点表现出随机性和多样性的特点,因各种网络故障引起的路由节点的破坏,使得不能建立一个有效的故障定位模型。基于特征匹配的传统随机故障定位方法在分析网络节点时难以获得准确的定位值。为解决上述问题,提出一个基于网络的检测算法中的随机行走的失效检测方法。该方法检测随机种子点和原始节点失效特征点,将登记特征的随机变异的检测位置与其匹配的故障点结合起来分析,提高其变化检测的效率问题。

在网络上找到异常的常用方法有基于距离的定位方法和非测距定位方法。它们都是基于一个特定节点定位其网络外部目标的方法。其中基于距离的定位方法是根据距离和角度测量其信息相应的硬件节点,即利用三角测量法和最大似然法预测及计算节点的位置来进行定位。该算法可得到精确的节点位置信息,但存在问题是,需要消耗大量的网络硬件资源,成本较高。非测距定位方法是基于运行时间的定位,通常认为非距离网络中相关的位置信息要到其他节点中去分析,目前主要有 DV-Hop、Centroid 等方法。由于网络应用的增

加,有效定位一个特定网络节点的问题已成为当前问题的理论分析重点。

1 随机网络的概述与特性

1.1 随机网络模型

网络是节点和其连线的集合。如果节点按确定的规则连接,由此产生的网络称为正则网络。如果节点不是按确定的规则连线,如通过纯粹的随机方式连接,所得到的网络被称为随机网络。早期的研究主要集中在一个小规模的正规网络。20 世纪 50 年代末,匈牙利科学家斯 Erdos 和 Renyi 将随机性引入到网络中,提出了著名的随机网络模型,简称 ER 模型。该模型描述了从多个随机分布的点通过相同概率随机相连而形成网络的过程。由于具有复杂拓扑结构和未知组织规则的大规模网络所表现出的随机性,所有 ER 随机网络模型常常被用于复杂网络研究中。随机网络是由一些节点通过随机连接而组成的一种复杂网络,在图论中,随机网络构成有两种等价方法,

本文受全国教育信息技术研究课题(126240673)资助。

高建明(1975-),男,硕士,讲师,网络工程师,主要研究方向为计算机网络及安全。

①ER模型: 给定 N 个节点, 最多可以存在 $N(N-1)/2$ 条边, 从这些边中随机选择 M 条边就可以得到一个随机网络, 显然一共可产生 $C_{N(N-1)/2}^M$ 种可能的随机图, 且每种可能的概率相同;

②二项式模型: 给定 N 个节点, 每一对节点的连接概率设为 p 。则所连线的数目是一个随机变量, 其平均值定义为 $M = pN(N-1)/2$ 。若 G_0 是一个节点为 v_1, v_2, \dots, v_N 和 M 条边组成的图, 则得到该图的概率为 $P(G_0) = p^M(1-p)^{N(N-1)/2 - M}$, 其中 p^M 是 M 条边同时存在的概率, $(1-p)^{N(N-1)/2 - M}$ 是其他边都不存在的概率, 二者是独立事件, 故二概率相乘即得图 G_0 存在的概率。

ER模型的一个特点是: 当连接概率 p 超过某个临界概率 $p_c(N)$, 许多性质就会突然改变。例如, 针对随机图的连通性, 若 p 大于临界值 $(\ln N)/N$, 那么几乎每一个随机图都是连通的。若当 $N \rightarrow \infty$ 时, 连接概率 $p = p(N)$ 的增长比 $p_c(N)$ 慢, 则几乎所有连接概率为 $p(N)$ 的随机图都不会有性质 Q 。相反, 若连接概率 $p(N)$ 的增长比 $p_c(N)$ 快, 则几乎每一个随机图都有性质 Q 。因此, 一个有 N 个节点和连接概率 $p = p(N)$ 的随机图有性质 Q 的概率满足:

$$\lim_{N \rightarrow \infty} P_{N,p}(Q) = \begin{cases} 0, & P(N)/P_c(N) \rightarrow 0 \\ 1, & P(N)/P_c(N) \rightarrow \infty \end{cases}$$

1.2 随机网络的度分布

在连接概率为 p 的 ER 随机图中, 可知其平均度为 $\langle k \rangle = p(N-1) \approx pN$ 。而某节点 v_i 的度 k_i 等于 k 的概率, 则遵循参数为 $N-1$ 和 p 的二项式分布 $p(k_i = k) = C_{N-1}^k p^k (1-p)^{N-1-k}$ 。值得注意的是, 若 v_i 和 v_j 是不同的节点, 则 $P(k_i = k)$ 和 $P(k_j = k)$ 是两个独立的变量。为了找到随机图的度分布, 需得到度为 k 的节点数 X_k 。为此, 需要得到 X_k 等于某个值的概率 $P(X_k = r)$ 。连接度为 k 的平均节点数为 $\lambda_k = E(X_k) = N \cdot P(k_i = k)$, 即 $\lambda_k = NC_{N-1}^k p^k (1-p)^{N-1-k}$ 。 X_k 值的概率接近泊松分布 $P(X_k = r) = \frac{e^{-\lambda_k} \lambda_k^r}{r!}$, 这样一来, 度为 k 的节点数目 X_k 满足均值为 λ_k 的泊松分布。上式意味着 X_k 的实际值和近似结果 $X_k = N \cdot P(k_i = k)$ 并没有很大偏离, 只是要求节点相互独立。这样, 随机图的度分布可近似为二项式分布。在 N 比较大的条件下, 它可以被泊松分布 $p(k) = \frac{e^{-pN} (pN)^k}{k!} = \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}$ 取代。

由于网络节点之间的连接是等概率的, 因此大多数节点在平均值 $\langle k \rangle$ 附近时, 网络中没有度特别大的节点。对于大范围内的 p 值, 最大和最小的度值都是确定性的和有限的。若 $p(N) \propto N^{-1-1/k}$, 几乎没有图有度大于 k 的节点。另外一个极值情况是, 若 $p = [\ln(N) + k \ln(\ln(N)) + c]/N$, 几乎每一个随机图都至少有最小的度 k 。图 1 给出 $N = 1000, p = 0.0015$ 时随机网络的度分布, 其中图中的点代表 X_k/N (度分布), 而连续曲线代表期望值 $E(X_k)/N = p(k_i = k)$, 可以发现两者偏离确实很少。

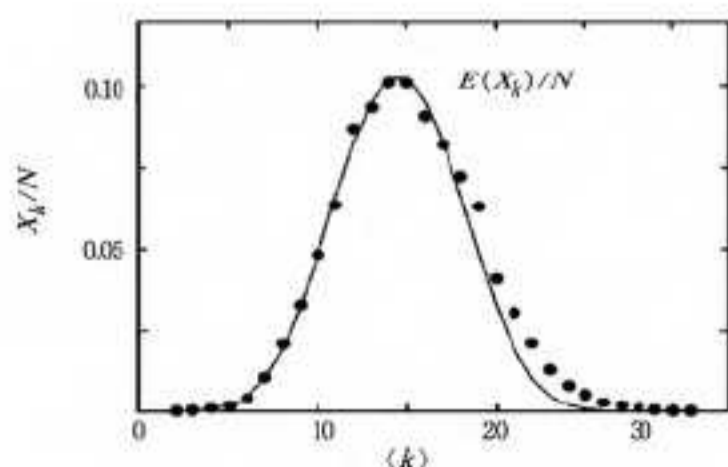


图 1 ER模型的度分布 ($N=1000, p=0.0015, k=15$)

1.3 随机网络的特征谱

考查连接概率 $p(N) = cN^{-z}$ 的随机网络 $G_{N,p}$ 的特征谱。该网络的平均度为 $\langle k \rangle = Np = cN^{1-z}$ 。当连接概率中的参数变化时, 随机网络的特征谱会发生转变, 具体表现如下所述。

若 $0 \leq z < 1$, 图 $G_{N,p}$ 中将出现无限聚类体, 并且当 $N \rightarrow \infty, \langle k \rangle \rightarrow \infty$ 时, 任何节点都是几乎完全属于无限的聚类体。在这种情况下, 随机图的频谱密度发散到如下半圆形分布, 如图 2 所示。图中 p 值固定为 0.05。

$$\rho(\lambda) = \begin{cases} \frac{\sqrt{4Np(1-p) - \lambda^2}}{2\pi Np(1-p)}, & |\lambda| < 2\sqrt{Np(1-p)} \\ 0, & \text{其它} \end{cases}$$

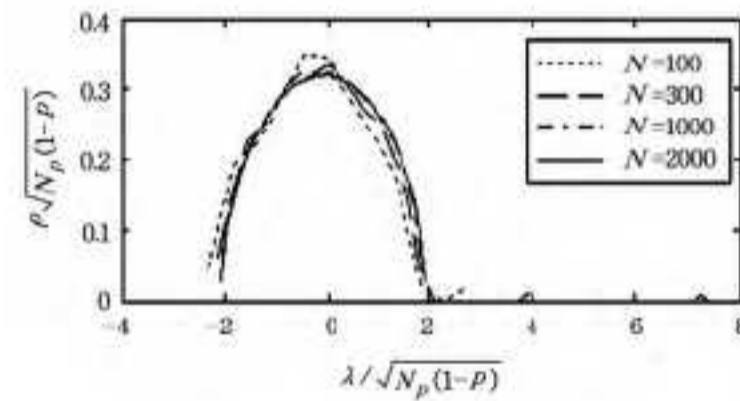


图 2 ER网络谱密度(50次运行的平均值曲线)

由图 2 可见, 最大的特征值 λ_1 是和频谱孤立的, 并且随着网络大小衰减为 pN 。

当 $z > 1$ 时 (N 取 3000), $\rho(\lambda)$ 偏离半圆形分布, 如图 3 中的点划线所示, 而且当 $N \rightarrow \infty$ 时, $\langle k \rangle \rightarrow 0$, 此时 $\rho(\lambda)$ 的奇数阶矩等于 0, 这意味着要回到原节点的路径只能是沿来时经过的相同节点返回, 这正好表明网络具有树状结构。

当 $z=1$ 且 $N \rightarrow \infty$ 时, 节点的平均度数 $\langle k \rangle = c$ 。此时, 若 $c \leq 1$, 网络基本上为树状结构; 而若 $c > 1$, 它的谱密度的奇数阶矩远大于 0, 说明网络的结构发生了显著的变化, 出现了网络环和网络分支。当 $z=1, N=3000$ 时的谱密度如图 3 所示。

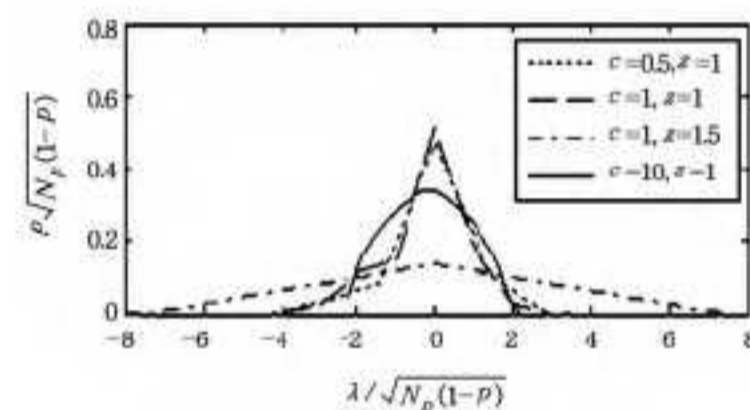


图 3 ER网络谱密度(50次运行的平均值曲线)

2 随机行走的网络故障节点定位算法

2.1 定义随机网络的直径和平均距离

对于大多数的 p 值, 几乎所有的图都有同样的直径值。这就意味着连接概率为 p 的 N 阶随机图的直径的变化幅度非常小, 通常集中在 $D = \ln(N)/\ln(\langle k \rangle) = \ln(N)/\ln(pN)$ 。

若 $\langle k \rangle$ 小于 1, 则图由孤立树组成, 且其直径等于树的直径。若 $\langle k \rangle$ 大于 1, 则图中会出现连通子图。当 $\langle k \rangle$ 大于等于 3.5 时, 图的直径等于最大连通子图的直径且正比于 $\ln(N)$ 。若 $\langle k \rangle$ 大于等于 $\ln(N)$, 则几乎所有图都是完全连通的, 其直径集中在 $\ln(N)/\ln(pN)$ 左右。

随机网络的平均最短距离可以估算如下: 考虑平均度 $\langle k \rangle$ 的随机网络, 任何节点上, 一阶邻居节点数为 $\langle K \rangle$, 二阶的邻居节点数为 $\langle k \rangle^2$ 。也就是说, 在 ER 随机图中随机选择一个节点 v_i , 网络中大约有 $\langle k \rangle L_{rand}$ 个节点与节点 v_i 的距离为 L_{rand} 。依此类推, l 步后达到网络的总节点数目 N , 有 $N = \langle k \rangle^l$

故 $L_{rand} \propto \ln(N)/\ln(\langle k \rangle)$, 可以看出, 平均最短距离的随机网络随着网络规模的增加呈对数级增长, 这是典型的小世界效应。因为 $\ln N$ 随 N 增长得很慢, 所以即使是一个很大规模的网络, 它的平均距离也很小。

2.2 计算随机网络的集聚系数

因为任何两个随机的节点之间的网络连接概率相等, 所以对于一个节点 V_i , 在相邻节点之间的连接的概率也是 P , 使得随机网络的聚类系数为 $C_{rand} \approx p = \langle k \rangle / N$ 。

然而, 真实网络并不遵循随机图分布的规律, 相反, 其集聚系数并不依赖于 N , 而是依赖于节点的邻居数目。通常, 在具有相同的节点数和相同的平均度的情况下, ER 模型的集聚系数 C_{rand} 比真实复杂网络的要小得多。这意味着大规模的稀疏 ER 随机图一般没有集聚特性, 而真实网络一般都具有明显的集聚特性。则网络的共同特征和平均聚类系数大且平均距离长, 其特点是随机网络的聚类系数低, 平均距离小。假设有两个网络节点定位点 S_1, S_2 , 在两个节点收集相同的位置, 且采集时间为 T_1 和 T_2 。在故障定位算法的基础上提出随机行走的目标是创建一个网络位置的非监督变化检测模型, 其中包含了注册信息和信息的相应变化的故障特征提取的位置信息。该方法所描述的流程如图 4 所示, 包含了主要的特征提取值, 基于三部分的迭代最小化和故障数据的权重模型。

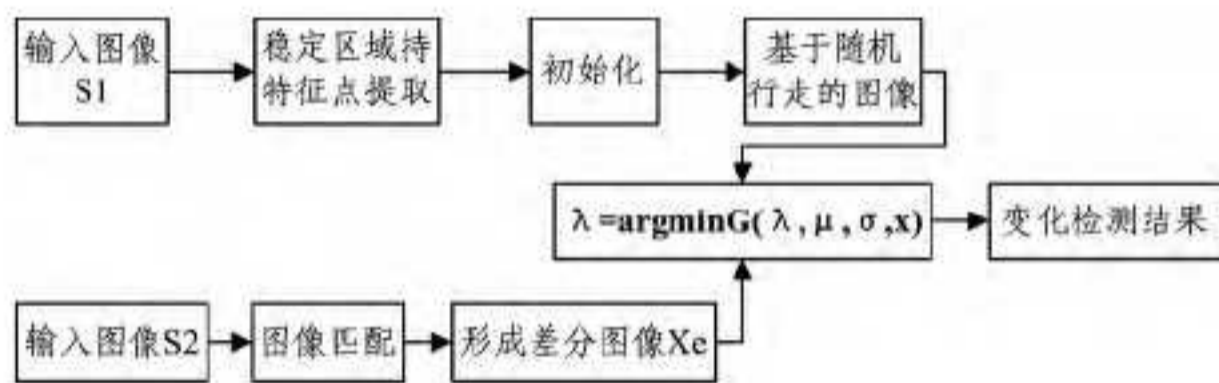


图 4 基于随机行走的变化检测流程

2.3 提取随机网络故障特征的有效参数

通常两个或两个以上的节点的距离没有一个统一稳定的值, 通过路由节点的有效性值可以定位故障和描述这些特征。使用收集到的数据提取故障特征值, 然后找到一个基于随机行走方法的原始种子点的关键特征值。先设置关键特征值 $L=50$ 的经验阈值, 然后决定保留故障值。因为这种方法容易混淆, 故将节点的权重纳入相应的干扰因素进行分析。

2.4 建立故障数据的特征模块

在获得 S_1 和 S_2 固定数据通信节点的差分模型的精确故障特征数据后, 构造 S_E 模型对应的权重计算函数。网络通信节点 S_1 和 S_2 的 S_E 差模型定义如式(1)所示:

$$S_E = S_1 - S_2 \quad (1)$$

对微分节点模型 S_E 建模以优化模型之间的差异。先构造形成微分方程区域稳定性差模型。结果是差分图像检测得到一个 $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)$, 透明度值划分结果是确定的, 有 $\lambda_N \in \{0, 1\}$, 设置值“1”表示在该地区是变化区, “0”代表稳定区。 δ 代表描述一个高斯混合模型参数。

为了保证独特的功效, 建立高斯混合模型, 有效提取联合高斯模型的一个独特的组成部分——高斯混合模型分量 u_n , 且有 $u = \{u_1, \dots, u_n, \dots, u_N\}$, $u_n \in \{1, \dots, m\}$ 。检测变化的能量函数, 如式(2)所示:

$$G(\lambda, \mu, \sigma, x) = F(\lambda, \mu, \sigma, x) = \sum_{n=1}^N Z(\lambda_n, \mu_n, \sigma, x_n) \quad (2)$$

式中, $G(\cdot)$ 用于描述透明度为 λ 的分布节点特性和输入数

据 x 相似性特征。 $\pi(\cdot)$ 是用来描述节点定位的后验概率。 Σ 代表协方差, η 表明分配的均值。最大后验概率预测函数 G 对应一个二分类问题的解决方案, 其由模型和稳定区域的变化值决定。因此, 减少最小化能量函数可以得到故障特征的变化检测结果, 如式(3)所示。

$$\lambda = \arg \min G(\lambda, u, \delta, x) \quad (3)$$

以上所述的权值最小化的过程由随机行走的一种迭代算法来实现。

2.5 迭代权值最小化的节点故障定位

故障定位的过程就是实现随机行走的一种迭代算法, 以确保最佳的故障定位结果。把原种子点作为区域 Z_w 稳定值为第一关键特征, 并最终对区域的种子点的 C_D 变化模型区域以外的稳定区域连续模型的建立进行初始化。如果阈值大于改变区域稳定区的权值, 可以设置该区域的随机种子点。经过随机行走后得到的种子点即为分析故障定位的节点。改变首次使用原始种子点区域后, 一个随机的节点被标记为操作点, 每个节点的操作对应于第一次到达时间的随机种子点的概率点。第一个到达的随机种子点的操作点被标记为每个节点的身份标识。当最初的种子点少时, 算法还不能有效地定位结果, 应当调整权重, 定位的结果是使迭代最小化。该算法分析如下:

- 1) 改变该区域的种子点稳定性区域以及透明度 λ_n ;
- 2) 在一个给定的节点和对应的向量区域稳定性模型中定义 $n \in C_w, u_n = \arg \min Z_n(\lambda_n, u_n, \delta, x_n)$;
- 3) 基于随机行走位置的定位算法的结果为 λ , 且其权值为 $G(\lambda, u, \delta, x)$;
- 4) 建立基于定位的区域变化的结果的模型参数 δ ;
- 5) 改变稳定区域的种子点模型区域。区域的变化作为一个种子点的定位结果, 将该区域中的高稳定性的节点作为种子点, 实现种子不断更新, 其值为 $G(\lambda, U, \delta, x)$, 达到最低值时结束。相应的故障节点的最小权值就是最好的定位结果。

3 实验仿真结果及分析

实验验证算法的有效性, 通过 PC P4 intel e7500, 4G 内存的计算机, 使用 matlab7.0 相应的实验仿真软件。建立一个二维传感器网络环境, 路由节点包括了一个通信网络, 其节点分布在一个 $K \times K$ 的正方形的面积中, 预先设定锚节点的坐标, 不存在干扰因素。在通信网络本身是位于区域中心时, 它的通信节点圆半径设为 R , 锚节点与未知节点具有相同的性能特性, 对称性的通信每个节点具有均匀的收发功能。其平均位置误差设置如式(4)所示:

$$Error = \frac{\sum_{i=1}^{EKnown} \sqrt{(X_{real} - X_{est})^2 + (Y_{real} - Y_{est})^2}}{EKnown} \quad (4)$$

式中, $EKknown$ 表示未知节点的数目, (X_{real}, Y_{real}) 代表未知节点的坐标, (X_{est}, Y_{est}) 是预测的未知坐标节点。在 500×500 区域模拟实验 500 个节点的任意 1000 组, 这些组节点的通信半径为 50。当锚节点服从均匀分布和随机分布时, 传统的 DV-Hop 算法仿真结果与平均相对误差定位算法所描述的一样。分析可以得出, 该算法的定位误差小于传统的 DV-Hop 算法, 并在节点数为 400 时, 误差小于 40%。传统的 DV-Hop

(下转第 292 页)

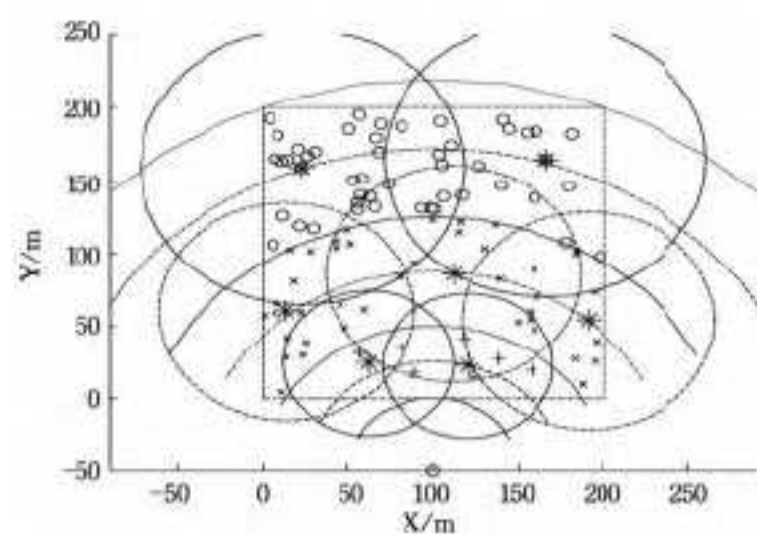


图7 本文算法随机一轮生成的拓扑结构图

(2) 在随机生成的一种节点分布模型下, 3种分簇算法的网络生存时间如图8所示。从图可以看出, 新算法延长了网络的生存时间。在簇头选举时, 新算法先对节点分层, 然后在每一层各自展开簇头的选举, 选举方法比EEUC算法简单高效, 减小了能耗。

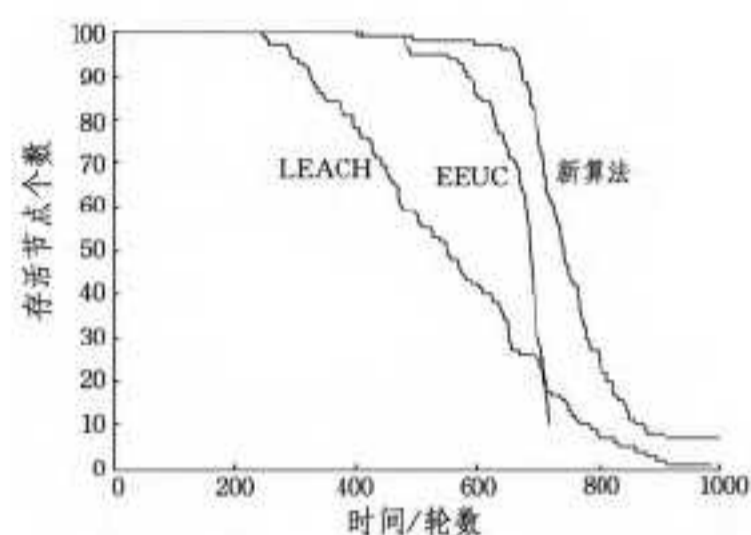


图8 3种分簇算法的网络存活时间比较

结束语 本文研究了无线传感器网络接近实际的一种网络模型, 通过对网络非均匀分层并且将各层节点组织成不同规模的簇, 使靠近基站的节点簇规模较小, 以此来达到均衡簇头能耗的效果。文中重点优化了簇头选举方法, 使得新算法运行简单, 产生的拓扑结构更为合理。仿真实验证明新算法能降低成簇开销, 均衡网络节点能耗和延长网络寿命, 是解决

(上接第288页)

算法误差率在700时达到40%。本文算法的效率和精度比传统的DV-Hop算法要高。当节点的数量达到400时, 锚节点所占比例增加, 结果提高了算法的平均定位误差率, 如图5所示。通过图5可知, 相比DV-Hop定位算法, 本文算法的误差率减小了。

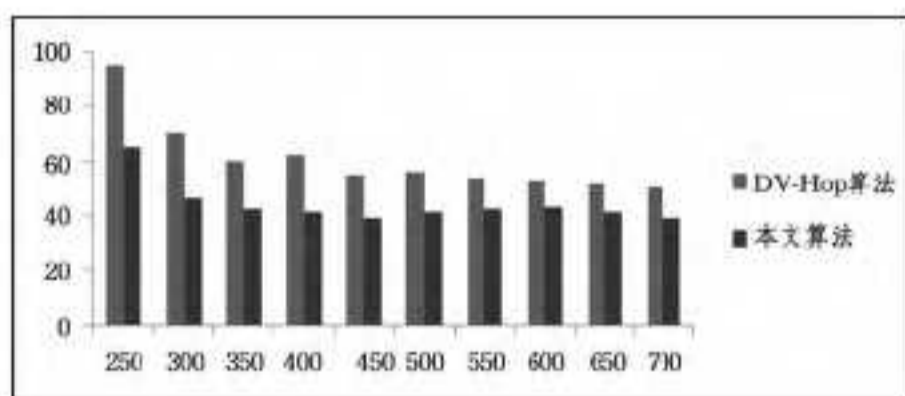


图5 锚节点平均分布的平均定位误差

结束语 基于一个随机行走的非监督变化检测方法是一个典型的变化检测方法。该方法的特点是提取遥感图像的关键点作为一个随机行走的初始种子点, 使用一个策略来减小像素提取的不透明度。它和特征点的图像配准变化检测相结合, 提高了变化检测的效率。本文算法适用于现实网络, 具有定位误差小、范围广等优点, 比传统的DV-Hop定位算法具有更好的精度, 是一种高效的定位算法, 具有较强的应用价值。

参考文献

[1] 江山. 有关计算机网络路由的分析与探讨[J]. 科技资讯, 2012

无线传感器网络多跳通信中“热区”问题的一种有效方法。

参考文献

[1] Soro S, Heinzelman W. Prolonging the lifetime of wireless sensor networks via unequal clustering [C] // Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium. San Francisco: IEEE Computer Society Press, 2005: 236-240

[2] 李成法, 陈贵海, 叶懋等. 一种基于非均匀分簇的无线传感器网络路由协议[J]. 计算机学报, 2007, 30(1): 27-36

[3] 刘志. 无线传感器网络中的能量高效覆盖与路由算法研究[D]. 北京: 北京交通大学, 2011

[4] 董杨, 郭拯危, 王青正. 一种基于非均匀分簇的WSN路由协议[J]. 郑州轻工业学院院报, 2009, 24(6): 70-74

[5] 李超良, 胡春华. 无线传感器网络中面向动态多跳的非均匀分簇路由[J]. 中南大学学报, 2011, 42(7): 2048-2053

[6] Heinzelman W, Chandrakasan A, Balakrishnan H. Energy-efficient Communication Protocol for Wireless Microsensor Networks [C] // Proceedings of the 33rd Hawaii International Conference on System Sciences. Maui, Hawaii, USA: IEEE Computer Society, 2000

[7] 马慧颖. 无线传感器网络路由协议研究[D]. 哈尔滨: 哈尔滨工程大学, 2008

[8] 黄琛, 房鼎益, 陈晓江. 传感器网络中基于非均匀分簇负载均衡路由算法[J]. 计算机应用研究, 2009, 26(9): 3475-3477

[9] 林观康, 程良伦. 基于地理信息静态分簇的无线传感器网络路由算法[J]. 计算机应用与软件, 2011, 28(2): 37-39

[10] Heinzelman W, Chandrakasan A, Balakrishnan H. An Application-Specific Protocol Architecture for Wireless Microsensor Networks [J]. IEEE Transactions on Wireless Communications, 2002, 1(4): 660-670

[28] 曹付元. 面向分类数据的聚类算法研究[D]. 太原: 山西大学, 2010

[3] Ghosh A K, Michael C, Schatz M. A real-time intrusion detection system based on learning program behavior [C] // Algorithms for Finding Paths with Multiple Constraints Networks. 1984: 95-116

[4] 于兴敏, 彭学平, 林培芳. 现代预分解窑设计产量初探 [J]. 中国建材装备, 1998(4)

[5] Celik T. Unsupervised change detection in satellite image using principal component analysis and k-means clustering [J]. IEEE Geoscience and Remote Sensing Letters, 2009, 6(1): 33-37

[6] Rouhollah D, Shohreh K. Change detection in optical remote sensing images using difference-based methods and spatial information [J]. IEEE Geoscience and Remote Sensing Letters, 2010, 7(1): 215-219

[7] Kulpers R, Van Mlegheem P, Korkmaz T, et al. An overview of constraint based path selection algorithms for Qos routing [C] // IEEE Communications Magazine. December 2002

[8] Korkmaz T, Krunz M. A randomized algorithm for finding a path subject to multiple Qos requirements [J]. Computer Networks, 2001, 36: 251-268