一种新的容忍恶意节点攻击的无线传感器网络安全定位方法

叶 苗 王宇平

(西安电子科技大学计算机学院 西安 710071)

摘 要 无线传感器节点位置定位正确与否对整个网络传感器起着至关重要的作用.当无线传感器网络暴露在恶意危险环境中时,攻击者会攻击节点定位的过程,使其定位到错误位置,从而导致整个网络应用完全失效.基于最大似然估计的传感器定位概率模型是一种常用的定位模型,但是它有两个缺点:(1)为了降低计算复杂性,通常将RSS(接收信号强度)信号标准差看成常数,影响定位精度;(2)安全性差,在有恶意节点攻击时模型常常会定位失效.文中首先通过拟合测试数据归纳出了 *RSS* 信号标准差随距离变化的函数关系,克服了第一个缺点.针对第二个问题,在分析其受攻击时定位失败的具体原因后,对节点定位的概率计算公式进行了改进,设计了一种新的基于变方差特征的传感器节点定位概率模型.该模型属于高度非线性全局优化问题.针对其难以求解的特点,文中设计了一个新的有效的进化算法,并证明了该算法的全局收敛性.最后通过对公开数据集的测试和实际实验,验证了该模型和求解算法能在保证定位精度的前提下,完成节点的安全定位.

关键词 无线传感器网络;变方差;过滤;投票法;进化算法;物联网

中图法分类号 TP393 **DOI 号** 10.3724/SP.J.1016.2013.00532

A New Malicious Nodes Attack-Resistant Security Location Method in Wireless Sensor Network

YE Miao WANG Yu-Ping

(*School of Computer Science and Technology, Xidian University, Xi'an* 710071)

Abstract It is crucial that wireless sensor nodes should be properly located to facilitate the operation of the whole network. When wireless sensor network is exposed in malicious and dangerous environment, attackers may attack the nodes in the location process and cause incorrect location results which may lead to the complete breakdown of the entire network. The sensor location probability model based on maximum likelihood estimation is one of the commonly-used location models. However, it has two flaws. First, it usually treats the standard deviation of the received signal strength (RSS) as constant to lower the calculation complexity, which affects the accuracy of location. Second, it is not secure enough. Under malicious node attack, this model usually cannot fulfill its location function. This research generalizes the functional relations between RSS standard deviation and distance through fitting test data and thus fixes the first problem. To tackle the second problem, this research analyzes the reasons behind the failure of location under attack and thus improves the probability formula of node location, and designs a new sensor node location probability model based on the characteristics of variant variance. As this new model is a highly nonlinear characteristic global optimization problem that is difficult to work out, this

收稿日期:2011-12-05;最终修改稿收到日期:2012-09-19.本课题得到国家自然科学基金(61272119,61203372)资助.**叶 苗**,男,1977 年生,博士研究生,主要研究方向为无线传感器网络、信息安全、进化算法.E-mail:ym@mail.xidian.edu.cn.**王宇平**,男,1961 年生,博士,教授,博士生导师,主要研究领域为进化计算、最优化理论、网络和工程设计中的优化方法、数据挖掘.

research has designed a new and effective evolutionary algorithm and proven its global convergence. In tests using public datasets and actual experiments, the designed model and algorithm are finally proven to be able to fulfill secure location of the nodes on the premise that the accuracy of location is guaranteed.

Keywords　wireless sensor networks; variant variance; filtering; voting-based algorithm; evolutionary algorithm; Internet of Things

1　引　言

无线传感器网络是由大量成本低廉、存储能量有限、分布广泛的传感器节点组成的网络，这些节点之间能在短距离内进行无线通信. 它在医疗健康监视、危险环境数据采集和军队战场操作等领域中都有广泛应用[1]. 节点定位技术是无线传感器网络应用(如地理路径路由协议 GPSR 和 GEAR、环境检测、目标跟踪等)的基础，一个没有位置信息的节点，在传感器网络中不会起到任何作用[1-2]. 目前，已经设计出很多位置发现协议(也称为定位技术)[3]. 这些定位技术通常会使用一些已知自身位置信息的节点(信标节点，也称为锚点)进行定位，步骤如下：首先，待定位的普通节点(非信标节点)接收到从信标节点发送过来的无线信号数据包，这些数据包中携带着对应信标节点的位置信息；然后，从接收到的信号数据包中包含的信息(比如接收信号强度 RSS[3-4]、到达时间差[5-6]、数据包记录的路径跳数等[7])中计算出发送节点和接收节点之间的距离，再依据这些测量来估计普通节点的位置. 常用的估计算法有三边测量法[8]、三角测量法[9]、最大似然估计法(概率可能性最大)[10]、质心法等. 其中，依据接收信号强度(RSS)与位置距离之间的概率分布关系，利用最大似然估计法(概率可能性最大)来估计普通节点位置的方法是一种常用的定位方法[10].

但这些定位方法通常都忽略了存在恶意节点攻击的情况，在恶意节点攻击环境中会发生定位错误[1,11-12]. 比如，攻击者在不同位置通过拦截并伪造信标节点数据包，提供虚假位置信息，或者通过捕获控制信标节点，更改信标节点信号值等行为，将导致普通节点的定位错误，从而颠覆无线传感器网络正常的应用. 虽然可以通过加密和认证等传统安全技术(比如给每个数据包加密，待定位的普通节点只接收通过加密认证的信标节点数据包)提高节点安全定位的准确性，但这样传统的认证机制在无线广播环境中很难保证足够的安全，攻击者很容易采用复制、阻塞和改变传播路径等手段破坏这些传统的安全技术，导致定位失败. 而且有些定位方法是由于本身设计缺陷、缺乏辅助安全机制导致安全问题的，概率因子连乘形式的最大似然估计法就属于这样的例子. 因此必须设计一些辅助的安全机制为节点定位提供安全保障.

传感器网络中基于最大似然估计[10]的节点定位概率模型是一种常用的定位模型，但是它有两个缺点：(1)为了降低计算复杂性，通常将 RSS 信号标准差看成常数，不能反映实际情况，影响定位精度；(2)在有恶意攻击节点时，模型常常定位失效. 为了克服这些缺点，本文首先针对基于最大似然估计的传感器定位概率模型，通过拟合 RSS 信号标准差随距离的变化规律，给出了它们的函数关系，克服了该标准差取常数值带来的缺陷，从而对原模型进行了改进. 其次，针对改进模型没有考虑恶意节点攻击时会出现定位失效的问题，采用投票法的思想，对模型中的关键技术：定位节点的概率计算公式进行了重新设计，解决了原模型在有恶意节点攻击时出现定位失效的问题，实现安全定位. 然后，针对设计的模型为高度非线性的全局优化问题而难以求解的特点，设计了一个新的有效的进化算法，并利用随机论[13-14]证明了设计算法的全局收敛性. 最后通过对公开数据集[15]①的模拟和实际实验，验证了本文模型和求解算法能在保证定位精度的前提下，完成节点的安全定位. 只要大多数的信标节点为正常节点，该定位模型和求解算法就可以很好抵御共谋攻击，过滤掉恶意节点的定位参考信息，能解决传统概率似然定位方法在攻击环境下彻底失效的问题. 按照文献[11]的分类方法，节点定位安全措施按照设计目标分可以分为三类：预防[16]、检测[17]②和过滤[18-19]. 本文提出的安全方法属于一种过滤算法，除了能解决传统最大似然估计定位方法在攻击环境下

① http://www.eecs.umich.edu/~hero/localize/
② http://www.cis.temple.edu/~wu/DRBTS_JournalVersion_JoATC.pdf

定位彻底失效的问题之外，相比投票过滤算法[18]能避免投票法网格大小和网格数目影响求解时间性能的缺陷，相比求最大一致信标集的恶意节点过滤算法(如 Attack-Resistant Minimum Mean Square-Estimate，AR-MMSE)[19]能避免其由于组合数爆炸而不得不使用贪婪算法获得近似解的缺陷.

2 传感器网络中基于变方差特征的最大似然估计定位模型

无线信号是一种电磁波信号，考虑各向同性球面波的 RSS(接收信号强度，单位：dB)信号，其在传播过程中会被传播介质吸收部分能量，强度会随距离成指数衰减，衰减规律可以表示为式(1)

$$RSS(d)=PL0-10\eta\log_{10}(d/d_0) \tag{1}$$

其中 $RSS(d)$表示接收端信号强度值(单位是 dB)，d 表示发射点到测量点的距离，d_0 表示参考距离，$PL0$ 是在参考距离为 d_0 处的接收信号强度值(单位是 dB)(通常取 $d_0=1$)，η 表示路径损失因子，即衰减因子.

由于干扰噪声的存在，测量的 RSS 值往往有偏差，由此式(1)可修改成(取 $d_0=1$)

$$RSS(d)=PL0-10\eta\log_{10}d+N_{\sigma_N} \tag{2}$$

其中 $N_{\sigma_N}\sim N(0,\sigma_N^2)$表示噪声干扰. 它是零均值、标准差为 σ_N 的高斯分布的随机变量，则在实际距离为 d 的条件下信号值为 RSS 的条件概率分布就可表示为

$$P(RSS|d)=\frac{1}{\sigma_N\sqrt{2\pi}}\mathrm{e}^{\frac{-(r_1-0)^2}{2\sigma_N^2}} \tag{3}$$

其中，$r_1=RSS-PL0+10\eta\log_{10}(d)$. 在实际测量表示中，式(2)通常表示为 $RSS(\tilde{d})=PL0-10\eta\log_{10}\tilde{d}$，其中，$RSS(\tilde{d})$表示由测量得到的 RSS 值，$\tilde{d}$ 表示测量得到的距离，从而有

$$\tilde{d}=g(RSS)=10^{\frac{RSS-PL0}{10\eta}} \tag{4}$$

用 $P(d)$表示距离为 d 的先验概率，$\tilde{d}$ 表示通过式(4)计算得到的测量距离，$P(d|\tilde{d})$表示在测量距离为 $\tilde{d}$ 条件下实际距离为 d 的概率，由式(3)和式(4)可得

$$\begin{aligned}P(d|\tilde{d})&=P(\tilde{d}|d)P(d)/P(\tilde{d})\\&=P(RSS|d)\cdot\left|\frac{\mathrm{d}g}{\mathrm{d}RSS}\right|\cdot P(d)/P(\tilde{d})\\&=\frac{1}{\sigma_N\sqrt{2\pi}}\mathrm{e}^{\frac{-(10\eta\log_{10}(d/\tilde{d}))^2}{2\sigma_N^2}}\cdot\tilde{d}\cdot\frac{\ln 10}{10\eta}\cdot P(d)/P(\tilde{d})\end{aligned} \tag{5}$$

由于随机事件的概率分布要求随机事件发生的初始条件必须相同，在不同地点对 RSS 进行测量，场景不一样，概率事件发生的初始条件也不一样，因此式(3)和(5)中的 σ_N 通常不是常量. 在很多文献，如文献[20-21]中，将 σ_N 作为常数处理，仅仅是为了减少计算复杂度、方便求解而对模型进行的简化，但这往往影响模型的精度. 一些文献讨论了 σ_N 不为常数的情形，如文献[15]取 σ_N 的上界作为 σ_N 的估值，这相当于还是将其看成常数；文献[22]直接给出了 σ_N 随距离 d 的变化关系，但没给出理由，建模依据不足；文献[23]通过实验证实了 σ_N 不为常数的结论，但只是利用此结论给出了一个选择信标节点的原则(依此原则设计了一种模糊贴近度的间接定位方法)，没有讨论如何将 σ_N 的变化规律直接用在定位方法的设计上. 采用类似文献[23]的实验方法，本文在感知 RF2-210 传感器网络实验教学平台上，利用实验方法拟合出两个传感器节点之间 RSS 信号标准差与距离之间的函数关系，从而给出了 σ_N 变化规律的定量描述. 感知 RF2-210 是国内公司开发的一种无线传感器网络实验教学平台，传感器节点采用 ZigBee 协议. 在室内办公室环境中保持 15 dB 发射功率，每改变一次两个节点之间的距离，采集 1000 次 RSS 信号值，求出标准差. 图 1 统计出了 RSS 信号标准差与距离的变化关系，所得结果和文献[23]的测量结果类似. 其实图 1 中标准差 σ_N 与距离间的变化规律其实是有共性的，在越靠近信标节点的地方，受到的噪声影响越小，标准差 σ_N 也越小；随着距离增加，噪声影响增加，σ_N 也变大；再随着距离的增加，RSS 信号会减弱，在有限噪声影响下，σ_N 也跟着减小. 结合图 1 所示实验结果及以上分析，RSS 信号的标准差与距离的变化关系接近高斯函数，可以用高斯函数 $\sigma_N(d)=a\mathrm{e}^{-\frac{(d-d_0)^2}{b^2}}$ 进行拟合，其中参数 a,b,d_0 的值可以从训练测试数据中进行曲

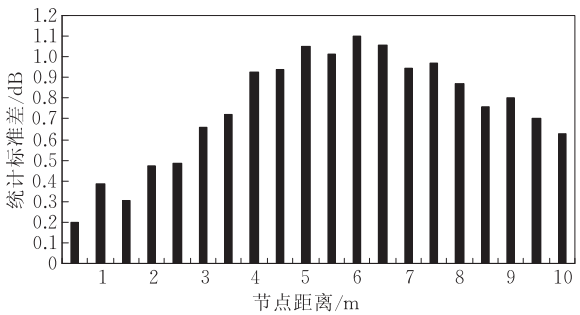

图 1 RSS 信号标准差随距离变化关系图

线拟合得到(本文的实验部分给出了拟合方法).因此,式(5)可以表示为

$$\begin{aligned}P(d|\tilde{d})&=P(\tilde{d}|d)P(d)/P(\tilde{d})\\&=P(RSS|d)\cdot\left|\frac{\mathrm{d}g}{\mathrm{d}RSS}\right|\cdot P(d)/P(\tilde{d})\\&=\frac{1}{\sigma_N(d)\sqrt{2\pi}}\mathrm{e}^{\frac{-(10\eta\log_{10}(d/\tilde{d}))^2}{2\sigma_N^2(d)}}\cdot\tilde{d}\cdot\frac{\ln 10}{10\eta}\cdot P(d)/P(\tilde{d})\end{aligned}\tag{6}$$

其中 $d=\sqrt{(x-x_i)^2+(y-y_i)^2}$,$P(d)$表示与信标节点坐标$(x_i,y_i)$的距离为 d 的先验概率,$\sigma_N(d)=a\mathrm{e}^{-\frac{(d-d_0)^2}{b^2}}$为标准差拟合函数.对 $P(d)$的计算,采用类似文献[24]中的做法,$P(d)=\dfrac{m(d)}{\int m(d)\mathrm{d}d}$,$m(d)$近似表示为 $m(d)\approx\begin{cases}2\pi\cdot d, & d\le R\\0, & R<d\end{cases}$,$R$ 表示通信最大尺寸,假设节点通信范围大于 R,这样做的依据是让 $P(d)$表示落在半径为 d 的环形区域占整个区域的比例.

假设总共有 n 个信标节点参与定位,第 i 个信标节点发送过来的定位参考信息用三元组$\langle x_i,y_i,\tilde{d}_i\rangle$表示,其中$(x_i,y_i)$是表示第 i 个信标节点的位置坐标,$\tilde{d}_i$表示待定位节点与第 i 个信标节点的测量距离.由于有 n 个信标节点参与定位,相当于 n 个随机事件独立发生,利用与文献[10,24]类似的思想,依据概率论中的乘法法则和贝叶斯法则,在待定位节点到各信标节点的测量距离为 $\tilde{d}_i$的条件下实际距离为 d_i的条件概率$(i=1,\cdots,n)$最大,即满足

$$\begin{aligned}&(x^*,y^*)=\arg\max_{x,y}\{PM(x,y)\},\\&PM(x,y)=\prod_{i=1}^{n}PM_i(x,y)=\prod_{i=1}^{n}P(d_i|\tilde{d}_i)\\&=\prod_{i=1}^{n}P(\tilde{d}_i|d_i)P(d_i)/P(\tilde{d}_i)\\&=\prod_{i=1}^{n}\frac{1}{\sigma_N(d_i)\sqrt{2\pi}}\mathrm{e}^{\frac{-(10\eta\log_{10}(d_i/\tilde{d}_i))^2}{2\sigma_N^2(d_i)}}\tilde{d}_i\frac{\ln 10}{10\eta}P(d_i)/P(\tilde{d}_i)\end{aligned}\tag{6'}$$

其中,$d_i=\sqrt{(x-x_i)^2+(y-y_i)^2}$,表示当前位置点$(x,y)$距离第 i 个信标节点的距离为 d_i,$\sigma_N(d_i)=a\mathrm{e}^{-\frac{(d_i-d_0)^2}{b^2}}$表示距离第 i 个信标节点 d_i 长度的地方接收到这个信标节点 RSS 信号的标准差,$P(d_i)\propto 2\pi\cdot d_i$表示处于距离第 i 个信标节点长度为 d_i 位置的先验概率.其余的量中,n 表示信标节点个数,$\tilde{d}_i$表示测量估计距离,$P(\tilde{d}_i)$表示测量估计距离为 $\tilde{d}_i$的先验概率,η是路径损失因子(即衰减因子),这些量都是与变量(x,y)无关的常数.为方便引用,记式(6′)为模型(6′).模型(6′)表示要定位到的位置点能使得在测量得到与各个信标节点距离为$(\tilde{d}_1,\tilde{d}_2,\cdots,\tilde{d}_n)$条件下该点的条件概率取最大值.

3　传感器网络中基于变方差特征的安全定位概率模型和求解算法

3.1　安全定位概率模型

假设攻击者可以改变三元组$\langle x_i,y_i,\tilde{d}_i\rangle$中的任何值,可以恶意地声称一个错误的位置$\langle x_i,y_i\rangle$,也可以恶意地调整信号值改变测量距离 $\tilde{d}_i$.发起的攻击可以是单个恶意节点的攻击,也可以是多个恶意节点彼此独立的非共谋攻击,还可以是多个恶意信标节点协同合作的共谋攻击.在没有攻击的安全情况下,上节概率连乘的最大似然估计定位模型(6′)是有效的.相当于每个信标节点都为定位提供一个条件概率值,总概率为这些条件概率的乘积.n 个概率事件相互独立发生,依据概率论中的乘法定理,要求概率连乘值最大.但一旦有恶意攻击存在,概率连乘的模型(6′)注定会失败.原因很简单,安全环境下会使每个条件概率 $PM_k(x,y)$在正确位置定位点都取一个大概率值;而在恶意节点攻击环境下,由于恶意信标节点(比如信标节点 k)的存在,所提供的定位三元组$\langle x_k,y_k,\tilde{d}_k\rangle$参考信息有误,则在某个定位位置点,当定位概率值 $PM_k(x,y)$取值为大概率值时,其余正常信标节点提供的定位概率值 $PM_i(x,y)$在该点上取值几乎为 0;或者在正确定位位置点,正常信标节点提供的定位概率值 $PM_i(x,y)$取值为大概率值时,恶性信标节点 k 提供的定位概率值 $PM_k(x,y)$在该位置点上取值几乎为 0.无论哪种情况,由于有几乎为 0 的连乘因子存在,模型(6′)中最后相乘起来得到的结果也就几乎总为 0,定位也就失效了.为此,文献[18]提出可以先识别并过滤掉这样的恶性信标节点再进行定位,并设计了求最大一致信标集的恶意节点过滤算法 AR-MMSE(Attack-Resistant Minimum Mean Square-Estimate),但由于要准确求出最大一致信标集需要列举所有节点的组合情况而产生组合数爆炸,不得不使用了贪婪算法,并且这样只能获得近似解,不仅如此,还将定位过程机械地分成了滤除恶意节点和安全定位两个阶

段，增加了时耗. 本文认为，只要将定位模型(6′)中的概率值连乘 $PM(x,y)=\prod_{i=1}^{n}PM_i(x,y)$ 改成连加的形式 $PM(x,y)=\sum_{i=1}^{n}PM_i(x,y)$，恶性信标节点提供几乎为 0 的定位概率值 $PM_k(x,y)$ 不是作为乘法因子项出现，而是作为加法加数项出现，从而免除了对最终目标函数值的影响，完成安全定位，不但避免了由于组合数爆炸求近似解的缺点，而且滤除恶意节点和安全定位可以同时完成，减少耗时.

将式(6′)中概率值连乘 $PM(x,y)=\prod_{i=1}^{n}PM_i(x,y)$ 改成连加形式 $PM(x,y)=\sum_{i=1}^{n}PM_i(x,y)$ 的可行性还有一个依据，就是基于网格划分法求解的投票法的思路. 投票法是文献[18-19]提出的一种典型的恶意节点过滤方法. 图 2 是文献[18]中投票法示意图，将定位平面区域划分成网格，每个信标节点 i 和测量距离 $\tilde{d}_i$ 对应一个圆环，对每个圆环经过的网格投票计数加 1，得票最多的网格中心为定位位置点. 这样即使存在恶意节点的攻击，只要大多数信标节点为正常节点，正确定位的网格得到的投票最多，实现安全定位. 文献[25]证明了当恶意节点的个数大于等于 $(n-2)/2$ 时，若不采取特殊措施，则没有算法可以绝对保证安全的定位到指定精度. 投票法实质上是对空间网格进行加法投票计数，以此达到过滤恶意节点的目的，思路是连加计数. 因此，将模型(6′)中的概率值连乘 $PM(x,y)=\prod_{i=1}^{n}PM_i(x,y)$ 改成连加的形式 $PM(x,y)=\sum_{i=1}^{n}PM_i(x,y)$ 是可行的.

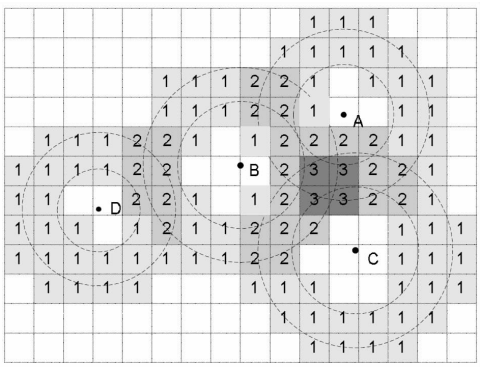

图 2 基于网格划分的投票定位方法

综合以上分析，在模型(6′)中去掉与坐标变量 (x,y) 无关的常数项 $\tilde{d}$，$P(\tilde{d})$ 和 $\frac{\ln 10}{10\eta}$ 等，得到能过滤恶意节点完成安全定位的新模型为

$$(x^*,y^*)=\arg\max_{x,y}\{PM(x,y)\},$$

$$PM(x,y)=\sum_{i=1}^{n}PM_i(x,y)=\sum_{i=1}^{n}\frac{1}{\sigma_N(d_i)}e^{\frac{-(10\eta\log_{10}(d_i/\tilde{d}_i))^2}{2\sigma_N^2(d_i)}}P(d_i) \quad (7)$$

其中 $\sigma_N(d_i)=a\,e^{-\frac{(d_i-d_0)^2}{b^2}}$，$P(d_i)\propto 2\pi\cdot d_i$，$d_i=\sqrt{(x-x_i)^2+(y-y_i)^2}$，待优化的自变量为位置坐标变量 (x,y). 为方便引用，记式(7)为模型(7).

3.2 求解算法

接下来的问题是如何求解模型式(7)的最优解. 这是一个非线性全局最优化问题，由于变方差项的存在，目标函数是一个关于变量 x,y 高度非线性复杂的函数表达式，x,y 的偏导数项非常复杂，用数学中传统的基于导数的优化方法，如最速下降法、牛顿法、拟牛顿法、共轭梯度法等求解. 用网格划分法来求解，其精度要受网格划分大小的影响，文献[18-19]设计了一种循环迭代求精改进方法，在找到投票最多的单元格后，再以该单元格作为目标区域，进入下一轮投票，反复迭代直到满足定位精度要求，但这类方法计算量太大，且计算量明显与定位区域大小有关，区域越大网格数量也越多，耗时也越多. 网格法还有其它的变种形式，如模糊贴近度[26]等，也都有类似的缺点. 因此必须设计一种求解算法，既能避免传统优化方法的缺陷，又能避开网格法中的缺陷. 进化算法作为一种随机搜索的自然仿生算法，求解时不需要函数导数信息，目标函数表达式复杂性对求解的影响相对较低，在传感器网络上的应用越来越受到重视[27-28]. 文献[29-31]在无线传感器定位问题中使用了智能优化算法，但仅仅是对目标函数为最小误差平方和 $\sum_{j\in H(i)}(\tilde{d}_{ij}-d_{ij})^2$ 的形式，由于节点分布的限制还需要加上修正步骤，这是比较麻烦的. 目前对目标函数为概率形式使用智能优化算法的文献还很少，关于变方差形式的讨论更少. 针对模型(7)中目标函数关于变量 x,y 高度非线性、难于求解的问题，结合无线传感器通信特征，本文运用智能优化中的进化计算理论设计了模型(7)的求解算法，具体算法如下.

求解算法 1.

第 1 步. 种群初始化：

在定位平面上随机产生 μ 个点作为初始解 $P(0)=$

$\{P^1(0),P^2(0),\cdots,P^\mu(0)\}$，其中 $P^j(0)=(x^j(0),y^j(0))^{\mathrm{T}}$ 表示平面上的第 j 个点坐标.

对任意的初始个体 $P^j(0),j=1,\cdots,\mu$，对每个信标节点 $A_q,q=1,\cdots,n$，分别在连线 $\overline{P^j(0)A_q}$ 方向上运用传统数学方法求出单自变量函数

$$PM_q(x,y)=\frac{1}{\sigma_N(d_q)}\mathrm{e}^{\frac{-(10\eta\log_{10}(d_q/\hat{d}_q))^2}{2\sigma_N^2(d_q)}}P(d_q)$$

的函数最大值点（由于 $d_q=\sqrt{(x-x_q)^2+(y-y_q)^2}$，函数 $PM_q(x,y)$ 中的两个自变量 x,y 都在 d_q 中，因此可以将函数 $PM_q(x,y)$ 看作是 d_q 的单自变量函数），令 $k=1$，记找到 $PM_q(x,y)$ 最大值的点为 $L^{qj}(k)$，比如在图 3 中的个体 $P^1(1)$ 对应信标点 A_1 而言找到的最大值点为 $L^{11}(1)$.

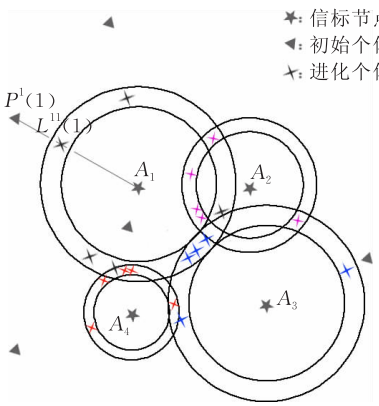

图 3　初始群体和局部搜索示意图

分析：这里没有使用 $PM(x,y)=\sum_{i=1}^{n}PM_i(x,y)$ 作为目标函数对变量 x,y 求导，而是使用单个的 $PM_i(x,y)$ 对 d_i 求导，可以避开对函数 $PM(x,y)$ 的复杂处理，这样简单而有效的方法很多，比如梯度法、插值法. 对个体 $P^j(0)$ 和函数 $PM_q(x,y)$ 而言，同一信标节点 A_q，函数值 $PM_q(x,y)$ 只与到圆心 A_q 的距离有关，对不同的个体 $P^j(0)$ 对应的最优点 $L^{qj}(k)$（其中 $k=1$）一定分布在以 A_q 为圆心的某个圆上，如图 3 所示. 若总共有 n 个信标节点，则会对应有 n 个圆，每个圆上分布了 μ 个局部搜索的点 $L^{qj}(1),j=1,\cdots,\mu$，初始化的结果是在每个圆上都初始化分布了一个规模为 μ 的子群体.

第 2 步（第 1 阶段进化）. 对各个圆心为 A_q 的圆上的子群体，令 $k=1$，使用 $PM(x,y)=\sum_{i=1}^{n}PM_i(x,y)$ 作为适应度函数，做第 1 阶段的交叉、变异和选择：

半角交叉. 在子群体中任意取两个点 $L^{qj_1}(k)$，$L^{qj_2}(k)$，用平分圆心角 $\theta=\angle L^{qj_1}(k)A_qL^{qj_2}(k)$ 的方式产生两个子代 $L_c^{qo_1}(k)$，$L_c^{qo_2}(k)$，如图 4 所示. 如果用 $x^{qo_1}(k)$ 和 $y^{qo_1}(k)$ 表示子代 $L_c^{qo_1}(k)$ 的坐标，则具体计算表达式为

$$x^{qo_1}(k)=x^{A_q}(k)+|\overline{L^{qj_2}(k)A_q}|\cdot\cos(\theta_{qj_2}+\theta/2),$$

$$y^{qo_1}(k)=y^{A_q}(k)+|\overline{L^{qj_2}(k)A_q}|\cdot\sin(\theta_{qj_2}+\theta/2),$$

其中 $x^{A_q}(k)$ 和 $y^{A_q}(k)$ 表示点 A_q 的坐标，$|\overline{L^{qj_2}(k)A_q}|$ 表示

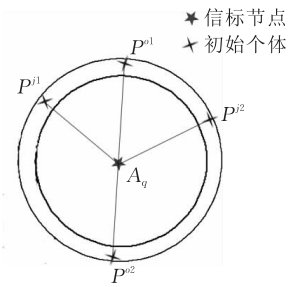

图 4　半角交叉

连线 $\overline{L^{qj_2}(k)A_q}$ 的长度，θ_{qj_2} 为连线 $\overline{L^{qj_2}(k)A_q}$ 的水平夹角，θ 表示圆心角 $\angle L^{qj_1}(k)A_qL^{qj_2}(k)$，由于 3 个点 $L^{qj_1}(k)$，A_q，$L^{qj_2}(k)$ 的坐标都已知，因此 θ_{qj_2} 和 θ 可以分别通过反正切函数和反余弦函数求出. 类似方法可以得到另一个子代 $L_c^{qo_2}(k)$ 的坐标 $x^{qo_2}(k),y^{qo_2}(k)$ 的表达式.

均匀变异. 设定变异概率为 P_m，对经过以上交叉后任一子代个体 $L_c^{qoj}(k)$ 点，绕圆心 A_q 随机转动一个角度 θ_{rnd} 得到变异后的子代个体 $L_m^{qoj}(k)$ 点，其中 θ_{rnd} 表示 $[0,2\pi)$ 之间的一个均匀分布的随机数.

选择. 在所有原子群体个体、经过交叉后所有个体和经过变异后所有个体组成的集合 $\{L^{qj}(k)\mid j=1,\cdots,\mu\}\cup\{L_c^{qoj}(k)\mid j=1,\cdots,\mu\}\cup\{L_m^{qoj}(k)\mid j=1,\cdots,\mu\}$ 中，以 $PM(x,y)=\sum_{i=1}^{n}PM_i(x,y)$ 作为适应度函数，令 $k=k+1$，选择最好的 μ 个个体作为下一代子群体 $\{L^{qj}(k)\mid j=1,\cdots,\mu\}$，若 k 达到事先给定的进化代数 G_1，则进入第 3 步；否则转入第 2 步开始的半角交叉继续进化.

分析：这一步中的进化实际上是在 n 个不同的子群体中独立进行的进化，每个限制于各个圆上的子群体进化方式（交叉、变异、选择）都相同，使用的适应度函数形式也相同，进化结果是各子群体的优良个体分布在 n 个圆交叉区域附近.

第 3 步（第 2 阶段进化）. 以 $PM(x,y)=\sum_{i=1}^{n}PM_i(x,y)$ 作为适应度函数，在第 2 步进化得到 n 个子群体的所有个体中（总共有 $n\cdot\mu$ 个）选择最好的 μ 个个体作为初始群体，记为 $P^j(t),j=1,\cdots,\mu$，令 $t=0$，做第 2 阶段的交叉、变异和选择：

随机凸组合交叉. 在当前群体中随机选取 3 个个体 $P^{i_1}(t)$，$P^{i_2}(t)$，$P^{i_3}(t)$，计算 $P^i(t)=\sum_{j=1}^{3}\theta_{i_j}\cdot P^{i_j}(t)$ 得到一个新个体，其中 θ_{i_j} 为元素在 0 和 1 之间的随机数，并且满足凸组合的约束条件 $1=\sum_{j=1}^{3}\theta_{i_j}$. 采用这样的交叉方式是由于圆的交叉区域可以是一个凸区域.

两重高斯变异. 设定变异概率为 P_m，对上面的个体 $P^i(t)$ 做变换得到 $Q^i(t)=P^i(t)+\Delta P$，其中 $\Delta P\sim N(0,\sigma_k^2)=(N(0,\sigma_{k_1}^2),N(0,\sigma_{k_2}^2))$ 表示二维高斯标准正态分布的扰动，这样的变异方式相当于对个体 $P^i(t)$ 的两个位置分量 $P^i(t)=$

$(x^i(t),y^i(t))$做高斯变异后得到新个体 $Q^i(t)$，这样的变异方式符合传感器 *RSS* 信道噪声服从高斯分布的特点. 扰动项中的方差 $\sigma_{k_1}^2$，$\sigma_{k_2}^2$ 为事先设定的一个常数.

选择. 以 $PM(x,y)=\sum_{i=1}^{n}PM_i(x,y)$作为适应度函数，从$\{P^j(t)\mid j=1,\cdots,\mu\}\cup\{Q^j(t)\mid j=1,\cdots,\mu\}$中选 μ 个最好的位置点个体作为下一代群体 $P(t+1)$. 进化到一定代数 G_2 后算法停止，输出最优个体 $P^*(x,y)=(x^*,y^*)$作为定位结果；否则转入第 3 步中的交叉算子步骤继续进化.

分析：第 3 步开始进化前的群体是从第 2 步进化结束后得到的 $n\cdot\mu$ 个个体中，以 $PM(x,y)=\sum_{i=1}^{n}PM_i(x,y)$作为适应度函数，选择最好的 μ 个个体组成的. 由于恶意信标节点对应的圆是和正常信标节点对应的圆的重叠区域是分离的(如图 3 所示)，这样可以去除恶意信标节点对应的子群体个体的影响，此后进行的随机凸组合交叉和两重高斯变异的进化就是在正常信标节点圆上的个体中进行的，这一步不但能完成精细求解，还屏蔽掉了恶意节点信息，保证了定位的安全性. 和其它一些先求最大一致信标集来过滤恶意节点再进行定位的安全定位算法(如 Attack-Resistant Minimum Mean Square-Estimate，AR-MMSE[18-19])相比，这一步可以让算法在安全定位过程中过滤掉恶意节点信息.

4 算法收敛性证明和时间复杂度分析

4.1 算法收敛性

求解算法 1 是一种随机搜索算法，主要由两个进化阶段组成. 如果能证明两个阶段收敛，那么由收敛的充分条件可知求解算法 1 是收敛的. 为此先明确数学中随机序列以概率收敛于最优解的确切定义，再证明求解算法 1 在数学上以概率收敛于最优解.

为了数学上表述方便，先给出如下的记号 $X^*=\{x^*\in S\mid f(x^*)=\max f(x),x\in S\}$，表示全局最优解对应的解；对 $\forall\varepsilon>0$，$M_\varepsilon=\{x\in S\mid f(x^*)-f(x)\leq\varepsilon,x^*\in X^*\}$，用 $m(M_\varepsilon)$ 表示该 ε 领域集合 M_ε 的测度，对欧式平面测度空间来讲，$m(M_\varepsilon)$ 表示集合 M_ε 的测度.

定义 1. 用 $x^*\in\Omega$(Ω 表示搜索可行域)表示目标函数 $f(x)$的最优解，若有 $Prob\{\lim_{t\to\infty}x^*\in P(t)\}=1$ 成立，则称该进化算法以概率 1 收敛于全局最优解；若有 $\exists x\in P(t)$，$\forall\varepsilon>0$，$Prob\{\lim_{t\to\infty}x\in M_\varepsilon\}=1$ 成立，则称该进化算法以概率 1 在 ε-精度上收敛于全局最优解.

显然，如果 $\varepsilon\to0$，两种以概率 1 收敛的定义是一致的.

定义 2. 对进化群体中的两个个体 a 和 b，如果个体 a 经过交叉、变异后进化成子代个体 b 的概率满足 $Prob\{MC(a)=b\}>0$，则称个体 b 可由个体 a 可达；若 $Prob\{MC(a)=b\in M_\varepsilon\}>0$，则称个体 b 可由个体 a 在 ε-精度可达.

显然，如果 $\varepsilon\to0$，两种可达的定义是一致的.

Bäck[13] 和 Rudolph[14] 已经证明了若设计的进化算法满足以下假设，则该进化算法以概率 1(或 ε-精度)收敛于全局最优解，且与初始群体分布无关.

假设 1. 可行域空间中的任意两个个体 a 和 b，个体 b 可由个体 a(ε-精度)可达.

假设 2. 对于群体序列 $P(0),P(1),\cdots,P(k),\cdots$单调，即对 $\forall k$，满足

$$\max\{f(x)\mid x\in P(k+1)\}\geq\max\{f(x)\mid x\in P(k)\}.$$

基于以上定义和结论，对上一节提出的进化求解算法有以下结论成立.

定理 1. 对求解算法 1 第 2 步中处于可行域为圆 A_q 上的群体，按照算法中的半角交叉、平均变异和选择策略进行进化，以概率 1 收敛于可行域圆 A_q 上的全局最优解，而且与初始群体分布无关.

证明. 算法第 2 步中适应度函数为$PM(x,y)=\sum_{i=1}^{n}PM_i(x,y)$，可行域 Ω 为以 A_q 为圆心的圆，由于适应度函数在可行域上连续，可确定上下界、最优解集不为空集. 任意两个个体 a,b，如果用 c 表示从父代 a 通过半角交叉得到的任一子代个体，发生变异的概率为 $p_m>0$，则个体 a 到达个体 b 的概率可以表示为 $Prob\{MC(a)=b\}=p_m\cdot Prob\{M(c)=b\}$，若 $Prob\{M(c)=b\}>0$ 成立，则说明个体 a,b 是可达的.

而可行域 Ω 为以 A_q 为圆心的圆，记为 C_q，由于采用的是$[0,2\pi)$均匀变异，$\forall c\in C_q$变异到另一个体 b 的概率为 $Prob\{M(c)=b\}=1/2\pi>0$，所以假设 1 成立.

又由于采用保留 μ 个最好个体作为下一代群体的选择策略，每次进化的当前代最好个体肯定比上代要好，满足单调要求，因此假设 2 也成立. 从而定理得证. 证毕.

定理 2. 对求解算法 1 第 3 步中按照算法中的随机凸组合交叉、高斯变异和选择策略进化的群体，以概率 1 在 ε-精度收敛于全局最优解.

证明. 适应度函数为$PM(x,y)=\sum_{i=1}^{n}PM_i(x,y)$，

可行域 Ω 为传感器有限传输区域，由于适应度函数在可行域上连续，可确定上下界、最优解集不为空集. 先证明假设 1 成立. 对任意的两个个体 a,b，用 c 表示从父代 a 通过随机凸组合交叉得到的任一子代个体，再以一定变异概率变异到 b.

记个体 c 的坐标分量为 x_1,x_2，变异成个体 b 的分量为 x_1',x_2'，由于采用的是分量相互独立的高斯变异，$x_i'=x_i+\eta_i,i=1,2$. 其中 $\eta_i\sim N(0,\sigma_t^2)$，$\eta=x'-x$，所以 η_i 的密度函数为 $p(y_i)=1/(\sqrt{2\pi}\sigma_t)\cdot\exp(-y_i^2/(2\sigma_t^2))$，$\eta$ 的密度函数为 $p(y)=1/(\sqrt{2\pi}\sigma_t)^2\cdot\exp(-(y_1^2+y_2^2)/(2\sigma_t^2))$，由此得 $Prob\{M(c)=b\in M_\varepsilon\}=\int_{M_\varepsilon'}p(y)\mathrm{d}y=\int_{M_\varepsilon'}1/(\sqrt{2\pi}\sigma_t)^2\cdot\exp(-(y_1^2+y_2^2)/(2\sigma_t^2))\mathrm{d}y$，其中 $M_\varepsilon'=\{y|x'-x,x'\in M_\varepsilon\}$，由 y_i 有界知 $y_i\leq\theta$ 为常数，上式有

$$Prob\{M(c)=b\in M_\varepsilon\}$$
$$\geq\int_{M_\varepsilon'}1/(\sqrt{2\pi}\sigma_t)^2\cdot\exp(-(2\theta^2)/(2\sigma_t^2))\mathrm{d}y,$$
$$=1/(\sqrt{2\pi}\sigma_t)^2\cdot\exp(-(\theta^2)/(\sigma_t^2))\cdot m(M_\varepsilon')>0,$$

因此假设 1 满足.

又由于采用保留 μ 个最好个体作为下一代群体的选择策略，每次进化的当前代最好个体肯定比上代要好，满足单调要求，假设 2 也成立. 从而定理得证.　　证毕.

定理 3.　求解算法 1 以概率 1 在 ε-精度收敛于全局最优解，ε 为任一小的数.

证明.　PMEA 进化算法由 3 个阶段组成. 第 1 阶段随机生成群体后，由于使用确定的局部搜索算法，可以保证以概率 1 收敛于第一阶段的最优解；由定理 1 知第 2 阶段以概率 1 收敛于该阶段可行域上的最优解，由定理 2 知第 3 阶段以概率 1 在 ε-精度收敛于整个定位区域的全局最优解，依据收敛的充分条件性质可判断算法以概率 1 在 ε-精度收敛于整个定位区域的全局最优解.　　证毕.

4.2　时间复杂度分析

对于进化算法的收敛速率分析，目前还没有通用的数学分析方法[32]，多数还是通过实验进行分析，这里仅仅对算法时间性能做简单分析与比较，直接的实验及分析在后面进行.

算法分 3 个步骤，第 1 步初始化所花时间与信标节点个数 n、种群规模大小 μ 成正比，时间复杂度为 $O(\mu\cdot n)$；对第 2 步子群体的进化，假设一次半角交叉、一次均匀变异和一次选择操作的时间为 T_1，则第 2 步所花时间为 $O(n\cdot G_1\cdot T_1)$，其中 G_1 为该阶段的最大进化代数；对第 3 步群体的进化，假设一次随机凸组合交叉、一次高斯变异和一次选择操作的时间为 T_2，则第 3 步所花时间为 $O(G_2\cdot T_2)$，G_2 为该阶段的最大进化代数. 整个算法时间复杂度就为 $O(\mu\cdot n)+O(n\cdot G_1\cdot T_1)+O(G_2\cdot T_2)$. 严格讲，这样简单的分析是不够严谨的，比如关于 T_1,T_2 的假设. 在目前对进化算法没有通用的数学分析方法[32]情况下，只能通过下节的实验进行验证.

再来对比分析模型(7)及设计的求解进化算法与基于网格划分的投票法[18]、基于最小均方差的最大一致信标集(Attack-Resistant Minimum Mean Square-Estimate，AR-MMSE)过滤算法[19]的时间性能. 对于基于网格划分的投票法运行时间依赖于网格数目的多少，在固定网格尺寸(即定位精度)时，场地尺寸越大，网格数目也越多，运行时间也越长；对于文献[19]中基于最小均方差的最大一致信标集过滤算法，为避免组合数爆炸使用贪婪算法求解最大一致信标集，运行时间与恶意节点数量有关，文献[19]给出分析并给出其时间复杂度为 $1+\binom{n}{m+1}+\cdots+\binom{n}{n}$，其中 m 为恶意节点的数目. 而从前面关于模型(7)及求解进化算法的简单分析已经可以看出本文的安全定位方法与场地区域大小无关，与恶意攻击节点的数目也无关(当然要正确定位，必须保证正常信标节点个数占多数). 这是相比投票法和最大一致信标集过滤算法好的方面.

5　实验及分析

本节对安全定位模型(7)和求解算法 1 在定位准确性、定位时间性能和安全性方面做了测试实验，包括在公开数据集上的仿真实验和在实验平台上的实际实验，以保证实验结果的真实性和正确性，方便检验.

5.1　实验数据集和安全环境下定位精度测试

先确定仿真实验方式. 和很多用 MatLab 软件进行定位仿真的文献不同，本文的变方差概率特征模型更加接近实际信道模型，再用 MatLab 软件仿真不具备说服力，较为合理的是对实际的 RSS 数据集进行仿真测试. 为此采用了文献[15,33]中提供的公开数据集. 该数据集由佛罗里达摩托罗拉通信实验室发布，传感器由摩托罗拉公司提供. 实验环境为室内办公室(如图 5)，任何一个节点都在所有其余节点的通信范围内. 该数据集总共采集了 44 个传感器相互之

间的 *TOA* 和 *RSS* 两种测试数据，本文只需要 *RSS* 数据. 为对比定位效果，本文采用和文献[15,33]中一样的实验方式，选择 3、10、35、44 号节点作为信标节点对其余节点进行定位. 由普通节点与这 4 个信标节点的 *RSS* 信号值计算得到该普通节点到这 4 个信标点的估计距离，再利用这些估计距离对其余 40 个普通节点定位. 设对第 i 个普通节点定位得到的最优位置为 (x_i^*, y_i^*)，计算其与真实位置 (x_i, y_i) 之间的偏离距离 $err_i = \sqrt{(x_i^* - x_i)^2 + (y_i^* - y_i)^2}$ 作为第 i 个节点的绝对估计误差，总共有 40 个普通节点，将平均误差 $err = 1/40 \cdot \sum_i err_i$ 作为衡量算法精度的指标. 文献[15,33]中用克美罗-雷界估计 *RSS* 平均定位误差为 2.18 m. 很多文献给出的实际定位测量误差结果都在 2 m 以上. 本文用 C 程序实现了安全定位模型(7)和求解算法 1，经过多次测试得到合适的群体规模参数值为 $\mu = 30$，变异参数 $\Delta P \sim N(0, \sigma_k^2) = (N(0, \sigma_{k_1}^2), N(0, \sigma_{k_2}^2))$ 中的 $\sigma_{k_1}^2 = 5^2$，$\sigma_{k_2}^2 = 5^2$；算法进化代数为 $G_1 = 100$，$G_2 = 100$，传感器参数取公开数据集文献[15,33]提供的值，通信半径 $R = 80$，路径损失因子 $\eta = 2.3$.

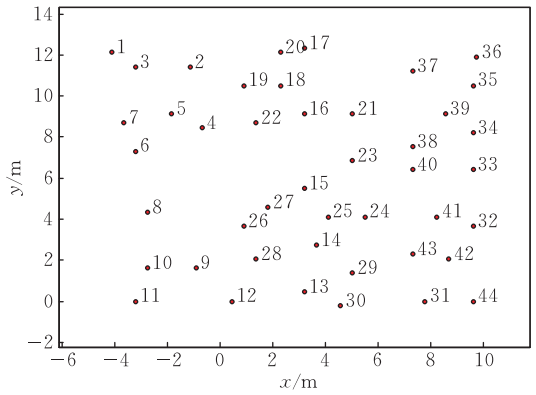

图 5 数据集文献[15,19,33]中的室内节点分布

确定公开数据集和仿真实验方式后，下面先在公开数据集上检验 *RSS* 信号标准差随距离高斯函数变化的概率特征是否成立，安全定位模型(7)和求解算法 1 是否可以提高定位精度，以证实多数文献中假定标准差 $\sigma_N(d)$ 为常数时结果不够理想的判断. 图 6 是取标准差 $\sigma_N(d) = C$ 为常数(C 分别取值为 0.5,0.9,1.0,1.5,2.5,4.0,6.0)时，用求解算法 1 对公开数据集[15,33]进行定位仿真的平均误差 *err* 随不同常数 C 变化的折线图. 可以看出常数 C 取 0.9 时，达到 2.67 m 最小平均误差(仍然比文献[15,33]的 2.18 m 大). 再考虑 *RSS* 标准差 $\sigma_N(d)$ 不为常数的情况. 对公开数据集中的 *RSS* 信号值标准差 $\sigma_N(d) = a\,e^{-\frac{(d-d_0)^2}{b^2}}$ 参数 a, b, d_0 可以通过以下实验步骤找到合适取值. 先固定某个 d_0 值，然后给 a, b 取不同值，运行算法程序，观察实验结果. 比如固定 $d_0 = 8.0$，a 从 0.8 开始取值(a 其实表示标准差 $\sigma_N(d) = a\,e^{-\frac{(d-d_0)^2}{b^2}}$ 的最大值)，引入 b' 满足 $b = d_0/sqrt(\ln(a/b'))$，不同的 b' 对应了不同的 b 值，b' 表示是 $\sigma_N(d) = a\,e^{-\frac{(d-d_0)^2}{b^2}}$ 曲线与纵轴的截距(理论上是距离 d 为零的标准差，当然实际中 $d = 0$ 是不存在的，所以只是理论值)，引入 b' 是由于 b' 比 b 更加直观. 让 b' 从 0.1 开始取值. 再固定某个 b' 值，变化 a, d_0，观察实验结果，然后再固定 d_0，变化 a, b'，重复这样的步骤. 通过多次的实验，查找平均误差最小时参数 a, b', d_0 分别的取值. 最后得到在参数 a, b', d_0 分别取值 0.9,0.3,8.0 时平均误差最小，为 1.86 m. 这不但比方差为常数的最小平均误差 2.67 m 小，而且比文献[15,33]用克美罗-雷界得到的 2.18 m 结果还要好，达到了 2 m 以下的平均误差. 说明多数文献中假定标准差 $\sigma_N(d)$ 为常数确实是不够理想的.

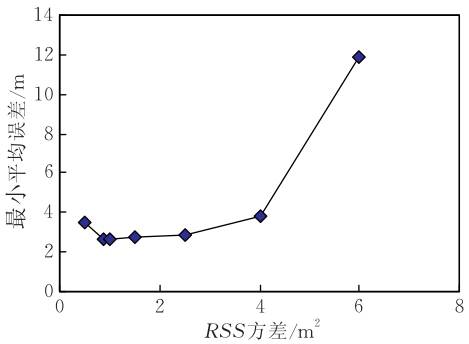

图 6 不同方差常数值对应的平均误差

再在公开数据集上[15,33]进行仿真实验以比较安全定位模型(7)和最大似然估计的概率连乘定位模型(6′)的定位精度. 采用算法 1 求解模型(6′)，进行同样实验步骤，结果显示还是在参数 a, b', d_0 取值为 0.9,0.3,8.0 时，概率连乘定位模型(6′)达到最小的估计误差 1.87 m. 说明安全定位模型(7)与最大似然估计的概率连乘定位模型(6′)相比，(7)并没有影响(6′)的定位精度.

以上仿真结果说明 *RSS* 信号标准差随距离高斯函数变化的概率特征是成立的，概率连加的安全定位模型(7)确实可以达到很好的定位精度.

以下再通过实际实验测试安全定位模型(7)和求解算法 1 的定位精度. 为此，在感知 RF2-V210 传感器网络实验教学平台上实现了模型(7)和

求解算法 1. 以第 2 节图 1 中采集的实验数据为例，通过如下步骤确定标准差 $\sigma_N(d)=a\,e^{-\frac{(d-d_0)^2}{b^2}}$ 中参数 a,b,d_0 的取值. 对第 2 节中采集到的实验数据(图 1 中数据采集方式为：在室内办公室环境下保持 15 dB 节点发射功率，每改变一次两个节点之间的距离，采集 1000 次 *RSS* 信号值并统计出标准差)的标准差取自然对数后进行基于最小二乘法的二次函数曲线拟合，得到的拟合曲线及拟合公式系数如图 7 所示. 曲线拟合效果可以用样本决定系数 r^2 表示，该值介于 0 和 1 之间，r^2 越大，说明曲线和样本数据拟合程度越好. 图 7 中 r^2 值为 0.9486，很接近 1，说明拟合程度是不错的. 由于以上拟合是对标准差的对数值进行的，对拟合公式中的系数进行指数变换就可以得到标准差公式 $\sigma_N(d_i)=a\,e^{-\frac{(d_i-d_0)^2}{b^2}}$ 中的 a,b,d_0 3 个参数的取值，分别为 1.05，4.75，6.30. 然后再代入模型(7)在室内环境中进行定位精度的测试实验，测试方案设计如下：在大小为 5 m×11 m 室内办公室处于同一高度的 4 个角上放置 4 个信标节点，在办公室中心位置放置一个待定位的普通节点，采集接收到的 4 个信标节点发送过来的 *RSS* 值，在普通节点上嵌入模型(7)和求解算法 1 的 C 语言代码. 群体规模参数 μ 取 30，为保证进化到足够精度，G_1 与 G_2 都取 100，信标节点个数 $n=4$，多次运行结果显示中心位置节点的定位位置仅仅偏离实际位置 0.63 m. 说明在没有攻击时定位模型(7)和求解算法的定位精度是不错的. 以上进行的数据采集、参数训练和定位测试都是在室内常温环境下进行. 图 8 表示在室内环境下重复以上三次数据训练得到的拟合曲线，三次拟合的结果都非常接近，三次拟合的样本决定系数 r^2 分别为 0.9486、0.9432 和 0.9477，都很接近 1. 说明在相同环境下反应标准差特性的 a,b,d_0 3 个参数值是可以保持稳定的.

标准差随距离的变化可以用 $\sigma_N(d_i)=a\,e^{-\frac{(d_i-d_0)^2}{b^2}}$ 表示，但 a,b,d_0 这 3 个参数值是和环境相关的①. 在不同环境下，即使是对同一类型号的传感器器件，在不同传播介质中的路径损耗不同，器件内部的热噪声干扰程度不同以及室内室外等环境下的反射等干扰程度不同，因此在不同环境(比如室内和室外、潮湿和干燥、高温和低温等)下 a,b,d_0 3 个参数值也会不同. 合理的做法是在某种环境下的具体定位应用前，先在相同环境中进行数据采集和训练数据的曲线拟合，得到反应标准差变化特性的 a,b,d_0 3 个参数的训练值，再在该环境下进行定位应用，以保证拟合结果的准确性和稳定性. 虽然麻烦了些，但从理论分析、仿真实验和实际测试的效果来看比方差为常数时的定位精度有了明显提高，在实际应用中也是可行的.

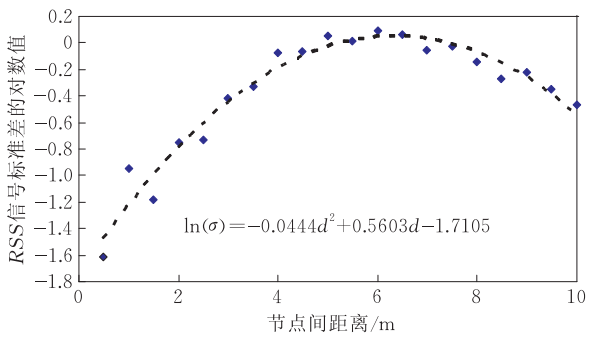

图 7　标准差自然对数值对节点距离的二次拟合曲线

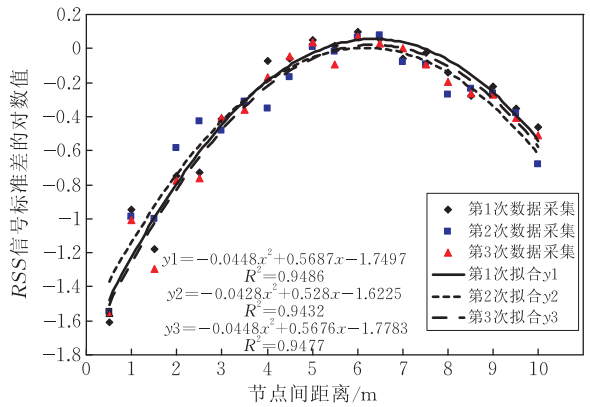

图 8　室内环境下标准差自然对数值随距离曲线拟合

5.2　安全环境下时间性能测试

先在公开数据集[15,33]上测试不同最大进化代数 G_1,G_2 取值对算法时间性能的影响. 算法 1 的第 2 步是一个加速收敛的子群体进化阶段，第 3 步相当于在所有子群体中选择最好的一部分个体再进行进化，这两个步骤的进化程度是通过最大进化代数 G_1,G_2 来控制的. 不同的最大进化代数 G_1,G_2 取值会对算法性能有直接的影响.

对公开数据集[15,33]，选择 3、10、35、44 号节点作为信标节点(即参考节点，总共 4 个，$n=4$)，单独测试 1 号节点定位的时间性能. 群体大小参数、变异参数和方差拟合参数 a,b',d_0 取值同前面实验. 每次进化找到群体中函数适应度值最好的个体位置点，

① 叶苗，王宇平. 基于变方差概率模型和进化计算的传感器网络定位算法. 软件学报，2012 年 6 月录用，其中比较详细地论述了对标准差变化关系式中的 a,b,d_0 3 个参数进行拟合训练的环境要与定位应用的环境相同或类似，以保证测量结果准确和稳定的观点. 本文的工作是在此基础上对安全性的进一步讨论.

计算该点偏离 38 号节点的误差. 图 9 表示第 2 步的 4 个子群体所有个体中的最好个体位置点对应的最小偏离误差与进化代数的变化关系,可以看出开始阶段进化效果比较明显,从第 40 代左右开始,进化比较缓慢. 图 10 表示第 2 步进化最大代数 G_1 取不同的值时(分别测试了 G_1 为 0,10,40,80,100 的情况),第 3 步进化群体的最好个体位置点对应的最小偏离误差与进化代数的变化关系. 最大进化代数 $G_1=0$,表示算法 1 直接跳过第 2 步进入第 3 步,从图 10 中可以看出收敛速度比较慢,差不多到第 90 代才稳定; $G_1=10$ 时发生类似的情况;而取 $G_1=40$,80 或 100 时,进化收敛速度明显加快,几乎在进化到第 30 代时就稳定了,特别是 $G_1=100$ 时,在进化到第 20 代时最小误差为 1.33m,此后稳定在这个值. 这说明适当 G_1 取值可以加速第 3 步的收敛速度. 假设第 2 步和第 3 步各自交叉、变异和选择所用时间复杂度相同,则可以用 G_1+G_2 作为时间复杂度的标准,从实验中可以看出 $G_1=40$, $G_2=50$ 比较合适.

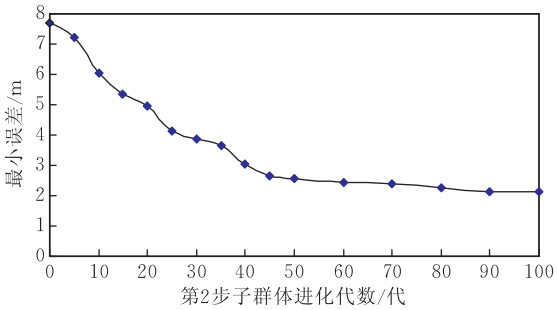

图 9 最小平均误差与第 2 步进化代数的关系图

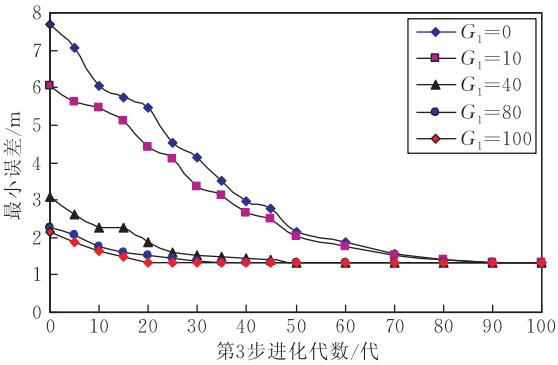

图 10 最小平均误差与第 3 步进化代数的关系图

再通过在实际平台上的程序运行时间来直接测试时间性能. 同 5.1 节实验,针对定位模型(7),在感知 RF2-V210 传感器网络实验教学平台上嵌入并运行求解算法 1 和网格划分法(不进行迭代),参数同 5.1 节实验中的取值,使用 API 时间函数统计出定位代码运行所花时间,得到图 11 结果. 可以看到网格划分比较粗时,网格划分法(不进行迭代)运行定位时间比求解算法 1 所花时间少,随着网格划分越来越细,网格边长越来越小,网格划分法花费时间也越多,已经明显超过了求解算法 1.

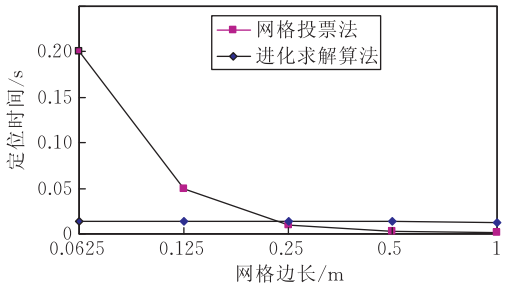

图 11 安全环境下定位的运行时间

5.3 攻击环境下定位精度测试

最后测试恶性信标节点攻击情况下安全定位概率模型(7)的准确性和安全性.

先在公开数据集[15,33]上建立如下 3 个测试实验场景. 第 1 个实验场景是模拟单个恶意信标节点提供偏离实际位置 e 长度的错误位置(直接改变数据集中三元组 $\langle x_i, y_i, \tilde{d}_i\rangle$ 中的估计距离 $\tilde{d}_i$,可以达到同样的效果). 具体设置为:选择 3、10、35、44 号节点作为信标节点,38 号节点为待定位的普通节点,任取一个信标节点(如指定 10 号节点)为恶意节点,在指向普通节点 38 号节点方向上偏离自身实际位置 e 长度. 第 2 个实验场景是模拟存在多个非共谋的恶意信标节点,每个恶意信标节点分别独立地声明一个偏离实际位置 e 长度的位置信息. 具体设置为:选择 3、10、11、31、35、36 和 44 号作为信标节点,38 号节点作为普通节点,11、31 和 36 号节点作为非共谋的恶意信标节点,分别在指向普通节点 38 号节点方向上偏离自身实际位置 e 长度. 第 3 个实验场景是存在多个共谋节点,提供一致的虚假定位信息,但是共谋节点个数不多于正常节点个数. 在这种情况中,恶意信标节点彼此协调好各自偏离距离 e 从而定位到一个错误位置. 具体设置为:选择 3、10、11、31、35、36 和 44 号作为信标节点,38 号节点作为普通节点,11、31 和 36 号节点作为共谋的恶意信标节点,每个恶意节点分别在垂直向上的方向上声明偏离自身实际位置 e 长度的位置,这样如果用 11、31 和 36 号信标节点定位就可以产生可以偏离定位位置 e 长度大小的合谋攻击. 显然,攻击力度随着偏离距离 e 的增加而增大,偏离距离 e 的大小可以代表攻击力度. 图 12 统计出以上 3 种场景中定位绝对误差随着恶意节点偏离实际位置 e 长度变化的情

况.从图 12 中看出在 3 种场景中,在 e 比较小的一个范围内,定位误差随着偏离距离 e 增加而增大,在 1 至 2m 时受攻击的定位精度略微差于安全环境下的定位精度,这是由于在偏离距离 e 比较小时,恶意信标节点和正常信标节点的区别并不明显,很难分辨出是恶意节点还是正常节点;随后不管攻击力度 e 如何加大,定位误差几乎不变,有时还略微好于安全环境下的定位精度(注意:纵轴是对数坐标),说明定位模型(7)和求解算法 1 确实能过滤掉恶意节点的影响完成安全定位.文献[34]可以解释恶意攻击环境下定位误差还略微好于安全环境下的定位精度的现象.文献[34]证明了当距离测量误差较大时增加信标节点的数量反而会降低定位精度的结论.从图 12 可以分析出,在安全环境下,3、10、35 号这 3 个信标节点的定位精度略好于安全环境下 3、10、11、31、35、36 和 44 号这 6 个信标节点的定位精度,这正好符合文献[34]的结论.文献[23]还给出了在多个信标节点中选择参与定位的原则.这都说明过多的信标节点参与定位会带来过多的冗余信息,反而会影响定位精度.因此在图 12 出现了恶意攻击环境下定位误差略微好于安全环境下的定位精度的现象.但值得注意的是如果定位冗余信息太少,在恶意攻击环境下的定位安全性又会显得很脆弱,因此一般要有折中.从图 12 中还可以看出模型(6′)不具备任何抵抗攻击的能力,在只有单个恶意节点攻击存在时就失去了应有的定位精度.本文在感知 RF2-V210 传感器网络实验教学平台上也建立了类似以上的实际场景,并进行了相应实验,得到的结果和图 12 类似,限于论文篇幅,不再列举.综合以上分析,本文提出的定位模型(7)和求解算法在恶意攻击环境下确实能过滤掉恶意节点的攻击信息,完成安全定位.

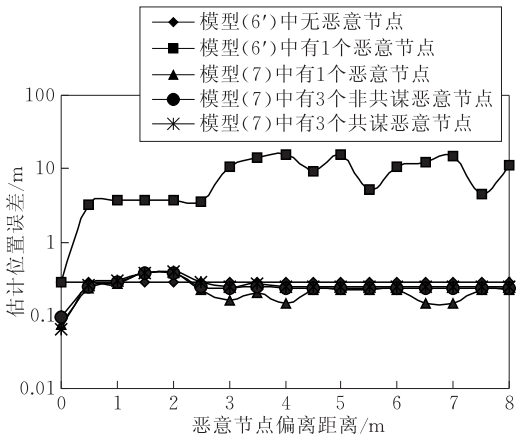

图 12　攻击环境下的定位精度

6　结　论

针对最大似然估计的定位法和提高安全的投票法,本文先建立更加符合实际情况的基于接收信号强度变方差特征的概率模型,在分析传统最大似然估计的定位方法的安全性缺陷后,采用投票法的思路,设计了变方差特征的传感器网络恶意节点过滤安全定位概率模型,考虑到模型目标函数中高度非线性不好求解的问题,运用进化计算理论设计出符合传感器通信特征的求解算法.只要是大多数的信标节点为正常节点,该定位模型和求解算法可以很好抵御共谋攻击,过滤掉恶意节点的定位参考信息.最后不仅在数学上证明了算法的收敛性,还在公开数据集的模拟和实际测试中验证了基于变方差特征的安全定位概率模型和求解算法能在保证定位精度的前提下过滤恶意节点信息,能避免投票法网格大小和网格数目影响求解时间性能的缺陷.相比投票过滤算法能避免投票法网格大小和网格数目影响求解时间性能的缺陷,相比求最大一致信标集的恶意节点过滤算法(如 Attack-Resistant Minimum Mean Square-Estimate,AR-MMSE)能避免由于组合数爆炸而不得不使用贪婪算法获得近似解的缺陷.

致　谢　在此,我们向对本文的工作给予支持和建议的同行,尤其是西安电子科技大学计算机学院王宇平教授领导的网络优化讨论班上的老师和同学表示感谢!

参　考　文　献

[1] Akyildiz I F, Su W, Sankarasubramaniam Y, Cayirci E. Wireless sensor networks: A survey. Computer Networks, 2002, 38(4): 393-422

[2] Karp B, Kung H T. GPSR: Greedy perimeter stateless routing for wireless networks//Proceedings of the 6th Annual International Conference on Mobile Computing and Network. New York, 2000: 243-354

[3] Wang Fu-Bao, Shi Long, Ren Feng-Yuan. Self-localization systems and algorithms for wireless sensor networks. Journal of Software, 2005, 16(5): 857-858(in Chinese)
(王福豹,史龙,任丰原. 无线传感器网络中的自身定位系统和算法. 软件学报, 2005, 16(5): 857-858)

[4] Girod L, Bychovskiy V, Elson J, Estrin D. Locating tiny sensors in time and space: A case study//Proceedings of the 2002 IEEE International Conference on Computer Design: VLSI in Computers and Processors. Freiburg, Germany, 2002: 214-219

[5] Harter A, Hopper A, Steggles P, Ward A, Webster P. The anatomy of a context-aware application//Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking. New York, USA, 1999: 59-68

[6] Girod L, Estrin D. Robust range estimation using acoustic and multimodal sensing//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Maui, USA, 2001: 1312-1320

[7] Nagpal R, Shrobe H, Bachrach J. Organizing a global coordinate system from local information on an ad hoc sensor network//Proceedings of the Conference on Information Processing in Sensor Networks. Palo Alto, USA, 2003, 2634: 333-348

[8] He T, Huang C D, Blum B M, Stankovic J A, Abdelzaher T. Range-Free localization schemes in large scale sensor networks//Proceedings of the 9th Annual International Conference on Mobile Computing and Networking. San Diego, USA, 2003: 81-95

[9] Hightower J, Borriello G. Location systems for ubiquitous computing. IEEE Computer, 2001, 34(8): 57-66

[10] Terwilliger Mark. Location in wireless sensor networks[Ph. D. dissertation]. Western Michigan University Kalamazoo, Michigan, 2006

[11] Zeng Yingpei, Cao Jiannong, Hong Jue, Xie Li. Secure localization and location verification in wireless sensor networks: A survey. The Journal of Supercomputing, 2010, 864-869

[12] Jiang Jinfang, Han Guangjie, Zhu Chuan, Dong Yuhui, Zhang Na. Secure localization in wireless sensor networks: A survey. Journal of Communications, 2011, 6(6): 460-470

[13] Bäck T. Evolutionary Algorithms in Theory and Practice, NewYork: Oxford University Press, 1996

[14] Rudolph G. Finite Markov chain results in evolutionary computation: A tour d'horizon. Fundamental Informaticae, 1998, 35(2): 67-89

[15] Wireless Sensor Network Localization Measurement Repository[EB/OL]. http://www.eecs.umich.edu/~hero/localize/, 2006, 10, 18

[16] Capkun S, Hubaux J P. Secure positioning in wireless networks. Selected Areas in Communications, 2006, 24(2): 221-232

[17] Distributed reputation-based secure localization in sensor networks[EB/OL]. http://www.cis.temple.edu/~wu/DRBTS_Journal Version_JoATC.pdf, 2007

[18] Liu D, Ning P, Du W K. Attack-resistant location estimation in sensor networks//Proceedings of the 4th International Symposium on Information Processing in Sensor Networks. Los Angeles, USA, 2006: 99-106

[19] Liu D, Ning P, Liu A, Wang C, Du W K. Attack-resistant location estimation in wireless sensor networks. ACM Transactions on Information and System Security, 2008, 11(4): 1-39

[20] Kuo Sheng-Po, Tseng Yu-Chee, Wu Fang-Jing, Lin Chun-Yu. A probabilistic signal-strength-based evaluation methodology for sensor network deployment. International Journal of Ad Hoc and Ubiquitous Computing, 2005, 1(12): 3-12

[21] Cheng Yu-Yi, Lin Yi-Yuan. A new received signal strength based location estimation scheme for wireless sensor network. IEEE Transactions on Consumer Electronics, 2009, 55(3): 1295-1299

[22] Zhang Yuan. Research on Gaussian mixture model based location estimation algorithms for WSN[Ph. D. dissertation]. Jilin University, Jilin, 2010(in Chinese)
(张原. 基于高斯混合模型的无线传感器网络节点定位算法的研究[博士学位论文]. 吉林大学, 吉林, 2010)

[23] Zhu Jian, Zhao Hai, Xu Jiu-Qiang, Li Da-Zhou. Research on a novel fuzzy theory based localization model in WSNs. Acta Electronica Sinica, 2010, 38(8): 1845-1851(in Chinese)
(朱剑, 赵海, 徐久强, 李大舟. WSNs 中一种新颖的模糊识别定位技术研究. 电子学报, 2010, 38(8): 1845-1851)

[24] Chang Chia-Hung, Liao Wanjinn. A probabilistic model for relative location estimation in wireless sensor networks. IEEE Communications Letters, 2009, 13(12): 893-895

[25] Sheng Zhong, Jadliwala Murtuza, Upadhyaya Shambhu, Qiao Chunming. Towards a theory of robust localization against malicious beacon nodes//Proceedings of the 27th Conference on Computer Communications. Phoenix, USA, 2008: 1391-1399

[26] Zhu Jian, Zhao Hai, Xu Jiu-Qiang, Li Da-Zhou. Localization model in wireless sensor networks. Journal of Software, 2011, 22(7): 1612-1625(in Chinese)
(朱剑, 赵海, 徐久强, 李大舟. 无线传感器网络中的定位模型. 软件学报, 2011, 22(7): 1612-1625)

[27] Kulkarni R V, Förster A, Venayagamoorthy G K. Computational intelligence in wireless sensor networks: A survey. IEEE Communications Surveys & Tutorials, 2011, 13(1): 68-96

[28] Nekooei S M, Manzuri-Shalmani M T. Location finding in wireless sensor network based on soft computing methods//Proceedings of the International Conference on Control, Automation and Systems Engineering. Singapore, 2011: 1-5

[29] Doherty L, Pister Kristofer S J, Ghaoui Laurent El. Convex position estimation in wireless sensor networks//Proceedings of the IEEE Conference on Computer Communications Twentieth Annual Joint Conference of the IEEE Computer and Communications Society. Anchorage, USA, 2001: 1655-1663

[30] Eren T, Goldenberg O K, Whiteley W, Yang Y R, Morse A S, Anderson B D O, Belhumeur P N. Rigidity, computation, and randomization in network localization//Proceedings of the 4th International Symposium on Information Processing in Sensor Networks. Los Angeles, USA, 2004: 2673-2684

[31] Kannan Anushiya A, Mao Guoqiang, Vucetic Branka. Simulated annealing based wireless sensor network localization with flip ambiguity mitigation. Journal of Computers, 2006, 1(2): 1022-1026

[32] Li Min-Qiang, Kou Ji-Song, Lin Dan et al. The Basic Theorem and Application of Genetic Algorithm. Beijing: Science Press, 2002(in Chinese)
(李敏强，寇纪淞，林丹等. 遗传算法的基本理论与应用. 北京：科学出版社，2002)

[33] Patwari Neal, Hero Alfred O, Perkins Matt, Correal Neiyer S, O'Dea Robert J. Relative location estimation in wireless sensor networks. IEEE Transactions on Signal Processing, 2003, 51(8): 2137-2148

[34] Zhang Songtao. Research on localization of wireless sensor networks [Ph. D. dissertation]. Huazhong University of Science and Technology, Wuhan, 2010(in Chinese)
(张松涛. 无线传感器网络定位问题研究[博士学位论文]. 华中科技大学，武汉，2010)

YE Miao, born in 1977, Ph. D. candidate. His research interests include wireless sensor networks, information security, and evolutionary computation.

WANG Yu-Ping, born in 1961, Ph. D., professor, Ph. D. supervisor. His major research interests include evolutionary computation, optimization theory, optimal design method for network and engineering, and data mining.

Background

It is crucial that wireless sensor nodes should be properly located to facilitate the operation of the whole network. When wireless sensor network is exposed in malicious and dangerous environment, attackers may attack the nodes in the localization process and cause incorrect location results which may lead to the complete breakdown of the entire network.

A voting-based algorithm (Vote) is one of the most typical location algorithms to improve network security. This paper addresses the drawbacks of the invalidity of the probability-based maximum likelihood location method under the circumstance of malicious attack and that of the uncertainty of the size of cells in grid-based Vote. It first establishes a more close-to-reality RSS-based variant variance (VV) probability model. After analyzing the security flaw of the probability-based maximum likelihood location method, this paper adopts Vote in designing featured VV probability-based sensor network security location model. Considering the difficulty in working out the object function with the highly nonlinear characteristic in this location model, this research takes a location approach using probability maximum with evolutionary algorithm which more corresponds to the communication characteristic of the sensors to find out the maximum likelihood point. In this way not only the malicious location references can be filtered, but also the malicious nodes should be detected. This paper continues to prove the convergence through stochastic process. When simulated in public datasets and implemented in realistic scenarios, the results confirm that this proposed VV probabilistic model and designed algorithm, under the premise of achieving location accuracy, can in effect accomplish the security location of nodes and the detection of malicious nodes and can also avoid the overload of computation in the Vote as a result of the overly divided cells for the purpose of increasing accuracy.

This work is supported by the National Natural Science Foundation of China under Grant Nos. 61272119, 61203372. The target of the research group is to design optimal method for network and engineering in Internet of Things and Cloud Computing.