基于高斯分布的传感器网络信誉模型

肖德琴[1]，冯健昭[1]，周权[2]，杨波[1]
(1. 华南农业大学 信息学院，广东 广州 510642;
2. 广州大学 信息安全研究所，广东 广州 510006)

摘 要：融合了密码学、经济学、统计学、数据分析等相关领域的知识来建立可信传感器网络，探讨了一种基于高斯分布的传感器网络信誉模型（GRFSN, Gauss reputation framework for sensor network）描述方法。通过对高斯概率分布与信誉分布的拟合分析与证明，证实了用高斯分布建立信誉模型的途径是可行的。通过仿真实验，说明了高斯分布可更好地保持信誉稳定性和表达信任更加直观等特点，实验也显示了 GRFSN 模型具有更强的识别故障和抵御信誉恶意攻击能力的优越性。

关键词：传感器网络; 信誉; 高斯分布
中图分类号：TP309.7 **文献标识码**：A **文章编号**：1000-436X(2008)03-0047-07

Gauss reputation framework for sensor networks

XIAO De-qin[1], FENG Jian-zhao[1], ZHOU Quan[2], YANG Bo[1]
（1. College of Informatics, South China Agricultural University, Guangzhou 510642, China;
2 Institute of Information Security, Guangzhou University, Guangzhou 510006, China）

Abstract: Knowledge from relative domains such as cryptography, economics, statistics, and data analysis was combined together for the development of trustworthy sensor networks, a Gaussian reputation framework for sensor networks (GRFSN) was proposed, which was developed by Jøsang's opinion, with the fitting analyses and test between Gaussian probability distribution and reputation distribution, the possibility to develop reputation model using Gaussian distribution has been proved. Through some preliminary simulation results, this framework is more stabile to describe reputation against other distributions and easier to express the trust directly, and it also has the powerful ability to prevent from malicious attacks or faulty nodes.

Key words: sensor networks; reputation; Gaussian distribution

1 引言

随着传感器网络越来越广泛的应用，保障传感器网络节点可信的信誉系统也逐渐呈现出越来越多的安全威胁，如节点故障、信誉欺骗与抵抗攻击等[1~3]。因此，信誉系统本身的描述方法及安全性成为关注的重点，描述途径的不断改进成为研究热点。早期的信誉系统采用累加平均的方法，这种方法存在很多缺陷，本身就难以抵抗信誉欺骗等攻击[4]，当前描述信誉分布主要有指数分布、二项分布和 β 分布等方法[5,6]。然而在现实世界中，许多物理量的分布概率是符合正态分布或近似正态分布的，Jøsang 等

收稿日期：2007-02-08；**修回日期**：2008-01-11
基金项目：广东省自然科学基金资助项目(06025838)；国家自然科学基金资助项目（60573043）
Foundation Items: The Natural Science Foundation of Guangdong Province (06025838); The National Natural Science Foundation of China (60573043)

人甚至认为在实际问题中遇见的几乎所有的连续变量，都可以满意地用正态分布（即高斯分布）来刻画[7]。借鉴 Jøsang 等人的思路，笔者考虑将节点信誉这一变量引用高斯分布来描述。与指数分布、二项分布和 β 分布相比，高斯分布是直接用数学期望和方差来表示的，在概率模型的信誉系统中信任就是信誉概率分布的数学期望，即用高斯分布来表示信任是相当直观的，免去了计算数学期望的过程。但是，用高斯分布对信誉建模的关键问题是对于一组独立的贝努利事件能否用高斯分布建模，以及信誉更新后是否继续符合高斯分布。

本文融合了密码学、经济学、统计学、数据分析等相关领域的知识来建立可信传感器网络，探讨了一种基于高斯分布的传感器网络信誉模型（GRFSN, Gauss reputation framework for network）描述方法，形式化地分析了符合高斯分布的传感器网络节点信誉的初始化、更新和整合的原理与方法。

作为本文所述方法的数学基础，首先在第 2 节简要描述统计学中的贝叶斯估计与高斯分布，然后分析证明利用高斯分布建立节点信誉的初始化、更新、整合和信任评估的原理与方法。最后，通过仿真测试此模型识别节点故障和抵抗恶意攻击的能力。

2 贝叶斯估计和高斯分布

2.1 贝叶斯估计

在贝叶斯统计学中，先验信息是用未知参数 θ 的概率分布形式给出的，即认为未知参数 θ 是一个随机变量，它就有一个概率分布，这个分布是在实验之前就有的，称为 θ 的先验分布，其密度函数用 $\pi(\theta)$表示。

引理 1(贝叶斯公式)[8] 贝叶斯公式用密度函数形式表示为

$$\pi(\theta \mid x_1,x_2,\cdots,x_n)=\frac{p(x_1,x_2,\cdots,x_n \mid \theta)\pi(\theta)}{\int_{\Theta} p(x_1,x_2,\cdots,x_n \mid \theta)\pi(\theta)\mathrm{d}\theta} \quad (1)$$

其中，$p(x_1,x_2,\cdots,x_n)=\int_{\Theta} p(x_1,x_2,\cdots,x_n \mid \theta)\pi(\theta)\mathrm{d}\theta$ 是样本 x_1，x_2，…，x_n 的边缘分布密度，其积分区域为参数空间 Θ，式(1)是贝叶斯统计学中最基本的公式，式中 $\pi(\theta|x_1，x_2，\cdots，x_n)$是利用样本信息得到的 θ 关于样本的条件密度函数，称之为 θ 的后验密度函数，或称为后验分布。

人们根据先验信息对未知参数 θ 已有一个认识，这个认识就是先验分布 $\pi(\theta)$，通过实验，获得样本，然后通过贝叶斯公式对 θ 的先验分布进行调整，调整的结果就获得了 θ 的后验分布 $\pi(\theta|\ x_1，x_2，\cdots，x_n)$，使人们对 θ 的认识由 $\pi(\theta)$调整到 $\pi(\theta|\ x_1，x_2，\cdots，x_n)$，对 θ 的统计推断就是建立在后验分布 $\pi(\theta|\ x_1，x_2，\cdots，x_n)$的基础上。

参数 θ 的后验分布综合了总体信息，样本信息和先验信息，如今要估计参数，就是要从后验分布密度函数中提取有关 θ 的信息。估计 θ 的方法有众数型贝叶斯估计、中位数型贝叶斯估计和期望型贝叶斯估计。

由于期望型贝叶斯估计在贝叶斯统计学的理论研究和实际应用中最为广泛，很多情形下直接将它简称为 θ 的贝叶斯估计，记为 $\hat{\theta}_{\mathrm{B}}$，即

$$\hat{\theta}_{\mathrm{B}}=\mathrm{E}(\theta \mid x_1,x_2,\cdots,x_n)=\int_{\Theta}\theta\pi(\theta \mid x_1,x_2,\cdots,x_n)\mathrm{d}\theta \quad (2)$$

显然 $\hat{\theta}_{\mathrm{B}}$ 是样本 x_1，x_2，…，x_n 的函数[8]。

2.2 高斯分布

根据概率学原理，高斯分布描述如定义 1 所示[8]。

定义 1 如果随机变量 X 的概率密度函数为

$$P(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),x\in(-\infty,+\infty) \quad (3)$$

其中，μ 和 σ 为常数，σ>0，则 X 服从高斯分布，记为 $X\sim N(\mu,\sigma^2)$，称 X 为正态随机变量。

可见高斯分布由 2 个参数 μ 和 σ^2 决定，当 μ=0，σ=1 时，称 X 服从标准正态分布，记作 $X\sim N(0,1)$。

3 基于高斯分布的信誉模型

很多分布如高斯分布、β 分布、泊松分布、二项分布等都能够表示节点的信誉，而事实上在实际应用中许多物理量的概率分布或者是高斯分布的，或者是近似高斯的。由于高斯随机变量刻画的随机现象比较普遍，符合人类社会认知事物的过程。受 Jøsang 等人观点的启发，探讨提出了一种基于高斯分布的传感器网络信誉模型（GRFSN）。在 GRFSN 中，用 X_{ij} 表示在节点 i 保存的节点 j 的信誉，信誉 X_{ij} 不是一个物理量，而是被看作一种概率分布。它只能在统计上预言其他节点将要发生的行为，但不

能确定地计算其他节点实际发生的行为，信誉更新模型如定义 2 所述。

定义 2　用 D_{ij} 表示节点 i 保存的与节点 j 通信成功与否的事件，X_{ij} 表示节点 i 对节点 j 的历史信誉，X'_{ij} 表示更新后的信誉，则可用函数 f 表示信誉更新过程为

$$X'_{ij} = f\left(D_{ij}, X_{ij}\right) \tag{4}$$

D_{ij} 是用来递归地更新保存在节点 i 中的节点 j 的信誉 X'_{ij}。根据 Jøsang 等人的观点，信誉分布也可以用高斯分布进行建模，框架描述如定义 3 所述。

定义 3　在基于高斯分布的传感器网络信誉系统中，节点 i 保存节点 j 的信誉通过下面的等式表示

$$X_{ij} = N(\mu_j, \sigma_j^2) \tag{5}$$

其中，μ_j 和 σ_j^2 分别表示节点 i 对节点 j 的信誉和方差。

下面将通过对高斯分布和信誉分布的拟合分析与证明，依序说明 GRFSN 模型的初始化、直接信誉更新、间接信誉整合和信任评估 4 个关键步骤实现的途径、原理与方法。

3.1　初始化

传感器网络初始化时，由于还没有进行通信，节点 i 要对所有的邻居节点 j 的信誉赋初值，由于没有先验知识，对邻居节点既不能完全信任，也不能完全不信任，很多相关模型都建议将初始信任值取 0.5 或 1[4~6,9]。但是，将信任初始值绝对地取 0.5 或 1 过于武断，没有考虑邻居节点可信度的偶发性，介于实际社会中大多数信誉分集中在信任（取值 1）与不信任（取值 0）中间，即在 0.5 附近取值更为合理，更加符合人类社会的信誉模型。因此，假定信誉初始化时服从μ_1=0.5，σ^2=0.5^2 的高斯分布，即 N（0.5，0.5^2）。另外，值得一提的是，笔者之所以如此取初始值，除考虑节点信誉服从一定规律又兼顾偶发性以外，同时也获得了信誉从初始到更新的一体化高斯分布描述，达到一种行文的完整和流畅。

3.2　直接信誉更新

直接信誉指节点 i 通过与邻居节点 j 直接进行通信并对其通信行为做出评价后得到的信誉。用高斯分布描述信誉系统需要解决的关键问题是对于一组独立的贝努利事件能否用高斯分布建模，然后变化之后的信誉是否继续符合高斯分布。下面的定理 1 和定理 2 分别给出了对贝努利事件进行高斯分布建模和基于高斯分布进行直接信誉更新的原理与方法。

定理 1　假设在 t 时段中节点 i 和节点 j 执行了 $m+n$ 次事件，其中 m 次成功和 n 次失败，则这组事件在[0,1]区间上满足高斯分布 $N\left(\frac{m}{m+n}, \frac{mn}{(m+n)^2}\right)$。

证明　对二项分布事件用高斯分布建模，其中 m 次成功发生和 n 次失败，则这组事件满足高斯分布 $N\left(m, \frac{mn}{m+n}\right)$。由于上述高斯分布主要集中在区间[0，$m+n$]之间，而信誉分定义在区间[0，1]之间，因此要把上述高斯分布映射到新的区间。则新的高斯分布为 $N\left(\frac{m}{m+n}, \frac{mn}{(m+n)^2}\right)$，可以把它的数学期望看作这组事件对应的信誉分。

根据定理 1 可对经历的一组事件用高斯分布建模，设 $v = \mu = \frac{m}{m+n}$，则 $u^2 = \sigma^2 = \frac{mn}{(m+n)^2} = v\frac{n}{m+n} = v(1-v)$。

定理 2(直接信誉的更新)　假设节点 i 关于节点 j 的信誉分布为 X_{ij}～$N(\mu_j, \sigma_j^2)$，$(X_{ij})_1$，$(X_{ij})_2$，…，$(X_{ij})_t$ 是 X_{ij} 的样本，参数μ_j 的先验分布为μ_j～N（v, u^2），则经历上述定理 1 的事件之后，信誉 X_{ij} 的后验分布服从高斯分布，且满足

$$\begin{cases} \mu'_j = \dfrac{\dfrac{t}{\sigma_j^2}\bar{X} + \dfrac{v}{u^2}}{\dfrac{t}{\sigma_j^2} + \dfrac{1}{u^2}} \\ \sigma'_j = \dfrac{1}{\dfrac{t}{\sigma_j^2} + \dfrac{1}{u^2}} \end{cases} \tag{6}$$

其中，$\bar{X} = \frac{1}{t}\sum_{n=1}^{t}(X_{ij})_n$。

证明　样本关于参数 μ_j 的条件密度为

$$p((X_{ij})_1, (X_{ij})_2, \cdots (X_{ij})_t \mid \mu_j) = \frac{1}{(2\pi)^{\frac{t}{2}}\sigma_j^t}\exp\left(-\frac{\sum_{n=1}^{t}((X_{ij})_n - \mu_j)^2}{2\sigma_j^2}\right) \tag{7}$$

而μ_j 的先验分布为

$$\pi(\mu_j)=\frac{1}{\sqrt{2\pi}u}\exp\left(-\frac{(\mu_j-v)^2}{2u^2}\right) \tag{8}$$

则得到后验密度为

$$\begin{aligned}&\pi(\mu_j\mid(X_{ij})_1,(X_{ij})_2,\cdots,(X_{ij})_t)\\&=\frac{p((X_{ij})_1,(X_{ij})_2,\cdots,(X_{ij})_t\mid\mu_j)\pi(\mu_j)}{\int_{-\infty}^{+\infty}p((X_{ij})_1,(X_{ij})_2,\cdots,(X_{ij})_t\mid\mu_j)\pi(\mu_j)\mathrm{d}\mu_j}\\&=C\exp\left(-\frac{(\mu_j-s)^2}{2\eta^2}\right)\end{aligned} \tag{9}$$

其中，$s=\dfrac{\dfrac{t}{\sigma_j^2}\bar{X}+\dfrac{v}{u^2}}{\dfrac{t}{\sigma_j^2}+\dfrac{1}{u^2}}$，$\eta=\dfrac{1}{\dfrac{t}{\sigma_j^2}+\dfrac{1}{u^2}}$，$C$ 是与μ_j 无关的常数，可以看出μ_j的后验分布是高斯分布，所以信誉的后验分布服从高斯分布。

又μ_j的贝叶斯估计就是后验分布的期望，即 $\hat{\mu}_{\mathrm{B}}$ 为

$$\hat{\mu}_{\mathrm{B}}=\mu'_j=\frac{\dfrac{t}{\sigma_j^2}\bar{X}+\dfrac{v}{u^2}}{\dfrac{t}{\sigma_j^2}+\dfrac{1}{u^2}} \tag{10}$$

此处 $\bar{X}$ 是 t 个时段的历史信誉分的样本均值，实际上在系统实现时不需要把每个时段的信誉分都记录下来。简化的方法是设定一个求和变量 *sum*，每经历一个时段就将当前时段的信誉分累加进去，再除以总时段数 t 就可以得到平均值，这样就大大减少了存储量。

在信誉先验概率服从高斯分布的前提下，其后验概率可由定理 2 计算直接信誉的更新。

3.3 整合信誉

如果节点只是依靠直接信誉，算法的收敛时间是很大的，而且将要耗费大量的代价。一种简单的优化方法是利用传感器网络中其他节点的经验值，这样节点之间可以相互交换信誉信息[9]，本文把这种间接获得的信息称为间接信息。

节点的信誉划分为 2 个子部分$(X_{ij})_{\mathrm{D}}$和$(X_{ij})_{\mathrm{ID}}$，如式(11)所示。

$$X_{ij}=(X_{ij})_{\mathrm{D}}\oplus(X_{ij})_{\mathrm{ID}} \tag{11}$$

其中用符号 $\oplus$ 来指明两者的加权求和运算，相应地称之为直接信誉和间接信誉。直接信誉$(X_{ij})_{\mathrm{D}}$通过通信的行为直接得到，间接信誉$(X_{ij})_{\mathrm{ID}}$ 利用间接信息得到。式(12)和式(13)描述了直接信誉和间接信誉的变化。

$$(X'_{ij})_{\mathrm{D}}=f(D_{ij},(X_{ij})_{\mathrm{D}}),\quad\forall j\in N_i \tag{12}$$

$$(X'_{ij})_{\mathrm{ID}}=h((X_{ij})_{\mathrm{ID}},X_{kj}),\quad\forall k\in N_i \tag{13}$$

虽然节点 i 的所有直接信息都是可靠的，但间接信息未必是可靠的。显然地，节点 i 应把从信誉好的节点处得到的间接信息赋予更高的权重。这样，带权重的间接信息表示如下

$$(X'_{ij})_{\mathrm{ID}}=(X_{ij})_{\mathrm{ID}}+(wX_{kj}),\quad\forall k\in N_i \tag{14}$$

$$w=g(X_{ik})\quad,\forall k\in N_i \tag{15}$$

此处 w 表示权重，它使用函数 $g(\cdot)$通过从 i 和 k 之间的信誉 X_{ik} 得到。显然，函数 $g(\cdot)$可以根据需要选择，关于 w 的选择策略将在后面详细描述。

从前面的分析可以看到，节点的信誉需要既考虑直接信誉，又考虑间接信誉。节点 i 接收到的节点 k 对节点 j 的间接信誉并不是全部认同的，因为这里还要考虑节点 k 的信誉，所以对间接信誉有一定程度的打折，即信誉高的节点 k 给予的间接信誉将获得更高的权重。采用下面的定理 3 进行整合节点 i 对节点 j 的直接信誉和所有间接信誉。

定理 3(整合信誉) 假设节点 i 关于节点 j 的先验信誉分布为 $X_{ij}\sim N(\mu_j,\ \sigma_j^2)$，节点 k 关于节点 j 的信誉分布为 $X_{kj}\sim N(\mu_{kj},\ \sigma_{kj}^2)$，则节点 i 通过节点 k 得到节点 j 的间接信誉，节点 i 对节点 j 的直接信誉和间接信誉进行整合可以由式(16)计算

$$\begin{aligned}X_{ij}&=wN(\mu_j,\sigma_j^2)+(1-w)N(\mu_{kj},\sigma_{kj}^2)\\&=N(w\mu_j+(1-w)\mu_{kj},w^2\sigma_j^2+(1-w)^2\sigma_{kj}^2)\end{aligned} \tag{16}$$

其中，w 是直接信誉的权重，取值范围是[0，1]。

证明 令 s=1−w，η=wX_{ij}+sX_{kj}，且 $X_{ij}\sim N(\mu_j,\ \sigma_j^2)$，$X_{kj}\sim N(\mu_{kj},\ \sigma_{kj}^2)$。先计算$\eta$的密度函数 $\varphi_\eta(x)$

$$\varphi_\eta(x)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}w\sigma_j}\exp\left(-\frac{(y-w\mu_j)^2}{2w^2\sigma_j^2}\right)\cdot\frac{1}{\sqrt{2\pi}s\sigma_{kj}}\exp\left(-\frac{\left(x-y-s\mu_{kj}\right)^2}{2s^2\sigma_{kj}^2}\right)\mathrm{d}y$$

令 $t=\dfrac{y-w\mu_j}{w\sigma_j}$，则 $y=w\sigma_j t+w\mu_j$

$$\varphi_\eta(x)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}w\sigma_j}\exp\left(-\frac{t^2}{2}\right)\frac{1}{\sqrt{2\pi}s\sigma_{kj}}\cdot \exp\left(-\frac{\left[x-\left(w\mu_j+s\mu_{kj}\right)-w\sigma_j t\right]^2}{2s^2\sigma_{kj}^2}\right)w\sigma_j\mathrm{d}t$$

$$=\int_{-\infty}^{\infty}\frac{1}{2\pi s\sigma_{kj}}\exp\left(-\frac{1}{2}\left\{\frac{t^2+\left[x-\left(w\mu_j+s\mu_{kj}\right)-w\sigma_j t\right]^2}{w^2\sigma_{kj}^2}\right\}\right)\mathrm{d}t$$

$$=\int_{-\infty}^{\infty}\frac{1}{2\pi s\sigma_{kj}}\exp\left(\left\{-\frac{1}{2}\frac{s^2\sigma_{kj}^2t^2+\left[x-\left(w\mu_j+s\mu_{kj}\right)\right]^2}{s^2\sigma_{kj}^2}+\frac{-2\left[x-\left(w\mu_j+s\mu_{kj}\right)\right]w\sigma_j t+w^2\sigma_j^2t^2}{s^2\sigma_{kj}^2}\right\}\right)\mathrm{d}t$$

$$=\int_{-\infty}^{\infty}\frac{1}{2\pi s\sigma_{kj}}\exp\left(-\frac{1}{2}\left\{\frac{\left(w\mu_j+s\mu_{kj}\right)t^2}{s^2\sigma_{kj}^2}-\frac{2\left[x-\left(w\mu_j+s\mu_{kj}\right)\right]w\sigma_j t}{s^2\sigma_{kj}^2}+\frac{\left[x-\left(w\mu_j+s\mu_{kj}\right)\right]^2\left(w^2\mu_j^2+s^2\mu_{kj}^2\right)}{s^2\sigma_{kj}^2\left(w^2\mu_j^2+s^2\mu_{kj}^2\right)}\right\}\right)\mathrm{d}t$$

$$=\int_{-\infty}^{\infty}\frac{1}{2\pi s\sigma_{kj}}\exp\left(-\frac{1}{2}\frac{\left[x-\left(w\mu_j+s\mu_{kj}\right)\right]^2}{w^2\mu_j^2+s^2\mu_{kj}^2}\right)\cdot \exp\left(-\frac{1}{2}\frac{w^2\mu_j^2+s^2\mu_{kj}^2}{s^2\mu_{kj}^2}\left\{t^2-\frac{2\left[x-\left(w\mu_j+s\mu_{kj}\right)\right]w\sigma_j t}{w^2\mu_j^2+s^2\mu_{kj}^2}+\frac{\left[x-\left(w\mu_j+s\mu_{kj}\right)\right]^2w^2\mu_j^2}{\left(w^2\mu_j^2+s^2\mu_{kj}^2\right)^2}\right\}\right)\mathrm{d}t$$

$$=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{\left[x-\left(w\mu_j+s\mu_{kj}\right)\right]^2}{w^2\mu_j^2+s^2\mu_{kj}^2}\right)\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}s\sigma_{kj}}\cdot \exp\left(-\frac{1}{2}\frac{w^2\mu_j^2+s^2\mu_{kj}^2}{s^2\mu_{kj}^2}\left\{t-\frac{\left[x-\left(w\mu_j+s\mu_{kj}\right)\right]w\sigma_j}{w^2\mu_j^2+s^2\mu_{kj}^2}\right\}^2\right)\mathrm{d}t$$

$$=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{\left[x-\left(w\mu_j+s\mu_{kj}\right)\right]^2}{w^2\mu_j^2+s^2\mu_{kj}^2}\right)\cdot \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sqrt{w^2\mu_j^2+s^2\mu_{kj}^2}}\sqrt{\frac{w^2\mu_j^2+s^2\mu_{kj}^2}{s^2\mu_{kj}^2}}\cdot \exp\left(-\frac{1}{2}\frac{w^2\mu_j^2+s^2\mu_{kj}^2}{s^2\mu_{kj}^2}\left\{t-\frac{\left[x-\left(w\mu_j+s\mu_{kj}\right)\right]w\sigma_j}{w^2\mu_j^2+s^2\mu_{kj}^2}\right\}^2\right)\mathrm{d}t$$

$$=\frac{1}{\sqrt{2\pi}\sqrt{w^2\mu_j^2+s^2\mu_{kj}^2}}\exp\left(-\frac{1}{2}\frac{\left[x-\left(w\mu_j+s\mu_{kj}\right)\right]^2}{w^2\mu_j^2+s^2\mu_{kj}^2}\right)$$

即 $\eta\sim N\left(w\mu_j+s\mu_{kj},w^2\mu_j^2+s^2\mu_{kj}^2\right)$。其中，对权重 w 的取值有几种方法：

1) 由用户根据系统需要赋予阈值。如果节点 i 只相信直接信誉，而完全忽略间接信誉，则取 w=1；如果节点 i 不相信自己，而只相信间接信誉，则取 w=0。一般来说，直接信誉是可靠的，间接信誉未必是可靠的，因此一般取 w 为大于 0.5 的常数，对于安全性要求很高的传感器网络关键节点可能取 w 接近 1。

2) 通过间接信誉获得。如果用户直接取 $1-w=\mu_{kj}/2$ 的话，那么从信誉较高的节点得到的间接信誉将获得较高的权重，信誉较低的节点则权重较小。

3) 采用相关系数的算法。如果节点 i 接收到的节点 k 对节点 j 的间接信誉与节点 i 整合后的信誉相关性比较大，那么节点 i 应该相信节点 k 多一些，否则如果相关性较小，那么应该不要过多地相信节点 k。因此，采用相关系数的计算方法能够比较科学地确定 w 的取值，可以采用下面的引理 2 来计算。

引理 2(相关系数)[8]　设 $\bar{X}$ 和 $\bar{Y}$ 分别是经过 t 个时段后 X_{ij} 和 X_{kj} 的均值，则 X_{ij} 和 X_{kj} 的样本相关系数为

$$w=\hat{\rho}=\frac{\sum_{n=1}^{t}((X_{ij})_n-\bar{X})((X_{kj})_n-\bar{Y})}{\sqrt{\sum_{n=1}^{t}((X_{ij})_n-\bar{X})^2}\sqrt{\sum_{n=1}^{t}((X_{kj})_n-\bar{Y})^2}}\tag{17}$$

为了简化计算，参考二维总体相关系数的估计

公式，可以采用如下公式进行计算

$$w=\frac{\sum_{n=1}^{t}(X_{ij})_n(X_{kj})_n-t\overline{X}\overline{Y})}{\sqrt{\sum_{n=1}^{t}(X_{ij})_n{}^2-t\overline{X}^2}\sqrt{\sum_{n=1}^{t}(X_{kj})_n{}^2-t\overline{Y}^2}} \tag{18}$$

其中，$\overline{X}$ 和 $\overline{Y}$ 是 t 个时段的历史信誉分的样本均值，简化的方法如前所述。

3.4 信任

在人类社会网络中，信任可以简单地定义为一个人对其他人行为的期望，当其他人的行为发生之前此人要采取的行动。在 GRFSN 中，信任 T_{ij} 是节点 i 对节点 j 将要发生行为的主观期望，即计算 2 个节点之间信誉概率分布的统计期望，T_{ij} 表示节点 i 对节点 j 的信任。与信誉不同的是，信任是一个数值，用信誉分的数学期望表示，即式（19）所示

$$T_{ij}=\mathrm{E}[X_{ij}] \tag{19}$$

在高斯分布的信誉系统中，信任就是高斯分布的第一个参数，即

$$T_{ij}=\mathrm{E}(N(\mu_j,\sigma_j^2))=\mu_j \tag{20}$$

4 仿真分析

根据以上原理，设计了一个基于 GRFSN 模型的传感器网络信誉系统，并在传感器网络仿真环境 TOSSIM[10]中进行了测试应用。模拟的环境是从 1 000 亩大规模农田水分传感器网络中，部署了 1 000 个节点，定时采集农田水分数据，水分信息数据以多层树形结构形式沿着传感器节点向基站传递，由父节点保存和计算子节点的信誉。通过仿真实验，假定初始时各节点信誉服从分布 $X_{ij}\sim N(0.5,0.5^2)$，在每个时段内发生的事件相互独立，然后模拟子节点 j 发生各种异常事件，测试了经历 100 个时段当中父节点 i 保存的对子节点 j 信誉的变化情况。为了评价高斯信誉模型，对比β分布、Beth 模型、算术平均信任更新算法的特点，本文设计了一系列的仿真情景，包括连续合作与连续不合作的信誉变化、诋毁攻击、合谋攻击、故障检测和身份攻击。每一种情景模拟委托者和授权者之间不同的行为模式，这些行为模式通过信誉算法反映了信誉的动态性，也显示了每一种信誉更新算法的特点。介于篇幅所限，在文献[11]中做出了详细的实验分析，下面仅列出应用最多的与β分布的对比结果。

1) 高斯分布能够描述信誉，且比β分布更接近最新信誉。如图 1 所示，若 j 不发送任何数据包，将保持初始值 0.5。本文总是选择阈值(此处选 0.95)远远高于初始信任 0.5，j 将被 i 分类为恶意节点。若 j 连续合作，信任呈稳定的上升趋势，高斯分布的信誉经历 9 个时段后稳定在 0.95 左右，β 分布的信誉需经历 30 个时段后才稳定在 0.95 左右[12]，与 β 分布相比高斯分布更加接近最新的信誉。若 j 连续不合作，信任呈下降趋势，并且没有接受其他节点的间接信誉，高斯分布的信誉经过 5 个时段后信任稳定在 0.08 左右，的信誉需经过 36 个时段后才稳定在 0.08 左右，与 β 分布相比高斯分布更加接近最新的信誉。

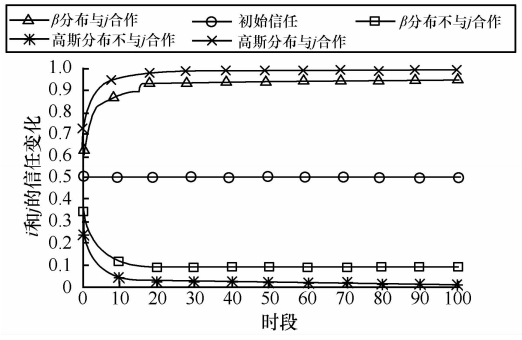


图 1 GRFSN 的信任变化

2) 能够对故障进行检测。如图 2 所示，若 j 经过 50 个时段以后发生了系统部件故障，信任将从 0.99 下降，i 不再与 j 合作。虽然高斯分布的下降速度不是太快，但是高斯分布的阈值选择应该比 β 分布的阈值更大，同样可以发现这种故障。

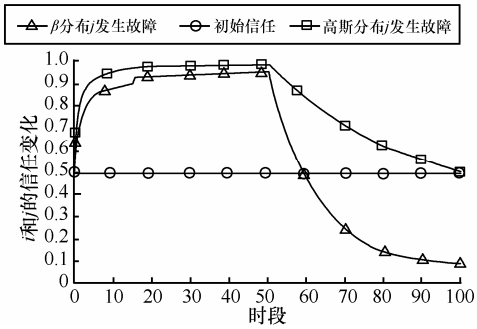


图 2 发生故障的信任变化

3) 能够检测恶意攻击。若 j 数次发送成功，又数次发送失败，如图 3 所示，j 的信誉呈抖动下降趋势，在 GRFSN 中可以发现这种攻击，并把 j 归类为恶意节点。

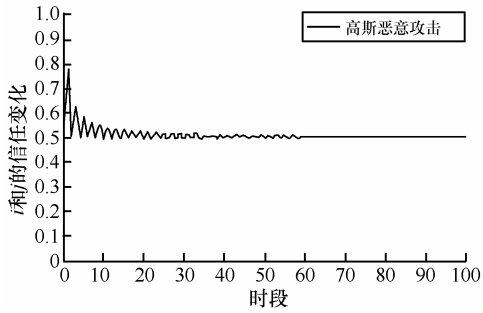


图 3　恶意攻击的信任变化

5　结束语

本文融合了密码学、经济学、统计学、数据分析等相关领域的知识，在概率模型的基础上提出了一种基于高斯分布建立传感器网络信誉系统的形式化模型（GRFSN）。在 GRFSN 模型中，信誉初始时假定服从 N（0.5，0.5^2）的高斯分布，经历一系列事件后信誉不断地进行调整，通过对高斯概率分布与信誉分布的拟合分析与证明，获得了信誉从初始到更新的一体化高斯分布描述，从而证实了用高斯分布建立信誉模型的途径是可行的。

值得一提的是，本模型在初始值确定和加权处理邻居信誉度方面还作了更加精确的描述。在系统开销方面，定理 1 的时间复杂度为 $O(1)$；对定理 2 直接信誉更新实现的运算量主要集中在求 σ 的开方运算和 $\bar{X}=\frac{1}{t}\sum_{n=1}^{t}(X_{ij})_n$ 这 2 个，而在本模型中 σ 很小，因此定理 2 实现的运算量就主要集中在求 $\bar{X}=\frac{1}{t}\sum_{n=1}^{t}(X_{ij})_n$，其运算时间复杂性为 $O(t)$，实现中，采用了以空间换时间的策略，时间复杂度可以为 $O(1)$；对定理 3 信誉整合的计算，与需要整合的邻居节点数 m 有关，其运算复杂性为 $O(m)$；引理 2 相关度计算方法的实现稍微复杂一些，其主要运算开销在开方运算，其时间复杂度最大为 $O(t^2)$。因此，在系统节点假冒发生概率少，仅需要数据认证情形下，为简化计算，也可以将初始信任值就简单地取 0.5 即可，在加权值取值方面也可不考虑采用相关系数权重而直接取经验值，可以大大简化计算量。

同时，仿真实验表明高斯分布比其他分布更能保持信誉稳定性，而且能更直观地表达信任，在识别故障和抵御信誉恶意攻击方面也显示了一定的优越性。当然，在本模型应用于传感器网络时，针对其他攻击的分析能力，以及其最佳运行规模问题需要进一步分析和测试，这正是我们下一步要研究的目标。

参考文献：

[1] PERRIG A, SZEWCZYK R, WEN V, *et al*. SPINS: security protocols for sensor networks[J]. ACM Wireless Networks, 2002, 8(5): 521-534.

[2] WATRO R, KONG D, CUTI S F, *et al*. TinyPK: securing sensor networks with public key technology[A]. To Appear in Second Workshop on Security in Sensor and Ad-hoc Networks[C]. 2004.

[3] GANERIWAL S, KUMAR R, HAN C C, *et al*. Srivastava. Location & Identity based Secure Event Report Generation for Sensor Networks[R]. NESL Techanical Report, 2004.

[4] RESNICK P, ZECKHAUSER R. Trust among strangers in Internet transactions: empirical analysis of eBay's reputation system[A]. NBER Workshop on Empirical Studies of Electronic Commerce[C]. 2000.

[5] BLAZE M, FEIGENBAUM J, LACY J. Decentralized trust management[A]. Proceedings of IEEE Conf. Security and Privacy[C]. Oakland, California, USA, 1996.

[6] JØSANG A, ISMAIL R. The beta reputation system[A]. Proceedings of the 15th Bled Electronic Commerce Conference[C]. 2002.

[7] JØSANG A. A logic for uncertain probabilities[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2001, 9(3): 279-311.

[8] HOGG R V, CRAIG A T. Introduction to Mathematical Statistics (5th Edition)(M). Prentice Hall ,2004.

[9] BUCHEGGER S, BOUDEC J L. A robust reputation system for P2P and mobile ad-hoc networks[A]. Proceedings of P2PEcon 2004[C]. Cambridge MA, U S A, 2004.

[10] 姚兰, 桂勋, 王保强. 无线传感器网络路由协议的研究和仿真，计算机应用与软件, 2006,23(9):128-130.
YAO L, GUI X, WANG B Q. Study and simulation of routing protocol for wireless sensor networks[J]. Computer Applications and Software, 2006,23(9):128-130.

[11] 冯健昭. 基于高斯分布的形式化信誉模型研究[D]，华南农业大学硕士毕业论文，2007.
FENG J Z. Formal Reputation Model Based on Gaussian Distribution[D]. South China Agriculture University, 2007.

[12] 冯健昭，肖德琴，杨波. 基于β分布的无线传感器网络信誉系统[J]，计算机应用，2007,27(1): 111-113,117.
FENG J Z, XIAO D Q, YANG B. Reputation system for wireless sensor networks based on β distribution[J]. Journal of Computer Applications, 2007,27(1): 111-113,117.

（下转第 62 页）

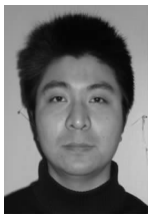
通信学报, 2006,27(1): 132-139.

ZHENG K, WANG N, LIU A F. New AODV based clustering routing strategy [J]. Journal on Communications, 2006，27（1）: 132-139 .

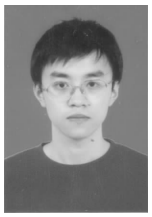
[10] CORSON S, MACKER J. Mobile ad hoc networking (MANET): routing protocol performance issues and evaluation considerations (RFC 2501) [EB/OL]. http://www.ietf.org/rfc/rfc2501.txt,1999.

[11] JiST user guide and SWANS user guide [EB/OL]. http://jist.ece.conell.edu/docs.html, 2004.

作者简介：



徐佳（1980-），男，江苏常州人，南京理工大学博士生，主要研究方向为无线网络路由技术。



李陟（1979-），男，江苏南京人，南京理工大学博士生，主要研究方向为无线网络拓扑控制。

李千目（1979-），男，江苏南京人，南京理工大学博士后，讲师，主要研究方向为信息安全和网络性能诊断。

刘凤玉（1943-），女，江苏江阴人，南京理工大学教授、博士生导师，主要研究方向为网络性能和信息安全。

（上接第 53 页）

作者简介：

肖德琴（1970-），女，重庆人，华南农业大学博士生、副教授、硕士生导师，主要研究方向为信息安全与无线传感器网络。

周权（1971-），男，四川广安人，广州大学讲师，主要研究方向为网络安全。

冯健昭（1981-），男，广东南海人，华南农业大学研究生，主要研究方向为计算机网络与安全。

杨波（1963-），男，陕西西安人，博士，华南农业大学教授、博士生导师，主要研究方向为密码与信息安全。