高效的攻击检测与数据融合算法

程宏兵[1,2,3]，容淳铭[3]，黄晓[4]，EGGEN Skjalg[3]，曾庆凯[1]

(1. 南京大学 软件新技术国家重点实验室，江苏 南京 210093；2. 江苏城市职业学院 信息工程系，江苏 南京 210017；
3. 斯塔万格大学 计算机科学与电子工程系，斯塔万格 挪威 4036；4. 江苏移动通信有限公司 互联网中心，江苏 南京 210013)

摘 要：给出了一种高效的无线多媒体传感器网络攻击检测和数据融合算法 EIDSART**。该算法从节点的多元属性方面对节点行为特征进行界定，通过选择合适的邻居节点集合，可以运用于任意规模的多媒体传感器网络；另外，在经过精确检测攻击行为的情况下，对传感数据进行了融合，降低了网络通信开销。仿真结果表明，**EIDSART**在攻击检测精度和误报率等方面具有优势，并能得到精确的数据融合结果。**

关键词：物联网；无线多媒体传感器网络；攻击检测；数据融合

中图分类号：TP393 **文献标识码：**A **文章编号：**1000-436X(2012)09-0085-10

Efficient attack detection and data aggregation algorithm

CHENG Hong-bing[1,2,3], RONG Chun-ming[3], HUANG Xiao[4], EGGEN Skjalg[3], ZENG Qing-kai[1]

(1. State Key Laboratory for Novel Software Technology ing University, Nanjing 210093, China;
2. Department of Information Engineering, Jiangsu City Vocation College, Nanjing 210017, China;
3. Department of Electronic Engineering & Computer Science, University of Stavanger, Stavanger, 4036, Norway;
4.Internet Center, Jiangsu Branch, China Mobile Company, Nanjing 210013, China)

Abstract: An efficient algorithm of attack detection and data aggregation for wireless multimedia sensor networks based on the previous work was proposed. The proposed algorithm concludes the action trait of sensor nodes from their attribute vectors without any prior knowledge, at the same time; it was scalable and could be applied in large scale networks. The simulation results show that the proposed algorithm can detect the attacks action more accurate than other technologies, and can make data aggregation efficiently. At the same time, the proposed algorithm can make the wireless multimedia sensor networks secure and reduce communication flow so that it will save a lot of resources in wireless multimedia sensor networks.

Key words: internet of things; wireless multimedia sensor network; attack detection; data aggregation

收稿日期：2012-02-22；**修回日期：**2012-05-25

基金项目：国家自然科学基金资助项目（61170070, 61021062, 60873231**）；江苏省六大人才高峰计划基金资助项目（**11-JY-009**）；江苏省高校自然科学基金资助项目（**11KJB510003**）；中国博士后基金资助项目（**2012M511252**）；江苏省博士后基金资助项目(**1102014C)

Foundation Items: The National Natural Science Foundation of China (61170070, 61021062, 60873); The Six Kinds Peak Talents Plan Project of Jiangsu Province (11-JY-009); The Nature Science Foundation of Jiangsu Normal Higher niversity (11KJB510003); China Postdoctoral Science Foundation Funded Project (2012M511252); Jiangsu Province Postdoctoral Science Foundation Funded Project (1102014C)

1 引言

攻击检测是物联网信息安全的重要保障技术，如感知终端受到恶意入侵者的非法攻击而被控制、采集的数据信息受到攻击者的攻击并泄露以及一些敏感监测数据信息的被操控等都是物联网实际应用中需要解决的关键安全问题，都需要在物联网应用中采用有效的安全攻击检测技术。无线多媒体传感器网络[1]（WMSN, wireless multimedia sensor networks）是物联网的主要组成部分，是在传统无线传感器网络的基础上发展起来由一组具有计算、存储和通信能力的多媒体传感器节点（包含摄像头、音视频以及温度、湿度等传感器）通过自组织的方式组成的分布式感知网络。WMSN 借助于节点上多媒体传感器感知周边环境的多种媒体信息（音频、视频、图形、图像等），通过多跳中继方式将数据传到信息汇聚中心，汇聚中心对监测数据进行分析，实现全面而有效的监测，并将监测到的多媒体数据发送给网络观察者和外部处理器进行相应处理。

目前，WMSN 已经在一些非商业或实验性质的环境得到了初步应用，尤其是一些传统网络难以涉及的场合，如战场中对敌监测以获取敌方高分辨率图像、生态环境监控以获取生物视频信息、原始森林防火以获取着火点源头、建筑及桥梁监控以及文物古迹的保护以获取异常音频信息等。同时，随着一些基础性技术如嵌入式处理技术、通信技术和存储技术的不断发展，WMSN 逐渐可以处理更加复杂的综合多媒体相关业务，如混合声音录制、多视频信息获取等。社会需求的不断增加必将使 WMSN 获得更大的应用空间。

尽管比普通传感器节点的资源有所改善，但 WMSN 中的节点处理多媒体信息还是显得很有限。同时由于节点的能量使用寿命决定网络的生命周期，因此节约能量是无线多媒体传感器网络工作必须考虑的问题，相关研究[2]表明在 100m 的通信线路上传输 1kbit/s 的数据分组和 CPU 执行 3Mbit/s 的指令所消耗的能量几乎是相同的，因此，传感节点应注重于本地的数据处理而减少远距离的数据传输，从而可以减少通信所带来的能量耗费巨大的负担。另一方面，在实际应用中，WMSN 所搜集的原始信息具有冗余和碰撞的特征，这样的数据传输会极大地影响有效数据的采集并降低网络的生存周期。因此，在保障网络安全的基础上设计有效的数据融合技术是非常重要的。有效的数据融合技术可以剔除冗余的传感数据，减少网络传输的通信开销，减少传输数据的碰撞，从而减轻了网络拥塞，节约了节点的宝贵能量，达到延长 WMSN 生命周期的目的。

2 预备知识

在提出 EIDSART 算法之前，首先对算法中需要涉及到的一些数学基础知识和理论进行简要介绍。

2.1 马氏平方距离

多媒体传感器网络由众多传感器节点协作进行感知工作，获取监测对象的信息。对于每个节点而言，都有多元属性，每个属性值表示节点的某一方面特征，在本文中，主要考虑节点的分组传输率、分组丢失率、传输时延和传感数据值等 4 个属性，所有的节点都按照 4 个属性进行排列形成一个节点属性矩阵。

节点属性矩阵中的列表示节点的某一单个属性，行表示每个节点的属性向量。多媒体传感器网络的行为是由所有节点共同体现的。向量表是一种有效的计算 2 个未知样本集相似度的方法，它考虑到各种特性之间的联系。而马氏平方距离则是一种基于变量（属性）平均值和一个或多个变量（属性）采样集协方差矩阵的多变量异常检测方法，是一种衡量某一随机向量、变量或属性与平均值分布距离的方法。

对一个 k 维的多变量样本集 $x_i=\{t_1,t_2,\L,t_k\}$ $(1\le i\le n)$，马氏平方距离（MD^2）定义如下。

$MD_i^2=\left((x_i-m)^{\mathrm{T}}\boldsymbol{S}^{-1}(x_i-m)\right)(1\le i\le n)$，其中，$m=\{m_1,m_2,\L,m_k\}$ 是估计多变量均值，$\boldsymbol{S}$ 是估计协方差矩阵。

2.2 卡方分布

卡方分布是随机统计学中的一个常用的概率分布方法，是对已知事件的一种总结。对于正整数 k，自由度为 k 的卡方分布是一个随机变量的概率分布。

设 $f(x_i)$ 满足 $N_k(m,\boldsymbol{S})$ 分布，即 k 维向量 $f(x_i)$ 符合均值向量为 m，协方差矩阵为 $\boldsymbol{S}$ 的多变量普通分布。马氏平方距离 $MD_i^2=\left((x_i-m)^{\mathrm{T}}\boldsymbol{S}^{-1}(x_i-m)\right)$ $(1\le i\le n)$ 满足 c_k^2 分布，k 是多变量中的变量数，算法中，k 表示采样集中的节点属性数。图 1 所示的卡方累积分布函数表示小于或等于 $c_k^2(a)$ 的样本概率 a，$c_k^2(a)$ 表示概率为 a % 的卡方分布值且使用 $c_k^2(a)$ 作为阈值 q，马氏平方距离 MD^2 超过该值的节点将被视为异常。

2.3 估计方法

正交 Gnanadesikan-Kettenring(OGK)估计方法

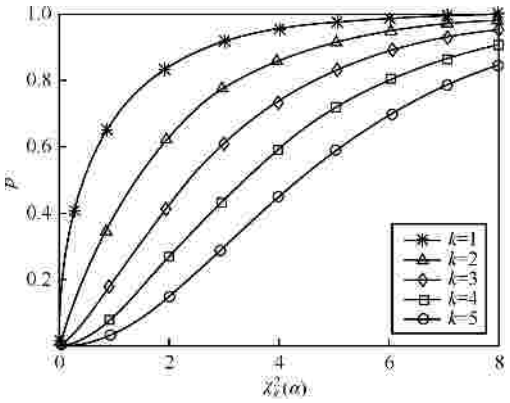

图 1　卡方累积分布函数（引自维基百科—卡方分布）

是一种稳健计算样本集均值 m 和协方差矩阵 S 的估计方法，该方法广泛运用于单个或多个变量（属性）的采样样本的均值和方差的稳健估计中。

本节中，对使用正交 Gnanadesikan-Kettenring (OGK) 估计方法计算单变量（属性）的均值 $\hat{m}$ 和 $\hat{s}^2$ 的方法进行介绍。

假设单变量（属性）向量 $Y=\{y_1,y_2,\text{L},y_n\}$ 符合均值为 m，方差为 s^2 的卡方分布。令 m_0 为向量 Y 的中间值，s_0 是向量 Y 的中值函数 $s_0=\text{med}(|Y-\text{med}(Y)|)$。

设计权重函数 $W(x)=\left(1-\left(\frac{x}{c_1}\right)^2\right)^2 I(|x|\leq c_1)$ 和 r 函数 $r(x)=\min(x^2,c_2^2)$，其中，$c_1=4.5$，$c_2=3$，于是 $\hat{m}$ 和 $\hat{s}^2$ 可以估算如下：$\hat{m}=\frac{\sum_{i=1}^{n}y_iW(v_i)}{\sum_{i=1}^{n}W(v_i)}$，$v_i=\frac{y_i-m_0}{s_0}$，$\hat{s}^2=\frac{s_0}{n}\sum_{i=1}^{n}r\left(\frac{y_i-\hat{m}}{s_0}\right)$。

限于论文的篇幅，使用正交 Gnanadesikan-Kettenring(OGK)估计方法估算多变量（属性）采样样本 $F(x)=\{f(x_i)=(f_1(x_i),f_2(x_i),\text{L},f_k(x_i))^{\text{T}}\mid x_i\in N(x)\}$，均值 $\hat{m}$ 和协方差矩阵 $\hat{S}$ 的方法不再赘述，详见文献[3]。

3　相关性研究分析与比较

传感器网络的安全[4]问题一直是业界研究的热点问题，Hung 等提出了基于时间片的选举机制[5]来应对针对数据融合的攻击，由于该方案在时间片的选定方面具有不确定性，很难获得理想的结果；在安全保密的设计方面，文献[6]给出了依赖密钥的隐蔽性和健壮性来保证传感网络安全的数据链路来达到对网络的保护，但实践表明，在目前资源受限的传感器节点之间进行大量的加、解密运算显然是行不通的；Francesco 等试图通过对网络协议[7]进行剖析和设计的基础上对网络安全进行保障，该思想符合传感器网络的发展趋势，但在多数情况下，一个安全协议的选择往往面临两难的选择，因为协议的复杂性和安全性往往是相伴的。文献[8~10]则从具体的传感器网络所面临的攻击方面进行了相关研究，分别从 DDoS 泛洪攻击、外部入侵者攻击以及寄生虫攻击等方面分别提出了相关的安全防护和应对机制，因为是针对具体的安全威胁，这些安全方法和机制不具有普遍性，难以全面应用。

在外部攻击和入侵检测方面，文献[11]为无线 ad hoc 网络提出了一种协作的安全入侵检测机制，该机制主要设计符合攻击者各种行为特征的规则来对攻击者进行匹配与归类，从而采取相应的安全应对措施，该机制由于在攻击者规则的更新和发现方面具有时滞性而无法获得有效应用；Onat 等提出了基于节点传输特性和节点行为[12]进行挖掘以获取节点异常行为的传感器网络入侵检测方法，该方法过于依赖节点之间的关联而忽视了一些普遍情况，因此，入侵检测的精确性难以得到保障；Mostarda 等提出了一种分布式的入侵检测系统来提高传感器网络的安全，该系统在网络内部安置了多个代理对网络异常行为进行检测，能取得较好的检测效果，但该方案在时间复杂度和节点资源消耗方面不适合传感器网络；文献[13~15]则分别从节点的异常检测、节点的异常行为统计分析和节点被俘获

而成为异常等方面对传感器网络的入侵检测进行了研究，由于节点的异常行为匹配规则选取具有一定的复杂性，该检测方法具有较高的误报率。

受到上述研究的启发，设计了一种基于前期工作[16]改进的攻击检测和数据融合算法 EIDSART，对原先方案的改进主要体现在：1）对算法中集合节点的选取体现了随意性；2）精练了算法，提高了算法的实现效率及可行性；3）构建了符合 WMSN 实际应用的网络模型和应用场景。文中把 IADART 算法[17]作为提出算法的主要对照和比较。IADART 算法是目前无线传感器网络入侵检测领域相对成熟的技术。该技术采用固定的阈值对异常节点进行比较，且算法中只采用 2 个邻居节点集合进行攻击检测，因此 IADART 算法无论是在检测的精度还是在异常检测覆盖率方面都有很大的改进空间；另外，由于 IADART 算法采用单一的卡方分布值对异常节点的阈值进行计算，所以检测中的误报率有时候会变得非常大。而最值得注意的是 IADART 算法根本没有涉及传感数据的融合问题，网络中的多媒体数据通信流量往往巨大，消耗了宝贵的网络和节点资源。图 2 和图 3 分别给出了在没有异常节点的情况下，传感器节点数目从 10 变化到 30 时，以卡方值 $c_4^2(97.5\%)$ 和卡方值 $c_4^2(99.9\%)$ 作为阈值时的

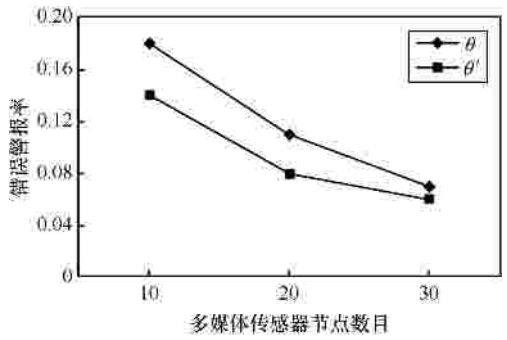

图 2 $c_4^2(97.5\%)$ 分布下的错误警报率趋势

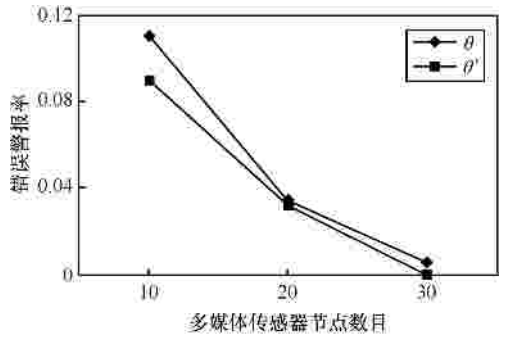

图 3 $c_4^2(99.9\%)$ 分布下的错误警报率趋势

基于动态阈值 q' 和 IADART 算法中静态阈值 q 进行测试的错误警报结果图。从测试结果可以看出，采用大的卡方值可以使错误警报率降低到可以接受的水准；同时，采用动态阈值 q' 比采用静态阈值 q 也能大大降低算法的错误警报率。

4 系统模型

本文所考虑的 WMSN 系统由连通图 $G=(Vet, Edg)$ 表示，图中的顶点 $v(v\in Vet)$ 表示 WMSN 的节点，边 $e(e\in Edg)$ 表示节点间的通信链路，且节点之间的通信是双向的，设 WMSN 中节点的数量为 $N=|Vet|$。

不失一般性，在本文的 WMSN 系统中，基站控制整个网络，并且负责簇的初始化和簇头的选择，节点初始状态下具有相同的能力，相同簇的成员节点具有相同的负载和传感任务；同时，在正常情况下，相邻节点对外表现相同的特征。对于受到外部攻击者操纵的网络内部攻击者一般与普通正常节点具有相同的资源，但其向外表现的行为与普通节点具有明显的差别，如内部攻击者可能会丢弃或广播多余的数据分组，报告错误的传感信息或报告与相邻节点有明显差异的传感数据等。

考虑到资源的有效利用性和 WMSN 对于网络系统中的每个传感器节点都工作在间歇性的选择模式，即每个节点只间歇地监听直接邻居节点的活动信息，如节点 x 可以选择性地监听或者忽略其直接相邻节点 x_i 所发送或接收的信息，因此，x 对 x_i 的监听是间歇性的。根据第 2 节的多元属性理论，可以对节点 x_i 的网络行为使用 q 元属性向量函数进行建模，建模结果为 $f(x_i)=(f_1(x_i), f_2(x_i), \cdots, f_k(x_i))^{\mathrm{T}}$，其中每个组成部分 $f_j(x_i)(1\le j\le k)$ 表示节点 x_i 某一方面属性的监测结果，例如某一个监测周期的分组丢失数量或者广播分组的数量，因此，$f_j(x_i)$ 的结果可以是连续或离散的值。节点建模结果是无线多媒体传感器网络检测并判断是否发生攻击的重要标准，一般情况下，攻击检测系统根据节点某一特征属性选定的属性阈值与检测对象的相应值进行比较以判断是否有攻击事件发生。

本文中，定义传感监测数据的融合函数为 $A(t)=(d_1(t), d_2(t), \cdots, d_N(t))$。$d_i(t)(i=1,2,\cdots,N)$ 表示传感器节点 i 在时刻 t 所获得的采集数据。目前，出现了很多与应用相关的典型数据融合算法，如 count、average、median、max 和 min 都可以采用求和 sum 函

数的方式进行处理，本文以 sum 函数作为数据融合的研究对象，记 $A(t)=\sum_{i=1}^{N} d_i(t)$。本文系统模型中的数据先由簇成员节点进行采集，其次传递给其父节点进行融合，然后再往上传递继续进行融合，最后传到基站，数据融合的具体过程如图 4 所示。

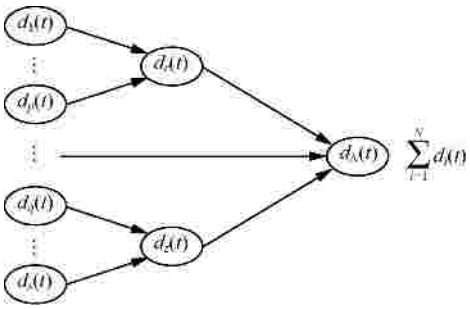

图 4　节点数据融合求和函数

5　EIDSART：高效的攻击检测和数据融合算法

由第 2 节的相关讨论已知，$MD_i^2=((x_i-m)^{\mathrm{T}}S^{-1}(x_i-m))$ 服从自由度为 k 的卡方分布，因此，一种简单计算阈值 q_0 的方法是计算 $c_k^2(a)$ 并使之作为阈值，其中，a 是高于 90%以上的卡方分布，可以选择 $a=0.975$，即 97.5%的马氏平方距离等于或小于 97.5%的卡方分布。在实际的多媒体传感器网络环境下，由于传感器节点的多属性，变量不一定满足如普遍分布，因此，这样选择的阈值 $c_k^2(a)$ 由于是固定值而会产生错误。在本算法中，采用一种基于权重函数对阈值 q_0 进行动态计算的方法，计算方法如下：

$$d_0=\frac{c_k^2(a)\times \mathrm{med}(MD_1^2,MD_2^2,\mathrm{L},MD_n^2)}{c_k^2(0.5)}$$

其中，n 是节点的数目。在攻击检测算法中，使用 d_0 替代阈值 q_0 进行检测，假如一个节点的马氏平方距离大于 d_0，则该节点将被看成内部攻击者。这种检测方式由于采用了动态计算该阈值 q_0 的方法而独立于马氏平方距离的计算，所以能取得较好的效果。下面对 EIDSART 算法进行描述。

5.1　信息收集阶段

1) 集合 $N_1(x)$、$N_2(x)$ 信息收集

令集合 $N_1(x)$ 是域 R^2 中的有界闭集，且集合中的所有节点都能直接被节点 x 监测到，即 $N_1(x)$ 是 x 的 1-跳邻居集合；

令集合 $N(x)(\supseteq N_1(x))$ 也是域 R^2 中的有界闭集，且集合包含节点 x 和与节点 x 最邻近的其他 $n-1$ 个节点；集合 $N_2(x)$ 表示不属于集合 $N(x)$ 的节点 x 的邻居节点。集合 $N_2(x)$ 的选取与网络中的节点密度相关，对于节点部署密集的网络，可以简单认为 $N_2(x)=N_1(x)$；而对于节点部署稀疏的网络，集合 $N_2(x)$ 则只包括节点 x 的 2-跳邻居节点。节点 x 监测集合 $N_1(x)$ 中的节点活动并且使用 k 元属性向量表示监测结果，然后监测结果在集合 $N_2(x)$ 中进行广播，这样节点 x 就获得一属性向量集，记为 $F_1(x)=\{f(x_i)=(f_1(x_i),f_2(x_i),\mathrm{L},f_k(x_i))^{\mathrm{T}}\mid x_i\in N_2(x)\}$，通过该属性向量集，节点 x 将获悉集合 $N_2(x)$ 中邻居节点的相关活动与属性。

2) 簇内信息收集阶段

令集合 $N_1^*(x)$ 是域 R^2 中的有界闭集，且集合中的所有节点都能直接被节点 x 监测到，即 $N_1^*(x)$ 是节点 x 所在的簇；节点 x 监测集合 $N_1^*(x)$ 中的节点活动并且使用 $k-1$ 元属性向量表示监测结果，然后监测结果在集合 $N_1^*(x)$ 中进行广播，这样节点 x 就获得一属性向量集，记为 $F_2(x)=\{f(x_i)=(f_1(x_i),f_2(x_i),\mathrm{L},f_{k-1}(x_i),f_k(x_i))^{\mathrm{T}}\mid x_i\in N_1^*(x)\}$，其中，$f_k(x_i)$ 是节点 x_i 的传感数据。通过该属性向量集，节点 x 将获悉集合 $N_1^*(x)$ 中邻居节点的相关活动与属性。

5.2　信息过滤阶段

集合 $N_1(x)$ 和 $N_2(x)$ 的信息收集完成之后，节点 x 将会获得一个表示邻居节点集合 $N_2(x)$ 活动属性的数据集 $F_1(x)$。然而，在集合 $N_1(x)$ 内部可能存在攻击者改变并转发集合 $N_2(x)-N_1(x)$ 中的一个或多个节点的监测结果。因此，为了更精确地检测外部攻击行为，节点 x 应尽可能多地过滤攻击者产生的异常结果。

基于对直接邻居节点的监测，节点 x 给其每个 1-跳的邻居节点 $x_i\in N_1(x)$ 赋予一个信任值，信任值 $T(x_i)$ 的取值范围为 $0\le T(x_i)\le 1$，信任值 $T(x_i)$ 越趋近于 1，说明节点 x_i 越可能是普通节点。考虑到临近节点之间行为的相似性，可以依据节点 x_i 与其相邻节点之间的活动属性偏差来计算信任值 $T(x_i)$。本文算法中的基于信任的信息过滤协议描述如下：记 $F_1(x)=\{f(x_i)=(f_1(x_i),f_2(x_i),\mathrm{L},f_k(x_i))^{\mathrm{T}}\mid x_i\in N_2(x)\}$ 为 $N_1(x)$ 的属性向量集，令 $\overline{m}_s$ 和 $\overline{s}_s$ 分别表示 $F_1(x)$ 的第 s 元集合 $F_{1,s}(x)=\{f_k(x_i)\mid x_i\in N_1(x)\}$ 的采样均值和采样标准差，计算如下：

$$\hat{m}_s=\frac{1}{n_1}\sum_{i=1}^{n_1}f_s(x_i)\text{，}\ \hat{s}_s=\mathrm{MAD}(f_j(x_i))$$

其中，n_1 是 $N_1(x)$ 中的节点数，$\mathrm{MAD}(T)=\mathrm{med}(|T-\mathrm{median}(T)|)$。

节点 x 首先对数据集 $F_{1,s}(x)(1\le k\le q)$ 进行标准化并计算绝对值以获得 $F_{1,s}(x)=\{f_s'(x_i)\,|\,x_i\in N_1(x)\}$，其中，$f_s'(x_i)=\left|\frac{f_s(x_i)-\hat{m}_s}{\hat{s}_s}\right|$。对于每一个节点 $x_i\in N_1(x)$，节点 x 计算其最大属性元 $f_M'(x_i)=\max(f_s'(x_i)\,|\,1\le k\le q)$，这样可以找到节点 x_i 与其相邻节点的最大绝对偏差，节点 x 赋予节点 x_i 的信任值为

$$T(x_i)=\frac{f_M^m}{f_M'(x_i)}$$

其中，$f_M^m=\min\{f_M'(x_i)\,|\,x_i\in N_1(x)\}$。

在算法中，当节点 $x_{sT}(1\le T\le t)$ 的信任值满足 $T(x_{sT})=\max\{T(x_{sr})\,|\,1\le r\le t\}$，$T(x_{sT})\ge T_{\min}$ 时，就认为节点 x_{sT} 是节点 x_s 的可靠中继。在实际应用中，$T_{\min}$ 的取值范围为 $0.1f_M^m\le T_{\min}\le 0.2f_M^m$，可以根据实际检测效果进行动态调节。

如果 $x_s\in N_2(x)-N_1(x)$ 在 $N_1(x)$ 中无法找到可靠中继时，经过对 $F_1(x)$ 的信息过滤后，节点 x 将会丢弃来自节点 x_s 的信息而只考虑接收来自集 $\hat{N}_2(x)$ 中节点的信息，其中，$\hat{N}_2(x)\in N_2(x)$。$\hat{N}_2(x)$ 包含 $N_1(x)$ 中的节点 x 的直接邻居以及属于 $N_2(x)-N_1(x)$ 但在 $N_1(x)$ 中存在可靠中继的节点 x 的邻居节点。于是，节点 x 重新计算属性向量集如下：$\hat{F}_1(x)=\{f(x_i)=(f_1(x_i),f_2(x_i),\text{L},f_k(x_i))^{\mathrm{T}}\,|\,x_i\in\hat{N}_2(x)\}$。

5.3 集合 $\hat{N}_2(x)$ 内部攻击检测阶段

节点 x 通过对 5.2 节中的属性向量集 $\hat{F}_1(x)$ 的测试就可以检测内部攻击者，随机选择部分(如 30%)节点集合进行检测，然后根据检测结果进行动态调节，如未发现入侵节点则停止检测，否则根据实际情况加大节点检测比例。具体检测过程如下。

1) 计算采样集 $\hat{F}_1(x)$ 的 OGK 稳健估计的平均值 m 和协方差矩阵 S。

2) 计算 $f(x_i)\,|\,x_i\in\hat{N}_2(x)$ 的马氏平方距离和阈值 $q_0=d_0=\frac{c_k^2(a)\times\mathrm{med}(MD_1^2,MD_2^2,\text{L},MD_n^2)}{c_k^2(0.5)}$。

3) 比较并标记马氏平方距离大于阈值 q_0 的节点为攻击者，其他节点标记为普通节点。

4) 在邻居集 $N'(x)\supseteq\hat{N}_2(x)$ 中广播 3)的结果，选择较大的邻居集 $N'(x)$ 的目的是确保更多节点参与投票；同时，节点 x 会收到其他广播结果并在 $N_1^*(x)$ 中登记邻居节点的选票。

5) 节点 x 登记广播结果后，统计并标识 $N_1^*(x)$ 中的马氏平方距离大于阈值 q_0 的节点为内部攻击者。

①如果节点 x 是簇头，它将把确认为攻击者的节点移除该簇并通知基站。

②如果节点 x 是普通节点，它将会等待并观察簇头节点是否移除被确认的攻击者。

③如果簇头节点不作为，节点 x 将指出攻击者并在簇内广播一警示信息；如果节点 x 从簇内其他节点收到警示信息后，它将会进行统计。

a) 如果簇内大多数节点，如 75%以上确认指定的节点为内部攻击者，簇头节点将会被视为内部攻击者并被移除该簇，启动簇头选择算法选择新的簇头并移除指定的内部攻击者；

b) 否则，认为节点 x 检测错误。

④ 如果簇头节点试图移除没有被簇内其他节点检测为内部攻击者的普通节点时，簇内其他节点将会同谋采用③的方式移除簇头节点。

5.4 簇 $N_1^*(x)$ 内部攻击检测阶段

节点 x 通过对 5.2 节中的属性向量集 $F_1^*(x)$ 的测试就可以检测内部攻击者，随机选择部分(如 30%)节点集合进行检测，然后根据检测结果进行动态调节，如未发现入侵节点则停止检测，否则根据实际情况加大节点检测比例。具体检测过程如下。

1) 计算采样集 $F_1^*(x)$ 的 OGK 稳健估计的平均值 m 和协方差矩阵 S；

2) 计算 $f(x_i)\,|\,x_i\in N_1^*(x)$ 的马氏平方距离和阈值 $q_0=d_0=\frac{c_k^2(a)\times\mathrm{med}(MD_1^2,MD_2^2,\text{L},MD_n^2)}{c_k^2(0.5)}$。

3) 比较并标记马氏平方距离大于阈值 q_0 的节点为攻击者，其他节点标记为普通节点。

4) 在邻居集 $N_1^*(x)$ 中广播 3)的结果，同时，节点 x 将会接收来自 $N_1^*(x)$ 中其他节点的广播结果。

5) 节点 x 登记广播结果后，统计并标识 $N_1^*(x)$ 中的马氏平方距离大于阈值 q_0 的节点为内部攻击者。

① 如果节点 x 是簇头，它将把确认为攻击者的节点移除该簇并通知基站。

② 如果节点 x 是普通节点，它将会等待并观察簇头节点是否移除被确认的攻击者。

③ 如果簇头节点不作为，节点 x 将把簇头作为内部攻击者并在簇内广播一警示信息。

a) 如果簇内大多数节点，如 75%以上都广播了警示信息，则簇头节点将会被视为内部攻击者并被移除该簇，启动簇头选择算法选择新的簇头并移除指定的内部攻击者；

b) 否则，认为节点 x 检测错误。

④ 如果簇头节点试图移除没有被簇内其他节点检测为内部攻击者的普通节点时，簇内其他节点将会同谋采用③的方式移除簇头节点。

5.5 簇 $N_1^*(x)$ 数据融合阶段

在对簇 $N_1^*(x)$ 进行攻击检测阶段，每个节点对采样集 $F_1^*(x)$ 进行 OGK 稳健估计以获取平均值 m 和协方差矩阵 $\boldsymbol{S}$。提出使用稳健估计计算的传感数据的均值作为数据融合的结果。然而，一旦在采样集 $F_1^*(x)$ 中发现了内部攻击者时，该攻击者的传感数据将要被丢弃并在去除攻击者后的采样集 $\overline{F}_1^*(x)$ 中重新进行稳健估计以获取新的数据融合结果。数据融合具体过程如下。

1) 如果没有发现内部攻击者，簇头将使用稳健计算的方法获取传感数据的均值。

2) 如果发现某节点为内部攻击者，则把该节点的属性向量从采样集 $F_1^*(x)$ 中除去并计算新的采样集 $\overline{F}_1^*(x)$ 的稳健估计均值作为数据融合结果。

由于攻击检测保证了簇 $N_1^*(x)$ 中所有节点能获得准确的相同采样集 $F_1^*(x)$，因此，通过稳健估计计算的传感数据均值也是完全一致的，所以当簇头节点是内部攻击者时，其所得到的数据融合结果由于与其他节点计算的融合结果不同而被识别出来。簇内其他节点把簇头视为攻击者并违背其指令。通过这种方式，可以有效地确保数据融合结果的正确性。

6 性能仿真与分析

在本节中主要从错误警报率、检测精度和数据融合结果与真实值之间的偏差这 3 个方面分析 EIDSART 算法的性能，IADART 方案是目前传感网络领域相对成熟的入侵检测技术，是一种采用静态阈值和粗略分集进行入侵检测的方法，另外，该方案不提供数据融合的功能，文中用它作为 EIDSART 的错误警报率、检测精度的对比项。

本文采用集成了多变量统计分析分组的 Mathematic 8.0 软件进行 IADART 和 EIDSART 算法的仿真。仿真中，采用文献[18]中的传感器网络分层模型，对于网络中的仿真节点 $x_i \in N(x)$，其邻居节点集合 $N_1^*(x)$ 只包括所在簇内的 1-跳邻居；邻居节点集合 $N_1(x)$ 包括网络中的所有 1-跳邻居；邻居节点集合 $N_2(x)$ 则包括网络中所有的 1-跳和 2-跳邻居。

不失一般性，在稀疏矩阵的环境中进行相关测试与实验，其中，簇邻居节点集 $N_1^*(x)$ 包含 10 个节点，1-跳邻居节点集 $N_1(x)$ 包含 25 个节点，邻居节点集 $N_2(x)$ 包含 40 个节点。仿真中，每个节点被建模成具有分组传输率、分组丢失率、传输时延和传感数据等 4 个属性的向量，考虑到效率问题，仿真只对 EIDSART 算法的传感数据属性进行检测。节点的属性生成值由仿真软件 Mathematic 8.0 中的随机实数函数产生，表示为 $N_4(m_i, \boldsymbol{S})$，其中 $m_i = (m_1, m_2, m_3, m_4)$，变量的协方差 $\boldsymbol{S}$ 由标准偏差 $s_i = (s_1, s_2, s_3, s_4)$ 表示，相关系数矩阵 $\boldsymbol{r} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$ 显示节点的 4 个属性之间没有相关性。最后，确定的协方差为 $\boldsymbol{S} = (\boldsymbol{S}_{ij}) = r_{ij} s_i s_j$。下面给出不同邻居节点集和数据融合结果偏差的仿真设置和结果，所有仿真重复次数为 50 次，结果为平均值。

1) 对于邻居节点集 $N_2(x)$ 包含 40 个节点的多媒体传感器网络系统，设异常节点的概率区间为 0~30%，对 IADART 和 EIDSART 算法进行错误警报率和检测精度的仿真比较。对于普通节点，属性生成值为 $m_i = (10, 20, 50, 100)$，$s_i = (1, 5, 10, 20)$；对于一个属性即传感数据出现异常的节点，属性生成值为 $m_i = (10, 20, 50, 150)$，$s_i = (1, 5, 10, 30)$；而 4 个属性均出现异常的节点属性生成值为 $m_i = (20, 40, 100, 150)$，$s_i = (1, 5, 10, 30)$。仿真结果如图 5 和图 6 所示。

图 5 显示了在 $N_2(x) = 40$，异常节点区间为 0~30%时，IADART 和 EIDSART 算法各在一个属性和 4 个属性出现异常的情况下错误警报率的趋势图。结果表明：多属性异常时两算法的误报率比单属性异常要低，但在同样条件下，EIDSART 算法的误报率比 IADART 明显低得多；图 6 显示了在 $N_2(x) = 40$，

异常节点区间为 0~30%时，IADART 和 EIDSART 算法各在一个属性和 4 个属性出现异常的情况下异常检测精度的趋势图。结果表明：多属性异常时两算法的检测精度比单属性异常要高，但在同样条件下，EIDSART 算法的检测精度比 IADART 算法要高。

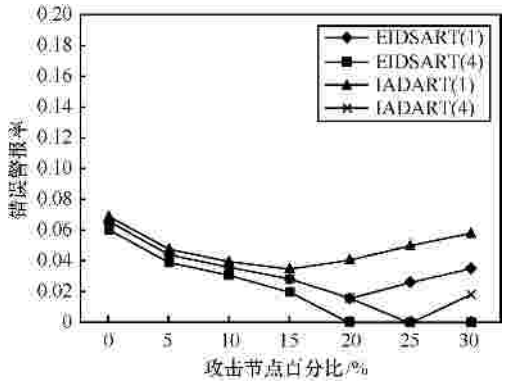

图 5 $N_2(x)=40$ 两算法错误警报率趋势

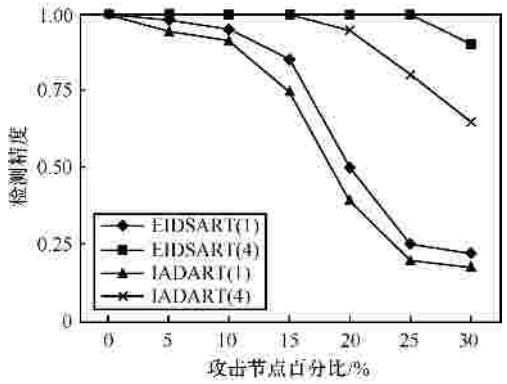

图 6 $N_2(x)=40$ 两算法检测攻击精度趋势

2）对于邻居节点集 $N_1^*(x)$ 包含 10 个节点的多媒体传感器网络系统，设异常节点的概率区间为 0~30%，对 IADART 和 EIDSART 算法进行错误警报率和检测精度的仿真比较。对于普通节点，属性生成值为 $m_i=(10,20,50,100)$，$s_i=(1,5,10,20)$；对于一个属性即传感数据出现异常的节点，属性生成值为 $m_i=(10,20,50,150)$，$s_i=(1,5,10,20)$；而 4 个属性均出现异常的节点属性生成值为 $m_i=(20,40,100,150)$，$s_i=(1,5,10,20)$。仿真结果如图 7 和图 8 所示。

图 7 显示了在 $N_1^*(x)=10$，异常节点区间为 0~30%时，IADART 和 EIDSART 算法各在一个属性和 4 个属性出现异常的情况下错误警报率的趋势图。结果表明：多属性异常时两算法的误报率比单属性异常要低，但在同样条件下，EIDSART 算法的误报率比 IADART 算法要低；图 8 显示了在 $N_1^*(x)=10$，异常节点区间为 0~30%时，IADART 和 EIDSART 算法各在一个属性和 4 个属性出现异常的情况下异常检测精度的趋势图。结果表明：多属性异常时两算法的检测精度比单属性异常要高，但在同样条件下，EIDSART 算法的检测精度比 IADART 算法要高。

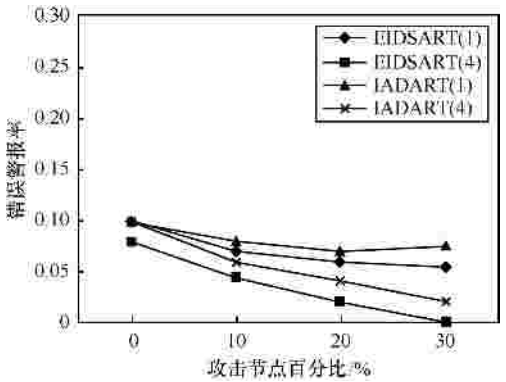

图 7 $N_1^*(x)=10$ 两算法错误警报率趋势

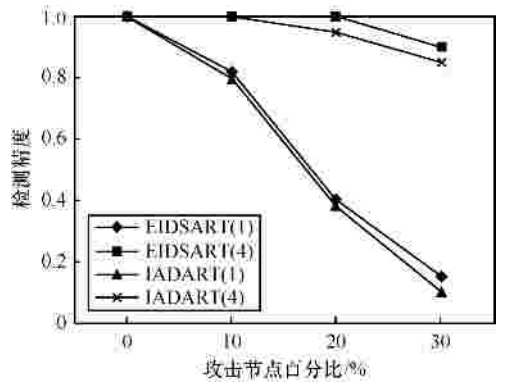

图 8 $N_1^*(x)=10$ 两算法检测攻击精度趋势

3) 在 $N_1^*(x)=10$，异常节点数为 0、1、2、5 的 4 种情况下对 EIDSART 算法获得的传感数据融合结果与真实传感数据结果之间的偏差进行仿真测试，设普通节点传感数据的标准差分为 1%、0.2% 和 0.02% 3 种情况，仿真结果如图 9 所示。

图 9 显示了 4 种情况下 EIDSART 算法数据融合结果与真实传感数据结果之间的偏差。结果显示：网内异常节点数对数据融合的结果有较大的影响，异常节点越多，数据融合结果偏差越大，另外，真实传感数据的标准差越大，数据融合的结果偏差也越大。由于 EIDSART 算法是异常检测后进行的数据融合，因此数据融合结果接近异常节点数为 0 的情况，近似于真实情况。

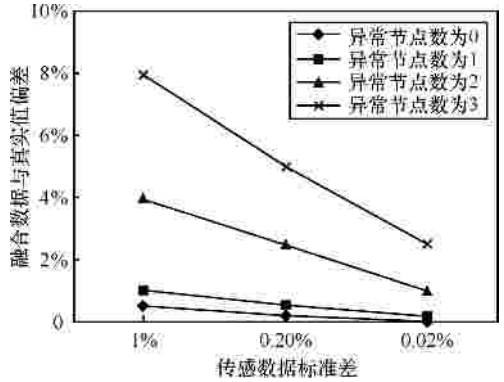

图 9　EIDSART 数据融合结果与真实值偏差

7　结束语

本文提出了一种改进有效的无线多媒体传感器网络攻击检测和数据融合算法 EIDSART，与已有成果 NIDAWSN[16]和目前该领域相对成熟的入侵检测技术 IADART 相比，提出的方案在有效降低错误警报率和提高攻击检测精度的基础上，能够获得精确的数据融合结果，减少了网络通信开销并节约了网络资源，为无线多媒体传感器网络乃至物联网的高效应用提供了有价值的参考。

表 1 对 3 个相关方案的优缺点进行了对比，EIDSART 算法中的动态阈值和灵活的邻居集选择使得算法的检测精度、适应性和误报率都得到了极大的优化。同时，数据融合阶段采用的健壮性估计方法有效地解决了攻击检测后精确数据融合的问题，大大节约了网络中的数据通信量和无线多媒体传感器网络宝贵的网络及节点资源。本文的下一步的工作是：对无线多媒体传感器网络中各种形式攻击检测预防机制、产生的异常以及有针对性的检测与应对方法等重要问题进行详细的研究与探讨。

表 1　IADART、NIDAWSN 和 EIDSART 性能比较

对比项	ADART	NIDAWSN	EIDSART
错误警报率	高	低	低
检测精度	低	一般	高
精确性	低	高	高
计算量	小	大	小
数据融合功能	无	有	有
资源消耗	多	多	少
算法适应性	弱	较强	强

参考文献：

[1] AKYILDIZ I F, MELODIA T, CHOWDURY K R, *et al*. Wireless multimedia sensor networks: a survey[J]. IEEE Wireless Communications, 2007, 14(6):32-39.

[2] GHIASI S, SRIVASTAVA A, YANG X, *et al*. Optimal energy aware clustering in sensor networks[J]. Sensors Magazine, 2002, 16(1): 258-269.

[3] MARONNA R A, MARTIN R D, YOHAI V J. Robust Statistics: Theory and Methods[M]. Wiley Publisher, 2006. 205-208.

[4] PERRIG A, STANKOVIC J. Security in wireless sensor networks[J]. Communications of the ACM-Wireless Sensor Networks, 2004,47(6): 53-57.

[5] PAI H T, DENG J, HAN Y S. Time-slotted voting mechanism for fusion data assurance in wireless sensor networks under stealthy attacks[J]. Computer Communications, 2010, 33(13):1524-1530.

[6] MEI A, PANCONESI A, RADHAKRISHNAN J. Unassailable sensor networks[A]. Proceedings of the 4th International Conference on Security and Privacy in Communication Networks[C]. Istanbul, Turkey, 2008. 241-248.

[7] BALLARDIN F, MERRO M. A calculus for the analysis of w reless network security protocols[A]. Proceedings of the 7th International Conference on Formal Aspects of Security and Trust[C]. Pisa, Italy, 2010. 206-222.

[8] GUO Y H, PERREAU S. Detect DDoS flooding attacks in mobile ad hoc networks[J]. International Journal of Security and Networks, 2010, 5(4):259-269.

[9] TRIPATHY S, NANDI S. Defense against outside attacks i wireless sensor networks[J]. Computer Communications, 2008, 31(4): 818-826.

[10] PAPADIMITRATOS P, LUO J, HUBAUX J P. A randomized countermeasure against parasitic adversaries in wireless sensor networks[J]. IEEE Journal on Selected Areas in Communications, 2010, 28(7): 1036-1045.

[11] HUANG Y A, LEE W. A cooperative intrusion detection system for ad hoc networks[A]. Proceedings of the 1st ACM Workshop on Security of Ad Hoc and Sensor Networks[C]. Fairfax, Virginia, 2003. 135-147.

[12] ONAT I, MIRI A. An intrusion detection system for wireless sensor networks[A]. 2005 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications[C]. 2005. 253-259.

[13] MOSTARDA L, NAVARRA A. Distributed intrusion detection systems for enhancing security in mobile wireless sensor tworks[J]. International Journal of Distributed Sensor Networks, 2008, 4(2):83-109.

[14] PASCHALIDIS I C, CHEN Y. Statistical anomaly detection with sensor networks[J]. ACM Transactions on Sensor Networks, 2010,7(2): 1-23.

[15] CONTI M, PIETRO R D, MANCINI L V, *et al*. Emergent properties: detection of the node-capture attack in mobile wireless sensor networks[A]. Proceedings of the First ACM Conference on Wireless Network security[C]. Alexandria, VA, USA, 2008. 214-219.

[16] RONG C, EGGEN S, CHENG H B. A novel intrusion detection algorithm for wireless sensor networks[A]. The 2nd International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems[C]. Chennai, India, 2011.1-7.

[17] LIU F, CHENG X Z, CHEN D C. Insider attacker detection in wirelss sensor networks[A]. INFOCOM 2007 the 26th IEEE International Conference on Computer Communications[C]. 2007. 1937-1945.

[18] CHENG H B, RONG C, YANG G. Design and analysis of a secure routing protocol algorithm for wireless sensor networks[A]. The 25th IEEE International Conference on Advanced Information Networking and Applications[C]. Biopolis, Singapore, 2011.475-480.

作者简介：

程宏兵（1979-），男，江西彭泽人，南京大学博士后、副教授，主要研究方向为物联网与传感器网络、信息安全与云计算等。

容淳铭（1969-），男，广东江门人，挪威斯塔万格大学教授、博士生导师，主要研究方向为传感器网络与通信、信息安全、云计算等。

黄晓（1970-），男，江苏启东人，博士，江苏移动通信有限公司高级工程师，主要研究方向为物联网与传感器网络、网络安全与信息安全等。

EGGEN Skjalg（1980-），男，挪威斯塔万格人，斯塔万格大学博士生，主要研究方向为传感器网络与通信、信息安全、云计算等。

曾庆凯（1963-），男，安徽来安人，南京大学教授、博士生导师，主要研究方向为信息安全、操作系统安全、网络安全、安全评估与测试。