A Group-Based Intrusion Detection Scheme in Wireless Sensor Networks *

WANG Ying[1*], *LI Guorui*[2]

(1. *Department of Information Engineering, Qinhuangdao Institute of Technology, Qinhuangdao Hebei* 066000, *China*;
2. *College of Computer Science and Technology, Beijing University of Technology, Beijing* 100124, *China*)

Abstract: In order to detect various kinds of malicious attacks in wireless sensor networks, we proposed a group based intrusion detection scheme which meets the efficient, lightweight and flexible property in this paper. The sensor networks are partitioned into many groups in which the sensors in each group are physically close to each other and equipped with the same sensing capability. Our intrusion detection algorithm takes into consideration of multiple attributes of the sensor nodes simultaneously to detect malicious attackers precisely. We show through experiments with real data that our algorithm can decrease the false alarm rate and increase the detection accuracy rate compared with existing intrusion detection schemes while lowering the power consumption.

Key words: wireless sensor networks; security; intrusion detection scheme; grouping algorithm
EEACC: 7230; 6150P

基于分组的无线传感器网络入侵检测方案 *

王 颖[1*], 李国瑞[2]

(1. 秦皇岛职业技术学院信息工程系, 河北 秦皇岛 066000;
2. 北京工业大学计算机学院, 北京 100124)

摘 要: 为了有效地检测无线传感器网络所面临的各种恶意攻击, 提出了一种轻量、高效、灵活的分组入侵检测方案。在该方案中, 整个传感器网络被划分成若干物理位置临近、具有相似观测结果的分组, 组内各传感器节点同时观测其它节点的多个属性, 以便精确地检测各种攻击行为。实验结果表明, 与传感器网络中现有的入侵检测方案相比, 本方案具有较低的误报率和较高的检测精度, 同时消耗更少的能量。

关键词: 无线传感器网络; 安全; 入侵检测; 分组算法
中图分类号: TN915.08 **文献标识码:** A **文章编号:** 1004-1699(2009)06-0878-05

无线传感器网络综合了微电子、嵌入式计算、现代网络及无线通信、分布式信息处理等先进技术, 能够协同地实时监测、感知和采集网络覆盖区域中各种环境或监测对象的信息, 并将处理后的结果通过无线多跳的方式发送给用户。由于具有自组织、高容错、高可靠、成本低廉、部署方便等特性, 它被广泛应用于环境监测、灾难响应、军事侦察、智能建筑以及工业质量控制等各个领域[1]。

传感器节点间使用无线通信链路进行数据传输, 节点的计算、存储、能量等性能都十分有限, 这些特性使得无线传感器网络很容易受到各种攻击方式的威胁。目前无线传感器网络的安全保障方案可以分为以下两大类: 基于预防的方案和基于检测的方案。基于检测的方案主要用在基于预防的方案失效时检测并隔离恶意的攻击节点, 它按照检测方式的不同可分为以下两种类型: 基于签名的检测方案和基于异常识别的检测方案。

Doumit 等人提出了一种基于安全级别和随机学习过程的入侵检测方案, 通过利用与位置信息相关的安全级别以及隐式马尔可夫模型来检测未知的

基金项目: 北京市教育委员会科技发展基金项目资助(KZ200610005003)
收稿日期: 2008-12-22 **修改日期:** 2009-03-30

异常行为[2]。Su 等人提出了 eHIP 即能量有效的混合入侵预防系统来提高分簇传感器网络的安全性[3]。此系统包括基于认证的入侵预防子系统 AIP 和基于合作的入侵检测子系统 CID,从而支持分簇传感器网络中不同级别的安全需要并降低能量的消耗。Agah 等人利用基于对抗的游戏理论来寻找传感器网络中最脆弱的节点并加以保护[4]。Silva 等人定义了用以检测入侵发生的多个规则,并在触发数超过阈值时自动报警[5]。内部攻击检测方案是最新提出的一种检测方案,它在识别节点行为时无需正常或异常行为的规则库,并且在缺失数据的情况下仍具有较高的精确度和较低的误报率[6],本方案将与之进行性能的比较。

1　基于分组的入侵检测方案

基于分组的入侵检测方案包括两个阶段:分组阶段和组内入侵检测阶段。具体的方案如下:

1.1　假设

基于分组的入侵检测方案适用于平面结构的无线传感器网络,无需簇头等特殊节点的参与。假设在传感器网络的部署初期网络中并不存在异常节点,所有传感器节点同时开始运行异常检测算法,在分组过程中各节点所监测的数据不存在大幅度或者异常的变化。

1.2　σ 分组算法

首先,整个传感器网络被划分成若干个分组,组内各传感器节点物理位置临近,并且采集的观察数据值接近,不超过阈值 σ。这一特性使得本方案比其它入侵检测方案能够得到更加精确的结果。文献[7]表明,σ 分组问题是一个 NP 完备问题,因此除非 P = NP,否则不可能在多项式时间内得到最优的结果。σ 分组算法中使用到的符号如表 1 所示。

表 1　δ 分组算法中使用到的符号

符号	含义
grouped	初始值为假的布尔变量,表示传感器节点是否已加入组
r_i	节点 i 所属组的根节点
p_i	节点 i 所属组中 i 的父节点
f_i	节点 i 采集到的数据值
f_{r_i}	根节点 r_i 采集到的数据值
$d(f_i, f_{r_i})$	f_i 和 f_{r_i} 之间的欧几里德距离
$hops(i, r_j)$	节点 i 和根节点 r_j 之间的跳数
h	组内预设的最大跳数值
$T_{random}(i)$	节点 i 初始化加入组消息等待的随机时间

在传感器节点 i 上,σ 分组算法如下所示:

```
If (Not grouped) And (receive < "Grouping", r_j, f_{r_j} >)
Then
    If (d(f_i, f_{r_j}) ≤ δ/2) And (hops(i, r_j) ≤ h/2) Then
        r_i := r_j;
        grouped := True;
        p_i := j;
        f_{r_i} = f_{r_j};
        broadcast < "Grouping", r_i, f_{r_i} >;
    Endif
Else If (Not grouped) And (T_random(i) is up) Then
        r_i : i;
        grouped := True;
        p_i := i;
        f_{r_i} := f_i;
        broadcast < "Grouping", r_i, f_{r_i} >;
    Endif
Endif
```

每个传感器节点 i 等待一个随机时间 $T_{random}(i)$ 初始化加入组消息。如果它在这段时间内从邻居节点接收到一个加入组消息并且尚未加入任何组,则计算它所采集的数据 f_i 与接收到的加入组消息中组长节点 r_j 所采集的数据 f_{r_j} 之间的欧几里德距离,以及这两个节点间的跳数,如果 $d(f_i, f_{r_j}) \leq \delta/2$ 并且 $\mathrm{hops}(i, r_j) \leq h/2$,则根据三角不等式可知组内任意两节点间采集数据的差距不超过 δ,并且它们之间的距离不超过 h。因此,同一组内的传感器节点物理空间上位置临近,并且具有相近的观测数据值。如果此传感器在 $T_{\mathrm{random}}(i)$ 时间后仍未加入任何组,它便构造一个新组并将自己作为组长节点,同时向邻居节点广播加入组消息。

在上述分组算法中,阈值 δ 需要根据无线传感器网络的具体应用和所采集数据的类型进行确定,而组内预设的最大跳数值 h 可以根据网络的规模进行确定。对于竞争时间 $T_{\mathrm{random}}(i)$ 的选择应该保证传感器节点在构造新组前等待足够长的时间。$T_{\mathrm{random}}(i)$ 可以通过利用随机数生成函数乘以传感器节点间平均消息传递时间 t 而计算得到。最多经过 $\max(T_{\mathrm{random}}(i)) + ht/2$ 时间后,网络中的所有传感器节点都将完成 δ 分组算法。

1.3　入侵检测算法

在应用分组算法划分的每个分组内,根据节点 ID、剩余能量、距离组长节点的跳数等属性可将此分组划分成若干相同大小的子组。假设分组 G_i 内有 N_i 个传感器节点,每个子组内含有 N_S 个节点,则此分组被划分为 $\lfloor N_i / N_S \rfloor$ 个子组。每个子组内的传

感器节点使用下述入侵检测算法同时监视整个分组。所有的$\lfloor N_i/N_S \rfloor$个子组轮流负责监视整个分组,以便降低整体的能量消耗,从而延长整个传感器网络的寿命。当监视节点发现组内某节点的数据与其他节点的数据间存在较大的差异时,它便广播以下格式的警告消息。

警告类型,异常节点,监视节点,时间戳

当某个传感器节点在一段时间内接收到来自同一个监视节点的 n_1 个警告消息或者关于同一个异常节点的 n_1 个警告消息后,它便将自身设置成混杂模式并亲自监测这个异常节点或者监视节点。如果它发现了 n_2 次关于被监测节点的异常行为,并且 $n_1 w_1 + n_2 w_2$ 大于预设的阈值 θ,其中 w_1 和 w_2 分别为其它节点和自身检测到异常行为的权重,则可以确定被监测节点为异常节点或者恶意报警节点,因此应更新路由表或者删除其对应的密钥以便将它从网络中隔离出去。

监视传感器节点可以收集表 2 中列出的各种信息以便检测各种类型的网络攻击行为:

表 2 收集的信息以及相应检测的攻击类型

收集的信息	检测的攻击类型
采集的数据值	伪造数据攻击
发包率	能量耗尽攻击
丢包率	选择转发攻击,黑洞攻击
包不匹配率	篡改消息攻击
收包率	采集点漏洞攻击
发包能量	Hello 消息攻击,虫洞攻击

假设传感器节点的多维监视属性构成的多维向量为 $f(x_i) = \{f_1(x_i), f_2(x_i), \cdots, f_q(x_i)\}$,其中 q 为监视属性的个数。则组内所有 n 个传感器节点 $x_1, \cdots, x_n$ 的多维向量构成了矩阵 $F(x) = \{f(x_1), f(x_2), \cdots, f(x_n)\}$。

假设所有的 $f(x_i)(x_i \in \{x_1, \cdots, x_n\})$ 服从多维正态分布 $N_q(\mu, \sum)$,其中 μ 和 $\sum$ 分别为均值和协方差,则马氏距离的平方 $(f(x_i) - \mu)^T \sum^{-1} (f(x_i) - \mu)$ 服从自由度为 q 的卡方分布 χ_q^2,因此 $f(x_i)$ 满足 $(f(x_i) - \mu)^T \sum^{-1} (f(x_i) - \mu) > \chi_q^2(\alpha)$ 的概率为 α,其中 $\chi_q^2(\alpha)$ 为检验水准为 α、自由度为 q 的卡方分布值。

假设 $\hat{\mu}$ 和 $\hat{\sum}$ 分别为均值 μ 和协方差 $\sum$ 的估计,则 $f(x_i)$ 满足 $(f(x_i) - \hat{\mu})^T \hat{\sum}^{-1} (f(x_i) - \hat{\mu}) > \chi_q^2(\alpha)$ 的概率可以估计为 α。令马氏距离 $d(x_i) = ((f(x_i) - \hat{\mu})^T \hat{\sum}^{-1} (f(x_i) - \hat{\mu}))^{1/2}$,如果 $d^2(x_i) > \chi_q^2(\alpha)$,那么传感器节点 x_i 将被识别为异常节点。

均值 μ 和协方差 $\sum$ 的估计不能简单的使用均值和协方差公式

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} f(x_i),$$

$$\hat{\sum} = \frac{1}{n-1}\sum_{i=1}^{n} (f(x_i) - \hat{\mu})(f(x_i) - \hat{\mu})^T$$

进行计算,因为异常数据会极大地干扰 μ 和 $\sum$ 的估计,从而产生错误的异常数据识别结果,本方案使用 OGK 估计均值 μ 和协方差 $\sum$[8]。

一维随机变量 $Y = \{y_1, y_2, \cdots, y_n\}$ 的均值估计 $\hat{\mu}$ 和协方差估计 $\hat{\sum}$ 的计算方法如下:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} y_i W(v_i)}{\sum_{i=1}^{n} W(v_i)}, \text{其中 } v_i = \frac{y_i - \mu_0}{\sigma_0}$$

$$\hat{\sigma}^2 = \frac{\sigma_0^2}{n}\sum_{i=1}^{n} \rho\left(\frac{y_i - \hat{\mu}}{\sigma_0}\right),$$

其中 μ_0 和 σ_0 分别为均值和中位数绝对偏差 MAD = median(| Y - median(Y) |),加权函数 $W(x) = (1 - (x/c_1)^2)^2 I(|x| \le c_1)$,$\rho(x) = \min(x^2, c_2^2)$,$c_1 = 4.5$,$c_2 = 3$。

多维随机变量 $F(x) = \{f(x_1), f(x_2), \cdots, f(x_n)\}$ 的均值估计 $\hat{\mu}$ 和协方差估计 $\hat{\sum}$ 的计算方法如下,其中 $f(x_i) = (f_1(x_i), f_2(x_i), \cdots, f_q(x_i))^T$:

(1) 计算 $G(x) = \{g(x_1), g(x_2), \cdots, g(x_n)\}$,其中 $g(x_i) = P^{-1} f(x_i)$,$P = \mathrm{diag}(\hat{\sigma}(行(x)), \hat{\sigma}(行(x)), \cdots, \hat{\sigma}(行(x)))$,行(x) 为 $F(x)$ 的第 j 行元素构成的向量。

(2) 计算 $q \times q$ 维矩阵 R:

$$R_{jk} = \begin{cases} \frac{1}{4}[\hat{\sigma}^2(G_j + G_k) - \hat{\sigma}^2(G_j - G_k)] & j \ne k \\ 1 & j = k \end{cases}$$

(3) 使用谱分解计算 $R = Q\Lambda Q^T$,其中 Q 是 R 的特征矩阵,Λ 是由 R 的特征值构成的对角矩阵。

(4) 计算 $H = \{h(x_i) \mid h(x_i) = Q^T g(x_i)\}$。然后计算 $\Delta = (\hat{\mu}(H_1(x)), \hat{\mu}(H_2(x)), \cdots, \hat{\mu}(H_q(x)))^T$ 和 $\Gamma = \mathrm{diag}(\hat{\sigma}^2(H_1(x)), \hat{\sigma}^2(H_2(x)), \cdots, \hat{\sigma}^2(H_q(x)))$。

(5) 令 $V = PQ$。则健壮的均值估计 $\hat{\mu} = V\Delta$,健壮的协方差估计 $\hat{\sum} = V\Gamma V^T$。

检测方法中使用马氏距离度量的原因是因为它可以包含多维随机变量之间的相关性,从而可以得到更加精确度的检测结果[9]。选择 OGK 估计的原因是因为即使在某些数据缺失的情况下,它仍然能够保证具有较高的溃点,而且计算量小,速度快[6]。

2　性能分析

实验中将使用部署在 Intel Berkeley 实验室的 54 个 Mica2Dot 传感器节点于 2004 年 2 月 28 日至 4 月 5 日所采集到的真实数据。这些数据每隔 31 s 采集一次,其中包括湿度、温度、光照和电压值。在这些数据的基础上增加服从正态分布的噪声值以模拟异常传感器节点的数据值。通过使用 R 2.6.0 统计计算软件和 robustbase、rrcov 工具包,将本方案和内部攻击检测方案进行比较。

图 1 展示了使用 δ分组算法对 Intel Berkeley 实验室的 54 个 Mica2Dot 传感器节点进行分组的一个实例。

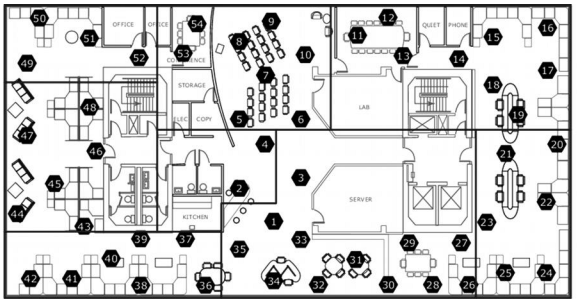

图 1　δ分组算法的实例

图 2 和图 3 分别显示了内部攻击检测方案与基于分组的入侵检测方案在统计记录数 $k=30$ 的条件下进行检测的结果。从图中可以看出,内部攻击检测方案触发了 4 条报警消息,分别对应着节点 4、6 和 10。然而,这 4 条记录都属于正常范围内的数据,被内部攻击检测方案误报为异常数据。图 3 显示了基于分组的入侵检测方案所做出的正确判断。内部攻击检测方案错误判断的原因在于传感器节点之间采集数据的差异会很大程度上影响入侵检测方案正确判断的结果。基于分组的入侵检测方案通过将具有相似观测值的传感器节点进行分组,有效地解决了上述问题。

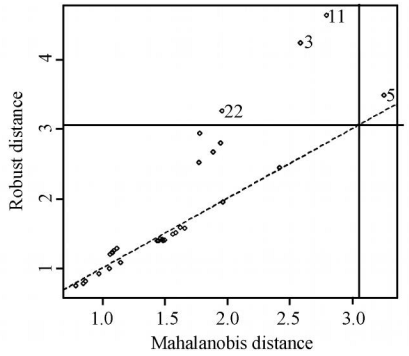

图 2　内部攻击检测方案异常检测的结果($k=30$)

误报率是正常传感器节点被误判为异常传感器节点的比率,检测精度是异常传感器节点被成功识别出的比率。它们是评测入侵检测算法性能的两个关键性指标。图 4 显示了基于分组的入侵检测方案与内部攻击检测方案误报率性能的比较,实验结果表明,基于分组入侵检测方案的误报率要远低于内部攻击检测方案,其原因在于分组算法有效地降低了分组内传感器节点间采集数据的差异,进而降低了检测算法的误报率。

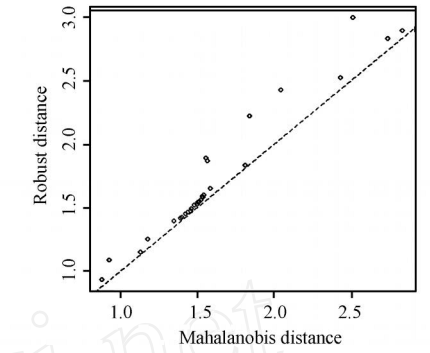

图 3　基于分组的入侵检测方案异常检测的结果($k=30$)

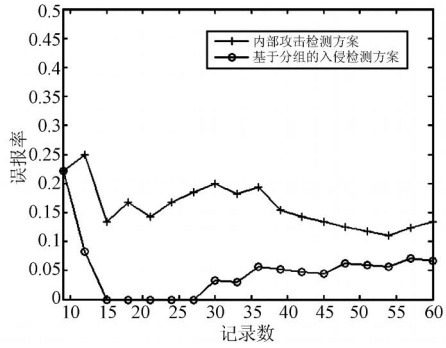

图 4　基于分组的入侵检测方案与内部攻击检测方案误报率性能的比较

图 5 显示了基于分组的入侵检测方案与内部攻击检测方案检测精度性能的比较。从图中可以看出,在统计记录数较少时,基于分组入侵检测方案的检测精度要远高于内部攻击检测方案。随着统计记录数的逐渐增加,这两种检测方案的检测精度趋于相同。

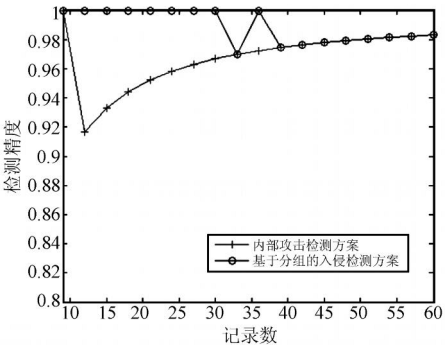

图 5　基于分组的入侵检测方案与内部攻击检测方案检测精度性能的比较

基于分组的入侵检测方案在降低误报率和提高

检测精度的同时,与内部攻击检测方案相比,还降低了传感器网络监测时所需的总体能耗。它通过入侵检测调度算法将监视能量的消耗分摊到每个传感器节点之上,组内的传感器节点间互相合作,周期性的轮换监视任务,其余的节点进入睡眠状态以节省能量,从而降低了总的监视能耗。

3 结论

本文中提出了一种适用于无线传感器网络的基于分组的入侵检测方案,用以有效地检测网络中的各种恶意攻击行为。实验结果表明,与现有的方案相比,本方案具有较低的误报率和较高的检测精度,同时消耗更少的能量用于监视整个网络。

参考文献:

[1] Akyildiz I, Su W, Sankarasubramaniam Y. Wireless Sensor Networks: a Survey[J]. Computer Networks, 2002, 38(4): 393-422.

[2] Doumit S, Agrawal D. Self-Organized Criticality & Stochastic Learning Based Intrusion Detection System for Wireless Sensor Network[C]// IEEE Military Communications Conference, Monterey, CA, USA, 2003: 609-614.

[3] Su W, Chang K, Kuo Y. eHIP: An energy-efficient Hybrid Intrusion Prohibition System for Cluster-Based Wireless Sensor Networks[J]. Computer Networks, 2007, 51(4): 1151-1168.

[4] Agah A, Das S, Basu K. Intrusion Detection in Sensor Networks: a Non-cooperative Game Approach[C]// IEEE International Symposium on Network Computing and Applications, Los Alamitos, CA, USA, 2004:343-346.

[5] Silva A, Martins M, Rocha B. Decentralized Intrusion Detection in Wireless Sensor Networks[C]// ACM International Workshop on Quality of Service & Security in Wireless and Mobile Networks, Montreal, QB, Canada, 2005:16-23.

[6] Liu F, Cheng X, Chen D. Insider Attacker Detection in Wireless Sensor Networks[C]// IEEE International Conference on Computer Communications, Anchorage, AK, USA, 2007: 1937-1945.

[7] Meka A, Singh A. Distributed Spatial Clustering in Sensor Networks[J]. Lecture Notes in Computer Science, 2006, 3896:980-1000.

[8] Maronna R, Martin R, Yohai V. Robust Statistics: Theory and Methods[M]. Wiley Publisher. 2006: 175-227.

[9] Hodge V, Austin J. A Survey of Outlier Detection Methodologies[J]. Artificial Intelligence Review, 2004, 22 (2):85-126.

王 颖(1979-),女,天津人,秦皇岛职业技术学院教师,硕士。主要研究方向为无线传感器网络和信息安全,wyqhd@hotmail.com。

李国瑞(1980-),男,山西夏县人,北京工业大学计算机学院博士研究生。主要研究方向为无线通信网络和信息安全,现已发表学术论文 20 余篇,liguorui@emails.bjut.edu.cn