Computer Science and Application 计算机科学与应用, 2017, 7(6), 590-602
Published Online June 2017 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2017.76070

Hans汉斯

Fuzzy Clustering Based Intrusion Detection Algorithm in Wireless Sensor Networks

Xiaoyong Zhao[1], Junde Ran[1], Rongyong Chen[1], Yuanchan Xia[1], Songtao Guo[2*]

[1]Maintenance Branch of State Grid Chongqing Electric Power Company, Chongqing
[2]College of Electronic and Information Engineering, Southwest University, Chongqing
Email: *songtao_guo@163.com

Received: Jun. 6th, 2017; accepted: Jun. 25th, 2017; published: Jun. 28th, 2017

Abstract

When wireless sensor nodes are applied to the transmission line testing or other areas, they are often easily attacked due to the limited computation capability and the open data transmission environment. Attack nodes can obtain the useful information of nodes (including node location, secret key, and node identity) by capturing the normal nodes in the network, and then copy the information to become clone nodes that can take various internal attacks so that they can obtain more secure information. To solve the problem, we propose an intrusion detection algorithm (IDA) based on detecting the existence of clone nodes. In this algorithm, firstly, we propose the weighted variation coefficient based fuzzy mean clustering algorithm and cluster the networks by the proposed clustering algorithm. Secondly, we choose some nodes with less energy consumption as witness nodes. The witness nodes will monitor the whole network to determine whether the data transmission nodes and the cluster head nodes are replicated. Then, when the witness nodes monitor the data transmission nodes, IDA algorithm will determine whether the data transmission nodes are cloned within the cluster by analyzing the miss detection probability and the effective throughput. In the monitoring of cluster head nodes, IDA algorithm will determine whether the cluster head nodes are replicated by setting the alarm threshold. The simulation results show that our IDA algorithm will decrease the miss detection probability greatly to 50% and reduce the average energy consumption to 20% by choosing appropriate coding function.

Keywords

Wireless Sensor Networks, Fuzzy Clustering, Clone Attack, Intrusion Detection, Missing Detection Probability

无线传感器网络中基于模糊分簇的入侵检测算法

赵晓勇[1]，冉军德[1]，陈荣勇[1]，夏远灿[1]，郭松涛[2*]

*通讯作者。

文章引用: 赵晓勇, 冉军德, 陈荣勇, 夏远灿, 郭松涛. 无线传感器网络中基于模糊分簇的入侵检测算法[J]. 计算机科学与应用, 2017, 7(6): 590-602. https://doi.org/10.12677/csa.2017.76070

[1]国网重庆市电力公司检修分公司，重庆
[2]西南大学电子信息工程学院，重庆
Email: *songtao_guo@163.com

收稿日期：2017年6月6日；录用日期：2017年6月25日；发布日期：2017年6月28日

摘　要

当无线传感器节点被应用到输电线路检测等领域时，由于其自身的计算能力有限、传输环境的开放性等因素，通常容易遭受到攻击。攻击节点通过捕获无线传感器网中的正常节点来获得节点中的有用信息(包括节点位置、密钥、节点身份)并加以复制构成一个能发起各种内部攻击的克隆节点，从而获取网络内部更为机密的信息。为了解决这类问题，我们提出了基于探测克隆节点存在的入侵检测算法(IDA)。在这种算法中，首先，我们提出基于加权变异系数的模糊均值分簇算法并对监测网络进行分簇。然后，我们选择功耗较小的节点作为监测节点(Witness node)，这些监测节点在簇内全覆盖地监测数据传输节点和簇头节点是否被克隆。在监测数据传输节点时，通过分析错失探测概率和有效吞吐量来确定簇内的数据传输节点是否被克隆。在检测簇头节点时，通过设置合适的报警阈值来确定簇头节点是否被克隆。仿真结果表明所提出的入侵检测算法在选择合适的编码函数时，错误探测概率会减小50%以上，网络平均能耗降低20%。

关键词

无线传感器网络，模糊分簇，克隆复制攻击，入侵探测，错失探测概率

1. 概述

被誉为当代三大高新科学技术之一的无线传感器网络[1]，为我们更好地理解逻辑上的信息世界提供了一种全新的平台。部署在特定监控区域的大批传感器节点构成了一个无线传感器网络，虽然这些传感器体积小、计算能力和信息处理能力比较有限，但价格相对低廉[2]。它们可以通过无线的方式彼此之间交互、转发信息完成通信，组成一个自组织的网络系统。由于无线传感器网络具有部署环境复杂，无人值守，能量有限，抗干扰低等特性[3]，使得它在数据感知，收集、传输、处理等各个环节都面临着非常大的安全挑战[4]。此外，单个传感器节点处理能力非常弱，容易被攻击。节点的身份，密钥和位置等信息会被盗用，造成网络中存在大量克隆节点，这些克隆节点看似与正常节点无异，实际上它们可以对网络发动各种各样的恶意攻击，从而达到瘫痪网络的目的。本论文将围绕无线传感器网络中存在的节点克隆攻击问题展开研究。

在无线传感网中如何检测存在的克隆节点，研究人员主要关注以下几个方面的问题，对克隆节点的检测率、网络的通信量等[5]。现在，越来越多的传感器趋于小型化，导致存储能力和电源能量十分有限。如果通信量较大会使电量消耗太快，过多的存储信息也会大量占据内存。克隆节点对于无线传感器网络的危害比较大，现在我们对克隆节点的研究主要集中在两个部分，分别是集中式的克隆攻击方法和分布

式的克隆检测方法。在集中式的方案中，最直接、最容易的方案就是传感器节点将自己的身份(ID)，位置信息和隐私信息一起发送给基站，基站通过检测判断是否具有相同的 ID，如果节点在相同位置 ID 却不同，就可以判断这个节点就是克隆节点。文献[6]介绍了一种通过计算不同节点群之间的(交，并)集合来实现对克隆节点检测的方法，简称为“SET”。如果是网络中不存在克隆节点，那不同的节点群之间没有交集。这种“SET”的简单过程就是在开始的时候就给网络分簇，在簇头节点内进行集合运算，再把得到最终的结果反馈给基站。这种方法通过簇头节点的分级运作，将基站的工作压力转移了一部分给簇头节点。分布式的克隆检测方法的提出是为了解决集中式方案中单节点故障的问题，其监测方式是在节点之间进行克隆检测，分布式方法是目前传感网安全研究中最热的方向。文献[7]提出了两种分布式的克隆攻击检测方案,其分别是 LSM (Line-Selected Multicast)和 RM (Randomized Multicast)检测算法。两者都是利用了相同的观察者来发现具有相同 ID 身份信息的节点，但是却在不同位置来进行克隆检测。两者的区别在于，LSM 是每次随机地选择一个邻居节点进行消息传输以此来形成一条线性可供选择的线路，而 RM 算法是通过泛洪的方式来向周围节点扩散消息。

文献[8]提出了基于压缩感知的克隆节点识别方法。该方法通过对节点感知数据的压缩融合来提高检测有效性。特别是该方法通过利用网络内克隆节点的稀疏性质，不仅可以实现较低的通信开销而且使得网络流更加均匀。在[9]，作者将随机漫步和网络划分相结合提出了一种分布式节点复制攻击检测算法(RAND)。该算法分为两个阶段，第一个阶段是网络配置阶段，即整个网络被分成几个区域；第二个阶段是复制检测节点，在该阶段克隆节点被检测，通过采用一种声明者－记录者－见证者架构并在每个区域内采用随机漫步模型来选择见证者节点。在[10]，作者提出了基于 Quorum 的多播(QBM)和星形线性选择多播(SLSM)来探测节点复制攻击算法。在[11]，为了快速地发现克隆节点，作者提出了一种改进的 LEACH 协议来减小簇的规模，然后通过应用监控节点提出了一种入侵检测算法来探测复制攻击从而大大减少了信息泄露的发生。

通过以上的分析，可以发现现有的方法中还存在一些问题，如检测精度和节点的能量，克隆节点检测率相关。正是由于这些因素的影响，无线传感网的应用得到了很大的限制。本文主要从如何发现克隆节点的角度，对无线传感器网络的安全监测展开研究，针对在探测克隆节点时，网络规模巨大，探测效率低等方面的问题，提出了基于变异系数的模糊分簇方案，然后从理论上说明了本文提出的探测克隆节点方法，即入侵检测算法的有效性。本文分别从网络模型的建构、数学公式的推导、实验仿真等多个方面来构建克隆节点模拟的攻击环境，并在此环境下找到合适的方法去应对这种威胁。通过这两种方法的结合，在探测克隆节点时对恶意节点的攻击，恶意转发信息具有较高的检测率和较低的误判率，网络的安全性得到了一定的保障。本文针对网络中的克隆节点问题，提出了入侵检测算法来检测网络中的克隆节点。在该算法中，主要过程分成四个部分，其分别是：网络的分簇，选择合适的监测节点集合，对数据传输节点的监测和对簇头节点的监测。下面对这四个过程做更详尽的描述。

2. 网络分簇

在现有的分簇算法中，*K* 均值分簇算法[12] [13]是一种使用极其广泛的方法。*K* 均值算法的基本思想是：首先随机地选取 *K* 个对象作为最原始的分簇中心，再计算网络中每个对象与开始选取的 *K* 个原始的聚类中心点之前的距离(采用欧式距离作为度量对象的差异性)，然后把对象分给离它最近的聚类中心。分配到每个聚类中心的所有对象代表了一个聚类。当所有的对象都被分配完成后，在每个聚类内根据对象重新计算聚类中心。在这个过程中会设置终止条件，以此来让聚类达到最优。其中终止的条件可以是以下的任何一种情况：

1) 误差的平方和局部最小

2) 没有(或者最小数目)聚类中心发生变化

3) 没有(或者最小数目)对象再次被重新分配到任何不同的聚类中去。

但是这种聚类的方式存在一个明显的不足，那就是在计算对象之间距离的时候通常认为这些属性对于距离维度的重要性是同等的，所以可能造成一些非常重要的数据被丢失，这种情况就是我们所说的“维数灾难”[14]。基于此，本文提出了一种有效解决这种问题的方法，即采用维度加权变异系数的分簇方式，其核心思想就是给数据中每个维度附加一个不同的权重，越重要的维度其所占的权重也就越大，这就让重要性不同维度的数据在分簇中起到不同的作用，更能合理科学的进行分簇。

2.1. 基于变异系数的模糊分簇

首先，对网络中所有的节点做一个数学的建模，以便在接下来的工作中能更好地来描述分簇算法。假定网络中存在 N 个 M 维度的节点即，其中每个节点 $A=\{a_1,a_2,a_3,\cdots,a_n\}$，$a_{il}$ 为特征向量 a_i 的第 l 个值。存在一个 i 使得 $1\le i\le n$。C 为集合 A 的簇的个数。

K 近邻： K 近邻指的是在距离对象 a 最短的 K 个对象(不包含自身)的集合。记对象 a 的 K 近邻为 $KNN(a)$。

变异系数：当需要对两组数据比较离散程度的时候，两组数据如果因测量的尺度问题而相差非常大或者数据的单位不同时，如果还是用标准差来比较两组数据就非常的不合适了，所以在比较的时候首先消除数据单位和测量尺度的影响。本文提出的变异系数就可以解决这个问题。这是因为变异系数没有量纲[15]，它是数据的标准差和平均数的比值。这样就可以对两组数据进相对客观的比较。其实，变异系数和标准差，极差，方差一样，都是反映数据的离散程度。变异系数的大小不仅受到变量值平均值的影响，而且还受到变量值离散程度的影响。其数学表达式为：

$$CV=\frac{\sigma}{\mu} \tag{1}$$

其中，

$$\begin{aligned}\mu_m&=\frac{1}{N}\sum_{n=1}^{N}a_{nm},\\ \sigma_m&=\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(a_{nm}-\mu_{nm}\right)^2}.\end{aligned} \tag{2}$$

从上面的定义中可以得到，数据的第 m 维属性变异系数值可以定义为：

$$CV_m=\frac{\sigma_m}{\mu_m}=\frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(a_{nm}-\mu_{nm}\right)^2}}{\frac{1}{N}\sum_{n=1}^{N}a_{nm}}. \tag{3}$$

在这里，m 表示数据的第 m 维属性，M 表示数据的维度，n 表示第 A 集合中第 n 个数据，N 表示数据集合 A 的数据个数。a_{nm} 表示第 n 个对象的第 m 维属性。μ_m 表示第 m 维的平均值，σ_m 表示第 m 维的标准差。CV_m 表示第 m 维的变异系数。

变异系数的权重：定义 $W=\{W_1,W_2,\cdots,W_m\}$ 为集合 A 里的 M 维空间权值，其中 m 维的权值表示为：

$$W_m=\frac{CV_m}{\sum_{m=1}^{M}CV_m} \tag{4}$$

M 维的欧氏距离：在 m 维的欧式空间里，其数据节点 a_i 到簇头节点 v_i 的距离可以表示为：

$$Dis\left(a_i,v_j\right)=\sqrt{\sum_{m=1}^{M}\left(a_{im}-v_{jm}\right)^2} \tag{5}$$

所以,带有变异系数的 m 维欧式距离可以表示为：

$$CVDis\left(a_i,v_j\right)=\sqrt{\sum_{m=1}^{M}W_m\left(a_{im}-v_{jm}\right)^2}. \tag{6}$$

本文提出的基于变异系数的模糊分簇算法就是寻找一个最小的目标函数，其主要思想是把数据集 A 中每一个节点对象分配给离它最近的质心节点，使得所有数据对象节点到其所属的质心节点距离之和为最小。其数学表达式可为：

$$\begin{aligned}&\min_{(U,V)}\left\{J_{cvfcm}=\sum_{l=1}^{C}\sum_{i=1}^{N}\left(U_{li}\right)^{\phi}*W_m*d_{li}^2\right\}\\&d_{li}^2=\left\{Dis\left(a_i,v_j\right)\right\}=\left\|a_i-v_l\right\|^2\end{aligned} \tag{7}$$

其中，U 是一个记录对象节点和簇头节点之间关系的 $C*N$ 型矩阵，在矩阵中的每个节点表示该列所在的对象节点属于该行簇头节点的隶属度。V 为网络中质心节点的集合，即为 $V=\left(v_1,v_2,\cdots,v_C\right)^{\mathrm{T}}$。

目标函数的约束条件为：

$$\sum_{l=1}^{C}U_{li}=1,i=1,2,3,\cdots,N.$$

在上面的公式中，ϕ 表示模糊化率($\phi>1$)，$\|*\|$ 表示 m 维欧式空间的距离。我们的目标就是计算在取得最小值时的 U_{li} 和 V_l。

我们知道目标函数的极值的约束条件为：

$$\sum_{l=1}^{C}U_{li}=1,i=1,2,3,\cdots,N.$$

通过拉格朗日乘数法分析，可知其问题是相当于寻找下面给出的方程极小值。

$$F\left(U,\lambda\right)=\sum_{l=1}^{C}\sum_{i=1}^{N}\left(U_{li}\right)^{\phi}*W_m*d_{li}^2+\sum_{i=1}^{N}\lambda\left(1-\sum_{l=1}^{C}U_{li}\right)$$

方程极小值的条件是一阶偏导为 0，即 $\partial F/\partial\lambda=0$ 和 $\partial F/\partial U_{li}=0$，对其展开可以得到：

$$\partial F/\partial\lambda=1-\sum_{l=1}^{C}U_{li}=0 \tag{8}$$

$$\begin{aligned}&\partial F/\partial U_{li}=0\\&\Rightarrow\phi\left(W_m\right)U_{li}^{\phi-1}\left\|a_i-v_l\right\|^2-\lambda=0\\&\Rightarrow U_{li}=\left(\lambda/\phi*W_m\left\|a_i-v_l\right\|^2\right)^{\frac{1}{\phi-1}}\end{aligned} \tag{9}$$

把(8)式代入到(9)式可以得到：

$$\begin{aligned}&\sum_{l=1}^{C}\left(\lambda/\phi*W_m\left\|a_i-v_l\right\|^2\right)^{\frac{1}{\phi-1}}=1\\&\Rightarrow\left(\lambda/\phi\right)^{\frac{1}{\phi-1}}=\left[1\Big/\sum_{l=1}^{C}\left(W_m\left\|a_i-v_l\right\|^2\right)\right]^{\frac{1}{\phi-1}}\end{aligned} \tag{10}$$

将公式(2.10)代入到(2.9)式，可以得到在目标函数取最小值时的U_{li}：

$$U_{li}=\frac{\left(W_m\left\|a_i-v_l\right\|^2\right)^{\frac{1}{\phi-1}}}{\sum_{l=1}^{C}\left(W_m\left\|a_i-v_l\right\|^2\right)^{\frac{1}{\phi-1}}} \tag{11}$$

使用类似的方法可以得到在目标函数取极小值时的V_l。通过计算$\partial F/\partial v_l=0$，可以得到：

$$\begin{aligned}&\partial F/\partial v_l=0\\&\Rightarrow -2\sum_{i=1}^{N}\left(U_{li}\right)^{\phi}*W_m\left(a_i-v_l\right)=0\\&\Rightarrow v_l=\frac{\sum_{i=1}^{N}\left(U_{li}\right)^{\phi}*W_m*a_i}{\sum_{i=1}^{N}\left(U_{li}\right)^{\phi}*W_m}\end{aligned} \tag{12}$$

2.2. 分簇算法的描述

基于变异系数的模糊分簇算法的核心思想是：根据每个对象的属性来计算每个维度的变异系数的不同权值，然后使用 K 邻近的思想在网路中选取 C 个初始的质心节点，其过程就是首先计算所有对象的 K 近邻，在所设定的阀值之下过滤掉低密度的离群点，在高密度对象的网络节点中选择基于属性加权最大距离的 2 个对象作为起始的质心节点，然后逐次选取一个与所有已经确定的质心节点距离最远的点作为另外一个质心节点。就这样每次在网络内生成一个新质心节点，直到完全确定 C 个质心节点才结束此过程。在确定完 C 个初始化质心节点以后，算法会根据每个维度的不同变异系数权值，对 C 个质心节点进行迭代运算，根据维度加权欧氏距离的不同来确定每个对象的隶属度，以此来更新每个簇的质心节点直到其收敛性满足我们所设定的要求。

3. 选择合适的监测节点集合

在无线传感器网络的安全检测中，由于传感器节点探测的数据不同，其节点的能量消耗率也会不一样，所以我们定义一个能量消耗函数$W\left(s_i\right)$来表示节点s_i的能量消耗，即

$$W\left(s_i\right)=1-\frac{e_i\left(t\right)}{e_i\left(0\right)} \tag{13}$$

在这里$e_i\left(0\right)$是节点的起始能量，$e_i\left(t\right)$表示节点s_i在t时刻的剩余节点能量。

在现有的工作[8] [9] [10]中，为了有效地减轻恶意节点的不当行为，提出了一些不当行为检测的算法来保护系统的信誉，算法的核心思想就是在一个簇内有一个监测节点来检测是否有恶意节点。但是其缺点就是单节点故障问题。通过研究发现，在簇内选择多个监测节点可以有效地减少探测错失概率和节点的总能耗。现在问题的关键是如何在一个簇内选择合适数量的监测节点来监测簇内不同类型的节点(数据传输节点和簇头节点)。本文算法的核心思想就是用多个监测节点来监测数据传输节点和簇头节点，要求所选择的监测节点之间需要有足够的能量。这就需要在选择监测节点的时候要考虑监测节点的总能耗最小，即为：

$$\min\left\{Cost\left(H\right)=\sum_{i=1}^{m}W\left(s_i\right)\right\} \tag{14}$$

4. 对数据传输节点的监测

一旦监测节点的集合(S)确定了，每一个监测节点就可以监测簇内的数据传输节点，为了更容易理解，我们考虑一个简单的数据模型来说明在簇内，算法是如何发现恶意节点并且降低克隆节点探测错失概率。

在图 1 中可以看到，有四类节点，其分别是源节点 T，簇头节点 D，攻击节点 R 和监测节点 M_1，M_1。图中的实线表示为数据传输方向在这里数据会通过 R 中继，虚线表示为监测节点对其他节点监测的信息流向。

假设单位时间内数据传输率为 1 并且从 T 到 D 的数据传输链路是可靠的，每个监测节点观察数据传输的概率都为 p。源节点 T 采用极大距离可分编码(MDS)，在这里长度为 y 的数据包被封装在长度为 x 的包内($y<x$)其封装是使用一个关于(x,y)的函数[16]，一个(x,y)的极大距离可分码其最小的汉明距离为 d 通过辛格顿界引理[17]可以知道 $d \le x-y+1$。所以如果攻击节点修改的信息长度超过 $x-y$，那么在编码/解码过程将会被发现。假设攻击节点没有在被编码/解码过程被发现，即攻击节点篡改的数据包不超过 $x-y+1$。可以很容易计算攻击节点不被检测到的概率为：

$$P_{miss}(x,y,p)=(1-p)^{x-y+1} \tag{15}$$

$P_{miss}(x,y,p)$表示错失探测概率即攻击节点没有被其中一个监测节点发现的概率。我们可以构造一个(x,y)的编码函数，

$$y=x+1-\frac{f(x,p)}{p} \tag{16}$$

通过公式(15)，我们可以得到 $P_{miss}(x,y,p)$ 为：$P_{miss}(x,y,p)\le e^{-p(x-y+1)}=e^{-f(x,p)}$。

以此，攻击节点没有同时被 g 个监测节点观测到的概率为：

$$P_{miss}(x,y,p,g)=e^{-g*f(x,p)}$$

为了使错失探测概率更小和编码函数 y/x 达到更优，本文构造了函数 $f(x,p)=\beta\ln x$，其中 β 为正常数。那么可以得到：

$$P_{miss}(x,y,p,g)\le e^{-g*\beta\ln x}=x^{-g\beta}\to 0 \tag{17}$$

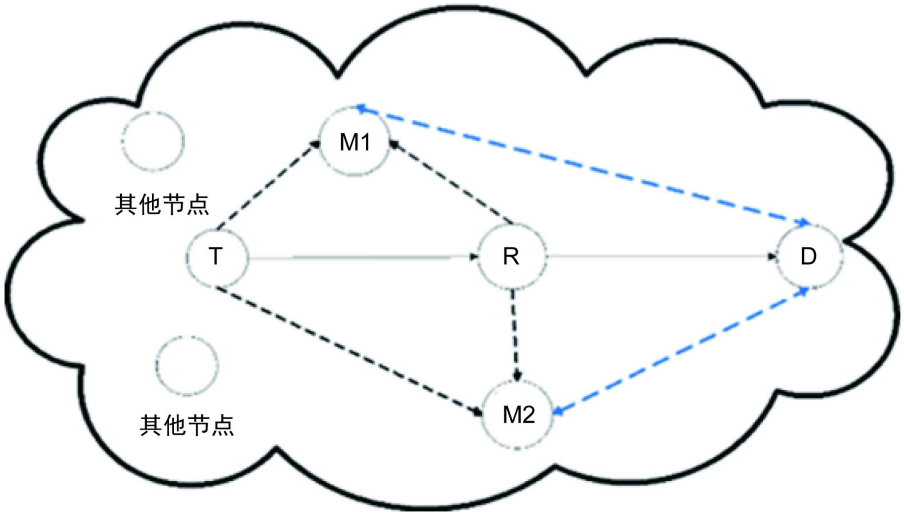

Figure 1. Data transmission traffic in WSNs
图 1. 网络中的数据传输流

因为 $g>0, x\to\infty$ 和 $x^{-g\beta}\to 0$ 所以 $P_{miss}(x,y,p,g)\to 0$。使用上述的编码函数，可以计算出 $x\to\infty$ 时的编码率 y/x 为：

$$\frac{y}{x}=\frac{x+1-\frac{\beta\ln x}{p}}{x}=1+\frac{1}{x}-\frac{\beta\ln x}{px}\to 1 \tag{18}$$

所以，寻找合适的 β，使 x 足够大和在簇内安插监测节点，可以让攻击节点篡改信息的机会将会被有效降低。

5. 对簇头节点的监测

在一个簇内仅仅只依靠一个监测节点去判断簇头节点是否已经被俘获这是不可靠的，因为被俘获的节点可以伪造信息。因此，当有多个监测节点去监测簇头的不当行为时，检测报警消息可能更可信。基于这样的考虑，本文利用 m 个监测节点合作监控簇内的簇头节点。此外，与以前的工作相比，发送节点撤销消息会在整个传感器网络传播，而在本文提出的监测方法中，撤销消息只是在一个簇内传播。在监测节点撤销不正常的簇头节点后，其将不再与其它传感器节点进行通信。不正常簇头节点的撤销过程可以表示如下。

5.1. 发出警报信息

这一步的核心思想就是簇头节点 $CH(u)$ 广播它的身份信息(ID),位置信息(l)，和加密密钥信息(K_C)给每一个监测节点(MN) S_i，其数学表达式可以表示为 $u\to S_i:\{Msg\}_{K_C}=\{ID,l\}_{K_C}$。在监测节点(MN)收到簇头节点传输过来的信息后，监测节点会在一段时间 t 内轮流监测簇头节点，当簇头节点 $CH(u)$ 被其中一个监测节点判断其行为异常时，它将会发送警告信息 $Alarm\{u\}$ 给其他的监测节点集合 S，其数学表达式为： $S_i\to S:\{Alarm\{u\},S_i,Y\}$。如果其中的一个监测节点其收到的警告信息超过 Y，其会撤销不正常的簇头节点。

为了记录获得的警告信息，每个监测节点都会保有一个警告表，如表 1 所示。这个报警表中有三个部分组成，其分别是，“不正常的节点”、“报警计数”、“当前监测节点”。“不正常的节点”的作用是记录疑似异常的簇头节点 CH。“报警计数”的作用是记录从其他监测节点发过来的报警信息次数。“当前监测节点”的角色是列出接收警告信息的监测节点。如果节点收到从其它监测极点发来的警告信息，它会检查自身的报警表，当报警计数超过给定的阈值 Y 的时候。异常的簇头节点 CH 和密钥信息 K_C 将会被撤销。

5.2. 报警阈值的确定

现在用 P_c 来表示监测节点被俘获的概率，其值的大小是由实际环境决定的。例如，在不同的监测现场 $P_D=\sum_{i=X}^{m}\binom{m}{i}(1-P_f)^{m-i}P_f^i$ 可能分别是 0.1，0.5，0.7。由于战争环境是最危险的，其 P_c 也应该为最高。正如上面提到的，如果一个监测节点想要撤销一个异常簇头节点 CH，其报警阈值必须超过 Y，意味着在

Table 1. The alarm table in a monitoring node
表 1. 在每个监测节点内的报警表

不正常的节点	报警计数	当前监测节点
U	1	S_i

监测节点清除簇头节点之前，攻击节点必须要俘获 Y 个监测节点。Y 个监测节点被同时俘获的概率可以表示为，$(P_c)^Y$。由于 $P_c<1$，Y 值越大，$(P_c)^Y$ 的值就会越小，被俘获的机会就会越小，从而更灵敏发现异常簇头节点的机会也就越少。所以报警阈值 Y 的大小是由监测节点被俘获的概率 P_c 和安全容忍度 θ 共同决定的，其表达式为：$(P_c)^X<\theta$。

5.3. 确定监测节点的数量

通过以上的分析，我们可以知道监测节点在网络安全和能量消耗中都发挥了非常重要的作用，如果监测节点的数量过少的话，每一个监测节点监测的时间就会增加，在监测簇头节点的行为时其消耗的能量也会更大。所以，合适的监测节点数 m 是由网络的安全和能量消耗共同决定的。攻击节点被 X 个监测节点发现的概率为 P_D，一个监测节点成功发现攻击节点的概率记为 P_f。我们知道 P_D 是由 m，Y 和 P_f 共同决定的，我们可以通过公式 $P_D=\sum_{i=X}^{m}\binom{m}{i}(1-P_f)^{m-i}P_f^i$ 来确定最优的监测节点数量。

6. 仿真分析

本文假设在一个 300 m×300 m 的区域内均匀的分布了 100 个无线传感器节点于该区域的输电线路上，这些节点的起始能量为 12 J。

6.1. 数据传输节点的检测性能

在这一部分，仿真的重点放在对分簇网络中的数据传输节点的研究，图 2 和图 3 显示了在不同的 β 和 x 情况下，错失探测的概率与观测概率 p 之间的关系。从图 2 和图 3 可以看得出错失探测概率随着观测概率 p 的增加而逐渐减小，这意味着攻击节点将会有更高的概率被探测到。此外，错失探测概率也与参数 β 和 x 有关。从图 2 和图 3 也可以看出随着 β 和 x 的增大，错失探测概率会减小。所以，可以不断地提高观察概率 p，参数 β 和 x 的值来使措施探测概率在实际环境中达到最优，能够在传感网中更好的发现克隆节点。

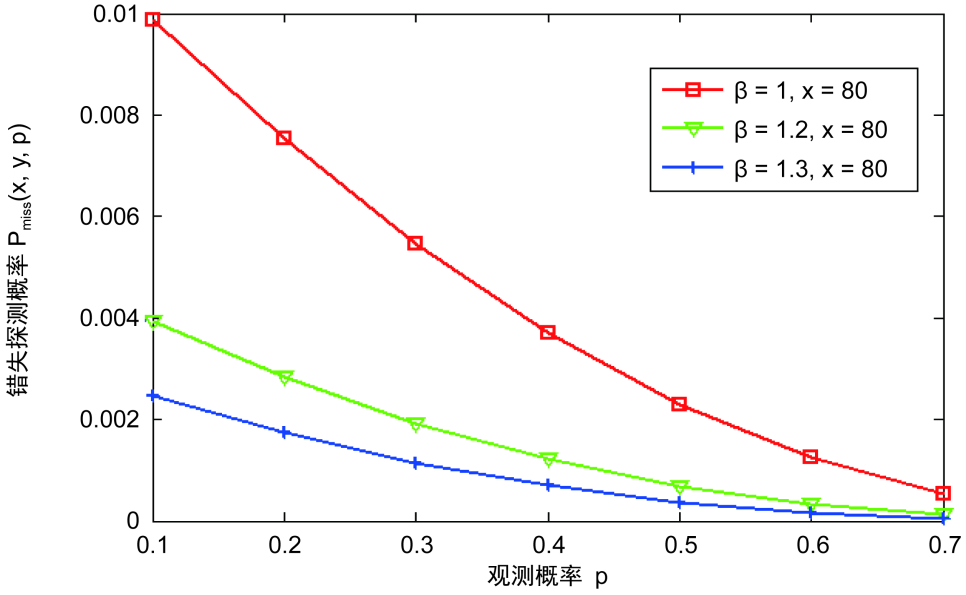

Figure 2. The relationship between miss detection probability and the observation probability for different β
图 2. 在不同的 β 下错失探测概率与观测概率 p 的关系

图 4 描述的是错失探测概率与数据包的长度 x 之间的关系。一个较大的数据包 x 意味着其编码和解码的复杂性也就越高。从仿真结果上可以看出，随着 x 的增加，错失探测概率会减少，这是因为编码的数据包越大，花在攻击数据包的时间也就越长。因此，攻击者需要篡改更多信息，以达到俘获该数据包，这样更容易被监控节点探测到它的恶意行为。此外，从图 5 中可以观察到，错失探测概率会随着监控节点的数量 g 的增加而迅速的减小。正因如此，为了防止节点被克隆，应该在传感网中选择合适数量的监测节点以此来监测整个网络，而不是以前的单个监测节点的模式，这样不仅可以避免单节点攻击，还可以提高探测克隆节点的概率。

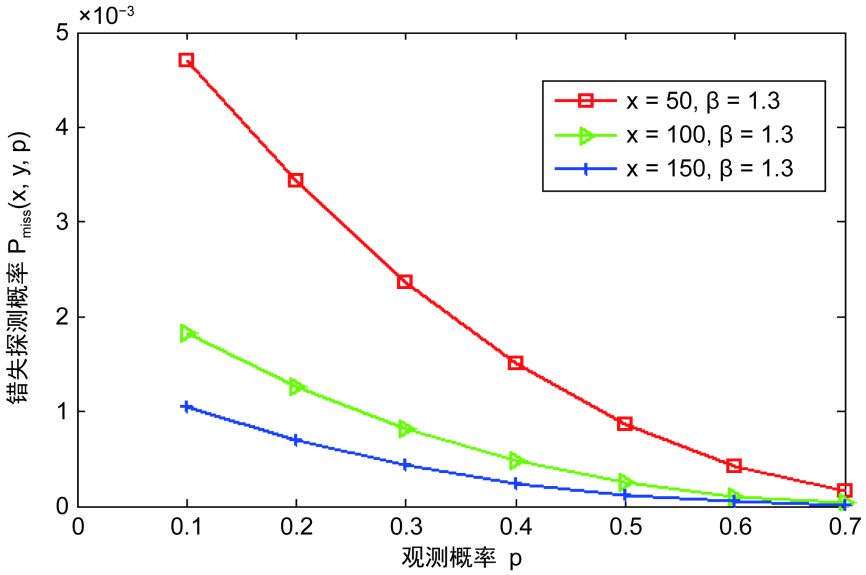

Figure 3. The relationship between miss detection probability and the observation probability for different x

图 3. 在不同的 x 下错失探测概率与观测概率 p 的关系

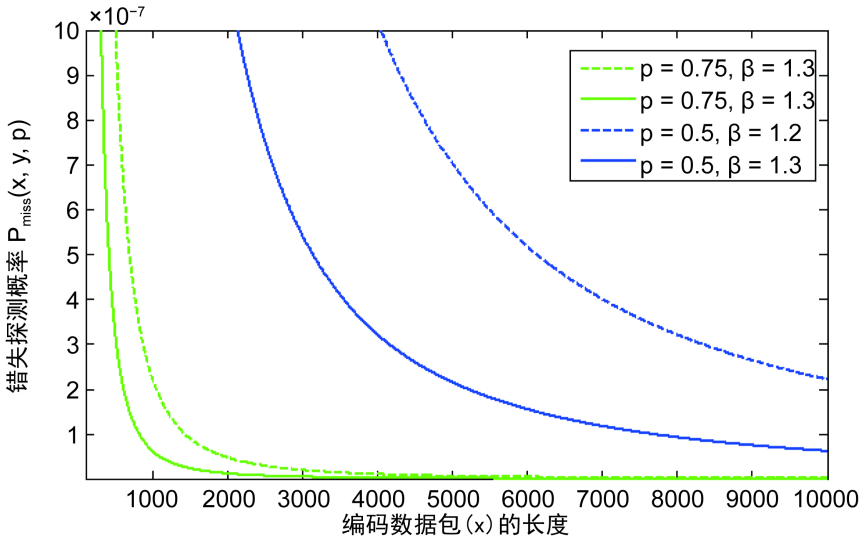

Figure 4. The relationship between miss detection probability and the length of encoding data packet x

图 4. 错失探测概率与编码数据包的长度 x 之间的关系

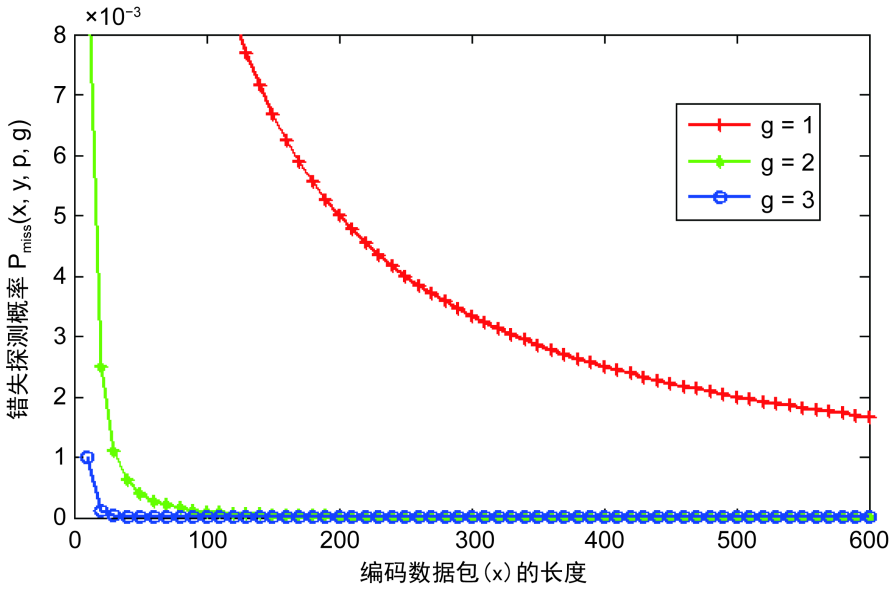

Figure 5. The relationship between miss detection probability and the number of monitoring nodes
图 5. 错失探测的概率与不同数量的监测节点 g 的关系

6.2. 簇头节点的检测性能

从图 6 中我们可以看出簇头被俘获的概率 $(P_c)^Y$ 与监测节点数量之间的关系。随着网路中监测节点的数量不断地增加，簇头被俘获的概率 $(P_c)^Y$ 也极大地降低了。此外，对于一个攻击者来说，如果要俘获簇头节点，其被俘获的概率不应该低于 $(P_c)^Y$。正是因为如此，我们可以得到一个实际网络中应该部署多少监测节点，例如，在战争环境中的传感器网络($P_c = 0.7$)，此时对 θ 的要求是不应小于 0.2，这样我们可以很容易从图 2~8 中得到 $Y = 5$。

6.3. 入侵检测算法的比较

这部分的仿真试验比较本文提出入侵检测算法(IDA)和单监测节点克隆检测协议如 RAND[9]以及 SLSM[10]在簇头节点的平均能量消耗以及网络中的节点存活数据方面的性能。从图 7 中可以看出单监测节点克隆检测协议 RAND 的簇头平均能耗比我们设计的入侵检测算法快得多。图 8 显示了本文提出的 IDA 算法与 SLSM 算法在运行相同的时间内网络中节点的存活数量。可以看出本文提出的入侵检测算法的节点存活数量有很大的提高。这是因为相比单监测节点，克隆检测协议不考虑网络的分层和节点工作负载有关，入侵检测算法会有更长的生命周期。

7. 结束语

本文针对如何快速有效地检测无线传感网中存在的恶意克隆攻击节点而展开了研究。首先，本文对现有无线传感网存在的安全问题做了分析，提出了现有算法存在的不足，即现有算法直接在整个网络里检测恶意克隆节点，这样做的缺点就是网络的能量浪费很大。所以本文提出了入侵检测算法，其主要过程分成四个部分：网络分簇、选择合适的监测节点集合、对数据传输节点的监测和对簇头节点的监测。该算法的核心思想就是通过簇内的多个监测节点去观测网络里的数据传输节点和簇头节点是否被俘获，从而减少错失探测克隆节点的概率。通过实验我们可以看出，本文提出的入侵检测算法很好地解决了如何在降低错失探测克隆节点的概率的同时，还能很好地降低网络的能量消耗和保障网络的吞吐量。

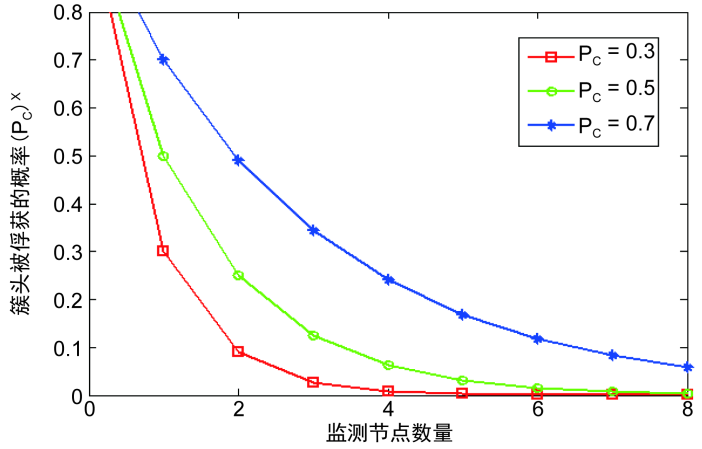

Figure 6. The relationship between capturing probability of cluster head and the number of monitoring nodes
图 6. 簇头的俘获概率与网络中监测节点数量之间的关系

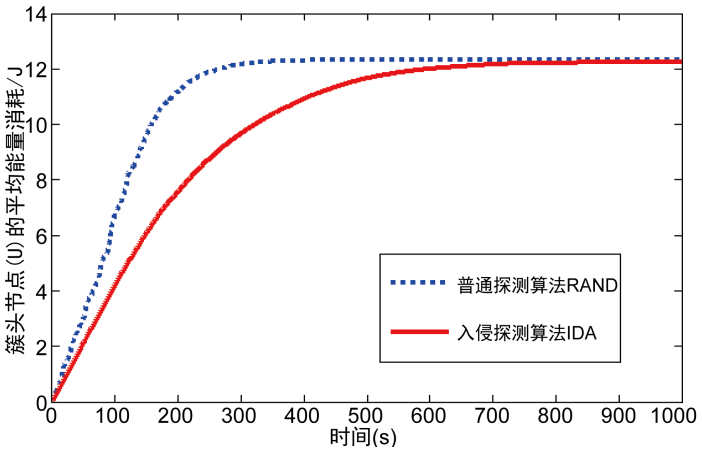

Figure 7. The relationship between average energy consumption of cluster head and the time
图 7. 簇头节点的平均能耗与时间的关系

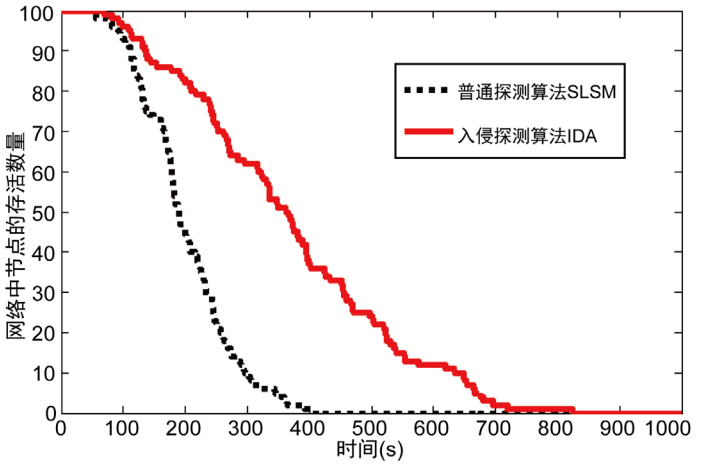

Figure 8. The relationship between the number of alive nodes and the time
图 8. 网络中节点的存活数量与时间的关系

基金项目

国家自然科学基金项目(61373179)。

参考文献 (References)

[1] 王殊, 阎毓杰, 胡富平, 等. 无线传感器网络的理论及应用[M]. 北京: 北京航空航天大学出版社, 2007.

[2] Fekete, S.P., Kroller, A., Pfisterer, D., *et al.* (2004) Locating and Bypassing Routing Holes in Sensor Networks. *Proceedings of International Workshop on Algorithmic Aspects of Wireless Sensor Networks*.

[3] 郎为民, 杨德鹏, 李虎生. 智能电网 WCSN 安全体系架构研究[J]. 信息网络安全, 2012(4): 19-22.

[4] Wang, Y., Lin, W. and Zhang, T. (2010) Study on Security of Wireless Sensor Networks in Smart Grid. *Proceedings of* 2010 *International Conference on Power System Technology*, Hangzhou, 24-28 October 2010, 267-273. https://doi.org/10.1109/POWERCON.2010.5666729

[5] Zouridaki, C., Mark, B.L., Hejmo, M. and Thomas, R.K. (2005) A Quantitative Trust Establishment Framework for Reliable Data Packet Delivery in MANETs. *Proceedings of the* 3*rd ACM Workshop on Security of Ad Hoc and Sensor Networks*, Alexandria, VA, 7 November 2005, 1-10. https://doi.org/10.1145/1102219.1102222

[6] Choi, H., Zhu, S. and La Porta, T.F. (2007) SET: Detecting Node Clones in Sensor Networks. 3*rd International Conference on Security and Privacy in Communications Networks and the Workshops*, Nice, 17-21 September 2007, 341-350. https://doi.org/10.1109/SECCOM.2007.4550353

[7] Parno, B., Perrig, A. and Gligor, V. (2005) Distributed Detection of Node Replication Attacks in Sensor Networks. 2005 *IEEE Symposium on Security and Privacy*, 8-11 May 2005, 49-63. https://doi.org/10.1109/SP.2005.8

[8] Yu, C.M., Lu, C.S. and Kuo, S.Y. (2016) Compressed Sensing-Based Clone Identification in Sensor Networks. *IEEE Transactions on Wireless Communications*, **15**, 3071-3084. https://doi.org/10.1109/TWC.2016.2516021

[9] Khan, W.Z., Aalsalem, M.Y., Saad, N.M., Xaing, Y. and Luan, T.H. (2014) Detecting Replicated Nodes in Wireless Sensor Networks Using Random Walks and Network Division. 2014 *IEEE Wireless Communications and Networking Conference* (*WCNC*), Istanbul, 6-9 April 2014, 2623-2628. https://doi.org/10.1109/WCNC.2014.6952822

[10] Ho, Y.S., Ma, R.L., Sung, C.E., Tsai, I.C., Kang, L.W. and Yu, C.M. (2015) Deterministic Detection of Node Replication Attacks in Sensor Networks. 2015 *IEEE International Conference on Consumer Electronics*, Taipei, 6-8 June 2015, 468-469. https://doi.org/10.1109/ICCE-TW.2015.7217002

[11] Cheng, G., Guo, S., Yang, Y. and Wang, F. (2015) Replication Attack Detection with Monitor Nodes in Clustered Wireless Sensor Networks. 2015 *IEEE* 34*th International Performance Computing and Communications Conference* (*IPCCC*), Nanjing, 14-16 December 2015, 1-8. https://doi.org/10.1109/pccc.2015.7410341

[12] 楼晓俊, 李隽颖, 刘海涛. 距离修正的模糊 C 均值聚类算法[J]. 计算机应用, 2012, 32(3): 646-648.

[13] 齐淼, 张化祥. 改进的模糊 C 均值聚类算法研究[J]. 计算机工程与应用, 2009, 45(20): 133-135.

[14] 罗会兰, 危辉. 一种基于聚类集成技术的混合型数据聚类算法[J]. 计算机科学, 2010, 37(11): 234-238.

[15] Jing, L., Ng, M.K. and Huang, J.Z. (2007) An Entropy Weighting k-Means Algorithm for Subspace Clustering of High-Dimensional Sparse Data. *IEEE Transactions on Knowledge and Data Engineering*, **19**, 1026-1041. https://doi.org/10.1109/TKDE.2007.1048

[16] Balli, H., Yan, X. and Zhang, Z. (2009) On Randomized Linear Network Codes and Their Error Correction Capabilities. *IEEE Transactions on Information Theory*, **55**, 3148-3160. https://doi.org/10.1109/TIT.2009.2018173

[17] Ngai, C.K., Yeung, R.W. and Zhang, Z. (2011) Network Generalized Hamming Weight. *IEEE Transactions on Information Theory*, **57**, 1136-1143. https://doi.org/10.1109/TIT.2010.2095233