无线传感器网络中基于规范的入侵检测算法研究*

张 帅, 张凤斌
(哈尔滨理工大学 计算机科学与技术学院, 哈尔滨 150080)

摘 要: 为了解决无线传感器网络的安全问题,提出了一种基于规范的入侵检测算法。该算法根据概率论的有关理论,对邻域节点的单位时间特征量设定阈值,阈值的设定方法具有通用性,并且阈值自适应更新,符合传感节点性能随着时间发生变化的特点。将检测节点获得的单位时间特征量值与阈值进行对比来判断入侵。通过仿真实验验证了该算法优于其他基于规范的入侵检测方案,不仅能够满足高检测率低误报率的要求,还具有较好的能效性。

关键词: 无线传感器网络; 入侵检测; 数理统计; 特征量
中图分类号: TP393.08 **文献标志码**: A **文章编号**: 1001-3695(2012)09-3464-03
doi:10.3969/j.issn.1001-3695.2012.09.070

Research on intrusion detection algorithm based on specification for WSN

ZHANG Shuai, ZHANG Feng-bin
(*College of Computer Science & Technology*, *Harbin University of Science & Technology*, *Harbin* 150080, *China*)

Abstract: To solve the security problem of wireless sensor networks, this paper proposed a kind of intrusion detection algorithm based on specification. The algorithm set threshold for characteristic quantity of neighborhood nodes in unit time according to the theory of probability. The method of threshold setting was versatile, and threshold could be updated adaptively to accord with the properties of sensor nodes changing over time. Compared the characteristic quantity in unit time from detection node with the threshold to judge invasion. Simulation results show the proposed algorithm is superior to other intrusion detection schemes based on specification, not only can meet high detection rates and low false positives, but also has good energy efficiency.

Key words: wireless sensor networks(WSN); intrusion detection; mathematical statistics; characteristic quantity

无线传感器网络由大量传感器节点组成,这些传感节点能够监测周围环境,并将传感数据通过多跳路由传向基站[1]。无线传感器网络起初主要应用于军事,随着成本降低将广泛应用于环境、医疗卫生、家庭、太空探测、救灾等其他领域。而缺乏有效的安全机制已经成为制约无线传感器网络应用的主要障碍[2]。

入侵检测是无线传感器网络的重要安全防护手段,目前主要有误用检测、异常检测和基于规范检测三种主要技术供无线传感器网络入侵检测系统应用[3]。基于规范的入侵检测系统对某些特征量设定阈值来描述正确的操作,特征量不在阈值范围内则被怀疑为攻击行为。文献[4]提出一种基于规范的入侵检测方案,通过学习阶段得到各特征量的阈值,当某个邻域节点的特征量值高于学习阶段得到的平均值时则认为发生了入侵行为,该方案符合无线传感器网络资源受限的特点,但响应较慢且平均值并不能够很好地对异常进行界定。文献[5]同样基于规范检测的思想,通过分析邻域节点报文能量和数据包接收率这两种特征量来对入侵进行判断,若接收包能量超过一定范围或数据包接收率比值超过一定门限,则可判断异常,该方案采用简单的统计学方法进行阈值设定,但此种阈值设定的方法不适用于其他特征量。文献[6]对丢包率这一特征进行阈值设定,在 w 窗口的时间内若丢包率超过一定百分比则认为发生异常产生警报,这种方法强调节点间进行协作来提高检测效果,但需要节点之间大量的通信,能效性较差。文献[7]采用序贯概率比检验的方法,对传感节点不与周围节点发生消息通信的时间进行阈值设定,检测动态攻击节点,该方案针对动态攻击节点较为新颖,然而很多军事和民用的无线传感器网络支持传感节点的移动,故该方案对那些节点位置可能发生改变的网络不适用,同时该方案并不支持对那些静态攻击节点的检测。本文基于规范检测提出入侵检测算法。

1 特征量的提取与分析

入侵检测节点将其接收端口设为混杂模式进行监听,收集邻域节点的数据信息。为了避免大量数据包的存储带来存储空间上的消耗,主要进行特征量的提取,并且每经过一段时间将数据包丢弃,这样既能达到收集信息的目的又不耗费过多的存储空间。可供收集的具体特征量有无线电的传播范围、信息的完整性、节点发生信息冲突的次数、连续信息的发送间歇、信息的最大重传次数、丢失数据包的个数、发送数据包的个数、接收数据包的个数、数据包的接收功率值、平均等待时间、路由表的变化次数等。

对某些特征量进行分析时,通常需要加入时间因素,如节点发生信息冲突的次数和丢包数,若不考虑时间,则很难有一个确定的阈值。本文参照传统网路的流量数据来讨论单位时间特征量,如单位时间发生的冲突次数、丢包个数和单位时间

收稿日期: 2012-02-15; **修回日期**: 2012-03-25 **基金项目**: 国家自然科学基金资助项目(60671049,61172168)

作者简介:张帅(1987-),男,黑龙江伊春人,博士研究生,主要研究方向为信息与网络安全、无线传感器网络入侵检测(shuaiflyfox@163.com);张凤斌(1965-),男,黑龙江哈尔滨人,教授,博导,博士,主要研究方向为信息与网络安全.

收到数据包的个数等。本文设计的入侵检测系统并不对所有的特征量进行提取,那样将耗费大量的通信和计算资源,主要对丢包数和收包数这两个特征进行提取,并对单位时间丢包数和收包数的阈值进行设定便可对选择转发、黑洞、能量耗尽、拒绝服务等大量常见的攻击进行检测。

中心极限定理揭示了正态分布的意义,在客观实际中有许多随机变量,它们是由大量的相互独立的随机因素的综合影响所形成的,而其中每一个因素在总的影响中所起的作用都是微小的,这种随机变量近似地服从正态分布,这种现象就是中心极限定理的客观背景。决定无线传感器网络单位时间特征量这个随机变量的因素很多,有信道误码、网络拥塞、接收装置、基站、传感节点自身、周围环境、电磁干扰、障碍物、信号强弱、各种恶意攻击等,而每个因素都相互独立且在总的影响中所起的作用是微小的。由此,单位时间特征量可表示为以上大量随机因素的函数,这个函数形式上可以展开成在某个中心值附近的一阶微分形式,就是 N 个因素的随机变量代数和。因此,无线传感器网络单位时间的特征量值可近似为服从正态分布[8]。

2　单位时间特征量阈值的设定

2.1　基于数理统计设定单位时间特征量阈值

假设 D_n 是 n 个单位时间得到的特征量值的累加和,d_{n+1} 是第 $n+1$ 个单位时间内的特征量值,因此有 $n+1$ 个单位时间内特征量的累加和为

$$D_{n+1}=D_n+d_{n+1} \tag{1}$$

且 D_n 和 d_n 的关系可表示为

$$D_n=\sum_{i=1}^{n}d_i \tag{2}$$

$n+1$ 个单位时间的特征量均值为

$$\overline{D_{n+1}}=\frac{D_{n+1}}{n+1} \tag{3}$$

样本标准差为

$$S_1=\sqrt{\frac{1}{n}\sum_{i=1}^{n+1}(d_i-\overline{D_{n+1}})^2}=\sqrt{\frac{1}{n}(\sum_{i=1}^{n+1}d_i^2-(n+1)\overline{D_{n+1}}^2)} \tag{4}$$

定理 1　设 $X_1,X_2,\cdots,X_n$ 为总体 $N(\mu,\sigma^2)$ 的一个样本,$\overline{X}$ 与 S^2 分别为样本均值和样本方差,则

$$\frac{(\overline{X}-\mu)\sqrt{n}}{S}\sim t(n-1) \tag{5}$$

其中:$S=\sqrt{S^2}$ 称为样本标准差。

由定理 1,$t=\dfrac{(\overline{x}-\mu)\sqrt{n}}{s}\sim t(n-1)$,对给定的 α,查 t 分布表得临界值 $t_{\frac{\alpha}{2}}(n-1)$ 使得

$$P[-t_{\frac{\alpha}{2}}(n-1)<t<t_{\frac{\alpha}{2}}(n-1)]=1-\alpha \tag{6}$$

将式(6)括号内不等式

$$-t_{\frac{\alpha}{2}}(n-1)<\frac{(\overline{x}-\mu)\sqrt{n}}{s}<t_{\frac{\alpha}{2}}(n-1) \tag{7}$$

转换为等价形式

$$\overline{x}-t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}}<\mu<\overline{x}+t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}} \tag{8}$$

则可得 μ 的置信区间为

$$(\overline{x}-t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}},\overline{x}+t_{\frac{\alpha}{2}}(n-1)\frac{s}{\sqrt{n}}) \tag{9}$$

由此,对于每个单位时间的观察结果,只需知道样本均值 $\overline{D_{n+1}}$ 与样本平方累加和 $\sum_{i=1}^{n+1}d_i^2$。而实际情况下,总体的标准差 σ 大多数情况下是未知的,则可构造总体均值的置信区间为

$$(\overline{D_{n+1}}-t_{\frac{\alpha}{2}}(n)\frac{S_1}{\sqrt{n+1}},\overline{D_{n+1}}+t_{\frac{\alpha}{2}}(n)\frac{S_1}{\sqrt{n+1}}) \tag{10}$$

其中:$t_\alpha(n)$ 为 $t(n)$ 的上侧 α 分位数。如果当前单位时间特征量值满足式(10)的阈值区间,则确定此特征量正常,否则为异常。这样,只需观察一定容量的 d_{n+1} 就可以很容易地判断某个传感节点的行为是否正常,而无须观察大容量的样本数据来进行判断。

2.2　阈值的动态更新

将收集到的某个传感节点的单位时间特征量值与该单位时间特征量的阈值进行对比,若在阈值范围内,则此量值正常,将其加入单位时间特征量值的总体样本。存储单位时间特征量总体样本的空间设置为一个先入先出的队列,最初得到的样本将随着新样本的加入被丢弃,而不在阈值范围内,判断为异常的量值将被丢弃,然后将这个量值所在的传感节点确定为可疑节点向基站报告。每隔一段时间,在总体样本中抽取一定数量的样本对阈值重新进行计算,这样阈值将随着整体网络的变化自适应改变。动态更新阈值的整个过程如图 1 所示。

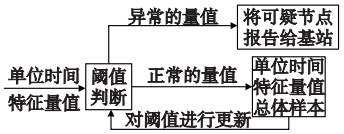

图1　动态更新阈值过程

3　入侵检测算法描述

根据以上工作,可设计单位时间特征量阈值自适应变化的基于规范入侵检测算法,这里以单位时间丢包数为例,算法具体过程如下:

a)入侵检测节点对邻域节点进行监听,分别用计数器记录各邻域节点的丢包数,设 Δt 为单位时间,每隔 Δt 时间统计一次各邻域节点的丢包数,统计后将其存储到先进先出队列中,计数器清零进行下一个单位时间丢包数的统计。

b)设定 t 为准备阶段的准备时间,通过时间 t 判断是否准备阶段结束。若准备时间未到则跳到 a),监督节点继续监视邻域节点并提取特征量,统计并存储单位时间特征量值;若准备时间结束则跳到下一步进入检测阶段。

c)对准备阶段得到的所有邻域节点的单位时间丢包数的总体样本进行采样,设定采样样本的容量为 m。

d)由式(3)计算采样样本的均值,根据式(4)计算样本的标准差。

e)给定 α 值,查 t 分布表得临界值,结合 d)中的样本均值和样本标准差,通过式(10)得出总体均值的置信区间,并将其确定为阈值。

f)入侵检测节点继续对邻域节点进行监听,用计数器记录各邻域节点的丢包数,Δt 时间后统计一次各邻域节点的丢包数。

g)将 f)中得到的单位时间丢包数与 e)中获得的阈值区间比较。若在阈值范围内则判断该单位时间丢包数的邻域节点正常,将这个单位时间丢包数存储到准备阶段获得的单位时间丢包数总体样本的先进先出队列中;若不在此区间内,则将单位时间丢包数丢弃,确定被监听的那个邻域节点异常并通知基站。

h)设定 T_1 为检测时间,若检测时间未到 T_1,则跳到 f)继续检测,若已到 T_1 时间,则跳到 c)重新抽取样本计算阈值区间,使得阈值自适应地变化。

由以上步骤可以看出,算法的时间复杂度主要集中在步骤 d)上,在根据式(4)计算样本的标准差时,需计算样本平方累加和,故幂运算量可作为评判算法复杂度的主要标准,算法的时间复杂度为 $O(n^2)$。而对此算法的空间复杂度进行度量时,主要考虑的是存储单位时间丢包数总体样本的队列,当新样本加入时,对最初进入队列的样本进行更新,故算法的空间复杂度为 $O(n)$。由此可见,该算法的时间复杂度和空间复杂度可被计算资源和存储资源都非常有限的无线传感器网络节点所接受。

4 仿真实验

本实验使用 NS2 仿真软件[9] 仿真一个大小为 15 m × 15 m、由 200 个节点组成的无线传感器网络,各节点最初的能量和通信范围相同,设定网络中数据包的传输速率为每秒 20 KB,文中的单位时间 Δt 为 10 s,采样的样本容量 m 为 50,准备时间 t 和重新计算阈值区间的时间 T_1 均为 10 min。在实验中保持平均每个检测节点周围有相同数目的普通邻域节点,使得检测节点能够对周围普通节点进行很好的覆盖。先后对单位时间丢包数和收包数这两个特征量通过提出的算法来分别检测随机模拟的选择转发攻击[10] 和拒绝服务攻击[11]。

检测率是由检测到的恶意节点数量与全部恶意节点数量的比值求得,检测率越高说明检测能力越强;误报率是正常传感器节点被误判为攻击节点的比率。图 2 和 3 左半部分分别是分析丢包数特征量检测选择转发攻击的检测率和误报率,右半部分是分析收包数特征量检测拒绝服务攻击的检测率和误报率。图 2 和 3 中 date1 和 date3 是 α 为 0.1 时的结果, date2 和 date4 是 α 为 0.02 时的结果。从图 2 可以看出,使用本文提出的算法对丢包数特征进行分析,能够很好地检测选择转发攻击,同样对单位时间收包数特征进行阈值设定能够对拒绝服务攻击进行很好的检测,并且 α 为 0.1 时比 α 为 0.02 时检测率高。图 3 显示出本文提供的算法产生的误报率较低,能够对攻击节点进行较准确的判断。误报率在 α 为 0.1 时比 α 为 0.02 时高。由此可见,α 为 0.1 时比 α 为 0.02 时的置信区间范围小,正常节点被误判为异常的可能性较大,检测到的异常节点数也较多。

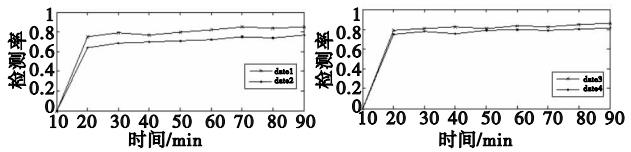

图2 检测率

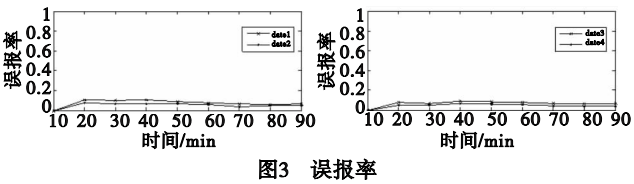

图3 误报率

图 4 中的平均能量指的是所有监督节点的平均能量和所有普通节点的平均能量。由于无线传感器网络无人看管并且能量有限,使用能量消耗很大的入侵检测技术将很快耗尽检测节点的电能使其失效,因此能效性是衡量入侵检测技术的重要指标。从图 4 可以看出,检测节点的能量并未较普通节点的能量消耗很多,说明此算法能效性较强。

实验结果显示,本算法与文献[6,7]中的方案均具有高检测率、低误报率、能效性好的特点。本算法使用统计知识使得特征量阈值的设定具有通用性,而文献[6,7]提出的方案分别用不同的方法对特定的几个特征量进行阈值设定,此方法不适于其他特征量,而且检测结果过分依赖于缓冲区的大小。

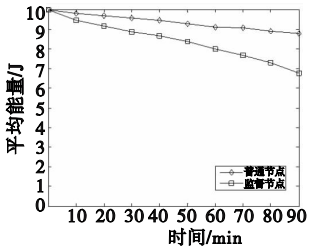

图4 平均能量

5 结束语

本文提出了一种基于规范的入侵检测算法,其具有良好的时间复杂度与空间复杂度,适用于资源受限的无线传感器网络节点。利用统计学的方法为邻域节点的单位时间特征量设定阈值,此设定阈值的方法具有通用性,适用于不同的特征量。同时阈值的动态更新方法使阈值可自适应变化,符合无线传感器网络传感节点的性能随着时间发生变化的特点。仿真实验结果表明,本算法不仅能够保持较高的入侵检测率和较低的误报率,而且能够使检测节点的能量消耗较低。

参考文献:

[1] 任丰原,黄海宁,林闯. 无线传感器网络[J]. 软件学报,2003,14(7):1282-1291.

[2] LEE J, KAPITANOVA K, SON S H. The price of security in wireless sensor networks[J]. Computer Networks, 2010, 54(17): 2967-2978.

[3] ABDUVALIYEV A, LEE S, LEE Y K. Energy efficient hybrid intrusion detection system for wireless sensor networks[C]//Proc of International Conference on Electronics and Information Engineering. Washington DC: IEEE Computer Society, 2010: 25-29.

[4] Da SILVA A P R, MARTINS M H T, ROCHA B P S, *et al*. Decentralized intrusion detection in wireless sensor networks[C]//Proc of the 1st ACM International Workshop on Quality of Service and Security in Wireless and Mobile Networks. New York: ACM Press, 2005: 16-23.

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

[6] KRONTIRIS I, DIMITRIOU T, FREILING F C. Towards intrusion detection in wireless sensor networks[C]//Proc of the 13th European Wireless Conference. 2007: 1-4.

[7] HO J W, WRIGHT M, DAS S K. Distributed detection of mobile malicious node attacks in wireless sensor networks[J]. Ad Hoc Networks, 2011, 10(3): 512-523.

[8] WALPOLE R, MYERS R. Probability and statistics for engineers and scientists[M]. New York: Macmillan, 1985.

[9] The network simulator: ns-2[EB/OL]. http://www.isi.edu/nsnam/ns.

[10] HAI T H, HUH E N. Detecting selective forwarding attacks in wireless sensor networks using two-hops neighbor knowledge[C]//Proc of the 7th IEEE International Symposium on Network Computing and Applications. Washington DC: IEEE Computer Society, 2008: 325-331.

[11] 曹晓梅,韩志杰,陈贵海. 基于流量预测的传感器网络拒绝服务攻击检测方案[J]. 计算机学报,2007,30(10):1798-1805.