一种基于无监督免疫优化分层的网络入侵检测算法

林冬茂[1] 薛德黔[2]
(湖州师范学院现代教育技术中心 湖州 313000)[1] (湖州师范学院信息工程学院 湖州 313000)[2]

摘 要 高校网络被外网访问时,外网访问数据没有类别标记,导致数据识别特征不明显,传统的入侵检测模型不能有效提取出无监督外网访问数据中的识别特征,无法准确训练入侵检测模型,造成高校网络入侵检测准确度不高。为了解决这一难题,提出一种基于无监督免疫优化分层的入侵检测算法,即在免疫网络中对数据进行学习,用小规模的网络完成数据压缩,集中增强数据的识别特征,运用分层聚类方法分析网络,完成数据模型的建立。仿真实验表明,这种无监督入侵检测模型方法克服了高校网络外网访问数据的识别特性不明显,提高了高校网络入侵检测的准确率,取得了满意的结果。

关键词 高校网络,入侵检测,无监督,免疫网络
中图法分类号 TP391 **文献标识码** A

Network Intrusion Detection Algorithm Based on Unsupervised Immune Hierarchical Optimization

LIN Dong-mao[1] XUE De-qian[2]
(Modern Education Technology Center, Huzhou Teachers' College, Huzhou 313000, China)[1]
(Information and Engineering Institute, Huzhou Teachers' College, Huzhou 313000, China)[2]

Abstract When the external network accesses university network, the external network data has no category tags, so the data recognition is unclear. The traditional intrusion detection model can not effectively extract the identifying characteristics of the unsupervised external network accessing data, and intrusion detection model can not be accurately trained, which makes accuracy of the college network intrusion detection is not high. To solve this problem, this paper proposed a intrusion detection algorithm based on unsupervised immune hierarchical optimization to learn data in the immune network, complete the data compression using the small-scale network, focus on improving the identifying characteristics of the data, and analyze the network using hierarchical clustering method to complete the establishment of the data model. Simulation results show that this unsupervised intrusion detection model method overcomes the obvious identifying characteristics of the university external network accessing data, and improves the accuracy of the university network intrusion detection, achieves satisfactory results.

Keywords University network, Intrusion detection, Unsupervised, Immune network

随着计算机网络技术的发展,高校的教育模式也日益科技化,通过构建高校网络,共享教学资源和信息传输,能够为高校提供方便的远程教学和管理。然而高校网络在给教育教学带来便利的同时,由于网络的开放性、资源共享性,网路会受到各种病毒攻击或入侵[1],高校网络的安全也成为困扰人们的问题。高校网络的安全成为人们关注的焦点,其中入侵检测是保证网络安全的重要分支。入侵检测是计算机通过建立网络模型对访问数据进行特征识别,检测出其中的入侵数据,以达到保护高校网络的目的。随着高校网络的发展,对入侵检测的检测准确度提出了更高的要求。

高校网络的数据访问分成校内访问和外网访问两类[2],其中校内访问数据被高校网络规定,即有高校网络的监督,使得校内访问数据具有类别标记,其访问数据识别特征非常明显,网络入侵较易检测。而高校网络被外网访问时,因外网数据是没有监督的、复杂的、随意的,其访问数据一般没有类别标记,使得外网访问数据的识别特征不明显[3],入侵较难检测。传统的入侵检测模型针对识别特征不明显的外网访问数据,不能有效提取出无监督外网访问数据中的识别特征,无法准确训练入侵检测模型,造成高校网络入侵检测准确度不高,极易出现漏检测问题。

为了提高高校网络入侵检测的准确度,提出无监督入侵检测模型的检测方法,首先在免疫网络中对数据进行学习,用小规模的网络完成数据压缩,通过数据学习和压缩增强访问数据的识别特征,然后基于分层聚类方法分析网络,完成数据模型的准确建立和训练,通过将高校网络的访问数据在数据模型中训练检测,来实现高校网络的入侵检测。这样通过无监督入侵检测模型,克服了高校网络外网访问数据的识别特性不明显,提高了高校网络入侵检测的准确率。

到稿日期:2012-05-20 返修日期:2012-09-01 本文受浙江省自然科学基金(Y1101237)资助。
林冬茂(1973-),男,硕士生,主要研究方向为计算机网络安全与应用,,E-mail:ldm@hutc.zj.cn;薛德黔(1959-),男,教授,主要研究方向为计算机网络、数据技术、智能控制。

1 入侵特征的有效检测

高校网络的建设和发展，极大地方便了教育教学，并且为各高校之间的教学交流提供了很好的平台，实现了高效的网络在线教学、经验在线交流、高校联网管理等。然而随着高校网络的发展和广泛应用，方便地进行教学数据交流的同时也出现了病毒攻击、非法访问等网络入侵问题，威胁着高校网络的安全，直接影响教学工作的顺利进行。因此，高校网络的入侵检测成为了人们研究的重点问题。高校网络的数据访问请求分为校内访问和外网访问两类，具体的高校网络访问模型如图 1 所示。

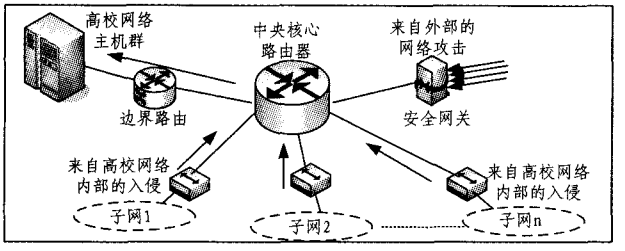

图 1 高校网络访问模型

由图 1 可以看出，高校网络的两类访问方式也带来了两类网络入侵：校内访问入侵、外网访问入侵。高校网络的入侵检测是通过获取网络访问数据，并建立入侵检测模型，利用访问数据的类别标记对模型进行训练优化，得到准确的分类模型。其中构建的入侵检测模型为：

$$R=K(x,x)-2\sum_j\beta_jK(x_j,x)+\sum_{i,j}\beta_i\beta_jK(x_j,x_i) \tag{1}$$

$$f(x)=syn(R-(K(x,x)-2\sum_j\beta_jK(x_j,x)+\sum_{i,j}\beta_i\beta_jK(x_j,x_i))) \tag{2}$$

式中，R 是提取的访问数据的识别特征，$K(x_i,x_j)$ 是高校网络的访问数据，β 是访问数据的类别标记。$f(x)$ 是模型的分类参数，其值直接影响入侵检测的准确度。由公式可以看出，通过对访问数据识别特征的提取，进而调整模型的参数，以提高模型分类识别的准确性[4,5]。根据访问数据的类别特征调整入侵检测模型的结构图如图 2 所示。

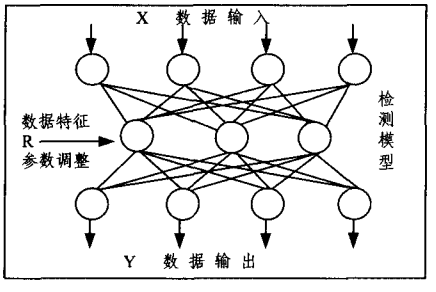

图 2 高校网络入侵检测模型训练调整结构图

由访问数据识别特征 R 和模型分类参数 $f(x)$ 的关系可以看出，访问数据的识别特征直接决定了此类参数的准确度，并且访问数据的识别特征越明显，即识别特征 R 提取得越充分、准确，则通过训练调整得到的模型越能得到有效的检测信息，以保证入侵检测的准确完成[6-8]。

2 基于无监督免疫优化分层的入侵检测算法

对于高校网络的内网访问，其访问数据具有类别标识，入侵检测的数据特征明显，极易检测，而无监督外网的访问数据没有类别标识，访问数据识别特征不明显而难于检测，为提高高校网络入侵检测的准确度，克服无监督入侵数据特征不明显的问题，提出无监督入侵检测方法，保证检测准确度，满足实际要求，下面将详细阐述这一方法。

2.1 免疫网络中对访问数据的学习和压缩

因为对于高校网络的外网访问是无监督的，其无监督访问数据没有类别标识，导致访问数据的识别特征不明显，影响入侵的检测。为克服无监督访问数据识别特征不明显，受网络调节算法的启发，通过模拟抗体的免疫过程构建免疫网络，对访问数据进行学习进化，初步实现访问数据的聚类分析。获取高校网络中的访问数据，并将得到的访问数据等效变换成免疫网络中的节点，通过在免疫网络中对节点计算分析实现对访问数据学习压缩的目的。设需要构建的免疫网络由矩阵 A 表示，此免疫网络矩阵是 N 行 M 列的，其中矩阵 A 的每一个行向量代表免疫网络中的每个节点，则免疫网络中两个节点 A_i 和 A_j 的相似性 s_{ij} 就是两个行向量之间的距离，表示为 $s_{ij}=d(A_i,A_j)$。并且 A_i 和 A_j 之间的差异越小，其相似性 s_{ij} 就越大。则在免疫网络中对访问数据进行学习压缩的具体步骤为：

1）免疫网络参数的初始化。设初始参数为 $i=1$，将获取的高效网络访问数据等效转换为初始免疫网络 A。

2）免疫网络学习。对于免疫网络 A 中的每个访问数据节点 x_i，计算其与周围所有免疫网络节点的相似度，然后在这些节点中选取与访问数据节点 x_i 相似度最高的 n 个网络节点进行克隆，设选取出的网络节点集合为 $\{A_{r_1},A_{r_2},\cdots,A_{r_n}\}$，则对这些节点进行克隆的具体运算公式为：

$$q_j=\mathrm{int}(N_c\times\frac{f(x_j,A_{r_j})}{\sum_{k=1}^{n}f(x_i,A_{r_i})}) \tag{3}$$

式中，N_c 表示克隆的规模，q_j 是对 A_{r_j} 克隆的结果。根据克隆的结果继续对节点进行变异操作：

$$CA_i=A_i-\alpha_i*(C\cdot A_i-x_i) \tag{4}$$

式中，α_i 是克隆个体 CA_i 的变异率，CA_i 就是克隆节点 A_{r_j} 的变异结果。在对免疫网络的所有节点完成选择、克隆、变异操作后，就完成了对访问数据的学习过程，然后重新选取相似度较高的 n 个节点重新构建免疫网络，初步增强访问数据的识别特征。

3）访问数据压缩。计算新的免疫网络中各节点之间的相似度 s_{ij}，选取一个压缩阈值 σ，将计算得到的所有节点相似度 s_{ij} 与压缩阈值对比，若 $s_{ij}>\sigma$，则删除网络节点 A_j，这样通过阈值比较和删除运算，就将访问数据中特征不明显的数据删除，并通过压缩运算提高访问数据的识别特征。

这样，根据免疫学的理论知识，通过模拟抗体的免疫刺激过程，将高校网络的访问数据等效成网络节点而构建免疫网络，并利用免疫网络的学习、压缩计算提高了无监督访问数据的识别特征，初步保证了入侵检测的准确性。

2.2 无监督入侵检测模型完成检测

在免疫网络中对高校网络访问数据学习和压缩后，增强了无监督访问数据的识别特征。利用这些识别特征明显的访问数据，并基于分层聚类算法，对高校网络的入侵检测模型进行训练优化。由分层聚类算法的对偶原则，将访问数据对偶成一个优化分类函数：

$$Q(A)=\sum_{i=1}^{n}A_i-\frac{1}{2}\sum_{i,j=1}^{n}A_iA_j(x_i,x_j) \tag{5}$$

式中，$Q(A)$就是高校网络访问数据的优化分类函数。根据式(2)的入侵检测模型构建参数函数，具体构建的入侵检测模型表示为：

$$f(x)=\mathrm{syn}(R-(A(x,x)-2\sum_j\beta_jA(x_j,x)))$$

$$=\mathrm{sgn}(\sum_{i=1}^{n}\alpha_iR(x_i\cdot x)+b) \tag{6}$$

式中，b是入侵检测模型的分类阈值。根据访问数据计算得到的优化分类函数，对入侵检测的模型进行优化训练：

$$A(x,x)=\sum_jQ(A_j)^2\beta_j-\sum_{i,j}\beta_i\beta_jQ(A_i)Q(A_j) \tag{7}$$

$$R=A(x,x)-2\sum_j\beta_jA(x_j,x)+\sum_{i,j}\beta_i\beta_jA(x_j,x_i) \tag{8}$$

$$A(x_i,x_j)=A(x,x)\exp\left(-\frac{\|x_i-x_j\|^2}{2\alpha^2}R\right)$$

$$=\tan Rh(a(x_i\cdot x_j)+c) \tag{9}$$

式中，R是优化训练得到的高校网络无监督访问数据的识别特征，$A(x_i,x_j)$是通过分层聚类算法并根据无监督访问数据训练得到的优化入侵检测模型。算法的描述如下所示。

算法利用上述原理和公式设计入侵检测分类器，输入训练数据集和检测数据集，输出检测数据的类型(正常、入侵或异常)。算法描述如下：

1)训练传统的入侵特征分类器，得到初始特征，并用其来构成免疫学习样本 $wx+b=0$；

2)计算中心向量 x_0；

3)根据学习分类面和 x_0 计算半径 R 和 $2R$；

4)根据式(10)计算准确度函数 $u(x_j)$；

5)获得模糊训练集：$\{(x_1,y_1,u(x_1)),(x_2,y_2,u(x_2)),\cdots,(x_l,y_l,u(x_l))\}$；

6)训练模糊训练点，构造最优决策函数，获得最终的入侵检测结果。

这样根据免疫网络学习压缩后得到的识别特征明显的高效网络访问数据，基于分层聚类方法训练得到的优化入侵检测模型，就完成了无监督入侵检测模型的最终构建。并且此模型通过无监督模型训练具有较高的检测准确度。在检测高校网络中的入侵时，通过数据采集器获取对于高校网络的多组访问数据，然后将访问数据输入无监督入侵检测模型中完成最终的入侵检测，得到准确检测结果。

3 仿真实验及结果分析

高校网络因其应用的特殊性和频繁性，常易受网络入侵攻击，直接影响高校网络的安全，干扰正常的教学秩序，因此高校网络的入侵检测研究就成为了必然。一般地，高校网络的入侵检测评价指标为准确度和漏检率，用以标识入侵检测方法的检测性能。其中，高校网络入侵检测的准确度、漏检率指标为：

$$\text{准确度 } z=\frac{\text{实验检出的攻击数}}{\text{实际的攻击数}} \tag{10}$$

$$\text{漏检率 } L=\frac{\text{实际攻击数}-\text{检测出的攻击数}}{\text{实际的攻击数}} \tag{11}$$

为验证文中提出的无监督入侵检测模型的检测性能，进行仿真实验。

设计仿真实验，验证高校网络入侵检测模型的检测准确度问题。实验利用 Visual C++软件构建高校网络的仿真环境，使用C语言实现入侵检测算法的编程。构造600个高校网络的访问数据包，其中200个访问数据来自于无监督的外网访问请求，且200个外网访问请求中有20个是攻击入侵，同时在剩余的400个校内访问数据包中有20个入侵攻击。分别利用传统的入侵检测方法和文中提出的无监督入侵检测模型的检测方法对构建的此高校网络进行入侵检测，在600个访问请求数据包中200个外网访问请求数据包是随机存放的，逐一将600个访问请求输入到仿真高校网络中，并采用两种入侵检测方法进行检测，分别记录两种方法检测出的攻击数，根据式(10)、式(11)计算两种方法的入侵检测准确度和漏检率，将实验结果数据列表对比，得到实验数据对比结果如图3所示。

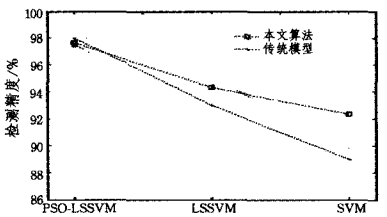

图3 不同算法下网络入侵检测精度比较

具体的统计结果如表1所列。

表1 实验数据对比表

检测方法	传统方法	文中方法
(校内)检出攻击数/个	20	20
(外网)检出攻击数/个	8	19
检测准确度/%	70	97.5
检测漏检率/%	30	2.5

由上述实验结果可知，对于高校网络的校内访问，由于其访问数据受高校网络的监督都有类别标识，访问数据的识别特征很明显，因此传统的入侵检测模型能够准确检测出校内访问中的入侵，由表1数据可知，对于校内访问中的20个如入侵攻击都能检出，然而对于高校网络的外网访问，因其无监督性访问数据没有类别标识，使得无监督访问数据识别特征不明显，传统方法在检测物件的外网访问数据时，不能有效提取出无监督外网访问数据中的识别特征，无法准确训练入侵检测模型，表1数据显示出20个外网访问中的入侵攻击只能检出8个，造成高校网络入侵检测准确度不高，仅为70%，且漏检率为30%，不能保证高校网络入侵检测的准确度要求。而文中提出的无监督入侵检测模型的检测方法，通过在免疫网络中对数据进行学习，用小规模的网络完成数据压缩，集中增强数据的识别特征，然后基于分层聚类方法分析网络，完成数据模型的建立，保证入侵检测的准确性。由实验结果可知，其不仅能够将高校网络校内访问中的20个入侵攻击全部检出，而且对于识别特征不明显的外网访问数据能够将其中的19个入侵攻击检测出来，使得高校网络入侵检测的准确度提高为97.5%，且漏检率大大降低，仅为2.5%。实验表明，这种无监督入侵检测模型方法，能够克服高校网络外网访问数据的识别特性不明显，提高高校网络入侵检测的准确率，取得了满意的结果。

结束语 提出一种无监督入侵检测模型的检测方法，首先在免疫网络中对数据进行学习，用小规模的网络完成数据压缩，集中增强数据的识别特征，然后基于分层聚类方法分析

(下转第191页)

20s 时，TxOutPower=－1dBm 条件下感染的节点数目接近 TxOutPower=－10dBm 情况下的一倍。

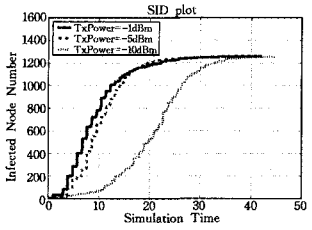

图 5　节点发射功率对病毒传播的影响

图 6 示出不同的节点监听时间周期比例(duty cycle=0.1,0.2,0.5)情况下，WSN 中的感染节点数目的变化情况，其中 TxOutPower=－10dBm。当 duty cycle 较大时，节点有更多的监听时间，能够减少数据包的延时，也同时加快了病毒数据包在 WSN 中的传播。当 duty cycle 取 0.1 时，病毒可以在 10s 左右扩散到整个 WSN 中，其余节点应为死亡节点，而且传播所消耗的时间大多为节点感染重启所用时间，可见 duty cycle 增大会极大地加快病毒传播速度。但一般的 WSN 应用中考虑到节约能量使用情况，会将 duty cycle 设置较小，这也会比较明显地降低病毒传播速度，对病毒传播可以起到一定的抑制作用。

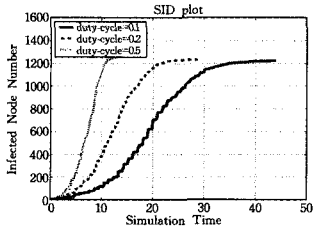

图 6　节点休眠时间比例对病毒传播的影响

结束语　通过分析 WSN 的基本特征，结合病毒的具体感染过程特点，建立了 WSN 中的病毒传播 SID 模型。重点研究了 WSN 的节点空间分布方式、无线通信机制等对病毒在 WSN 中传播的影响。仿真结果表明，本文中的 SID 模型能够较好地针对 WSN 特征描述病毒在 WSN 中传播的动态性。WSN 网络的散布情况会对病毒传播产生较大影响，休眠唤醒工作模式的采用能够相应抑制病毒的传播速度。文中 SID 对病毒在 WSN 中传播的模拟分析为病毒的防御机制研究提供了基础。WSN 中病毒的有效侦测以及对病毒的控制机制将是下一步研究的重点。

参 考 文 献

[1] Yang Yi, Zhu Sen-cun, Cao Guo-hong. Improving sensor network immunity under worm attacks: a software diversity approach [C]// Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing. 2008: 149-158

[2] Gu Qi-jun, Noorani R. Towards self-propagate mal-packets in sensor networks[C]// Proceedings of the first ACM Conference on Wireless Network Security. 2008

[3] Khayam S A, Radha H. Using Signal Processing Techniques to Model Worm Propagation over Wireless Sensor Networks [J]. IEEE Signal Processing Magazine, 2006, 23(2): 164-169

[4] Khayam S A, Radha H. A topologically-aware worm propagation model for wireless sensor networks [C]// Distributed Computing Systems Workshops. 2005. 25th IEEE International Conference. 2005: 210-216

[5] De P, Liu Yong-he, Das S K. Deployment-aware modeling of node compromise spread in wireless sensor networks using epidemic theory [J]. ACM Transactions on Sensor Networks, 2009, 5(3): 413-425

[6] Bo Sun, Yan Guan-hua, Yang Xiao. Self-propagating mal-packets in wireless sensor network: Dynamics and Defense Implications [J]. Ad Hoc Networks, 2009, 7(8): 1489-1500

[7] Wang Xia9o-ming, Li Ying-shu. An Improved SIR Model for Analyzing the Dynamics of Worm Propagation in Wireless Sensor Networks [J]. Chinese Journal of Electronics, 2009, 18(1): 8-12

[8] 宋玉蓉，蒋国平. 无线传感器网络中恶意软件传播研究[J]. 南京邮电大学学报：自然科学版，2009，4：1-7

(上接第 182 页)

网络，完成数据模型的建立，保证入侵检测的准确性。仿真实验表明，这种无监督入侵检测模型方法，克服了高校网络外网访问数据的识别特性不明显，提高了高校网络入侵检测的准确率，具有一定的使用价值。

参 考 文 献

[1] 徐子豪，张腾飞. 基于语音识别和无线传感网络的智能家居系统设计[J]. 计算机测量与控制，2012(01)

[2] Deng H, Zeng P-A, Agrawal D P. An Unsupervised Network Anomaly Detection System Using Random Projection Technique [C]// Proceedings of the 2003 International Workshop on Cryptology and Network Security. 2003: 593-598

[3] 李辉，等. 基于支撑向量机的网络入侵检测[J]. 计算机研究与发展，2003，40(6)：799-807

[4] Du H, Jiao L, Wang S. Clonal Operator and Antibody Clone Algorithms[J]. Proceedings of the Firs! International Conference on Machine Learning and Cybernetics, 2002, 4(11): 506-510

[5] 杨杰，李中文. 神经网络在高校科研能力中的评价研究[J]. 计算机仿真，2011(5)：384-387

[6] 田俊峰，张晶，毕志明. 基于改进 RBF 神经网络的入侵检测研究[J]. 计算机工程与应用，2008，44(31)：135-138

[7] 孙晓艳，郑淑丽，沈洪伟. 优化的 RBF 神经网络在入侵检测中的应用[J]. 合肥工业大学学报：自然科学版，2008，31(11)：1794-1797

[8] 康世瑜. 基于数据挖掘和特征选择的入侵检测模型[J]. 微电子学与计算机，2011(8)

[9] 唐鑫，高博. 高校校园网安全性研究与对策[J]. 重庆理工大学学报：自然科学版，2011，25(2)：91-95