一种改进蚁群聚类的入侵检测方法

姜 参,王大伟
(渤海大学 管理学院,辽宁 锦州 121013)

摘 要:入侵检测是网络信息安全的一个重要方面。针对现有的入侵检测对各类攻击不全面以及在检测率低误检率高的缺点,文中提出了一种改进的蚁群聚类的入侵检测方法。该方法对蚁群聚类算法的收敛速度方面和易陷入局部最优问题进行了改进,在优化过程中引进K-means算法以及信息熵,从而使其能够对信息素的更新进行自动的调整,提高了聚类速度和效果。进而设计了网络入侵检测系统。实验结果表明,该方法不仅提高了检测率,而且降低了误检率,对于各大类攻击都能够进行精确的检测。

关键词:网络安全;入侵检测;数据挖掘;蚁群聚类;聚类分析
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2013)12-0139-04
doi:10.3969/j.issn.1673-629X.2013.12.033

An Improved Ant Colony Clustering Method for Intrusion Detection

JIANG Shen, WANG Da-wei
(School of Information Science & Engineering, Bohai University, Jinzhou 121013, China)

Abstract: Intrusion detection is an important aspect of the network information safety. For the disadvantage that the existing intrusion detection method is not comprehensive of various kinds of attack and has lower detection rate and the higher fault detection rate, an improved ant colony clustering method for intrusion detection is proposed. The convergence rate of ant colony cluster algorithm is improved. In the optimization process, the information entropy is introduced to prevent into local optimal, and thus the method can adjust automatically the pheromone updating and improve the clustering speed. And follow on, the intrusion detection system is designed. The experimental results show that the method not only improves the detection rate, but reduces the fault detection rate, and can detect precisely the various kinds of attacks.

Key words: network safety; intrusion detection; data mining; ant colony clustering; cluster analysis

0 引 言

随着互联网的高速发展,人们在享受互联网带来的各种智能便捷服务的同时,也受到各种信息安全问题的困扰,其中最主要的是各种类型的入侵行为。入侵检测就是一种对入侵行为进行检测的安全防护技术,它能够提供误操作以及内外部攻击的实时保护,在网络系统被入侵之前进行拦截。入侵检测系统则能够保护重要资源和网络,一个性能比高的入侵检测系统不仅能够提高网络的安全性,而且可以节省能量,提高网络的快速性和精确性。入侵检测已经成为继防火墙后的最主要的安全防护技术,已经成为国内外学者的热门研究课题[1-3]。

目前,网络入侵检测方法被研究的范围主要包括数据挖掘和模式识别,研究工作主要有数据加密、防火墙、访问控制等[4-5]。网络的智能性使得入侵检测系统也被引入了智能性。汪淑丽等提出了一种基于蜜罐的入侵检测系统,并设计和实现了基于人工神经网络的HoneypotIDS,该系统通过引入蜜罐技术主要解决了对一些未知攻击的识别,但是网络配置较复杂[6-7]。安辉耀等人提出了一种用于网络入侵检测的群体协同人工淋巴细胞模型,该模型在入侵检测中主要引入了淋巴细胞的群体刺激机制,能够检测蠕虫、木马以及拒绝服务式攻击,不足是对非人工免疫网络的检测率就较低[8]。针对检测速度慢问题,李正洁等人将免疫原理、移动Agent技术和量子粒子群优化算法相结合,提出了组合入侵检测模型,提高了检测速度,但是量子粒子群算法存在陷入局部最优问题[9]。余小华等人提出了一种蚁群优化的WSN分布式入侵检测模型,能

收稿日期:2013-04-21 **修回日期**:2013-07-22 **网络出版时间**:2013-09-29
基金项目:国家自然科学基金资助项目(61273072)
作者简介:姜 参(1979-),男,硕士,讲师,研究方向为无线Mesh网、网络算法;王大伟,硕士,副教授,研究方向为网络与通信技术。
网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20130929.1548.060.html

够对无线传感器网络进行入侵检测,并且提高了对未知攻击的检测率,缺点也依然是存在陷入局部最优问题[10]。

根据入侵检测的研究基础,文中提出了一种改进的蚁群聚类的入侵检测方法,首先在对蚁群聚类方法进行研究的基础上,对其收敛速度慢以及容易陷入局部最优的缺点进行了改进;然后加上蚁群聚类算法具有可扩展、自组织和健壮性等特征,利用改进的蚁群聚类的特性,设计并实现了入侵检测方法以及体系结构;最后进行了实验测试,结果表明提出的方法不仅提高了检测率,而且误检率降低,具有较高的检测效率,并且对当下的五大类攻击都能够进行精确有效的检测。

1 改进的蚁群聚类的入侵检测方法

蚁群聚类算法[11](Ant Colony Clustering Algorithm,ACCA)是蚁群优化算法(Ant Colony Optimization,ACO)进行衍生后的一类适用于数据挖掘的算法,它是继模拟退火算法、遗传算法等之后的一个新的优化算法。它的主要原理是根据蚂蚁在找寻食物的过程中在路径上积累的信息素来发现最短路径,选择一条路径的蚂蚁越多,该条路径上积累的外激素就越多,因此就会有更多的蚂蚁选择这条路径,就此形成一种正反馈的机制,由此蚂蚁就能选择到最短路径。

蚁群优化算法的关键要点是信息素更新和蚂蚁概率选择路径。该算法具有发现最优解和鲁棒性高等优点,但是由于蚁群优化算法采用的是随机选择策略,从而导致了进化速度较慢,出现停滞现象。为了对算法的收敛速度进行提高,对信息素更新策略进行了改进,引入了信息熵对收敛速度进行更新。

改进的蚁群聚类算法在 1.1 节进行详细介绍,在改进的蚁群聚类算法基础上在 1.2 节给出入侵检测方法,最后对入侵检测体系结构进行描述。

1.1 改进的蚁群聚类算法

蚁群聚类算法在聚类中心进行隶属度分配时,经常出现隶属度相同但位置不同的情况,这就对计算的准确性造成影响。信息熵能够对数据信息进行利用,并结合数据的定性和定量信息,进行理想的聚类中心评价。

假设 τ_{ij} 为城市 i 和城市 j 之间路径上的信息素,则两城市间的信息素比例 p_{ij} 为:

$$p_{ij}=\frac{\tau_{ij}}{\sum_{r=1}^{n}\sum_{s=1}^{n}\tau_{rs}} \tag{1}$$

则种群的信息熵 H 为:

$$H=-\sum_{i=1}^{n}\sum_{j=1}^{n}p_{ij}\ln p_{ij} \tag{2}$$

通过实验发现,引入信息熵的蚁群聚类算法由于初始数据过于分散,导致了算法收敛速度太慢。为了解决这一问题,在信息熵蚁群聚类算法的预计算过程中又引入了 K-means 算法,该算法收敛速度较快,利用其快速确定食物源,然后聚类过程分为两个阶段:

① 比较聚类之间熵值,找出与对象熵值的最小类,把对象划入该类中,然后进行聚类再比较。这样改进了 K-means 算法的初始分类的粗糙。

② 进行信息蚁群聚类的全局聚类,其一有效地提高了聚类中心的精确度,其二由于计算的并行分布式特性,加快了收敛速度,提高了聚类的效率。

改进后的蚁群聚类算法初始阶段,所有的边都具有相同的信息素,信息熵此时最大,收敛速度也最慢;在信息素在较短边的积累,信息熵逐渐下降,收敛速度逐渐加快,同时也逐渐陷入局部最优,同时也可能停滞,此时可以让对比度减弱。对于熵值大收敛速度缓慢的边,采取增强对比度,放大较大概率缩小较小概率。这种信息熵的对比度调整策略不仅对选择机制的反馈度进行了加强利用,从而算法的收敛速度得到了加快,而且能对退化解进行接收并对新解进行搜索,避免了局部最优。

1.2 入侵检测方法

利用改进的蚁群聚类算法对入侵检测方法进行了设计。方法中的数据对应于蚁群中的蚂蚁,聚类中心对应于蚂蚁寻找的食物源,数据的聚类过程对应于蚂蚁对食物源的寻找过程。

假设初始数据集为 $D=\{D_i=(d_{i1},d_{i2},\cdots,d_{im}),i=1,2,\cdots,n\}$,则利用蚁群聚类算法的入侵检测方法基本流程如下:

(1)初始化聚类。初始化主要是对数据实例以及参数进行标准化,目的是消除因不同度量而影响聚类结果。初始化后的数据实例为:

$$d'_{ij}=\frac{d_{ij}-\overline{d_j}}{S_j} \tag{3}$$

其中,$\overline{d_j}=\frac{1}{n}\sum_{i=1}^{n}d_{ij}$,$S_j=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_{ij}-\overline{d_j})^2}$。

信息熵具有对聚类好坏评估的理论特性。初始化的同时,计算每个数据的信息熵并进行比较,把最小的信息熵值作为基准熵值。基准熵值越小,说明数据的相似度越大。根据相似标准把每个数据分别分配给每个聚类,形成初始聚类数据。

(2)计算每个数据记录之间的加权欧式距离:

$$W_{ij}=\sqrt{\sum_{k=1}^{m}p_k(d_{ik}-d_{jk})^2} \tag{4}$$

(3) 计算信息素。假设聚类的半径为 r, 则各条路径上的信息素为:

$$\tau_{ij}=\begin{cases}1 & W_{ij}\leqslant r\\ 0 & W_{ij}>r\end{cases} \tag{5}$$

(4)聚类归并。根据加入信息熵的信息素概率进行聚类归并。如果 $p_{ij}\geqslant p_0$,则将 D_i 归并到邻域 D_j。

(5) 确定聚类中心。根据计算得到的基准熵值计算蚁群聚类算法所需要的初始聚类中心:

$$\overline{O_j}=\frac{1}{J}\sum_{k=1}^{J}D_k \tag{6}$$

(6)全局聚类。首先确定各聚类间的偏离误差:

$$E_j^{'}=\sum_{k=1}^{j}\sqrt{\sum_{i=1}^{m}(d_{ki}-d_{ji})^2} \tag{7}$$

则整体误差为:

$$E=\sum_{j=1}^{k}E_j^{'} \tag{8}$$

(7)如果 $E\leqslant E_0$,对聚类中心和聚类个数进行输出,反之,转(2)步骤进行迭代。

(8) 确定正常行为和异常行为。聚类后根据数据记录数量进行排序。Portnoy 的实践假设:入侵行为的数目远小于正常行为,并且正常行为和入侵行为差异非常大。根据此假设,设定一个阈值,小于此阈值的聚类为异常行为,反之,都为正常行为。

因此,通过该方法得到的类划分能够对正常行为和异常行为进行准确的区分。

1.3　入侵检测体系结构

入侵检测体系主要包括四个模块,各模块具体结构如图 1 所示。

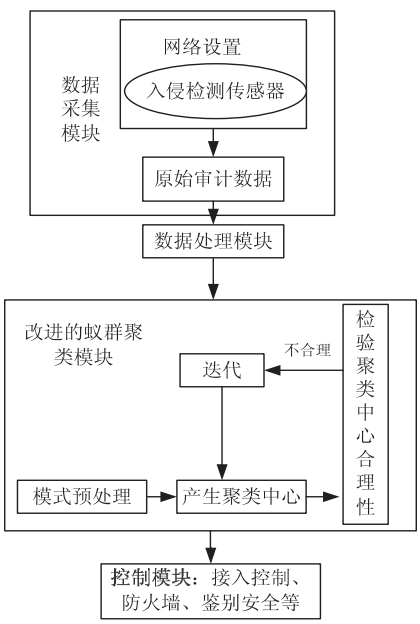

图 1　改进的蚁群聚类的入侵检测方法模块结构图

① 数据采集模块。这是入侵检测的基础模块,首先根据 TCP 数据包序列整合成连接记录,每个连接记录包括网络协议、服务号、IP 地址、连接起末时间等,这些特性用一个特征向量进行标示。

② 数据处理模块。该模块的主要目的是把原始数据映射成适合算法的分布,但模块关联度仍为同一类别。因为一些数据集记录的各个属性类型不同,有字符型、数值型、连续型、离散型等,不统一的标准对蚁群聚类算法相似度计算会产生影响。因此,就需要对数据进行预处理,分割以及过滤,进行标准化属性值。数据预处理不仅能够提高蚁群聚类算法的效率,而且可以对算法的适用范畴进行扩大。

③蚁群聚类模块。该模块主要是结合改进的蚁群聚类算法设计的入侵检测处理模块,目的在于区分同一类别的模式。聚类的对象为每个特征向量,用改进的蚁群聚类算法进行分析和处理,分离正常行为和入侵行为。

④ 控制模块。根据蚁群聚类模块的结果对网络行为进行入侵检测。如果检测结果为入侵行为,则自动启用防火墙等机制进行及时报警,同时进行控制防止入侵或留下病毒等;如果检测结果为正常行为,则自动继续进行检测。

2　实验结果及分析

2.1　实验环境

从美国麻省理工学院的 Lincoln 实验室的仿真测试数据集和 KDD 网络数据集中选择 600 万条数据记录,大概包括了两个月的网络流量。这些数据记录中包含多种网络环境中的模拟入侵,共有 22 种正常类型和入侵类型。每个数据实例包括 42 个特征向量,并且被标记为正常类型或者入侵类型。实验环境以及参数设置如下:

① 处理器:2.2 GHz Intel Pentium4;

② 内存:1 GB DDR2;

③ 操作系统:Windows XP;

④ 从 KDD99 数据集中选取 8 个数据子集:50,100,300,500,800,1 000,1 500,2 000,正常行为和异常行为比例为 18:1;

⑤ 检测的攻击主要有五大类:DoS[12],R2L,U2R,PROBE 和 Sinkhole 攻击。分别对这五大类攻击对各个层次的数据集进行检测率和误报率的测试和分析,这是对网络入侵检测的重要分析。其中:

检测率(DR)= 检测出的入侵实例记录数/总的入侵实例记录总数;

误报率(FAR)= 误判为入侵行为的正常实例记录数/总的正常实例记录数。

⑥ 在同样的环境和初始数据下,与原始的蚁群聚类方法进行了对比实验。

2.2　性能分析

检测率和误报率能对算法的入侵检测能力进行充

分的反映,对算法性能优劣进行有效分析评价。首先对该方法对五种攻击的检测率进行了数据综合测试,结果如图 2 所示。

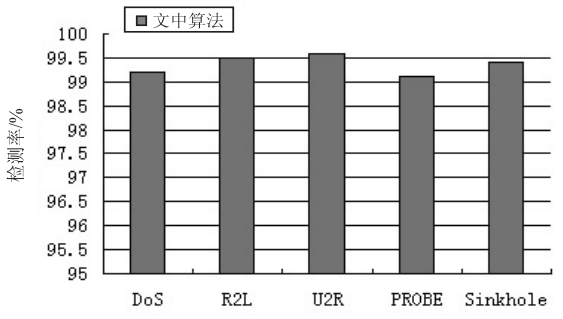

图 2 对各类入侵的检测率

从图 2 可以看出,文中算法对这几大类入侵攻击都能够进行有效检测,并且检测率都在 95% 以上。为了进一步对算法的性能进行测试,从 KDD99 数据集中选取 8 个数据子集:50,100,300,500,800,1 000,1 500,2000,分别采用文中方法以及蚁群聚类算法进行测试并比较分析了检测率和误检率,如图 3 和图 4 所示。

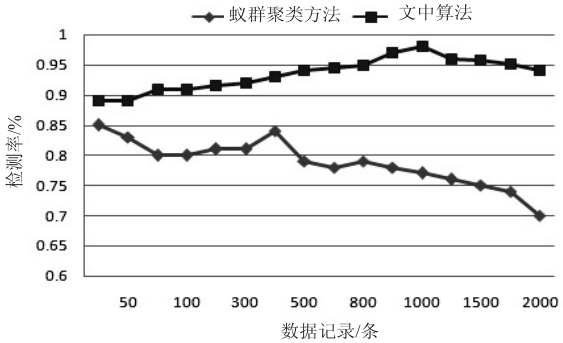

图 3 检测率对比图

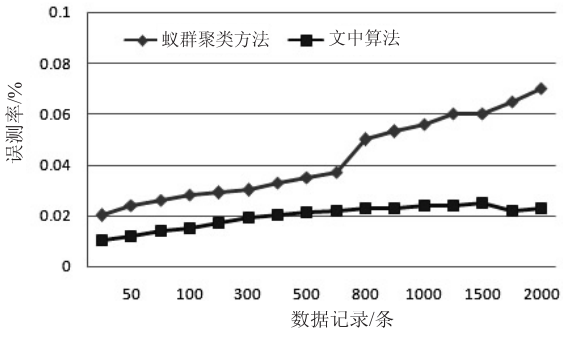

图 4 误检率对比图

从图 3 可以看出,在实验数据比例的入侵行为下,文中方法的检测率较高,并且随着实例记录条数的改变,检测率也很稳定,即该方法检测率不受实例记录条数的影响,检测率都保持在 90% 以上。而未改进的蚁群聚类算法的检测率较低,这是因为原始的蚁群聚类算法采用的是随机选择,收敛速度慢且容易停滞,会陷入局部最优,这也是造成算法检测率时高时低不稳定的原因。

从图 4 的误检率对比图可以看出,相对于原始的蚁群聚类方法,文中方法的误检率明显较低,引进的信息熵用于聚类的计算过程加快了收敛速度,防止了陷入局部最优情况,因此区分正常行为和入侵行为的误报就降低了很多,并且性能很稳定。

对于五大类攻击的检测率和误检率根据测试数据进行了综合整理,如表 1 所示。

表 1 五大类攻击的检测率和误检率

攻击类型	DR/%		FAR/%	
	文中方法	原始蚁群聚类方法	文中方法	原始蚁群聚类方法
DoS	99.2	98.1	2.5	2.6
R2L	99.5	98.9	2.2	2.4
U2R	99.7	97.9	1.7	2.2
PROBE	99.4	96.5	2.3	2.6
Sinkhole	98.8	94.3	2.6	3.9

通过表 1 的数据可以看出,对于 R2L 和 DoS 攻击,文中方法与原始的蚁群聚类算法的误检率基本相同,但是文中方法的检测率明显相对较高。在其他攻击中,文中的检测率均高于原始蚁群聚类方法,误检率则都较低,且稳定性较高。文中改进的蚁群聚类的入侵检测方法不仅提高了检测率,而且误检率减小,具有较高的检测效率。

3 结束语

文中在蚁群聚类算法入侵检测模型的基础上进行了改进,针对收敛速度过慢以及容易陷入局部最优导致的检测率低的问题,提出了 K-means 算法和信息熵的蚁群聚类的入侵检测方法,解决了现有入侵检测存在的检测率低和误报率高的问题,并且使用于目前的各大类攻击。与其他聚类算法相比,文中方法能够较准确地划分数据类型,提高收敛速度,且性能稳定。但是该方法对于未知攻击的检测效率较低,这有待进一步研究。

参考文献:

[1] Saravanakumar S, Kumar A, Anandaraj S, et al. Algorithms based on artificial neural networks for intrusion detection in heavy traffic computer networks[C]//Proc of 2011 international conference on advancements in information technology with workshop of ICBMG. [s. l.]:[s. n.],2011:6-12.

[2] Nakkeeran R, Albert T A, Ezumalai R. Agent based efficient anomaly intrusion detection system in Ad hoc networks[J]. International journal of engineering and technology, 2010, 2(1):52-56.

[3] 黄红艳,安素芳. 数据流聚类算法在入侵检测中的应用

(下转第 146 页)

line、Microsoft SharePoint Online、Microsoft Office Live Meeting、Microsoft Lync Online 以及 Microsoft Dynamic CRM 等软件即服务功能,也就是微软推出的 Microsoft Office365 项目,在操作系统与应用程序执行环境方面,Windows Azure 的 Guest OS 是以差异化磁盘进行操作系统部署,所有系统设置无法被应用程序更新。

4 结束语

云计算等开放网络服务环境的安全问题成为企业是否采用云计算等开放网络服务的重要疑虑之一。关注基于云计算等开放网络服务的 IT 应用中的安全设计,并去解决开放网络服务所引发的安全问题势在必行。文中基于 Windows Azure 云平台,结合微软 Office365 项目和微软 Azure 架构,研究 Windows Azure 云平台安全性要求,分别从 IaaS、PaaS、SaaS 三个角度分析研究了 Windows Azure 架构下云平台当下对云安全所提供的保护措施以及相应的解决方案,保证 Windows Azure 云平台的安全性,从而进一步促进云计算安全走向成熟。

参考文献:

[1] Jennings R. 云计算与 Azure 平台实战[M]. 王 鑫,丁斌,译. 北京:清华大学出版社,2011.

[2] 朱明中. Windows Azure 实战手记[M]. 北京:中国水利水电出版社,2011.

[3] 赵立伟,方国伟. 让云触手可及:微软云计算实践指南[M]. 北京:电子工业出版社,2010.

[4] 杨文志. 云计算技术指南-应用、平台与架构[M]. 北京:化学工业出版社,2010.

[5] 王 鹏. 云计算的关键技术与应用实例[M]. 北京:人民邮电出版社,2010.

[6] John R, James R. Cloud computing: Implementation, management, and security[M]. Beijing: China Machine Press, 2010.

[7] Rings T, Grabowski J, Schulz S. Grid and cloud computing: Opportunities for integration with the next generation network [J]. Grid computing, 2009(7): 375-393.

[8] Zhang Liangjie, Zhou Qun. CCOA: Cloud computing open-architecture[C]//Proc of 2009 IEEE international conference on Web Services. New York: IEEE Computer Society Press, 2009: 607-616.

[9] Yousef L, Bulrico M, Silva D. Toward a unified ontology of cloud computing [EB/OL]. 2010. http://www. collabogce. org/gce08/images/7/76/LamiaYousef. pdf.

[10] 冯登国,张 敏,张 妍,等. 云计算安全研究[J]. 软件学报,2011,22(1):71-83.

[11] 陈尚义. 浅谈云计算安全问题[J]. 网络安全技术与应用,2009(10):20-22.

[12] 李 玮. 云计算安全问题研究与探讨[J]. 电信工程技术与标准化,2012(4):44-49.

[13] 张爱玉,邱旭华,周卫东,等. 云计算与云计算安全[J]. 中国安防,2012(3):89-91.

[14] 杨 健,汪海航,王 剑,等. 云计算安全问题研究综述[J]. 小型微型计算机系统,2012,33(3):472-479.

[15] 陈 全,邓倩妮. 云计算及其关键技术[J]. 计算机应用,2009,29(9):2562-2567.

(上接第 142 页)

[J]. 计算机工程与应用,2012,48(20):112-116.

[4] Chen Lian. A kind of improved attribute reduction algorithm in intrusion detection application research based on rough sets theory[C]//Proc of international conference on computer science and automation engineering (CSAE). [s. l.]: [s. n.], 2011: 707-710.

[5] 樊爱京,杨照峰. 用于网络入侵检测的模式匹配新方法[J]. 计算机应用,2011,31(11):2961-2964.

[6] 汪淑丽. 基于支持向量机的无线传感器网络的入侵检测系统[J]. 传感器与微系统,2012,31(7):73-76.

[7] 汪 洁,杨 柳. 基于蜜罐的入侵检测系统的设计与实现[J]. 计算机应用研究,2012,29(2):667-671.

[8] 安辉耀,吴泽俊,王新安,等. 用于网络入侵检测的群体协同人工淋巴细胞模型[J]. 通信学报,2010,31(9):122-130.

[9] 李正洁,李永忠,徐 磊. 免疫 Agent 和量子粒子群优化的入侵检测方法研究[J]. 计算机工程与应用,2012,48(1):102-104.

[10] 余小华,黄灿辉,陈 瑛. 一种蚁群伏化的 WSN 分布式入侵检测模型[J]. 计算机工程与应用,2012,48(9):78-82.

[11] Ma C, Cao A, Zhou Y. Primary research on improved algorithm of ant colony clustering combination[J]. Journal of Shenyang Jianzhu university (natural science), 2011, 27(4): 798-803.

[12] Xi O, Bin T, Qi L, et al. A novel framework of defense system against DoS attacks in wireless sensor networks[C]//Proc of 2011 7th international conference on wireless communications, networking and mobile computing. [s. l.]: IEEE, 2011: 1-5.

一种改进蚁群聚类的入侵检测方法

作者： 姜参， 王大伟， JIANG Shen， WANG Da-wei
作者单位： 渤海大学 管理学院,辽宁 锦州,121013
刊名： 计算机技术与发展
ISTIC
英文刊名： Computer Technology and Development
年，卷(期)： 2013(12)