学校编码：10384　　分类号＿＿＿密级＿＿＿
学　　号：X2007223005　　UDC＿＿＿

厦门大学

工　程　硕　士　学　位　论　文

一种基于蚁群聚类的异常网络入侵检测算法

Detection of anomaly network intrusion based on Ant Colony Clustering

林肖莹

指导教师姓名：罗德林　副教授

专　业　名　称：控制理论与控制工程
论文提交日期：2013 年 05 月
论文答辩时间：2013 年 08 月
学位授予日期：2013 年　　月

答辩委员会主席：＿＿＿＿＿＿
评　　阅　　人：＿＿＿＿＿＿

2013 年 08 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行)》。

另外，该学位论文为（　　）课题（组）的研究成果，获得（　　）课题（组）经费或实验室的资助，在（　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

2013　年　月　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（　　）1.经厦门大学保密委员会审查核定的保密学位论文，于　　年　　月　　日解密，解密后适用上述授权。

（　　）2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

2013　年　　月　　日

摘　要

随着计算机的普及和网络的快速发展，安全威胁迅速增加，需要采取有效的措施保障计算机系统和网络的安全运行。入侵检测技术是近 20 年来出现的一种动态的监控、预防或抵御系统入侵行为的安全机制。它能在不影响网络性能的情况下对网络进行监测，从而提供对内部攻击、外部攻击和误操作的实时保护。

通过对现有聚类方法进行深入的分析、比较，得知聚类分析可扩展性强，实效性高，适用于大规模数据集。其既可以作为一个独立的工具类使用，也可以作为其它算法的预处理步骤。由于蚁群聚类算法不必预先设定聚类的数目，适用于无监督聚类的异常入侵检测，因此使用蚁群聚类作为异常入侵检测的预处理步骤，在不具备完整领域知识背景的情况下完成入侵检测。

本文系统地研究了入侵检测的基本理论，其中包括入侵检测的定义、种类和入侵检测的模型，并分析了当前存在的问题。从中发现现有的异常入侵检测算法难以检测包含混合属性的数据集及大规模数据集。本文根据异常数据会偏离正常数据的特性，提出利用类的异常因子解决以上不足。

针对聚类分析及异常检测方面的问题，本文提出一种基于蚁群聚类的异常网络入侵检测算法DANI。先从KDD Cup99数据集中选取合适的数据，分析这些网络数据的属性特征并对数据进行标准化处理。然后使用蚁群聚类算法进行聚类，并通过对这些类的异常因子大小的比较，检测出异常类。最后通过ROC曲线对DANI算法进行评估。

实验结果表明，DANI 算法的时间复杂度与数据集的规模为线性关系，与属性个数以及最终簇的个数成近似线性关系，这使得方法具有良好的扩展性，可用于大规模数据集中的异常挖掘，特别可望用于数据流中的异常挖掘。并且此算法具有较高检测率和较低的虚警率，整体检测性能优于 K-means 聚类算法。

关键词：蚁群聚类；K-means 算法；异常检测

Abstract

With the popularity of computers and the rapid development of Internet, security threats increase rapidly, we need to take effective measures to guarantee the security of computer systems and networks running. In recent 20 years, intrusion detection technology emerged. It is a security mechanism of dynamic monitoring, preventing and resisting the intrusion behavior. Intrusion detection technology does not affect network performance in the case of network monitoring, thereby it can provide real-time protection from internal attack, exterior attack and misoperation.

Based on in-depth analysis and comparison of the existing clustering methods, we know that the clustering analysis is suitable for large-scale data sets because of scalability and effectiveness.It not only can be used as an independent tool, but also can be used as a preprocessing step for other algorithms. Because the ant colony clustering algorithm apply to unsupervised clustering anomaly intrusion detection without prespecified number of clusters, this paper use the ant colony clustering algorithm as a preprocessing step for anomaly intrusion detection. Even though intrusion detection system does not have complete domain knowledge background, the anomaly intrusion detection algorithm also can complete intrusion detection.

This paper systematically studied the basic theory of intrusion detection, including definitions, types and models. Then it analyze the current problems. The problems is that the existing anomaly intrusion detection algorithm is difficult to detect the datasets containing mixed attributes and the large-scale datasets.According to the characteristics that abnormal data deviates from the normal data, this paper propose using outlier factor to solve the above problems.

In order to solve the problems of clustering analysis and anomaly detection, this paper presents a detection of anomaly network intrusion algorithm based on ant colony clustering，namely DANI algorithm. First，the algorithm selects the appropriate data from the KDD Cup99 dataset, analyzes the attribute characteristics of network data and standardizes data.Then it use the ant colony clustering algorithm to cluster data, compare the size of the outlier factors in other to detect the abnormaly

classes. Finally,this paper use the ROC to evaluate DANI algorithm.

The experimental results show that, the time complexity of the DANI algorithm and the scale of the dataset is linear relationship. The time complexity, the number of attributes and the number of the final cluster is approximate linear relationship, so the method has good scalability and can be used for outlier mining in large-scale dataset, especially for outlier mining in the data stream. The algorithm has a high detection rate and a low false alarm rate. The overall detection performance is better than that of K-means clustering algorithm.

Key Words: Ant Colony Clustering; K-means Algorithm; Anomaly Detection

目 录

Contents

第一章 绪论

1.1 选题背景

随着计算机和网络技术的发展，计算机网络已成为社会生活不可或缺的一部分，电子商务、电子政务、虚拟社区等建立在 Internet 网络上的电子在线服务呈快递增长的趋势，人类社会对数字信息的依赖达到前所未有的程度。信息之间的交换和交流越来越便利，由于 Internet 共享和开放的特性在给人们带来方便的同时也存在更多的安全隐患，对信息安全问题提出了严峻的挑战。由于系统存在一定的安全脆弱性，硬件设备、操作系统和应用软件这些不可避免地会存在一些安全方面的漏洞，网络协议本身的设计也存在一些安全隐患，这些都是为黑客采用非正常手段入侵系统提供了可乘之机。近些年来，个人组织所面临的安全问题越来越多，安全威胁正在迅速增加，尤其是混合威胁的风险，例如蠕虫的感染、病毒的破坏、间谍软件的使用、分布式拒绝服务攻击、垃圾邮件泛滥等，这些都对用户造成极大地困扰，严重破坏了个人或企业的网络信息安全。能否及时发现并成功拦截入侵攻击，保障计算机系统和网络系统的安全，并使它们正常运行，便成为一个重要的亟待解决的问题。

入侵（intrusion）是指任何企图危及计算机资源的完整性、机密性和可用性或试图越过计算机或网络的安全机制的行为[1]。入侵不仅包括攻击网络和计算机系统的人（例如黑客等）在合法的范围之外取得系统的控制权，对系统资源进行非授权操作，造成系统数据的丢失和破坏。入侵也包括收集计算机系统的漏洞信息，造成了计算机系统处于危险之中。入侵的发起者可能是通过互联网访问计算机系统的攻击者，也可能是计算机系统的某些授权的用户。这些用户在错误地行使系统授予他们的特权时，将导致对计算机系统入侵的后果产生。因此，入侵者可以分为两类：外部入侵者（一般指系统的非法用户，如黑客）和内部入侵者（越权使用系统资源行为的合法用户）。

1.2 研究现状

现有的各种安全防御机制都有其局限性。例如，防火墙虽然能够阻止对系统的许多非法访问，但是不能抵御某些入侵攻击，尤其是在防火墙系统存在配置上

的错误。因此，网络的安全不单单依靠单一的安全防护体系和防御机制。只有在对网络安全防护系统和各种有关网络安全的技术和工具等方面的研究基础上，制定具体的计算机系统的安全策略，设立多道的安全防线，集成各类可靠地安全机制（例如，网络防火墙、存取控制机制和身份验证体制、安全监控工具、计算机系统漏洞的扫描工具和入侵检测系统），进行有效的安全管理即培训等，建立完善的多层安全防护体系，这样才能够有效地抵御来自计算机系统内、外的入侵攻击，确保计算机系统及网络的安全。

为了确保计算机网络的安全，必须建立一整套的安全防护体系，进行多层次，多手段的检测和防护。入侵检测技术是继防火墙、数据加密等传统保护措施后新一代的网络安全保障技术，是一种主动的安全防护技术。入侵检测系统被认为是防火墙之后第二道安全闸门，是对防火墙的有益补充。它在不影响网络性能的情况下能对网络进行监听，从而提供对内部攻击、外部攻击和误操作的实时防护，大大提高了网络的安全性。