大数据技术在计算机网络入侵检测中的应用

黄鸿运

海南师范大学　海南海口　571127

摘　要：随着科技水平的不断提高，计算机与信息技术也在不断地发展，大数据技术应运而生。大数据技术是一种可以实现各种信息数据的采集、分析、整理和存储的一种技术，可以对各种类型的信息进行管理，将大数据技术运用到计算机网络入侵检测中，可以提升网络的准确性和精准度，并可以实现网络入侵检测的自动化和智能化，这对确保计算机网络的安全稳定运行有着重大的实际价值。

关键词：大数据技术；计算机网络入侵检测；应用

网络入侵检测是目前国际上最热门的研究课题之一。是一种先进的动态网络安全控制技术。但是，作为一种新兴的网络安全隐患检测技术，缺少可供参考的经验与启发，目前还存在很多问题需要解决。将大数据挖掘技术运用于计算机网络入侵检测中。可以迅速地确定入侵的地点和时间，并提出相应的安全防御策略，保障计算机网络的安全。因此，研究大数据技术在计算机网络入侵检测中的应用显得尤为重要。

1. 大数据技术概述

在今天的信息时代，大数据已经成为一个不可避免的需求，已成为促进我国经济发展的关键技术。大数据是一种可以对大量的信息进行分析、整理和存储的数据，其实质是信息数据的信息技术。同时，通过对大数据的信息挖掘，可以在理念上、模式上、技术上、应用上有所创新，使大数据技术不断地进行优化和创新，从而更好地满足当前信息化时代的需求，为当今社会的发展作出贡献。大数据本质上就是一个用来储存各种行业、各类数据信息的数据库。在相关人士从数据库收集信息时，可利用合理的技术手段搜集数据，从而改善数据搜集的时效性。大数据是以互联网、物联网、企业数据为基础，构建数据源，经过抽取、转换、装载等方式，实现对数据的采集与存储。在此基础上，实现了对各种数据的自动化管理，如果使用者希望搜集有关数据，可以从该数据库获取相应的授权，使数据成为可视化。

2. 网络入侵检测概述

2.1. 网络入侵检测概念

为了确保网络的安全性，网络入侵检测是一种有效的手段。网络入侵检测就是对网络的运行情况进行检查，基于计算机用户使用计算机的行为，判定该行为有没有可能对网络构成入侵的危险，如果使用者行为可能给互联网带来了入侵的危险，那么，网络入侵检测就可以对其进行拦截，并将其报告给网络使用者，因此可以最大限度地保障网络的安全，使网络可以顺利运作。网络入侵检测作为基于网络的内部体系，是一项非常有意义的网络安全管理技术，可以采集来自于各种系统源的信息，并且可以根据这些数据来对计算机网络工作状态进行分析，从而可以识别计算机网络有没有被入侵的风险，有没有被其他的黑客攻击。另外，通过网络入侵检测技术，可以对整个网络的运行情况进行全面地监测。在网络正常工作的时候，就会进行网络入侵检测，直至整个网络停止运行为止。该技术通过对网络运行状态的自动采集，建立相应的日志，并将其上载至系统。相关人员在对网络进行管理的同时，同时，还可以查询到有关的网络检测记录，从而对计算机网络的稳定程度作出评价。该系统还能在发现有入侵行为的情况下，自动产生相应的反应报告，并向有关人员汇报。对一些具有威胁性的网络入侵检测，可以使用防火墙自动侦测地域。当攻击者遇到难以抵御的攻击时，需要在收到入侵信息后，迅速地作出相应的应对措施，从而保证计算机网络的安全稳定运行。

2.2. 入侵检测技术的分类

根据入侵检测技术的不同，其分类可以划分为：异常监测和误用行为识别；按照检测目标的不同，可以将其划分为基于主机、基于网络和基于主机混合型。入侵检测主要由以下三部分组成：一是信息采集，采集设备运行状态、用户

行为等信息，并采集网络协议和网络流量等信息。二是信息分析，包含信息统计分析，当发现异常情况时，系统会立即报警，并向控制台发送故障日志。三是控制台依据所收到的非正常信息的种类，选择合适的处理方式。

3. 当前入侵检测系统的问题

入侵检测是一项有效地保障网络安全的技术，包含异常检测，特征检测，协议分析，状态检测等。实际应用中，入侵检测系统通常采用几种不同的方法组合在一起，但尚有很多方面有待进一步研究和完善。

3.1. 错误率较高

虚报和漏报是当前入侵检测系统（IDS）中最为突出的一项性能指标。据统计，每年有 3000—10000 个攻击者使用入侵检测系统漏洞，而现有入侵检测系统漏洞检测率只有 50%。很多的攻击都是针对这些弱点进行入侵的。因此，如何有效地提高入侵探测系统的检测率和降低错误识别率，是当前入侵探测系统研究中亟待解决的关键问题。传统的网络入侵检测方法通常是由多个入侵检测器同时检测一个目标计算机，而多个入侵检测器则只能够检测到一个目标计算机。目前常用的方法都是对多个目标主机进行扫描，但是存在着扫描耗时过久、不能涵盖全部可能的漏洞等问题。目前入侵检测系统的虚报率很高，其根本原因在于 IDS 的检测准确度不够高，且已有的检测手段均有缺陷。当前，对于大规模的混合分布式攻击，尚未有较好的解决方案。比如，常用的统计方法探测网络中的异常，但是其门限难以精确地判断出错误，门限值太小会引起很多误差，门限值太高也会引起很多差错。

3.2. 缺乏主动地保护

入侵检测技术是一种被动的、有局限性的技术，不能有效地进行入侵检测。近年来，越来越多的新技术如蠕虫，木马，黑客软件等对计算机网络的安全构成了极大的威胁。在已有的安全措施不能保证网络的安全时，利用 IDS 保证网络的安全是一种可行的方法。在 IDS 中，已有的检测规则可以通过预先定义的方式或特征描述等方式进行更新，而这类规则通常是滞后的。当新的漏洞被发现时，就会有一些方法和手段可以立刻对其进行攻击。但是，要花很长的时间才能发现合理的检测与防御规则，以处理这个漏洞。事实上，新的黑客技术更新也需要一定的时间，在这期间，足以使黑客做出攻击。

3.3. 缺乏准确的定位与加工机制

IDS 仅能够对 IP 进行识别，而无法对 IP 进行确定，从而无法判断出具体的数据来源。

比如，防火墙的作用就是扫描网络安全，发现不安全的因素，过滤不安全的流量，保证网络的正常运转。入侵检测系统的主要作用就是对网络中的数据进行实时监测，发现异常状态，及时报警。防火墙与 IDS 在本质上是一个整体，目的都是防止黑客入侵，保护网络的安全。而二者存在着显著的差异：防火墙的作用是对整个网络进行全面的管理；而 IDS 则是通过对网络的运行状态进行监测，从而发现可能存在的安全缺陷，从而对其进行有效的防护。当前，大部分的防火墙、IDS 都是各自独立工作的。这两种方法都有各自的缺陷，制约着其实用化。在检测到攻击的时候，只有几个端口或者是输出是可以被关闭的，这会对其他普通用户造成很大的影响，而且没有有效的反应机制。

4. 数据挖掘技术在网络入侵检测中的应用

数据挖掘技术是大数据技术的重要组成之一，结合已有的研究结果和工作经验，该技术在实际的网络入侵中取得了很好的效果。其具体如下：

4.1. 系统模型和检测方法

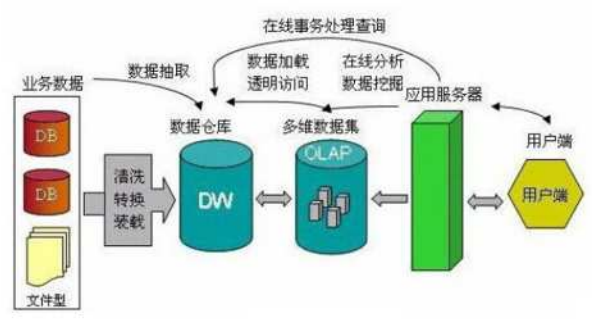
将数据挖掘技术运用到网络入侵检测中，可以构成基于大数据和云计算等技术的分布式网络入侵检测系统。该系统使用移动的 Agent 采集完整的探测内容，并将其传输到事件序列发生器中，通过数据挖掘技术实现对行为的有效识别。该方法通过对已发现的信息或有关规则之间的相似度进行分析，使决策者能够作出最后的决定，从而保证计算机网络系统的安全。

当前，网络入侵检测技术主要有两种。其中一种是针对不同操作系统特性的基于主机入侵检测系统，能够迅速地发现应用层的攻击。然而，该方法要求主机与审计系统协同工作，不具有实时性；另外一种则是以网络为基础的入侵检测，该方法利用网络中所搜集到的数据，对入侵检测可疑行为进行分析，其不依赖于主机，并且可以按照标准的网络协议运行。通过智能化网络入侵检测系统，能够将二者有机地融合在一起，从而更好地适应各种环境下的入侵检测要求。当检测环境发生改变时，仅需对相应的数据做相应的修改，而无需对硬件，软件，协议做全部的修改。在网络环境下，智能入侵检测系统能够迅速地识别出新的技术和应用，并能

够通过自主学习功能对规则库进行扩充，从而实现对网络的自适应。

4.2. 系统架构设计

当前，国际上普遍采用两种网络入侵系统构架，一种是构建统一的中心平台，实现对网络入侵的监控。然而，这个架构仅仅适合于小型网络管理，在大规模网络管理中，会出现检测到的信息不准确的问题。第二种网络架构是以子网络为核心的中心架构。每个区域都有入侵检测专业人员，每个系统都可以看作是一个独立的系统，该系统结构可以对各个子网进行全面、精确地探测，更符合网络入侵检测系统结构的要求。在进行网络 IDS 架构设计时，必须对经过预处理的数据源进行采集，并将其导入数据仓库。然后，通过数据挖掘技术，构建数据挖掘引擎，然后将这些数据信息，分别传送到检测模块和规则库两个部分之中，同时，将规则库所进行的分析、处理后的数据传送给检测模块。然后该检测模块对该数据进行检测，可以判定是否存在网络入侵。在此基础上，当被检测存在网络入侵行为时，必须将有关信息传送至对应的入侵模块，以达到检测到的目的。如图 1，数据挖掘技术。



如图 1，数据挖掘技术

4.3. 数据挖掘算法在网络入侵检测中的应用

关联规则是指在一系列数据集合中，展现一系列物件间的关联性与规则，数据挖掘算法发现网络漏洞时程序运行与用户行为存在关联性，而这些关联往往体现在关联数据集中。利用关联分析方法，能够迅速地获取多个数据要素间的相关度，进而确定各要素间的联系，同时，利用序列分析技术对多个事务进行关联描述，能够从交易中抽取出序列模式，并且能够满足用户对频繁序列的最小需求。

在审计过程中，通常无法事先决定使用者所需的数据，但是下列两种方式可用于对使用者所关心的规则进行分析。一是先获取合适的规则，再按用户的兴趣程度对每条规则进行排序，去除不需要的规则；二是利用已有的知识，将待处理的规则作为条件约束，然后再进行挖掘。将关联规则用于网络入侵检测的基本机制为：利用相关规则算法，可以迅速地识别出各种未知的入侵模式，为实现对网络算法监测与预防，提高网络的安全与稳定。与其他网络入侵检测方法比较，使用关联规则的算法可以迅速、高效地对一些不知名的网络入侵模式进行挖掘，并对当前计算机使用者的多种信息和行为模式进行展示，对当前的行为模式与历史数据和行为进行检验，若两者相差过大，可以判断为入侵。

5. 结束语

综上所述，利用大数据挖掘技术对计算机网络进行入侵检测，既能有效地增强其智能程度，又能提高其工作效率与质量。另外，随着我国经济和社会的发展，对网络安全的要求也越来越高。所以，要想有效地提升计算机网络入侵系统运作的效率与品质，需要有关部门加大对大数据技术的研究，并对其进行持续地完善与优化。另外，在实施计算机网络入侵检测时，相关人员也要对大数据技术的运用有充分地认识与分析，才能保证其成功实施。

参考文献

[1] 王震 . 计算机网络安全的入侵检测技术分析 [J]. 中国信息化 ,2021,(12):61–62.

[2] 王刚 . 基于数据挖掘技术的网络入侵检测系统 [J]. 电子设计工程 ,2021,29(13):15–19.

[3] 鹿鸣 . 探析计算机网络安全的入侵检测技术 [J]. 电子测试 ,2021,(10):54–55.

[4] 赵菲 . 网络入侵检测中数据挖掘技术的应用研究 [J]. 科技创新与生产力 ,2020,(12):58–60.

[5] 吕光铭 . 计算机网络入侵检测技术研究 [J]. 科学技术创新 ,2020,(34):79–80.

[6] 朱媛媛 . 基于机器学习与大数据技术的入侵检测方法研究 [D]. 太原理工大学 ,2020.