大数据技术在计算机信息安全中的应用研究

张 坤[1] 邓 郁[2]
1 黑龙江出入境边防检查总站哈尔滨警犬训练基地，黑龙江 哈尔滨 150000
2 国家移民管理局常备力量第二总队，云南 昆明 650214

[摘要]随着数字化时代的快速发展，计算机信息安全面临着前所未有的挑战。大规模的数据交互和存储带来了巨大的安全隐患，传统的安全防护手段已无法满足对抗复杂威胁的需求。因此，大数据技术的兴起为解决这一难题提供了全新的可能性。大数据技术的高速处理能力和强大的分析能力，使得安全专家可以通过对海量数据的深度挖掘和分析，实现对网络入侵、威胁检测和恶意行为的精确预测和防范。文章将探讨大数据技术在计算机信息安全领域的应用研究，并阐述其在提升网络安全性和降低风险方面的潜力。

[关键词]大数据技术；计算机；信息安全
DOI：10.33142/sca.v6i3.8824 中图分类号：TP3 文献标识码：A

Research on the Application of Big Data Technology in Computer Information Security

ZHANG Kun[1], DENG Yu[2]
1 Harbin Police Dog Training Base of HeiLongJiang General Station of Exit and Entry Frontier, Harbin, Heilongjiang, 150000, China
2 The Second Brigade of the Standing Force of the National Immigration Administration, Kunming, Yunnan, 650214, China

Abstract: With the rapid development of the digital era, computer information security is facing unprecedented challenges. Large scale data exchange and storage have brought huge security risks, and traditional security protection methods can no longer meet the needs of combating complex threats. Therefore, the rise of big data technology provides a new possibility to solve this problem. The high-speed processing and powerful analysis capabilities of big data technology enable security experts to accurately predict and prevent network intrusion, threat detection, and malicious behavior through deep mining and analysis of massive data. The article will explore the application research of big data technology in the field of computer information security, and elaborate on its potential in improving network security and reducing risks.

Keywords: big data technology; computer; information safety

大数据技术是应对海量、复杂、异构和动态变化的数据处理需求的新型技术和方法。它通过处理、存储和分析大规模数据，提取有价值的信息，支持决策和创新。大数据技术具有处理大规模数据、处理不同类型数据和实时处理速度的特点。它在计算机信息安全领域的应用研究中发挥着重要作用，包括实时威胁检测、安全日志分析、恶意代码检测和跨平台协同等措施。通过应用大数据技术，可以提高安全防御的准确性、效率和响应速度，提升计算机信息安全的水平。

1 大数据基本概念及特点

大数据技术是指在处理、存储和分析大规模、复杂、异构和动态变化的数据时，应用的一系列新型的数据处理技术和方法。其核心在于从海量数据中提取出有价值的信息，并加以挖掘和分析，以便支持决策和创新。其主要特点包括三个方面：首先，大数据技术处理的数据规模巨大，一般以TB或PB为单位。随着互联网和物联网的迅速发展，数据的规模呈指数级增长，传统的数据处理技术已无法胜任如此庞大的数据规模。其次，大数据技术处理的数据类型复杂，包括结构化数据、半结构化数据和非结构化数据。结构化数据是按照固定的模式和格式组织的数据，如关系型数据库数据；半结构化数据是部分具有结构化特征的数据，如XML文档、JSON等；非结构化数据则是没有明显结构的数据，如文本、图片和视频等。处理这些不同类型的数据需要灵活的数据处理技术和方法。最后，大数据技术的数据处理速度要求高，需要实现实时或近实时的数据处理和分析。传统的批处理方式已不能满足如此高速的数据处理要求，需要采用实时流式处理技术，以便实现对数据的快速响应和决策。

2 大数据技术在计算机信息安全中运用技术

2.1 云计算技术

云计算技术在大数据安全中的应用是为计算机信息安全提供高效的存储和处理能力。云计算平台通过虚拟化技术和分布式计算资源，可以提供弹性的计算环境，满足大规模数据处理的需求，同时提高计算效率和可靠性。首先，云计算技术提供了大规模的存储空间，使得安全专家能够将海量的安全数据进行集中存储和管理。这些数据可以包括网络流量日志、入侵检测系统产生的事件日志、恶意软件样本等。通过将这些数据存储在云平台上，安全专

家可以方便地进行数据的备份、恢复和共享，提高数据的可靠性和可用性。其次，云计算平台提供的弹性计算资源可以满足大规模数据处理的需求。大数据安全分析通常需要进行复杂的计算和算法运算，需要消耗大量的计算资源。云计算平台可以根据需求自动调配计算资源，弹性地扩展和收缩计算能力，从而提高计算效率和响应速度。这对于实时的威胁检测和应对是至关重要的。此外，云计算技术还可以提供分布式计算环境，实现并行处理和分布式存储。大数据安全分析通常需要对海量的数据进行深度挖掘和分析，而分布式计算和存储可以将数据分散存储在多个节点上，并利用并行计算的能力快速处理数据。这种分布式的方式不仅提高了计算速度，也增强了系统的容错性和稳定性[1]。

2.2 数据挖掘技术

数据挖掘技术在大数据安全中的应用是通过对海量数据进行深度分析和挖掘，以发现潜在的威胁和异常行为。数据挖掘技术结合了统计学、机器学习和人工智能等领域的方法和算法，可以从大数据中提取有价值的信息和模式，帮助安全专家进行威胁检测、行为分析和风险评估。首先，数据挖掘技术可以用于威胁检测。通过分析网络流量数据、入侵检测系统产生的事件日志等，数据挖掘技术可以识别出异常的网络活动、恶意代码和未知的攻击模式。基于机器学习算法的分类和异常检测模型可以对正常和恶意行为进行区分，并发现新的威胁和攻击手段。其次，数据挖掘技术可以用于行为分析。大规模的安全日志和事件数据包含了丰富的信息，数据挖掘技术可以从中挖掘出用户行为模式和规律。通过分析用户的登录模式、访问模式和操作行为，可以识别出异常的用户活动和可疑行为，及时发现内部威胁和数据泄露的风险。此外，数据挖掘技术还可以用于风险评估和预测。通过对历史安全数据的挖掘和分析，可以建立风险评估模型和预测模型，预测未来可能发生的安全事件和攻击。这可以帮助安全专家作出相应的防御和应对策略，提前做好安全准备，降低风险发生的概率和影响[2]。

2.3 可视化技术

可视化技术在大数据安全中的应用是将复杂的安全数据转化为直观的图形展示，以帮助安全专家进行数据分析、决策和沟通。通过可视化技术，安全专家可以更好地理解和解释数据，从而快速洞察安全事件和趋势。首先，可视化技术可以帮助安全专家进行数据分析和探索。大数据安全分析通常涉及大量的数据维度和指标，难以通过传统的数据表格和报告来全面理解。可视化技术可以将这些数据以图形、图表或热力图的形式展示出来，使得安全专家可以更直观地观察和分析数据之间的关系和趋势。通过交互式的可视化界面，安全专家可以自由地进行数据探索和发现，发现隐藏在数据中的有价值的信息。其次，可视化技术可以帮助安全专家进行决策和预测。通过将安全数据可视化，安全专家可以更清楚地了解当前的安全状况和风险。例如，通过将网络攻击的来源和目标地理位置可视化在地图上，安全专家可以迅速发现攻击的热点区域和目标集中区域，从而调整安全策略和资源分配。此外，通过可视化安全数据的历史趋势和模式，安全专家可以预测未来可能的安全威胁和趋势，提前采取相应的防范措施。此外，可视化技术还可以帮助安全专家进行数据沟通和共享。通过将安全数据以可视化形式展示，安全专家可以更清晰地向管理层或其他团队成员传达安全状况和风险情况。可视化技术提供了直观、易懂的方式，使非技术人员也能理解和参与安全决策，从而提高整个组织对安全问题的认识和响应能力。

3 大数据技术在计算机信息安全中的具体应用措施

3.1 实时威胁检测

实时威胁检测是大数据技术在计算机信息安全中的重要应用措施。随着网络威胁的不断演变和增长，传统的静态防御手段已经无法满足对抗复杂威胁的需求。实时威胁检测利用大数据分析技术，可以快速、准确地分析大量的网络流量数据、入侵检测系统产生的事件日志等，实现对实时威胁和异常行为的监测和识别。首先，实时威胁检测依靠大数据分析技术处理海量的网络流量数据。通过大数据平台的高速处理能力和分布式计算环境，可以对大规模的网络流量数据进行深度挖掘和分析。这包括对网络会话、数据包和通信行为进行实时监测和分析，以发现潜在的恶意活动、异常行为和攻击模式。其次，实时威胁检测借助机器学习和数据挖掘技术来构建模型并进行实时判断。通过对大量的已知恶意行为和正常行为进行学习，可以训练机器学习模型和算法，实现对实时流量数据的分类和识别。这样，安全专家可以快速得到实时威胁的警报和报告，及时采取相应的防御和应对措施。最后，实时威胁检测结合了实时事件响应和自动化处理的能力。当发现异常活动或威胁时，系统可以自动触发预定的响应机制，如警报通知、封锁网络访问或启动紧急补丁更新。这大大缩短了威胁检测到响应的时间，有助于及时阻止攻击并减轻潜在的损害[3]。

3.2 安全日志分析

安全日志包含了各种安全事件、操作记录和系统状态信息，通过对大量安全日志进行分析，可以发现潜在的安全威胁和异常行为，从而帮助安全专家进行决策和应对。首先，安全日志分析利用大数据技术处理海量的安全日志数据。随着计算机信息系统规模的扩大和网络活动的增加，产生的安全日志数据呈指数级增长。传统的手动分析方法已无法应对如此庞大的数据量。而大数据技术的高速处理能力和分布式存储系统，可以高效地处理和存储大规模的安全日志数据，为后续的分析和挖掘提供强有力的支持。其次，安全日志分析利用数据挖掘和机器学习技术对安全

日志进行深度分析。通过应用数据挖掘算法，可以从海量的安全日志中发现隐藏的安全威胁、异常行为和攻击模式。例如，通过关联分析技术可以发现不同事件之间的关联性，识别攻击链的路径；通过异常检测算法可以检测到不符合正常行为模式的活动，及时发现潜在的入侵行为。再次，安全日志分析借助可视化技术将复杂的安全日志数据转化为直观的图表、图形或仪表盘展示，以便于安全专家进行数据理解和决策。通过可视化界面，安全专家可以直观地观察和分析安全事件、行为趋势和异常模式，及时发现和定位安全问题，并采取相应的应对措施。最后，安全日志分析通过持续监测和分析安全日志数据，可以建立和更新安全模型和规则。基于历史安全日志数据的分析，可以建立起不同类型攻击的特征模型，实现对未知威胁的预警和检测。安全模型的不断优化和迭代，可以提高安全日志分析的准确性和实时性，为安全专家提供更精确的安全决策和响应。

3.3 恶意代码检测

随着恶意代码的不断演进和增长，传统的防病毒软件和签名检测已无法有效应对新型的恶意代码攻击。恶意代码检测利用大数据技术中的机器学习和数据挖掘技术，可以对海量的恶意代码样本进行分类和识别，实现精准的恶意代码检测和分析。首先，恶意代码检测利用大数据技术处理大量的恶意代码样本。恶意代码样本呈指数级增长，传统的人工分析和特征提取方法已无法满足检测的速度和规模。大数据技术的高速处理能力和分布式存储系统可以快速处理和存储大规模的恶意代码样本，为后续的分析和模型训练提供强大的支持。其次，恶意代码检测利用机器学习和数据挖掘技术构建恶意代码分类模型。通过对已知的恶意代码和正常代码样本进行学习，可以训练机器学习模型和算法，实现对新的恶意代码样本的自动分类和识别。常用的机器学习算法如支持向量机、决策树和深度学习等可以从恶意代码中学习到特征模式，进而判断未知代码的恶意程度。再次，恶意代码检测结合了静态和动态分析技术。静态分析通过对恶意代码的静态特征进行分析，如代码结构、API 调用和字符串等，可以检测出一些已知的恶意代码。动态分析通过在虚拟环境中运行恶意代码，监测其行为和系统交互，可以发现一些新型的恶意行为和变种。结合静态和动态分析，可以提高恶意代码检测的准确性和全面性。最后，恶意代码检测通过持续更新和优化恶意代码库和检测规则，不断适应新的恶意代码威胁。随着恶意代码的不断变化和演进，及时更新和升级恶意代码库，以及通过大数据技术分析和挖掘新的特征和行为模式，可以提高恶意代码检测的准确性和实时性，为安全专家提供更可靠的防御手段[4]。

3.4 跨平台协同

在多平台和分布式环境中，安全威胁往往不局限于单一系统或网络，而是跨越多个平台和设备。跨平台协同利用大数据技术的分析和集成能力，实现不同安全设备和系统之间的信息共享、协作和集成，提高安全威胁的检测精度和响应速度。首先，跨平台协同实现了安全数据的整合和集中存储。不同的安全设备和系统产生的数据包括网络流量数据、入侵检测系统的事件日志、防火墙日志等，数据分布在不同的平台和设备上。通过大数据技术，可以将这些分散的安全数据进行整合和集中存储，建立统一的安全数据湖或数据仓库。这样，安全专家可以方便地访问和分析这些数据，实现全面的安全威胁监测和分析。其次，跨平台协同实现了安全事件的跨系统关联和分析。不同安全设备和系统产生的事件数据往往存在关联性，需要进行跨系统的关联和分析。通过大数据技术中的数据挖掘和关联分析技术，可以将来自不同系统的事件数据进行关联和匹配，从而发现攻击链的路径和行为规律。这样，安全专家可以全面了解攻击事件的全貌，采取相应的防御和应对措施。最后，跨平台协同实现了安全策略和事件响应的协调和统一。不同平台和系统可能有各自的安全策略和事件响应机制，导致信息孤岛和响应不协调的问题。通过跨平台协同，可以实现安全策略的统一制定和管理，确保一致性和协调性。同时，跨平台协同也可以实现安全事件响应的协同和自动化，即当一个平台发现异常或攻击事件时，可以自动触发其他平台的响应措施，实现联动防御和迅速响应。

4 结束语

大数据技术在计算机信息安全领域的应用研究为我们提供了强大的工具和方法，用于解决日益复杂和持续增长的安全威胁。通过实时威胁检测、安全日志分析、恶意代码检测和跨平台协同等措施，大数据技术能够提高安全防御的准确性、效率和响应速度。随着大数据技术的不断发展和创新，我们将能够更好地保护计算机信息系统免受威胁，并提升整体的网络安全水平。

[参考文献]

[1]墙浩煊.大数据技术在计算机信息安全中的应用研究[J].自动化应用,2022(12):97-100.
[2]张成.大数据技术在计算机信息安全中的应用分析[J].黄河科技学院学报,2022,24(11):49-54.
[3]李春毅.计算机信息安全中大数据技术的应用研究[J].电脑知识与技术,2022,18(14):19-21.
[4]薛俊海,李晋泰,张承,等.大数据技术在计算机信息安全中的应用研究[J].网络安全技术与应用,2022(2):70-71.

作者简介：张坤（1983.4-）女，汉族，本科学历，黑龙江出入境边防检查总站哈尔滨警犬训练基地，二级主任科员，长期从事信息通信网络管理及运行支撑工作；邓郁（1984.8-）女，广东中山人，汉族，本科学历，国家移民管理局常备力量第二总队，警务技术二级主管（副高），长期从事通信技术教学研究工作。