大数据在医院网络安全威胁识别中的作用

段鸿飞

新疆喀什地区泽普县人民医院，新疆 喀什 844800

摘要：随着医院网络环境的复杂性和网络安全威胁的日益增长，大数据技术在威胁识别中展现出显著优势。本文综述了大数根据医院网络安全威胁识别的作用，分析了其理论基础、关键技术及应用实践。医院面临勒索软件、数据泄露等威胁，传统安全机制难以应对，而大数据通过海量数据采集、实时分析和机器学习技术，显著提升了威胁检测的效率与精度。文章探讨了数据处理、异常检测和自动化响应等机制在医院中的应用，并通过案例分析验证了其实效性。同时，研究指出了隐私保护、资源限制等挑战，并展望了低成本解决方案、隐私保护技术和跨学科研究的未来方向。本文为医院网络安全领域的大数据应用提供了系统参考，为技术部署和政策制定提供了指导。

关键词：大数据，医院网络安全，威胁识别，隐私保护

一、引言

（一）医院网络安全威胁的现状与挑战

近年来，医院作为医疗服务和敏感数据管理的核心机构，面临日益严峻的网络安全威胁。医疗机构高度依赖电子健康记录（EHR）、互联网医疗设备（IoMT）以及远程医疗平台，导致其成为网络攻击的主要目标。勒索软件、数据泄露、钓鱼攻击和恶意软件等威胁显著增加，尤其在 COVID-19 疫情期间，攻击者利用远程工作和未加密的远程医疗平台漏洞，针对医院发起大规模攻击。例如，2020 年，美国超过 400 家医院被列入网络犯罪分子的攻击目标清单，涉及勒索软件和数据窃取。此外，医院内部管理和技术上的不足进一步加剧了威胁。许多医院缺乏完善的业务连续性计划，高级管理人员对网络威胁的认识不足，导致资源分配和员工培训严重滞后[1]。医疗设备的安全性问题尤为突出，互联设备通常缺乏内置防火墙或防病毒保护，易被攻击者利用[2]。这些挑战不仅威胁患者数据隐私，还可能导致医疗服务中断，甚至危及患者安全。例如，英国 NHS 的 WannaCry 攻击导致医院日均住院率下降 6%，急诊和择期手术受阻[3]。医院网络安全威胁的复杂性和多样性要求医疗机构采取全面的防御策略，以应对技术、政策和人员管理方面的多重挑战。

医院网络安全的现状还受到行业特性的制约。医疗行业相较于金融等其他领域，在网络安全投入上长期不足，资源分配偏向患者护理而非 IT 安全[4]。此外，医院系统的复杂性和异构性增加了防护难度，传统安全机制难以应对新型攻击方式，如利用 5G 网络的物理层攻击[5]。低收入和中等收入国家（LMICs）的医院尤为脆弱，因基础设施落后和缺乏专业人员，难以实施有效的网络安全措施。这些现状表明，医院需在技术升级、人员培训和政策制定上投入更多资源，以应对不断演变的网络威胁。

（二）大数据技术在网络安全领域的兴起

大数据技术近年来在网络安全领域迅速兴起，为威胁检测和响应提供了新的解决方案。医疗机构生成的海量数据，包括系统日志、患者记录和设备交互数据，为大数据分析提供了丰富的素材。通过特征工程和机器学习模型，大数据技术能够实时监控网络活动，识别异常行为并预测潜在威胁。例如，基于大数据的风险预测模型通过分析多源数据（如系统日志和外部威胁情报），可在企业信息安全管理中实

现高达 0.95 的 AUC（曲线下面积），显著提升威胁检测的准确性[6]。在医院场景中，大数据分析可用于检测勒索软件攻击的早期迹象，通过关联分析发现隐藏的恶意活动模式[7]。

大数据技术在网络安全领域的应用还推动了跨行业的最佳实践借鉴。医疗行业可借鉴企业信息安全管理中的大数据分析框架，通过整合历史数据和实时数据，构建自适应的安全模型。例如，基于图神经网络的大数据分析能够揭示网络攻击的传播路径，为医院提供更精准的防御策略。

二、大数据技术在网络安全威胁识别中的理论基础

（一）大数据的基本概念与技术框架

大数据技术以处理海量、异构和高动态数据为核心，为网络安全威胁识别提供了强大的技术支撑。大数据的核心特性包括体量大（Volume）、速度快（Velocity）、多样性（Variety）、真实性（Veracity）和价值（Value），这些特性使之能够应对医院网络环境中复杂的日志数据、患者记录和设备交互信息[8]， 大数据技术框架通常包括数据采集、存储、处理和分析四个关键环节。数据采集依赖于传感器、日志系统和外部威胁情报的整合；存储则采用分布式系统如 Hadoop HDFS 或云存储以确保高效扩展；处理环节利用 MapReduce 或 Apache Spark 进行并行计算；分析阶段则结合机器学习和深度学习算法提取潜在威胁模式[9]， 在医院场景中，大数据框架能够处理来自电子健康记录（EHR）和互联医疗设备（IoMT）的多源异构数据，生成实时威胁情报。例如，基于 Spark 的流处理技术可在毫秒级内分析网络流量，识别异常行为，此外，大数据技术的开源生态系统，如 Apache Kafka 和 Elasticsearch，降低了医院部署复杂分析系统的成本，同时提升了系统的可扩展性。大数据技术的应用还需考虑医疗行业的特殊需求[10]。

（二）网络安全威胁识别的关键技术

网络安全威胁识别依赖于一系列关键技术，以应对日益复杂的攻击形式，如高级持续威胁（APT）、零日攻击和内部威胁。异常检测是威胁识别的核心技术之一，通过分析网络流量、用户行为和系统日志的偏差，识别潜在的恶意活动。基于机器学习的异常检测模型，如孤立森林和自编码器，能够处理高维数据并发现隐藏的攻击模式， 其次，入侵检测系统（IDS）通过签名匹配和行为分析技术，实时监控网络活动。现代 IDS 结合深度学习算法，能够在医院网络中检测到复杂的多阶段攻击，例如利用医疗设备漏洞的横向移动攻击， 此外，威胁情报分析技术通过整合外部数据源（如暗网数据和行业报告）与内部日志，构建攻击者的行为画像，从而预测潜在威胁，在医院环境中，这些技术需要处理高并发的数据流，同时保持低误报率以避免干扰医疗服务。

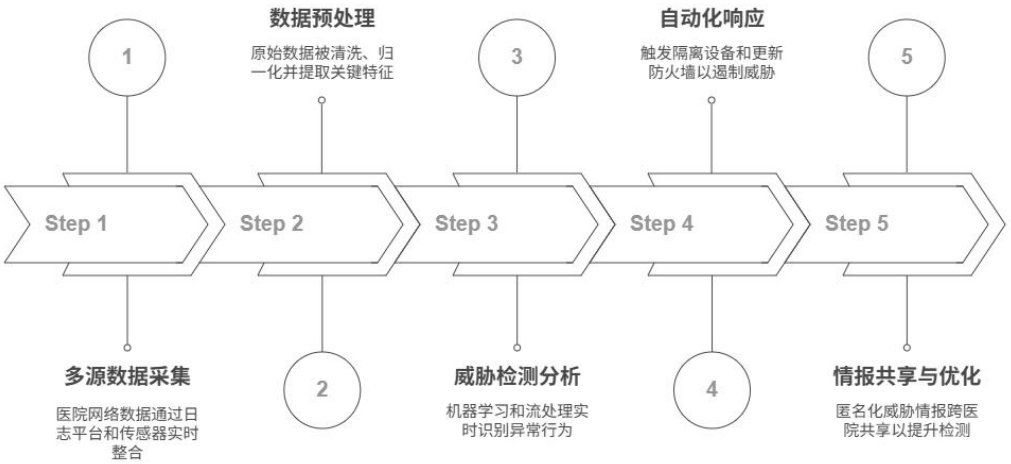
（三）大数据与网络安全威胁识别的结合机制

大数据技术与网络安全威胁识别的结合通过数据驱动的分析和自动化响应机制，显著提升了医院网络安全的防护能力（图 1）。核心结合机制之一是多源数据融合，大数据技术通过整合网络日志、设备数据和外部威胁情报，构建全面的威胁视图。例如，基于图计算的大数据分析能够揭示攻击者在医院网络中的传播路径，从而实现对复杂攻击的早期预警， 其次，机器学习和深度学习算法在大规模数据处理中发挥关键作用。卷积神经网络（CNN）和长短期记忆网络（LSTM）可用于分析时间序列数据，检测勒索软件或数据渗漏的微弱信号，此外，大数据支持的自动化响应机制能够在威胁检测后迅速采取行动，如隔离受感染设备或更新防火墙规则，从而减少攻击的破坏性。

大数据与威胁识别的结合还体现在自适应安全模型上。传统安全系统依赖静态规则，难以应对新型攻击，而大数据驱动的动态模型能够通过持续学习适应新的威胁模式， 例如，基于联邦学习的大数据系统可以在不共享患

者数据的情况下，跨医院协作训练威胁检测模型，从而提升整体防御能力，然而，这种结合机制需解决数据隐私、计算复杂度和模型可解释性的挑战。医院需在大数据分析与法规合规之间找到平衡，同时优化计算资源以支持实时威胁检测。

图 1 大数据与网络安全威胁识别的结合机制



三、大数据在医院网络安全威胁识别中的应用

（一）数据采集与处理在医院网络安全中的作用

数据采集与处理是大数据库威胁识别的基础，为医院网络安全提供全面的数据支持。医院网络环境生成多种数据，包括电子健康记录（EHR）、医疗设备日志、网络流量和员工操作记录。这些数据的采集需要高效的系统，如日志管理平台和物联网传感器，以确保数据的完整性和实时性。数据处理环节则通过清洗、归一化和去噪，将异构数据整合为可分析的格式。例如，医院可利用分布式存储系统处理海量日志数据，确保高并发环境下的数据可用性。处理后的数据为后续威胁识别提供可靠输入，例如通过分析设备通信模式发现异常流量。这些环节还支持跨部门数据整合，打破数据孤岛，提升医院整体安全态势感知能力。

（二）大数据分析技术在威胁检测中的应用

大数据分析技术通过处理海量数据，显著提升医院网络威胁检测的效率与精度。机器学习算法，如聚类和分类模型，可分析网络流量和用户行为，识别异常模式，例如未授权访问或数据泄露的早期迹象。流处理技术能够在毫秒级内分析实时数据，适合检测快速传播的恶意软件。关联分析技术则通过挖掘多源数据间的关系，揭示复杂的多阶段攻击路径，例如利用医疗设备漏洞的横向移动。此外，大数据分析支持预测性建模，通过历史数据预测潜在威胁，协助医院提前部署防御措施。这些技术能够自动化处理高维数据，降低人工干预需求，同时保持低误报率，确保不干扰医疗服务。

（三）医院网络安全中的大数据案例分析

大数据在医院网络安全中的应用已在多个实际案例中展现成效。例如，某大型医院通过部署基于大数据的入侵检测系统，成功识别并阻止了一次针对 EHR 系统的钓鱼攻击，系统通过分析员工邮件行为模式发现异常登录尝试。另一案例中，医院利用流处理技术监控互联医疗设备，实时检测到一台设备异常高频发送数据，及时隔离避免了数据泄露。此外，某地区医院联盟通过共享匿名化威胁数据，构建区域性威胁情报平台，显著提升了对新型勒索软件的响应速度。这些案例表明，大数据技术能够适配医院复杂网络环境，通过实时分析和自动化响应有效应对威胁。然而，案例也反映出部署成本和隐私保护的挑战，需要医院在

技术与合规间寻求平衡。

结论

大数据技术在医院网络安全威胁识别中发挥了关键作用，为应对复杂威胁提供了高效解决方案。通过数据采集与处理，大数据整合了电子健康记录、网络流量和设备日志等多源数据，为威胁检测奠定了基础。分析技术如机器学习和流处理能够实时监控异常行为，精准识别勒索软件、钓鱼攻击等威胁，显著提升响应速度。案例分析表明，大数据支持的入侵检测系统和威胁情报平台已在医院实践中有效减少数据泄露和服务中断风险。其自动化和智能化特性降低了人工干预需求，适配了医院高并发、复杂网络环境。大数据不仅增强了威胁识别的准确性，还通过预测性建模为主动防御提供了可能。尽管面临隐私保护和部署成本的挑战，大数据仍是医院网络安全不可或缺的支柱。

参考文献

[1]Wasserman, L., Wasserman, Y.: Hospital cybersecurity risks and gaps: Review (for the non-cyber professional).[J] Frontiers in Digital Health 4, (2022) 1-20.

[2]Coventry, L., Branley-Bell, D.: Cybersecurity in healthcare: A narrative review of trends, threats and ways forward. Maturitas 113, (2018) 48-52 .

[3]Ghafur, S., Kristensen, S., Honeyford, K., Martin, G., Darzi, A., Aylin, P.: A retrospective impact analysis of the WannaCry cyberattack on the NHS.[J] npj Digital Medicine 2,(2019) 98-99 .

[4]Jalali, M. S., Kaiser, J. P.: Cybersecurity in hospitals: A systematic, organizational perspective.[J] Journal of Medical Internet Research 20(5), (2023) e100-121

[5]齐欢庆.大数据在医院网络安全防御中的应用与研究[J].网络安全技术与应用,2022,(06):118-120.

[6]庄一峰.基于大数据背景的医院网络信息安全防范分析[J].网络安全技术与应用,2020,(11):138-140.

[7]Vilakazi, K., Adebesin, F.: A systematic literature review on cybersecurity threats to healthcare data and mitigation strategies.[J] Proceedings of Society 5.0 Conference 93, (2023)240-251.

[8]Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I., Guizani, M.: A survey of machine and deep learning methods for Internet of Things (IoT) security.[J].IEEE Communications Surveys & Tutorials 22(3), (2020) 1646-1685.

[9]Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., Ng, A.: Cybersecurity data science: An overview from machine learning perspective.[J] Journal of Big Data 7, (2020) 41-45.

[10]曾运强.大数据时代医院网络安全防御架构研究与设计[J].现代信息科技,2020,4(06):161-162+166.