

网络安全态势感知技术研究：数据整合与可视化系统设计

李绍铭¹ 李宗阳¹ 周岳亮¹

LI Shaoming LI Zongyang ZHOU Yueliang

摘要

传统医院信息系统在面对网络安全威胁时存在安全规划不足、互联网接入防护薄弱以及人员权限管理混乱的问题，缺乏有效的数据整合与态势感知能力。针对上述问题，提出了一种基于网络安全态势感知的数据整合与可视化系统方法。首先，通过多源异构数据融合、数据预处理、数据关联与存储以及粒子群优化算法构建融合模型，提高数据权重分配的精度。其次，构建基于特征与行为的安全态势感知模型，采用特征提取与隐马尔可夫模型识别安全威胁。系统设计分为前端与后端，利用 D3.js、Spring Boot 和 Node.js 等技术实现实时监控、警报通知、交互式图表。所设计的系统能够有效融合多源数据，实时检测和展示网络安全态势，提升医院信息系统的安全管理水平。

关键词

网络安全；态势感知；数据整合；可视化系统；系统设计

doi: 10.3969/j.issn.1672-9528.2024.10.040

0 引言

随着互联网技术的迅猛发展和医院信息化建设的不断深入，医院信息系统面临的网络安全威胁日益严重。现有医院系统运行过程中仍旧无法完全克服网络安全问题，难以应对复杂多变的网络攻击环境。在此背景下，研究基于网络安全态势感知的数据整合与可视化系统具有重要意义。该研究旨在通过多源异构数据的融合、特征提取与行为分析，构建高效、准确的安全态势感知模型，并通过可视化技术直观展示复杂的数据关系与安全态势，从而提高医院信息系统的安全管理水平。通过实现对网络安全威胁的实时监控和动态分析，该系统不仅能够及时发现并应对潜在的安全威胁，还能为安全决策提供强有力的数据支持，对提升医院整体安全防护能力具有重要的应用价值与研究意义。

1 医院面临的网络安全风险

为应对日益复杂的网络环境，医院作为落实民生工程的重要载体，纷纷采取一系列措施，用于构建网络安全防护体系，以此面对外部威胁。在互联网医疗背景下，存在核心业务系统安全规划不足、互联网接入端防护薄弱以及人员管理等多方面风险，需要在原有平台技术构架基础上进一步加强数据整合与处理，通过实时流量分析、深度智能感知技术增强医院对网络威胁的监测、预警与响应，满足网络安全等级保护 2.0 标准要求。针对当前存在的问题，主要包括以下三点。

第一，医院核心业务系统，如医院信息系统（HIS）、

电子病历系统（EMRS）、医院影像系统（PACS）、实验室信息系统（LIS）等，存在由于底层架构问题导致的安全隐患。这些系统在建设初期缺乏完善的安全规划，导致在网络安全等级保护定级备案时面临修复难题。传统安全防护设备无法实时监测并主动防御现有的漏洞、安全隐患问题，缺乏快速响应机制，威胁通过非核心业务系统攻入后直接危害核心业务系统^[1]。此外，系统接入终端错综复杂，资产监控不足，增加了安全管理的难度。

第二，互联网接入端风险也有所增加。医院混合云平台建设后，互联网医院系统面向患者提供多场景服务，导致内外网交互需求增加，内网存储的医疗业务数据暴露在互联网上，成为非法攻击的目标。安全漏洞普遍存在，互联网医院系统在三级等保测评时发现暴力破解、弱口令、用户越权访问和 SQL 注入等问题，存在后台被植入远程操控木马的风险，网络安全防御能力不足。

第三，人员管理方面，运维人员权限管理混乱，缺乏细化的权限划分，未授权访问、越权访问问题频发。运维人员直接连接服务器进行操作，医院缺乏主动发现和防御这些问题的能力，这些风险使医院的网络安全面临严峻挑战。

2 基于网络安全态势感知的数据整合技术

2.1 多源异构数据融合

2.1.1 数据预处理技术

数据预处理是保障数据质量的关键步骤，涉及数据清洗、归一化、降维等技术。在医院管理系统中，这些技术对于处理来自不同来源的数据至关重要^[2]。数据清洗阶段，通过去

1. 右江民族医学院附属医院 广西百色 533000

除噪声、修复缺失值、处理重复数据，确保数据的准确性与完整性。使用正则表达式检测并纠正格式错误，例如电话号码、身份证号等。对于缺失值，采用均值填补法，公式为：

$$x_{\text{new}} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

式中： x_{new} 表示填补后的值， x_i 表示相邻数据点， n 表示有效数据点。对于重复数据，则使用哈希函数快速检测去重。

数据归一化采用 Min-Max 归一化和 Z-score 标准化技术，统一不同尺度的数据。Min-Max 归一化公式为：

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

式中： x 表示原始数据， $x_{\min}$ 和 $x_{\max}$ 分别表示数据最小值与最大值。

Z-score 标准化公式为：

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

式中： z 表示标准化后的值， μ 表示数据均值， σ 表示数据的标准差。归一化步骤消除量纲差异，提高数据处理的一致性，适用于不同来源数据，如患者信息、实验室结果、影像数据。

完成预处理与归一化后的数据还需进一步优化，可采用降维技术，如主成分分析（PCA）、线性判别分析（LDA）用于提取主要特征，减少数据维度^[3]。此次研究选择主成分分析法，PCA 通过协方差矩阵计算主成分，其公式为：

$$C = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T \quad (4)$$

式中： C 表示协方差矩阵， x_i 表示数据向量， μ 表示均值向量。计算特征值与特征向量，选择最大特征值对应的特征向量作为主成分。LDA 通过计算类内散布矩阵和类间散布矩阵，优化类间差异与类内相似性。

在医院管理系统中，数据预处理通过 ETL（提取、转换、加载）工具实现，确保数据的实时性，提高处理效率。通过上述预处理步骤，将来自不同系统的数据统一处理，为后续的数据融合与分析提供高质量的输入基础，有助于提高网络安全态势感知的准确性与可靠性。

2.1.2 数据关联与存储

数据关联与存储是实现多源异构数据融合的关键环节。在医院管理系统中，不同数据源（如 EMR、LIS、PACS）需要进行关联分析与高效存储^[4]。数据关联采用时间同步、事件匹配技术。时间同步通过网络时间协议（NTP）确保各系统时间一致性，避免因时间差异导致数据不匹配。事件匹配则依赖于特征标记、上下文分析，实现不同数据源事件的确切关联。

数据关联部分采用时间同步与事件匹配技术，其中网络事件协议 NTP 的时间同步公式为：

$$T = T_0 + \frac{(T_1 - T_0) + (T_2 - T_3)}{2} \quad (5)$$

式中： T_0 表示客户端请求时间， T_1 表示服务器响应时间， T_2 表示服务器接收请求时间， T_3 表示客户端接收响应时间。通过该公式，可以计算出同步时间 T ，确保所有系统时间一致。

事件匹配依赖特征标记与上下文分析，例如在医院实际应用中，患者的电子病历数据需要与实验室测试结果与影像数据进行匹配与关联^[5]。使用患者 ID 以及检查时间作为特征标记，通过 Jaccard 相似系数计算相似度，实现数据的精准关联，其公式为：

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (6)$$

式中： A 、 B 表示两个数据集，通过计算器相似度来确定事件匹配度。

例如 A 表示电子病历数据集， B 表示实验室数据集，则通过计算其相似度，确定事件的匹配度。当患者在 EMR 系统中进行了一次检查时，系统会记录时间戳和患者 ID，这些数据将用于匹配 LIS、PACS 系统中的相应记录，确保不同系统的数据可以准确关联。

针对数据库选择与优化方面，采用 NoSQL 数据库处理大量非结构化数据，如日志文件、影像数据，是基于当前医院数据库在安全态势感知中的具体问题。传统关系型数据库在处理这些非结构化数据时表现出查询效率低下、扩展性差、灵活性不足的问题^[6]。在医院环境中，日志文件、影像数据占据大量存储空间，并且需要快速查询、实时分析。关系型数据库在面对如此庞大的数据量和复杂的数据结构时，难以满足实时性要求，主要表现为查询速度下降，安全事件监控滞后，无法快速响应。对此，在数据库选择阶段选用 MongoDB 和 Cassandra 具有高扩展性与灵活性特征，能够有效处理非结构化数据，支持横向扩展，可以轻松增加存储节点以处理不断增长的数据量，确保系统在处理高并发查询时仍能保持高性能。其中哈希分片公式为：

$$S = H(k) \bmod n \quad (7)$$

式中： S 表示分片编号； k 表示患者 ID； $\bmod$ 表示取模运算。

通过哈希函数将数据分配到不同的分片。索引设计则使用 B 树索引提高查询速度，特别是在处理大量查询时。B 树索引的查找时间复杂度为 $O(\log n)$ ，提升查询效率。

为实现数据关联，采用基于图的数据库（如 Neo4j），通过节点与边的结构化存储，实现复杂关系的数据查询与分析。在医院系统中，医生、患者、医疗设备、检测结果等数据通过节点表示，关系通过边表示。例如，通过图数据库查询患者与其所有检测结果的关系，运行代码如下：

```
MATCH (p:Patient)-[:HAS_RESULT]->(r:Result)
WHERE p.patientID = '12345'
RETURN p, r
```

在数据存储过程中,采用数据压缩技术如 LZ4、Zstandard,减少存储空间占用,提升数据传输效率。LZ4 压缩算法通过快速压缩与解压缩,提供高效的数据处理能力,压缩率达到 2.1:1。

总体而言,数据关联与存储技术通过时间同步、事件匹配、数据库优化,能够实现医院管理系统中多源异构数据的高效融合与存储,为网络安全态势感知提供坚实数据基础。此类技术的应用能够保证数据的准确性、实时性以及可用性,有助于提高网络安全态势感知的精确度与反应速度^[7]。

2.1.3 融合模型构建

粒子群优化算法(PSO)是一种基于群体智能的优化算法,通过模拟鸟群觅食行为进行全局优化搜索。其优点包括计算简单、收敛速度快、参数设置较少,适用于多维度、多特征的数据优化。在医院管理系统中,PSO 算法被选用于优化数据权重分配,以提高多源异构数据的融合精度,确保不同来源数据的有效整合与关联分析^[8]。

PSO 算法的基本步骤如下。

(1) 初始化粒子群,每个粒子代表一个潜在解决方案,其位置和速度通过随机分布设定。每个粒子的适应度函数用于评估其质量。在医院数据融合中,适应度函数可以基于数据匹配度和关联度进行设计。适应度函数公式为:

$$f(x) = \sum_{i=1}^n w_i \cdot d_i \quad (8)$$

式中: w_i 表示权重, d_i 表示数据匹配度或关联度, n 表示数据维度数量。

(2) 在算法迭代过程中,每个粒子根据自身与全局最优位置调整速度与位置。速度更新公式为:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1 (p_i^{\text{best}} - x_i(t)) + c_2 r_2 (g^{\text{best}} - x_i(t)) \quad (9)$$

式中: $v_i(t)$ 表示粒子 i 在时间 t 的速度,表示医院数据权重在当前迭代中的变化速度; ω 表示惯性权重,用于平衡全局搜索与局部搜索,影响医院数据融合权重的更新幅度; c_1 、 c_2 是学习因子,表示粒子在自己的经验中学习的程度,用于调整医院系统中当前粒子与历史最优位置的距离; r_1 、 r_2 是随机数,在 $[0,1]$ 之间变化,增加随机性,避免医院数据融合过程中陷入局部最优解; p_i^{best} 表示粒子 i 的历史最优位置,表示在医院数据融合中,当前粒子所找到的最优权重分配方案; g^{best} 是全局最优位置,表示在医院数据融合过程中,所有粒子共同找到的最优权重分配方案^[9]。

(3) 位置更新公式为:

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (10)$$

式中: $v_i(t)$ 表示粒子速度, $x_i(t)$ 表示粒子位置。

在医院管理系统中,融合模型将患者的电子病历、实验室结果、影像数据权重进行优化分配,确保数据融合的准确性和一致性。例如,考虑一个实际案例,某患者在不同系统中有血液检测结果、影像数据以及详细的病历记录。这些数据需要被综合分析,以提供全面的安全态势感知。通过 PSO 算法优化后的权重分配,可以使这些数据更准确地反映患者的健康状况,支持医疗决策与网络安全态势感知。

具体应用过程中,通过实时采集患者的多源数据,PSO 算法动态调整每种数据源的权重,以获得最优的融合结果。融合模型构建过程中,采用交叉验证方法评估融合模型的性能,通过划分训练集与验证集,确保模型的泛化能力。例如,使用 80% 的数据进行训练,20% 的数据进行验证,以检验模型稳定性与可靠性。

2.2 安全态势感知模型构建

2.2.1 基于特征的检测

基于特征的检测通过提取与分析数据中的特征来识别安全威胁^[10]。在医院系统中,这些特征包括网络流量模式、用户登录行为和系统调用频率等。特征提取过程利用机器学习算法,如支持向量机(SVM)、决策树。SVM 通过构建高维空间中的超平面进行分类,而决策树则通过递归分割数据集来建立分类模型。医院网络流量的数据特征可通过以下公式提取:

$$f(x) = \sum_{i=1}^n w_i \cdot x_i \quad (11)$$

式中: $f(x)$ 表示特征函数, w_i 表示权重, x_i 表示输入数据。通过计算不同特征权重与数据点特征值,可以准确识别异常行为。

2.2.2 基于行为的检测

基于行为的检测通过分析系统与用户行为模式,识别潜在的安全威胁。在医院系统中,行为模式包括用户访问日志、系统操作日志、设备使用记录等。行为检测模型采用序列模式挖掘并识别正常或异常行为模式。隐马尔可夫模型(HMM)是常用的行为检测方法,通过构建状态转移矩阵与观测概率矩阵来捕捉行为序列。状态转移矩阵 A 、观测概率矩阵 B 的定义为:

$$A = \{a_{ij}\} = P(S_{t+1} = j | S_t = i) \quad (12)$$

$$B = \{b_j(o_t)\} = P(O_t = o_t | S_t = j) \quad (13)$$

式中: A 、 B 分别表示状态转移矩阵与观测概率矩阵; a_{ij} 表从状态 i 转移到状态 j 的概率; $b_j(o_t)$ 表示在状态 j 下观测值为 O_t 的概率。通过训练与推断, HMM 可以有效检测出医院系统中的异常行为。

3 可视化系统设计

3.1 框架结构

3.1.1 前端设计

可视化系统的设计在网络安全态势感知中扮演着关键角色，能够直观呈现复杂的数据关系及安全态势，为决策提供支持。系统框架分为前端、后端两个部分，紧密协作实现数据的实时处理，并展示具体操作过程的数据信息变动。前端设计旨在提供用户友好的界面，通过图形化手段展示网络安全态势^[11]。前端采用响应式设计，确保在不同设备上都能良好显示。使用的技术栈包括 HTML5、CSS3、JavaScript 以及 D3.js 等可视化库。D3.js 通过数据驱动文档技术，将复杂的数据转化为互动性强、视觉效果丰富的图表、图形。前端界面包含多个组件，如实时监控面板、警报通知系统以及交互式图表。实时监控面板显示当前的网络安全状态，警报通知系统及时告知安全事件，交互式图表支持用户深入分析数据。

3.1.2 后端设计

后端设计负责数据处理、存储、传输，确保前端能实时获取最新的安全态势数据。后端架构基于微服务，采用 Spring Boot 和 Node.js 等技术构建。数据存储使用 NoSQL 数据库（如 MongoDB）以及关系型数据库（如 MySQL），分别处理非结构化和结构化数据^[12]。后端通过 RESTful API 接口与前端进行数据交互，确保数据传输的高效性与安全性。实时数据处理采用 Apache Kafka、Flink，实现高吞吐量与低延迟的数据流处理。数据安全通过加密传输与访问控制机制保障，确保敏感数据不被未授权访问。

3.2 可视化技术

可视化技术是整个系统的核心，通过将复杂的数据转化为易于理解的图形，帮助用户快速掌握网络安全态势。主要采用的技术包括以下几方面。

（1）图表可视化：使用折线图、柱状图、饼图、热力图等多种图表类型，直观展示不同维度的数据，如攻击趋势、事件分布、风险等级。

（2）地理信息系统（GIS）：通过 GIS 技术，将网络攻击来源与目标呈现在地理信息中，实现可视化，帮助用户了解被攻击区域以及走势。GIS 技术集成了 OpenLayers、Leaflet 等开源库，支持丰富的地理数据展示。

（3）交互式可视化：提供用户与数据的交互功能，通过点击、悬停、缩放等操作，用户可以深入探索数据细节，进行自定义分析。交互式图表通过绑定事件监听器，回调函数，实现动态更新与及时响应。

（4）时序分析：采用时间轴、趋势图等工具，分析安全事件的时间变化规律，帮助用户识别潜在威胁与异常行为。

4 结语

综合来看，为保证医院业务运行安全，应进一步加强对网络安全态势感知技术的深入研究，旨在解决医院信息系统在数据整合、安全监测方面的诸多难题。文章提出在原有基础上采用粒子群优化算法的方式进行数据权重分配，并在前端、后端系统框架设计中引入 D3.js 等可视化技术，进一步提高互动性，提供强有力的决策支持。该研究成果能够有效填补当前理论空白，在未来实践中具有一定的应用前景。

参考文献：

- [1] 王延培,汪偲.大数据背景下网络信息安全防护设计[J].信息与电脑(理论版),2024,36(7):212-214.
- [2] 唐丽萍.基于大数据与智能技术的信息安全态势感知系统分析[J].信息与电脑(理论版),2024,36(7):215-217.
- [3] 王天平,李珍.智能时代在线课程的应然样态、实然困境与实践路向[J].教育与教学研究,2024,38(4):20-31.
- [4] 彭永倩.融合计算迁移模型的态势感知策略实现方法研究[J].现代信息科技,2024,8(7):172-178.
- [5] 叶栋.大数据时代计算机网络安全技术的优化策略[J].网络安全和信息化,2024(4):124-126.
- [6] 周文桀,徐顺航,刘丽红.马尔可夫攻防模型下网络边缘态势监控仿真[J/OL].计算机仿真,1-6[2024-05-30].http://kns.cnki.net/kcms/detail/11.3724.TP.20240327.1648.002.html.
- [7] 薛永平,郭治豪,周展利.网络安全态势感知与响应的自适应方法[J].信息与电脑(理论版),2024,36(6):185-187.
- [8] 颜唐林.基于态势感知技术的广电智能网络安全调度预警架构设计[J].电视技术,2024,48(3):189-192+205.
- [9] 吕华辉,明哲,樊凯,等.基于数据感知融合的电力通信网络智能运维[J].电信科学,2024,40(1):136-143.
- [10] 何文雯,张涛涛,吴若无.基于改进的模糊神经网络的网络运行态势感知技术[J].通信技术,2024,57(1):47-53.
- [11] 徐言海.基于大数据与智能技术的信息安全态势感知系统分析[J].电子技术,2024,53(1):378-380.
- [12] 贺再平.大数据在网络安全态势感知平台的应用[J].家庭影院技术,2024(2):31-34.

【作者简介】

李绍铭（1992—），男，广西百色人，本科，初级工程师，研究方向：网络体系结构、网络和系统安全、网络应用及安全。

周岳亮（1985—），通信作者（email: 577668868@qq.com），男，湖南衡山人，硕士，工程师，研究方向：网络与信息安全。

（收稿日期：2024-07-16）