基于改进哈希算法的网络信息安全溯源体系研究

康　茜 [1]
KANG Qian

摘　要　针对复杂网络环境下信息安全溯源精度不足、抗篡改能力薄弱，且在大规模数据场景下溯源效率低下的问题，文章对网络信息安全溯源方法进行了深入探索与优化。以深度哈希算法为核心框架，重点优化了哈希码生成、数据一致性检测及节点验证三大关键环节，尤其对 DRH 算法的特征提取层结构与汉明空间映射机制进行了针对性改进，构建了高效可靠的溯源模型。实验结果表明，与现有主流溯源模型相比，所提方法在溯源精度与运行效率上均展现出显著优势：信息溯源检索精度最高可达 92.3%，信息检索一致性稳定在 85% 以上，响应时间仅为 98.67 ms，吞吐量最高达到 215.55 请求 /s。研究结果证实，所提方法在真实复杂网络环境中具备优异的效率与可靠性，可为复杂环境下的网络信息安全溯源提供新的技术路径与解决方案。

关键词　哈希算法；DRH；网络信息；溯源；生成平衡

doi：10.3969/j.issn.1672-9528.2026.03.028

0 引言

随着数字化和智能化的发展，网络信息安全问题日益突出，信息的追溯和篡改检测成为保障数据可信性的重要研究课题 [1]。信息溯源能够有效跟踪数据的来源和流通过程，从而及时发现数据异常，保障信息的完整性和真实性。在此背景下，溯源技术在供应链管理、物联网等领域逐渐受到广泛关注 [2]。李达等 [3] 研究发现，在能源电力项目工程数据生成过程中，数据易遭篡改，导致数据追溯和存证验证难度较高的问题，设计了一种基于区块链的能源电力数据验证与追溯模型。实验结果表明，与传统方法相比，该模型能够有效解

1. 山西工程科技职业大学 山西晋中 030619

率达到 97.44%，召回率和精确率均超过 97%，显著优于传统 SMOTE-RF 模型及其他先进算法。表明 DDPM 生成的高质量平衡数据能够有效缓解分类模型对少数类样本的学习不足，提升模型在实际不平衡网络环境中的攻击检测能力。

同时，基于层次分析法的态势评估模块实现了网络安全状态的多层次量化评估与可视化。通过服务层、主机层到网络层的自底向上态势计算，能够精准定位安全威胁源头，直观展示网络安全状态的动态变化趋势，为网络安全管理和应急响应提供了科学的决策依据。

综上所述，本文提出的 DDPM-RF 网络安全态势感知算法通过高质量数据生成、精准攻击识别和系统化态势评估三个环节的有机结合，有效提升了在不平衡数据场景下的网络安全态势感知性能。该方法不仅为网络安全防护提供了新的技术思路，也为相关领域的数据不平衡问题解决提供了可借鉴的解决方案。未来研究可进一步探索 DDPM 在其他网络安全数据集上的泛化能力，并尝试将深度学习模型与集成学习方法相结合，以应对日益复杂的网络攻击威胁。

参考文献：

[1] 黄健 . AI 在网络安全态势感知中的异常检测技术研究 [J]. 互联网周刊 , 2025(17):34-36.

[2] 耿美月 . 基于深度学习的计算机网络安全态势感知技术 [J]. 信息记录材料 , 2025,26(9):143-145.

[3] 康会娟 , 张腾飞 . 基于机器学习的网络安全态势感知系统研究 [J]. 电脑知识与技术 ,2025,21(20):88-90.

[4] 杨宇 , 谷宇恒 . 网络安全态势感知综述 [J]. 科学技术与工程 , 2022,22(34):15011-15019.

[5] 朱光剑 , 李阳冬 , 钱平 . 基于智能化网络模型的信息安全态势感知算法 [J]. 信息技术 ,2025(5):135-139.

[6] 林广朋 , 李闯 . 入侵攻击下无线网络安全态势感知算法 [J]. 计算机仿真 , 2023,40(12):451-454.

【作者简介】

胡琦涛（1997—），男，黑龙江佳木斯人，本科，研究方向：网络安全、应用电子。

（收稿日期：2025-12-25　修回日期：2026-03-11）

决能源电力场景中存在的数据存取验证与追溯问题。Ezhil 等 [4] 发现大多数传统方法没有考虑控制用户访问云数据存储和共享的问题，为此，研究人员提出了一种可审计的基于属性的数据溯源和加密模型。实验结果表明，该模型有效避免了数据完整性丢失，数据溯源一致性较高。然而上述方法在处理大规模数据时往往面临计算瓶颈和精度下降的问题。为此，研究分别对哈希码生成、特征提取层和汉明空间映射进行改进，提出了一种新型网络信息安全溯源模型。研究创新点在于引入了深度哈希算法（deep-re-hash, DRH）生成具有特定相似度的哈希码。研究旨在为网络信息安全溯源提供一种高效、可靠的解决方案，有望在物联网、智能制造等领域实现更广泛的应用。

1 哈希算法

哈希算法是一种将任意长度的数据输入映射到固定长度的散列值的函数。其中，DRH 是哈希算法在深度学习中的扩展，利用深度神经网络生成具有特定相似度特性的哈希码，广泛应用于图像检索和相似性匹配 [5-6]。DRH 首先将数据通过卷积神经网络等模型提取特征，然后将特征嵌入低维空间并输出二进制码 [7-8]。DRH 的特征提取计算公式为：

$$h(x) = f(x;\theta) \tag{1}$$

式中：x 表示输入数据；θ 表示神经网络参数；f 表示特征提取函数。

将输出的数据特征向量映射到一个低维空间中的二进制哈希码，计算公式为：

$$b(x) = \mathrm{sign}(h(x)) \tag{2}$$

式中：$b(x)$ 表示二进制哈希码；sign(·) 表示符号函数，将特征值量化为 +1 或 −1，以便生成二进制位。

2 网络信息安全溯源模型构建

2.1 改进深度哈希算法设计

与传统哈希方法不同，DRH 在保持高检索精度的同时大幅提升了计算效率，非常适合大规模数据的快速匹配和信息检索。基于此特性，DRH 算法通常对卷积网络的全连接层进行优化，增加 n+1 个全连接层，用以控制神经元数量和哈希码长度 [9-10]。DRH 算法通常有两种检索策略，即汉明排序和汉明检索。其中，汉明排序通过计算查询哈希码与数据库哈希码之间的汉明距离来进行相似度排序；而汉明检索则基于指定汉明距离范围内的快速匹配，直接找到与查询哈希码最接近的结果 [11]。汉明距离和汉明检索的计算公式为：

$$\begin{cases} D_H(\mathrm{HM}(X),\mathrm{HM}(Y)) = \\ \sum_{i=1}^{L}[\mathrm{HM}(X)_i \neq \mathrm{HM}(Y)_i] \\ \mathrm{Search}D_H(\mathrm{HM}(Q),\tau) = \\ \{X \mid D_H(\mathrm{HM}(Q),\mathrm{HM}(X) \leq \tau)\} \end{cases} \tag{3}$$

式中：D_H 表示汉明距离；L 表示哈希码的长度；HM(X) 和 HM(Y) 分别表示输入 X 和 Y 的哈希码；Search D_H 表示汉明检索结果；HM(Q) 表示给定的查询哈希码；τ 表示既定阈值。考虑到信号的单类和多类特性，研究针对单信号和多信号场景下 DRH 哈希中心的生成提出了一种加权平均策略，该策略如图 1 所示。

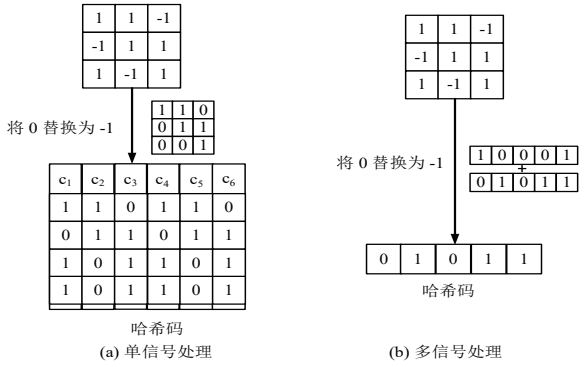

图 1 哈希生成平衡策略示意图

图 1（a）为单信号的哈希中心处理生成策略；图 1（b）为多信号的哈希中心处理生成策略。由图 1 可知，对于单信号数据，哈希中心可视为该类别的典型代表哈希码，并且可以通过计算同一类别中所有信号的哈希码平均值来直接得到。对于多信号数据，每个信号可能同时属于多个类别，过程更为复杂，因为需要综合考虑每个信号的多标签影响。此时，单标签或多标签图像数据的哈希中心都应该具有独立性和平衡性，用公式表示为：

$$C_j^* = \frac{\sum_{i=1}^{N} y_{ij} \cdot H(x_i) + \lambda \sum_{j=1}^{M} \omega_{ij} \cdot H(z_j)}{\sum_{i=1}^{N} y_{ij} + \lambda \sum_{j=1}^{M} \omega_{ij}} \tag{4}$$

式中：C_j^* 表示经优化后的第 j 类哈希中心；N 和 M 分别表示单信号数据和多信号数据的数量；z_j 表示多信号数据；λ 表示平衡系数；ω_{ij} 表示多信号数据的权重系数；$H(x_i)$ 表示数据点 x_i 的哈希表示；y_{ij} 表示一个指示变量。

2.2 基于改进哈希算法的网络信息安全溯源模型构建

在对哈希算法进行改进后，研究将其应用于网络信息安全溯源。以产品网络信息溯源为例，首先，溯源系统会对每个产品的生产、流通、销售等环节的信息进行记录和跟踪 [12]。产品网络信息的溯源示意图如图 2 所示。

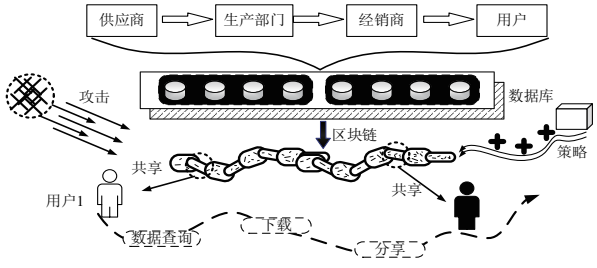

图 2 产品网络信息的溯源示意图

由图 2 可知，产品网络信息的溯源主要由信息采集模块、数据存储模块以及用户访问模块组成。信息采集模块从供应商、生产部门、经销商和用户等各环节收集产品信息，并存储在数据库中。在数据存储模块中，所有产品溯源信息被集中保存，并通过区块链技术实现多节点共享。用户访问模块则为各类用户提供了数据查询、下载和分享等功能。若区块链发生节点失效或遭受攻击，将会使产品的网络信息安全受到威胁，可能引发例如数据篡改、信息丢失等安全性事故。综上，研究结合改进 DRH 构建了一种产品网络信息安全溯源策略模型，该策略模型如图 3 所示。

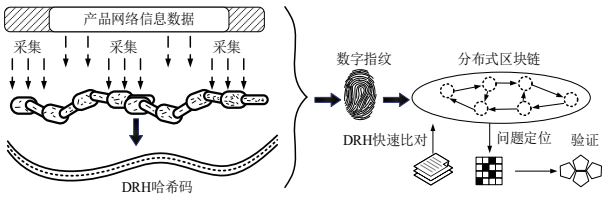

图 3 产品网络信息安全溯源策略模型

由图 3 可知，在信息采集环节，数据存储节点对每个产品的核心特征进行提取，并使用改进的 DRH 算法生成唯一的哈希码作为该产品的数字指纹。在数据传输环节，这些哈希码被写入区块链的分布式网络中，实现多节点共享和不可篡改性。另外，若某节点的数据遭受篡改或攻击，系统可以通过快速比对改进 DRH 生成的哈希码检测出异常，并实时定位问题节点。在验证环节，验证节点会对传输的数据与其原始哈希指纹进行比对，确保数据真实性。验证过程中哈希对比的计算公式为：

$$S(x)=\sum_{t=1}^{T}\gamma^{t}H(x_t) \tag{5}$$

式中：$S(x)$ 表示在时间段 T 内的数据传输完整性验证值；$H(x_t)$ 表示在时间 t 生成的哈希码；γ 表示衰减因子。

此外，采用加权差异检测机制，对每个节点传输的数据进行相似性校验，以识别出异常节点或篡改信息。该机制的计算公式为：

$$D(x,y)=\frac{1}{n}\sum_{n}^{1}\left|H^{*}(x)_i-H^{*}(y)_i\right|\cdot\mu_i \tag{6}$$

式中：$D(x, y)$ 表示节点 x 和 y 的加权差异值；μ_i 和 $H^{*}(y)_i$ 分别表示两个节点的第 i 位哈希码；μ_i 表示该点位的权重因子。若 $D(x, y)$ 超出预定阈值，则认为节点之间的传输存在异常，可能涉及信息篡改或数据损坏。

3 网络信息安全溯源模型测试分析

研究设置 CPU 为 Intel Xeon E5-2680 v4，GPU 为 NVIDIA Tesla V100 32 GB，内存为 128 GB。此外，实验所需的数据库运行在 MySQL 8.0 上，区块链框架采用 Hyperledger Fabric 2.2。采用 GS1 物流追踪数据集和 TPCx-IoT 数据集作为测试数据来源。其中，GS1 包含了供应链中不同环节的物流追踪数据，涉及产品的来源、生产日期、物流状态、位置和时间戳等信息。TPCx-IoT 模拟了物联网环境中物流和生产管理场景的海量数据，包括产品 ID、生产工序、传输位置、时间戳信息等。实验测试环境如图 4 所示。

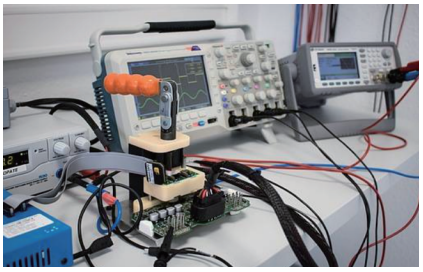

图 4 网络信息安全溯源测试环境

在此环境下，研究引入了同类型较为先进的溯源模型进行了对比，例如基于区块链的溯源模型（blockchain-based traceability model, BBTM）、基于图神经网络的溯源模型（graph neural network-based traceability model, GNN-TM）和基于深度对抗网络的溯源模型（deep adversarial network-based traceability model, DAN-TM）。测试结果如图 5 所示。图 5（a）为 GS1 物流追踪数据集中四类模型的检索精度测试结果；图 5（b）为 TPCx-IoT 数据集中四类模型的检索精度测试结果。

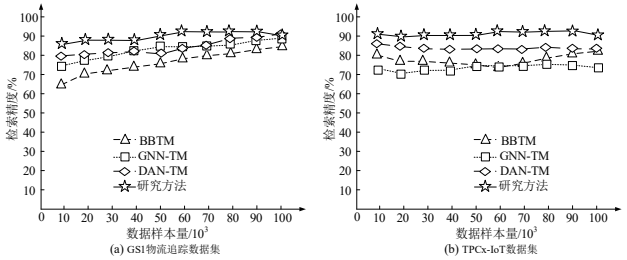

图 5 不同溯源模型的检索精度测试结果

由图 5（a）可知，研究方法在 GS1 数据集中的精度稳定在 90.7% 以上，相比之下，BBTM 和 GNN-TM 的精度在样本量较大时略有下降，表明研究方法在大数据环境下的表现更为稳定和高效；图 5（b）可知，研究方法在样本量增加的情况下仍保持在 92.3% 以上的精度。相比之下，BBTM 和 DAN-TM 的精度随样本量增加有轻微波动，这说明研究方法在不同数据集上均表现出较强的鲁棒性和精度优势，适合在大规模产品网络信息溯源中应用。以检索一致性为指标验证模型在不同的数据传输时间段内的抗篡改能力，测试结果如图 6 所示。

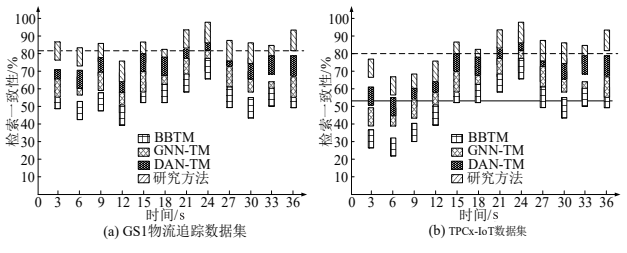

图 6 不同模型的检索一致性测试结果

图 6（a）为 GS1 物流追踪数据集中不同模型的检索一致性测试结果；图 6（b）为 TPCx-IoT 数据集中不同模型的检索一致性测试结果。

由图 6（a）可知，研究所提方法的检索一致性表现始终保持在 80% 以上。相比之下，BBTM 和 GNN-TM 的检索一致性随着时间增加波动较大，尤其在传输时间超过 30 s 后，一致性显著下降；图 6（b）可知，研究所提方法在 20 s 以内的传输时间段内，检索一致性稳定在 85% 以上。而其他模型的检索一致性在 24 s 以后多数低于 70%，由此可知，研究所提方法在长时间的数据传输过程中具有更强的抗篡改能力。以智能手表和高端手机为例，对比四种模型在响应时间、抗攻击性、吞吐量和溯源 F_1 值等指标下的表现，结果如表 1 所示。

表 1 产品网络信息溯源多指标测试结果

产品对象	模型	响应时间 /ms	抗攻击性 /%	吞吐量 /(请求 ·s^{-1})	溯源 F_1 值
智能手表	BBTM	125.35	87.62	180.12	0.88
	GNN-TM	110.48	83.21	195.34	0.85
	DA-TM	102.29	78.55	160.27	0.81
	研究方法	98.67	92.73	210.49	0.91
高端手机	BBTM	130.45	88.12	175.89	0.87
	GNN-TM	115.76	84.43	190.22	0.84
	DA-TM	105.89	80.45	165.33	0.82
	研究方法	100.12	93.45	215.55	0.92

由表 1 可知，在响应时间方面，研究模型在智能手表和高端手机上的平均响应时间分别为 98.67 ms 和 100.12 ms，显著低于其他模型，显示出较高的响应效率。其次，在抗攻击性方面，研究模型表现出优异的抗篡改能力，智能手表和高端手机的抗攻击性分别为 92.73% 和 93.45%，远高于 BBTM、GNN-TM 和 DA-TM 模型。此外，研究模型智能手表和高端手机的吞吐量分别达到了 210.49 请求 /s 和 215.55 请求 /s，显示出其在大规模数据查询场景下的出色并发处理能力。最后，在溯源 F_1 值方面，研究模型的 F_1 值在智能手表和高端手机上分别为 0.91 和 0.92，略高于其他模型，表明其在溯源检索的精度和召回率方面具有良好的平衡性。

4 结论

研究针对大规模数据环境下的溯源需求，通过改进 DRH 在哈希码生成、数据一致性检测和节点验证等方面的性能，提出了一种新型网络信息安全溯源模型。实验结果表明，相比于其他三类先进算法模型，研究所提模型的溯源检索精度最高为 92.3%，表现出较优的数据处理能力。同时，其信息检索一致性稳定在 85% 以上。在智能手表和高端手机对象上的响应时间分别为 98.67 ms 和 100.12 ms；抗攻击性在两个产品对象上分别达到 92.73% 和 93.45%；吞吐量分别为 210.49 请求 /s 和 215.55 请求 /s；溯源 F_1 值在智能手表和高端手机上达到 0.91 和 0.92。相较于其他模型，研究方法在响应时间、抗攻击性和吞吐量等指标上均表现出显著优势，证明了该方法在真实环境中的高效性和可靠性。然而，研究仍存在一定局限性，例如在超大规模数据条件下的适应性和对多样化攻击手段的进一步防护效果还需优化。未来的研究将进一步提升模型的扩展性和鲁棒性，以适应更多实际应用场景，并探索在多源异构数据环境中的溯源能力提升。

参考文献：

[1] 李文举 , 王子杰 , 崔柳 . 基于 SIFT 和感知哈希改进的 CamShift 跟踪算法 [J]. 电子测量技术 , 2023, 46(4): 184-192.

[2] 张世良 , 王尧 , 邵晶晶 . 基于感知哈希的能源数据安全访问控制方法研究 [J]. 自动化仪表 , 2024, 45(5): 89-93.

[3] 李达 , 郭庆雷 , 王栋 , 等 . 基于区块链的能源电力数据验证与溯源模型 [J]. 中国电力 ,2023,56(5):203-208.

[4]EZHIL ARASI V, INDRA GANDHI K, KULOTHUNGAN K. Auditable attribute-based data access control using blockchain in cloud storage[J].The journal of supercomputing, 2022, 78(8):10772-10798.

[5] 张宇 , 贾正森 , 龙波 , 等 . 时间戳服务到 UTC(NIM) 的溯源方法研究 [J]. 仪器仪表学报 , 2024, 45(2): 176-183.

[6] 叶远波 , 王吉文 , 汪伟 , 等 . 基于哈希和编辑距离算法的 SCD 双层向量化与变更校验技术 [J]. 中国电力 , 2024, 57(1): 255-262.

[7] 萧展辉 , 张世良 , 宋云奎 . 基于感知哈希算法的能源数据安全快速检索 [J]. 自动化技术与应用 , 2024, 11(1):1-5.

[8] LI J H, CAO L F , ZHAO S C, et al. LC-PBFT: layered cross-chain consensus algorithm based on forest topology[J].The journal of supercomputing, 2024, 80(12):17849-17873.

[9] 童鑫红 , 高伟 , 李平辉 , 等 . 漏洞预测下级联网络安全态势感知研究 [J]. 自动化仪表 , 2024, 45(8): 69-73.

[10] 陈祖歌 , 毛冬 , 饶涵宇 , 等 . 面向物联网云计算的能源大数据存储优化算法 [J]. 浙江电力 , 2023, 42(8): 19-26.

[11] 章光东 , 孙楷淇 , 廖根宇 , 等 . 基于大数据的电网设备供应链溯源管理评价体系研究 [J]. 自动化技术与应用 , 2024, 10(21): 11-17.

[12]RASHEED A, BAZA M, BADR M M, et al. Efficient crypto engine for authenticated encryption, data traceability, and replay attack detection over CAN bus network[J]. IEEE transactions on network science and engineering, 2023, 11(1): 1008-1025.

【作者简介】

康茜（1989—），女，山西临汾人，硕士研究生，讲师，研究方向：计算机网络、5G 移动通信方向。

（收稿日期：2025-08-26 修回日期：2026-03-05）