arXiv:2601.16976v1 [cs.LG] 23 Jan 2026

用于入侵检测的物联网攻击数据生成潜在扩散模型*

预印本

Estela Sánchez-Carballo[†1], **Francisco M.Melgarejo-Meseguer**[1],和**JoséLuis Rojo-Álvarez**[1,2]1胡安·卡洛斯国王大学信号与通信理论、电信系统与计算系，马德里，西班牙。
[2]胡安·卡洛斯国王大学数据、网络与网络安全研究所，马德里，西班牙。

摘要

入侵检测系统（IDS）是保护物联网环境的关键组成部分。然而，在基于机器学习的入侵检测系统中，良性流量与攻击流量之间严重的类别不平衡往往会导致性能下降。尽管数据增强已被广泛用于缓解这一问题，但现有方法通常依赖简单的过采样技术或生成模型，难以同时实现高样本保真度、多样性和计算效率。为解决这些局限，我们提出将潜在扩散模型用于物联网入侵检测中的攻击数据增强，并与最先进的基线方法进行全面比较。我们针对三种具有代表性的物联网攻击类型开展实验，具体包括分布式拒绝服务攻击（DDoS）、Mirai和中间人攻击，并使用分布、依赖关系和多样性度量，评估下游入侵检测系统性能及生成质量本身。结果表明，使用潜在扩散模型生成的样本平衡训练数据，可显著提升入侵检测系统性能；对于分布式拒绝服务攻击和Mirai攻击，F1分数最高可达0.99，且始终优于其他竞争方法。此外，定量和定性分析表明，潜在扩散模型能够在生成多样化样本的同时有效保留特征依赖关系；与直接在数据空间运行的扩散模型相比，采样时间缩短约25%。这些发现凸显了潜在扩散是一种有效且可扩展的合成物联网攻击数据生成方案，能够显著减轻类别不平衡对物联网场景中基于机器学习的入侵检测系统造成的影响。

关键词 数据增强 · 扩散模型 · 物联网 · 入侵检测系统 · 潜在空间 · 表格数据合成。

1 引言

物联网（IoT）设备的快速普及已经改变了医疗保健、智慧城市、工业自动化和智能交通系统等多个领域 [Perwej 等人，2019，Lohiya 和 Thakkar， 2021]。然而，智能环境的增加也显著提高了现代网络的攻击率，使物联网生态系统尤其容易受到网络威胁 [Shafiq 等人， 2022]。因此，入侵检测系统（IDS）已成为保护物联网基础设施的关键 [Khraisat 等人， 2019]。近年来，基于机器学习（ML）和深度学习（DL）的入侵检测系统展现出强大的性能，因为它们能够从网络流量数据中学习复杂模式 [Shenfield 等人，2018，Melgarejo-Meseguer 等人，2021，Huang 等人， 2025]。然而，其性能高度依赖于大规模、具有代表性且均衡良好的数据集。在这一背景下，一个关键挑战是物联网环境面临多种攻击家族，从高流量的分布式拒绝服务攻击（DDoS）和僵尸网络驱动的泛洪攻击，到低速且隐蔽的中间人攻击（MitM）；这些攻击会分别造成不同的流量模式和检测挑战 [Kaur 等人， 2023]。此外，物联网攻击数据集还存在诸多局限。首先，收集真实世界攻击流量成本高、耗时长，并且通常受到隐私和监管方面的限制 [Alex 等人，2023，Dritsas 和 Trigka， 2025]。其次，与良性流量相比，这些数据集通常高度不平衡，其中并不常见但十分关键的攻击类型严重缺乏代表性

*本文已提交至 IEEE，可能会被发表
†estela.sanchezc@urjc.es

与良性流量相比 [Ma 等人，2025，Mallidi 和 Ramisetty， 2025]。第三，物联网流量数据本质上具有异构性，同时包含表格形式的连续特征（例如数据包大小、时间统计量）和分类特征（例如协议标志、连接状态）[Kaur 等人，2023，Rahman 等人， 2025]。这些特征使得稳健的入侵检测系统训练更加困难，也促使人们将合成数据生成技术作为数据收集的补充策略。

为解决物联网攻击数据集中的数据稀缺和类别不平衡问题，文献中提出了多种合成数据生成技术。经典过采样方法（如合成少数类过采样技术（SMOTE））通过在少数类样本之间插值来生成新样本 [Chawla 等人， 2002]。这类方法计算效率较高，在特征空间中进行局部操作，通常能够保持成对特征依赖关系，但往往生成与原始数据高度相似的新样本，导致多样性有限 [Azhar 等人， 2022]。近年来，深度生成模型已被用于网络安全领域的表格数据合成，包括变分自编码器（VAE）[Zhang 等人， 2023]和生成对抗网络（GAN） [Wang 等人， 2017]。这些模型能够学习全局数据分布并捕捉高阶相关性，但也存在局限。由于施加了潜在正则化，变分自编码器可能生成过度平滑的样本；而生成对抗网络则容易出现训练不稳定性、模式崩溃以及对超参数调优问题敏感等情况 [Wang 等，2017、2024，Li 等，2025a]。扩散模型（DM）近来已成为一种强大的生成建模方法，通过逐步对纯噪声样本进行去噪，在多个领域取得了最先进的成果 [Ho 等人， 2020]。其坚实的理论基础和稳定的训练过程使其适用于表格数据生成，但将其直接应用于表格型物联网数据仍面临挑战。直接在特征空间中操作，需要谨慎处理异质变量；物联网流量数据维度较高且扩散步数较多，导致采样速度较慢；此外，模型还可能难以处理统计尺度各异的特征[Wang 等，2024，Li 等，2025b]。

在本研究中，我们提出利用扩散模型相较于现有生成方法的优势，同时解决其在物联网流量数据场景中的已知局限。为此，我们引入一种潜在扩散框架：首先采用混合变量自编码器（自编码器）将异构网络流量特征编码到紧凑且经过归一化的潜在空间中，然后训练扩散模型以捕捉这些潜在表示的分布。在潜在域中执行扩散过程可提升训练稳定性，并显著降低计算复杂度，同时为具有不同统计属性的特征提供更统一的表示。与此同时，自编码器解码器保留了原始数据的语义结构，从而能够生成真实且连贯的合成物联网流量样本。潜在扩散模型（潜在扩散模型）最初用于图像合成 [Rombach 等人， 2022] ，近期在表格数据生成方面也展现出良好前景 [Sattarov 等，2023，Villaizán-Vallelado 等人，2025，Zhu 等人， 2025]。本研究的一项关键贡献，是将这一潜在扩散范式迁移至物联网流量领域，并证明其在入侵检测中的合成物联网攻击数据生成方面行之有效。

2 相关工作

类别不平衡是物联网入侵检测系统面临的一个众所周知的挑战，其中良性流量通常占主导地位，而许多攻击类型的样本严重不足 [Mallidi 和 Ramisetty， 2025]。SMOTE 等经典过采样技术已被广泛用于应对这一问题。例如，在 [Soe 等人， 2019] 和 [Dash 等人， 2020], 的研究中，研究者在训练基于深度学习的入侵检测系统之前，使用 SMOTE 解决不同物联网攻击数据集中的严重数据不平衡问题。然而，据报道，对于包含混合连续特征和类别特征的高度不平衡数据集，经典过采样技术带来的性能提升有限[Wang 等人， 2024], ，这促使研究者采用能够提升全局数据分布学习能力的深度生成模型。

变分自编码器作为经典过采样技术的替代方案，开始用于合成表格数据生成。它们通过学习数据的概率潜在表示，并借助解码器网络对其进行重构，从而生成新样本 [Pinheiro-Cinelli 等人， 2021]。已有多项研究探讨了使用变分自编码器生成合成物联网流量数据，以缓解入侵检测中的类别不平衡问题。例如，在 [Liu 等人， 2022], 的研究中，研究者评估了两种不同的变分自编码器变体，以平衡物联网流量数据集，进而提升入侵检测系统性能。同样，在 [Zhang 和 Liu， 2022], 的研究中，研究者提出了一种基于变分自编码器的方法来生成合成物联网攻击样本，并报告称数据增强后入侵检测系统性能有所提升。然而，众所周知，由于潜在空间正则化，变分自编码器生成的样本往往过度平滑，这可能会限制其合成高度不平衡且复杂的表格型物联网攻击数据的效果 [Pinheiro-Cinelli 等人， 2021]。

生成对抗网络是另一种基于深度学习的经典过采样技术替代方案，可通过生成器与判别器之间的对抗训练学习潜在数据分布，从而生成合成表格数据 [Wang 等人， 2017]。在物联网安全领域，生成对抗网络已被用于缓解入侵检测数据集中的类别不平衡。例如，在[Park 等人， 2023], 中，作者发现，在类别严重不平衡的情况下，机器学习模型难以学习有意义的攻击模式，因此使用生成对抗网络生成合成物联网流量样本。同样，在 [Rahman 等人， 2024], 中，作者使用生成对抗网络合成物联网攻击数据，并报告称入侵检测系统性能显著提升。然而，生成对抗网络众所周知存在训练

不稳定性、模式崩溃以及对架构和优化选择敏感等问题，这些问题可能会限制其鲁棒性 [Wang 等人，2017]。

近年来，扩散模型已成为一类强大的生成模型，克服了变分自编码器和生成对抗网络的若干局限[Wang 等人， 2024]。扩散模型通过学习去噪过程生成新数据，该过程逐步将纯噪声样本转换为从目标数据分布中抽取的样本 [Ho 等人， 2020]。在物联网安全领域，近期已有研究探索使用扩散模型进行合成攻击数据生成。例如，在 [Camerota 等人， 2024], 中，研究者提出了一种扩散模型来生成合成物联网攻击样本，并发现将生成数据纳入训练后，恶意软件检测性能有所提升。同样，在 [Wang 等人， 2025], 中，研究者提出了一种扩散变换器框架，先使用扩散模型平衡数据集，再通过基于变换器的分类器提升入侵检测系统性能。然而，扩散模型通常计算开销较大，且对特征异质性较为敏感，这可能限制其在大规模、高维物联网攻击数据集中的实用性 [Li 等，2025b]。

综述文献普遍得出的结论是，合成数据生成对物联网入侵检测至关重要，因为攻击类别通常代表性不足，导致入侵检测系统性能不佳；已有研究表明，加入合成样本后，系统性能可显著提升。在所研究的生成方法中，扩散模型是最先进的技术，并已展现出生成高质量物联网攻击数据的能力 [Camerota 等，2024，Wang 等， 2025]。然而，其计算成本高昂，且对混合类型特征以及统计尺度不同的变量较为敏感，因此仍有进一步改进的空间 [Wang 等， 2024]。为解决这些局限，研究者提出了潜在扩散模型 [Rombach 等人， 2022] ，该模型通过自编码器获得低维空间，并在其中学习扩散过程。潜在扩散模型已成功应用于其他领域的表格数据合成，例如金融数据 [Sattarov 等人， 2023] ，以及包括人口统计、住房和健康相关信息在内的敏感个人数据 [Zhu 等人， 2025], ，但其在物联网网络流量数据中的应用尚未得到探索。在这些领域中，潜在扩散模型已被证明能够与最先进的生成模型相媲美，这促使研究者探索其在物联网攻击数据生成中的应用。

3 方法

本节详细介绍所提出的合成物联网流量数据生成方法。首先介绍所研究的问题；其次，阐述能够处理混合类型物联网表格数据的自编码器设计，并描述潜在扩散过程。随后说明潜变量采样和解码过程，最后分析所提出框架的计算复杂度。

3.1 问题定义

设 $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ 表示一个物联网网络流量数据集。每个输入样本 $\mathbf{x}_i$ 由连续变量和类别变量组成，即 $\mathbf{x}_i = (\mathbf{x}_i^{(c)}, \mathbf{x}_i^{(b)})$，其中 $\mathbf{x}_i^{(c)} \in \mathbb{R}^{d_c}$ 由 d_c 个连续特征构成（例如数据包大小、到达间隔时间和统计摘要），而 $\mathbf{x}_i^{(b)} \in \{0, 1, \cdots, N_b\}^{d_b}$ 由 d_b 个二元或类别特征构成（例如协议标志和连接状态）。相应的标签记为 $y_i \in \mathcal{Y}$，表示良性流量或特定攻击类型。本文重点研究类别特定合成数据生成。

给定真实数据集 $\mathcal{D}_{real}$，本研究旨在开发一种生成模型，用于生成合成样本 $\hat{\mathbf{x}} \sim p_\theta(\mathbf{x}|y)$，使得到的增强数据集 $\mathcal{D}_{aug} = \mathcal{D}_{real} \cup \mathcal{D}_{syn}$ 能够保留真实物联网流量的统计结构、特征依赖关系和语义一致性。为此，生成的样本应符合原始特征的联合分布，与目标类别保持一致，并提升数据覆盖度，同时避免生成不真实的样本或落在真实数据流形之外的点。

由于物联网流量数据具有异质性、各特征的统计尺度存在差异且数据维度较高，直接在原始特征空间中对联合分布 $p(\mathbf{x})$ 建模颇具挑战 [Kotelnikov 等，2023，Wang 等人， 2024]。此外，在数据空间中运行的扩散模型需要执行迭代去噪过程，计算成本高，且对特征归一化较为敏感 [Nichol 和 Dhariwal，2021，Karras 等人， 2022]。为应对这些挑战，我们提出遵循潜在扩散范式，在学习得到的潜在空间中执行样本生成，并针对混合类型物联网流量数据的具体特性进行调整。首先通过自编码器编码器降低原始数据的维度，然后训练扩散模型以近似潜在分布；随后可从该分布中抽取新的潜在样本，并通过自编码器解码器将其解码回原始特征空间。在这一表述下，合成数据生成任务可简化为学习潜在空间中的基于扩散的生成过程，随后执行确定性解码步骤。

3.2 混合类型表格数据自编码器

为使扩散模型能够在更低维的空间中运行，必须先对原始数据进行降维。为此，我们采用自编码器。自编码器的作用是学习原始混合类型

表 1：所考虑攻击的主要特征概述。

攻击	类别	流量规模	可见性	主要影响
分布式拒绝服务攻击-*ICMP*_分片	流量洪泛	非常高	显眼	资源耗尽
*Mirai-greip*_洪泛	僵尸网络洪泛	高强度、分布式	中等可见度	带宽/CPU耗尽
MITM-ArpSpoofing	协议操纵	Low	隐蔽	流量拦截

物联网流量数据的紧凑且连续的潜在表示，使其适用于基于扩散的生成建模。自编码器编码器 $f_\phi(\cdot)$ 将输入向量 $\mathbf{x}_i$ 映射为潜在表示 $\mathbf{z}_i = f_\phi(\mathbf{x}_i) \in \mathbb{R}^{d_z}$，其中 $d_z \ll d_c + d_b$。编码器由包含全连接层和非线性激活函数的多层感知机实现，使模型能够捕捉异构特征之间的复杂交互。为确保潜在空间适合扩散过程，编码器的输出会经过归一化，使其在潜在维度上的均值为零、方差为一；这些统计量基于训练集计算。此归一化可缓解潜变量之间的尺度失衡，并提高后续扩散训练的稳定性。解码器 $g_\phi(\cdot)$ 将潜在表示映射回原始特征空间 $\hat{\mathbf{x}}_i = g_\phi(\mathbf{z}_i)$，并通过适当的输出参数化方式重构连续特征和离散特征。连续特征使用实值输出进行重构，而二元特征和类别特征则分别通过 Sigmoid函数和 Softmax函数进行重构。这种设计使自编码器能够在重构过程中明确考虑物联网流量数据的混合类型特性。

训练自编码器时，目标是最小化一种重构损失，该损失反映连续变量和离散变量不同的统计特性。具体而言，损失函数定义为

$$\mathcal{L}_{AE} = \mathcal{L}_c\big(\mathbf{x}^{(c)}, \hat{\mathbf{x}}^{(c)}\big) + \mathcal{L}_b\big(\mathbf{x}^{(b)}, \hat{\mathbf{x}}^{(b)}\big), \tag{1}$$

其中，$\mathcal{L}_c(\cdot)$ 表示连续特征的均方误差（MSE）损失，$\mathcal{L}_b(\cdot)$ 表示离散特征的二元或类别交叉熵损失。该公式确保不同类型特征得到均衡学习，并防止连续变量或离散变量占据主导地位。学习得到的潜在表示用作扩散模型的输入空间。通过将异构表格数据映射到紧凑、连续且归一化的潜在空间，自编码器简化了扩散模型的学习任务，并提高了训练稳定性。此外，将表示学习（自编码器）与生成建模（扩散模型）分离，使各组件能够专注于明确的目标，从而构建出更具可扩展性和鲁棒性的合成数据生成框架。

3.3 潜在扩散

混合类型的物联网流量数据经自编码器映射为紧凑的潜在表示后，采用扩散模型来学习潜变量的分布。令$\mathbf{z}_0 \in \mathbb{R}^{d_z}$ 表示由编码器 $f_\phi(\cdot)$ 得到的一般潜在表示。扩散模型的目标是通过逐步向潜在样本添加噪声，再学习逆转这一过程，从而学习潜在的数据分布 $p(\mathbf{z}_0)$。扩散过程分为两个阶段，即前向扩散过程和反向（生成）扩散过程。

前向扩散过程定义为一个固定的马尔可夫链，在T个时间步内逐步向潜变量添加高斯噪声。在每个时间步 $t \in \{1, 2, \cdots, T\}$，潜变量$\mathbf{z}_t$由$\mathbf{z}_{t-1}$得到，表达式如下：

$$\mathbf{z}_t = \sqrt{\alpha_t}\mathbf{z}_{t-1} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{2}$$

其中 $\{\alpha_t\}_{t=1}^T$ 是预先定义的噪声调度，且 $0 < \alpha_t < 1$。定义 $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$ 后，前向过程可写成闭式允许直接从$\mathbf{z}_t$采样的表达式直接由 $\mathbf{z}_0$, 得到，如下所示，

$$\mathbf{z}_t = \sqrt{\bar{\alpha}_t}\mathbf{z}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{3}$$

随着 t 增大，潜变量越来越受噪声支配；当 T足够大时，$\mathbf{z}_T$ 会趋近于高斯分布。

反向扩散过程（也称为生成过程）对应于学习反向马尔可夫链，该链逐步去除$\mathbf{z}_T$ 中的噪声，以恢复来自 $p(\mathbf{z}_0)$的样本。该反向过程由神经网络 $\boldsymbol{\epsilon}_\theta(\mathbf{z}_t, t)$参数化，该网络经过训练以预测在时间步 t添加的噪声。反向转移定义为

$$p_\theta(\mathbf{z}_{t-1}|\mathbf{z}_t) = \mathcal{N}\big(\mathbf{z}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{z}_t, t), \boldsymbol{\Sigma}_t\big), \tag{4}$$

其中，均值 $\boldsymbol{\mu}_\theta(\mathbf{z}_t, t)$定义为

$$\boldsymbol{\mu}_\theta(\mathbf{z}_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{z}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{z}_t, t)\right), \tag{5}$$

且 $\boldsymbol{\Sigma}_t$ 是由噪声调度确定的固定方差。

扩散模型通过最小化去噪目标进行训练，该目标旨在提升每个时间步 t的噪声预测准确性。具体而言，损失函数定义为

$$\mathcal{L}_{DM} = \mathbb{E}_{\mathbf{z}_0,\boldsymbol{\epsilon},t}\left[\left|\left|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{z}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t\right)\right|\right|_2^2\right], \tag{6}$$

其中， t从 $\{1, 2, \cdots, T\}$中均匀采样。该公式使扩散模型能够利用直接从前向过程中抽取的样本，通过随机梯度下降高效训练。

3.4 潜在空间采样与解码

合成物联网流量样本通过从学习到的反向扩散过程中采样生成，随后将得到的潜在表示解码回原始特征空间。首先对潜变量$\mathbf{z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$进行采样，然后根据学习到的反向转移$\mathbf{z}_{t-1} \sim p_\theta(\mathbf{z}_{t-1}|\mathbf{z}_t)$，经过 T个时间步迭代去噪，其中 $t = T, T-1, \cdots, \mathbf{1}$。完成反向扩散过程后，得到合成潜在表示$\hat{}\mathbf{z}_0$，其符合学习到的潜在数据分布 $p_\theta(\mathbf{z}_0)$。这一迭代去噪过程使扩散模型能够生成多样的潜在样本，同时保留自编码器所学习到的潜在空间的全局结构。随后，使用解码器将合成潜在表示$\hat{}\mathbf{z}_0$ 映射回原始特征空间，得到$\hat{}\mathbf{x} = g_\phi(\hat{}\mathbf{z}_0)$，从而生成合成样本$\hat{}\mathbf{x} = (\hat{}\mathbf{x}^{(c)}, \hat{}\mathbf{x}^{(b)})$，其中$\hat{}\mathbf{x}^{(c)}$ 和$\hat{}\mathbf{x}^{(b)}$ 分别表示重构的连续特征和离散特征。

3.5 计算复杂度分析

令 N表示训练样本数， $d = d_c + d_b$ 表示原始特征空间的维度， d_z 表示潜在空间的维度， T表示扩散时间步数。自编码器仅训练一次，计算成本为 $\mathcal{O}(N \cdot C_{AE})$，其中 C_{AE} 表示通过编码器-解码器网络进行一次前向-反向传播的成本。在推理过程中，每个样本仅编码和解码一次，而扩散采样需要多个连续步骤，因此与之相比，编码和解码的计算成本可以忽略不计。因此，主要计算成本来自扩散模型的训练和采样。对于数据空间扩散模型，该成本随原始特征维度 d 增长，复杂度为 $\mathcal{O}(T \cdot C_{DM}(d))$。相比之下，本文提出的方法在潜在空间中运行，因此复杂度降低至 $\mathcal{O}(T \cdot C_{DM}(d_z))$。由于 $d_z \ll d$，训练和采样成本均大幅降低。因此，尽管所提出的框架增加了一步自编码器训练，但该步骤只需执行一次，而在潜在空间中进行扩散所带来的效率和可扩展性显著提升，足以抵消这项成本。

4 实验设置

本节介绍用于评估所提方法的实验设置。首先，详细说明实验所用的数据集以及对输入数据实施的预处理步骤。然后，介绍所有深度学习模型采用的超参数优化策略。接着，阐述用于与潜在扩散模型进行比较的基线模型的实现。最后，报告潜在扩散模型的架构和配置详情。

4.1 数据集

本研究使用了由加拿大网络安全研究所、新不伦瑞克大学发布的 CICIoT2023 数据集 [Pinto- Neto et al., 2023], 。该数据集记录了在受控环境中由多种物联网设备产生的真实物联网网络流量，其中既包括良性流量，也涵盖了多种攻击场景。本研究基于以 CSV 格式提供的预提取流级特征开展分析，这些特征专为机器学习评估设计。有关这些特征的详细说明，请参阅 [Pinto-Neto et al., 2023]。

我们重点关注三种攻击类型，即 *DDoS-ICMP_*分片、*Mirai-greip_*洪泛和 *MITM-ArpSpoofing*。选择这些攻击是因为它们代表不同的攻击家族，并通过不同机制运作，从而在物联网环境中产生不同的流量模式。*DDoS-ICMP_*分片是一种 DDoS攻击，它通过向目标传输大量分片 ICMP 数据包来滥用互联网控制消息协议（ICMP）。数据包分片需要重新组装，这会显著增加处理和内存需求，并可能导致资源耗尽 [Sonar 和 Upadhyay, 2014]。*Mirai-greip_*洪泛 属于 Mirai 僵尸网络家族，该家族是最著名的针对物联网设备的恶意软件家族之一。在此变体中，遭入侵的物联网设备充当僵尸网络节点，向目标产生高强度洪泛流量，导致带宽饱和和中央处理器（CPU）崩溃 [Abbas 等, 2021]。最后，*MITM-ArpSpoofing* 是一种中间人攻击，利用地址解析协议中的漏洞，在局域网内拦截或篡改流量。这种攻击通常速率较低且隐蔽 [Fereidouni 等, 2025]。表 1 总结了所考察攻击类型的主要特征。

表2：各生成模型的攻击特定超参数。

模型超参数		DDoS	Mirai	MitM
VAE	潜在维度	5	11	14
	学习率（$\cdot 10^{-3}$）	0.95	1.72	1.40
	KL权重（$\cdot 10^{-2}$）	9.77	1.31	1.01
	梯度裁剪	4.22	4.71	0.06
GAN	潜在噪声维度	13	15	12
	生成器学习率（$\cdot 10^{-4}$）	4.53	1.41	2.03
	判别器学习率（$\cdot 10^{-4}$）	4.90	0.98	1.16
	真实标签平滑	0.80	0.80	0.91
DM	扩散步数	200	200	200
	网络深度	5	8	8
	学习率（$\cdot 10^{-4}$）	3.36	0.61	3.00
LDM	自编码器潜在维度	13	16	11
	自编码器学习率（$\cdot 10^{-3}$）	2.22	0.46	2.66
	扩散步数	782	461	200
	扩散网络深度	2	5	5
	扩散学习率（$\cdot 10^{-4}$）	2.31	0.89	0.52

4.2 预处理

我们创建了一个数据集，其中包含良性流量样本以及对应于所考察三种攻击的样本。首先，移除了数据集中的常量特征，并将其余特征分为连续数值变量和离散变量。异常值检测仅应用于连续特征。为避免破坏原始数据分布，异常值移除过程经过专门设计，确保丢弃的数据少于总数据量的 1%。具体而言，对于每个数值特征，首先将数值归一化到 $[0, 1]$ 范围并排序。随后，该方法识别相邻有序数值之间的间隔，并依据预定义阈值判断间隔是否足够大。根据此标准，计算每个特征的下界和上界。如果某个数据点至少有一个数值特征值超出相应界限，则将其归类为异常值。移除异常值后，将攻击样本与良性流量分开，因为只有攻击实例用于训练生成模型。随后，将攻击数据划分为训练集（80%）、验证集（10%）和测试集（10%）。训练集和验证集分别用于生成模型的训练和验证，而测试集则专门用于评估生成的合成样本。从此开始，分别独立处理每种攻击的样本。对于每种攻击，首先使用对数函数变换训练集和验证集中的数值特征，以缓解重尾分布问题；然后使用相应训练集的均值和标准差对其进行标准化。离散特征采用标签编码，并在整个建模过程中作为离散变量处理。

4.3 超参数优化

为确保比较公平且可复现，所有基于深度学习的生成模型的超参数均使用 MATLAB® 中的贝叶斯优化进行选择，每个模型的目标函数评估次数固定为 50 次。对于所有模型，均在验证集上监测性能，目标函数则采用各模型使用的标准形式。针对每种攻击场景分别独立优化超参数，因此得到的特定于攻击的模型配置如表2所示。

4.4 基线生成模型

为评估潜在扩散模型的有效性，我们将其与多种数据增强和生成建模方法进行比较，具体包括 SMOTE、变分自编码器、生成对抗网络和扩散模型。基线模型仅使用攻击样本进行训练，并配置为处理连续特征与离散特征的混合数据，其损失函数由连续损失分量和离散损失分量组合而成。

表3：基线生成模型的配置。

模型	超参数	值
SMOTE	邻居数量 (k)	5
VAE	编码器隐藏层宽度	[128 64]
	解码器隐藏层宽度	[64 128]
	激活函数	ReLU
	小批量大小	128
	最大训练轮数	300
	优化器	Adam
GAN	生成器隐藏层宽度	[128 128]
	判别器隐藏层宽度	[128 128]
	激活函数	ReLU
	小批量大小	128
	最大训练轮数	300
	优化器	Adam
DM	扩散调度	线性
	$[\beta_{start}, \beta_{end}]$	$[10^{-4}, 0.02]$
	隐藏层宽度	256
	时间嵌入维度	128
	小批量大小	128
	最大训练轮数	400
	梯度裁剪	1
	优化器	Adam

表 4：不同分类器在各种攻击类型下的行归一化混淆矩阵（%），其中以粗体标出最佳结果，以斜体标出次佳结果。每个混淆矩阵中，行对应真实类别（良性、攻击），列对应预测类别（良性、攻击）。

攻击	基线	SMOTE	VAE	GAN	DM	LDM
DDoS	$\begin{bmatrix} 100.0 & 0.0 \\ 69.9 & 30.1 \end{bmatrix}$	$\begin{bmatrix} 100.0 & 0.0 \\ 19.4 & 80.6 \end{bmatrix}$	$\begin{bmatrix} 100.0 & 0.0 \\ 7.5 & 92.5 \end{bmatrix}$	$\begin{bmatrix} 100.0 & 0.0 \\ 11.8 & 88.2 \end{bmatrix}$	$\begin{bmatrix} \textit{100.0} & \textit{93.0} \\ \textit{6.5} & \textit{5} \end{bmatrix}$	$\begin{bmatrix} \mathbf{100.0} & \mathbf{98.0} \\ \mathbf{1.1} & \mathbf{9} \end{bmatrix}$
Mirai	$\begin{bmatrix} 100.0 & 0.0 \\ 90.5 & 9.5 \end{bmatrix}$	$\begin{bmatrix} 100.0 & 0.0 \\ 51.0 & 49.0 \end{bmatrix}$	$\begin{bmatrix} 100.0 & 0.0 \\ 25.2 & 74.8 \end{bmatrix}$	$\begin{bmatrix} 100.0 & 0.0 \\ 28.6 & 71.4 \end{bmatrix}$	$\begin{bmatrix} \textit{100.0} & \textit{85.7} \\ \textit{14.3} & \end{bmatrix}$	$\begin{bmatrix} \mathbf{99.6} & \mathbf{100.} \\ \mathbf{0.0} & \mathbf{0} \end{bmatrix}$
MitM	$\begin{bmatrix} 100.0 & 0.0 \\ 100.0 & 0.0 \end{bmatrix}$	$\begin{bmatrix} 99.7 & 0.3 \\ 78.3 & 21.7 \end{bmatrix}$	$\begin{bmatrix} 39.3 & 60.7 \\ 10.0 & 90.0 \end{bmatrix}$	$\begin{bmatrix} 35.5 & 64.5 \\ 8.3 & 91.7 \end{bmatrix}$	$\begin{bmatrix} \mathbf{81.5} & \mathbf{18.5} \\ \mathbf{26.7} & \mathbf{3} \end{bmatrix}$	$\begin{bmatrix} \textit{75.7} & \textit{24.3} \\ \textit{31.7} & \textit{3} \end{bmatrix}$

作为非深度学习基线，我们采用了一种基于 SMOTE 的单类数据增强方法。通过在随机选取的攻击样本及其 k 个最近邻之间进行插值，生成合成样本；最近邻依据混合距离度量计算，该度量结合了连续变量的欧几里得距离和离散变量的汉明距离。连续特征通过线性插值合成，离散特征则采用伯努利采样或类别采样策略生成。我们使用变分自编码器作为潜变量生成基线。变分自编码器学习攻击流量的概率潜在表示，并通过解码从标准正态先验中抽取的潜变量来生成合成样本。我们采用生成对抗网络作为基线生成模型，并使用对抗训练框架。生成器将潜在噪声向量映射为合成攻击样本，而判别器则经过训练以区分真实数据和生成数据。连续特征由生成器直接生成，离散特征则通过 Sigmoid 激活后进行伯努利采样或类别采样生成。我们还使用了一个直接在数据空间中运行的扩散模型，作为基于扩散的基线。该模型通过在预先设定的扩散步数内逐步向真实数据添加高斯噪声来学习攻击样本的分布，并训练神经网络通过预测所添加的噪声来逆转这一过程。基线模型的具体配置见表3。

表 5：各个分类器的 F1分数 针对每种攻击类型，最佳（粗体）和次佳（斜体）结果已突出显示。

攻击	基线	SMOTE	变分自编码器	生成对抗网络	扩散模型	LDM
DDoS	0.46	0.89	0.96	0.94	*0.97*	**0.99**
Mirai	0.17	0.66	0.86	0.83	*0.92*	**0.99**
MitM	0.00	0.35	0.34	0.33	**0.53**	*0.44*

4.5 潜在扩散模型配置详情

本小节详细介绍了实验评估中所提出的潜在扩散模型（LDM）所采用的具体架构选择、训练设置和实现细节。所提出的潜在扩散模型中使用的确定性自编码器采用多层感知机（MLP）实现，并为数值特征和类别特征分别设置了处理分支。编码器由两个并行分支组成：数值分支包含宽度分别为 64 和 32 的全连接层，类别分支包含宽度分别为 32 和 16 的层，二者均采用 ReLU 激活函数。这两个分支的输出经过拼接后，输入一个包含 32 个单元的共享全连接层，随后线性投影至潜在空间。解码器采用镜像结构，由包含两个全连接层的共享主干组成，层宽分别为 32 和 64，后接两个输出头。数值输出头生成连续特征的实数重构值；类别输出头则分别通过 Sigmoid 函数或 Softmax 函数输出伯努利概率或多项式概率。自编码器使用重构损失进行训练，该损失由数值变量的均方误差项和类别变量的类别交叉熵项组成。训练使用 Adam优化器，批次大小为 128，共训练 200 个周期。自编码器训练完成后，将训练集中所有攻击样本编码为潜在表示。为确保扩散训练稳定，使用根据训练集计算出的均值和标准差对每个潜在维度进行归一化。采样时沿用这些统计量，在解码前对生成的潜在向量进行反归一化。

扩散模型仅在归一化潜在空间中运行。采用线性方差调度，其中 β_{start} 设为 10^{-4}，β_{end} 设为 0。02。反向去噪网络采用时间条件多层感知机实现。维度为 128 的正弦时间嵌入经由一个两层多层感知机处理后，与含噪潜在输入拼接。所得表示随后通过一个深度经过优化、隐藏层宽度为 256 的主干网络，再经由线性输出层预测干净的潜在表示。扩散模型使用预测潜在向量与真实潜在向量之间的均方误差损失进行训练。优化采用带有梯度裁剪的 Adam优化器，训练共进行 400 个周期。随后，通过从潜在空间中的高斯噪声初始化反向扩散过程，并迭代应用学习到的去噪转移，生成合成样本。所得潜在样本经过反归一化，并通过自编码器解码器解码，以获得合成样本。所有生成的样本均对应于训练中使用的攻击类别。

5 结果

本节全面评估所提出的潜在扩散模型在合成物联网攻击数据生成方面的表现，并将其与最先进的方法进行比较。评估旨在考察数据增强对入侵检测系统性能的实际影响，以及生成样本的质量、保真度和效率。

首先，构建了一个二分类入侵检测系统（良性与攻击），以评估扩充少数类攻击数据是否能够提升分类性能。用于评估的数据中，分布式拒绝服务攻击、Mirai 和中间人攻击这几种攻击类型的攻击流量与良性流量之比分别约为 1:5.30、1:5.57 和 1:5.63；这些数据被划分为训练集（80%）和测试集（20%）。检测模型采用集成分类器，使用子空间方法训练，共进行 100 个学习周期。为评估合成数据增强的影响，将各生成模型生成的合成攻击样本加入基线训练数据，直至实现类别平衡，从而构建增强训练集。对于每种增强策略，均采用相同配置从头重新训练分类器。所有模型均在同一不平衡测试集上进行评估，性能分别以按行归一化的混淆矩阵（CM）和 F1 分数汇总于表 4 和表 5 中。

表 4 显示，在原始不平衡数据集上训练分类器，会导致其对所有攻击类型的攻击检测性能严重下降。在基线设置下，分类器严重偏向多数类，虽然能够完美识别良性流量，但对攻击的召回率却很低。这种现象与极端的类别不平衡相符，也凸显了在少数类样本稀缺时，分类器无法学习相关攻击模式。当通过添加合成攻击样本使训练集达到平衡后，所有增强策略的攻击检测性能均显著提升，这一点既体现在表 4 的混淆矩阵中，也反映在表 5 报告的相应 F1分数中。在所评估的增强方法中，SMOTE 在所有攻击类型上的性能提升始终最小。尽管与基线相比，其召回率有所提高，但性能仍明显低于基于深度学习的模型。这可能是因为 SMOTE 通过在相邻点之间进行线性插值来生成新样本，生成的合成实例往往无法为攻击类别带来足够的多样性。相比之下，基于深度学习的模型生成的合成样本能够带来更高的检测

表6：针对每种攻击类型生成的合成样本的定量评估。

指标	DDoS					Mirai					MitM				
	SMOTE	变分自编码器	生成对抗网络	扩散模型	潜在扩散模型	SMOTE	变分自编码器	生成对抗网络	扩散模型	潜在扩散模型	SMOTE	变分自编码器	生成对抗网络	扩散模型	潜在扩散模型
MMD	**0.00**	0.10	*0.01*	0.05	0.06	**0.00**	0.16	*0.01*	0.08	0.08	**0.00**	0.51	*0.01*	0.02	0.02
平均KL散度	**0.03**	0.65	*0.53*	0.78	0.70	**0.02**	1.13	*0.48*	0.55	0.61	**0.04**	1.18	*0.55*	0.62	0.67
互信息误差	**1.19**	7.59	*3.86*	4.66	4.67	**0.95**	10.50	*4.91*	5.24	5.01	**1.39**	9.16	5.16	4.17	*4.05*
精确率	*0.99*	0.84	**1.00**	**1.00**	*0.99*	**1.00**	0.87	0.89	0.82	*0.99*	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
召回率	**1.00**	0.84	**1.00**	0.98	*0.99*	**1.00**	0.48	*0.99*	0.69	0.75	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
N_{med}	0.47	**2.85**	0.62	0.83	*0.88*	0.28	**3.78**	0.82	0.83	*1.96*	1.49	**3.23**	1.98	1.66	*2.33*

性能。对于分布式拒绝服务攻击和 Mirai攻击，潜在扩散模型取得了最佳结果，两种情况下的 F1分数均达到 0.99。这表明，潜在扩散能够有效扩充攻击样本空间，生成真实且多样的实例，从而增强分类器泛化能力。直接在数据空间中运行的扩散模型在这些攻击上的表现与潜在扩散模型接近，并在最具挑战性的中间人攻击场景中取得最佳结果，F1分数为 0.53。考虑到基线分类器在该场景下完全失效，这一提升十分显著。潜在扩散模型在中间人攻击上的表现也颇具竞争力，F1分数达到 0.44。变分自编码器和生成对抗网络的表现优于 SMOTE，但仍逊于基于扩散的方法。尽管这两种方法都提高了 F1分数，但其合成样本在捕捉复杂攻击模式方面的效果似乎较差，尤其是对于中间人攻击流量。这些结果表明，所提出的潜在扩散模型非常适用于合成物联网攻击数据生成，因为它不仅优于经典过采样技术，而且与最先进的基于深度学习的生成模型相比，也能取得具有竞争力甚至更优的性能。

为了定量评估LDM生成的合成攻击样本的质量，并将其与其他数据增强技术进行比较，我们评估了分布相似性、样本多样性以及潜在的记忆化效应。具体而言，我们计算了真实攻击样本与各模型生成的合成数据之间的最大均值差异（MMD）、边际平均库尔贝克-莱布勒散度（KL）、互信息（MI）误差、生成精确率和召回率，以及最近邻距离的中位数（N_{med}）。这些指标在三种不同的增强比例下进行评估，即相对于可用真实攻击样本数量的50%、100%和200%。表6报告了增强比例为100%时的结果，因为所有比例下的指标值均相近，这表明这些定量指标主要受生成样本质量的影响，而非生成的合成实例数量。

表6显示，没有任何单一技术在所有指标和攻击类型上均占据优势，凸显了分布相似性、样本多样性与记忆化之间存在的权衡。SMOTE始终取得最低的MMD和边际平均KL值，表明真实样本与合成样本的经验分布几乎完全匹配。其互信息误差也最低，说明特征依赖关系得到了较好的保留。然而，SMOTE在所有攻击类型上的 N_{med} 也最低，这表明其存在显著的记忆化风险。这种表现符合预期，因为SMOTE通过在相邻真实实例之间进行线性插值来生成新样本，使生成点仍然非常接近原始数据流形，并且只能提供有限的新颖信息。GAN增强表现出类似但不太明显的趋势。GAN取得了相对较低的MMD和KL值，表明其分布对齐效果良好；然而，其 N_{med} 值仍然较小，尤其是在分布式拒绝服务攻击和Mirai中，这表明生成样本仍集中在真实数据附近，可能无法扩展攻击空间。这种表现与生成对抗网络在建模复杂或碎片化分布时常见的模式集中效应一致。相比之下，VAE的MMD和平均KL值最大，表明真实分布与合成分布之间的匹配程度较低；同时，其互信息误差也最高，说明特征依赖关系的保留程度较差。然而，VAE在所有攻击类型上的 N_{med} 值最大，表明其记忆化程度最低，并且具有较强的生成新颖样本倾向。这反映了变分自编码器的平滑效应：样本从连续潜在分布中采样，可能会牺牲细粒度结构以换取多样性。基于扩散的模型在这些相互冲突的表现之间实现了更有利的平衡。两种方法均取得较低的MMD和KL值，同时保持适中的互信息误差，表明它们能够较好地保留边际分布和特征依赖关系。就 N_{med}而言，在所有攻击类型上，LDM始终取得高于DM的值，表明潜在空间扩散在保持分布保真度的同时，进一步提升了多样性。所有方法的生成精确率和召回率通常都较高，这证实大多数方法生成的样本均位于真实数据分布的支撑集内，并能够提供充分的覆盖。结果表明，LDM在真实性、多样性和鲁棒性之间实现了良好的折中。经典过采样方法在分布匹配方面表现出色，但容易产生记忆化；变分自编码器则以牺牲保真度为代价追求新颖性。然而，LDM能够有效平衡这些相互竞争的目标。

为定性评估合成样本对真实攻击几何结构的影响，我们在三维嵌入空间中分析了真实样本和合成样本的统一流形近似与投影（UMAP）投影。针对每种攻击类型和增强方法，我们构建了一个包含真实攻击样本及其对应合成样本的组合数据集，并训练了一个UMAP模型，以可视化它们的联合流形结构。代表性嵌入

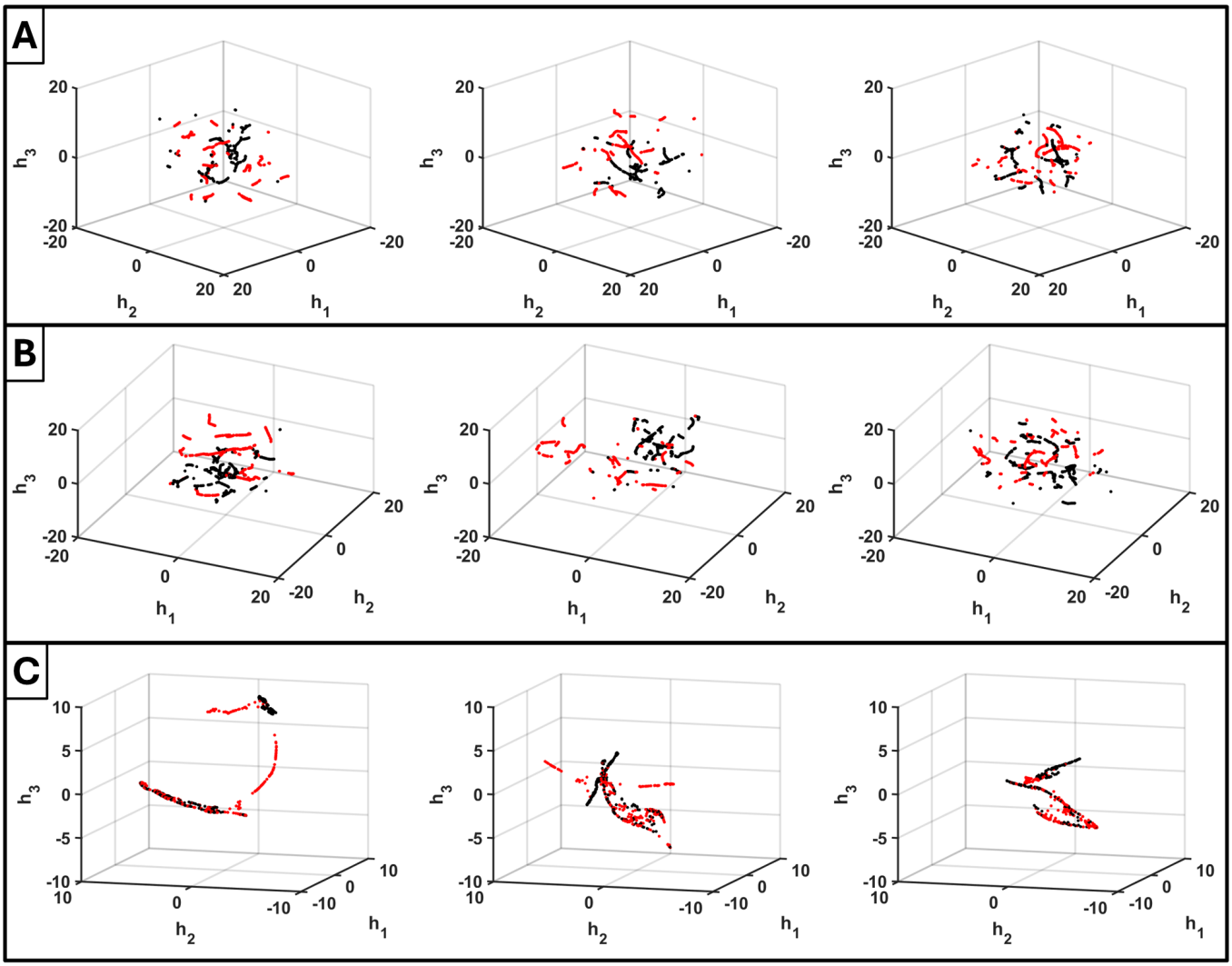


图 1：由 GAN（第一列）、DM（第二列）和 LDM（第三列）生成的 DDoS（面板 A）、Mirai（面板 B）和中间人攻击（面板 C）攻击的真实样本（黑色）与合成样本（红色）的 UMAP 投影。

如图1所示，其中分别展示了分布式拒绝服务攻击（面板A）、Mirai（面板B）和中间人攻击（面板C）这几种攻击类型，并重点呈现了使用生成对抗网络（第一列）、扩散模型（第二列）和潜在扩散模型（第三列）获得的流形。不出所料，基于SMOTE的流形中，合成样本与真实数据距离很近，往往几乎重叠。这一现象反映了SMOTE基于插值的特性，也与表6中观察到的较低 N_{med} 值一致，说明其多样性有限且记忆化程度较高。另一方面，变分自编码器生成的流形与生成对抗网络生成的流形表现相似，如图1所示。

对于 DDoS 攻击类型（图 1，A 面板），GAN 和 DM 生成的样本看起来比 LDM 生成的样本更稀疏。GAN 和 DM 在真实数据周围引入了额外的变异性，但有若干合成样本位于真实流形的边界上，表明其覆盖不完整；而对于 DM，嵌入的上方和左侧区域尤其存在偏离流形的样本。相比之下，LDM 生成的样本有效填补了流形中的空隙，且不会坍缩到现有点上，这表明潜在扩散能够更准确地捕捉攻击空间的局部几何结构。对于 Mirai 攻击（图 1，B 面板），GAN 和 DM 的偏离流形现象更加明显。这些样本增加了多样性，但可能会改变真实数据分布的内在结构。相比之下，LDM 生成的合成样本始终位于真实流形边界内，同时避免与真实点重叠，表明它在保真度和多样性之间实现了良好平衡。MitM 攻击（图 1，C 面板）是最具挑战性的场景，因为其真实数据流形呈碎片化分布。在这种情况下，GAN 生成的样本试图连接流形中彼此断开的区域，从而改变原有的几何结构。DM 也呈现出类似倾向，若干合成样本偏离了真实簇。相比之下，LDM 生成的样本分布于真实流形的各个碎片区域，强化了已有结构。

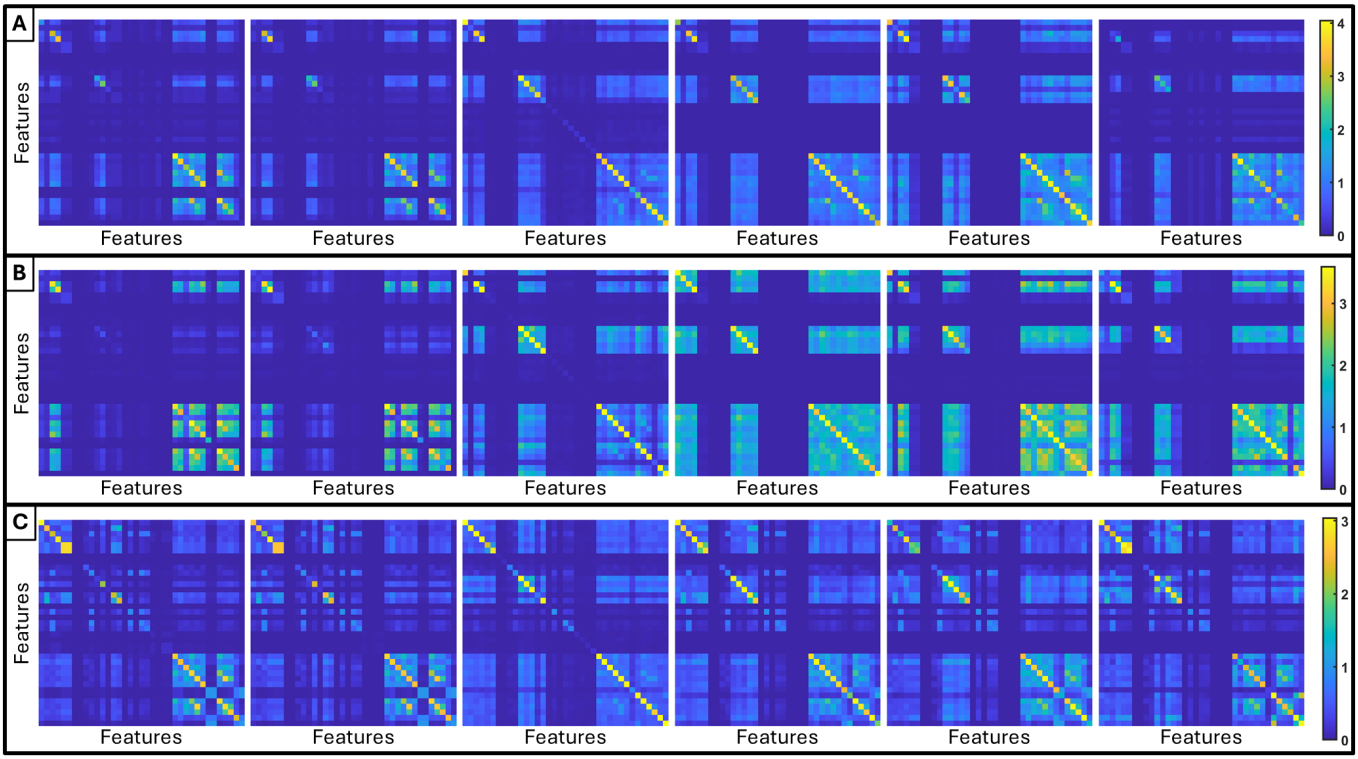


图2：分布式拒绝服务攻击（A 面板）、Mirai（B 面板）和中间人攻击（C 面板）的攻击特征互信息矩阵。从左到右，各列分别展示了根据完整真实攻击数据集计算的互信息矩阵，以及根据一个包含20个真实攻击样本、并利用 SMOTE、变分自编码器、生成对抗网络、扩散模型和潜在扩散模型生成的合成数据进行增强后的精简子集计算的互信息矩阵。颜色深浅反映成对特征依赖关系的强弱。

为进一步分析不同增强方法是否保留了特征间的统计依赖关系，我们针对每种攻击类型和生成方法计算了互信息矩阵。对于每种攻击（图 2 的 A、B、C 面板分别对应 DDoS、Mirai 和 MitM 攻击），首先计算完整真实攻击数据集的互信息矩阵（图 2 第一列）。随后，从中随机选取 20 个真实攻击样本作为缩减子集，并使用 SMOTE、VAE、GAN、DM 和 LDM 生成的合成样本对其进行增强，直至达到原始攻击数据集的规模。根据这些增强数据集计算得到的互信息矩阵显示在图 2 的第 2 至第 6 列中。

正如预期的那样，使用 SMOTE 增强精简的真实攻击子集后，所有攻击类型的互信息矩阵几乎都与根据完整真实攻击数据集计算的矩阵完全一致。这一现象与表6中报告的极低互信息误差值相符。由于 SMOTE 通过对现有实例进行线性插值来生成合成样本，因此它在很大程度上保留了原始特征关系，但引入的多样性有限。相比之下，深度生成模型引入了更大的变异性，这反映在互信息矩阵的差异增大上。对于所有攻击类型，使用生成对抗网络和扩散模型增强的数据集与真实互信息矩阵之间的差异都更为明显，尤其表现为局部依赖模式减弱或失真。在这两种方法中，扩散模型通常比生成对抗网络更好地保留特征依赖关系，但仍有部分区域的相关性减弱或发生偏移。基于变分自编码器的增强方法，在若干区域生成的互信息矩阵比生成对抗网络和扩散模型更接近真实攻击依赖关系。具体而言，矩阵中央特征块和左上区域中可见的较精细依赖结构得到了部分保留，而在生成对抗网络和扩散模型的结果中，这些结构则缺失或变得模糊。在所有方法中，潜在扩散模型生成的互信息矩阵始终最能兼顾保真度与多样性。对于所有攻击类型，潜在扩散模型都保留了整体依赖结构，同时保留了其他深度学习方法会丢失或扭曲的细粒度特征。这些定性证据支持了定量结果，并凸显了潜在扩散在生成多样化合成样本的同时保留有意义特征依赖关系的能力。

最后，通过测量生成合成样本所需的时间，评估了不同生成方法的计算效率。由于数据增强通常在线下进行，但可能涉及大量样本，因此采样效率仍是一个重要的实际考量。针对 $10^2, 10^3,$ 和 10^4 个合成样本测量了生成时间，每项实验均重复五次。表 7 报告了采样时间的均值和标准差。正如预期，GAN 在所有样本规模下都最快，即使生成 10^4 个样本也只需几毫秒。这反映出 GAN 训练完成后，采样只需对生成器网络进行一次前向传播。VAE 也

表 7：各方法在样本数量递增时生成合成样本所需的时间（均值 ± 标准差），其中最快（粗体）和第二快（斜体）的方法已标出。

	10^2 个样本	10^3 个样本	10^4 个样本
SMOTE	137.17 ± 3.12 毫秒	139.45 ± 5.49 毫秒	154.07 ± 2.86 毫秒
VAE	*5.97 ± 1.56* 毫秒	*9.19 ± 1.97* 毫秒	*54.85 ± 9.44* 毫秒
GAN	**2.62 ± 0.67 毫秒**	**4.00 ± 0.78 毫秒**	**4.85 ± 0.87 毫秒**
DM	487.16 ± 6.59 毫秒	496.01 ± 3.08 毫秒	553.86 ± 9.31 毫秒
LDM	213.24 ± 4.45 毫秒	259.06 ± 7.81 毫秒	411.23 ± 8.34 毫秒

表现出较短的采样时间，不过随着生成样本数量增加，由于随机解码和采样操作，其开销的增长更为明显。这些结果证实，依赖直接解码的潜变量模型在推理时效率很高。SMOTE 在所有样本规模下的生成时间较长，且相对稳定。这是因为其计算开销主要来自原始数据集中的最近邻搜索，而非生成的合成样本数量。基于扩散的模型计算开销最大。扩散模型的采样时间最长，这反映出它需要在原始高维特征空间中执行大量迭代去噪步骤。相比之下，对于所有样本规模，潜在扩散模型都大幅降低了相较于扩散模型的采样开销。对于 10^4 个合成样本，潜在扩散模型的生成时间约减少 25%，这凸显了在较低维空间中执行扩散所带来的效率提升，同时保留了基于扩散的生成优势。

6 讨论

我们的结果证实，物联网流量数据集中固有的极端类别不平衡，显著限制了有监督入侵检测系统学习有意义攻击模式的能力。当分类器使用原始的不平衡数据进行训练时，良性流量占据压倒性多数，使分类器受到主导，并倾向于收敛到偏向良性预测的决策规则，因而良性流量识别率较高，但攻击检测能力有限。这是少数类表示不足的常见后果：攻击样本稀缺，导致分类器无法学习定义清晰的判别性决策边界。通过合成攻击数据增强来平衡训练集后，学习动态发生了变化。攻击样本得到充分表示，能够对训练过程产生实质性影响，使决策边界不再偏向良性多数类，并更好地捕捉攻击类别的结构。因此，在所有评估的攻击类型中，攻击检测性能均得到持续提升。不过，提升幅度并不一致，而是取决于攻击本身的复杂程度。与良性流量差异更大的攻击类型，例如表现出明显泛洪行为的分布式拒绝服务攻击和 Mirai，从数据增强中获益最大。相比之下，流量模式与良性活动更相似的攻击（如中间人攻击），即便在平衡数据之后，检测起来仍然更具挑战性。这一观察表明，数据增强显著缓解了微妙且复杂的攻击行为所带来的困难，但并未将其完全消除。

除了平衡训练集所带来的一般性益处外，结果还揭示了不同增强策略在丰富少数类攻击类别方面的效果存在明显差异。使用 SMOTE 进行经典过采样能够持续保留边际数据统计量和成对特征依赖关系，这一点体现在其极低的 MMD 值、平均KL散度和互信息误差上。然而，由于 SMOTE 依赖在相邻样本之间进行线性插值，因此引入的多样性有限，并表现出类似记忆化的行为，这一点由最小的 N_{med} 值所表明。因此，尽管 SMOTE 的分布相似性指标表现出色，但它对入侵检测系统性能的提升有限，因为它基本上只是复制现有攻击模式，而没有扩展分类器可利用的有效决策空间。与之相反，变分自编码器生成的样本多样性很高， N_{med} 值较大便证明了这一点，但它难以保留真实的数据分布和特征依赖关系，尤其是对于 MitM 等更复杂的攻击。这一表现与观测到的较高 MMD、KL 散度和互信息误差值一致，可能是由于高斯潜在先验，以及变分自编码器倾向于过度平滑多模态或碎片化数据流形所致。生成对抗网络的表现介于两者之间，提升了样本的真实性，且 MMD 和 KL 散度低于变分自编码器。然而，生成对抗网络的对抗训练动态可能导致数据流形覆盖不均，并偶尔发生模式崩溃；这表现为低维投影中的流形外样本以及互信息结构失真。直接在数据空间中运行的扩散模型通过迭代去噪显式地对数据分布进行建模，克服了其中若干局限，从而生成高保真样本，并显著提升分类性能。然而，这也增加了计算复杂度：在高维特征空间中所需的大量扩散步数导致采样缓慢，而且仍偶尔会出现偏离流形的情况。所提出的潜在扩散模型有效结合了基于扩散的生成方法的优势，以及学习所得潜在表示所带来的效率和正则化效果。通过在紧凑且条件良好的潜在空间中执行扩散，潜在扩散模型不仅在真实性、覆盖范围、多样性和依赖关系保留之间实现了良好平衡，而且与数据空间扩散相比，还显著缩短了采样时间。

这种效率提升与生成样本的质量相结合，在定量指标、流形可视化、互信息分析以及下游入侵检测系统性能方面均得到了一致体现。

分布式拒绝服务攻击和 Mirai攻击呈现出相对有序且同质的流量模式，因此更易于建模，也更容易与良性流量区分。分布式拒绝服务攻击流通常具有一致的数据包级统计特征，例如较高的请求速率、重复出现的通信模式以及有限的协议多样性，因此会在特征空间中形成密集且连接良好的流形。类似地，Mirai攻击虽然比分布式拒绝服务攻击更多样，但其驱动因素是自动化的僵尸网络行为，会产生反复出现的扫描与漏洞利用特征，因此攻击流形仍相对紧凑且连续。这些特性有利于生成建模和下游分类。相比之下，中间人攻击的情况要复杂得多。中间人攻击流量本质上稀疏、异质且高度依赖上下文，往往与良性通信十分相似，只在数据包时序、路由行为或协议语义方面略有变化。因此，相关特征分布呈碎片化且不连续的状态，在潜在空间中形成少数彼此孤立的区域，而非单一簇。基于扩散的模型更适合捕捉这类复杂的多模态分布，但在建模中间人攻击流量时也存在局限。扩散模型在中间人攻击检测上的 F1分数最高，优于本文提出的潜在扩散模型，这表明在特征空间中直接进行扩散有助于提升这一困难场景下的分类性能。然而，UMAP投影显示，数据空间扩散偶尔会生成落在真实攻击流形之外的样本，说明其可能与真实数据分布存在几何不一致。相比之下，潜在扩散模型生成的合成样本更连贯地重构了中间人攻击流形的碎片化结构，使样本分布于彼此断开的子流形上，同时总体上与真实数据的几何结构保持一致。这凸显了提升检测性能与保留复杂攻击分布内在结构之间的权衡。

尽管本研究取得了令人鼓舞的结果，仍需指出若干局限性。首先，所有实验均基于单一物联网攻击数据集开展，并聚焦于三种具有代表性的攻击类型。尽管这些攻击在统计特征和几何特征方面各不相同，但所提方法在其他数据集、网络环境和攻击场景中的泛化能力仍有待系统评估。其次，实验分析仅限于二分类设置。虽然这种选择适合用于隔离并评估少数类数据增强的影响，但未能充分反映真实世界入侵检测系统的复杂性，因为其中可能同时存在多个攻击类别，且这些类别之间可能部分重叠。因此，未来工作将重点把所提框架扩展至跨数据集评估场景，以评估其在领域偏移和流量分布变化下的鲁棒性。此外，将基于潜在扩散的数据增强集成到多分类入侵检测流程中，是本研究一种自然且具有实际意义的拓展。最后，进一步提高采样效率，并探索在线或增量增强策略，或可提升潜在扩散模型在大规模实时物联网安全部署中的适用性。

7 结论

现实世界中的物联网入侵检测数据集普遍存在严重的类别不平衡，这给数据驱动的安全模型带来了根本性挑战，因为少数类攻击模式往往代表性不足，难以被传统分类器充分学习。现有的数据增强策略要么依赖简单的插值机制，要么难以捕捉网络流量复杂且异质的分布，因此难以有效扩充代表性不足的攻击类别。为应对这一挑战，本文提出了一种用于合成物联网攻击数据生成的基于潜在扩散的框架。该框架将确定性自编码器与在紧凑潜在空间中运行的去噪扩散过程相结合，使所提出的潜在扩散模型在保留对复杂多模态攻击分布建模能力的同时，克服了数据空间扩散在可扩展性方面的局限。此设计能够利用有限的真实攻击数据生成多样且逼真的合成样本，且生成过程中无需访问良性流量。实验结果表明，将潜在扩散用于数据增强时，能够持续提升攻击检测性能。与经典过采样技术和成熟的深度生成模型相比，所提出的方法在分布相似性、样本多样性、依赖关系保留和计算效率之间实现了良好平衡。具体而言，潜在扩散显著提升了分布式拒绝服务攻击和 Mirai攻击的入侵检测系统性能，并对更具挑战性的中间人攻击场景带来了切实改善；定量指标、流形分析、互信息保留及分类结果均证实了这一点。因此，本研究表明，潜在扩散模型是物联网安全领域合成表格数据生成的一种有效且可扩展的解决方案。除入侵检测外，所提出的框架还为解决其他网络安全和网络分析应用中的数据稀缺与类别不平衡问题提供了一种通用方法，有望推动构建更稳健、数据效率更高的学习系统。

致谢

本研究部分获得以下资助：欧盟通过 INCIBE 提供的 CyberFold（ETD202300129）项目（欧盟下一代计划、复苏、转型与韧性计划）；由 MI- CIU/AEI/10.13039/501100011033andbyFEDER/UE 资助的 LATENTIA（PID2022-140786NB-C31）和 HERMES（PID2023-152331OA-I00）项目；以及 ELLIS 马德里节点（马德里自治区）的资助。

参考文献

Yusuf Perwej、Kashiful Haq、Firoj Parwej 和 Mumdouh M. Mohamed Hassan。物联网（IoT）及其应用领域。*International Journal of Computer Applications*，182(49):36–49，2019。Ritika Lohiya 和 Ankit Thakkar。物联网的应用领域、评估数据集与研究挑战：系统综述。*IEEE Internet of Things Journal*，8(11):8774–8798，2021。Muhammad Shafiq、Zhaoquan Gu、Omar Cheikhrouhou、Wajdi Alhakami 和 Habib Hamam。“物联网”的兴起：与物联网安全攻击检测和防御相关的综述及开放研究问题。*Wireless CommunicationsandMobile Computing*，2022(8669348):1–12，2022。Ansam Khraisat、Iqbal Gondal、Peter Vamplew 和 Joarder Kamruzzaman。入侵检测系统综述：技术、数据集与挑战。*Cybersecurity*，2(20):1–22，2019。Alex Shenfield、David Day 和 Aladdin Ayesh。采用人工神经网络的智能入侵检测系统。*ICTExpress*，4(2):95–99，2018。Francisco M. Melgarejo-Meseguer、Andrés Lorenzo-Bleda、Sergio Eduardo-Abbenante、Francisco Javier Gimeno-Blanes、Estrella Everss-Villalba、Sergio Muñoz-Romero、José Luis Rojo-Álvarez 和 Rafael Maestre Ferriz。利用家庭环境传感器，从低维潜在流形中进行异常检测。*IEEE Internet of Things Journal*，11(8):13411–13423，2021。Haoqi Huang、Ping Wang、Jianhua Pei、Jiacheng Wang、Shahen Alexanian 和 Dusit Niyato。异常检测领域的深度学习进展：全面综述。*IEEE Internet of Things Journal*，12(21):44318–44342，2025。Barjinder Kaur、Sajjad Dadkhah、Farzaneh Shoeleh、Euclides Carlos Pinto-Neto、Pulei Xiong、Shahrear Iqbal、Philippe Lamontagne、Suprio Ray 和 Ali A. Ghorbani。物联网安全数据集的演进：挑战与未来方向。*InternetofThings*，22(100780):1–23，2023。Christin Alex、Giselle Creado、Wesam Almobaideen、Orieb Abu Alghanam 和 Maha Saadeh。物联网安全数据集分类体系、分类方法与机器学习机制的全面综述。*Computers& Security*，132(103283):1–21，2023。Elias Dritsas 和 Maria Trigka。物联网网络安全综述。*Future Internet*，17(30):1–32，2025。Huan Ma、Wan Zhang、Dalong Zhang 和 Baozhan Chen。一种基于特征选择和大语言模型微调的物联网入侵检测框架。*Scientific Reports*，15(21158):1–21，2025。S Kumar Reddy Mallidi 和 Rajeswara Rao Ramisetty。优化物联网入侵检测：结合特征选择与数据平衡的机器学习和深度学习方法系统综述。*WIREs Data Mining and Knowledge Discovery*，15(2):1–41，2025。Md Mahbubur Rahman、Shaharia Al Shakil 和 Mizanur Rahman Mustakim。物联网网络入侵检测系统综述。*Cyber Security and Applications*，3(100082):1–19，2025。Nitesh V. Chawla、Kevin W. Bowyer、Lawrence O. Hall 和 W. Philip Kegelmeyer。SMOTE：合成少数类过采样技术。*Journal of Artificial Intelligence Research*，16(1):321–357，2002。Nur Athirah Azhar、Muhammad Syafiq Mohd Pozi、Aniza Mohamed Din 和 Adam Jatowt。对用于不平衡数据集的 SMOTE 方法及数据复杂度分析的研究。*IEEE Transactions on Knowledge and DataEngineering*，35(7): 6651–6672，2022。Shuo Zhang、Jinyi Chen、Jiayuan Chen、Xiaofei Chen 和 Hejiao Huang。利用时空变分自编码器进行物联网数据插补。*Neurocomputing*，529:23–32，2023。Kunfeng Wang、Chao Gou、Yanjie Duan、Yilun Lin、Xinhu Zheng 和 Fei-Yue Wang。生成对抗网络：简介与展望。*IEEE/CAAJournalofAutomatica Sinica*，4(4):588–598，2017。Alex X. Wang、Stefanka S. Chukova、Colin R. Simpson 和 Binh P. Nguyen。表格数据生成模型面临的挑战与机遇。*Applied Soft Computing*，166(112223):1–15，2024。Xiang Li、Yang Peng 和 Hao Zheng。变分自编码器、生成对抗网络和扩散生成模型的研究与分析。*Science andTechnology of Engineering, ChemistryandEnvironmental Protection*，1(6):1–5，2025a。Jonathan Ho、Ajay Jain 和 Pieter Abbeel。去噪扩散概率模型。载于*Proceedings ofthe34thInternationalConference on NeuralInformationProcessingSystems (NeurIPS)*，第 6840–6851 页，2020。Zhong Li、Qi Huang、Lincen Yang、Jiayang Shi、Zhao Yang、Niki van Stein、Thomas Back 和 Matthijs van Leeuwen。用于表格数据的扩散模型：挑战、当前进展与未来方向。*arXiv*，2025b。Robin Rombach、Andreas Blattmann、Dominik Lorenz、Patrick Esser 和 Björn Ommer。利用潜在扩散模型进行高分辨率图像合成。载于*Proceedings ofthe 35th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*，第 10684–10695 页，2022。

Timur Sattarov、Marco Schreyer 和 Damian Borth。FinDiff：用于金融表格数据生成的扩散模型。载于第四届 *ACM* 金融人工智能国际会议（*ICAIF*）论文集，第 64–72 页，2023 年。Mario Villaizán-Vallelado、Matteo Salvatori、Carlos Segura 和 Ioannis Arapakis。用于表格数据插补和合成数据生成的扩散模型。*ACM* 数据知识发现汇刊，19(125)：1–32，2025 年。Chaoyi Zhu、Jiayi Tang、Juan F. Pérez、Marten van Dijk 和 Lydia Y. Chen。DP-TLDM：差分隐私表格潜在扩散模型。载于第*20*届可用性、可靠性与安全国际会议（*ARES*）论文集，第 337–357 页，2025 年。Yan Naung Soe、Paulus Insap Santosa 和 Rudy Hartanto。基于简单人工神经网络与 SMOTE 的物联网环境 DDoS攻击检测。载于第四届信息学与计算国际会议（*ICIC*）论文集，第 1–5 页，2019 年。Pandit Byomakesha Dash、Janmenjoy Nayak、Bighnaraj Naik、Etuari Oram 和 SK Hafizul Islam。采用多类别自适应提升与 SMOTE 的基于模型的物联网安全框架。*Security and Privacy*，3(5)：1–15，2020 年。Lucas Pinheiro-Cinelli、Matheus Araújo-Marins、Eduardo Antúnio Barros da Silva 和 Sérgio Lima-Netto。变分自编码器。载于用于机器学习及深度网络应用的变分方法，第 5 章，第 111–149 页。Springer，Cham，柏林，德国，2021 年。Chang Liu、Ruslan Antypenko、Iryna Sushko 和 Oksana Zakharchenko。采用基于 VAE 和 CVAE 的数据增强方案后的入侵检测系统。*IEEE* 可靠性汇刊，71(2)：1000–1100，2022 年。Ying Zhang 和 Qiang Liu。基于数据增强以提升不平衡样本学习效果的物联网入侵检测。*Future Generation Computer Systems*，133：213–227，2022 年。Cheolhee Park、Jonghoon Lee、Youngsoo Kim、Jong-Geun Park、Hyunjin Kim 和 Dowon Hong。采用生成对抗网络的增强型 AI 网络入侵检测系统。*IEEE Internet of Things Journal*，10(3)：2330–2345，2023 年。Saifur Rahman、Shantanu Pal、Shubh Mittal、Tisha Chawla 和 Chandan Karmakar。SYN-GAN：一种使用基于 GAN 的合成数据保障物联网安全的强健入侵检测系统。物联网，26(101212)：1–13，2024 年。Chiara Camerota、Lorenzo Pappone、Tommaso Pecorella 和 Flavio Esposito。解决物联网数据安全问题：用于改进恶意软件检测的最小样本量与去噪扩散模型。载于第*20*届网络与服务管理国际会议（*CNSM*）论文集，第 1–7 页，2024 年。Peng Wang、Yafei Song 和 Xiaodan Wang。DIFT：一种用于物联网不平衡学习入侵检测的扩散变换器。*Journal of Network and Systems Management*，33(48)：1–24，2025 年。Akim Kotelnikov、Dmitry Baranchuk、Ivan Rubachev 和 Artem Babenko。TabDDPM：使用扩散模型对表格数据建模。载于第*40*届国际机器学习会议（*ICML*）论文集，第 17564–17579 页，2023 年。Alex Nichol 和 Prafulla Dhariwal。改进的去噪扩散概率模型。载于第*38*届国际机器学习会议（*ICML*）论文集，第 2319–2329 页，2021 年。Tero Karras、Miika Aittala、Samuli Laine 和 Timo Aila。阐明基于扩散的生成模型的设计空间。载于第*36*届神经信息处理系统国际会议（*NeurIPS*）论文集，第 26565–26577 页，2022 年。Euclides Carlos Pinto-Neto、Sajjad Dadkhahand Raphael Ferreira、Alireza Zohourian、Rongxing Lu 和 Ali A. Ghorbani。CICIoT2023：用于物联网环境中大规模攻击的实时数据集和基准。*Sensors*，23(13)：1–26，2023 年。Krushang Sonar 和 Hardik Upadhyay。综述：针对物联网的 DDOS 攻击。*International Journal of EngineeringResearch and Development*，10(11)：58–63，2014 年。Syed Ghazanfar Abbas、Fabiha Hashmat、Ghalib A. Shah 和 Kashif Zafar。物联网僵尸网络家族的通用特征开发。*Forensic Science International: Digital Investigation*，38(301224)：1–14，2021 年。Hamidreza Fereidouni、Olga Fadeitcheva 和 Mehdi Zalai。物联网与中间人攻击。*Forensic Science International:Digital Investigation*，8(2)：1–19，2025 年。