边缘网络中的分布式流量合成与分类：一种联邦自监督学习方法

Yong Xiao，高级会员，*IEEE*，Rong Xia、Yingyu Li、Guangming Shi，会士，*IEEE*，Diep N. Nguyen，高级会员，*IEEE*，Dinh Thai Hoang，高级会员，*IEEE*，Dusit Niyato，会士，*IEEE*，以及 Marwan Krunz，会士，*IEEE*

arXiv:2302.00207v1 [cs.LG] 1 Feb 2023

摘要—随着无线服务需求的不断增长以及人们对数据保护必要性的认识日益增强，现有的网络流量分析和管理架构在分类和合成日益多样化的服务与应用方面正面临前所未有的挑战。本文提出 FS-GAN，这是一种联邦自监督学习框架，可支持对大量异构数据集进行自动流量分析与合成。FS-GAN 由多个分布式生成对抗网络（GAN）组成，其中包含一组生成器，每个生成器都经过专门设计，可生成符合单个业务流量分布的合成数据样本；每个判别器则经过训练，用于区分合成数据样本与本地数据集中的真实数据样本。本文采用基于联邦学习的框架，协调不同 GAN 在不同数据集上的本地模型训练过程。FS-GAN 能够对未知类型服务的数据进行分类，并创建能够捕捉这些未知类型流量分布的合成样本。我们证明，FS-GAN 能够最小化所有数据集中真实数据分布与合成数据样本分布之间的詹森-香农散度（JSD）。FS-GAN 还最大化……之间的 JSD 不同生成器所创建的数据样本分布各异，因此每个生成器生成的合成数据样本都遵循某一种特定服务类型的分布。大量仿真结果表明，与最先进的基于聚类的流量分析算法相比，FS-GAN 的分类准确率平均提高了超过20%。FS-GAN 还能够合成高度复杂的流量类型混合分布，且无需任何人工标注的数据样本。

索引词—流量分类、自监督学习、边缘计算、联邦学习、生成对抗网络。

1 引言

过去几十年来，在众多垂直领域中不断涌现的创新服务与应用推动了电信技术的飞速发展 [1],[2]。尤其是，近期标准化的 5G 技术提出了三大主要应用场景：增强型移动宽带（eMBB）、超可靠低时延通信（URLLC）和海量机器类通信（mMTC），并有望支持包括物联网 [3]–[5] 和自动驾驶车辆 [6]在内的新型应用。根据近期报告 [7], [8], ，下一代移动技术（如 B5G 和 6G）有望通过提供更多样化、更具创新性的服务，进一步拓展 5G 的应用场景，例如全息通信、触觉互联网 [9], 、语义通信 [10], [11],等[12]。

随着越来越多的服务和应用在不同地区于不同时间推出并得到采用，网络流量分析与管理面临着前所未有的挑战。尤其是，不同服务的多样化需求与要求大幅增加了网络流量的动态性，使流量预测、网络规划、资源分配和调度比以往更加困难 [13], [14]。大多数现有解决方案依赖回归或基于卷积神经网络（CNN）的监督学习来拟合历史数据，并据此对现有服务的流量进行分类 [15]。这些解决方案通常需要大量人工标注数据集，而实际中可能难以获取 [16]。近期研究采用基于聚类的无监督学习解决方案来分析并聚类未知流量 [17]。然而，这些解决方案往往准确性有限，也无法用于预测或跟踪高度混合且不断演变的业务

*本文已被 *IEEE*移动计算汇刊 接收。版权可能会在未事先通知的情况下转让，届时此版本可能不再可供访问。*Y. Xiao* 获得了国家自然科学基金项目（批准号：*62071193*）及湖北省重点研发计划项目（批准号：*2021EHB015* 和 *2020BAA002*）的部分资助。*G. Shi* 获得了国家自然科学基金项目（批准号：*62293483*、*61871304* 和 *61976169*）的部分资助。*Y. Xiao* 和 *G. Shi* 获得了鹏城实验室重大重点项目（项目编号：*PCL2021A12*）的部分资助。*M. Krunz* 获得了美国国家科学基金会（项目编号：*1910348*、*1731164*、*1813401*、*2229386* 和 *IIP-1822071*）以及宽带无线接入与应用中心（*BWAC*）的资助。*Y. Xiao* 任职于中国武汉 *430074* 华中科技大学电子信息与通信学院，同时任职于中国广东深圳 *518055* 鹏城实验室以及中国广东广州 *510555* 琶洲实验室（黄埔）（电子邮箱：*yongxiao@hust.edu.cn*）。*R. Xia* 任职于中国武汉 *430074* 华中科技大学电子信息与通信学院（电子邮箱：*rongxia@hust.edu.cn*）。*Y. Li* 任职于中国武汉 *430074* 中国地质大学机械工程与电子信息学院（电子邮箱：*liy- ingyu29@cug.edu.cn*）。*G. Shi* 任职于中国广东深圳 *518055* 鹏城实验室，同时任职于中国陕西西安 *710071* 西安电子科技大学人工智能学院以及中国广东 *510555* 琶洲实验室（黄埔）（电子邮箱：*gmshi@xidian.edu.cn*）。*D. Nguyen* 和 *D. T. Hoang* 任职于澳大利亚悉尼科技大学电气与数据工程学院（电子邮箱：{*diep.nguyen, hoang.dinh*}*@uts.edu.au*）。*D. Niyato* 任职于新加坡南洋理工大学计算机科学与工程学院（电子邮箱：*dniyato@ntu.edu.sg*）。*M. Krunz* 任职于美国亚利桑那州图森亚利桑那大学电气与计算机工程系（电子邮箱：*krunz@arizona.edu*）。

流量。我们迫切需要开发智能流量分析与分类解决方案，以自动识别已知服务，并对不同地点的去中心化数据集中出现的未知服务和数据样本进行分类或标注。

应对上述挑战的一种有前景的解决方案是自监督学习，这是一种新颖的算法框架，无需任何有标签的数据样本即可训练机器学习（ML）模型，并从中学习特征和表征。自监督学习首先通过解决一些精心设计的前置任务，自动为无标签数据样本创建伪标签，从而结合监督学习和无监督学习的优势。随后，它采用基于监督学习的下游训练任务，基于伪标注数据集构建机器学习模型 [18]–[20]。鉴于自监督学习有望显著提高数据效率和模型泛化能力，人们普遍认为，它是推动“下一次人工智能革命”的关键因素之一 [21]。

尽管具有潜力，现有大多数自监督学习研究都侧重于设计利用图像或视频属性的前置任务，例如图像方向识别、灰度图像着色和图像拼图求解；这些任务均无法用于分析网络流量数据集。此外，与图像数据样本不同，网络数据包通常尺寸小得多，且不具备图像数据样本丰富的（像素级）信息以及众所周知的表征特征和属性。网络中传输的数据流包含高度混杂的流量类型，与不同服务相关的数据包可能呈现非常相似、难以分类的帧模式。目前，人们仍不了解与各个服务相关的数据包有哪些普遍特征。

智能服务对响应速度日益严格的要求，以及人们对数据保护和网络安全意识的不断增强，进一步加剧了上述挑战。尤其值得注意的是，现有的大多数自监督学习方案都基于云端，在这些方案中，每个移动设备都必须将其本地收集的数据样本上传至云服务器，并等待反馈后才能做出决策 [22], [23]。众所周知，这些方案存在通信时延长的问题，并且由于暴露原始数据而面临隐私泄露风险。为了解决这些问题，近年来引入了边缘智能，使大量低成本且通常为去中心化的边缘设备（例如移动边缘服务器）能够进行分布式数据处理和学习 [24]。它被认为是 B5G/6G 中最受追捧的功能之一。

实施边缘智能面临的挑战之一，是为大型、资源受限的无线网络开发简单有效的算法方案，以实现去中心化数据的学习与处理 [12]。作为一种新兴的分布式人工智能方案，联邦学习（FL）能够在不暴露去中心化数据集中的任何原始数据的情况下，协作构建机器学习模型。将联邦学习与自监督学习相结合，有望显著降低通信开销、增强数据保护，并提高模型训练效率。遗憾的是，传统的联邦学习方法往往会遭遇收敛缓慢和偏差问题——当不同数据集中的流量分布高度异质且不平衡时。事实上，最常用的联邦学习方案 FedAvg 仅适用于所有去中心化数据集都具有相同特征集的情况。此外，现有的大多数联邦学习方案都基于监督学习，因此要求所有去中心化数据集都包含高质量的带标签样本 [25]。目前，还没有简单有效的自监督学习方案能够基于异质数据集联合构建模型。

本文提出了联邦自监督生成对抗网络（FS-GAN），这是一种面向去中心化数据集上的分布式流量分析与合成的新型框架，其数据集由高度异构且不平衡的流量类型组成。FS-GAN 由多个去中心化的生成对抗网络组成，这些网络部署在一组边缘服务器上，每个边缘服务器都只能访问与本地服务组合相关的一组专属数据样本。不同的边缘服务器可以访问不同的服务组合。生成器部署在能够提供计算功能的边缘服务器或虚拟机上。相比之下，每个判别器均部署在一台边缘服务器上，并被授予访问本地数据集的专属权限；这些本地数据样本需要得到一定程度的隐私保护。FS-GAN 学习创建能够捕获真实数据统计特征（即其分布）的合成数据样本，这些真实数据来自每种独立的服务，用于支持自动流量分类。与传统生成对抗网络相比，FS-GAN 具有三个独特优势。首先，它采用基于联邦学习的方法来协调不同边缘服务器上的模型训练，从而支持基于去中心化数据集的协同模型构建。其次，它通过为每个本地数据集关联一个分类器，对不同生成器生成的样本进行分类，从而解决许多基于生成对抗网络的方法所面临的模型坍塌问题。第三，FS-GAN 不需要任何带标签样本来训练模型；相反，它采用基于自监督学习的方法，自动为不同生成器生成的合成数据样本分配不同的伪标签。与传统的网络数据分类/聚类方案相比，所提出的 FS-GAN 训练多个生成器，其中每个生成器都能够为给定服务合成数据流量。FS-GAN 的这些特性使其能够进一步提升数据流量分类性能，尤其是在不同服务的流量样本高度不平衡时，即某些服务的数据样本远少于其他服务。它们还使 FS-GAN 适用于涉及流量预测与合成的广泛场景和用例，例如涉及多服务流量跟踪与预测的动态网络切片、未知攻击检测，以及基于预测的网络规划。我们基于与十种热门服务应用相关的真实世界流量开展了广泛仿真，这些应用包括电子邮件、FTP、视频和音频聊天等 [26]。结果表明，与当前最先进的基于聚类的流量分析算法相比，FS-GAN 的分类准确率平均提升了20% 以上。

为突出 FS-GAN 的优势，图1展示了来自

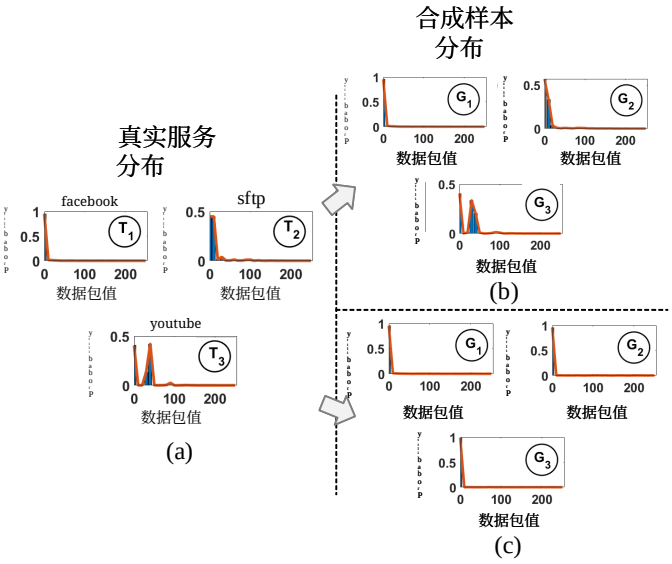


图1：比较 (a) 以数据包值（数据包的十进制转换值）衡量的三种服务流量分布T_1，T_1，T_3 与 (b) FS-GAN 生成的合成流量数据，以及 (c) F-GAN 生成的合成流量。

3 类（未知）应用的真实流量数据包的流量分布（图1(a)），以及 FS-GAN 经过 50 轮训练后生成的流量分布（图1(b)），并与将传统多生成器生成对抗网络直接扩展到联邦学习环境中所生成数据的分布进行比较，后者称为 F-GAN（图1(c)）。可以看出，FS-GAN 能够成功生成符合不同类型应用分布的合成数据样本。然而，F-GAN 无法捕捉不同服务之间的数据分布差异，只能生成多样性有限的样本。本文的主要贡献如下：

1) 我们提出了一种名为 FS-GAN 的新型框架，该框架融合了生成对抗网络、联邦学习和自监督学习的概念，可在去中心化数据集上实现异构流量的自动学习、合成与分类。与基于回归的聚类方案相比，这类方案依据单个或有限数量的预选特征分离数据样本；FS-GAN 则自主生成带有伪标签的合成样本，以捕捉流量类型本地混合分布的分布，随后采用类监督学习方案进一步提升分类性能。据我们所知，FS-GAN 是首个采用自监督学习的网络流量合成与分类方案。

2) 我们证明，由 FS-GAN 训练的模型能够最小化真实流量分布与 FS-GAN 生成器生成的合成数据样本分布之间的差异。FS-GAN 具有独特的能力，能够学习并生成捕捉分散式数据集上高度混合但未知的流量分布的合成样本，因此可应用于各种新颖的应用场景，远超传统的流量分类。我们还讨论了一些潜在应用，并指出了 FS-GAN 在不同场景下的局限性。3) 我们使用真实流量数据集进行了广泛的仿真。结果表明，FS-GAN 在流量分类和数据合成方面均取得了显著提升。

本文其余部分安排如下。第2节首先回顾与 FS-GAN 相关的近期研究。随后，我们在第3节介绍其架构并讨论其应用场景。第4节将介绍所提架构、算法的详细内容以及主要理论结果。我们在第5节开展广泛的仿真并评估 FS-GAN 的性能。最后，第6节总结全文。

2 相关工作

本文提出的方案与基于深度学习的流量分类，以及基于联邦学习的数据分析与合成方面的最新进展密切相关。本节回顾了大多数相关工作。

2.1 基于深度学习的流量分类

深度神经网络（DNN）是网络流量分类的一种有前景的解决方案 [15], [17], [27]–[34]。例如，Wang等人提出了一种混合神经网络，结合循环神经网络（RNN）和 CNN，学习移动数据的动态特征以进行流量分类 [15]。Liang等人提出了一种名为 NeuroCuts 的基于深度强化学习的方法，用于数据包分类 [30]。Zhang等人提出了一种基于深度学习的流量聚类方案，可对与未知类别相关联的数据包进行分类，并自动构建新的训练数据集，以更新用于识别未知流量的分类器 [17]。Liu等人采用基于 RNN 的流序列网络方法对加密流量进行分类 [28]。Aouedi 等人采用基于集成学习的方法来提高流量分类准确率等人 [35] 和 Jin等人 [36]。

2.2 基于联邦学习的网络数据分析

联邦学习及其在网络数据分析中的应用近来引起了广泛关注 [37]–[44]。联邦学习有望在去中心化数据集上进行分布式训练和协作期间保护数据隐私。因此，网络领域的 Wen等人提出了一种两步联邦学习框架，以实现兼具高通信效率的隐私保护模型训练。Wang等人提出了一种由实证驱动的方案，用于优化参与模型更新的客户端设备选择 [39]。近期，联邦学习被扩展至资源受限网络系统中的数据分析。例如，Wang等人研究了基于梯度下降的联邦学习方案的收敛性，并在给定资源约束下优化本地更新与全局模型聚合之间的权衡 [45]。Luo等人提出了一种低成本的采样算法，用于学习与收敛相关的参数，并尽量降低学习所需的时间和能耗 [46]。Xiao等人提出了一种联邦边缘智能（FEI）框架，用于实现

在通信和计算资源受限的边缘计算辅助物联网网络中实现联邦学习 [47]。该框架进一步扩展至边缘智能网络中的实时学习 [48], 以及语义通信 [10] 和语义感知网络系统 [11], [49], [50]。 Zhang等人 [51] 研究了基于同态加密的联邦学习在隐私医疗数据分析中的通信和计算成本，Lu 等人 [52] 研究了基于联邦学习的工业数据类别不平衡分类。

2.3 基于自监督学习的数据分析

近年来，自监督学习被引入，用于提高数据效率和模型泛化能力 [18]–[20], [53]–[56]。具体而言，Araslanov 等人提出了一种基于伪监督目标生成的预测进行语义分割的领域自适应方法 [20]。Li 等人提出了一个用于异常检测的两阶段框架，通过学习自监督深度表征模型来构建生成式分类器 [53]。Sun等人提出了一种用于图卷积网络（GCNs）的新型模型训练算法。其中的自监督模型[54] 提升了模型泛化性能，同时减少了所需带标签样本的数量。Zhang 等人提出了一种结合自适应记忆网络的自监督学习方法，以提升模型泛化能力并增强特征代表性 [57]。Shi 等人研究了一种用于自监督学习的新型掩码图像模型，其中运用了对抗训练的思想 [58]

我们提出的 FS-GAN 与传统的基于聚类的数据分析技术不同，它采用基于自监督学习的方案，首先创建带有伪标签、在统计上与真实流量类型的混合分布相匹配的合成样本，然后使用伪标记数据集训练机器学习模型，以分类和识别各类业务流量。

3 架构概述与应用场景

3.1 架构概述

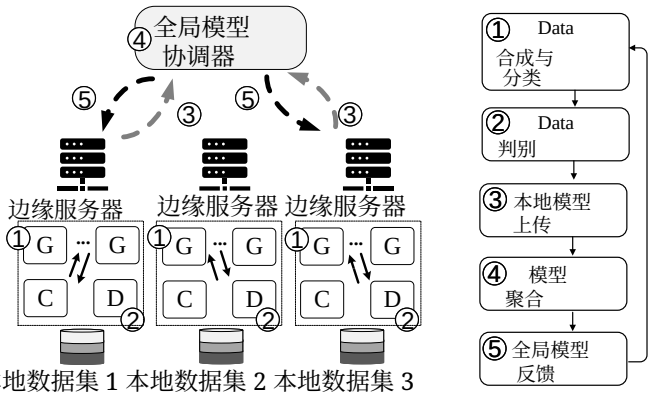


图 2：FS-GAN 架构，包括生成器（G）、判别器（D）、分类器（C）和全局模型协调器等成分（左图）；以及主要算法流程（右图）。

FS-GAN 的整体架构，包括其关键成分和算法流程，如图所示。

2. FS-GAN 由以下部分组成：(1) 多个生成器，用于创建合成样本，其统计分布与本地数据集中的真实流量相似；(2) 多个判别器，每个判别器都尝试区分真实流量与生成器生成的合成样本。我们提出的 FS-GAN 框架采用以下三个流程：

(1) 本地数据合成：每台边缘服务器都部署一个判别器，该判别器独占访问本地数据集的权限。每个判别器都与若干相关联的生成器共同训练。尽管生成对抗网络通常具有很强的合成样本生成灵活性 [59]–[61]，但它们会遇到所谓的模型坍塌问题：生成器学会生成多样性极其有限的合成样本，例如仅捕获单一流量类型或给定流量类型集合的固定混合分布的数据分布的样本。为解决这一问题，我们借鉴多生成器生成对抗网络的思路 [62]–[64] ，为每个生成的样本分配一个伪标签，以标明该样本来自哪个生成器。正如后文将证明的那样，与判别器在本地进行训练的生成器能够创建与本地数据集中真实流量的数据分布相匹配的合成数据样本。

(2) 全局模型协调：判别器和生成器训练得到的本地模型将定期上传至全局模型协调器。协调器可部署在某台边缘服务器上，也可部署在云数据中心。模型协调的一种直接实现方式是采用联邦学习方案 [40], [41]，对生成器和/或判别器的本地模型参数进行聚合，例如使用 FedAvg 联邦平均算法 [65], 求取平均值，以形成更新后的全局模型，然后将其广播至边缘服务器。需要注意的是，来自边缘服务器的模型参数数量会影响通信效率。本文考虑两种模型协调方案：

(i) 协调方案一 (C-I)：将本地训练的判别器及其对应的生成器上传至全局协调器。**(ii) 协调方案二 (C-II)：**仅将判别器上传并由协调器进行聚合。

C-II 需要在边缘服务器与协调器之间交换较少的模型相关数据，因此通信效率更高。然而，正如本文后文所示，C-I 的收敛速度快于 C-II，在某些情况下，这可以弥补其额外的通信开销。

在本地数据合成过程中，每个判别器只判断数据样本是真实流量，还是生成器生成的合成数据。经过全局模型协调后，所有判别器都应能够区分任意本地数据集中的真实数据与任意生成器生成的合成数据。同样，生成器也将能够捕获不同本地数据集中的更丰富特征。为解决模型坍塌问题，我们在每个本地数据集中引入一个分类器，用于对来自不同生成器的样本进行分类。如第 4 节所示，该分类器增大了不同生成器所生成合成数据分布之间的差异度，从而缓解传统生成对抗网络的模型坍塌问题。

（3）自监督学习：上述经过训练和更新的分类器为不同生成器生成的合成样本创建伪标签。随着生成器越来越擅长生成接近真实数据的合成数据，与每个判别器关联的分类器自然也将支持对未来到达的真实流量进行数据分类和自标记。我们表明，该分类器实际上与本地训练的判别器共享相同的模型参数，因此无需额外投入进行模型训练。

我们给出了理论分析，证明了 FS-GAN 框架的有效性。具体而言，我们证明 FS-GAN 具有两项重要性质：（1）生成器生成的合成数据混合分布与真实流量数据分布之间的差异度最小；（2）不同生成器生成的合成数据样本分布之间的 JSD散度最大。基于这些性质，我们证明了本文的主要结论：如果与不同服务相关联的流量数据分布彼此可分，那么每个生成器生成的合成数据样本都能捕捉相应流量类型的真实流量分布，并且分类器能够区分来自不同生成器的所有合成样本。

与现有的自监督学习方案相比，FS-GAN 具有以下独特优势，因此更适用于网络流量分类问题。首先，FS-GAN 可应用于流量分布高度分散且不均衡的大型网络，同时满足数据隐私保护要求。其次，FS-GAN 无需对数据进行任何预标记，因此适用于当今许多网络系统；这些系统涉及大量未标记网络流量数据集，而对其进行标记的成本高得令人难以承受。第三，FS-GAN 的数据合成能力使其能够应用于许多新兴网络服务，这些服务涉及流量预测与合成。

3.2 应用场景示例

与传统流量分类方案相比，FS-GAN 采用了一种截然不同的方法：首先训练多个生成器网络，生成能够捕捉与单项服务相关的真实流量分布的合成样本；随后利用已训练模型的参数，对不同服务的流量进行分类。FS-GAN 的这一独特之处使其适用于一系列新的应用场景和用例，包括以下内容：

1) 5G/B5G 中的动态网络切片：5G NR release16 [66] 引入了网络切片的概念，着重于隔离和保留网络资源与功能的分区，这些分区通常称为切片。这些切片经过定制和编排，以支持具有不同服务质量要求的应用。网络切片面临的一项关键挑战，是跟踪潜在未知应用的流量需求，从而在满足这些应用需求的同时预留适量资源。FS-GAN 可直接用于分类并跟踪不同服务的需求，即使流量高度混杂。它还可以帮助网络切片管理器识别新观测到的服务流量类型，这些类型缺乏足够的预标注数据。

2) 未知攻击检测：流量分类的一个重要应用场景是攻击检测。现有方案大多侧重于检测已知攻击。检测未知攻击一直是一个公认的难题 [67]。FS-GAN 有望识别未知攻击，并模拟各种攻击场景及其混合分布。

3) 基于流量预测的网络规划：由于 FS-GAN 得到的最终模型能够生成在统计特性上与各种新兴流量真实数据相似的合成数据样本，因此可将 FS-GAN 用于预测未来流量趋势，并协助规划网络基础设施，以更好地满足预期需求。

4 FS-GAN 设计与理论分析

本节介绍 FS-GAN（C-I 和 C-II）的主要流程算法细节：本地数据合成、全局模型协调和自监督学习。详细架构如图 3 所示。本节末尾将进行理论分析。

表 1：符号列表

符号说明	描述
d	本地数据集的索引
D	FS-GAN 的判别器
G	FS-GAN 的生成器
P	数据分布
z	高斯分布的噪声
m	生成器索引
$\mathcal{L}$	损失函数
x	真实数据样本
$\hat{x}$	合成数据样本
λ	多样性超参数
ω	生成器参数
θ	判别器和分类器参数

4.1 LocalData 合成

首先，我们考虑在第 d 个本地数据集 $d = 1, ..., N$ 上对单个生成器进行本地模型训练。如前所述，该过程的主要目标是训练一个模型，使其能够生成合成数据样本，以捕捉本地数据集真实流量的分布。设 G_d 和判别器 D_d 分别为与第 d 个数据集关联的生成器和判别器。生成对抗网络的训练可以表述为 G_d 与 D_d[59] 之间的极小极大博弈。生成器和判别器通常是深度神经网络。设 P_{data} 为真实数据集的分布，并假设 $\boldsymbol{x}$ 取自 P_{data}，即记作 $\boldsymbol{x} \sim P_{data}$。生成器 G_d 旨在训练一个多层神经网络，将其输入变量映射为输出。首先生成随机噪声 $\boldsymbol{z}$，以合成数据样本，其分布记为 $P_{\boldsymbol{z}}$。不失一般性地，我们滥用符号说明，将 $G_d(\boldsymbol{z})$ 记作训练输入 $\boldsymbol{z}$ 下训练所得生成器的输出。同时，判别器 D_d 尝试区分真实数据样本与合成

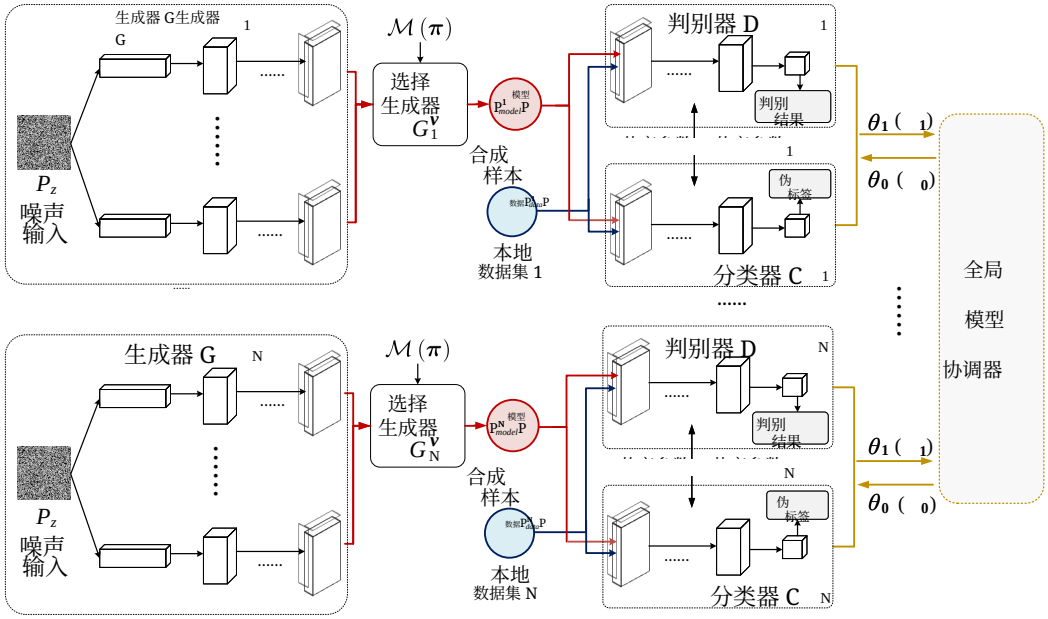


图 3：FS-GAN 的详细架构组件

样本。我们用 $D_d(\cdot)$ 表示判别器的输出，该输出代表判别器判定给定输入样本来自真实数据集而非由生成器合成的概率。更正式地说，单个生成器与单个判别器之间的对抗训练过程可通过以下极小极大问题表示:

$$\min_{G_d}\max_{D_d}\mathbb{E}_{\boldsymbol{x}\sim P_{data}}[\log D_d(\boldsymbol{x})]+\mathbb{E}_{\boldsymbol{z}\sim P_z}[\log(1-D_d(G_d(\boldsymbol{z})))]. \tag{1}$$

FS-GAN 的核心思想是使用多个生成器 $\boldsymbol{G}_d=\{G_d^1,G_d^2,...,G_d^M\}$，而非单个生成器，共同训练一个模型，使其能够捕获给定本地数据集 d中不同流量类型混合分布的特征。为了解决模型崩溃问题，我们引入一个分类器网络 C_d，用于对生成器生成的合成数据进行分类。具体而言，分类器首先为每个合成数据样本生成一个伪标签，指明该样本来自哪个生成器。随后，分类器通过最小化自身损失，促使生成器生成彼此不同的样本。正如我们将在本节后文证明的那样，采用 C_d 后，只要不同流量类型的分布之间存在足够大的差异度，每个生成器就会合成代表特定流量类型的数据样本。

现在，我们聚焦于多生成器本地模型的训练过程。我们使用下标表示本地数据集的索引，使用上标表示生成器的索引。假设一组 $\boldsymbol{G}_d=\{G_d^1,G_d^2,...,G_d^M\}$ M生成器已分配给判别器D_d，其中 $d\in\{1,2,...,N\}$ 表示判别器的索引。每个判别器 D_d 还与一个本地数据集 X_d交互，并且每个生成器 G_d^m，$m\in\{1,2,...,M\}$都将输入噪声 $\boldsymbol{z}$ 映射为 $\hat{}\boldsymbol{x}=G_d^m(\boldsymbol{z})$。分类器 C_d与 $\boldsymbol{G}_d$ 和 D_d共同训练。令 $P_{G_d^m}$ 表示由 G_d^m生成的合成样本的分布。下面我们描述每个生成器与判别器交互的规则。我们设置一个索引向量 $\boldsymbol{v}$，它从预先定义的多项分布 $\boldsymbol{v}\sim\mathcal{M}(\boldsymbol{\pi})$ 中随机抽取，其中 $\boldsymbol{\pi}=\langle\pi_{\boldsymbol{v}}\rangle_{\boldsymbol{v}\in\boldsymbol{G}_d}$ 是混合分布的加权系数。将 $G_d^{\boldsymbol{v}}\triangleq\{G_d^i:i\in\boldsymbol{v}\}$ 设为 $\boldsymbol{G}_d$的子集后，我们可以将多生成器生成对抗网络转换为一个具有生成器向量 $G_d^{\boldsymbol{v}}$ 和判别器 D_d的标准生成对抗网络。我们用 P_{data}^d 表示数据集 X_d的真实分布。随后，我们可以将采用多生成器进行本地数据合成的模型表述为以下极小极大博弈:

$$\begin{aligned}\min_{\boldsymbol{G}_d,C_d}\max_{D_d}\quad&\mathbb{E}_{\boldsymbol{x}\sim P_{data}^d}[\log D_d(\boldsymbol{x})]\\&+\sum_{m=1}^{M}\mathbb{E}_{\boldsymbol{z}\sim P_z}[\log(1-D_d(G_d^m(\boldsymbol{z})))]\\&-\lambda\sum_{m=1}^{M}\left(\pi_m\mathbb{E}_{\hat{\boldsymbol{x}}\sim P_{G_d^m}}[\log C_d^m(\hat{\boldsymbol{x}})]\right)\end{aligned} \tag{2}$$

其中，$C_d^m(\hat{}\boldsymbol{x})$ 是分类器 C_d的输出，表示数据样本 $\hat{}\boldsymbol{x}$ 由 G_d^m生成的概率。式（2）中的最后一项是多分类情形下的标准 Softmax 损失 [68]。换言之，在固定 D_d 的情况下，根据式（2）最小化 C_d 会使 C_d的熵最小化，从而增大生成器之间的差异度。我们引入多样性超参数 $\lambda>0$，用于指定分类器对生成器的影响程度。我们在第5节比较不同 λ 条件下的模型性能。

令 ω_d 为与判别器 D_d相关联的生成器 $\boldsymbol{G}_d$的模型参数，令 ω_d^m为生成器 G_d^m，$m\in\{1,...,M\}$的模型参数。分类器 C_d 和判别器 D_d 除最后一层外，共享相同的模型参数。令 θ_d 为分类器和判别器 D_d之间共享的模型参数。请注意，C_d 的最后一层会为每个输入数据样本自动输出学习得到的伪标签，而 D_d 的最后一层则会输出二元结果。与标准生成对抗网络类似，我们交替使用随机梯度

下降（随机梯度下降）方法进行学习。本地数据合成框架如图 3 所示。

设 B 为训练过程中选定的小批量大小。在每次本地训练迭代开始时，每个判别器 D_d 都会从真实数据分布 P_{data}^d 中采样一个包含 B 个数据点的小批量 $(\boldsymbol{x}_d^{(1)}, \boldsymbol{x}_d^{(2)}, ..., \boldsymbol{x}_d^{(B)})$。此外，生成器混合 G_d^v 会在 $\boldsymbol{G}_d$中生成一个包含 B 个合成数据样本的小批量 $(\hat{}\boldsymbol{x}_d^{(1)}, \hat{}\boldsymbol{x}_d^{(2)}, ..., \hat{}\boldsymbol{x}_d^{(B)})$。该混合分布可根据预先设定的采样概率 π_m生成。在每次训练迭代中，我们沿着各自损失函数的梯度更新 C_d、 D_d和 $G_d^{\boldsymbol{v}}$。根据式（2），在上述设置下，损失函数可写为:

$$\mathcal{L}(C_d, \hat{\boldsymbol{x}}_d) = -\frac{1}{B}\sum_{b=1}^{B} \log C_d^m\left(\hat{\boldsymbol{x}}_d^{(b)}\right) \tag{3a}$$

$$\mathcal{L}(D_d, \boldsymbol{x}_d, \hat{\boldsymbol{x}}_d) = -\frac{1}{B}\sum_{b=1}^{B}\Big(\log D_d\left(\boldsymbol{x}_d^{(b)}\right) + \log\left(1 - D_d\left(\hat{\boldsymbol{x}}_d^{(b)}\right)\right)\Big) \tag{3b}$$

$$\mathcal{L}(G_d^{\boldsymbol{v}}, \hat{\boldsymbol{x}}_d) = -\frac{1}{B}\sum_{b=1}^{B}\left(\log D_d\left(\hat{\boldsymbol{x}}_d^{(b)}\right) + \lambda \log C_d^m\left(\hat{\boldsymbol{x}}_d^{(b)}\right)\right) \tag{3c}$$

我们在算法1中总结了详细流程。

4.2 全局模型协调

在上一节中，与本地判别器关联的生成器会生成合成数据样本，以捕获本地数据集中不同流量类型的分布。然而，不同的本地数据集可能包含不同的流量混合分布。因此，理想的做法是训练一个能够捕获所有本地数据集中多样化流量的全局模型。我们采用基于联邦学习的框架，在无需交换本地数据集的情况下协调本地模型训练过程。联邦学习的一个关键思想是通过模型参数协调不同的本地模型训练过程。在一种名为 FedAvg 的常用联邦学习方案中，每台边缘服务器会定期将其本地训练的模型上传至全局协调器。随后，对上传的模型进行平均，形成更新后的全局模型，并将其广播给所有边缘服务器。

形式化地，令Ω′ 表示协调方案 C-II 中（或方案 C-I 中判别器与生成器对的）被选中用于上传本地模型的判别器子集。注意Ω′ $\subseteq$ Ω。令 ω_0 和 θ_0 分别为相应的全局模型参数，并令 p_d 为第 d个本地数据集的权重，使得 $p_d \geq 0$ 和 $\sum_{d=1}^{N} p_d = 1$。全局模型协调过程可表示如下:

$$\begin{cases} \omega_0 = \sum\limits_{d\in\Omega'} p_d\omega_d \text{ and } \theta_0 = \sum\limits_{d\in\Omega'} p_d\theta_d, & \text{if Scheme C-I,} \\ \theta_0 = \sum\limits_{d\in\Omega'} p_d\theta_d, & \text{if Scheme C-II.} \end{cases} \tag{4}$$

算法1 本地数据合成**输入：** 本地数据集 X_d; 噪声分布 $P_{\boldsymbol{z}}$; 采样概率 π_m; **输出：** 更新后的本地参数 ω_d 生成器 $\boldsymbol{G}_d$; 判别器 D_d 和分类器C_d的更新后本地参数 θ_d ; **初始化：** $\lambda > 0$; 小批量大小 B; 网络参数 ω_d和 θ_d; 最大迭代次数 I; **对于** 每次迭代 $i \leq I$个生成器 $\boldsymbol{G}_d$ **执行**: 1）根据 π_m选择生成器 $G_d^{\boldsymbol{v}}$; 2）每个生成器 $G_m^d \in G_d^{\boldsymbol{v}}$ 同时**执行**:

- 从 $P_{\boldsymbol{z}}$采样一个噪声输入变量 $\boldsymbol{z}$;
- 合成一批数据样本 $\hat{}\boldsymbol{x}(b)_d, b \in \{1, 2, ..., B\}$; 3) 将合成的数据样本 $\hat{}\boldsymbol{x}(b)_d$ 发送给判别器 D_d 和分类器 C_d; 4) 沿其损失函数梯度上升，更新 ω_d： $\nabla_{\omega_d}\mathcal{L}(G_d^{\boldsymbol{v}})$; 判别器 D_d 执行: 1) 从本地数据集 X_d采样一批数据 $\boldsymbol{x}(b)_d$, $b \in \{1, 2, ..., B\}$; 2) 区分 $\boldsymbol{x}(b)_d$ 和 $\hat{}\boldsymbol{x}(b)_d$; 3) 沿其损失函数梯度下降，更新 θ_d : $\nabla_{\theta_d}(\mathcal{L}(C_d) + \mathcal{L}(D_d))$; 分类器 C_d 执行: 1) 根据不同的生成器对合成样本 $\hat{}\boldsymbol{x}(b)_d$ 进行分类; 2) 沿其损失函数梯度下降，更新 θ_d : $\nabla_{\theta_d}(\mathcal{L}(C_d) + \mathcal{L}(D_d))$;$i = i + 1$; 返回本地训练参数

4.3 自监督学习

如前述流程所述，分类器与本地判别器共享相同的网络参数。因此，在全局模型协调过程中，可以更新该分类器，根据合成样本对应的生成器对其进行分类并生成伪标签。该分类器还能够对所有数据集中与所有流量类型相关的真实数据进行分类。由于该分类器是 FS-GAN 训练过程的一部分，且基于带有伪标签的合成数据集进行训练，因此不会引入额外的训练工作。

4.4 实现复杂度

由于 FS-GAN 复用了判别器的模型参数，因此与传统的 GAN 联邦学习实现相比，无需引入任何额外的计算复杂度。换言之，FS-GAN 的实现复杂度与两类实现成本密切相关：通信成本和计算成本。FS-GAN 的计算复杂度主要包括训练一组本地生成

算法 2 全局模型协调（方案 C-I）

输入： D_d 和 C_d的本地训练参数 θ_d；本地训练参数 ω_d（属于 $\boldsymbol{G}_d$）；
输出： 全局判别器的聚合参数 θ_0 —器 D_0 和分类器 C_0；聚合参数 ω_0，用于全局生成器 $\boldsymbol{G}_0$
初始化： 本地数据集数量 N；最大通信轮次数 J;
对于通信轮次数 $j \leq J$

全局模型协调器**执行**:
1) 根据聚合参数 θ_0 和 ω_0计算方程 (4):
$\theta_0, \omega_0 \leftarrow$ 本地数据合成(X_d, θ_d, ω_d)
2) 将 θ_0 广播给本地判别器和 ω_0，分类器和生成器;

每个本地模型**执行**:
通过以下方式初始化本地参数 θ_d 和 ω_0：
$\theta_d \leftarrow \theta_0$; $\omega_d \leftarrow \omega_0$;
$j = j + 1$;

结束循环

算法 3 全局模型协调（方案 C-II）

输入： D_d 和 C_d的本地训练参数 θ_d；
输出： 全局判别器的聚合参数 θ_0 —器 D_0 和分类器 C_0;
初始化： 本地数据集数量 N；最大通信轮次数量 J;
对于 通信轮次数量 $j \leq J$

全局模型协调器**执行**:
1) 计算聚合参数 θ_0，根据方程 (4b):
$\theta_0 \leftarrow$ 本地数据合成(X_d, θ_d)
2) 将 θ_0 广播至本地判别器和分类器;

每个本地模型**执行**:
初始化本地参数 θ_d：
$\theta_d \leftarrow \theta_0$;
$j = j + 1$;

结束循环

基于随机梯度下降（SGD）并行训练生成模型和判别模型，即边缘服务器执行的本地 SGD 轮数总计为 $O(MKT)$。通信成本取决于模型大小和协调轮数。换言之，FS-GAN 在不增加任何额外计算和通信复杂度的情况下，引入了新功能并解决了传统联邦生成对抗网络的问题。

4.5 理论分析

我们从以下两个方面分析 FS-GAN：数据合成能力，以及各本地模型的分类准确率。与之前的符号说明保持一致，我们用 P^d_{data}表示数据集 X_d中真实数据样本的混合分布，用 P^d_{model} 表示由生成器 G^1_d、G^2_d.....G^M_d生成的合成数据样本组合的分布。首先，我们给出方程 (2) 中分类器 C_d 的最优解。

命题 1。 对于一组生成器 $\boldsymbol{G}$ 以及给定的采样概率 π_m，分类器 ${C_d}^*(\boldsymbol{x};\ \theta_d)$ 学习到的最优分布具有以下形式:

$$C_d^{m*}(\boldsymbol{x};\theta_d) = \frac{\pi_m P_{G_d^m}(\boldsymbol{x})}{\sum_{i=1}^{M} \pi_i P_{G_d^i}(\boldsymbol{x})}, m \in \{1, 2, ..., M\} \tag{5}$$

证明：见附录 A。 □

最优分类器${C_d}^*(\boldsymbol{x};\theta_d)$ 的输出结果可视为不同合成样本分布的加权归一化。

我们采用一种常见的设定，并使用詹森-香农散度 (JSD) [69]来量化两个分布之间的差异。对于最优生成器${\boldsymbol{G}_d}^*(\boldsymbol{x};\omega_d)$，我们有以下命题:

*Pr*命题 *2.* 给定最优判别器和分类器，
生成器的目标是最小化:

$$\begin{aligned} {\boldsymbol{G}_d}^*(\boldsymbol{x};\omega_d) \quad = \quad & \arg\min_{G}(2 \cdot \mathrm{JSD}(P^d_{data} \| P^d_{model}) \\ & -\lambda \cdot \mathrm{JSD}_{\pi_1,\pi_2,...,\pi_M}(P_{G_d^1}, P_{G_d^1}, ..., P_{G_d^M}). \end{aligned} \tag{6}$$

证明：参见附录 B。 □

可以看出，式 (6) 右侧的两项分别对应于分布 P^d_{data} 与 P^d_{model}之间的差异，以及不同生成器生成的合成数据之间差异的倒数。换言之，最小化这两项会直接使 P^d_{data} 与 P^d_{model} 之间的差异最小化，同时使生成器之间的差异最大化。

在陈述主定理之前，我们作出如下假设:

假设 *1*。 数据集 X_d的真实数据分布 P^d_{data} 可以写成以下形式:

$$P^d_{data}(\boldsymbol{x}) \quad = \quad \sum_{m=1}^{M} \pi_m p_d^m(\boldsymbol{x}), \tag{7}$$

其中，对于任意给定的 $\boldsymbol{x}$ $p_d^m \boldsymbol{x} > 0$ $m \in$，如果 () 对于，则，()。 $\{1, 2, ..., M\}$ $m' \neq m$ $p_d^{m'} \boldsymbol{x} = 0$

方程（7）表示，数据样本的数据分布是由 $p_d^m(x)$ 给出的 M 个可分离分布的混合分布，其中 $m = 1, 2, ..., M$。我们可以证明以下定理:

定理 1。 假设假设 1成立。在方程（2）定义的极小极大博弈的均衡点处，最优分类器和生成器满足以下方程:

$$C_d^{m*}(\boldsymbol{x}) \to \begin{cases} 1, \text{if } x \sim P_{G_d^m}, \\ 0, \text{if } x \sim P_{G_d^{m'}}, \text{ for } m' \neq m, \end{cases} \tag{8a}$$

$$P^*_{G_d^m}(\boldsymbol{x}) = p_d^m(\boldsymbol{x}),\ \forall m = 1, 2, ..., M, \tag{8b}$$

$$P^d_{model}(\boldsymbol{x}) = P^d_{data}(\boldsymbol{x}). \tag{8c}$$

Proof: 见 A pp附录 C。 □

从 (8a) 可以看出，分类器能够正确区分由不同生成器生成的合成样本，即分类器识别出

生成每个给定合成样本的正确生成器 G_d^m 的概率 x 趋近于1。如 (8b) 和 (8c) 所示，最优生成器 G_d^{m*} 生成的合成样本遵循与 p_d^m相同的分布。此外，所有生成器根据 π_m 加权后共同生成的样本，与真实数据集具有相同的数据分布。

在假设 1 不满足的情况下，定理 1 可被视为本地数据合成和分类性能的上界。现有研究以及我们自己的实验表明，在许多实际场景中，即使假设 1 不成立，FS-GAN 仍然能够生成高质量的合成数据，并实现较高的分类准确率。我们将在第5节回到这一问题。

到目前为止，我们已经分析了本地模型的最优解。接下来，我们将重点关注全局模型的收敛。在全局模型协调过程中，我们使用联邦平均（*FedAvg*） [65], 这一在联邦学习中广泛使用的算法，定期聚合本地模型。更具体地说，假设每个参与的本地模型在接收到最新的全局模型后执行 I 次本地更新。令 T 表示每个客户端执行的本地迭代总次数。令 B 表示小批量大小。令 $\mathcal{L}_d(\boldsymbol{\varepsilon}, \xi_t^d)\, d \in \{1, 2, ..., N\}$表示本地模型在第 t次迭代中、网络参数为$\boldsymbol{\varepsilon}$ 且小批量包含 B 个样本时的损失函数 ξ_t^d。我们将 $\mathcal{L}_d(\boldsymbol{\varepsilon})$ </style}定义为 $(\boldsymbol{\varepsilon}, \xi_t^d)$ $t = 1$的平均值，即有$\mathcal{L}_d(\boldsymbol{\varepsilon}) = \mathbb{E}[\mathcal{L}_d(\boldsymbol{\varepsilon}, \xi_t^d)]$。考虑如下优化模型:

$$\min_{\boldsymbol{\varepsilon}} \{\mathcal{L}(\boldsymbol{\varepsilon}) \triangleq \sum_{d=1}^{N} p_d \mathcal{L}_d(\boldsymbol{\varepsilon})\}. \tag{9}$$

按照与 [43], [70], [71] 相同的方法，可以得到式 (9) 的理论保证。

我们可以看到，定理 1 已证明，所提出的 FS-GAN 解决了联邦学习和自监督生成对抗网络的两个根本问题。具体而言，与要求所有去中心化数据集共享相同数据特征组合的联邦学习相比，FS-GAN 允许具有不同服务流量组合的边缘服务器联合构建一个共享模型，从而识别和合成所有边缘服务器的服务流量。此外，FS-GAN 通过重复使用判别器训练得到的模型，解决了生成对抗网络的模式崩溃问题。进一步地，如第4.4节所述，与传统联邦生成对抗网络相比，FS-GAN 不会引入任何新成分，也不会增加复杂度。这进一步证明了 FS-GAN 在实际网络系统中的灵活性和实用性。

5 性能评估

本节中，我们使用真实世界流量数据集进行大量实验，以验证定理 1 所证明的 FS-GAN 的两项关键能力：对未知服务流量进行分类，以及合成不同类型的服务流量。本节其余部分将首先在第 5.1 节介绍仿真设置，然后分别在第 5.2 节和第 5.3 节给出详细的仿真结果，以评估 FS-GAN 的分类性能和流量数据合成性能。

5.1 仿真设置

为评估 FS-GAN 的性能，我们采用真实世界数据集“VPN-nonVPN数据集”（ISCXVPN2016）[26], ，评估由多种类型的未知服务构成的高度混合流量数据流。在该数据集中，我们选取了与 10 种常见服务相关的数据样本，如表 2 所示。由于我们关注的是有效信息的分类，因此移除数据包中的物理层和媒体访问控制层报头，并将每个有效载荷的长度限制为相同的 2500 字节，以在训练效率和分类准确率之间取得适当的平衡。

表 2：用于评估FS-GAN的数据集

服务	# 个数据包	服务	# 个数据包
电子邮件	32,566	YouTube	12,738
FTPS（上传）	47,795	SFTPS（上传）	107,234
SCP（下载）	85,018	Skype 视频	140,569
Facebook 音频	91,815	Skype 聊天	53,996
Tor-YouTube	54,294	VPN-Vimeo	215,102

我们在一台工作站上进行实验，该工作站配备 Intel(R) Core(TM) i9-9900K 中央处理器@3.60GHz、64.0 GB 随机存取存储器@2133 MHz、2 TB 硬盘以及两块 NVIDIA Corporation GP102 [TITAN X] GPU。每台边缘服务器均通过一块运行 Ubuntu 16.04、Python 3.6、CUDA 10.0 和 PyTorch 1.3.1 的 TITAN X GPU 进行仿真。边缘服务器之间的训练过程使用 PySyft 框架 [72]进行仿真。我们将所有选定流量类型的数据包合并到一条轨迹中，并将数据均分为六个本地数据集，分别分配给一个判别器（即一台边缘服务器）。生成器、判别器和分类器的网络架构与参数设置汇总于表 3。

表 3：仿真设置

架构设置	网络类型		
	生成器	判别器	分类器
输入尺寸	100×1	2500×1	2500×1
输出尺寸	2500×1	2×1	# 生成器的数量
激活函数	LeakyReLU	Sigmoid	LeakyReLU
#层数	8	6	
优化器	Adam 与 $\beta_1 = 0.5$ 和 $\beta_2 = 0.9$		

5.2 分类性能

如前所述，与现有聚类方案相比，FS-GAN 通过自主生成带伪标签的合成数据，为模型训练引入了一种新颖的聚类方案。本小节将 FS-GAN 的聚类性能与现有聚类方案进行比较。我们考虑了三种常用的性能指标:

（1）兰德指数（RI）： 用于衡量被正确聚类的样本对所占的比例，其中包括被划分到同一类别的相似样本，以及被划分到不同类别的不同样本。形式化地，RI 定义如下:

$$\mathrm{RI} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}}, \tag{10}$$

其中，TP 表示被分配到同一流量中的同类型数据包数量。TN 表示被分配到不同服务类型的数据包对数量。FN 表示同类型数据包对中

表4：FS-GAN 与最先进方案的聚类性能

方案	训练损失	训练方法	RI	NMI	ACC
K-means++	均方误差	迭代方法	0.71	0.30	0.32
DEC	重构与簇分配	预训练和微调	0.82	0.39	0.39
DCN	重构和 K-means损失	预训练和联合训练	0.81	0.31	0.35
IDEC	重构和簇分配	预训练和联合训练	0.85	0.42	0.40
FS-GAN	对抗与分类损失	联合对抗训练	**0.89**	**0.62**	**0.58**

表5：不同方案下的模型训练时间和聚类性能

N	n	E	B	Time 每轮（秒） C-I	C-II	RI C-I	C-II	NMI C-I	C-II	ACC C-I	C-II
1	12,000	4	100	31.19		0.91		0.64		0.63	
6	2,000	4	100	34.32	39.16	0.89	0.89	0.62	0.43	0.58	0.57
3	2,000	4	100	34.72	35.43	0.89	0.88	0.61	0.42	0.60	0.58
6	2,000	4	100	34.32	39.16	0.89	0.89	0.62	0.43	0.58	0.57
9	2,000	4	100	38.03	37.21	0.89	0.88	0.60	0.40	0.63	0.57
6	500	4	100	10.14	11.36	0.89	0.88	0.62	0.41	0.60	0.51
6	1000	4	100	18.33	18.14	0.89	0.89	0.60	0.47	0.56	0.55
6	2000	4	100	34.32	39.16	0.89	0.89	0.62	0.43	0.58	0.57
6	2000	2	100	20.32	19.73	0.89	0.88	0.62	0.43	0.60	0.53
6	2000	4	100	34.32	39.16	0.89	0.89	0.62	0.43	0.58	0.57
6	2000	8	100	68.15	75.32	0.89	0.90	0.64	0.47	0.61	0.54
6	2000	4	50	61.17	62.73	0.90	0.89	0.63	0.46	0.57	0.60
6	2000	4	100	34.32	39.16	0.89	0.89	0.62	0.43	0.58	0.57
6	2000	4	200	21.50	24.13	0.89	0.89	0.57	0.43	0.57	0.57

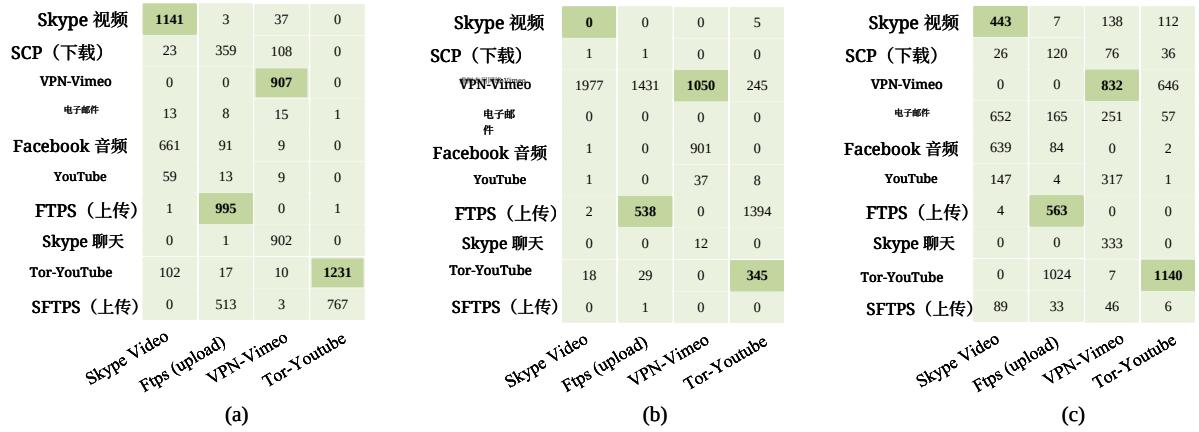


图 4：使用以下方法对 10 组服务流量数据进行聚类的结果：(a) FS-GAN、(b) K-means++ 和 (c) IDEC。

被分配到不同服务类型的数据包对数量。最后，FP 表示被分配到同一种服务类型的不同类型数据包对数量。RI 通过判断同类型样本是否被划分到同一类别来评估聚类性能。具体而言，RI 的值必须大于或等于 0；当 FN 和 FP 的值趋近于零时，RI 的值趋近于 1。

(2) 归一化互信息（NMI）： 该指标衡量在给定聚类方案后，真实聚类结果的不确定性可降低的程度。设U 为真实类别，V 为所提出的聚类算法得到的标签。NMI 定义为：

$$\mathrm{NMI}(U,V)=\frac{2\mathrm{MI}(U,V)}{H(U)+H(V)}, \tag{11}$$

其中，$\mathrm{MI}(U,\ V)$ 表示 U 与V之间的互信息，而 $H(*)$ 表示熵。我们有

(3) 无监督准确率（ACC）： 衡量所得标签与真实类别之间的最佳一对一对应关系。ACC 的定义如下：

$$\mathrm{ACC}=\frac{1}{n}\max_{\sigma}\sum_{i=1}^{n}1_{y_i=\sigma(x_i)} \tag{12}$$

其中，x_i 和 y_i 分别表示每个数据样本i对应的聚类结果标签和真实类别标签。 σ 表示任意两个类别标签之间所有可能的一对一排列。

请注意，上述三个指标的取值范围均为0 到1。数值越大，聚类性能越好。

在表4中，我们将 FS-GAN 与四种聚类方案进行比较：K-means++、DEC [73]、IDEC [74], 和 DCN [75]。可以观察到，在所有三项性能指标上，FS-GAN 在所有受测方案中表现最佳

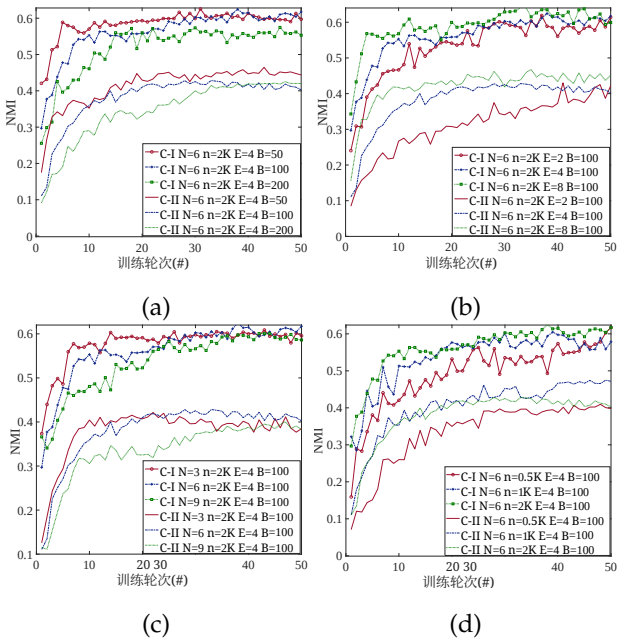


图 5：基于真实世界交通数据集的 FS-GAN-I 和 FS-GAN-II 的聚类性能。

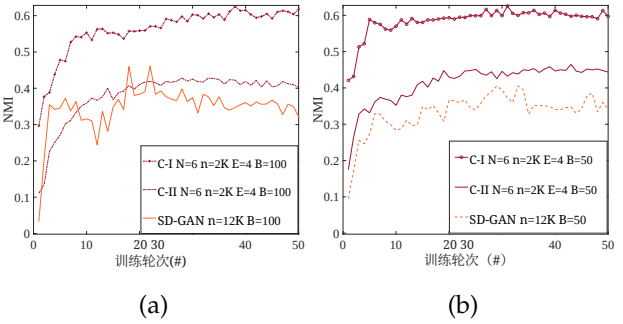


图 6：基于真实世界交通数据集的 FS-GAN 和 SD-GAN 的聚类性能。

。具体而言，DEC、IDEC 和 DCN 使用深度神经网络来提取不同的潜在表示，以用于聚类。与 K-means++ 直接搜索数据样本的聚类中心或质心相比，基于深度学习的方案通常表现更好。FS-GAN 采用了一种新颖的方法：先生成带有伪标签的合成数据样本，再基于有标签数据集训练分类器。从某种意义上说，这种方法弥补了因缺少有标签数据集而导致的性能不佳；当伪标记合成数据样本的质量较高时，还能提升性能。

为了将 FS-GAN 与最先进的自监督学习方案进行比较，我们考虑了 DEC 算法近期的一个扩展版本，称为半监督 DEC（SDEC）[76]，。该算法将从有标签数据样本中获得的半监督信息融入 DEC，以进一步提升聚类性能。我们将 SDEC 用作前置任务，然后把伪标签数据样本纳入自监督 GAN算法中训练模型，并将这种方法称为 SD-GAN。请注意，与 FS-GAN 相比，SD-GAN 需要有标签数据集来学习先验信息，以引导学习过程；它还需要额外的神经网络模型来执行数据聚类和伪标签生成。在图 6 中，我们比较了不同训练轮次下采用不同设置的 FS-GAN 和 SD-GAN。可以观察到，FS-GAN-I 的表现优于 SD-GAN，而 FS-GAN-II 在 NMI 方面的表现与其相近。

为比较不同设置下模型训练参数对 FS-GAN 计算负载的影响，我们在表5中展示了固定硬件和软件环境下，FS-GAN 在各种性能指标下进行 50 轮模型训练时每轮计算的运行时间。我们主要评估四个关键参数的影响，包括本地数据集的数量、每个数据集的大小、两次连续全局模型协调之间的本地训练步数（N,n, E），以及小批量大小 B。同时还展示了在单个集中式数据集上的学习结果。可以看出，与集中式设置相比，FS-GAN 的学习时间和性能都有所下降，这是因为不再需要模型聚合过程。可以观察到，计算负载对数据集大小最为敏感。与方案 C-I 相比，方案 C-II 的训练计算负载始终略高。这是因为在方案 C-I 中，生成器和判别器都参与协调，从而加快模型训练的收敛速度，并降低总体计算时间。NMI 和 ACC 指标进一步体现了方案 C-I 相比方案 C-II 的性能优势。具体而言，在训练轮数固定时，C-I 的聚类性能始终比 C-II 更准确。C-I 和 C-II 在兰德指数方面的表现相近。这是因为兰德指数衡量的是所有聚类结果中正确解的综合表现，而不是像 NMI 和 ACC 那样衡量各个类别中的最高表现。因此，不同方案在兰德指数上的性能差异始终小于其他两个性能指标上的差异。

在图 4 中，我们展示了使用 FS-GAN 对表 2 中的 10 项服务进行分类时的聚类结果。再次强调，FS-GAN 的性能优于 IDEC 和 K-means++。需要注意的是，由于 K-means++ 直接依据质心分离数据样本，因此往往会将大多数样本聚类到单个簇或数量有限的几个簇中。

表 6：不同分类器级别下的性能 λ

多样性参数		$\lambda = 0.1$	$\lambda = 0.5$	$\lambda = 5$
FS-GAN（C-I）	RI	0.89	0.89	**0.90**
	NMI	0.60	0.60	**0.64**
	ACC	0.57	0.58	**0.60**
FS-GAN（C-II）	RI	0.88	**0.89**	0.88
	NMI	0.43	**0.47**	0.40
	ACC	0.52	**0.57**	0.51

在图 5 中，我们比较了不同模型参数下基于 NMI 的收敛性能。可以观察到，C-I 的收敛性能始终优于 C-II。这一结果与之前的观察一致。当本地数据集的大小相同时，收敛性能与其规模密切相关。具体而言，数据集规模越大，模型训练过程的收敛速度就越快

表 7：不同场景下的合成质量

指标	F-GAN(C-I)		F-GAN(C-II)		FS-GAN(C-I)		FS-GAN(C-II)	
	Three-G	Six-G	Three-G	Six-G	Three-G	Six-G	Three-G	Six-G
KL散度	5.643	0.360	0.226	0.224	0.091	**0.043**	0.188	0.099
JS散度	0.086	0.062	0.054	0.051	0.022	**0.011**	0.041	0.025
Wasserstein距离	0.647	0.619	1.013	0.578	0.353	**0.286**	0.484	0.387

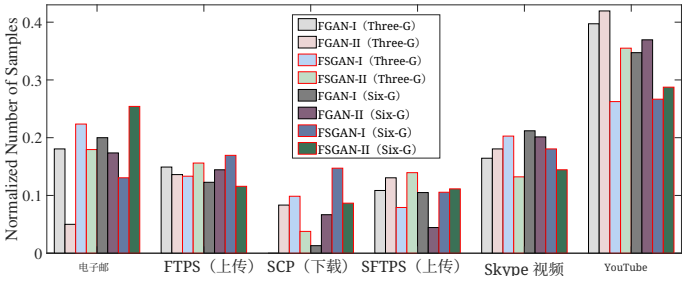


图 7：不同方案下分配标签的分布。

。请注意，NMI 的波动是由于本地数据合成过程中生成器与判别器之间的交互，以及全局模型协调所致。

如前所述，分类器有助于提高生成器所生成合成样本的多样性。同时，它也会干扰生成器与判别器之间的对抗训练过程。我们通过调整方程 (2) 中的多样性超参数 λ，研究分类器对数据聚类性能的影响。直观地说，当分类器带来的惩罚大于判别器带来的惩罚时，即使生成的样本可能不符合真实数据的分布，生成器也会受到鼓励，生成更多样化的样本。表 6 展示了 FS-GAN 在 λ取不同值时的聚类性能。我们可以观察到，λ 的值过大或过小通常都无法取得最佳性能。最优的 λ 因模型设置而异。如何找到能够在分类与判别之间实现适当平衡的最优λ，将留待未来研究。

5.3 流量数据合成的性能评估

如前所述，FS-GAN 不仅是一种流量分类方法。它还能够从去中心化数据集中学习，并生成能够反映与未知服务相关的真实数据分布的合成数据样本。本小节将评估 FS-GAN 在这方面的表现。

我们将 FS-GAN 的数据合成能力与采用联邦学习的多生成器生成对抗网络的各种扩展进行比较，并将其称为 F-GAN。具体而言，我们比较以下 4 种不同方案：

1) F-GAN（三生成器）：每个数据集使用 3 个生成器，不使用分类器。2) FS-GAN（三生成器）：每个数据集使用 3 个生成器。3) F-GAN（六生成器）：每个数据集使用 6 个生成器，不使用分类器。4) FS-GAN（六生成器）：每个数据集使用 6 个生成器。

如前所述，基于生成对抗网络的解决方案面临的一个关键挑战是可能触发模式崩溃问题；在这种情况下，不同的生成器无论输入如何，都只会生成种类有限的样本。在图 7 中，我们评估了不同方案所生成合成数据样本的多样性。我们比较了各方案在使用相同输入时，为不同服务生成的默认数量的合成样本。可以观察到，不同生成器倾向于生成数量不同的合成样本。例如，在六种服务中，SCP（下载）是最不受欢迎的待合成服务，生成的合成数据样本数量也最少；例如，F-GAN（C-I，六生成器）生成的全部合成样本中，归属于此服务类别的不到 0.02%。然而，我们观察到，FS-GAN，尤其是采用六个生成器的 C-I方案，为不同服务生成的样本数量相对均衡。这意味着 FS-GAN 在平衡合成数据样本方面表现最佳。

为了量化合成样本的多样性，我们使用统计距离指标来衡量合成数据样本的分布与真实服务流量数据分布之间的差异度。我们考虑三种距离指标：Kullback-Leibler（KL）散度、Jensen-Shannon（JS）散度 [69], 和 Wasserstein（W）距离[77],，考虑两个随机变量 p和 q，其分布分别为 $p(x)$ 和 $q(x)$。表 7 给出了 FS-GAN 和 F-GAN 下的距离结果。FS-GAN 在合成数据与真实数据之间实现了最小距离，其中 FS-GAN-I 下的 JS散度 低至 0.011，几乎是 F-GAN 所达到的最低距离结果的五分之一。这意味着所提出的 FS-GAN 非常适合对包含混合服务的高度异构流量进行分类和合成。

6 结论

本文提出了FS-GAN，这是一种联邦自监督学习框架，能够在大量去中心化数据集上自动识别、分类和合成不同类型的流量。FS-GAN由三个成分组成：本地数据合成、全局模型协调和自监督学习。我们采用了类联邦学习方法，并证明联合训练的全局模型能够同时最小化所有数据集上真实数据分布与合成数据样本分布之间的JSD，同时最大化不同生成器所创建的数据样本分布之间的JSD。仿真结果表明，与当前最先进的数据聚类方案相比，FS-GAN能够实现显著的性能提升；在数据合成多样性方面，其性能几乎是联邦生成对抗网络方案的五倍。

参考文献

[1] K. Pentikousis、Y. Wang和W. Hu，“Mobileflow：面向软件定义移动网络”，IEEE通信杂志，第51卷，第7期，第44–53页，2013年7月。[2] V. Yazıcı、U. C. Kozat和M. O. Sunay，“面向5G网络架构的新型控制平面：以统一切换、移动性和路由管理为例”，IEEE通信杂志，第52卷，第11期，第76–85页，2014年11月。[3] M. Agiwal、A. Roy和N. Saxena，“下一代5G无线网络：全面综述”，IEEE通信调查与教程，第18卷，第3期，第1617–1655页，2016年2月。[4] A. Al-Fuqaha、M. Guizani、M. Mohammadi、M. Aledhari和M. Ayyash，“物联网：赋能技术、协议与应用综述”，IEEE通信调查与教程，第17卷，第4期，第2347–2376页，2015年6月。[5] Y. Xiao、M. Krunz和T. Shu，“面向大规模物联网的多运营商网络共享”，IEEE通信杂志，第57卷，第4期，第96–101页，2019年4月。[6] T. Luettel、M. Himmelsbach和H. Wuensche，“自主地面车辆——概念及通向未来的路径”，IEEE会议录，第100卷，第1831–1839页，2012年4月。[7] “国际电信联盟2019年网络2030”，日内瓦，2018年7月。[在线]。可访问：

h

t

t

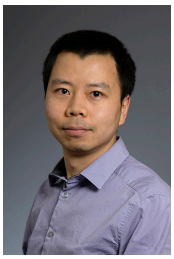
p

s://www.itu.int/en/ITU-T/focusgroups/net2030/Pages/default.aspx

[8] I. FG-NET2030，“网络2030的新业务与能力：描述、技术差距与性能目标分析”，2019年10月。[9] Y. Xiao和M. Krunz，“触觉互联网中面向能效雾计算的分布式优化”，IEEE精选通信领域期刊，第36卷，第11期，第2390–2400页，2018年11月。[10] G. Shi、Y. Xiao、Y. Li和X. Xie，“从语义通信到语义感知网络：模型、架构与开放问题”，IEEE通信杂志，第59卷，第8期，第44–50页，2021年8月。[11] Y. Xiao、Z. Sun、G. Shi和D. Niyato，“基于模仿学习的隐式语义感知通信网络：多层表示与协同推理”，IEEE精选通信领域期刊，第41卷，第3期，2023年3月。[12] Y. Xiao、G. Shi、Y. Li、W. Saad和H. V. Poor，“迈向6G中的自学习边缘智能”，IEEE通信杂志，第58卷，第12期，2020年12月。[13] M. Reiter和R. Steinberg，“通信网络互补部分的远期合约”，载于IEEE INFOCOM，圣迭戈，美国，2010年3月，第1–9页。[14] Y. Xiao和M. Krunz，“AdaptiveFog：智能交通系统中雾计算的建模与优化框架”，IEEE移动计算汇刊，第21卷，第12期，第4187–4200页，2022年12月。[15] X. Wang、S. Chen和J. Su，“App-net：用于加密移动流量分类的混合神经网络”，载于IEEE INFOCOM，虚拟会议，2020年7月，第424–429页。[16] M. H. Almannaa、M. Elhenawy和H. A. Rakha，“一种面向交通系统应用的新型监督聚类算法”，IEEE智能交通系统汇刊，第21卷，第1期，第222–232页，2020年1月。[17] J. Zhang、F. Li、F. Ye和H. Wu，“面向基于深度学习的流量分类器更新的自主未知应用过滤与标记”，载于IEEE INFOCOM，虚拟会议，2020年7月，第397–405页。[18] X. Liu、F. Zhang、Z. Hou、L. Mian、Z. Wang、J. Zhang和J. Tang，“自监督学习：生成式还是对比式”，IEEE知识与数据工程汇刊，2021年。[19] X. Wang、S. Zhang、Z. Qing、Y. Shao、C. Gao和N. Sang，“用于半监督时序动作提议的自监督学习”，载于IEEE/CVF计算机视觉与模式识别会议论文集，2021年6月。[20] N. Araslanov和S. Roth，“用于适应语义分割的自监督增强一致性”，载于IEEE/CVF计算机视觉与模式识别会议论文集，2021年6月。[21] Y. LeCun，“自监督学习”，AAAI 2020会议主题演讲，2020年2月。[22] S. B. Baker、W. Xiang和I. Atkinson，“智慧医疗中的物联网：技术、挑战与机遇”，IEEE Access，第5卷，第26521–26544页，2017年。[23] 基于机器学习方法的多提供商区基础设施网络流量分类问题”，布德瓦，2021年6月。[24] Z. Zhou、X. Chen、E. Li、L. Zeng、K. Luo 和 J. Zhang，“边缘智能：利用边缘计算铺设人工智能的最后一公里”，IEEE会议录，第107卷，2019年，第1738–1762页。[25] W. Y. B. Lim、N. C. Luong、D. T. Hoang、Y. Jiao、Y.-C. Liang、Q. Yang、D. Niyato 和 C. Miao，“移动边缘网络中的联邦学习：全面综述”，IEEE通信调查与教程，第22卷，第3期，第2031–2063页，2020年。[26] G. Draper-Gil、A. H. Lashkari、M. S. I. Mamun 和 A. A. Ghorbani，“利用时间相关特征表征加密流量和虚拟专用网络流量”，载于 ICISSP，葡萄牙，2016年2月，第407–414页。[27] P. Tang、Y. Dong 和 S. Mao，“利用粒进行在线流量分类”，载于 IEEE INFOCOM，虚拟会议，2020年7月，第1135–1140页。[28] C. Liu、L. He、G. Xiong、Z. Cao 和 Z. Li，“Fs-net：用于加密流量分类的流序列网络”，载于 IEEE INFOCOM，巴黎，法国，2019年4月，第1171–1179页。[29] E. Arestr¨om 和 N. Carlsson，“利用多重分形特征对加密流流进行早期在线分类”，载于 IEEE INFOCOM，巴黎，法国，2019年4月，第84–89页。[30] E. Liang、H. Zhu、X. Jin 和 I. Stoica，“神经数据包分类”，《ACM Sigcomm会议录》，中国北京，2019年8月。[31] R. Li、X. Xiao、S. Ni、H. Zheng 和 S. Xia，“用于网络流量分类的字节片段神经网络”，载于 IEEE ACM IWQoS，加拿大艾伯塔，2018年6月，第1–10页。[32] A. Nascita、A. Montieri、G. Aceto、D. Ciuonzo、V. Persico 和 A. Pescap´e，“可解释人工智能与移动流量分类相遇：理解并改进多模态深度学习架构”，IEEE网络与服务管理汇刊，2021年。[33] A. M. Sadeghzadeh、S. Shiravi 和 R. Jalili，“对抗性网络流量：评估基于深度学习的网络流量分类鲁棒性的探索”，IEEE网络与服务管理汇刊，第18卷，第2期，第1962–1976页，2021年。[34] G. Aceto、D. Ciuonzo、A. Montieri 和 A. Pescap´e，“Distiller：通过多模态多任务深度学习进行加密流量分类”，网络与计算机应用杂志，第183卷，文章编号102985，2021年。[35] O. Aouedi、K. Piamrat 和 B. Parrein，“用于网络流量分类的基于集成的深度学习模型”，IEEE网络与服务管理汇刊，第1–12页，2022年。[36] D. Jin、Y. Lu、J. Qin、Z. Cheng 和 Z. Mao，“Swiftids：基于 LightGBM 和并行入侵检测机制的实时入侵检测系统”，计算机与安全，第97卷，第1–12页，2020年。[37] Q. Yang、Y. Liu、T. Chen 和 Y. Tong，“联邦机器学习：概念与应用”，ACM智能系统与技术汇刊，第10卷，第2期，第1–19页，2019年1月。[38] H. Wen、Y. Wu、C. Yang、H. Duan 和 S. Yu，“面向无线通信的统一联邦学习框架：兼顾隐私、效率与安全”，载于 IEEE INFOCOM，虚拟会议，2020年7月，第653–658页。[39] H. Wang、Z. Kaplan、D. Niu 和 B. Li，“利用强化学习优化非独立同分布数据上的联邦学习”，载于 IEEE INFOCOM，虚拟会议，2020年7月，第1698–1707页。[40] Q. Li、Z. Wen 和 B. He，“联邦学习系统：数据隐私与保护的愿景、炒作与现实”，ArXiv：1907.09693，2019年7月。[41] T. Li、A. K. Sahu、A. Talwalkar 和 V. Smith，“联邦学习：挑战、方法与未来方向”，IEEE信号处理杂志，第37卷，第3期，第50–60页，2020年5月。[42] J. Shi、H. Zhao、M. Wang 和 Q. Tian，“基于联邦学习的信号识别”，载于 IEEE INFOCOM，虚拟会议，2020年7月，第1105–1110页。[43] X. Li、K. Huang、W. Yang、S. Wang 和 Z. Zhang，“关于非独立同分布数据上的 FedAvg 收敛性”，ICLR，新奥尔良，2019年4月。[44] X. Zhu、N. Shu、H. Wang 和 T. Wu，“基于联邦学习的分布式流量分类模型”，载于 IEEE大数据计算与通信国际会议（BigCom），2021年，第75–81页。[45] S. Wang、T. Tuor、T. Salonidis、K. K. Leung、C. Makaya、T. He 和 K. Chan，“资源受限边缘计算系统中的自适应联邦学习”，IEEE精选通信领域期刊，第37卷，第6期，第1205–1221页，2019年6月。

[46] B. Luo、X. Li、S. Wang、J. Huang 和 L. Tassiulas，“经济高效的联邦学习设计”，载于 IEEE INFOCOM，虚拟会议，2021年。[47] Y. Xiao、Y. Li、G. Shi 和 H. V. Poor，“物联网网络中联邦边缘智能的资源效率优化”，载于无线通信与信号处理国际会议（WCSP），中国南京，2020年10月。[48] Y. Xiao、X. Zhang、Y. Li、G. Shi、M. Krunz、D. N. Nguyen 和 D. T. Hoang，“异构联邦边缘智能的时效敏感学习”，已被 IEEE 移动计算汇刊接收，2023 年 1 月。[49] Y. Xiao 、 Y. Li 、 G. Shi 和 H. V. Poor，“空中推理：一种隐式语义通信架构”，载于 IEEE ICC网络与系统数据驱动智能研讨会论文集，韩国首尔，2022年5月。[50] Y. Xiao、X. Zhang、Y. Li、G. Shi 和 T. Basar，“策略性语义通信的率失真理论”，载于 IEEE信息论研讨会论文集，印度孟买，2022年11月。[51] L. Zhang、J. Xu、P. Vijayakumar、P. K. Sharma 和 U. Ghosh，“基于同态加密的物联网赋能医疗保健系统隐私保护联邦学习”，IEEE网络科学与工程汇刊，第1–17页，2022年。[52] S. Lu、Z. Gao、Q. Xu、C. Jiang、A. Zhang 和 X. Wang，“面向去中心化故障诊断并采用生物特征认证的类别不平衡隐私保护联邦学习”，IEEE工业信息学汇刊，第1–11页，2022年。[53] C.-L. Li、K. Sohn、J. Yoon 和 T. Pfister，“CutPaste：用于异常检测与定位的自监督学习”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2021年6月。[54] K. Sun、Z. Lin 和 Z. Zhu，“面向带有少量带标签节点的图的图卷积网络多阶段自监督学习”，载于 AAAI人工智能会议论文集，第34卷，第04期，美国纽约，2020年2月。[55] A. Saeed、F. D. Salim、T. Ozcelebi 和 J. Lukkien，“面向嵌入式智能的多传感器表征联邦自监督学习”，IEEE物联网期刊，第8卷，第2期，第1030–1040页，2021年。[56] J. Z. Bengar、J. van de Weijer、B. Twardowski 和 B. Raducanu，“减少标签标注工作量：自监督与主动学习相结合”，载于 IEEE/CVF国际计算机视觉会议论文集，2021年，第1631–1639页。[57] Y. Zhang、J. Wang、Y. Chen、H. Yu 和 T. Qin，“用于无监督异常检测的自适应记忆网络与自监督学习”，IEEE知识与数据工程汇刊，第1–13页，2022年。[58] Y. Shi、N. Siddharth、P. Torr 和 A. R. Kosiorek，“用于自监督学习的对抗掩码”，载于第39届国际机器学习会议论文集，机器学习研究论文集系列，第162卷，2022年7月17–23日，第20 026–20 040页。[59] I. Goodfellow、J. Pouget-Abadie、M. Mirza、B. Xu、D. Warde-Farley、S. Ozair、A. Courville 和 Y. Bengio，“生成对抗网络”，载于 NIPS，加拿大蒙特利尔，2014年12月，第2672–2680页。[60] L. Yu、W. Zhang、J. Wang 和 Y. Yu，“SeqGAN：采用策略梯度的序列生成对抗网络”，载于 AAAI人工智能会议，美国加利福尼亚州旧金山，2017年2月。[61] X. Chen、Y. Duan、R. Houthooft、J. Schulman、I. Sutskever 和 P. Abbeel，“InfoGAN：通过信息最大化生成对抗网络实现可解释表征学习”，载于 NIPS，西班牙巴塞罗那，2016年12月，第2172–2180页。[62] Q. Hoang、T. D. Nguyen、T. Le 和 D. Phung，“MGAN：使用多个生成器训练生成对抗网络”，载于 ICLR，加拿大温哥华，2018年5月。[63] A. Ghosh、V. Kulharia、V. P. Namboodiri、P. H. Torr 和 P. K. Dokania，“多智能体多样化生成对抗网络”，载于 ICPR，中国北京，2018年8月，第8513–8521页。[64] H. Zhang、S. Xu、J. Jiao、P. Xie、R. Salakhutdinov 和 E. P. Xing，“Stackelberg生成对抗网络：通过多生成器架构实现可证明的极小极大均衡”，2018年11月。[65] B. McMahan、E. Moore、D. Ramage、S. Hampson 和 B. A. y Arcas，“基于去中心化数据的深度网络通信高效学习”，载于 AISTATS，第54卷，美国劳德代尔堡，2017年4月，第1273–1282页。[66] “Release 16”，2020年7月。 [在线]。可查阅：https://www.3gpp. org/release-16

[67] N. Mangrulkar、A. Bhagat Patil 和 A. Pande，《网络攻击及其检测机制：综述》，《国际计算机应用杂志》，第 90 卷，2014年2月。[68] K. Duan、S. S. Keerthi、W. Chu、S. K. Shevade 和 A. N. Poo，《通过二元分类器的 soft-max 组合进行多类别分类》，载于多分类器系统，柏林、海德堡，2003年6月，第 125–134 页。[69] J. Lin，《基于香农熵的散度度量》，《*IEEE* 信息论汇刊》，第 37 卷，第 1 期，第 145–151 页，1991年9月。[70] M. Rasouli、T. Sun 和 R. Rajagopal，《FedGAN：用于分布式数据的联邦生成对抗网络》，*ArXiv* 预印本 *arXiv: 2006.07228*，2020年6月。[71] Y. Zhao、M. Li、L. Lai、N. Suda、D. Civin 和 V. Chandra，《使用非独立同分布数据的联邦学习》，*ArXiv* 预印本 *arXiv: 1806.00582*，2018年6月。[72] T. Ryffel、A. Trask、M. Dahl、B. Wagner、J. Mancuso、D. Rueckert 和 J. Passerat-Palmbach，《一种用于隐私保护深度学习的通用框架》，*ArXiv* 预印本 *arXiv:1811.04017*，2018年。[73] J. Xie、R. Girshick 和 A. Farhadi，《用于聚类分析的无监督深度嵌入》，载于*ICML*，美国纽约，2016年6月，第 478–487 页。[74] X. Guo、L. Gao、X. Liu 和 J. Yin，《采用局部结构保持的改进型深度嵌入聚类》，载于*IJCAI*，澳大利亚墨尔本，2017年8月，第 1753–1759 页。[75] B. Yang、X. Fu、N. D. Sidiropoulos 和 M. Hong，《迈向适用于 k 均值的空间：深度学习与聚类的同步进行》，载于*ICML*，澳大利亚悉尼，2017年8月，第 3861–3870 页。[76] Y. Ren、K. Hu、X. Dai、L. Pan、S. C. Hoi 和 Z. Xu，《半监督深度嵌入聚类》，《神经计算》，第 325 卷，第 121–130 页，2019年。[77] S. Vallender，《计算直线上概率分布之间的 Wasserstein距离》，《概率论及其应用》，第 18 卷，第 4 期，第 784–786 页，1973年2月。[78] J. Lin，《基于香农熵的散度度量》，*IEEE Trans. Inf. Theor.*，第 37 卷，第 1 期，1991年1月。



Yong Xiao （IEEE高级会员）于2002年在中国武汉的中国地质大学获得电气工程学士学位，2006年在香港科技大学获得电信硕士学位，并于2012年在新加坡南洋理工大学获得电气与电子工程博士学位。他现任华中科技大学（HUST）电子信息与通信学院教授，工作地点在中国武汉。他还在中国深圳的鹏城实验室和中国广州的琶洲实验室（黄埔）任职。他担任IMT-2030（6G推进组）网络智能组副组长，并任华中科技大学5G垂直行业创新实验室副主任。他的研究兴趣包括机器学习、博弈论、分布式优化，以及它们在语义通信和语义感知网络、云/雾/移动边缘计算、绿色通信系统、无线通信网络和物联网（IoT）中的应用。



RongXia 于2018年和2022年分别在中国武汉的华中科技大学获得信息与通信工程学士学位和硕士学位。她的研究兴趣包括联邦边缘智能和边缘计算网络。



Yingyu Li （IEEE会员）于2012年和2018年分别在中国西安的西安电子科技大学获得电子信息工程工学学士学位和电路与系统博士学位。2014年至2016年，她在美国得克萨斯州休斯敦大学电子与计算机工程系担任研究学者。2018年起，她在华中科技大学电子信息与通信学院从事博士后研究

至2021年。她现任中国地质大学（武汉）机械与电子信息学院副教授。她的研究兴趣包括语义通信、边缘智能、绿色通信网络和物联网。



Dinh Thai Hoang （IEEE高级会员）目前是澳大利亚悉尼科技大学电气与数据工程学院的教职人员。他于2016年在新加坡南洋理工大学获得计算机科学与工程博士学位。他的研究兴趣包括新兴无线通信与网络领域，尤其是网络、边缘计算和网络安全中的机器学习应用。他曾获得多项奖项，包括

澳大利亚研究理事会和IEEE TCSC可扩展计算卓越奖（早期职业研究人员）。他是《IEEE无线通信汇刊》《IEEE认知通信与网络汇刊》《IEEE车辆技术汇刊》的编辑，也是《IEEE通信调查与教程》的副编辑。



Guangming Shi （IEEE会士）于1988年和2002年分别获得西安电子科技大学计算机控制专业理学硕士学位和电子信息技术专业博士学位。2018年至2022年，他担任西安电子科技大学副校长。目前，他担任鹏城实验室副院长，并任西安电子科技大学人工智能学院教授。他是IEEE会士、IEEE电路与系统学会西安分会主席、计算机协会和中国计算机学会高级会员、中国

电子学会会士及工程技术学会会士。2012年，他获教育部长江学者讲座教授称号。2017年，他获得国家自然科学奖二等奖。他的研究兴趣包括人工智能、语义通信和人机交互。



DusitNiyato （IEEE会士）是新加坡南洋理工大学计算机科学与工程学院教授。他于1999年在泰国蒙库特国王科技大学叻甲挽校区（KMITL）获得工学学士学位，并于2008年在加拿大曼尼托巴大学获得电气与计算机工程博士学位。他的研究兴趣包括物联网、机器学习和激励机制设计。



Diep N. Nguyen （IEEE高级会员）于2008年在美国加利福尼亚州圣迭戈拉霍亚的加利福尼亚大学圣迭戈分校（UCSD）获得电气与计算机工程工学硕士学位，并于2013年在美国亚利桑那州图森的亚利桑那大学（UA）获得电气与计算机工程博士学位。他目前是澳大利亚新南威尔士州悉尼悉尼科技大学（UTS）工程与信息技术学院的教职人员。在加入UTS之前，他曾担任

麦考瑞大学（澳大利亚新南威尔士州麦考瑞园区）的DECRA研究员，同时也是博通公司（美国加利福尼亚州圣何塞）和ARCON公司（美国马萨诸塞州波士顿）的技术人员，并为联邦航空管理局（美国华盛顿特区）提供无人机和飞机转弯检测方面的咨询，也为美国空军研究实验室（美国俄亥俄州赖特-帕特森空军基地）提供抗干扰方面的咨询。他的研究兴趣包括计算机网络、无线通信和机器学习应用，尤其关注系统性能以及安全与隐私。Nguyen博士曾获得LG电子、加州大学圣迭戈分校、亚利桑那大学、美国国家科学基金会和澳大利亚研究理事会颁发的多项奖项。他目前担任编辑，并担任《IEEE移动计算汇刊》、《IEEE通信调查与教程》（COMST）、《IEEE通信学会开放期刊》和《科学报告》（《自然》旗下）的副编辑。



Marwan Krunz （IEEE 会士）是亚利桑那大学的校董教授。他担任电气与计算机工程领域的Kenneth VonBehren冠名教授席位，同时也是计算机科学教授。他负责领导宽带无线接入与应用中心（BWAC）。该中心是由多所大学与美国国家科学基金会及业界共同设立的中心，专注于下一代无线技术。他还兼任悉尼科技大学教授。此前，他曾担任 Connection One 的站点主任，

由美国国家科学基金会/业界资助、由五所大学和 20+ 家业界附属机构组成的中心。Krunz博士的研究领域包括无线通信、网络和安全，近期重点是运用人工智能和机器学习技术进行协议自适应、资源管理和信号情报。他已发表320余篇期刊论文和经同行评审的会议论文，并以发明人身份列名于12项专利。他最新的H指数为60。他是IEEE会士、亚利桑那州工程学院院士，也是IEEE通信学会杰出讲师（2013—2015年）。他获得了美国国家科学基金会职业奖。他曾担任《IEEE移动计算汇刊》的主编，也曾担任多种IEEE期刊的编辑。他曾担任INFOCOM'04、SECON'05、WoWMoM'06和Hot Interconnects 9的技术程序委员会主席。他曾担任WiOpt 2016大会副主席和WiSec'12大会共同主席。Krunz博士曾担任两家初创公司的首席科学家/技术专家，这些公司专注于5G及后续无线系统。

附录 A 命题 1 的证明

证明：根据方程 (2) 中的多参与者零和博弈，分类器 C_d 试图最大化 softmax 量 $V(C)$，其中

$$\begin{aligned} V(C) &= \int_{\boldsymbol{x}} \sum_{m=1}^{M} \pi_m P_{G_d^m}(\boldsymbol{x}) \log C_d^m(\boldsymbol{x}) dx \\ &= \int_{\boldsymbol{x}} \pi_1 P_{G_d^1}(\boldsymbol{x}) \log(1 - \sum_{m=2}^{M} C_d^m(\boldsymbol{x})) dx \\ &\quad + \int_{\boldsymbol{x}} \sum_{m=2}^{M} \pi_m P_{G_d^m}(\boldsymbol{x}) \log C_d^m(\boldsymbol{x}) dx. \end{aligned} \tag{13}$$

为了得到最优分类器，我们首先计算关于 C_d^m，$m = 2, 3, ..., M$ 的泛函导数，得到：

$$\frac{\partial V(C)}{\partial C_d^m(\boldsymbol{x})} = \frac{\pi_m P_{G_d^m}(\boldsymbol{x})}{C_d^m(\boldsymbol{x})} - \frac{\pi_1 P_{G_d^1}(\boldsymbol{x})}{C_d^m(\boldsymbol{x})}. \tag{14}$$

然后，将上述方程对 $m \in \{2, 3, ..., M\}$ 置为零，可得：

$$\frac{\pi_1 P_{G_d^1}(\boldsymbol{x})}{C_d^1(\boldsymbol{x})} = \frac{\pi_m P_{G_d^m}(\boldsymbol{x})}{C_d^m(\boldsymbol{x})}, \forall m \in \{2, 3, ..., M\}. \tag{15}$$

鉴于分类器使用的所有训练样本必定都来自生成器，即 $\sum_{m=1}^{M} C_d^m(\boldsymbol{x}) = 1$，我们可以得到分类器 C_d^m的最优解，即：

$$C_d^{m*}(\boldsymbol{x}; \theta_d) = \frac{\pi_m P_{G_d^m}(\boldsymbol{x})}{\sum_{i=1}^{M} \pi_i P_{G_d^i}(\boldsymbol{x})}, m \in \{1, 2, ..., M\}. \tag{16}$$

证明至此完成。 □

附录 B 命题 2 的证明

证明：设最优判别器为 $D_d{}^*$。这里，我们直接采用 [59] $D_d{}^* = \frac{P_{data}^d \boldsymbol{x}}{}$中的结论，即 $() P_{data}^d(\boldsymbol{x}) + P_{model}^d(\boldsymbol{x})$。将生成器的量函数记为 $V(\boldsymbol{G})$。根据式 (2) 中的多参与者零和博弈，给定上述最优判别器 $D_d{}^*$ 和命题 1 中的分类器 $C_d{}^*$，生成器试图最小化：

$$\begin{aligned} V(\boldsymbol{G}(\boldsymbol{x}; \omega_d)) =& \mathbb{E}_{\boldsymbol{x} \sim P_{data}^d}[\log \frac{P_{data}^d(\boldsymbol{x})}{P_{data}^d(\boldsymbol{x}) + P_{model}^d(\boldsymbol{x})}] \\ &+ \mathbb{E}_{\boldsymbol{x} \sim P_{model}^d}[\log \frac{P_{model}^d(\boldsymbol{x})}{P_{data}^d(\boldsymbol{x}) + P_{model}^d(\boldsymbol{x})}] \\ &- \lambda \{ \sum_{m=1}^{M} \pi_m \mathbb{E}_{\boldsymbol{x} \sim P_{G_d^m}}[\log \frac{\pi_m P_{G_d^m}(\boldsymbol{x})}{\sum_{i=1}^{M} \pi_i P_{G_d^i}(\boldsymbol{x})}] \}. \end{aligned} \tag{17}$$

（17）右侧的前两项与标准 GAN 中相同，[59]，其等于 $2 \cdot \text{JSD}(P_{data}^d \| P_{model}^d) - \log 4$。这里，我们重点关注该方程的最后一项，以阐明辅助分类器的作用。

根据标准 GAN 中的结论，我们可以将（17）表述如下：

$$\begin{aligned} V(\boldsymbol{G}(\boldsymbol{x}; \omega_d)) =& 2 \cdot \text{JSD}(P_{data}^d \| P_{model}^d) - \log 4 \\ &- \lambda \{ \sum_{m=1}^{M} \pi_m \mathbb{E}_{\boldsymbol{x} \sim P_{G_d^m}}[\log \frac{\pi_m P_{G_d^m}(\boldsymbol{x})}{\sum_{i=1}^{M} \pi_i P_{G_d^i}(\boldsymbol{x})}] \} \end{aligned} \tag{18}$$

为了更清楚地阐述证明，我们用 ^ ^$-\lambda\{V\}$ 表示式 (18) 中的最后一项，并令 $P_G(\sum_{m=1}^{M} \pi_m P_{G_d^m}$ $\boldsymbol{x}$) ()。于是，我们有：

$$\hat{V} = \sum_{m=1}^{M} \pi_m \mathbb{E}_{\boldsymbol{x} \sim P_{G_d^m}}[\log \frac{\pi_m P_{G_d^m}(\boldsymbol{x})}{\sum_{i=1}^{M} \pi_i P_{G_d^i}(\boldsymbol{x})}] \tag{19a}$$

$$= \sum_{m=1}^{M} \pi_m \mathbb{E}_{\boldsymbol{x} \sim P_{G_d^m}}[\log \frac{P_{G_d^m}(\boldsymbol{x})}{\hat{P}_G(\boldsymbol{x})}] + \sum_{m=1}^{M} \pi_m \log \pi_m \tag{19b}$$

$$\begin{aligned} &= \pi_1 \mathbb{E}_{\boldsymbol{x} \sim P_{G_d^1}}[\log \frac{P_{G_d^1}(\boldsymbol{x})}{\hat{P}_G(\boldsymbol{x})}] + ... \\ &\quad + \pi_M \mathbb{E}_{\boldsymbol{x} \sim P_{G_d^M}}[\log \frac{P_{G_d^M}(\boldsymbol{x})}{\hat{P}_G(\boldsymbol{x})}] + \sum_{m=1}^{M} \pi_m \log \pi_m. \end{aligned} \tag{19c}$$

根据 KL 散度的标准定义，上式可以表示为多个 KL 散度与 $\boldsymbol{\pi}$的负熵之和，即：

$$\begin{aligned} \hat{V} &= \pi_1 \cdot \text{KL}(P_{G_d^1}(\boldsymbol{x}) \| \hat{P_G}) + ... \\ &\quad + \pi_M \cdot \text{KL}(P_{G_d^M}(\boldsymbol{x}) \| \hat{P_G}) + \sum_{m=1}^{M} \pi_m \log \pi_m \end{aligned} \tag{20a}$$

$$= \sum_{m=1}^{M} \pi_m \text{KL}(P_{G_d^m}(\boldsymbol{x}) \| \hat{P_G}) + \sum_{m=1}^{M} \pi_m \log \pi_m \tag{20b}$$

$$= \text{JSD}_{\pi_1, \pi_2, ..., \pi_M}(P_{G_d^1}, P_{G_d^1}, ..., P_{G_d^M}) + \sum_{m=1}^{M} \pi_m \log \pi_m, \tag{20c}$$

其中，$\text{JSD}_{\pi_1,\ \pi_2,\ ...\pi_M}(P_{G_d^1},\ P_{G_d^1},\ ...,\ P_{G_d^M})$ 是广义詹森-香农散度 [78]。因此，我们可以将式 (17) 中生成器的量函数改写为：

$$V(\boldsymbol{G}(\boldsymbol{x}; \omega_d)) = 2 \cdot \text{JSD}(P_{data}^d \| P_{model}^d) - \log 4 - \lambda \{\hat{V}\} \tag{21a}$$

$$\begin{aligned} =& 2 \cdot \text{JSD}(P_{data}^d \| P_{model}^d) \\ &- \lambda \cdot \text{JSD}_{\pi_1, \pi_2, ..., \pi_M}(P_{G_d^1}, P_{G_d^2}, ..., P_{G_d^M}) \\ &- \lambda \sum_{m=1}^{M} \pi_m \log \pi_m - \log 4. \end{aligned} \tag{21b}$$

忽略最后两个常数项，我们最终可以得到方程 (6) 中生成器的目标函数。因此，优化生成器等价于最小化 P_{data}^d 与 P_{model}^d 之间的 JD 散度，同时最大化这些生成器之间的差异。至此，证明完毕。

□

附录 C 定理 1 的证明

证明：为进一步简化命题 2 中生成器的目标函数，我们首先重新表述 (19) 中 V的方程ˆ：

$$\hat{V} = \sum_{m=1}^{M} \pi_m \mathbb{E}_{\boldsymbol{x}\sim P_{G_d^m}}[\log \frac{\pi_m P_{G_d^m}(\boldsymbol{x})}{\sum_{i=1}^{M} \pi_i P_{G_d^i}(\boldsymbol{x})}]. \tag{22}$$

ˆ可以看出， $V \leq 0$，1且仅当$\pi_m P_{G_d^m}(\boldsymbol{x})$ $P_{G_d^m}\ \boldsymbol{x}\ \neq$ () $\frac{\sum_{i=1}^{M} \pi_i P_{G_d^i}\ \boldsymbol{x}}{\sum_{i\neq m} \pi_i P_{G_d^i}\ \boldsymbol{x}} = 0 \rightarrow \forall i \neq m\ \ P_{G_d^i}\ \boldsymbol{x}\ = 0$ 时等号成立。因此，我们有() () ,() 。在这种情况下，不同的生成器充分分离且互不重叠。此外，我们可以在 $V = 0$下重写命题 2 中生成器的目标函数 ˆ：

$$\boldsymbol{G}_d{}^*(\boldsymbol{x};\omega_d) = \arg\min_{\boldsymbol{G}} 2 \cdot \mathrm{JSD}(P_{data}^d \| P_{model}^d). \tag{23}$$

这是一个简单得多的表述，我们可以直接得到其最优解，即: $P_{data}^d = P_{model}^d$。当假设 1 成立时，真实数据分布被限制为由 M 个成分 $p_d^m(\boldsymbol{x})\ m\ = 1, 2, ..., M$构成的混合分布，我们可以轻松得到式 (8b) 和 (8c) 中的结果，即:

$$P_{G_d^m}^*(\boldsymbol{x}) = p_d^m(\boldsymbol{x}),\ \forall m = 1, 2, ..., M, \tag{24a}$$

$$P_{model}^d(\boldsymbol{x}) = \sum_{m=1}^{M} \pi_m P_{G_d^m}^*(\boldsymbol{x}) = P_{data}^d(\boldsymbol{x}). \tag{24b}$$

到目前为止，我们已经证明，在多参与者零和博弈的均衡点处，生成器能够合成一种流量类型的混合分布，其分布与本地数据集中的真实流量相同。为了评估分类器的性能，我们将充分分离的最优解 $P_{G_d^m}$代入命题 1。可以看出，当 $\boldsymbol{x}$ 是从 $P_{G_d^m}(\boldsymbol{x})$中抽取时， C_d^{m*} 等于 1 ； 否则，C_d^{m*} 为 0。因此，最优分类器能够正确区分由不同生成器生成的所有样本。形式化地说，我们有:

$$C_d^{m*}(\boldsymbol{x}) = \begin{cases} 1, & x \sim P_{G_d^w},\ w = m \\ 0, & x \sim P_{G_d^w},\ w \neq m \end{cases} \quad \forall m = 1, 2, \ldots, M, \tag{25}$$

此外，当 $P_{data}^d = P_{model}^d$时，这意味着生成器正在合成高质量样本，分类器也将能够对本地数据集中的真实样本进行分类，例如，为不同的本地服务标注类别。这就完成了我们的证明。

□