可持续边缘计算：挑战与未来方向

Patricia Arroba[a]、Rajkumar Buyya[b]、Román Cárdenas[a]、José L. Risco-Martín[c]、José M. Moya [a]

a 综合系统实验室 (LSI) CCS—计算仿真中心 电信工程学院 马德里理工大学，西班牙 b 云计算与分布式系统 (CLOUDS) 实验室 计算与信息系统学院 墨尔本大学，澳大利亚 c 计算机体系结构与自动化系 马德里康普顿斯大学，西班牙

摘要

物联网（IoT）应用正向网络中注入越来越多的数据。许多旨在改善社会生活质量的此类应用对延迟要求极高，并向网络注入大量数据。物联网应用的这些需求促使边缘计算范式应运而生。目前，数据中心的能源消耗占全球能源使用量的 2% 至 3%。然而，这一趋势难以为继，因为将计算基础设施部署到更靠近网络边缘的位置，会带来一系列能源效率方面的挑战。 本文提出了我们对未来计算基础设施可持续性的方案，以实现：（i）节能且经济可行的部署；（ii）容错的自动化运维；以及（iii）协同资源管理，从而提高资源效率。我们指出了在靠近数据源的位置应用基于云的方法所面临的主要局限，并阐述了这些限制给边缘计算的可持续性带来的研究挑战。我们提出将两阶段浸没式冷却、形式化建模、机器学习和以能源为核心的联邦管理作为赋能边缘计算的技术。我们展示了在实现边缘基础设施可持续性方面取得的初步成果，以证明我们的方法对未来计算环境和部署的益处。

I. Introduction

边缘计算范式的出现，旨在解决数据密集型应用导致的数据速率不断增长的问题。物联网应用产生的数据量日益增多，这些应用被部署用于提升医疗、交通、能源和农业等各个领域与系统的效率和效能。如此海量的数据注入网络可能导致网络饱和，从而造成严重延迟。然而，这是一个关键问题，因为许多物联网应用对延迟有严格要求，以确保及时完成。

并能够提供准确的响应，因为它们涉及实时数据处理和决策。例如，支持物联网的设备可以远程监测患者的健康状况，并在发现任何异常时向医疗专业人员发出警报，从而实现更快速、更准确的治疗。边缘计算通过将计算资源置于更靠近数据源的位置，减少了需要发送到云进行处理的数据量，从而有助于解决这一问题。由于数据能够在更本地化的基础上进行处理和分析，因此带来的益处众多：(i)用户感知到的延迟得以降低，速度得到提升；(ii) 网络成本降低，所需带宽减少；(iii) 在出现连接问题时，可以提供本地可靠性[1]。

根据Gartner关于边缘计算的预测，[2], 到2025年，企业管理的数据中将有超过50%在云之外生成和处理。根据Cisco的说法，边缘计算技术涵盖多接入边缘计算（MEC）、云端和微型数据中心，以及雾计算（位于云与边缘设备之间）[3]。但是，撇开理论概念不谈，如今云服务提供商是如何触达边缘的？这个问题的答案是，超大规模运营商如今无处不在。拥有海量云资源的云服务提供商，其云资源在大多数情况下位于偏远地区，正通过多种方式向用户靠近。然而，首先，它们采取了一种风险较低的方法，即利用本地数据中心作为托管数据中心，托管其边缘节点（例如IT设备，从小型机架到整个数据机房）。因此，云、混合-云和边缘计算之间的界限正在逐渐模糊。

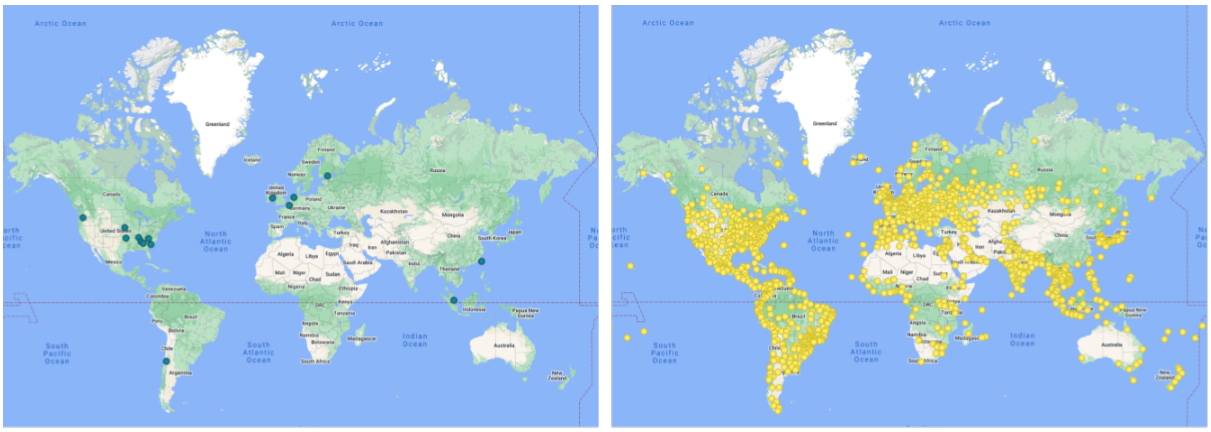
A. 向边缘计算转型的可持续性影响

对于 Google、Amazon 或 Microsoft 等领先的云服务提供商 [4]，，我们可以看到它们拥有并运营着非常可持续且高效的云数据中心。如今，它们声称已实现碳中和，并计划在 2030—2040 年实现净零碳排放（即移除与其排放量等量的碳）[5]。另一方面，它们在能源使用方面也非常高效。为了评估这一性能，我们可以参考电源使用效率（PUE）。这是一项用于描述数据中心能源使用效率的比率（设施总能耗除以信息技术设备所使用的能源）。因此，该指标有助于我们了解每单位用于计算的能源所对应的能源开销，这些开销主要用于冷却信息技术资源。根据 Uptime Institute 的数据，2022 年数据中心行业的平均 PUE 约为 1。55 [6], 这表明，总能源预算中约有 35% 被冷却系统消耗。

Google、Amazon 和 Microsoft 声称其电源使用效率约为 1.1 [7] [8] [9], 这意味着它们的能源开销比行业平均值低六倍。但它们是如何做到如此高效的呢？例如，如果我们查看图 1.a 所示 Google在 2022 年拥有并运营的数据中心的位置，就会发现其中大多数位于：(i) 寒冷地区，在这些地区制冷效率高得多，因此电能使用效率会更好；以及 (ii)支持可再生能源生产的地区，从而减少碳足迹。因此，它们使用更清洁的能源，并且使用效率更高。如果我们考察这些云服务提供商最新建成的数据中心，可以看到它们有几个共同点。它们占据了很大的区域，这不仅是为了容纳数据机房，也是为了进行可再生能源生产（风力发电机和太阳能电池板）以及建设制冷基础设施；同时，它们还需要大量的资金投入。因此，无论云是否位于经过优化、凉爽且空间充足的位置，只要资金充足，它就可以非常可持续且高效。但当信息技术基础设施需要更加靠近

数据源?

图 1.b 表示距离用户最近的 Google 边缘节点（超过 1,300 个）。边缘节点是云服务提供商的服务器，部署在本地运营商和互联网服务提供商的网络内部。其中许多节点部署在托管数据中心中，范围从机架到整个数据机房，遍布200多个国家和地区。边缘节点是云服务提供商的服务器，部署在本地运营商和互联网服务提供商的网络内部，其中许多节点部署在托管数据中心中，范围从小型机架到整个数据机房。2020 年，由于这些超大规模运营商的收入 [10]，托管市场增长了 22%；其中批发托管和零售托管分别增长了9%。这些边缘节点提供深受本地主机用户群体欢迎的服务，从而降低延迟和网络拥塞。



a. Google's Cloud DCs　　b. Google's Edge Nodes

图 *1*：*Google* 的云数据中心和边缘计算节点。*[https://peering.google.com/#/infrastructure]*

但是，鉴于全球平均电源使用效率约为 1.55，可以看出，将信息技术迁移至更靠近边缘计算的位置后，能源效率会显著下降。尽管本地数据中心运营商已实施了大量能源效率政策，但由于经济、技术或地理约束，在大多数情况下不可能达到超大规模云服务商的电能使用效率 [6]，，导致平均电源使用效率陷入停滞，如图 2 所示。此外，由于即使在新建数据中心中，空气冷却仍是主导技术，热功耗更高的新型处理器将对冷却效率造成不利影响。因此，预计行业平均电源使用效率将在未来几年上升，之后才会继续下降。所以，我们越进一步向边缘计算发展，就越会遇到可持续性和能源效率方面的障碍。

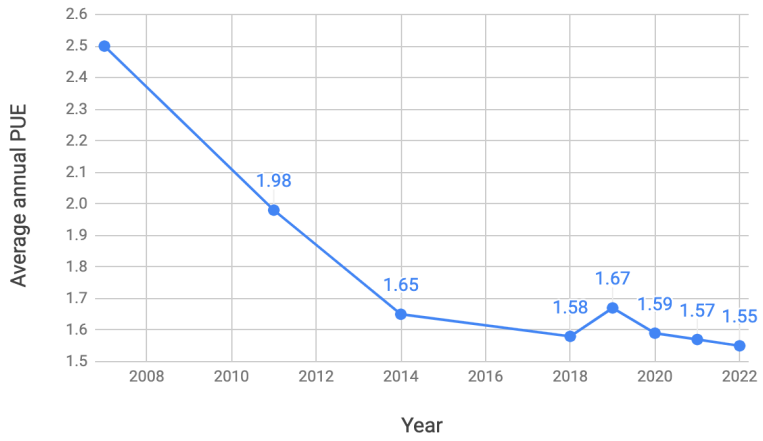


图 *2*：数据中心行业电源使用效率（*PUE*）的演变。

在此背景下，本文介绍了我们为实现未来边缘计算基础设施的可持续性所采取的初步措施。**我们的工作旨在实现**：(i) 高能源效率且具备经济可行性的部署；(ii) 容错的自动化运维；以及 (iii) 协同资源管理，以提高资源效率。

本文其余部分安排如下。第 2 节指出了在靠近数据生成源的位置应用基于云的方法所面临的主要局限。我们阐述了这些限制带来的边缘计算可持续性研究挑战，并提出一种创新方法来应对这些挑战。随后，第 3 至第 5 节介绍两相浸没式冷却、形式化建模、机器学习和以能源为中心的联邦管理等边缘计算赋能技术。为展示结合这些技术的方法所带来的潜在益处，第 6 节和第 7 节介绍了不同的用例。第 8 节总结了主要结论并展望了未来方向。

II. 开放挑战与设想的方法

据 Gartner [2], 预测，物联网（IoT）设备数量预计将从 2020 年到 2030 年增长至三倍，推动注入网络的数据量不断增加。增长最快的领域将是制造业和自然资源行业，年复合增长率（CAGR）为 18%；医疗服务提供者（CAGR 为 13%）、智能建筑（CAGR 为 12%）和汽车行业（CAGR 为 10%）也将快速增长。基于这些趋势，未来几年大量涌现的新型数据密集型应用将再次使现有设施不堪重负。因此，预计边缘计算将在不久的将来通过部署新的基础设施来提高其计算密度。

因此，在此背景下必须考虑三个因素。首先，这种新的边缘计算基础设施将更靠近用户。其次，越接近边缘，能源效率就越低，因为当地条件可能更差。第三，由于边缘计算将推动社会数字化，我们应以满足未来应用需求为目标进行部署。

驾驶辅助系统、个性化医疗和在线游戏只是边缘计算将促进并增强的应用中的几个例子。这些应用需要

强大的计算能力，以便通过机器学习进行海量数据处理。而且，这必须在靠近用户的地方完成，因为数据传输速率将非常高，延迟也至关重要。在这种情况下，可以设想未来的边缘部署将采用小型、计算密集型边缘数据中心（EDC），并将其部署在更靠近数据源的位置。

A. 当前方法的局限性与亟待解决的关键挑战 s

在此背景下，边缘计算带来的变革远超我们的预期，它并非云计算的渐进式解决方案。当前技术推动云计算实现了高效发展，但将其应用于网络边缘的部署、运行和管理时，却存在诸多局限性。

边缘基础设施部署相关问题：
正如我们此前所讨论的，越接近网络边缘，基础设施的能源效率就越低（主要是冷却方面），平均电源使用效率会从 1.1 大幅升至 1.55。此外，效率最高的数据中心占地面积很大，而当我们将计算能力部署到更靠近用户的位置时，这些区域往往无法使用，因为它们大多位于城市环境中。

随着大量边缘数据中心的部署，我们还必须在非理想选址中兼顾能源效率和占地面积，同时降低成本。然而，提升效率、减少物理占地所需的技术解决方案通常价格昂贵。基于这些原因，我们需要应对的首要关键挑战是：如何实现环境与经济层面均可持续的边缘计算部署。

与分布式动态运行相关的问题：
云端运行是在现场进行的，因为这些大型集中式基础设施配有专职人员，负责维护数据中心，并根据其日常运行情况作出决策。与之相反，边缘数据中心将构成规模庞大且高度分布式的计算基础设施，因此本地运行并不切实际。

这使得以学习为中心的自动运行成为动态分布式基础设施面临的主要研究问题之一。但当出现超出我们控制范围的异常情况时（例如安全攻击、数据机房中的热点、系统故障、服务中断等），会发生什么？在云端，这些异常由运维人员处理；但如果边缘数据中心无人值守，我们就需要相应工具，以避免设备遭受不可修复的损坏并造成重大经济损失。此外，健康与安全关键型应用（即个性化医疗和驾驶辅助）发生服务中断，会使用户的人身安全面临风险。因此，未来边缘计算需要解决的第二项关键挑战，是如何实现容错自动运行。

可用资源的高效利用相关问题：
大型超大规模运营商的云汇聚了整个区域内的用户请求。因此，到达数据中心的需求曲线更加平滑，因为其不会受到用户移动性的显著影响。在边缘计算中，用户需求在本地进行处理，而移动性会导致同一区域内的小型数据中心之间发生切换（例如，在无线接入网（RAN）层级之间或跨越骨干网的无线接入网之间传输工作负载）。这就是为什么每个边缘数据中心感知到的处理需求将大幅

受数据生成源移动性的影响，呈现出非常高的可变性（即每日或季节性变化模式）。

实际上，按照其峰值需求扩展边缘数据中心是不可行的，因为基础设施将无法得到充分利用，而且成本极高。但另一方面，如果边缘计算支持关键应用，我们就不能牺牲最终用户所感知的服务质量。因此，要确保边缘计算的可持续性，第三个关键挑战是实现有效的协同资源管理。

B. 设想的方法

可持续部署、容错自动运行和协同资源管理是实现可持续边缘计算的三大关键挑战。随着未来竞争力强的边缘基础设施的发展取决于其经济与能源可持续性，这些尚待解决的研究问题亟需得到解决。本研究提出通过三种不同的方法应对这些挑战，并通过将其结合起来，打造出所需的密集型、计算密集型边缘计算，以推动社会数字化商品化。图3展示了驱动我们研究方法的高层架构框架。

我们提出通过采用新型冷却系统（例如两相浸没式冷却），改善部署对环境和经济的影响。这项技术能够降低能耗、物理占地和成本，且不受环境条件（如温度和湿度）的影响，因此对提升网络边缘的可持续性尤为有吸引力。

在基于模型的系统工程（MBSE）框架中，形式化建模具有稳健性、模块化和可集成性，能够显著改善数据中心等复杂环境中的决策。因此，我们旨在以学习为中心的方法，将数学建模与机器学习算法相结合，为边缘计算固有的分布式自动化运行提供容错性。

计算、冷却和可再生能源发电系统的多样性，基础设施位置的差异，以及用户需求、能源价格和天气等因素的波动，都会影响边缘计算的能源消耗和成本。因此，我们建议利用这种异构性构建边缘联邦，通过协作管理动态提升资源利用效率。

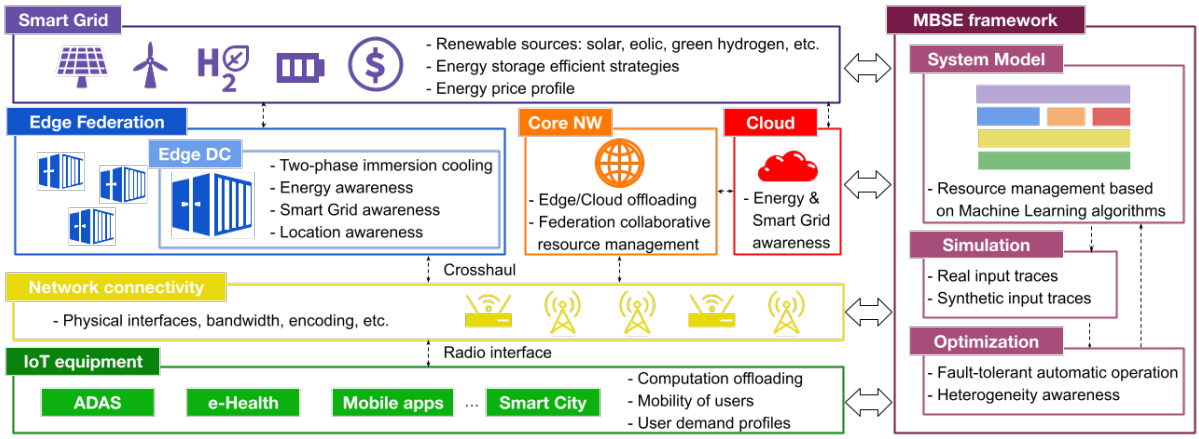


图*3*：一种面向物联网应用的分布式环境，集成了两相浸没式冷却以及用于边缘联邦的建模、仿真与优化框架。

III. 面向可持续边缘部署的两相浸没式冷却

两相浸没式冷却是一种新兴技术，能够应对边缘计算面临的诸多挑战，并提升现有数据中心的可持续性。在两相浸没式冷却中，信息技术设备浸没在一个装有工程流体的密封罐中；这种流体会在极高的工作温度（高于 60°C）下周期性地蒸发和冷凝。

当流体接触到非常热的表面（例如 CPU 或 GPU 芯片）时，流体会蒸发。随后，冷凝器（通过二次回路循环乙二醇水以与外界进行热交换）会将这些蒸气重新冷凝为液相，如图4所示。这些相变能够非常高效地从系统中带走热量[11]，，与传统方法相比，冷却能耗最多可降低 95%。

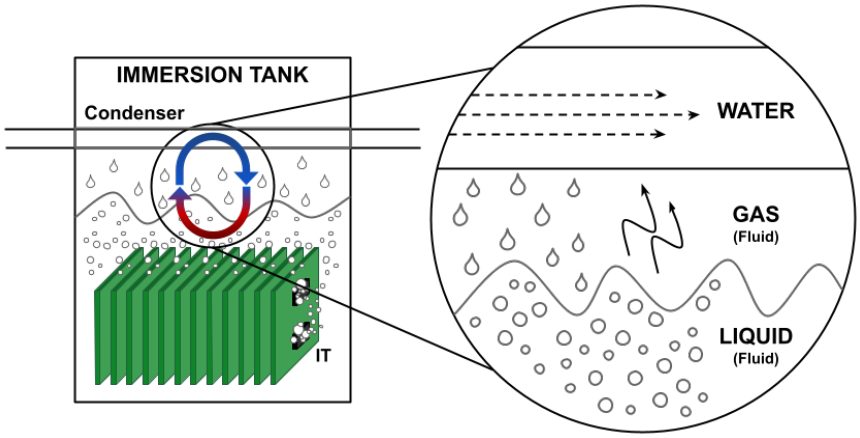


图*4*：两相浸没式冷却。

A. 两相浸没式冷却对边缘计算的益处

更高的能源效率和气候适应性：
通道　　该技术的气相与液相之间的转换在提取方面非常高效。

g

来自浸没式系统的热量，并在高于 60°C 的温度下产生。这使冷却能耗显著降低，最高可减少 95%，电能使用效率达到 1.02-1.03，这一水平此前仅在自然冷却中出现。此外，其全球变暖潜能值较低（低于 1），因此符合严格的环境法规。

两相浸没式冷却为边缘计算带来的主要优势之一，是无论基础设施位于何处、室外温度如何，都能发挥这些优势。由于这项技术的工作温度高于 60°C，冷却系统的设定温度可以设得很高。因此，即使在气候炎热的边缘位置，电源使用效率也可达到约 1.03，显著优于行业平均水平（电源使用效率=1.55），也高于超大规模运营商的平均水平（电源使用效率=1.1）。

更高的功率密度与更小的物理占地：
功率密度是设计数据中心计算与冷却基础设施时的一项关键指标 [6]。近年来，机架功率并未显著提升，工业界仍在使用典型的 4–6 kW 机柜。这种停滞的主要原因之一是空气冷却。空气冷却仍是工业界最广泛采用的方式，但受限于空气较低的传热系数。此外，这种传统技术需要采用冷热通道布局，并配备大量设备（例如冷水机组、计算机房空调与空气处理机组等），导致物理占地显著增加。

两相浸没式冷却凭借出色的热传递能力，可使机架功率密度提升 60 倍，每个机架的功率最高可达 250kW [11]。它还能显著改善基础设施的物理占地。该技术可将数据中心所需的地面空间缩减至原来的十分之一，在更小的空间内提供更强的计算能力：传统冷却方式的功率密度为 10kW/m2，而采用双相技术后可达 100kW/m2。这样一来，边缘数据中心便可缩小规模，部署在靠近用户的城市环境中，同时提供极强的计算能力，以支持要求严苛的应用。

低成本：
两相浸没式冷却的热量提取性能和效率可显著降低基础设施成本。冷却能耗最多可减少 95%，从而大幅改善能源预算（能源开销仅为 1%）。此外，与传统冷却方式相比，我们所需的冷却设备更少（也能减少用水量），占用的地面空间也更小，因此资本支出也将大幅降低。对于边缘计算而言，这种成本削减至关重要，因为要满足本地应用需求，需要部署大量边缘数据中心。

IV. 基于模型的系统工程与机器学习驱动的边缘计算自动化运行

如今，数据中心已将人工智能技术融入其中，以优化计算和冷却系统的管理。这些系统会监测环境，并将收集到的数据输入算法，以学习如何优化性能决策。根据 Uptime Institute 最近的一项调查，[6], 超过半数的数据中心所有者和运维人员愿意依靠经过适当训练的机器学习模型来制定

运维决策。对于由运维人员现场管理的数据中心而言，自动化运维是一种选择；但鉴于边缘计算的分布式特性，信息技术越贴近用户，就越难在基础设施中部署工作人员。因此，在边缘计算中，运维自动化变得至关重要，而环境标准化必须成为运维模式不可或缺的一部分，才能对边缘系统进行控制和管理 [1]。

在此背景下，我们旨在将基于模型的系统工程、基于仿真的系统优化与形式化规范及机器学习相结合，构建具有两个目标的边缘计算形式化模型。第一个目标是运用数学形式化方法，对基础设施的设计与运行进行标准化。第二个目标是帮助优化算法探索系统行为，并自动做出更优决策。

A. 运行数字化的形式化建模

本文提出一种新方法，将建模、仿真与优化（M&S&O）方法融入复杂系统开发的系统工程（SE）技术流程。我们的建模、仿真与优化型系统工程（MSOBSE）策略强调使用形式化建模工具，构建更稳健、更可靠的系统。仿真是验证所提模型是否满足系统需求的主要工具。此外，该策略还融入优化技术，以自动化设计决策，从而提升开发中系统对关键绩效指标（KPIs）的符合程度。图5展示了所提出的MSOBSE工作流程。

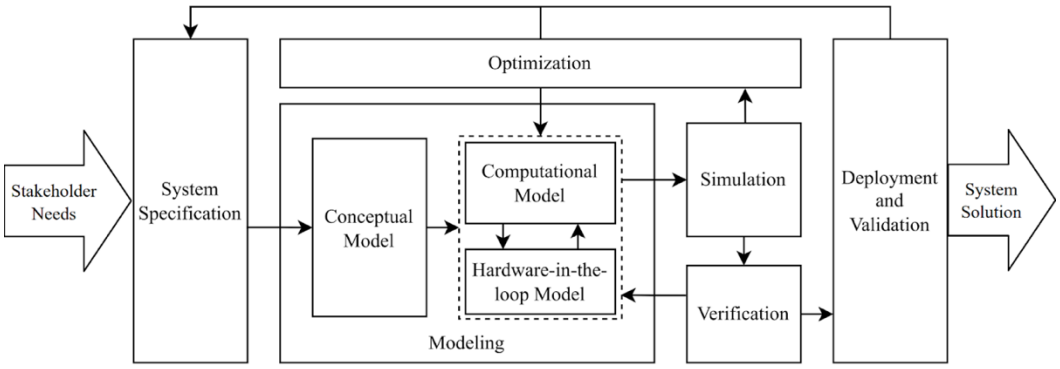


图*5*：所提出的*MSOBSE*技术流程的工作流程。

与之前的系统工程技术流程一样，我们首先必须研究所有系统利益相关者（例如客户、承包商和开发人员）的需求。根据研究结果，我们将定义构成系统需求的关键绩效指标。明确目标系统后，我们便开始系统建模流程。首先，我们建立解决方案的概念模型。然后，将概念模型转换为计算模型。计算模型是计算机程序，它们使用虚拟表示而非物理原型来仿真和分析系统行为。我们可以对计算模型运行仿真，以验证模型的行为是否符合概念模型的规定。通常，这一流程采用增量式方法：将概念模型分解为多个子模型。每个子组件都会被转换为计算模型，经过仿真后再独立验证。随着各个子模型通过验证，我们将它们合并，并验证所得

模型。最终，我们将得到一个计算模型，能够令人满意地捕捉系统形式化模型的行为。

复杂系统（例如数据中心）具有众多自由度。在这种情况下，对场景中的所有元素进行最优设计并非易事。为解决这一问题，我们建议在系统开发过程中集成自动化优化算法。这些算法将调整经过验证的计算模型，以获得满足系统规范的解决方案，同时避免产生高昂的开发成本。如果无法获得满足所有系统需求的设计方案，我们将修订规范，并重复建模、仿真和验证过程。在优化计算模型后，我们开始系统实施过程。硬件在环（HIL）模型会逐步用所建模系统组件的实际实现来替代计算模型的各个部分 [12]。然后，我们通过在 HIL 模型上运行实时仿真，对已实现的元素进行测试和验证。如果组件的行为与其对应模型不一致，我们将检查概念模型，以修正潜在错误。最终，我们将完成所研究系统的完整实现。所开发的解决方案是经验证的概念模型的产物，其中各组件经过优化，既能满足所有系统规范，又能降低开发成本。

B. 面向容错决策的机器学习

基于机器学习的解决方案如今已在数据中心中成为现实。机器学习目前能够为基础设施要素（例如温度、湿度、能耗等）提供非常准确的动态行为模型。它还有助于运维人员优化控制决策（例如温度设定点、工作负载分配和调度）。要恰当地训练机器学习模型，我们只需要监控系统收集的大量数据。这对于生产环境中真实系统的常规行为非常有效。但当异常情况发生时（即设备或监控系统故障），又会怎样呢?

数据中心故障可能导致服务中断，并造成严重的经济后果。在过去三年中，接受调查的运维人员中约有 60% [6] 遭遇过服务中断，原因主要与电力（37%）、计算机系统（22%）和冷却系统（13%）有关。这些数据中心中，45% 最近一次经历的服务中断造成了 10 万至 100 万美元的损失，另有 25% 的损失超过 100 万美元。大多数运维人员认为，通过改进流程、管理、配置和培训，本可避免这些服务中断。对于分布式、无人值守的边缘数据中心而言，这个问题会更加严重，因为它们可能面临更长的停机时间，进而产生更高的成本。

我们的目标是改进这些基础设施中的容错管理，以便在预测、避免或自动缓解停机和设备故障的同时做出决策。由于这类情况十分罕见，采集到的数据中只有极小一部分存在异常。因此，用于训练机器学习算法的数据集存在严重偏差，管理系统无法学会在发生故障时做出决策。在实际数据中心复现异常可能会危及设备安全，还可能耗费大量时间和能源。本文提出使用机器学习来优化系统管理，并通过生成能够复现异常情况的逼真合成数据，提供平衡数据集。由于.....的复杂性，这是一项重大挑战。

多变量数据中心场景中，要使合成数据集具有真实性，在生成按需异常时，必须保持变量之间的非线性相关性。

C. 集成 MSOBSE 和机器学习对边缘计算的益处

边缘计算需要一种目前处于设计与部署阶段的基础设施，因此可借助建模、仿真与优化标准化框架探索其可持续发展。我们提出的方法还通过将机器学习融入异常情况下的决策，实现对其动态运行的优化。这为边缘数据中心的去中心化无人值守管理提供了所需的容错自动化。我们提出的方法的主要优势如下。

形式化基础：
依赖数学形式化的建模方法能够为系统模型提供更高的完整性和稳健性。因此，它有助于及时检测概念模型中的错误。这一策略使我们能够更早地开发系统，并降低成本。

模型模块化：
如前所述，对复杂系统进行建模是一个迭代过程。系统组件分别进行建模和验证。然后，我们逐步合并经过验证的模型，并再次证明所得模型。模块化建模技术是这些场景中的绝佳选择，因为它简化了边缘基础设施与现有技术（例如云、物联网设备和网络元素）的集成与验证过程。

硬件在环集成：
所提出的工作流程在系统实现与验证阶段整合了硬件在环建模技术。我们的目标是找到合适的建模框架，使计算模型能够轻松过渡到硬件在环模型，同时不增加额外的开发时间。因此，任何计算模型都可以由真实硬件替代，从而让 MSOBSE 框架能够在纯仿真环境、真实场景或混合环境中运行。因此，该框架可以探索整个系统的行为，并了解其对形式化模型和实际设备的影响（例如资源使用、能耗以及用户感知到的服务质量）。

容错决策：
最后，MSOBSE工作流程必须提供一个完整的接口，用于优化开发中的系统。如图6所示，可以将机器学习集成到MSOBSE框架中，以生成无偏/有偏的训练环境。基于机器学习的合成数据生成器使MSOBSE框架能够利用来自标准系统行为和异常的充足数据进行训练。这样，资源管理优化算法将根据形式化模型中观察到的影响，学习针对系统故障场景进行决策。我们提出的方法旨在提供经过训练的基于机器学习的模型，使真实边缘基础设施能够实现容错决策。

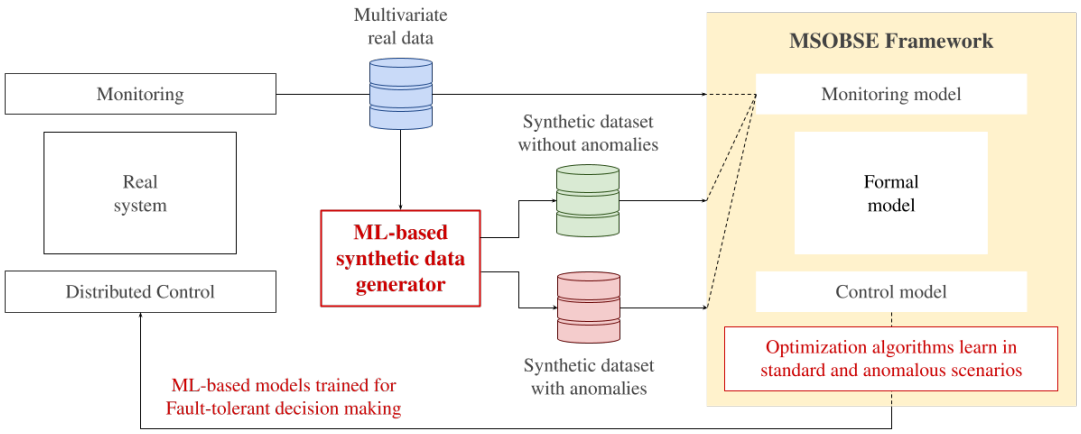


图*6*：容错决策集成方法的工作流程。

V. 面向协同资源管理的边缘联邦

由于用户移动性，在网络边缘本地感知到的服务请求特征具有显著的可变性。针对峰值需求对边缘数据中心进行容量规划并不切实际，因为这不仅成本高昂，而且资源使用效率低下。然而，鉴于边缘目标应用的重要性，必须确保服务可用性。因此，有必要开展协作，使边缘计算能够吸收峰值需求，同时保证服务质量。在本文中，我们提出对边缘数据中心实施联邦管理，以提高能源与成本可持续性，并维持用户感知到的服务质量。

A. 利用异构性提高能源效率

异构性是联邦管理面临的一项固有挑战。如图7所示，工业界认为新型冷却系统和可再生能源是可持续性的主要驱动因素[6]。新兴冷却技术（例如液冷、单相浸没式冷却和两相浸没式冷却）以及绿色能源（例如太阳能光伏、风能或绿氢）的逐步采用，将增加数据中心在物理层面的异构性。此外，边缘数据中心的地理分布更贴近用户，也是联邦管理异构性的另一个来源。天气、能源价格、智能电网能力、本地用户需求及其波动特征，将取决于每个边缘基础设施所在的位置。

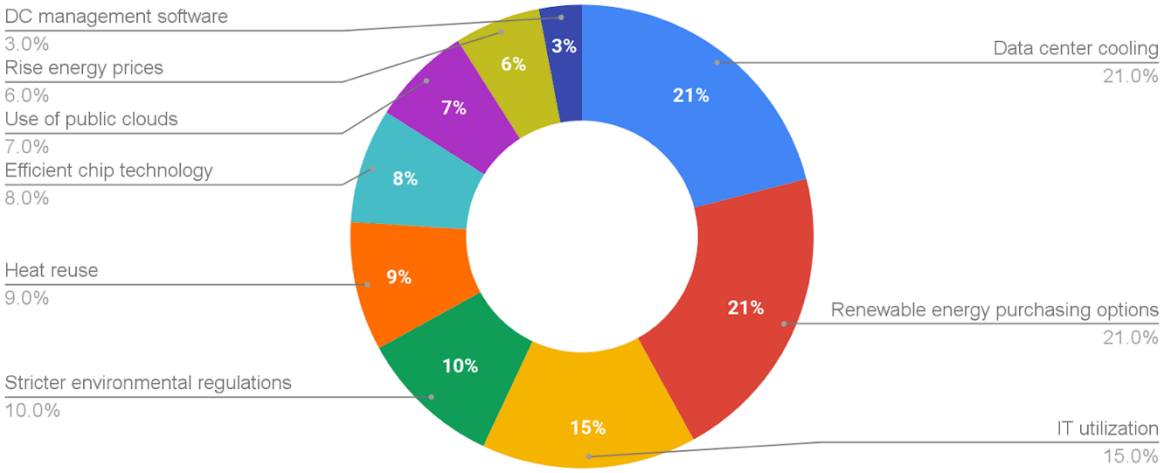


图7：数据中心行业可持续性提升的驱动因素。

这种异构性为通过联邦管理提升边缘计算的可持续性提供了绝佳机会。我们建议扩展资源管理系统，将联邦中调度和配置决策所带来的能源影响纳入考量。这样，联邦便能在任何给定时间利用各个边缘数据中心在其特定位置的异构性，同时了解其状态如下。

- **IT设备的利用率：**由分配的工作负载动态驱动。
- **IT功耗：**取决于IT技术（例如，CPU、GPU等）、资源利用率以及其低功耗策略的配置。
- **制冷功耗：**随着IT设备产生的热量、制冷技术以及室外气候条件（即温度和湿度）的变化而变化。
- **可再生能源生产：**这取决于部署的能源发电技术和天气（即太阳辐射强度、风力强度）。
- **能源存储：**储能技术特性以及充放电策略的函数。
- **电网能源定价：**由能源公司根据用户的日常消费模式设定。
- **服务质量要求：**针对联邦所支持的不同应用和服务专门达成一致。

B. 利用二级市场模型提升资源利用率

如图7所示，信息技术利用率是工业界认为推动可持续性改善的第三大因素。如今，云数据中心在这一方向上仍投入了大量精力，因为其平均利用率不超过60%[13]。然而，对于网络边缘基础设施而言，这一问题可能更加严重。提高信息技术资源利用率是支持关键应用的联邦边缘基础设施所面临的另一项固有挑战。为确保服务质量而过度配置计算设备会降低资源利用率。这会对能源效率产生负面影响，因为即使完全闲置的服务器也会消耗其最大功率的70%[14]。

在本文中，我们还提议将二级市场模式纳入联邦边缘计算管理：租户将闲置资源租赁给其他公司，以提高资源利用率。该模式利用现有边缘数据中心的计算资源，减少公司部署边缘基础设施的需求。然而，这一二级市场也带来了挑战，因为工作负载必须与对服务质量和安全性有严格要求的应用共存。联邦可通过针对每种类型的工作负载了解以下信息，从二级市场模式中受益。

- **应用资源利用率概况：**例如 CPU、内存、I/O、网络等，以及它们因用户交互而产生的波动。
- **用户需求特征：**会以不同模式变化，例如每日、每周等。
- **约定的服务质量：**根据应用的性质（尽力而为型与关键应用）。

能够对这些模式进行详细预测的模型，将是确保用户服务质量的关键。此外，基于资源需求的演变预测与能源相关的变量，将有助于实施主动式管理策略。我们的目标是利用异构性和二级市场机会，在维持应用服务质量的同时，提高边缘计算运营的可持续性和效率。

VI. 用例：面向高级驾驶辅助系统的可持续边缘计算

高级驾驶辅助系统（高级驾驶辅助系统）源于对更安全驾驶体验的需求，以减少事故和伤亡人数。高级驾驶辅助系统的目的是根据从车辆、车辆周围环境及乘员收集的数据，预测并避免紧急情况。据 Nvidia[15], 该类应用每辆车每秒生成约 2 千兆像素的数据，同时要求每秒具备 250 万亿次运算的处理能力。在车辆中嵌入信息技术系统以实现这种计算能力，会使技术解决方案极其昂贵，从而减缓其商业化发展。另一方面，将数据发送到云端进行处理会增加服务延迟，并可能危及车辆及其乘员的人身安全。

因此，在网络边缘部署计算基础设施对于高级驾驶辅助系统管理海量数据和应对延迟的关键性至关重要 [16]。此外，在靠近数据源的位置进行计算卸载，将有助于车辆与智慧城市协作，从而改善驾驶体验。因此，作为自动驾驶汽车的先驱，高级驾驶辅助系统是推动边缘计算发展的关键应用之一，这也凸显了探索其部署与运行可持续性的重要性。本节介绍我们为实现高级驾驶辅助系统边缘基础设施的可持续性所迈出的第一步。这些步骤包括：（i）设计采用两相浸没式冷却的原型边缘数据中心，并探索其热行为和功率行为；（ii）对原型边缘数据中心进行建模，并仿真其联邦运行，以利用制冷异构性；（iii）对联邦中的智能电网感知进行建模和仿真。

A. 用于高级驾驶辅助系统的两相浸没式冷却边缘数据中心原型

本用例旨在证明，我们的方法有助于为社会数字化发展可持续边缘基础设施。自2017年以来，我们一直与工业界合作开发数据中心原型，以研究两相浸没式冷却技术在边缘计算中的可行性和优势。我们的原型位于西班牙马德里理工大学电信工程技术学校，是一个改造为数据机房的小型集装箱船，如图8所示。集装箱内的浸没槽容纳了一套适用于运行基于机器学习的应用的GPU计算基础设施。集装箱顶部的干式冷却器是与环境进行热交换的热交换器。我们的原型边缘数据中心专为最高50 kW的信息技术基础设施而设计。



图 *8*：两相浸没式冷却边缘数据中心原型。

图9展示了机柜外部冷却系统的安装情况，包括膨胀容器、驱动泵、干式冷却器以及其运行所需的管道系统。在此原型中，冷却能耗的唯一来源是驱动泵，该泵在闭式回路中将水从冷凝器循环至干式冷却器。因此，冷却模型仅取决于浸没槽出口温度与回流温度之间的差值（分别为 T_{out} 和 T_{in} ）以及环境温度（T_{amb}）。

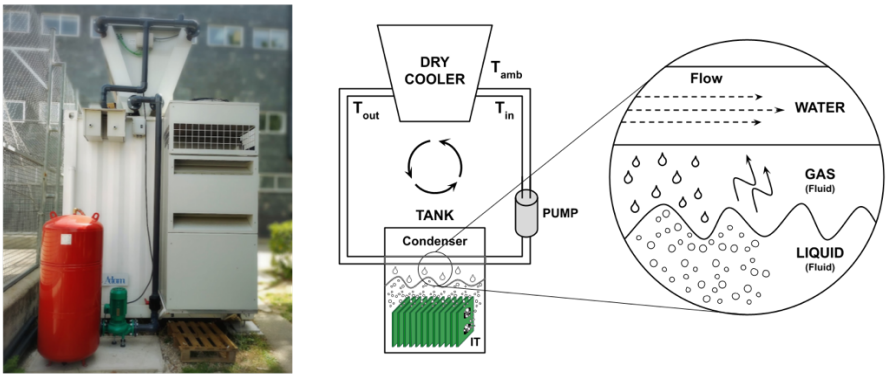


图 *9*：简化的两相浸没式冷却边缘数据中心。

根据原型中所使用的 IntersamIDV-130V 型冷却器的规格，已规定以下连接计算系统和冷却系统的温度范围：

$\Delta T_1 \in [0, 20]$ (oC)
$\Delta T_2 \in [0, 8]$ (oC)
其中：$\Delta T_1 = T_{in}$ -
$T_{amb}\Delta T_2 = T_{in} - T_{out}$

Aero-cooler 专为风冷环境设计，该环境的温度远低于浸没式冷却环境中的温度。因此，我们将在支持范围的上限运行（即，如果 Tamb = 32oC，那么 Tin = 52oC，Tout = 44oC）。在此场景中，我们使用了 Wilo IPL 50/115-0.75/2 驱动泵。由于驱动容量过大，冷却功率主要由其可提供的最小流量决定，而该流量足以满足系统的工作温度范围。因此，在大多数情况下，冷却功率约为 1300 W，正如我们此前的研究所述，该研究介绍了本工作采用的泵功耗模型（P_{EDC_COOL}）[11]。这使我们能够在最先进的两相浸没技术下实现约 1.03 的理论电源使用效率。此外，即使资源利用率达到云中的水平（60%），电源使用效率仍约为 1.04。

在计算基础设施方面，我们设计了一个定制机架，用于将 GPU（Sapphire Pulse Radeon RX 570）浸没在介电液体中。该机架可以将 GPU 及其控制器机械固定。该设计支持交错放置两个机架，每个机架配备 9 个 GPU，从而提高计算密度。通过在浸没槽中垂直和水平叠放多个单元，该设计能够实现较高的计算密度。图 10.a 展示了一个为 空气冷却设计的 9 GPU 机架。在我们的原型中，我们移除了 IT 设备中的空气冷却组件，例如风扇和散热器；这些组件通常体积较大，尤其是用于 GPU 时。这样一来，通过如图 10.b 所示将两个机架交错组合，我们便能在相同空间内安装两倍数量的 GPU。该配置可使机柜容纳高达 50 kW 的功率。

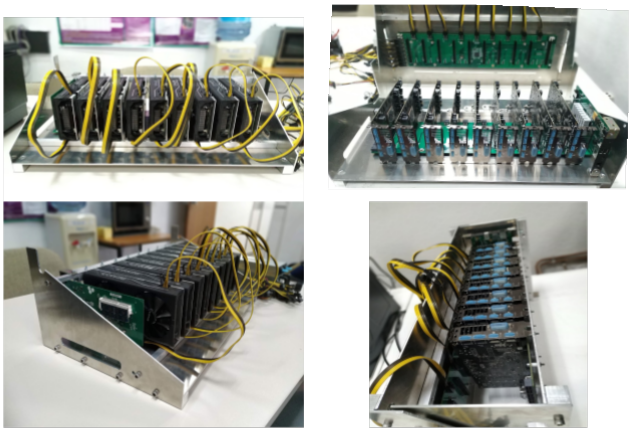


图 *10*：每个机架的计算密度比较。

GPU 运行深度学习（DL）算法时的性能显著高于 CPU[17], 因此，GPU 集群成为实时视频处理应用的最佳 IT基础设施之一。我们的高级驾驶辅助系统应用 [18] 是一项基于 DL 的服务，可在检测到驾驶员分心时向用户发出提醒。该系统收集车内视频，并通过卷积神经网络（CNN）进行处理，以评估驾驶员是否已无法专注于道路。由于 DL 算法的训练计算量很大，因此会将其卸载到边缘层，从而降低车载硬件系统的成本。这是工业界和研究领域的常见做法，相关人员正在探索联邦学习的优势，以提高边缘计算训练的效率。在 EDC 上利用用户的视频片段训练的模型会被发送到车辆中，以提高预测质量。由于边缘基础设施靠近用户，因此可以以极低的延迟更频繁地更新模型，从而改善道路安全。

该工作负载在边缘数据中心的IT基础设施上运行时，其功耗曲线会动态变化，并取决于用户数量。此外，当达到沸点时，两相浸没式冷却剂的性能会变得非常不稳定。在本研究中，我们使用了Novec7100，这是一种沸点为61oC的工程流体[11]。因此，机器学习是获取非常精确的功耗模型的绝佳替代方案（无需事先了解流体动力学），有助于我们估算工作负载在边缘数据中心中的行为。

我们在两相浸没式冷却中运行了高级驾驶辅助系统工作负载，同时修改主要控制变量（例如 GPU 时钟频率设置），并设置不同数量的用户并行会话。我们使用这些数据，通过前馈神经网络[11]对信息技术功耗进行建模。如图11所示，我们获得了准确的功耗模型P_{IT} (t)，其归一化均方根偏差（NRMSD）约为 3%，R^2 为 98%。

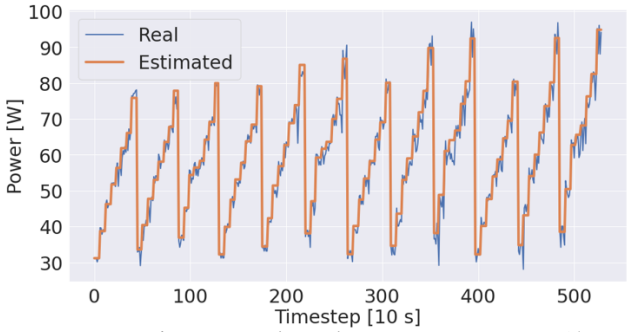


图 *11*：浸没式冷却边缘数据中心的信息技术设备功耗模型估算 *t.*

B. 两相浸没式冷却边缘数据中心的建模与仿真

为实现边缘数据中心作为联邦一部分的自动运行，我们提出应用我们的M&S&O 方法论。在本文中，我们将使用 Mercury[19], 我们的 MSOBSE框架，帮助优化算法探索系统行为，并基于形式化建模自动改进决策。我们的框架为系统中的各个元素提供了详细模型，从车辆到边缘数据中心联邦。我们对边缘数据中心进行了建模，使其表现类似于我们的两相浸没式冷却原型的行为，其中包括所获得的信息技术功率和冷却模型。图12展示了定义完整系统的模型，以及边缘数据中心联邦所需的不同粒度级别。P_{EDC_IT} (t) 是动态聚合能耗

浸没槽中所有GPU 的。P_{EDC_COOL} (t)表示动态冷却能耗，它是泵流量(ϕ_{EDC}(t)) 的函数，取决于干式冷却器提供的温度差 (T_{in} - T_{out})和 P_{EDC_IT} (t)[11]。然后，我们使用 DEVS形式化方法[20] 来定义概念模型，并据此生成计算模型[21]。我们的框架使用 xDEVS 仿真引擎对计算模型进行仿真[22]。

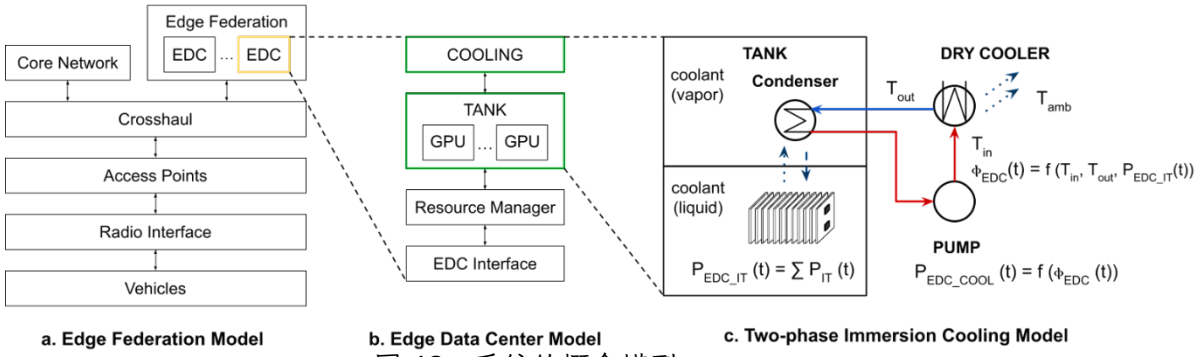


图 *12*：系统的概念模型。

利用冷却异构性的性能评估：

基于我们之前的工作 [11]，我们模拟了具有异构冷却的数据中心在边缘联邦中的运行，因为浸没式冷却将逐步发展，并与空气冷却共存。高级驾驶辅助系统应用请求取决于接入点覆盖范围内的车辆数量，其特征会随每日和每周模式显著变化。该联邦中的数据中心以协调的方式协作，以吸收这一需求。我们优化了工作负载分配，同时考虑了数据中心的异构冷却功率和用户移动性。

我们采用强化学习的演员-评论家配置来分配工作负载，并为更高效的场景提供奖励。主要优化目标是根据边缘数据中心的能耗状态，选择最合适的边缘数据中心来分配新到达的会话，从而最小化联邦的能耗。另一方面，我们的基线方法将工作负载分配给距离车辆最近的数据中心，优先最小化时延，而不考虑能耗。有关场景配置、用户移动性和算法细节的更多信息，请参见我们之前的工作 [11]。

为了评估我们用于高级驾驶辅助系统的原型模型的性能，我们扩展了 Mercury，将我们提出的两相浸没式冷却原型概念模型纳入其中。我们提出的 MSOBSE 方法有助于我们轻松地在框架中纳入并验证模型。图 13 展示了异构场景下联邦的总能耗对比。结果表明，异构性越高（结合风冷和两相浸没式冷却的边缘数据中心），效率越高；与基于延迟的优化策略相比，节能幅度超过 20%。

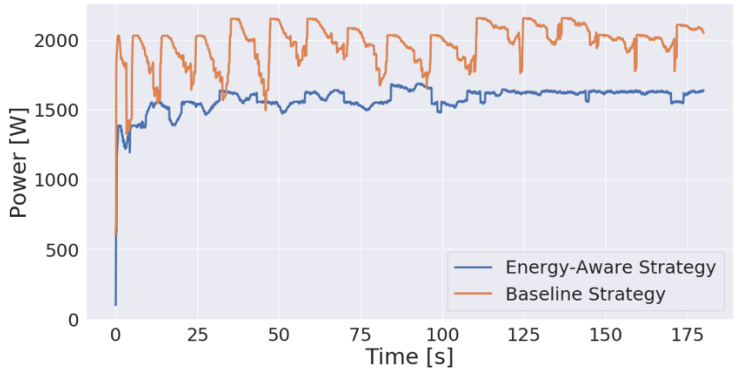


图 *13*：边缘联邦的能耗比较。

C. 智能电网感知的建模与仿真

我们已在 MSOBSE框架中纳入智能电网形式化模型，如图 14 所示。能源供应商（PROVR）以可变价格（$E_{price}(t)$）向智能电网供电。太阳能电池板产生免费的绿色电力（$P_{solar}(t)$），其变化曲线取决于太阳辐照度。如果所产生的电力超过边缘数据中心的总电力需求（$P_{EDC}(t) = P_{EDC_IT}(t) + P_{EDC_COOL}(t)$），则预计会产生剩余能源（$P_{surplus}(t)$）。联邦中的每个边缘数据中心都包含一个智能电网用户模型（CONSR），其中涉及太阳能电池板、电池和控制器。存储控制器根据电池的最大容量（$B_{capacity}(t)$）、能源价格和剩余能源（$P_{surplus}(t)$），决定电池的充电和放电（$P_{charge}(t)$）。控制器还负责管理从电网消耗的电量（$P_{CONS}(t)$）。控制器管理工作流程和边缘数据中心资源管理器策略的详细信息可参见我们先前的工作 [23]。

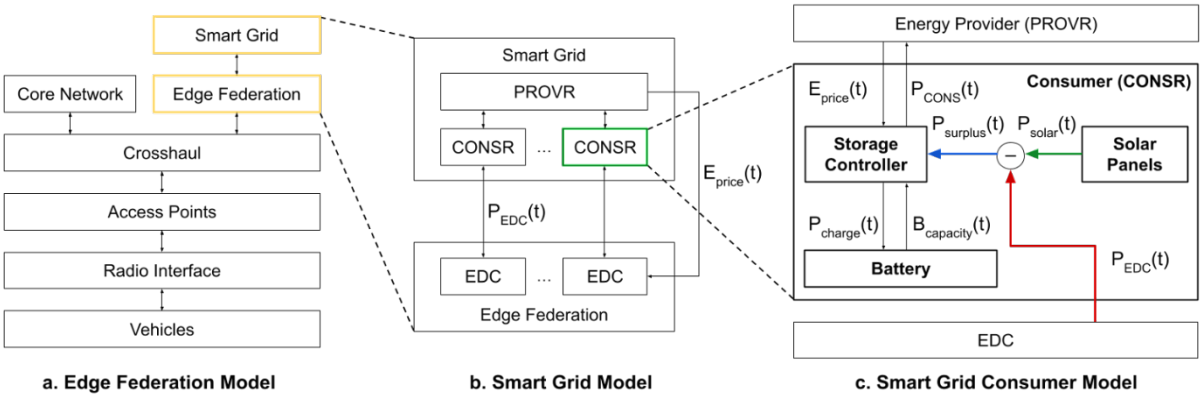


图 *14*：智能电网的概念模型。

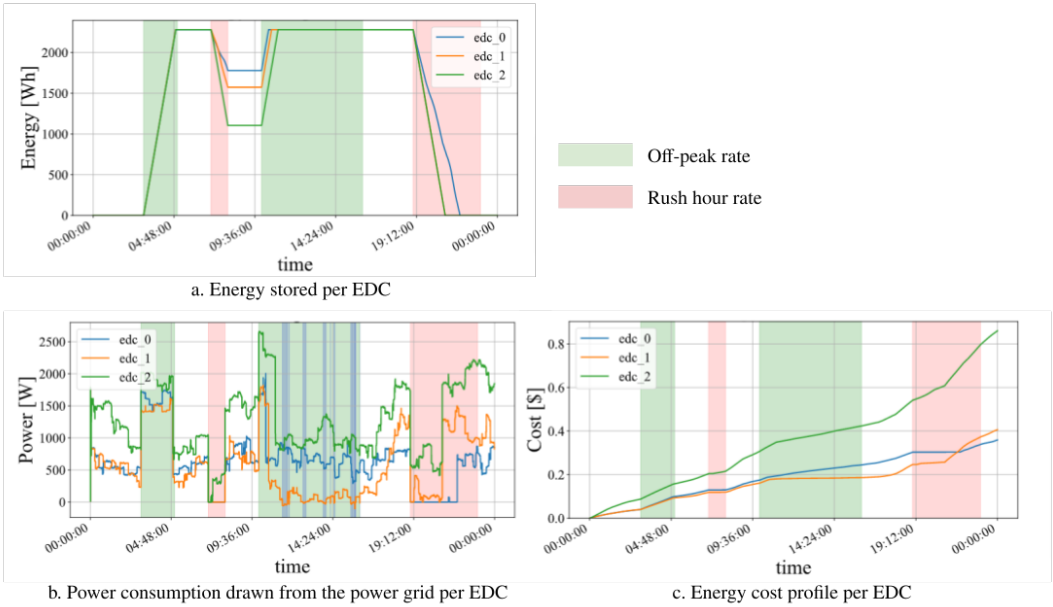
该模型使我们能够探索针对高级驾驶辅助系统的发电、储能、制冷和计算能力的协调管理。我们的目标是通过利用边缘数据中心位置、其环境条件以及能源价格的异构性，降低提供高级驾驶辅助系统服务的联邦的能耗。

利用智能电网异构性的性能评估：

在本研究中，我们通过工作负载分配优化了联邦的能耗，同时考虑了两相浸没式冷却、用户移动性、配备太阳能发电和电池的智能电网以及能源价格。我们采用了与先前工作 [23], 相同的太阳能发电和能源存储场景及模型，但不包含云

计算模型。通过这种方式，我们将利用智能电网异构性的优势限定在对边缘计算进行形式化建模的范围内，以分析其可持续性。

我们运行了一个为期24小时的场景：高级驾驶辅助系统应用由边缘基础设施提供服务，其需求曲线会根据系统接入点覆盖范围内的车辆数量而变化。在该实验中，为了隔离智能电网的效益，高级驾驶辅助系统服务均由距离各车辆最近的边缘数据中心提供。图15.a展示了联邦中三个边缘数据中心的电池充放电曲线。存储控制器利用非高峰时段，在能源价格最低时为电池充电。在高峰时段，当价格较高时，控制器优先使用存储的能源和太阳能。这使得非高峰时段对电网电力的需求增加，同时最大限度地减少高峰时段的电网电力使用量，如联邦中三个边缘数据中心的图15.b所示。因此，图15.c所示的联邦边缘数据中心成本曲线，仅在车辆所需的计算导致能耗超过太阳能发电量和电池储能容量时才会出现陡增。我们的结果表明，利用智能电网的异构性为高级驾驶辅助系统提供服务，可以使联邦的能源消耗降低20%，运营成本降低30%，从而显著提升边缘基础设施的能源可持续性和经济可持续性。



a. Energy stored per EDC

b. Power consumption drawn from the power grid per EDC

c. Energy cost profile per EDC

图 *15*：电池储能、电网电力消耗和能源成本。

VII. 用例：通过基于机器学习的合成数据生成进行容错冷却建模

除了地理位置和可再生能源生产外，数据科学也是超大规模运营商的云中提升能源效率的重要驱动力。据 DeepMind [24], 称，基于人工智能的解决方案可使 Google 数据中心的冷却能耗降低 40%，并将电源使用效率降低 15%。然而，工业界在纳入人工智能驱动的优化时面临的最大挑战，是获取海量数据以适当地训练算法（尤其是基于深度学习的方法），因为这在成本、时间和资源方面都十分昂贵。因此，这项

技术仅限数据所有者使用，因为公司通常不愿分享这类技术，以避免泄露用户信息。当机器学习驱动的优化必须在异常情况下做出决策时，这一问题会更加突出，因为此类情况偶发，且数据量有限。在这一用例中，我们旨在展示，包含逼真合成场景（包括按需生成的异常）的人工智能生成数据集，如何提升容错数据中心冷却模型的训练效果。在本节中，我们提出将基于生成对抗网络（GAN）的合成场景生成器集成到 MSOBSE框架中，以提高用于动态控制冷却基础设施的预测性温湿度模型的鲁棒性。

预测数据中心的温度对于捕捉数据机房的热行为尤为重要。热模型会自动管理冷却系统的设定温度，并确保IT设备处于安全工作范围内。我们需要可靠且稳健的预测热模型，以准确刻画冷却优化的影响。这样便可以探索更激进的优化方案，从而显著降低冷却能耗。然而，在风冷基础设施中，热变化会伴随数据机房内相对湿度的非线性漂移。控制数据中心的湿度对于避免设备遭受不可修复的损坏至关重要。如果湿度过高，露点很容易达到；如果湿度过低，静电放电的风险就会增加。因此，为了获得可靠的预测模型，有必要使用来自能够保持温度与湿度之间相关性的场景的数据集对模型进行训练。此外，如果模型针对异常情况进行训练，就能实现更好的自动容错控制。在此用例中，我们提出使用生成对抗网络生成逼真的合成场景，同时保持变量之间的相关性。生成对抗网络通过潜在空间实现对所生成场景的高度控制，这使其特别适合按需为有偏数据集扩充异常样本。

图16展示了我们系统的工作流程：获取预测性温湿度模型，并将其整合到控制系统的决策过程中。首先，基于从数据中心收集的真实数据，生成对抗网络生成器会生成包含和不包含异常的合成场景。然后，使用真实数据集和合成数据集训练温湿度模型，使其学会在各种场景下（包括发生异常时）进行稳健预测。训练完成后，我们的模型可以整合到 MSOBSE 框架中。我们的模型与真实数据和合成数据的注入相结合，将帮助控制系统探索优化对机房温度和湿度的影响，即使发生异常也不例外。因此，可以提升容错决策能力。

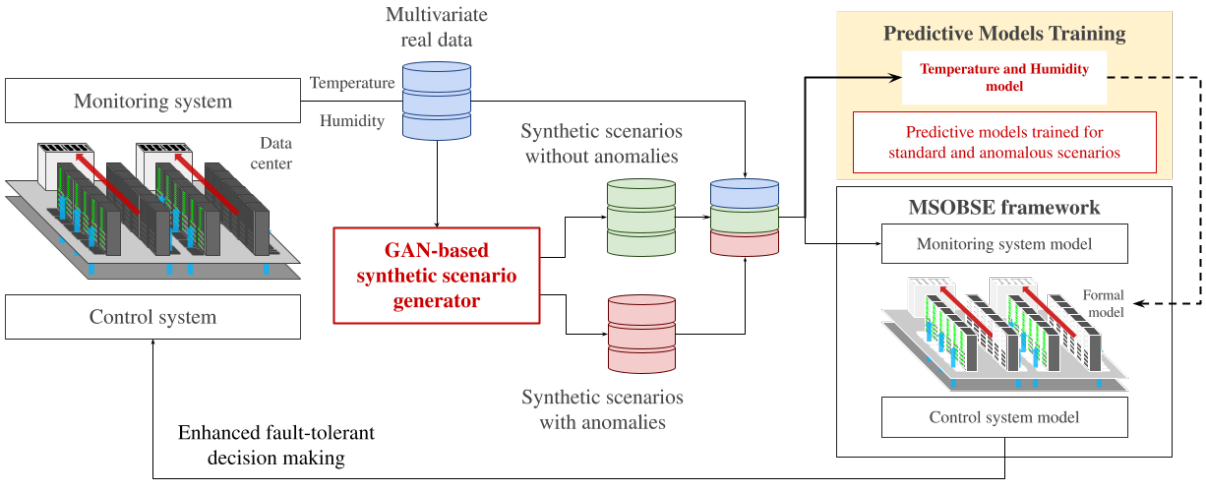


图*16*：用于容错冷却建模的集成方法工作流程。

容错冷却建模的性能评估：

为了评估我们提出的建模方法，我们使用了来自西班牙马德里 Adam Data Centers 设施的 15,000 个温度和湿度样本构成的数据集。这些数据样本来自分布在冷热通道机架不同高度的 35 个传感器，每 10 分钟采集一次。该数据集还包含一个分类变量，用于标识传感器节点的唯一标识符。这些数据已输入我们在先前工作 [25], 中介绍的基于生成对抗网络的合成场景生成器，以生成包含和不包含异常的合成数据样本，如图17所示。

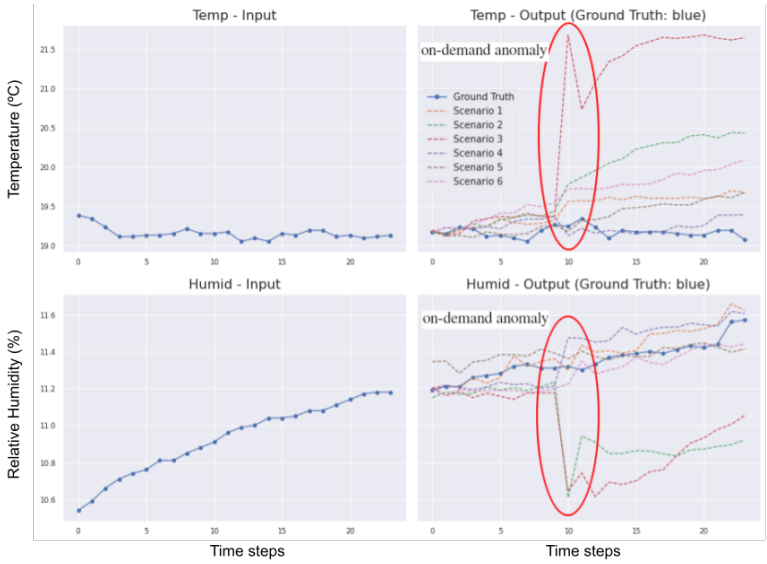


图 *17*：单个传感器的合成场景生成。

图 17 展示了输入生成对抗网络（GAN）的双变量时间序列，以及对应的单个传感器输出。每种颜色代表温度和湿度的一种不同的 4 小时合成场景。在生成场景的第 10 个时间步中，按需注入了一个异常。在场景 3 中，我们可以看到温度骤升，同时湿度大幅下降，保留了该场景的非线性相关性。场景的后续步骤与这一事件相符，呈现出逼真的演变（例如，服务器故障导致

出现热点，因此风扇开始全速运转以降低温度，但它们无法足够快地散热，温度便持续上升）。

为评估使用合成数据时的训练性能，我们采用一个以 10 分钟为预测窗口、用于获取温度和湿度的模型。待训练模型是一个基于长短期记忆神经元的前馈神经网络，包含 2 个隐藏层（分别有 256 个和 32 个神经元），其 2 个输出神经元分别提供我们关注的两个变量。与我们先前工作 [25]，中介绍的神经网络一样，我们使用了 AdaBelief优化器、线性输出激活函数、双时间尺度更新规则和丢弃法。

首先，我们使用 15,000 个真实数据样本和 150,000 个未注入异常的合成样本对模型进行训练。在真实数据上的测试表明，与仅使用真实数据训练的模型相比，温度和湿度的预测误差（均方误差）分别降低了 28.9% 和 7.4%。随后，我们使用 15,000 个真实数据样本和 150,000 个合成样本对模型进行训练，其中 5% 的样本含有异常。与仅使用真实数据训练的模型相比，在这种情况下，使用真实数据进行测试所得的温度和湿度误差分别降低了 27.9% 和 17%。这些结果表明，整合合成数据和异常数据有助于提升模型稳定性和预测鲁棒性。

VIII. 结论与未来方向

边缘计算能够推动社会数字化，促进关键应用的部署，从而改善人们的生活质量。但在实践中，由于当前基于云的策略应用于网络边缘时存在局限性，边缘范式的实施所带来的颠覆性远超我们的预期。我们的主要目标是增强未来边缘基础设施的可持续发展能力，使其在环境和经济方面都具备竞争力，从而加快其应用。本文提出了我们应对边缘部署和运营关键挑战的愿景，例如能源效率、容错自动化和协同编排。我们提出了一种创新方法，将两相浸没式冷却、形式化建模、仿真与优化技术、机器学习以及联邦管理相结合，以推动边缘计算的可持续性。

为展示我们方法的优势，我们介绍了在推进面向高级驾驶辅助系统应用的边缘计算基础设施可持续性方面所做的初步工作。我们设计了一个采用两相浸没式冷却、具备先进电源使用效率的原型边缘数据中心，并探索其热行为和功率行为。随后，我们对原型数据中心进行了建模，以仿真和优化其联邦运行。最后，我们通过仿真具备智能电网感知能力的边缘联邦，分析了基于两相浸没式冷却的原型的性能。结果表明，我们的方法可使边缘联邦的能耗降低 20%、运营成本降低 30%，显著提升边缘基础设施的能源和经济可持续性。我们还展示了通过提供合成的平衡数据集来训练更稳定、更稳健的模型，从而支持自动化管理的潜力。借此，优化算法能够学会在异常情况下做出决策，从而增强未来无人值守边缘部署的容错性。

这些初步结果展现了我们对边缘计算迈向更可持续未来这一愿景的潜在益处。朝着这一方向，我们未来的工作将围绕以下研究方向展开。

- 探索两相浸没式冷却的极限，以支持更高的IT功率密度，并使其更接近冷却剂临界热通量。
- 分析新型流体冷凝方法，以开发自然冷却两相浸没系统。
- 研究安全的工作负载分配和整合策略，并考虑两相浸没式冷却剂的换热容量限制。
- 提升边缘联邦的容错运行能力，以增强其抵御异常和漏洞的安全性。
- 根据新型关键应用的动态资源需求和需求曲线，为其设计可持续的边缘联邦容量规划与部署方案。
- 利用地理分布式边缘联邦和云的异构性，为它们之间的协作提供新颖的能源优化方案。
- 在竞争激烈的多租户环境中，为关键应用制定协作策略。

致谢

我们谨向 3M、Adam Data Centers、ImesAPI 和 Tychetools 致谢，感谢它们对本研究的支持。

本研究部分由 HiPEAC6网络、欧盟 Horizon2020 研究与创新计划以及墨尔本大学提供支持。本项目还获得了工业技术发展中心（CDTI）和面向社会挑战的国家研发计划（Retos Colaboración 2017）的支持，相关合同编号为西班牙科学与创新部的 IDI-20171194 和 RTC-2017-6090-3；此外还获得了西班牙国家研究署（Agencia Estatal de Investigación）项目 PID2019-110866RB-I00/AEI/10.13039/501100011033 的资助。

参考文献

[1] Cisco.2022. 建立边缘基础设施。面向服务提供商的新型基础设施模型。白皮书

[2] Thomas Bittman、BobGill、Tim Zimmerman、Ted Friedman、Neil MacDonald、KarenBrown。

[3] L. Lin、X. Liao、H. Jin 和 P. Li，《面向边缘计算的计算卸载》，载于 IEEE会刊，第 107 卷，第 8

[4] Raj Bala、Dennis Smith、Kevin Ji、David Wright、MiguelAngel Borrega。《云基础设施和平台

[5] Mytton,D. 在云中隐藏温室气体排放。《自然·气候变化》，10，701（2020）。

[6] JacquelineDavis、Daniel Bizo、Andy Lawrence、Owen Rogers、MaxSmolaks、Lenny Simon、

[7] Jesse Dodge、Taylor Prewitt、Remi Tachet des Combes、Erika Odmark、Roy Schwartz、Emma Strubell、Alexandra Sasha Luccioni、Noah A. Smith、NicoleDeCario 和 WillBuchanan。2022年。测量云实例中的人工智能碳强度。载于2022ACM公平性、问责制与透明度会议（FAccT '22）。ACM，纽约，美国，1877–1894。

[8] S. Vos、P. Lago、R.Verdecchia和 I. Heitlager，《在公有云中通过架构策略优化软件以提升能源效率》，2022年国际信息通信技术可持续发展会议（ICT4S），保加利亚普罗夫迪夫，2022 年，第 77–87 页。

[9] Li, H.,Dai,H., Liu, Z.等。Storm 中流处理应用的动态节能调度。Computing 104, 413–432（2022）。

[10] Synergy Research Group。托管市场大幅受益于超大规模运营商客户。2020

[11] Sergio Pérez、PatriciaArroba、JoséM.Moya，通过深度强化学习和两相浸没式冷却实现边缘计算的能耗感知优化，《未来一代计算机系统》，第125卷，2021年，第891-907页。

[12] J. A. Ledin，《硬件在环仿真》，《嵌入式系统编程》，第 12 卷，第 42–62 页，1999 年。

[13] Tahseen Khan、Wenhong Tian、Shashikant Ilager、Rajkumar Buyya，《云数据中心的工作负载预测与能耗状态估计：以机器学习为中心的方法》，《未来一代计算机系统》，第 128 卷，2022 年，第 320–332 页。

[14] Yadav、R.、Zhang、W.、Li、K. 等。云数据中心中面向能效的过载主机管理。《集群计算》，24，2001–2015（2021）。

[15] NVIDIA，《自动驾驶安全报告》，NVIDIA 公司，技术报告，2018 年。

[16] D.A. Chekired、M.A. Togou、L. Khoukhi、A. Ksentini，《支持 5G 切片的可扩展 SDN 核心网络：迈向自动驾驶服务的超低时延》，IEEE J. Sel. Areas Commun. 37 (8) (2019) 1769–1782。

[17] G. Ananthanarayanan、P. Bahl、P. Bodk、K. Chintalapudi、M. Philipose、L. Ravindranath 和 S. Sinha，《实时视频分析：边缘计算的杀手级应用》，Computer，第 50 卷，第 10 期，第 58–67 页，2017 年。

[18] S. Pérez、J. Pérez、P. Arroba、R. Blanco、J. Ayala、J. Moya，面向节能智慧城市的预测性 GPU 高级驾驶辅助系统管理，载于：2019 年 IEEE 国际智慧城市会议（ISC2），2019 年，第 349–354 页。

[19] Román Cárdenas、Patricia Arroba、Roberto Blanco、Pedro Malagón、José L. Risco- Martín、José M. Moya，Mercury：面向数据流的物联网应用的建模、仿真与优化框架，《仿真建模实践与理论》，第 101 卷，2020 年。

[20] B. Zeigler、Yoonkeon Moon、Doohwan Kim 和 G. Ball，“用于 高性能建模与仿真的 DEVS 环境”，《IEEE 计算科学与工程》，第 4 卷，第 61–71 页，1997 年 9 月。

[21] Román Cárdenas、Patricia Arroba、José L. Risco Martín（2022），将人工智能引入边缘计算：部署高效物联网架构的形式化建模与仿真规范，《仿真杂志》，165，494 511：-。

[22] Risco-Martín，JL、Mittal，S、Henares，K、Cardenas，R、Arroba，P。xDEVS：用于形式化离散事件系统互操作建模与仿真的工具包。Softw Pract Exper。2022；1-42。

[23] Cárdenas，R.、Arroba，P.、Risco-Martín，J.L. 等。智能电网感知的边缘计算联邦建模与仿真。Cluster Comput（2022）。https://doi.org/10.1007/s10586-022-03797-8

[24] Evans R、Gao J（2016）DeepMind AI 将 Google 数据中心冷却费用降低了 40%。DeepMind 博客

[25] Pérez，J.、Arroba，P. 和 Moya，J.M.。利用生成对抗网络对数据中心进行多变量场景预测，实现数据增强。Appl Intell 53，1469–1486（2023）。