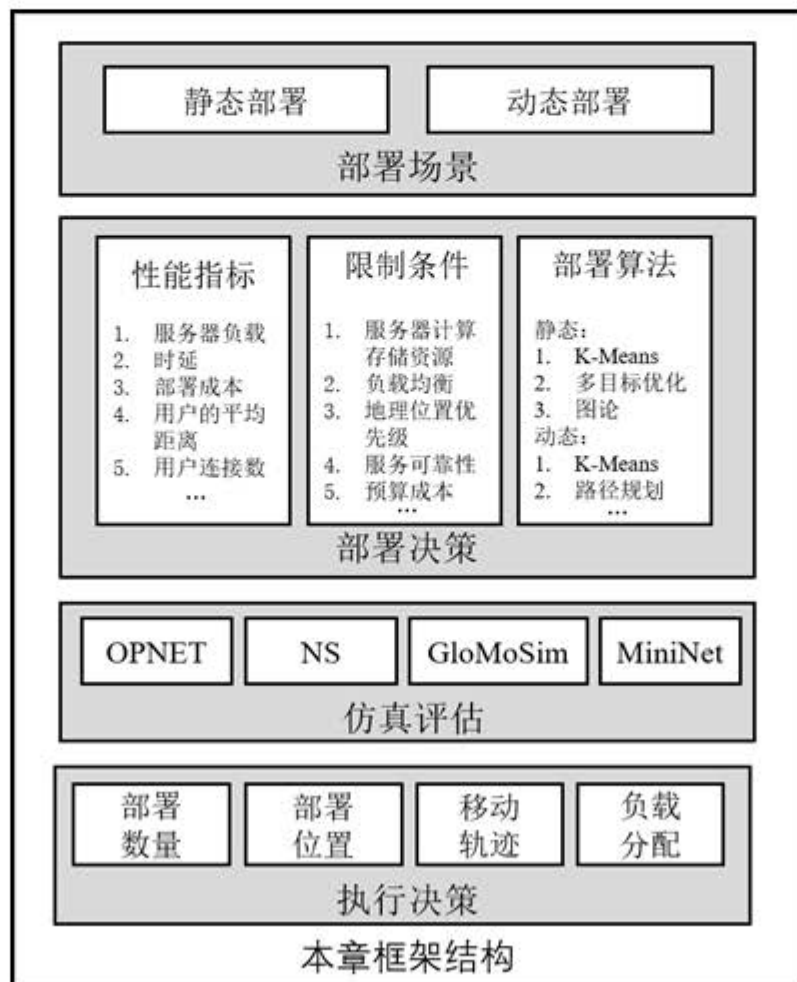


边缘系统部署

- ▶ 课程简介
- ▶ 边缘系统部署场景
- ▶ 边缘服务器部署问题

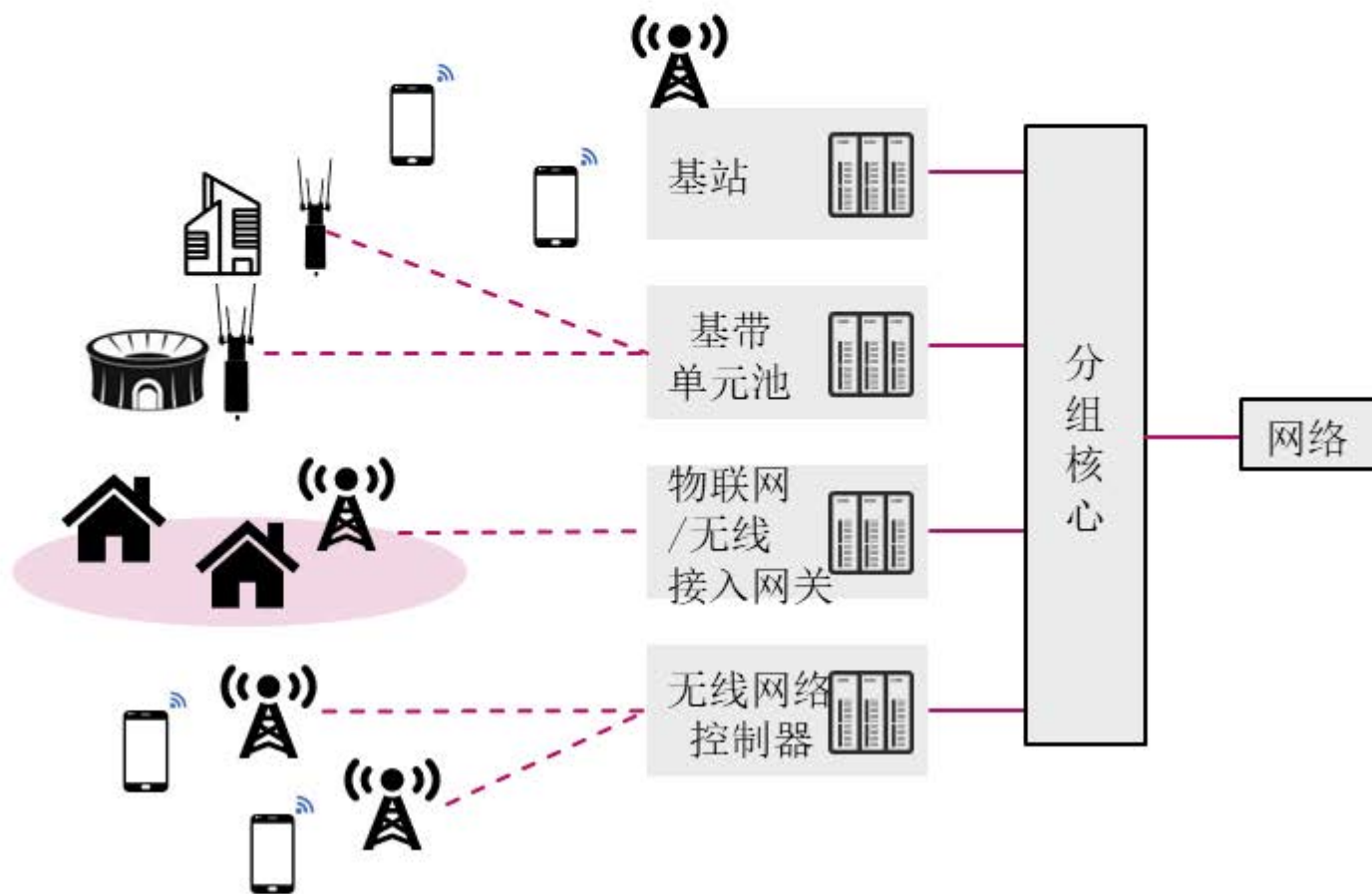
► 本章围绕边缘系统服务器部署问题的展开介绍。



- ▶ 课程简介
- ▶ 边缘系统部署场景
- ▶ 边缘服务器部署问题

- ▶ 静态部署场景
- ▶ 动态部署场景

- ▶ 静态部署指的是服务器位置一旦确定，后续难以改变的部署方式。



- ▶ **静态部署场景可根据具体环境的差异，细分为室内和室外两类场景。**
 - 室外场景中，边缘计算服务器可以直接部署在无线接入网或其周边，目的是便于和无线接入网进行应用协调，且其部署位置根据网络环境的不同而有所差异。
 - 室内场景中，边缘计算服务器通常表现为无线接入网关（Wireless Access Gateway），根据具体部署场景的差异，该网关内可部署多类场景通用的边缘应用服务。

- ▶ 静态部署场景
- ▶ 动态部署场景

► 静态部署场景有如下问题:

不能很好应对用户任务请求量随时空变化的情况

不利于节点扩展后的场景

静态部署服务器负载无法支撑突发性事件

服务器覆盖范围有限

- 上述问题在动态部署场景或混合部署场景下得到有效解决，即让服务器或部分服务器装载到移动小车、无人机或移动机器人让使其获得移动能力，服务器可以通过移动设备的请求动态调整其部署位置。

- ▶ 课程简介
- ▶ 边缘系统部署场景
- ▶ 边缘服务器部署问题

- ▶ 部署问题的评价指标及限制条件
- ▶ 静态部署问题
- ▶ 动态部署问题

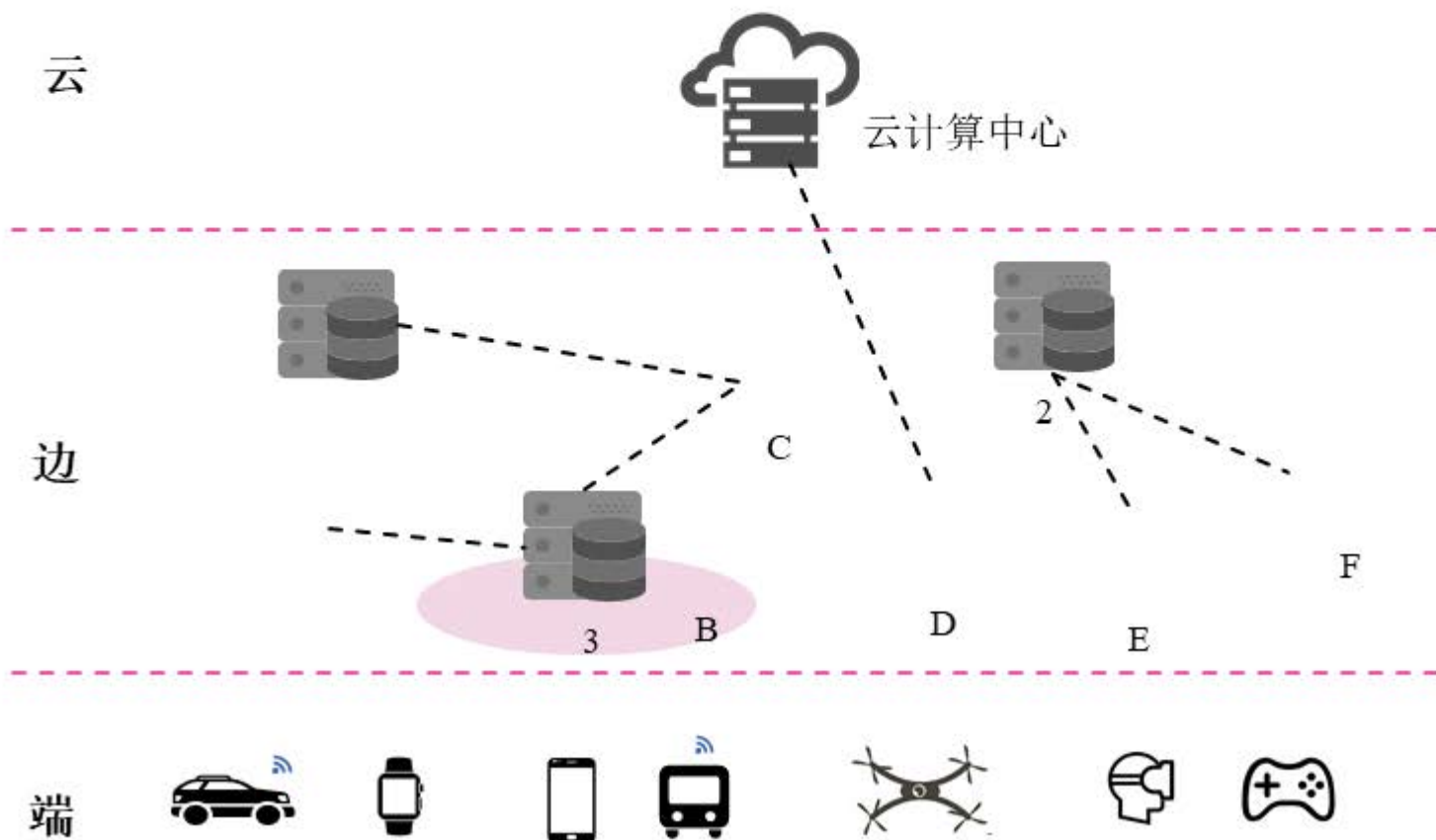
► 部署问题的评价指标：

时延	服务器资源利用率	服务器负载	部署财务成本
边缘计算服务器和其负责的接入点或用户的平均时延是部署问题主要考虑的KPI	资源利用率指服务器负载占整体计算资源和存储资源的利用率	负载是指边缘服务器部署后负责的在线用户的应用服务请求预期出现的计算负担	边缘服务器基础设施的部署和使用以及运行时能耗均会影响财务成本

- ▶ **部署问题的限制条件：**由于现实环境、商业用途和基础设施硬件本身的限制，导致边缘计算服务器部署问题在优化上节介绍的评价指标时需要考虑多种限制。

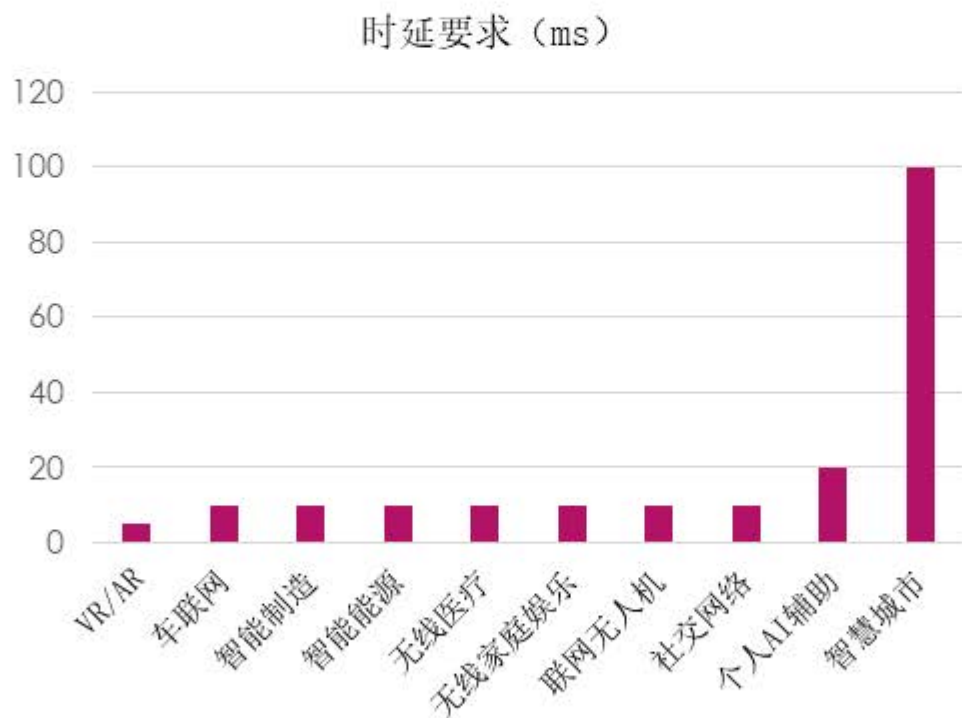
部署位置地理特点	服务器计算存储能力	负载均衡
任务时延要求	预算成本	服务可靠性

- 部署位置地理特点：部署问题中服务器的选址存在两种情况，即在现有接入点选择和和环境地理中独立选择。



- ▶ **服务器计算存储能力：**相比于云计算中心，边缘服务器拥有的计算资源和存储资源是有限的。因此服务器部署时需要对当前部署位置所产生的应用服务请求量进行估计。
- ▶ **负载均衡：**本着公平性和区域整体服务器资源利用率最大化原则，部署若干同构的边缘服务器需尽可能做到服务器负载均衡。
- ▶ **服务器的负载有多种衡量指标：**
 - 服务器在特定时长内的最大用户连接数
 - 终端设备向服务器发起的请求时长
 - 服务器在特定时长内的数据请求量

- **任务时延要求：**不同的应用程序对时延和带宽的要求有所差异。



- ▶ **预算成本：**从服务器提供商的角度看，财务收益是主要关注的指标。收益主要来源于，通过提供边缘计算服务来改善用户服务体验，向用户收取服务成本。财务主要的开销分为初始部署和后期维护两部分。
- ▶ **服务可靠性：**服务器部署时还需要考虑诸如可靠性、安全性和隐私等指标。可靠性难以给出一个确定的量化方法，通常会用系统在一段时间内出现故障并恢复的能力来衡量。

- ▶ 部署问题的评价指标及限制条件
- ▶ 静态部署问题
- ▶ 动态部署问题

- ▶ 概括
- ▶ 边缘服务器部署问题的独特性
- ▶ 区域离散化
- ▶ 部署策略：聚类算法
- ▶ 部署策略：最优化算法
- ▶ 部署策略：图论相关算法

- ▶ **边缘服务器静态部署问题的解决步骤可以概括为如下四步：**
 - 1. 区域离散化，确定服务器部署位置选择空间；
 - 2. 数学建模，明确问题的评价指标和限制条件的量化模型；
 - 3. 提出算法，解决研究的边缘服务器静态部署问题；
 - 4. 仿真测试，评估部署方案的优良。

- ▶ 概括
- ▶ 边缘服务器部署问题的独特性
- ▶ 区域离散化
- ▶ 部署策略：聚类算法
- ▶ 部署策略：最优化算法
- ▶ 部署策略：图论相关算法

相比于无线传感网传感器部署问题，边缘计算服务器部署问题主要有如下区别。

边缘计算中的终端节点数量更多、流量需求更大且节点异构性更强

边缘计算中节点较为密集，对无线信号干扰更为敏感

边缘计算中的终端节点即可以采集数据，也具备本地计算任务的能力

边缘计算中处于同一服务器的不同计算任务需要竞争CPU和内存资源

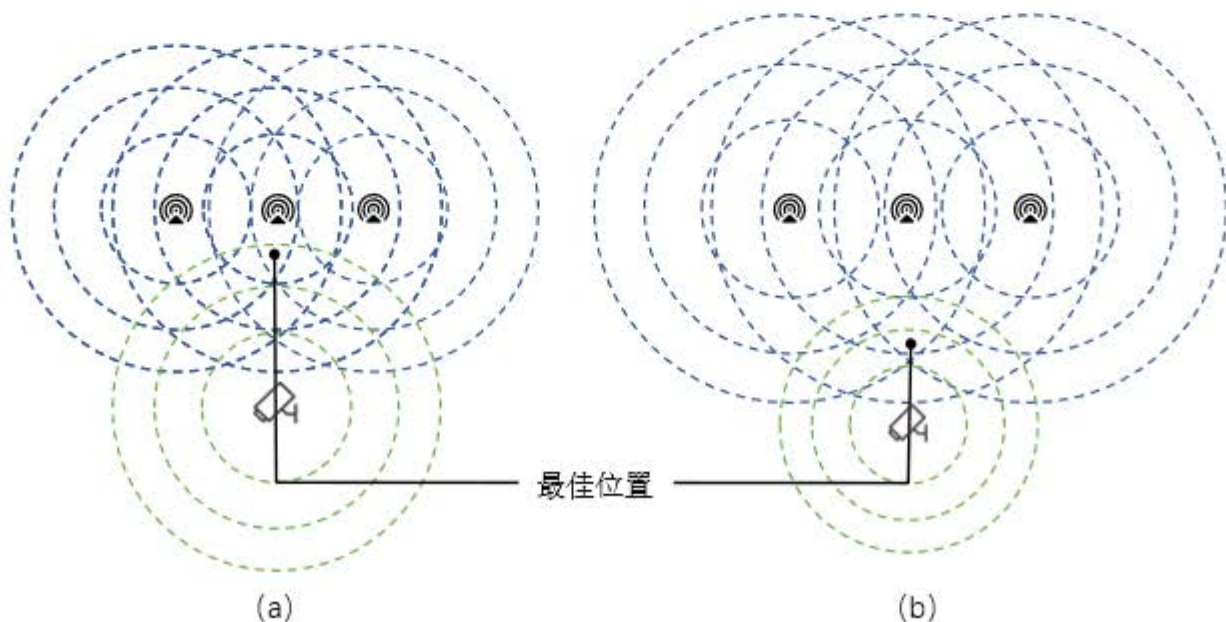
室内场景的WiFi接入点部署和边缘服务器部署问题的主要差异在于，WiFi接入点只负责提供数据传输的服务，不涉及任务的计算处理，因此在部署时主要考虑无线链路质量相关因素。而边缘计算服务器既拥有数据传输的功能，也具备数据处理计算的能力，部署时需要将通信和计算流程结合考虑。

室内场景WiFi接入点部署问题。



- ▶ 概括
- ▶ 边缘服务器部署问题的独特性
- ▶ 区域离散化
- ▶ 部署策略：聚类算法
- ▶ 部署策略：最优化算法
- ▶ 部署策略：图论相关算法

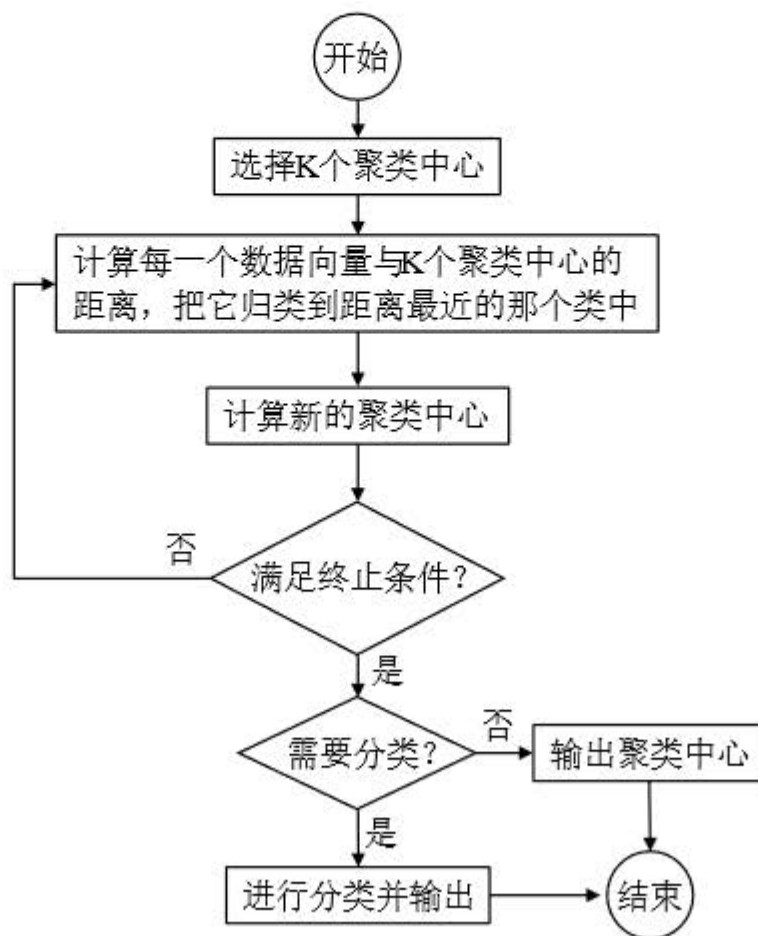
- ▶ 当服务器部署位置独立于接入点，或者研究接入点等基础设施部署问题时，首先需要执行区域离散化的步骤。
- ▶ 现有区域离散化方式可以归纳为两种：网格法和圆形区域求交。
- ▶ 圆形区域求交。



- ▶ 概括
- ▶ 边缘服务器部署问题的独特性
- ▶ 区域离散化
- ▶ 部署策略：聚类算法
- ▶ 部署策略：最优化算法
- ▶ 部署策略：图论相关算法

- ▶ 在边缘服务器部署位置的可选择空间确定后，进行部署算法的设计步骤。
- ▶ 常用的聚类算法：
 - k-means
 - DBSCAN
 - 高斯混合模型聚类
 - 层次聚类

- K-means算法是基于点之间的几何距离创建聚类的方法。



- ▶ DBSCAN (Density-based spatial clustering of applications with noise) 是一种基于密度的聚类算法，可以有效处理噪声点。
- ▶ 优势在于不需要事先确定簇的数量，可以很好地找到任意大小任意形状的簇。

- ▶ 高斯混合模型聚类（Gaussian Mixture Mode, GMM）是一种概率式的聚类算法。算法利用期望和标准差来描述簇的形状，每个簇使用一个高斯分布表示。
- ▶ 层次聚类算法的优势在于不需要指定簇的数量，算法对于距离度量标准的选择不敏感，另外，该算法对于具有层次结构的基础数据聚类效果更好，也可以用于恢复基础数据的层次结构。

- ▶ 概括
- ▶ 边缘服务器部署问题的独特性
- ▶ 区域离散化
- ▶ 部署策略：聚类算法
- ▶ 部署策略：最优化算法
- ▶ 部署策略：图论相关算法

- ▶ 大多数边缘计算服务器部署问题可建模为一个最优化问题。
- ▶ 优化目标：服务器和其负责接入点在单位距离下的平均负载最小化。
- ▶ 限制条件：
 - 服务器的候选位置限制在 n 个接入点附近
 - 每个接入点的负载由一个或多个服务器负责
 - 服务器负责不可超出其能力限制

目前使用最优化理论工具，解决边缘计算服务器静态部署问题的最优化问题可概括如下：

整数线性规划问题	混合整数线性规划问题
混合整数非线性规划问题	多目标优化问题

► 整数线性规划：

- 在介绍整数线性规划问题前，需要先明确线性规划（Linear Program, LP）问题的定义。
- LP问题是目标函数是线性函数，变量的约束条件也是线性约束，最终的决策变量可以取任何的实数的优化问题。
- 整数线性规划（Integer Linear Program, ILP）问题则在LP问题的基础上增加了决策变量是整数的限制。

- ▶ **混合整数线性规划**（Mixed Integer Linear Program, MILP）可视为LP问题和MLP问题的混合形式，决策变量中不仅有连续实数变量，也有离散整数变量。
- ▶ ILP和MILP问题都可以使用一个线性函数来表示优化目标，但是现实生活中大多数优化目标都是非线性的，这衍生了混合整数非线性规划（Mixed Integer Non-linear Program, MINLP）问题。
- ▶ 多目标优化指优化目标不止一个的最优化问题。需要注意多个优化目标之间可存在耦合，这类问题难以直接用单目标优化算法求解。

研究工作	优化目标	服务器能力	接入点负载	服务器数目	最优化问题	求解算法
Efficient Algorithms for Capacitated Cloudlet Placements	时延	上限	单服务器	固定	ILP	启发式
A Framework for the Joint Placement of Edge Service Infrastructure and User Plane Functions for 5G	综合KPI	上限	单服务器	可调	ILP	启发式
Cost Aware cloudlet Placement for big data processing at the edge	综合KPI	上限	多服务器	固定	MILP	启发式
Mobile edge computing resources optimization: A geo-clustering approach	资源利用率	最大化	多服务器	可调	MILP	图论算法
A hybrid cloudlet placement framework over passive optical access networks	财务成本	上限	多服务器	固定	MINLP	转换为MILP
Adaptive Edge-Centric Cloud Content Placement for Responsive Smart Cities	时延&负载均衡	上限	多服务器	固定	多目标优化	转换为单目标

- ▶ 概括
- ▶ 边缘服务器部署问题的独特性
- ▶ 区域离散化
- ▶ 部署策略：聚类算法
- ▶ 部署策略：最优化算法
- ▶ 部署策略：图论相关算法

大多数边缘计算服务器部署位置的选择问题，可以转换图论中的问题。如将边缘计算部署问题转换为图论中最小生成树（minimum spanning tree, MST）和最小支配集（minimum dominating set, MDS）问题。

- ▶ 部署问题的评价指标及限制条件
- ▶ 静态部署问题
- ▶ 动态部署问题

- ▶ 高移动性带来的挑战
- ▶ 动态服务器部署方案

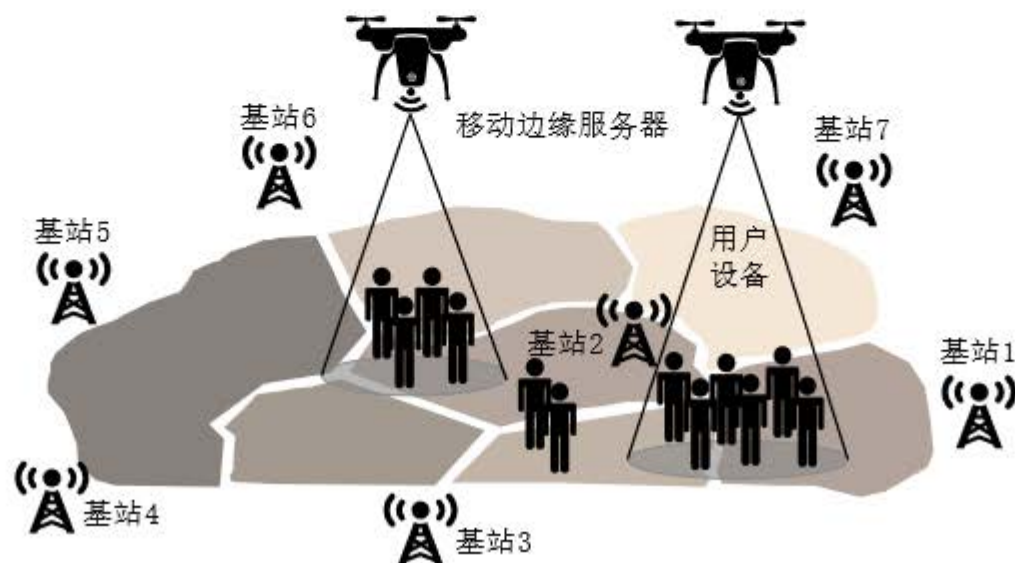
动态服务器为边缘计算场景服务器部署带来灵活性和便利性的同时，高移动性的引入也会带来新的挑战。

算法需要有自适应性	动态服务器移动时间通常远大于任务计算时间，时延敏感型任务的时间要求需要保证
服务器的路径规划（即动态服务部署）问题和资源分配、任务调度等经典问题耦合	动态服务器的计算能力不如静态服务器

- ▶ 高移动性带来的挑战
- ▶ 动态服务器部署方案

- ▶ 动静服务器混合部署方案
- ▶ 动态服务器路径规划
- ▶ 动态服务器移动性管理
- ▶ 部署方案性能评估工具

- ▶ 混合部署指既部署动态服务器也部署固定的静态服务器，利用动态服务器的灵活性对其进行调度，使得静态服务器无法满足用户需求的情况下动态服务器可为用户提供可靠服务。
- ▶ 维多利亚大学的Jingrong Wang团队利用UAV装载边缘服务器实现了服务器的移动性功能。

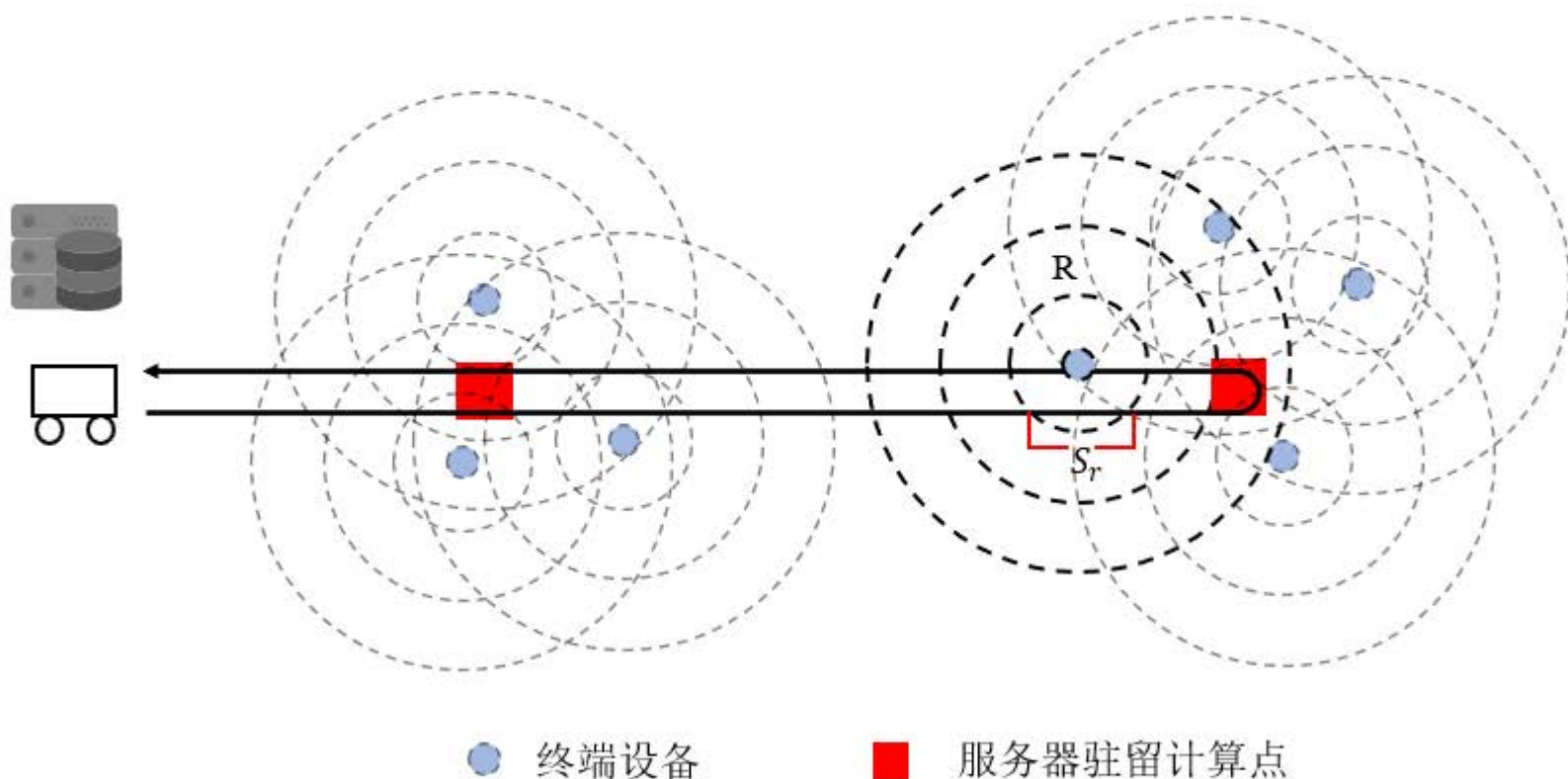


- ▶ 动静服务器混合部署方案
- ▶ 动态服务器路径规划
- ▶ 动态服务器移动性管理
- ▶ 部署方案性能评估工具

- ▶ 服务器移动路径是移动服务器需要重点规划的内容之一。
- ▶ 清华大学的Yu Liu团队提出了针对移动边缘计算服务器联合优化路径规划和资源分配的算法。
- ▶ 电子科技大学的Rong Cong等人提出动静服务器混合部署框架EdgeGO。

- ▶ 动静服务器混合部署方案
- ▶ 动态服务器路径规划
- ▶ 动态服务器移动性管理
- ▶ 部署方案性能评估工具

电子科技大学的赵志为团队提出了服务区速度控制算法。



- ▶ 动静服务器混合部署方案
- ▶ 动态服务器路径规划
- ▶ 动态服务器移动性管理
- ▶ 部署方案性能评估工具

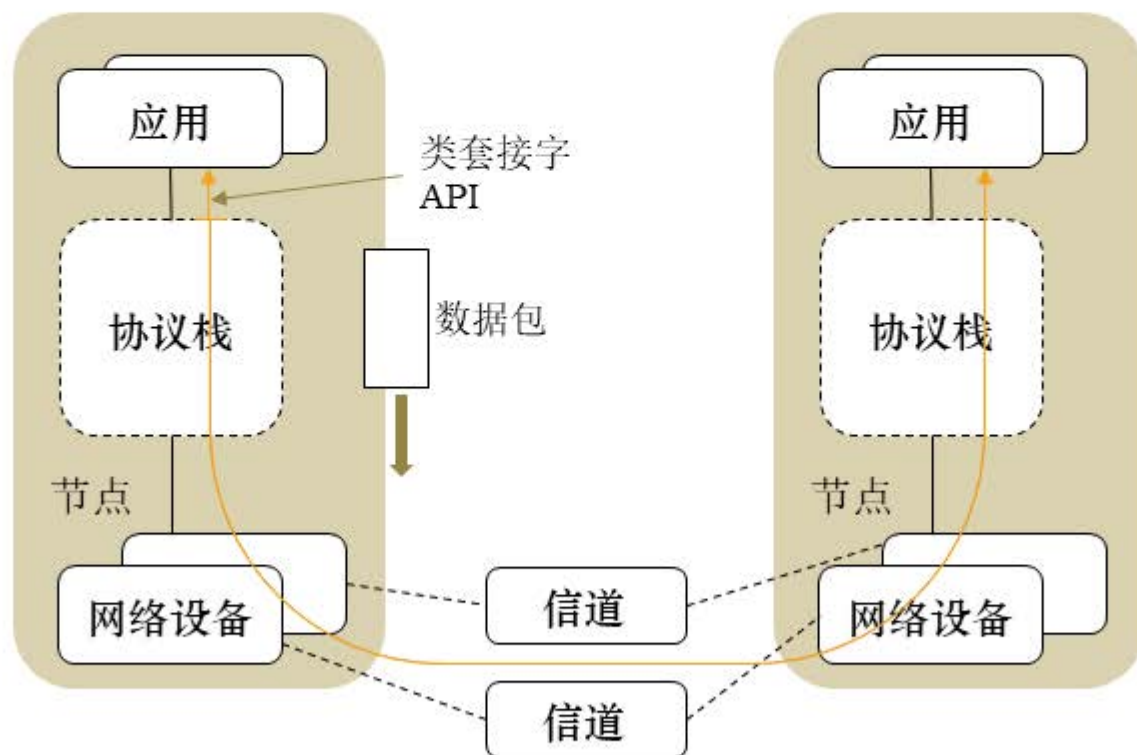
- ▶ **OPNET**：是最常用的网络仿真软件之一，由美国MIL3公司研发，主要针对有大型复杂网络仿真需求的网络设计专业人员而开发，功能非常完备强大。

ServiceProviderGuru	OPNET Modeler
ITGuru	WDM Guru

- ▶ NS (network simulator) 起源于1989年美国军方的Real Network Simulator项目, 由UC Berkeley开发, 是最早的网络仿真软件之一。
- ▶ 目前有NS-2和NS-3两个比较成熟的版本。

NS-2	NS-3
开源软件	开源软件
用户自定义模型困难	用户自定义模型容易
工具由大量用户贡献	工具由少量用户贡献
双语言编写 (C++、Otcl)	单语言编写 (C++)
编程能力要求高	编程能力要求适中
现有模型支持大量的协议	现有模型支持少量的协议

- ▶ NS-3具有高内聚低耦合的特点，主要包含六类网络组件。



- ▶ OPNET和NS仿真软件存在的内存开销极大问题，在SSFNet中得到有效处理。
- ▶ SSFNet（Scalable Simulation Framework Network）是一个基于Java SSF（Simple Seam Framework）组件合集，主要用于仿真IP层以上的网络模型，具有良好的扩展性，且支持对物理层和链路层的单独仿真。
- ▶ MiniNet是目前主流仿真SDN的平台，由斯坦福大学在2010年开发。软件基于Linux CONTAINER架构开发，可以创建一个包含主机、交换机、控制器和链路的虚拟网络，为SDN提供高灵活度的仿真支持。

- ▶ 边缘系统的部署场景可分为哪几类？每类部署场景的特点是什么？
- ▶ 请简要阐述边缘系统部署中常用的评价指标及限制条件。
- ▶ 边缘服务器部署问题和无线传感器网络中传感器节点的部署问题，有何异同？
- ▶ 简要概括动态部署策略的主要使用场景。
- ▶ 试针对边缘服务器静态部署问题进行数学建模，给出问题的最优化表示。

优化目标：最小化服务器和相关接入点单位距离下的平均负载。

限制条件：1) 接入点地点的优先级各异；2) 每个接入点的负载可由多个服务器协作负责；3) 每个服务器的负载能力有限。