数据中心的网络布局

研究成果：Data center network design for Internet-related service and cloud computing

作者：梁湧，Mengshi Lu, Zuojun Max Shen, 唐润宇

发表期刊：*Production and Operations Management*, 2021, 30(7), 2077-2101

近些年来，互联网相关服务和云计算产业得到了迅速的发展，对其背后的数据中心基础设施建设也提出了更高的需求。云计算和其他互联网相关服务市场机遇吸引了包括谷歌，亚马逊，微软和脸书在内的互联网巨头，这些巨头们已经开始逐渐开放其数据中心，向其他公司提供云计算服务。服务对象包括小型初创公司，也包括类似优步，网飞等估值数十亿的独角兽公司。

激烈的竞争迫使基础设施提供商提供质量更高，成本效益更优的服务。为了实现这一目标，服务提供商需要设计合理的数据中心网络，在保证服务质量的同时降低运营成本，从而增强其竞争优势。

尽管数据中心网络和传统供应链网络设计具有许多共同的特征，例如，两者都具有大型且昂贵的设施、运营成本对设施的位置以及设施与需求点之间的距离敏感等等，但数据中心网络的设计具备很多独有特点，也更具挑战性。首先，数据中心需要消耗大量电量，电能成本与数据中心内部的计算存储资源存量呈现非线性关系。其次，针对不同区域的居民特点，不同数据中心中的硬件资源配置需求不尽相同。最后，数据中心网络需要考虑需求点与数据中心见的网络传输延迟以及数据中心内部的处理数据所需的终端延迟时间。这些数据中心独有的特点为数据中心网络布局提出了挑战。

本研究通过整合优化数据中心的位置，需求分配以及数据中心内部计算、存储等资源供给决策，以达到总运营成本和服务延迟损失的最小化。在模型的拓展中也考虑了诸如非线性功耗、托管数据中心、多种资源配置限制以及相互依赖的需求等问题。本研究使用排队模型估计了数据中心内部的服务延迟，并对延迟函数进行变形，将网络设计的优化模型转化为混合整数二阶锥优化问题。为了提高大规模问题的计算效率，本文开发了基于拉格朗日松弛方法的两种优化算法，并利用问题的结构特性来生成加强割平面进一步加快求解速度。

通过收集相关的实际数据进行数值研究可以发现，整合优化数据中心选址、需求分配和计算、存储等资源供应的模型相比分阶段优化，可以显着降低成本并提高服务质量。通过敏感性分析，我们也得到了丰富的管理启示。比如当需求之间存在高度相互依赖关系时，最优的网络布局倾向将这些数据中心集中起来；单位网络传输延迟成本的增高总是倾向于建设更多的数据中心，然而单位终端延迟成本的增高会取决于数据中心的最大容量带来更多或更少的数据中心数量。此外，本文提出的两种优化算法相比于直接使用商业软件可以大幅度提高求解速度。

供稿：科研事务办公室　编辑：高晨卉　责编：吴淑媛 赵霞

School of Economics and Management
Tsinghua University
Beijing, China 100084
Tel 8610-6278 5515
Fax 8610-6278 5876
清华大学经济管理学院
中国·北京 100084
电话 8610-6278 5515
传真 8610-6278 5876