

The ever increasing demand for content is straining operators' networks, thereby necessitating development of alternative content delivery mechanisms. Distributed caching architectures constitute a promising solution for mitigating the effects of demand growth by placing popular content files in proximity to users, rather than in a central site. In this paper, we study the content placement problem in multiple-level hierarchical caching networks where user requests for files are routed upwards toward content servers, satisfied by intermediate nodes if the latter have cached the requested files. Our goal is to reduce the server load by serving as many requests as possible by the caches. We show that the problem is NP-Hard in its general form, but it can be solved optimally in polynomial-time when the caches are installed on a single hierarchy path. For the general case, we develop an algorithm achieving a provably better approximation ratio than the best-known counterparts. Numerical experiments show up to 56% reduction in server load over existing algorithms, with gains increasing as the content popularity distributions get steeper and more diverse across nodes, and the cache capacities at the upper hierarchy levels increase. Our evaluation code is publicly available for the benefit of research community