

Cache Networks: Opportunities and Challenges

Caching is an essential technique to improve throughput and latency in a vast variety of applications. There is a rich and beautiful theory, developed mostly in the computer science community during the 80s and 90s, for systems with a single cache. However, when it comes to networks of caches the existing theory falls short, and engineers instead rely on heuristics and the intuition gained from the analysis of single-cache systems. In addition, the contribution the underlying communication network (wireless or wired), connecting cache memories and users, has been basically unexplored.

The goal of this tutorial is to discuss some of the very recent results in this area, and highlights various opportunities and challenges in designing and using cache networks. In particular, we will argue that when the caching strategy is designed jointly with the delivery scheme, considering the properties and limits of the underlying communication networks, the overall performance of the system can substantially improved. This improvement can even scale with the size of the networks. The new designs are based on developing and exploiting the state of art in information theory, coding, communications, and networking. These recent developments have motivated researchers in academia and industry to investigate the role caching in achieving the demanding requirements of 5G, the next generation of wireless communication.