

**Utility-Centric Networking: Balancing Transit Costs With Quality of Experience**

**Abstract:**

This paper is focused on techniques for maximizing utility across all users within a total network transit cost budget. We present a new method for selecting between replicated servers distributed over the Internet. First, we introduce a novel utility framework that factors in quality of service metrics. Then we design an optimization algorithm, solvable in polynomial time, to allocate user requests to servers based on utility while satisfying network transit cost constraints, mapping service names to service instance locators. We then describe an efficient, low overhead distributed model which only requires knowledge of a fraction of the data required by the global optimization formulation. Next, a load-balancing variant of the algorithm is explored that substantially reduces blocking caused by congested servers. Extensive simulations show that our method is scalable and leads to higher user utility compared with mapping user requests to the closest service replica, while meeting network traffic cost constraints. We discuss several options for real-world deployment that require no changes to end-systems based on either the use of SDN controllers or extensions to the current DNS system.