Title: ''Content Delivery Systems with Coded Caching; algorithms for asynchronous coded caching and lower bounds on the synchronous case''

Abstract:

Caching is often used in content delivery networks as a mechanism for reducing network traffic. Recently, the technique of coded caching was introduced whereby coding in the caches and coded transmission signals from the central server were considered. Prior results in this area demonstrate that carefully designing the placement of content in the caches and designing appropriate coded delivery signals from the server allow for a system where the delivery rates can be significantly smaller than conventional schemes.

However, matching upper and lower bounds on the transmission rate have not yet been obtained. In the first part of this thesis we derive tighter lower bounds on the coded caching rate than were known previously. We demonstrate that this problem can equivalently be posed as a combinatorial problem of optimally labeling the leaves of a directed tree. Our proposed labeling algorithm allows for significantly improved lower bounds on the coded caching rate. Furthermore, we study certain structural properties of our algorithm that allow us to analytically quantify improvements on the rate lower bound for general values of the problem parameters. This allows us to obtain a multiplicative gap of at most four between the achievable rate and our lower bound.

The original formulation of the coded caching problem assumes that the file requests from the users are synchronized, i.e., they arrive at the server at the same time. Several subsequent contributions work under the same assumption. Furthermore, the majority of prior work does not consider a scenario where users have deadlines. In the second part of this thesis we formulate the asynchronous coded caching problem where user requests arrive at different times. Furthermore, the users have specified deadlines. We propose a linear program for obtaining its optimal solution. However, the size of the LP (number of constraints and variables) grows rather quickly with the number of users and cache sizes. To deal with this problem, we explore a dual decomposition based approach for solving the LP under consideration. We demonstrate that the dual function can be evaluated by equivalently solving a number of minimum cost network flow algorithms.

Moreover, we consider the asynchronous setting where the file requests are revealed to the server in an online fashion. We propose a novel online algorithm for this problem building on our prior work for the offline setting (where the server knows the request arrival times and deadlines in advance). Our simulation results demonstrate that our proposed online algorithm allows for a natural tradeoff between the feasibility of the schedule and the rate gains of coded caching.