Joint Data Compression and Caching: Approaching Optimality with Guarantees

Jian Li (University of Massachusetts Amherst), **Faheem Zafari** (Imperial College London), **Don Towsley** (University of Massachusetts Amherst), **Kin K. Leung** (Imperial College London), **Ananthram Swami** (US. Army Research Laboratory).

Contributions

We consider the problem of optimally compressing and caching data across a communication network:

- We propose a formal mathematical framework for joint data compression and cache optimization. We formulate the problem of finding optimal data compression ratios and caching locations that minimize average delay in serving requests subject to an energy constraint.
- We analyze the complexity of the problem and show that it is NP-hard in general. The hardness is caused by data allocation to the caches.
- We propose polynomial time solvable algorithms for the formulated problem. Since the original problem is NP-hard and non-convex, we relax the constraints and show that the relaxed problem can be transformed into an equivalent convex optimization problem that can be solved in polynomial time. We then show that combining this solution with greedy caching allocation achieves a solution with ½-approximation to the optimum. Moreover, we construct a polynomial-time 1-1/e approximation algorithm for the problem.
- We conduct extensive simulations using synthetic network topologies and compare our proposed algorithm with benchmark techniques. Our results show that the proposed algorithm achieves near-optimal performance, and significantly outperforms benchmarks Generic algorithm, Bonmin, and NOMAD by obtaining a feasible solution in less time for various network topologies.

Formulation

$$\min \quad L(\delta, b) = \sum_{k \in \mathcal{K}} \sum_{i=0}^{h(k)-1} \prod_{m=i+1}^{h(k)} \delta_{k,m} y_k R_k l_{i,i+1}^{k} \prod_{j=0}^{i} (1 - b_{k,j})$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{K}} \sum_{i=0}^{h(k)} y_k \left\{ R_k f(\delta_{k,i}) \prod_{m=i+1}^{h(k)} \delta_{k,m} + \left(\prod_{m=i}^{h(k)} \delta_{k,m} \right) b_{k,i} \cdot (w_{ca} T + (R_k - 1)\varepsilon_{kT}) \right\} \leq W,$$

$$b_{k,i} \in \{0,1\}, \forall k \in \mathcal{K}, i = 0, \cdots, h(k),$$

$$\sum_{k \in C_v} b_{k,h(v)} y_k \prod_{j=h(k)}^{h(v)} \delta_{k,j} \leq S_v, \forall\, v \in V,$$

$$\sum_{i=0}^{h(k)} b_{k,i} \leq 1, \forall k \in \mathcal{K}.$$

Evaluation

Nodes	Proposed		GA		Nomad		Bonmin	
	Obj. Value	Time(s)	Obj. Value	Time(s)	Obj. Value	Time(s)	Obj. Value	Time(s)
7	480000	3.30	479820	273.29	Infeasible	9.45	Infeasible	1.01
15	1440000	6.33	1440000	15.12	1439900	16.18	Infeasible	> 4000
31	3840000	29.28	3839000	3501.10	Non-Convergence	98.90	Non-Convergence	1232.31
63	9599900	538.17	8792100	158.56	Non-Convergence	966.16	Non-Convergence	2.04