Relaxations of the Delay-Constrained Maximum Concurrent Flow Problem

Guillaume Beraud-Sudreau[1,3], Walid Ben-Ameur[2], Sébastien Martin[1], and Hervé Kérivin[3]

[1]Huawei Technologies France, Paris Research Center, Boulogne-Billancourt, France, ✉ guillaume.beraud-sudreau@doctorant.uca.fr
[2]Samovar, Télécom Sud-Paris, Saclay, France,
[3]LIMOS, Université Clermont-Auvergne, Clermont-Ferant, France

04/2026

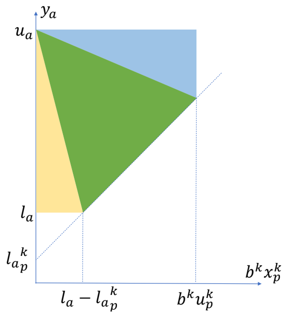
The Delay-constrained Multi-commodity Flow Problem

The expansion of cloud computing, from real-time gaming to distributed LLM training, has made predictable end-to-end latency a critical telecommunications requirement. This requirement can be formulated as constraints of a traffic-engineering problem [1] on a directed graph $D = (V, A)$, where commodities $k \in K$ with sizes b^k are routed from source s^k to destination t^k via a set of paths P^k. Flow proportions x_p^k determine the load $y_a = \sum_{p \ni a} x_p^k b^k$ on arcs with capacity c_a, driving latency according to the M/M/1 model $d_a = \frac{1}{c_a - y_a}$. Ensuring that the total delay on any active path remains within a bound d^k, the Delay-Constrained Maximum Concurrent Flow (DCMCF) problem maximizes the throughput factor γ:

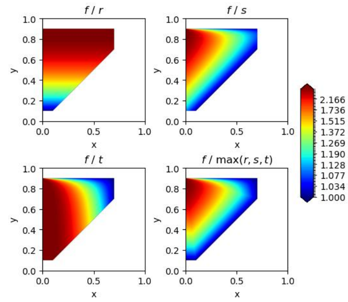
$$
\begin{aligned}
\text{DCMCF:} \quad & \max \gamma \\
& \sum_{p \in P^k} x_p^k \geq \gamma && \forall k \in K && (1) \\
& y_a \geq \sum_{k \in K, p \in P^k | a \in p} x_p^k b^k && \forall a \in A && (2) \\
& \sum_{a \in p} \frac{1}{c_a - y_a} \leq d^k \qquad \text{if } x_p^k > 0 && \forall k \in K, p \in P^k && (3) \\
& y_a \leq c_a && \forall a \in A && (4) \\
& x_p^k \geq 0 && \forall k \in K, p \in P^k. && (5)
\end{aligned}
$$

The objective maximizes the throughput factor γ (Eq. 1). Constraints aggregate path flows into arc loads (Eq. 2) and enforce capacity limits (Eq. 4), while (Eq. 3) ensures that cumulative delays on active paths adhere to the specified bounds, implicitly keeping loads strictly below capacity to prevent infinite latency.

The conditional nature of the delay constraints (3) (which are active only if the corresponding path is actually used) is particularly interesting. Such conditional constraints have been addressed in previous work by introducing a binary variable z_p^k for every commodity $k \in K$ and path $p \in P^k$, that equals to 1 if $x_p^k > 0$ and zero otherwise ; this can be achieve via constraints $x_p^k \leq z_p^k u_p^k$, with u_p^k an upper-bound of variable x_p^k. These conditional constraints make the problem NP-hard, as shown in [1]. Some convex relaxations are also proposed there. In [3], a big-M reformulation of the DCMCF is solved via an outer-approximation: the delay constraints are replaced by constraints $\sum_{a \in p} \frac{1}{c_a - y_a} \leq d^k + (1 - z_p^k)M$, with M large enough. This relaxation is relatively weak but proves efficient. In [2], a tighter relaxation, relying on a disjunctive programming approach, is considered: the authors compute the convex envelope of the feasible solutions $\left\{(y_a)_{a \in p}, z_p^k : \sum_{a \in p} \frac{1}{c_a - y_a} \leq d^k, z_p^k = 1\right\} \cup \left\{(y_a)_{a \in p}, z_p^k : y_a \leq c_a, z_p^k = 0\right\}$ and leverage it to obtain a convex relaxation of the DCMCF problem.



(a) Presentation of the $\mathcal{X}$ domain as the union of 3 subsets



(b) Multiplicative errors of the 3 convex under-estimators of f_a combined to build $\hat{f}_{k,p,a}$; r, s and t respectively represent the 1^{st}, 2^{nd} and 3^{rd} terms of the max.

Relaxation of the delay constraints

We propose to build a relaxation of the delay constraints directly in the space of the original variable, by reformulating them as $\sum_{a \in p} \frac{x_p^k}{c_a - y_a} \leq d^k x_p^k$.

This reformulation is equivalent to the original delay constraints (3) - indeed, if $x_p^k = 0$ the x_p^k terms simplify and the delay must be satisfied, and if $x_p^k = 0$ the constraints is trivially verified. Transitioning from a conditional to an algebraic constraints facilitates the use of convex analysis; specifically, $f_a(x_p^k, y_a) = \frac{x_p^k}{c_a - y_a}$ permits the derivation of a tight relaxation despite its inherent non-convexity.

In order to build a tight relaxation of the delay constraints, the domain of variables (x_p^k, y_a) is restricted: $\mathcal{X}_{k,p,a} = \left\{ (x_p^k, y_a) \mid y_a \in [l_a, u_a], x_p^k \in [0, u_p^k], y_a \geq b^k x_p^k + {l_a}_p^k \right\}$, with l_a, u_a, u_p^k and ${l_a}_p^k$ the bounds of variables y_a, x_p^k and $(y_a - b^k x_p^k)$. Figure 1a provides a visualization of the domain $\mathcal{X}_{k,p,a}$ and its partition in 3 triangular subsets. Based on this decomposition of the delay's functions domain, we compute the convex envelope $\hat{f}_{k,p,a}$ of $f_{k,p,a}$ as:

$$\hat{f}_{k,p,a}(x_p^k, y_a) = \max \left(\frac{x_p^k}{c_a - l_a}, \frac{x_p^k}{c_a - {l_a}_p^k - \frac{(u_a - {l_a}_p^k)}{u_a - (y_a - b^k x_p^k)} b^k x_p^k}, \frac{{x_p^k}^2}{c_a x_p^k - u_p^k y_a + (u_p^k - x_p^k) u_a} \right) \tag{6}$$

This relaxation allows to relax the delay constraints into the convex constraints $\sum_{a \in p} \hat{f}_{k,p,a}(x_p^k, y_a) - x_p^k d^k \leq 0$. We prove that this relaxation dominate the one computed in [2], as the later is equivalent to omitting the second term of the maximum in the definition of $\hat{f}_{k,p,a}$.

Experimental results allow us to compare the different relaxations, both as upper-bounds and within a branch-and-bound scheme to solve the exact solutions of the optimization problem.

Approximation algorithms

An heuristic with performance guarantees is proposed. After solving a convex relaxation based on the constraints (6) above, we only keep paths whose x_p^k values exceed some optimized threshold in the relaxed solution. Then we solve the convex problem by integrating the delay constraints for the selected paths, thereby obtaining a feasible solution. Properties of the newly proposed relaxation described above allows us to prove that the worst-case performance guarantee is at most $\frac{1}{\alpha\beta|K|(|P|-1)+1}$, where α is any fixed scalar greater than 1, $\beta = \frac{\max_{k \in K} b^k}{\min_{k \in K} b^k}$, and $|P| = \max_{k \in K} |P^k|$.

References

[1] Walid Ben-Ameur and Adam Ouorou. Mathematical models of the delay constrained routing problem. *Algorithmic Operations Research*, 1(2):94–103, 2006.

[2] Hassan Hijazi, Pierre Bonami, Gérard Cornuéjols, and Adam Ouorou. Mixed-integer nonlinear programs featuring "on/off" constraints. *Computational Optimization and Applications*, 52(2):537–558, 2012.

[3] Hassan Hijazi, Pierre Bonami, and Adam Ouorou. An outer-inner approximation for separable minlps. *LIF*, 2010.