Paper

Integrated Routing and Network Flow Control Embracing Two Layers of TCP/IP Networks – Methodological Issues

Andrzej Karbowski[a,b]

[a] Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland
[b] Research and Academic Computer Network (NASK), Warsaw, Poland

Abstract—A cross-layer network optimization problem is considered. It involves network and transport layers, treating both routing and flows as decision variables. Due to the nonconvexity of the capacity constraints, when using Lagrangian relaxation method a duality gap causes numerical instability. It is shown that the rescue preserving separability of the problem may be the application of the augmented Lagrangian method, together with Cohen's Auxiliary Problem Principle.

Keywords—decomposition, flow control, Lagrangian relaxation, networks, optimization, routing, TCP/IP.

1. Introduction

In the standard approach TCP congestion control together with active queue management (AQM) algorithms attempted to maximize aggregated utility over source rates, assuming that routing is given and fixed at the timescale of interest. However, it seems that it would be more profitable, when we will treat TCP and IP layers together and maximize cross-layer utility at the timescale of route changes. The integrated routing and network flow control problem was first addressed by Wang, Li, Low and Doyle [1] and independently by Jaskóła and Malinowski [2]. Unfortunately, due to the nonconvexity of the constraints' functions, the algorithm based on the price method (Lagrangian relaxation) is numerically unstable. Duality gap is the reason of problems [1]. The paper shows how this gap can be overcome, while not losing separability of the problem.

2. Problem Formulation

Our goal is to maximize the sum of utilities of all connections with respect to routing and flows over the whole network, taking into account the capacities of links.

Formally, the optimization problem can be described as follows:

$$\max_{x\in X, R\in \mathcal{R}} \sum_{s\in S} U_s(x_s)\,, \tag{1}$$

$$Rx \le c,\; R = [r_{ij}]_{\overline{\overline{L}}\times\overline{\overline{S}}}\,, \tag{2}$$

where:

- x_s – flow from the source s to a (single) destination node;
- $x \in X \subset \mathbb{R}^{\overline{\overline{S}}}$ – vector of all flows;
- S – the set of all sources;
- X – the set of admissible flows; it is a Cartesian product of intervals X_s belonging to nonnegative half lines;
- U_s – the sources' (connections') utility functions; it is assumed, that they are strictly concave and continuous;
- L – the set of all links;
- R – the matrix of binary elements with the number of rows equal the number of links $\overline{\overline{L}}$ and the number of columns equal the number of sources (active connections at a given time); the element r_{ls} equals 1 when the link l belongs to a path from the source s to a given destination node;
- R_s – s-th column of the matrix R;
- $\mathcal{R}_s$ – the set of all possible vectors representing paths from s to a given destination node;
- $\mathcal{R}$ – the set of all possible matrices, that is all possible combinations of vectors from the sets $\mathcal{R}_s$;
- $c \in \mathbb{R}_+^{\overline{\overline{L}}}$ – links capacity vector.

3. The Standard Price Decomposition Method

The Lagrangian for the problem (1)–(2) is as follows:

$$L(x,R,\lambda) = \sum_{s\in S} U_s(x_s) - \sum_{l\in L}\lambda_l\left(\sum_{s\in S} r_{ls}x_s - c_l\right)$$
$$= \sum_{s\in S}\left(U_s(x_s) - x_s\sum_{l\in L}\lambda_l r_{ls}\right) + \sum_{l\in L}\lambda_l c_l\,, \tag{3}$$

where λ_l are nonnegative Lagrange multipliers. Due to the duality theory, this Lagrangian will be further maximized with respect to x and R and minimized with respect to λ [3]. The iteration in the standard price method consists of two-steps [1]:

1. Solve the primal problem

$$(x(t),R(t)) = \arg \max_{x\in X, R\in\mathcal{R}} \sum_{s\in S}\left(U_s(x_s) - x_s \sum_{l\in L}\lambda_l(t) r_{ls}\right). \tag{4}$$

   Let us notice that owing to the specific structure and the nonnegativity of x_s for all s the overall optimization problem (4) can be decomposed in the following way:

$$\max_{x\in X}\max_{R\in\mathcal{R}} \sum_{s\in S}\left(U_s(x_s) - x_s \sum_{l\in L}\lambda_l(t) r_{ls}\right) =$$
$$= \sum_{s\in S}\max_{x_s\in X_s}\left[U_s(x_s) + \max_{R_s\in\mathcal{R}_s}\left(-x_s\sum_{l\in L}\lambda_l(t) r_{ls}\right)\right]$$
$$= \sum_{s\in S}\max_{x_s\in X_s}\left[U_s(x_s) - x_s \min_{R_s\in\mathcal{R}_s}\left(\sum_{l\in L}\lambda_l(t) r_{ls}\right)\right]. \tag{5}$$

   From the final form of Eq. (5) it is seen, that [1]:

   – the primal problem (4) can be decomposed into a family of problems assigned to subsequent sources s with local variables $x_s, r_{1s}, r_{2s}, \dots$ which can be solved independently,
   – the inner optimization $\min_{R_s\in\mathcal{R}_s}\sum_{l\in L}\lambda_l(t) r_{ls}$ for the given source index s (and its connection to a destination node) is simply the shortest path problem with metrics defined by Lagrange multipliers $\lambda_l(t)$, $l = 1,2,\dots$.

   Summing up, the problem (4) may be solved by solving for every source $s \in S$:

   - The shortest path problem:

$$R_s(t) = \arg\min_{R_s\in\mathcal{R}_s}\sum_{l\in L}\lambda_l(t) r_{ls}. \tag{6}$$

     Let us denote the optimal value of the performance index in Eq. (6) as $d_s(t)$, that is:

$$d_s(t) = \sum_{l\in L}\lambda_l(t) r_{ls}(t). \tag{7}$$

   - The flow optimization problem:

$$\max_{x_s\in X_s}\left(U_s(x_s) - x_s d_s(t)\right). \tag{8}$$

2. Modify Lagrange multipliers so as to get a better approximation of the solution of the dual problem $\min_{\lambda\geq 0}[L_D(\lambda) = \max_{x\in X, R\in\mathcal{R}} L(x,R,\lambda)]$

$$\lambda_l(t+1) = \max\left(0, \lambda_l(t) + \rho\left(\sum_{s\in S} r_{ls}(t)x_s(t) - c_l\right)\right),\ l\in L, \tag{9}$$

   where $\rho > 0$ is a properly chosen step coefficient.

Unfortunately, this algorithm is unstable [1]. The reason is a duality gap caused by the nonconvexity of capacity constraint (2) and the discrete character of variables r_{ls}.

4. Augmented Lagrangian Approach and Auxiliary Problem Principle in Cross-Layer Optimization

In optimization problems where the duality gap is present, we use augmented Lagrangian or, in other words, shifted penalty function method [3], [4], [5]. For the problem (1)–(2) it will have the form:

$$L_a(x,R,\lambda) = \sum_{s\in S}U_s(x_s) - \frac{1}{2}\sum_{l\in L}\rho_l\left\{\left[\max\left(0, \left(\sum_{s\in S} r_{ls}x_s - c_l\right) + \frac{\lambda_l}{\rho_l}\right)\right]^2 - \left(\frac{\lambda_l}{\rho_l}\right)^2\right\} = \sum_{s\in S}U_s(x_s) +$$
$$-\frac{1}{2}\sum_{l\in L}\frac{\rho_l}{\rho_l^2}\left\{\left[\rho_l \max\left(0, \left(\sum_{s\in S} r_{ls}x_s - c_l\right) + \frac{\lambda_l}{\rho_l}\right)\right]^2 - \rho_l^2\left(\frac{\lambda_l}{\rho_l}\right)^2\right\} = \sum_{s\in S}U_s(x_s) +$$
$$-\frac{1}{2}\sum_{l\in L}\frac{1}{\rho_l}\left\{\left[\max\left(0, \lambda_l + \rho_l\left(\sum_{s\in S} r_{ls}x_s + -c_l\right)\right)\right]^2 - \lambda_l^2\right\}, \tag{10}$$

where $\rho_l, l \in L$ are penalty coefficients.

The solution of the problem (1)-(2) is sought, as before, by solving the minimax problem:

$$\min_{\lambda\geq 0}\max_{x\in X, R\in\mathcal{R}} L_a(x,R,\lambda). \tag{11}$$

Augmented Lagrangians have one serious drawback – due to the quadratic terms (in our case – squares of the sums of products of variables) they are not separable, that is the optimization problem is not decomposable.

The easiest way to transform the augmented Lagrangian to a separable form consists in the application of so-called Auxiliary Problem Principle proposed by Cohen [6], [7]. This principle says, that if we want to solve the problem:

$$\max_{u\in U} J_1(u) + J_2(u), \tag{12}$$

where J_1 is an additive (that is separable), strictly concave functional, while J_2 is a differentiable, nonadditive, not necessarily strictly concave, functional, we may instead solve a sequence of auxiliary problems:

$$u(t+1) = \arg\max_{u\in U}\left[G_\varepsilon^{u(t)}(u) = \varepsilon J_1(u) + \varepsilon < J_2'(u(t)), u > - K(u) + < K'(u(t)), u >\right]. \tag{13}$$

In the above expression, $< .,. >$ denotes the scalar product, $\varepsilon > 0$ – a constant parameter, t is the index of iteration and

$$K(u) = ||u||_2^2. \tag{14}$$

In short, the idea of this transformation lies in the linearization of the nonseparable component and addition of

a regularizing, strictly concave, proximal component (more precisely, the subtraction of a strictly convex proximal component $||u-u(t)||_2^2$, with accuracy to the constant $||u(t)||_2^2$, which does not influence the optimization).
In the case of our problem (1)–(2) with the augmented Lagrangian Eq. (10):

$$u = \begin{bmatrix} x \\ R \end{bmatrix}, \tag{15}$$

$$J_1(u) = \sum_{s\in S} U_s(x_s), \tag{16}$$

$$J_2(u) = -\frac{1}{2}\sum_{l\in L}\frac{1}{\rho_l}\left\{\left[\max\left(0,\lambda_l+\rho_l\left(\sum_{s\in S} r_{ls}x_s - c_l\right)\right)\right]^2 - \lambda_l^2\right\} \tag{17}$$

and

$$\begin{aligned} G_\varepsilon^{u(t)}(u) = \varepsilon\sum_{s\in S}U_s(x_s) - \varepsilon\sum_{s\in S}\left\{\sum_{l\in L}\left[\max\left(0,\lambda_l + \right.\right.\right. \\ \left.+\rho_l\left(\sum_v r_{lv}(t)x_v(t)-c_l\right)\right) r_{ls}(t)x_s + \\ \left.\left.+\max\left(0,\lambda_l+\rho_l\left(\sum_v r_{lv}(t)x_v(t)-c_l\right)\right)x_s(t)r_{ls}\right]\right\} + \\ -\sum_{s\in S}\left(x_s^2+\sum_{l\in L}r_{ls}^2\right) + 2\sum_{s\in S}x_s(t)x_s + 2\sum_{s\in S}\sum_{l\in L}r_{ls}(t)r_{ls}. \end{aligned} \tag{18}$$

5. Decomposition Scheme and the Algorithm

Grouping together and rearranging terms dependent on the same variables in (18), we will get:

$$\begin{aligned} G_\varepsilon^{u(t)}(u) = \sum_{s\in S}\left[\varepsilon U_s(x_s) - x_s^2 + 2x_s(t)x_s + \right. \\ \left. -\varepsilon\sum_{l\in L}\max\left(0,\lambda_l+\rho_l\left(\sum_v r_{lv}(t)x_v(t)-c_l\right)\right)r_{ls}(t)x_s\right] + \\ -\sum_{s\in S}\sum_{l\in L}\left[r_{ls}^2 - 2r_{ls}(t)r_{ls} + \varepsilon\max\left(0,\lambda_l + \right.\right. \\ \left.\left. +\rho_l\left(\sum_v r_{lv}(t)x_v(t)-c_l\right)\right)x_s(t)r_{ls}\right]. \end{aligned} \tag{19}$$

Let us notice that for $r_{ls}\in\{0,1\}$, $r_{ls}^2 = r_{ls}$, so we will finally get:

$$\begin{aligned} G_\varepsilon^{u(t)}(u) = \sum_{s\in S}\left\{\varepsilon U_s(x_s) - x_s^2 + \left[2x_s(t) + \right.\right. \\ \left.\left. -\varepsilon\sum_{l\in L}\max\left(0,\lambda_l+\rho_l\left(\sum_v r_{lv}(t)x_v(t)-c_l\right)\right)r_{ls}(t)\right]x_s\right\} + \\ -\sum_{s\in S}\sum_{l\in L}\left\{\left[1-2r_{ls}(t)+\varepsilon\max\left(0,\lambda_l + \right.\right.\right. \\ \left.\left.\left. +\rho_l\left(\sum_v r_{lv}(t)x_v(t)-c_l\right)\right)x_s(t)\right]r_{ls}\right\}. \end{aligned} \tag{20}$$

Let us denote now:

$$\begin{aligned} V_s^{u(t)}(x_s,\lambda,\varepsilon,\rho) = \varepsilon U_s(x_s) - x_s^2 + \\ +\left[2x_s(t) - \varepsilon\sum_{l\in L}\max\left(0,\lambda_l+\rho_l\left(\sum_v r_{lv}(t)x_v(t)-c_l\right)\right)r_{ls}(t)\right]x_s, \end{aligned} \tag{21}$$

$$\begin{aligned} \varphi_{ls}^{u(t)}(\lambda_l,\varepsilon,\rho_l) = 1 - 2r_{ls}(t) + \\ +\varepsilon\max\left(0,\lambda_l+\rho_l\left(\sum_v r_{lv}(t)x_v(t)-c_l\right)\right)x_s(t). \end{aligned} \tag{22}$$

With this notation the function $G_\varepsilon^{u(t)}(u)$ can be written as:

$$G_\varepsilon^{u(t)}(u) = \sum_{s\in S}V_s^{u(t)}(x_s,\lambda,\varepsilon,\rho) - \sum_{s\in S}\sum_{l\in L}\varphi_{ls}^{u(t)}(\lambda_l,\varepsilon,\rho_l)r_{ls}, \tag{23}$$

and the primal optimization problem $\max\limits_{x\in X, R\in\mathcal{R}} L_a(x,R,\lambda)$ with the augmented Lagrangian (10) is equivalent to the following auxiliary problem:

$$\begin{aligned} \max_{x\in X, R\in\mathcal{R}}\left[G_\varepsilon^{u(t)}(u) = \sum_{s\in S}V_s^{u(t)}(x_s,\lambda,\varepsilon,\rho) - \sum_{s\in S}\sum_{l\in L}\varphi_{ls}^{u(t)}(\lambda_l,\varepsilon,\rho_l)r_{ls}\right] = \\ = \max_{x\in X}\sum_{s\in S}V_s^{u(t)}(x_s,\lambda,\varepsilon,\rho) - \min_{R\in\mathcal{R}}\sum_{s\in S}\sum_{l\in L}\varphi_{ls}^{u(t)}(\lambda_l,\varepsilon,\rho_l)r_{ls} = \\ = \sum_{s\in S}\left[\max_{x_s\in X_s}V_s^{u(t)}(x_s,\lambda,\varepsilon,\rho) - \min_{R_s\in\mathcal{R}_s}\sum_{l\in L}\varphi_{ls}^{u(t)}(\lambda_l,\varepsilon,\rho_l)r_{ls}\right]. \end{aligned} \tag{24}$$

Let us notice that the structure of the problem (24) is very similar to the problem (4), but the decomposition scheme goes further, because actually for a given λ we got a complete separation of the shortest path problems (variables r_{ls}), from the flow optimization problems (variables x_s).
The simplest gradient steepest descent algorithm of modification of the Lagrange multipliers due to Eq. (10) will be the following:

$$\begin{aligned} \lambda_l(t+1) = \lambda_l(t)\left(1-\frac{\beta}{\rho_l}\right) + \\ +\frac{\beta}{\rho_l}\max\left(0,\lambda_l(t)+\rho_l\left(\sum_{s\in S}r_{ls}(t)x_s(t)-c_l\right)\right). \end{aligned} \tag{25}$$

The values of parameters should be chosen from the intervals [7]:

$$0<\beta\le\min_{l\in L}\rho_l,\quad 0<\varepsilon<\frac{b}{\tau^2\max\limits_{l\in L}\rho_l}, \tag{26}$$

where b,τ are, respectively, Lipschitz constants of the function K (14) and the constraint function (2).
Summing up, the iteration of the modified, based on augmented Lagrangian approach, algorithm will be as follows:

1. Solve the primal problem, decomposed into the family of independent problems for every source $s\in S$:

$$x_s(t) = \arg\max_{x_s\in X_s}V_s^{u(t)}(x_s,\lambda(t),\varepsilon,\rho), \tag{27}$$

$$r_{ls}(t) = \arg\min_{R_s\in\mathcal{R}_s}\sum_{l\in L}\varphi_{ls}^{u(t)}(\lambda_l(t),\varepsilon,\rho_l)r_{ls}. \tag{28}$$

Functions $V_s^{u(t)}$ and $\varphi_{ls}^{u(t)}$ are defined by Eqs. (21) and (22).

2. Modify Lagrange multipliers for all links $l \in L$

$$\lambda_l(t+1) = \lambda_l(t)\left(1 - \frac{\beta}{\rho_l}\right) + \\ + \frac{\beta}{\rho_l} \max\left(0, \lambda_l(t) + \rho_l\left(\sum_{s\in S} r_{ls}(t)x_s(t) - c_l\right)\right). \quad (29)$$

The presented approach was implemented and thoroughly tested on many big networks generated by Netgen [8]. The results proved its high effectiveness [9], [10].

References

[1] J. Wang, L. Li, S. H. Low, and J. C. Doyle, "Cross-layer optimization in TCP/IP networks", *IEEE/ACM Trans. Networking*, vol. 13, no. 3, pp. 582–595, 2005.

[2] P. Jaskóła and K. Malinowski, "Two methods of optimal bandwidth allocation in TCP/IP networks with QoS differentiation", in *Proc. Summer Simulation Multiconf. SPECTS 2004*, San Jose, California, 2004, pp. 373–378.

[3] D. P. Bertsekas, *Lagrange Multiplier Methods in Constrained Optimization*. Academic Press, 1982.

[4] A. P. Wierzbicki, "A penalty function shifting method in constrained static optimization and its convergence properties", *Archiwum Automatyki i Telemechaniki*, vol. 16, pp. 395–416, 1971.

[5] R. T. Rockafellar, "Augmented Lagrange multiplier functions and duality in nonconvex programming", *SIAM J. Control*, vol. 12, no. 2, pp. 268–285, 1974.

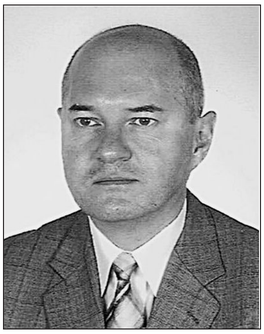
[6] G. Cohen, "Optimization by decomposition and coordination: a unified approach", *IEEE Trans. Autom. Control*, vol. AC-23, no. 2, pp. 222–232, 1978.

[7] G. Cohen, D. L. Zhu, "Decomposition coordination methods in large scale optimization problems. The nondifferentiable case and the use of augmented Lagrangians", in *Advances in Large Scale Systems, Vol. I*, J. B. Cruz, Ed. Greenwich, Connecticut: JAI Press, 1984, pp. 203–266.

[8] D. Klingman, A. Napier, and J. Stutz, "NETGEN: a program for generating large scale capacitated assignment, transportation, and minimum cost flow network problems", *Management Sci.*, vol. 20, no. 5, pp. 814–821, 1974.

[9] P. Paluch, "Cross-layer optimization in TCP/IP networks – the augmented Lagrangian approach assessment", B.Eng. thesis, Warsaw University of Technology, 2011.

[10] A. Karbowski and P. Paluch, "Integrated routing and network flow control embracing two layers of TCP/IP networks – implementation and tests results" in preparation.



Andrzej Karbowski received the M.Sc. degree in Electronic Engineering (specialization automatic control) from Warsaw University of Technology (Faculty of Electronics) in 1983. He received the Ph.D. in Automatic Control and Robotics, in 1990. He works as adjunct both at Research and Academic Computer Network (NASK) and at the Faculty of Electronics and Information Technology (at the Institute of Control and Computation Engineering) of Warsaw University of Technology. His research interests concentrate on data networks management, optimal control in risk conditions, decomposition and parallel implementation of numerical algorithms.
E-mail: A.Karbowski@ia.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Research and Academic Computer Network (NASK)
Wąwozowa st 18
02-796 Warsaw, Poland