Two approaches to dynamic power management in energy-aware computer networks - methodological considerations

Andrzej Karbowski

NASK, Research and Academic Computer Network
ul. Wąwozowa 18
02-796 Warszawa, Poland
and
Institute of Control and Computation Engineering
Warsaw University of Technology
ul. Nowowiejska 15/19
00-665 Warszawa, Poland
E-mail: A.Karbowski@elka.pw.edu.pl

Przemysław Jaskóła

NASK, Research and Academic Computer Network
ul. Wąwozowa 18
02-796 Warszawa, Poland
E-mail: pjaskola@nask.pl

Abstract—The paper compares two formulations of dynamic power management in energy-aware computer networks. In the first approach the only criterion is energy consumption, in the second there is an additional one - the quality of service. It is shown, that the second approach is appropriate when the routing problem with fixed demands is inadmissible. Fortunately, by some optimization modeling transformations it still allows for using the same standard mixed integer solvers as the first approach.

I. INTRODUCTION

METHODS for increasing energy efficiency of computer networks gained much attention last years. The reason is, that we are witnessing a rise of energy costs, customer increase, more on-demand services using cloud architectures, mobile Internet, a diffusion of broadband access and a growing number of services offered by Internet service providers. The growth of the energy consumption by network infrastructure may be well illustrated by the overall energy requirements of European Internet operators: in 2005 they needed 14 TWh, in 2010 - 21 TWh, and the forecast for 2020 is 36 TWh [1].

At the same time the capacity surplus becomes a standard in almost all networks. Consequently, so-called green network technologies are quickly becoming a high-priority issue for the Internet [1], [2].

In the European Union an additional motivation is 2020 Energy Strategy, which, among other goals, assumes achieving by 2020 a 20% improvement in energy efficiency [3].

Efforts to reduce power consumption in telecommunication networks follow in two mutually related directions – design of a more efficient equipment and development of energy-aware network control strategies and protocols. Initial efforts were aimed at assessment of energy characteristics of network equipment and building elementary models [4]. Upon this knowledge some local, i.e., concerning single device, strategies were built – see e.g., [5].

However, it is possible to save even more energy by employing network-wide solutions.

The paper [6] presented a model of energy-aware router, an architecture of a control framework and various formulations of a network-wide energy saving optimization problem. They start from the exact mixed integer linear programming (MILP) formulation, which is aimed at solving the problem of a minimum energy routing. The objective is the minimization of the total power utilized by network components while ensuring end-to-end Quality of Service (QoS). The basic link-node formulation (*LNb*) is a network management problem with binary decision variables describing full routing in a network and corresponding energy state assignments to all routers, line cards and communication ports. A more advanced version may be proposed, in which some parts of the network can be shifted to low energy mode as a result of the optimization algorithms, where both paths and flow rates are decision variables. It will exploit the fact, that Internet traffic used to be elastic in a large part, which means, that a quality of service is little aggravated by small deviations from assumed flow rate.

In this paper first a model with fixed flow rates inspired by the paper [6] will be shortly presented and assessed. Then, an improved version of it with flexible rates - in the authors' opinion much more practical - will be proposed and discussed.

II. A MODEL WITH GIVEN DEMANDS

A hierarchical network model proposed in [6] considers every single communication port $p \in \{1, \ldots, P\}$ of every line card of a router $r \in \{1, \ldots, R\}$. In our paper, for simplification, we do not consider individual cards of the router, because they do not bring anything into the model except additional summations.

Directed links connecting pairs of ports are denoted by $l \in \{1, \ldots, L\}$; any network component can operate in

$k \in \{1, \ldots, K\}$ energy states, but two ports connected by a link are in the same state. A demand $d \in \{1, \ldots, D\}$ is characterized by its source s_d and the destination t_d node (router) and the volume V_d.

The topology of the physical network is described by four matrices of binary indicators: g_{pr}, a_{lp}, b_{lp}, which indicate, whether, respectively: port p belongs to the router r, link l is incoming to the port p and link l is outgoing from the port p. If l is a link outgoing from the port p, the link $\tilde{l}$ denotes its partner link in the edge going in the opposite direction, that is:

$$b_{lp} = 1 \iff a_{\tilde{l}p} = 1 \qquad (1)$$

We assume that the numbers l, $\tilde{l}$ have different parities (e.g., they are consecutive numbers). The decision variables are two vectors of binary indicators x_p, z_r - whether the port p or router r is used for data transmission and two incidence matrices with elements: y_{lk} - whether the link l is in the state k and u_{dl} - whether the demand d uses the link l.

The full optimization problem is as follows:

$$\min_{\substack{x_p, y_{lk}, z_r, u_{dl} \\ p \in \overline{1,P}, l \in \overline{1,L}, \\ k \in \overline{1,K}, r \in \overline{1,R}, \\ d \in \overline{1,D}}} \left[F_{LNb} = \sum_{l=1,3,5,\ldots}^{L-1} \sum_{k=1}^{K} \xi_{lk} y_{lk} + \right.$$

$$\left. + \sum_{p=1}^{P} W_p x_p + \sum_{r=1}^{R} T_r z_r \right] \qquad (2)$$

subject to the constraints:

$$\forall_{\substack{d=1,\ldots,D, \\ p=1,\ldots,P}} \quad \sum_{l=1}^{L} a_{lp} u_{dl} \leq x_p \qquad (3)$$

$$\forall_{\substack{d=1,\ldots,D, \\ p=1,\ldots,P}} \quad \sum_{l=1}^{L} b_{lp} u_{dl} \leq x_p \qquad (4)$$

$$\forall_{\substack{r=1,\ldots,R, \\ p=1,\ldots,P}} \quad g_{pr} x_p \leq z_r \qquad (5)$$

$$\forall_{l=1,\ldots,L} \quad \sum_{k=1}^{K} y_{lk} \leq 1 \qquad (6)$$

$$\forall_{\substack{d=1,\ldots,D, \\ r=1,\ldots,R}} \sum_{p=1}^{P} \sum_{l=1}^{L} g_{pr} a_{lp} u_{dl} - \sum_{p=1}^{P} \sum_{l=1}^{L} g_{pr} b_{lp} u_{dl} =$$

$$= \begin{cases} -1 & r = s_d \\ 1 & r = t_d \\ 0 & \text{otherwise} \end{cases} \qquad (7)$$

$$\sum_{d=1}^{D} V_d u_{dl} \leq \sum_{k=1}^{K} M_{lk} y_{lk}, \ l = 1, 3, \ldots, L-1 \qquad (8)$$

$$\sum_{d=1}^{D} V_d u_{dl} \leq \sum_{k=1}^{K} M_{\tilde{l}k} y_{\tilde{l}k}, \ l = 2, 4, \ldots, L \qquad (9)$$

$$x_p, z_r \in \{0,1\} \ p \in \overline{1,P}, r \in \overline{1,R}, \qquad (10)$$

$$y_{lk}, u_{dl} \in \{0,1\} \ l \in \overline{1,L}, k \in \overline{1,K}, d \in \overline{1,D} \qquad (11)$$

where M_{lk} and ξ_{lk} are, respectively, the capacity and the power consumption of the edge comprising links $l, \tilde{l}$ in the state k, and W_p and T_r are power cost coefficients of the port p and the router r.

As in the paper [6] constraints (3)-(5) determine the number of ports and routers that are used for data transmission. The conditions (6) assure, that each edge can be in one energy-aware state. The constraints (7) are formulated according to 1st Kirchhoff's law applied to source, destination and transit nodes, and finally, the constraints (8),(9) assure, that the flow will not exceed the capacity of a given edge.

Any MILP algorithm can be used to solve (2)-(11) problem. Popular solvers such as CPLEX and Gurobi may be applied.

III. SHORTCOMINGS OF THE *LNb* MODEL AND POSSIBILITIES TO OVERCOME THEM

In the *LNb* model (2)-(11) the issue of determining the flow demand matrix, which is crucial for implementing the control system, is not addressed. The simplest approach is to measure the actual flow rates carried by the network (averaged over a selected timescale), and adopt it as optimization constraints. Then, this model can help to answer the question whether it is possible and worthwhile to temporarily reduce network capacity and maybe also reroute some paths in order to save some energy without sacrificing the QoS, that is maintaining the current values of the flow rates. It should work especially in the case, when the network is underutilized.

Shortcomings of the method may manifest themselves however in more complex scenarios, when the controlled network, or at least a part of it, works on its capacity limit. The measured traffic is no longer a good estimate of the demanded bandwidth, because flows are already truncated by the network. The demand matrix must be provided in some other way, e.g., by predictive model or educated guess.

It is then crucial, that demands, coupled through link capacity constraints, do not exceed the bandwidth offered by the network in any spot, which can be hard to assure in nontrivial network topologies. Otherwise, a feasible solution will not exist.

In such cases, a modified approach can be proposed, in which flow rates are represented by variables rather than constants. It exploits the fact, that Internet traffic used to be elastic in a large part, which means, that the QoS is only a little aggravated by small deviations from the assumed flow rate. The combined routing and rate control problem has to be solved, which leads to the solution feasible in terms of the formulated model, even when the traffic demand is greater than the capacity offered by the network.

Moreover, in some cases a minor reduction of flow rates, which is accepted by the comprehensive model taking into account the elasticity of a demand, may allow to accommodate

the traffic in a smaller number of links, thus allowing for further great reduction of power consumption.

IV. A GENERALIZATION – TWO-CRITERIA ROUTING PROBLEM

The most important modification of the *LNb* model consists in relaxing flow rates - from now on they are variables denoted by v_d. In a consequence, the objective function F_{LNb} (2) has to be augmented with a QoS related criterion Q_d, which represents a penalty for not achieving the assumed flow rate V_d by the flow d. $Q_d(v_d)$ is a convex and continuous function, decreasing on interval $[0, V_d]$. It is reaching minimum (zero) at V_d, the point in which user expectations are fully satisfied. The convexity of $Q_d(v_d)$ is associated with the conviction, that small deviations from the nominal throughput $\Delta = V_d - v_d$ are neglected by network users, while large deviations are noticed and should be avoided. Moreover, since $Q_d(v_d)$ is monotonically decreasing, it assures that the slope of the curve becomes steeper, as the rate v_d approaches zero.

A two criteria - i.e., reflecting energy costs and QoS - mixed integer network problem of simultaneous optimal bandwidth allocation and routing may be formulated in the following way:

$$\min_{\substack{x_p, y_{lk}, z_r, u_{dl}, v_d, \\ p \in \overline{1,P}, l \in \overline{1,L}, k \in \overline{1,K} \\ r \in \overline{1,R}, d \in \overline{1,D}}} \left\{ F_{2C} = \alpha F_{LNb} + (1-\alpha) \sum_{d=1}^{D} Q_d(v_d) = \right.$$

$$= \alpha \left[\sum_{l=1,3,5,\ldots}^{L-1} \sum_{k=1}^{K} \xi_{lk} y_{lk} + \sum_{p=1}^{P} W_p x_p + \sum_{r=1}^{R} T_r z_r \right] +$$

$$\left. + (1-\alpha) \sum_{d=1}^{D} Q_d(v_d) \right\} \qquad (12)$$

subject to constraints (3)-(7), (10)-(11) from *LNb* and

$$\sum_{d=1}^{D} v_d u_{dl} \leq \sum_{k=1}^{K} M_{lk} y_{lk}, \ l = 1, 3, \ldots, L-1 \qquad (13)$$

$$\sum_{d=1}^{D} v_d u_{dl} \leq \sum_{k=1}^{K} M_{\tilde{l}k} y_{\tilde{l}k}, \ l = 2, 4, \ldots, L \qquad (14)$$

$$0 \leq v_d \leq V_d, \ d \in \overline{1,D} \qquad (15)$$

The parameter $\alpha \in [0, 1]$ is a scalarizing weight coefficient, which can be altered to emphasize any of the objectives.

Such a problem in the particular case when $\alpha = 0$, i.e., without the energy components, was first addressed by Jaskóła and Malinowski [7] and independently by Wang et al. [8]. A decomposed algorithm to solve it was proposed by Karbowski [9].

A two-criteria optimal routing and bandwidth allocation problem, taking into account the energy component, for a completely different network and cost model than that of Section II was presented in [10].

In general, the formulation (12)-(15), (3)-(7), (10)-(11) has some drawbacks: it defines a mixed-integer nonlinear programming problem with nonconvex, bilinear link capacity constraints (13),(14). At present the leading solvers - e.g., CPLEX, Gurobi - can solve efficiently convex quadratic mixed-integer quadratically constrained problems MIQCP, with positive semidefinite matrices of constraints quadratic forms, which is not the case of (13),(14) constraints. The general nonlinear, mixed-integer, nonconvex solvers are very slow.

Fortunately, the problem (12)-(15), (3)-(7), (10)-(11) can be quite easily transformed to the form accepted by fast mixed-integer solvers, what we describe below. The reformulation of nonlinear network problems to the input format of standard MILP or MIQP solvers seems to be a very promising approach nowadays [11].

V. ELIMINATION OF THE NONLINEARITY FROM CONSTRAINTS

From the QoS components of the objective function F_{2C} (12) it is usually expected, that they assure so-called proportional-fairness of the allocations of the bandwidth, when the network is subject to a congestion [12]. Quadratic functions may be used to achieve it [13] (unfortunately, linear - not), so the objective function F_{2C} can be quadratic and convex.

The only problem that still remains to solve is nonconvex nonlinearity of the constraints (13), (14). It can be eliminated by a transformation proposed in [14].

It consists in the introduction of auxiliary variables $w_{dl} = v_d u_{dl}$, $d \in \overline{1,D}$, $l \in \overline{1,L}$ (denoting the part of a traffic rate in the link l assigned to the flow d) and the substitution of these inequalities with subsequent set of linear inequalities:

$$\forall_{l=1,3,\ldots,L-1} \quad \sum_{d=1}^{D} w_{dl} \leq \sum_{k=1}^{K} M_{lk} y_{lk} \qquad (16)$$

$$\forall_{l=2,4,\ldots,L} \quad \sum_{d=1}^{D} w_{dl} \leq \sum_{k=1}^{K} M_{\tilde{l}k} y_{\tilde{l}k} \qquad (17)$$

$$\forall_{\substack{d=1,\ldots,D \\ l=1,\ldots,L}} \quad w_{dl} \leq V_d u_{dl} \qquad (18)$$

$$\forall_{\substack{d=1,\ldots,D \\ l=1,\ldots,L}} \quad w_{dl} \leq v_d \qquad (19)$$

$$\forall_{\substack{d=1,\ldots,D \\ l=1,\ldots,L}} \quad w_{dl} \geq v_d - V_d(1 - u_{dl}) \qquad (20)$$

$$\forall_{\substack{d=1,\ldots,D \\ l=1,\ldots,L}} \quad w_{dl} \geq 0 \qquad (21)$$

VI. THE FINAL FORMULATION OF THE PROBLEM

Summing up, the final formulation of our two criteria energy-aware integrated routing and flow control problem is as follows:

$$\min_{\substack{x_p, y_{lk}, z_r, u_{dl}, v_d, w_{dl} \\ p \in \overline{1,P}, l \in \overline{1,L}, k \in \overline{1,K} \\ r \in \overline{1,R}, d \in \overline{1,D}}} \left\{ F_{2C} = \alpha F_{LNb} + (1-\alpha) \sum_{d=1}^{D} Q_d(v_d) = \right.$$

$$= \alpha \left[\sum_{l=1,3,5,\dots}^{L-1} \sum_{k=1}^{K} \xi_{lk} y_{lk} + \sum_{p=1}^{P} W_p x_p + \sum_{r=1}^{R} T_r z_r \right] +$$

$$+ (1-\alpha) \sum_{d=1}^{D} Q_d(v_d) \Bigg\} \tag{22}$$

subject to the constraints:

$$\forall_{\substack{d=1,\dots,D, \\ p=1,\dots,P}} \quad \sum_{l=1}^{L} a_{lp} u_{dl} \le x_p \tag{23}$$

$$\forall_{\substack{d=1,\dots,D, \\ p=1,\dots,P}} \quad \sum_{l=1}^{L} b_{lp} u_{dl} \le x_p \tag{24}$$

$$\forall_{\substack{r=1,\dots,R, \\ p=1,\dots,P}} \quad g_{pr} x_p \le z_r, \tag{25}$$

$$\forall_{l=1,\dots,L} \quad \sum_{k=1}^{K} y_{lk} \le 1 \tag{26}$$

$$\forall_{\substack{d=1,\dots,D, \\ r=1,\dots,R}} \sum_{p=1}^{P} g_{pr} \sum_{l=1}^{L} (a_{lp} - b_{lp}) u_{dl} = \begin{cases} -1 & r = s_d \\ 1 & r = t_d \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

$$\sum_{d=1}^{D} w_{dl} \le \sum_{k=1}^{K} M_{lk} y_{lk}, \ l = 1, 3, \dots, L-1 \tag{28}$$

$$\sum_{d=1}^{D} w_{dl} \le \sum_{k=1}^{K} M_{\tilde{l}k} y_{\tilde{l}k}, \ l = 2, 4, \dots, L \tag{29}$$

$$\forall_{\substack{d=1,\dots,D, \\ l=1,\dots,L}} \quad w_{dl} \le V_d u_{dl} \tag{30}$$

$$\forall_{d=1,\dots,D,} \quad w_{dl} \le v_d \tag{31}$$

$$\forall_{\substack{d=1,\dots,D, \\ l=1,\dots,L}} \quad w_{dl} \ge v_d - V_d(1 - u_{dl}) \tag{32}$$

$$\forall_{\substack{d=1,\dots,D, \\ l=1,\dots,L}} \quad w_{dl} \ge 0 \tag{33}$$

$$\forall_{d=1,\dots,D} \quad 0 \le v_d \le V_d \tag{34}$$

$$x_p, z_r \in \{0,1\} \ p \in \overline{1,P}, r \in \overline{1,R}, \tag{35}$$

$$y_{lk}, u_{dl} \in \{0,1\} \ l \in \overline{1,L}, k \in \overline{1,K}, d \in \overline{1,D} \tag{36}$$

When QoS components $Q_d(v_d)$ are quadratic and convex, the obtained mixed-integer quadratic problem can be solved by effective MILP/MIQP solvers, such as CPLEX, Gurobi.
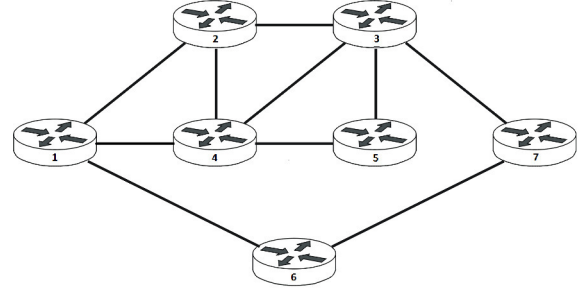

Fig. 1. Test network

VII. NUMERICAL EVALUATION

The problem (22)- (36) was formulated, implemented and solved with the help of the CPLEX solver for the test network presented in Fig. 1, with: $R = 7$, $L = 20$, $P = 20$. Every edge between two routers had two unidirectional links. We performed the experiments for three demands $(D = 3) : 1 \to 7, 1 \to 7, 5 \to 6$. It was possible for each link to operate in five energy-aware states $(K = 5)$. As their (penalty for not achieving) QoS functions we took [7]:

$$Q_d(v_d) = \frac{1}{2} V_d^2 - v_d \cdot (V_d - v_d/2) \tag{37}$$

The throughput of a given link $l \in \overline{1, L}$ and the power consumption in energy-aware state $k \in \overline{1, K}$ were as follows: $(M_{l1} = 20, \xi_{l1} = 15)$, $(M_{l2} = 40, \xi_{l2} = 30)$, $(M_{l3} = 60, \xi_{l3} = 40)$, $(M_{l4} = 80, \xi_{l4} = 60)$, $(M_{l5} = 100, \xi_{l5} = 75)$. As the power cost coefficients we took for ports $W_1 = W_2 = \dots = W_{20} = 30$ and for routers $T_1 = T_2 = \dots = T_7 = 300$. The maximum demand volumes were: $V_1 = 100, V_2 = 50, V_3 = 150$. The scalarizing coefficient was taken $\alpha = \frac{1}{2}$.

From calculations we got the following optimal paths: $d = 1, 2 : 1 \to 6 \to 7$; $d = 3 : 5 \to 3 \to 7 \to 6$ and the objective function $F_{2C} = 1957.499221$. It means, that only routers $1, 3, 5, 6, 7$ were used and the links and ports connecting them, that is, the rest of the network was in the sleeping mode. The obtained optimal transmission rates were: $\hat{v}_1 = 75, \hat{v}_2 = 25, \hat{v}_3 = 100$. All the used links worked at the highest energy levels.

VIII. CONCLUSIONS

We modified the dynamic power management of energy-aware computer networks model presented in [6] to capture these situations, when the users' demands are so high, that there is no admissible solution of the problem *LNb* (2)-(11). We suggest to use then a two-criteria model with flow rates as additional decision variables. When the QoS functions are quadratic and convex, it is possible to reformulate the problem in such a way, that the same standard solvers, e.g., CPLEX or Gurobi, can be used to find the solution. The resulting mixed-integer programming problem has more variables, but the new are only real, not binary, what should not influence too much the time of calculations.

The performed numerical test confirmed the appropriateness of the formulation.

REFERENCES

[1] D.G. Recupero, "Toward a Green Internet", *Science*, vol. 339, 2013, pp. 1533–1534, http://dx.doi.org/10.1126/science.1235623

[2] R. Bolla, R. Bruschi, F. Cucchietti, and F. Davoli, "Setting the Course for a Green Internet", *Science*, vol. 342, 2013, pp. 1316, http://dx.doi.org/10.1126/science.342.6164.1316-a

[3] European Commission "Energy 2020 A Strategy For Competitive, Sustainable And Secure Energy", *Communication From The Commission To The European Parliament, The Council, The European Economic and Social Committee and The Committee Of The Regions*, COM(2010) 639, Brussels, 2010, http://dx.doi.org/10.2833/78930

[4] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, "Energy efficiency in the Future Internet: A survey of existing approaches and trends in energy-aware fixed network infrastructures", *IEEE Communications Surveys & Tutorials*, vol. 13, 2011, pp. 223–244, http://dx.doi.org/10.1109/SURV.2011.071410.00073

[5] S. Nedevschi, L. Popa, G. Iannacone, D. Wetherall, S. Ratnasamy, "Reducing network energy consumption via sleeping and rate adaptation", in *Proc. of 5th USENIX Symposium on Networked Systems Design and Implementation*, 2008, pp. 323–336.

[6] E. Niewiadomska-Szynkiewicz, A. Sikora, P. Arabas, M. Kamola, K. Malinowski, P. Jaskóła, and M. Marks, "Network-wide power management in computer networks", in *Proc. 22nd International Teletraffic Congress Specialist Seminar on Energy Efficient and Green Networking* (Best Paper Award), 2013, http://dx.doi.org/10.1109/SSEEGN.2013.6705398

[7] P. Jaskóła, and K. Malinowski, "Two methods of optimal bandwidth allocation in TCP/IP networks with QoS differentiation", in *Proc. Summer Simulation Multiconference (SPECTS'04)*, 2004, pp. 373–378.

[8] J. Wang, L. Li, S. H. Low, J. C. Doyle, "Cross-layer optimization in TCP/IP networks", *IEEE/ACM Transactions on Networking*, vol. 13, 2005, pp. 582–595, http://dx.doi.org/10.1109/TNET.2005.850219

[9] A. Karbowski, "Integrated routing and network flow control embracing two layers of TCP/IP networks - methodological issues", *Journal of Telecommunications and Information Technology*, 2012, pp. 51–54, http://www.itl.waw.pl/czasopisma/JTIT/2012/2/51.pdf

[10] P. Jaskóła, P. Arabas and and A. Karbowski, "Combined Calculation of Optimal Routing and Bandwidth Allocation in Energy Aware Networks", in *Proceedings of the 26th International Teletraffic Congress (ITC)*, 2014, http://dx.doi.org/10.1109/ITC.2014.6932962

[11] M. Tvorogova, "Efficient models for special types of non-linear maximum flow problems", in *Proc. Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2013, pp. 409–416.

[12] F. Kelly, "Charging and rate control for elastic traffic", *European Transactions on Telecommunications*, 1997, vol. 8, pp. 33–37, http://dx.doi.org/10.1002/ett.4460080106

[13] C. Touati, E. Altman, J. Galtier, "On fairness in bandwidth allocation", *Tech. Rep. 4269, Unité de recherche INRIA, Sophia Antipolis, France*, 2001.

[14] J. Bisschop, *AIMMS Optimization Modeling*, Paragon Decision Technology B.V., 2007.