On Cross-layer Design of Wireless Mesh Networks Using Network Coding

Nikolay Petrov[1], Vladimir Poulkov[2] and Georgi Iliev[3]

Abstract – **This paper presents an overview of network coding and its application in cross-layer design. It includes a simple network coding example. A reasoning is given for choosing the cross-layer design approach over a layered design approach. An example of cross-layer design scenario, based on primal-dual interior point method, is briefly reviewed in a simplified and comprehensible manner, using a basic data model representation.**

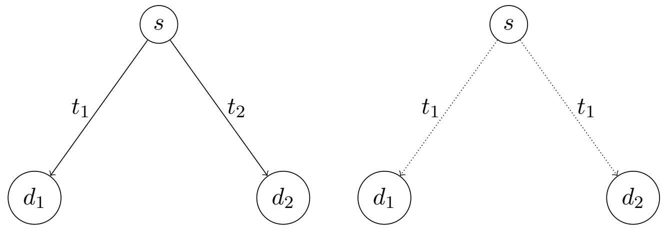
Keywords – **Network coding, Cross-layer design, Wireless mesh networks.**

I. INTRODUCTION

Nowadays, in the majority of existing computer networks, each node functions as a switch in a sense that it either forwards or replicates information from an input link to an output link or to a certain set of output links. Nevertheless, the specifics of the information flow in today's networks and the emergence of new services is a good reason to reconsider and reconstruct the idea of how the network should work.

In 2000, a paper by Ahlswede, Cai, Li and Yeung [1] introduced the idea of network coding. It suggests that the intermediate nodes in a network are allowed not only to route but also to combine incoming data from different nodes with coding operations. The complexity of coding is an important issue, related to additional processing time at some of the intermediate nodes, which are not interested in the transmitted information, in general. Consequently, in practice, the simplest and preferred method for such coding is to apply XOR operations to the input packets at the node that performs the network coding, and to output the result to the destination adjacent nodes.

Network coding has been proven to be effective for multicast applications [2, 3], and according to recent studies [4–6] in many scenarios of multiple unicast applications when such coding is employed in a multi-hop wireless environment. Leading direction of the research on network coding is related to mobile networks. One of the main reasons for that is the broadcasting feature of the wireless channel, which plays an important role in network coding over multiple unicasts. As shown in Fig. 1a in the case of wireline communication the transmission is limited to one destination d_i (one port) at a time, and requires number of channels or time slots (t_1, t_2) equivalent to the number of data sessions ($s \rightarrow d_1$, $s \rightarrow d_2$). On the contrary, the wireless channel Fig. 1b allows concurrent, at the same time t_1, transmission of packets to multiple destinations or sinks, situated in the coverage of the transmit antenna.



(a) Wireline communication (b) Wireless communication

Fig. 1. Advantage of broadcasting in aggregate capacity gain.

II. AN EXAMPLE OF NETWORK CODING

Consider the network topology presented in Fig. 2a. The indices of s and d relate to three independent unicast sessions, and b_1, b_2 and b_3 correspond to the information in each of these sessions. Every node is equipped with one radio; therefore, it can either transmit or receive information at the same time. Using a traditional approach as shown in Fig. 2b it would take 6 time slots in order to transmit the data of all of the three sessions to its destinations.

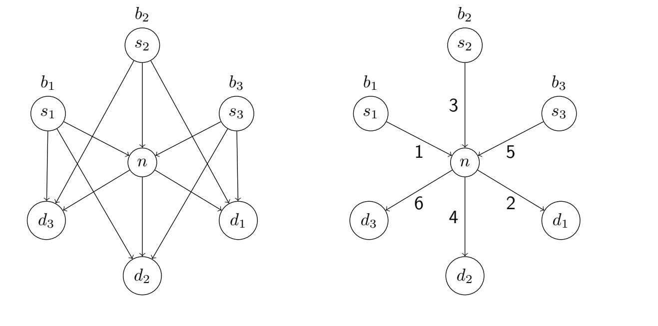
As known, in a network with one source and one sink the maximum information flow is limited by the weakest set of links which cut the source form the sink completely. In this case node n acts as a bottleneck and has to be used three times. Fig. 3 presents the case when network coding is enabled. In this example the network coding is actually the utilization of node n to perform XOR operations and later on decoding the information at the destination nodes by XORing the received data bits.

In details, the communication goes as follows. In the first step (Fig. 3a) the message b_1 sent from node s_1 reaches nodes d_3, d_2, and n. In the second step (Fig. 3b) s_2 sends the message b_2 to nodes d_3, n and d_2. In the third step (Fig. 3c) the nodes n, d_2 and d_1 receive b_3 from the source node s_3. After node n has received b_1, b_2 and b_3 it performs the network coding and broadcasts the message ($b_1 \oplus b_2 \oplus b_3$) to d_3, d_2 and d_1 in the fourth step (Fig. 3d). Now by applying XOR operation to the three messages received at each of the destination nodes the original message intended for the particular node is retrieved. The achieved network throughput gain due to the network coding in this case is 1.5.

[1]Nikolay Petrov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: petrovn@ieee.org.

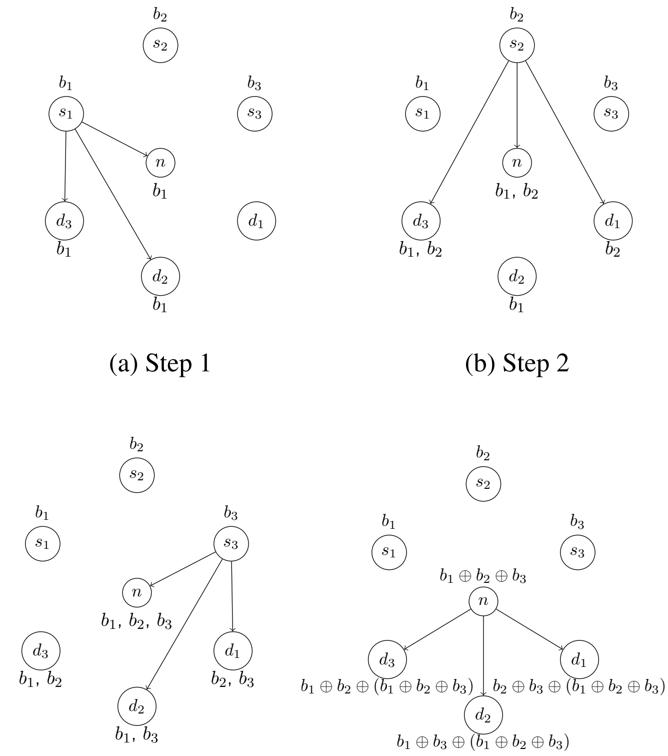
[2]Vladimir Poulkov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: vkp@tu-sofia.bg.

[3]Georgi Iliev is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: gli@tu-sofia.bg.



(a) Network topology (b) Sequence of channel occupation

Fig. 2. A three-source/three-destination unicast wireless network and the sequence of channel occupation in the case when network coding is disabled



(a) Step 1 (b) Step 2

(c) Step 3 (d) Step 4

Fig. 3. Example of network coding technique utilized in the network given in Fig. 2 (a).

III. Motivation for Choosing Cross-Layer Design Over Layered Design

The widely used layered architecture models as the OSI model and the Internet Protocol Suite (TCP/IP) facilitate handling the complexity of large communication networks. Each layer fulfills a limited, well-defined purpose and delivers a digest and simpler model of the network to its upper layer. This allows splitting the communication network design into several smaller design problems which are easier to solve.

Nevertheless, with the emergence of wireless networks and other new networking technologies in the past decade, environments and circumstances have changed. The characteristics of the wireless networks are quite different from wireline systems. System developers and researchers face different problems and challenges compared to the wireline networks. In this altered situation, a layered design approach can be in general suboptimal to a cross-layer design approach, where several layers are designed jointly.

In a cross-layer design scheme several different aspects of the layered model can be taken into account. Some of these can be congestion control, energy-conservancy, optimal channel use (according to achievable rates for transmission over the wireless medium), resource allocation, interference management, transmission scheduling and media access schemes, delay constrains. Such an algorithm is presented in [7].

IV. Problem Formulation and Data Model

The algorithm presented in [7] is considered. In this method, a pseudobroadcasting technique is taken into account. It complicates the data representation in the means of adding new virtual nodes and corresponding to them links. For the purpose of simplicity and in order to reproduce clearer the essence of the algorithm, pseudobroadcasting will not be explained in this paper.

The network is modeled as a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$. Where $\mathcal{N} = \{1, 2, ..., N\}$ represents the nodes in the network and $\mathcal{L} = \{1, 2, ..., L\}$ the links in the network. $\boldsymbol{I} = \{I_1, I_2, ..., I_N\}$ is defined as the radio distribution vector, where I_i is number of radios at the i^{th} node. The network topology can be represented by an N x L indicator matrix $\boldsymbol{D}$, whose entries $D(n, l)$ satisfy $D(n, l) = 1$ if the n^{th} node is the start node of the l^{th} link, $D(n, l) = -1$, if the n^{th} node is the end node of the l^{th} link, $D(n, l) = 0$, otherwise. There are T independent data sessions in the network and for each one R_t defines the data rate. They form a unicast data rates vector for the T sessions: $\boldsymbol{R} = \{R_1, R_2, ..., R_T\}$. s_t and d_t will denote respectively the source node and the destination node of the t^{th} network session. The aggregate flow rate over the l^{th} link, where $(1 \leq l \leq L)$ is denoted with f_l. The data flows are assumed to be lossless across the links and the traffic flow can be split arbitrarily at nodes as long as the flow conservation law is satisfied. $\mathcal{F}(\boldsymbol{R})$ forms the set of all possible network flow vectors that support $\boldsymbol{R}$. $\mathcal{C}(\boldsymbol{I})$ represents the region containing all achievable rate vectors $\boldsymbol{c} = \{c_1, c_2, ..., c_L\}$. By time-sharing over a large number of time slots of the available links [8], the link capacity set $\mathcal{C}$ can be represented as a convex hull. Therefore, $\mathcal{C}$ is fully determined by its vertices and it can be defined using another, approximate to $\mathcal{C}$, convex hull $\mathcal{C}' \subseteq \mathcal{C}$. Depending on different scenarios for the power used by the transmitter, and the interference, and noise at the receiver, a finite number of transmission rates may be defined. This leads to a finite number K of feasible link-rate vectors $\mathcal{V}' = \{\boldsymbol{c}_k \in \mathcal{C}, k = 1, 2, ..., K\}$. Now, using the set $\mathcal{V}'$, the corresponding convex hull

$$\mathcal{C}' = \left\{ \boldsymbol{c} | \boldsymbol{c} = \sum_{k=1}^{K} \alpha_k \boldsymbol{c}_k, \text{ s.t.} \sum_{k=1}^{K} \alpha_k = 1, \alpha_k \geq 0, \forall k \right\} \quad (1)$$

is fully defined. Behind the simple definition of the sets $\mathcal{F}(\boldsymbol{R})$ and $\mathcal{C}(I)$ given here underlies a detailed mathematical interpretation. It is discussed (in details) in [7, 8] and shows how resource allocation, data scheduling, routing, and network coding schemes characterize the two sets $\mathcal{F}(\boldsymbol{R})$ and $\mathcal{C}(\boldsymbol{I})$. Given these notations the optimization problem is defined as [7]:

$$\begin{aligned} &\text{maximize } U(\boldsymbol{R}, \boldsymbol{f}) \\ &\text{subject to: } \boldsymbol{f} \in \mathcal{F}(\boldsymbol{R}); \\ &\boldsymbol{R} \succeq 0; \\ &\boldsymbol{c} \in \mathcal{C}(\boldsymbol{I}); \\ &f_l < c_l, \forall l. \end{aligned} \quad (2)$$

The constraints in Eq. (2) have the meaning as follows. The first constraint enforces the dependence between the achievable rates $\boldsymbol{R}$ and the data flows $\boldsymbol{f}$. Rates R_t should be non-negative, as stated in the second constraint. The third constraint indicates the relationship between the achievable link capacity $\boldsymbol{c}$ and radio allocation along with the resource allocation, scheduling, and data routing schemes. Finally, the fourth constraint states that the sum of the flow rate on each link is bounded by the link capacity.

The choice for the utility function $U(\boldsymbol{R}, \boldsymbol{f})$ may vary depending on the design target. It is practical to aim maximization of the throughput ($U(\boldsymbol{R}, \boldsymbol{f}) = \sum_{t=1}^{T} R_t$, i.e. the sum of the data rates in all the sessions) or to pursue fairness, maximizing the minimum end-to-end communication rates ($U(\boldsymbol{R}, \boldsymbol{f}) = \min\{R_t\}$). The utility function is assumed to be concave.

The surveyed algorithm involves joint consideration of the physical-layer, wireless Media Access Control (MAC), and network-layer planning. In the physical layer the achievable rates for transmission over the wireless medium is taken into account, considering issues such as resource allocation, interference management. The wireless MAC layer includes transmission scheduling and media access schemes. The network layer accounts the actual aggregate load over a particular link. Also, in order to prevent adverse effect on the upper layer performance, the rate flow region $\mathcal{F}(\boldsymbol{R})$ includes delay constrains in its formulation.

The discussed algorithm is based on a primal-dual interior point method. This approach offers very good control of the distance to optimality. The key idea here is to start from a rather loose tolerance ϵ and to build as fast as possible the first rough approximation to the original problem. Subsequently, the optimality tolerance required for the solution of the restricted master problem is tightened until a predefined threshold ϵ_0 is achieved or a maximum iteration number is reached. Depending on the choice of ϵ_0, the algorithm may give fast some locally optimal solution for the original problem Eq. (2) or eventually, taking longer, to reach the global optimum.

In this approach the problem described above is relaxed to the following problem:

$$\begin{aligned} &\text{maximize } U(\boldsymbol{R}, \boldsymbol{f}) \\ &\text{subject to: } \boldsymbol{f} \in \mathcal{F}(\boldsymbol{R}); \\ &\boldsymbol{R} \succeq 0; \\ &\boldsymbol{c} = \sum_{k=1}^{K} \alpha_k c_{\boldsymbol{k}}, \sum_{k=1}^{K} \alpha_k = 1; \\ &f_l < c_l, \forall l. \end{aligned} \quad (3)$$

This problem Eq. (3) is referred as the restricted primal problem. The solution of this problem provides a lower bound U_{lower} for the original problem. On the other hand the dual of the original problem provides an upper bound U_{upper} of Eq. (2). In [7] it is defined as:

$$\max_{\boldsymbol{R} \succeq 0, \boldsymbol{f} \in \mathcal{F}(\boldsymbol{R})} \left\{ U(\boldsymbol{R}, \boldsymbol{f}) - \sum_{l=1}^{L} \lambda_l f_l \right\} + \max_{\boldsymbol{c} \in \mathcal{C}(\boldsymbol{I})} \sum_{l=1}^{L} \lambda_l c_l, \quad (4)$$

$$\text{subject to: } \lambda_l \geq 0, l = 1, 2, ..., L$$

Given the solution of these two formulations the gap between the two bounds can be obtained. This gap, as an indication of the accuracy of the current optimization result, is compared to ϵ. From this result, using an iterative algorithm with column generation [9], the distance to optimality can be controlled.

V. RESULTS

Consider the network topology shown in Fig. 2a. It has the following parameters: $N = 7$, $L = 12$, $T = 3$. The source nodes are denoted as s_1, s_2 and s_3. The destination nodes are d_1, d_2 and d_3. The model of the network is given by the graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where $\mathcal{N} = \{s_1, s_2, s_3, n, d_1, d_2, d_3\}$ and $\mathcal{L} = \{l_1, l_2, ..., l_{12}\}$. For this example the radio distribution vector will be defined as $\boldsymbol{I} = \{1, 1, 1, 4, 3, 3, 3\}$. This indicates that the number of radios at the source nodes remains 1, as in the previous example; the node n and the destination nodes are equipped with multiple radios.

The performance of the algorithm employed in this scenario can be seen from the results obtained in [7] on Fig. 4. A feasible solution of the problem Eq. (3), before the ϵ_0 threshold is reached, corresponds to a valid suboptimal network coding solution. In such a solution the system throughput has the value of the current utility function U_{lower}. Using the discussed iterative column generation interior-point method, the optimal primal solution of Eq. (2) gradually coincides with its dual solution. As it is seen from the figure that the defined tolerance here is $\epsilon_0 = 0$ and the final solution is actually the optimum solution for this particular network. The maximum achieved throughput is $U_{lower} = U_{upper} = 3$. In the case when network coding is disabled and only routing is allowed the optimized system throughput is 2.

VI. CONCLUSIONS

For the short period of its existence the network coding approach has captured the attention of a significant number of researchers. Different approaches have been proposed concerning

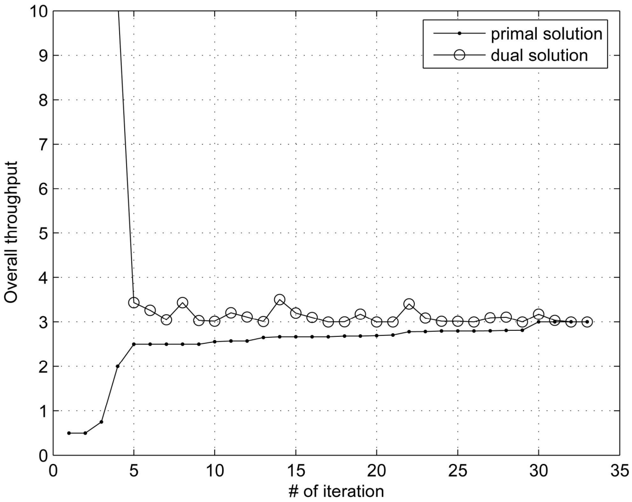


Fig. 4. Performance of the primal-dual algorithm [7] as a function of the increasing number of iterations for a three-source/three-destination network. The utility function of the optimization problem is the total system throughput.

different aspects of the network coding design. This paper has attempted to provide an overview of the cross-layer design for wireless mesh networks employing network coding to support multiple unicast applications. The reviewed approach considers the physical layer, MAC, network layer and sets up delay constraints. Generally, the process of optimization takes some time to converge to an optimal solution. However, the transmission can start before the convergence is done. This makes such an approach practical for application in wireless networks. Investigation in this direction is a future research topic.

ACKNOWLEDGEMENT

This work was supported by the Bulgarian National Science Fund Grant No. ДО-02-135/2008 "Research on Cross Layer Optimization of Telecommunication Resource Allocation".

REFERENCES

[1] R. Ahlswede, N. Cai, S.Y.R. Li, and R.W. Yeung, "Network Information Flow," *IEEE Trans. Information Theory*, vol. 46, no. 4, pp. 1204-1216, July 2000.

[2] R. Koetter and M. Médard, "An Algebraic Approach to Network Coding," *IEEE/ACM Trans. Networking*, vol. 11, no. 5, pp. 782-795, Oct. 2003.

[3] S.Y.R. Li, R.W. Yeung, and N. Cai, "Linear Network Coding," *IEEE Trans. Information Theory*, vol. 49, no. 2, pp. 371-381, Feb. 2003.

[4] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XORs in the Air: Practical Wireless Network Coding," *Proc. ACM SIGCOMM*, 2006.

[5] D. Traskov, N. Ratnakar, D.S. Lun, R. Koetter, and M. Médard, "Network Coding for Multiple Unicasts: An Approach Based on Linear Optimization," *Proc. IEEE Intl Symp. Information Theory (ISIT 06)*, July 2006.

[6] T. Ho, D. Lun, *Network Coding, An Introduction*, Cambridge University Press, 2008.

[7] K. Li, X. Wang, "Cross-Layer Design of Wireless Mesh Networks with Network Coding," *IEEE Trans. Mobile. Comp.*, vol. 7, no.11, pp.1363-1373, Nov. 2008.

[8] M. Johansson and X. Lin, "Cross-Layer Optimization of Wireless Networks Using Nonlinear Column Generation," *IEEE Trans. Wireless Comm.*, vol. 5, no. 2, pp. 435-445, Feb. 2006.

[9] J. Gondzio and R. Sarkissian, "Column Generation with a Primal-Dual Method," technical report, Logilab, HEC Geneva, Univ. of Geneva, Geneva, 1996.