QoSCM: QoS-aware Coded Multicast Approach for Wireless Networks

Amin Mohajer[1*], Morteza Barari[1] and Houman Zarrabi[2]
[1] Department of Information, Communications & Security, Malek Ashtar University of Technology, Tehran, Iran
[e-mail: mohajer@mut.ac.ir, barari@mut.ac.ir]
[2] Integrated Network Management Group, Iran Telecommunication Research Center (ITRC), Tehran, Iran
[e-mail: h.zarrabi@itrc.ac.ir]
*Corresponding author: Amin Mohajer

Received July 16, 2016; accepted September 21, 2016; published December 31, 2016

Abstract

It is essential to satisfy class-specific QoS constraints to provide broadband services for new generation networks. The present study proposes a QoS-driven multicast scheme for wireless networks in which the transmission rate and end-to-end delay are assumed to be bounded during a multiple multicast session. A distributed algorithm was used to identify a cost-efficient sub-graph between the source and destination which can satisfy QoS constraints of a multicast session. The model was then modified as to be applied for wireless networks in which satisfying interference constraints is the main challenge. A discrete power control scheme was also applied for the QoS-aware multicast model to accommodate the effect of transmission power level based on link capacity requirements. We also proposed random power allocation (RPA) and gradient power allocation (GPA) algorithms to efficient resource distribution each of which has different time complexity and optimality levels. Experimental results confirm that the proposed power allocation techniques decrease the number of unavailable links between intermediate nodes in the sub-graph and considerably increase the chance of finding an optimal solution.

Keywords: Wireless multicast, Quality of service (QoS), Power control, Network coding, Gradient power allocation (GPA)

1. Introduction

Nowadays, real time multimedia constitutes a major portion of IP-based broadcast services [1]. The difference between these broadband and other traditional telecommunication services is their high sensitivity to factors affecting quality of service (QoS) such as resource level, propagation condition and stability [2]. In such services user experience depends directly on the QoS level. So, not only making improvements of quality of service is important for better providing this type of services, but also guaranteeing quality of service is necessary.

In order to satisfy the QoS requirements in multimedia services, routing algorithms must be modified and adjusted to clarify the distinction between the data related to each service class to choose the best route to satisfy the QoS requirements [3]. Provision of QoS based on routing utilizes two common approaches:

- Selecting a route which is able to supply the QoS requirements in the relevant class;
- Using effective techniques to attain the maximum capacity of the network.

Previous studies have proven that network coding technique increases the capacities and effectiveness of the network to bring its available capacity closer to maximum theoretical capacity [4]. A proper efficient information exchange using network coding has been presented by [5]. Due to the numerous advantages of network coding, many algorithms which previously used routing are being modified to incorporate network coding [6]. As an example, using network coding has made it possible address the nondeterministic problem of achieving minimum-cost multicast routing to increase network capability [7]. Meanwhile, using this technique maximum network throughput between the source and each receiver based on the max flow-min cut algorithm in a multicast session will be achievable by choosing the optimal sub-graph [8].

The optimality of most existing multicast approaches based on seeking the optimal sub-graph does not guarantee satisfaction of hard QoS constraints, although; they rely on optimization schemes to determine the proper flow sub-graphs to minimize cost functions [9], [10]. So, it appears that using coded-based multicasting optimization the maximum capacity of network is achievable and a reasonable procedure to finding an optimal sub-graph that guarantees QoS constraints will be possible.

The present study uses the decomposition method to identify potential sub-graphs for a coded multicast session and applies route-selection mixed integer programming (MIP) method to tackle the path-flow framework [11] to determine the best sub-graph among all the potential sub-graphs which can satisfy QoS constraints with the minimum cost. However the path-flow algorithm used to find a coded multicast solution has a simple structure. Computational evaluations show that the approach presented in this paper, called “QoS-aware coded multicast network” (QoSCM) is an efficient method for solving the problem of optimal multicast routing in which the primal-dual algorithm considers only trees their QoS level is aligned with user QoS constraints.

Note that implementation of the model in distributed wireless networks having energy limits [12], Rician and Rayleigh fading, frequency interference and lack of centralized control over the network entities, faces critical challenges that do not exist in wired networks. In this regard, considering the nature of wireless environment, numerous studies have been carried out on the use of network coding in dynamic wireless networks [13], [14], [15]. In

[13], the authors tried to solve a multicast problem based on minimum energy using the advantages of dynamic multicast. The study in [15] proposed a distributed protocol supporting multiple unicast flows using the shared nature of interfaces in wireless networks. Also, Xi in [16] proposed a method of finding flow sub-graphs so that the minimum transmission rate is satisfied and the utilization function is at an optimum level. From this viewpoint, this scheme is somewhat similar to our proposed approach; however, some aspects of the introduced approach have not been considered in previous articles. The proposed scheme with the aim of guaranteeing the QoS constraints has been extended to distributed wireless networks using novel strategies and considering the dynamic characteristic of the network in dealing with interference.

The present study focuses on dynamic wireless networks in which the flexible capacity of network links can be interpreted as a function of the signal to noise-interference (SINR) in the receiver. Such a wireless network dynamically adjusts link capacities by modifying the power allocated to each link. In fact, in order to increase achievability of the optimal sub-graph, an optimization method is used for coded multicast based on a power control algorithm in the physical layer.

It is important to note that resource allocation-based optimization should not exert an excessive computational load on the network so that network performance remains optimal. With regard to the points raised, we propose our distributed scheme as an iterative gradient algorithm; in which after receiving the control messages of neighboring nodes, the network flow variables which emerge as optimization Lagrange coefficients are updated locally by each node in each iteration. After convergence of the algorithm, the optimal power obtained can be allocated to the outgoing links of each node. Clearly, it is evident that the number of iterations required to reach the optimal value relates to the network scale in term of number of nodes and links.

The continuation of this paper is structured as follows: first, Section II presents the problem and the formulation of guaranteeing the QoS constraints during coded multicast session based on minimum possible cost. Then, in Section III, in order to evaluate the functionality of QoSCM under actual conditions, we have introduced a framework to extending the proposed approach to wireless networks. Also two power allocation algorithms are proposed and the effectiveness of the scheme is analyzed using power control algorithm. Finally, after evaluation of the numerical results in part IV, the conclusions and recommendations for future study are presented.

2. QoS-aware Network Coding

In this section, the problem formulation and the coded-based flow algorithm are presented from the perspective of guaranteeing user QoS constraints.

2.1 Network Model and Notation

In this paper, the network is modeled with a $G:(V, E)$ graph where V is the set of nodes with $|V|=N$ and E is the set of network links (arcs) with $|E|=L$. In an individual multicast session, the source node $s \in V$ sends its data at the rate of r packets per unit time to each set of $K \subset V$ receivers. In this model, $X_l^{(t)}$ is the data packet sent from the transmitter towards the receiver $t \in V$ on link l. and the value of the final coded flow for each link shown as Z_l is obtained by combining the data flow sent from the transmitter to receivers. The characteristics of each communication link are specified with the predetermined weight coefficients to implement the proposed scheme in this network model.

The link coefficients represent total delay (*d*) and cost coefficient (*e*). The delay constraint and transmission rate of each link are preset fixed values. Each link delay is constant and calculated based on the propagation delay plus delay of the MAC layer. During transmission, each data stream belongs to a specific QoS class possessing specific quality constraints. Also, we consider *M* QoS class each of which has minimum transmission rate *R* and maximum threshold *D* for delay. Accordingly, the transmitter *s* only provides service where a sub-graph exists with the ability to guarantee the QoS constraints based on the class of requested service. Moreover, if more than one sub-graph is able to provide the QoS requirements, the sub-graph with the lowest cost is selected as the response sub-graph.

The data units and coded flows toward destination *t* belonging to QoS class *c* passing through link *l* are denoted as $X_l^{(t)(c)}$ and $Z_l^{(t)(c)}$ respectively. The total coded flow for class c is denoted by $Z_l^{(c)}$. These coded flows will be sent independently by each link. The total flow transmitted by link *l* for class *c* is denoted by $Z_l^{(c)}$. Consequently, the coded flows belonging to class c and the set of all packet data $(Z_l^{(c)}, X_l^{(t)(c)})$ are denoted as the *Z* and *X* matrix.

Taking into account the end-to-end delay experienced by network flow, it is assumed that each link incurs a constant delay with value d_l on the flows for a given period of time. It is assumed that before running the algorithm, each node has measured the delay of its outgoing links; therefore, d_l could be viewed as the short-term average delay of link *l*. To consider the constraints of the nodes in terms of QoS and cost, node splitting technique was applied. For this purpose, the processing delay and the cost of the nodes are also considered when calculating the delay and cost of end-to-end connections. **Fig. 1** shows that each node in the network is replaced by two dummy nodes and the connection between them is through a dummy link with a delay equal to the processing delay of the node [17]. Using this method, there are no alterations in the steps of algorithm implementation to guarantee QoS. Moreover, the cost for each coded data unit sent over link *l* denoted as e_l and the total cost can be calculated by multiplying the total passing flow (Z_l) by the link cost coefficient (e_l). To determine the final end-to-end cost for each sub-graph, a constant cost coefficient for each node is allocated to the dummy link.

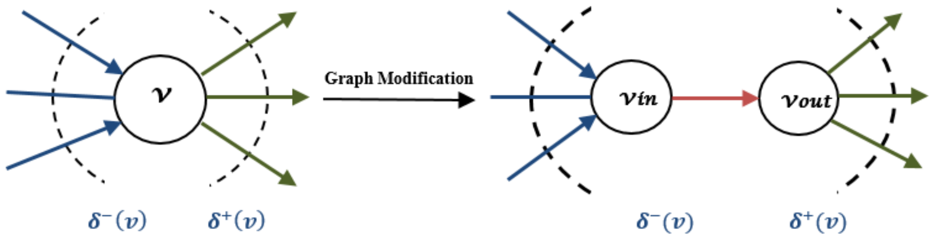

Fig. 1. Node splitting technique via graph modification

The solution sub-graph must guarantee that each flow experiences the end-to-end delay below the upper bound $D^{(c)}$ and should provide the minimum rate of $R^{(c)}$ for all the end-to-end connection relevant to class c. Utilizing network coding technique, the data flows belonging to the same QoS class are combined, although inter-class data combinations are prohibited. Accordingly, based on introduced framework, data passing through a link belongs to various classes in which, each class composed various data related to same class going toward different receivers. Using this approach, data belonging to each class experience a delay which is lower than the upper bound and will have a transmission rate larger than the minimum constraint specified for the class.

2.2 Problem Formulation

In line with the discussions presented in this paper, the target is to propose a solution for finding a sub-graph between the source and the receivers so that the end-to-end delay of each flow towards terminal t is not larger than upper bound $D^{(t)}$and the throughput for each flow is not less than lower bound $R^{(t)}$. In order to select the response sub-graph towards receiver t, if $P^{(t)}$ is the set of all directed routes from the source node s to this receiver in network G, $f(P)$ will be depicted as the number of flows passing through path p. In this case, if $\delta_l(p)$ denotes the path-link variable, it will be equal to 1 if link l belongs to path p; otherwise, $\delta_l(p)$ will be 0. In this way, the desired assessment of the cost calculation for each sub-graph can be carried out based on the links which make up that sub-graph. The sub-graph with minimum possible cost which satisfies the requirements of QoS will be selected as the solution for the problem. The link packet data ($X_l^{(t)}$) can be calculated as shown in Equation (1).

$$x_l^{(t)} = \sum_{p \in p^{(t)}} \delta_l(p) f(p) \tag{1}$$

Through end-to-end connection, weighted path $p \in P^{(t)}$ can be defined as shown in Equation (1.a):

$$W^{(t)}(p) = \sum\nolimits_{l \in p} w_l \tag{1.a}$$

In Equation (1.b), the upper bound of the weighted end-to-end path from transmitter s to receiver t is denoted as $U^{(t)}$ which this constraint guarantees the most tolerable end-to-end deviations.

$$\max_{p \in P^{(t)}} \left(W^{(t)}(p) \right) \leq U^{(t)}, \qquad \forall t \in K \tag{1.b}$$

Equation (1.b) can be written as:

$$\sum_{l\in p} w_l \le U^{(t)}, \qquad \forall t \in K, p \in P^{(t)} \tag{1.c}$$

Equation (1.c) should be guaranteed through each path included the optimal sub-graph. In order to use binary variables, the weighted constraint should be changed as Equation (1.d)

$$f(p) \le r\, y(p), \quad \forall t \in K, \forall p \in P^{(t)} \tag{1.d}$$

In the equation (1.d) the source data rate shown as *r* and $y(p) \in \{0,1\}$. Also, $y(p)$ is equal to 1 only if path *p* is part of solution sub-graph between source and receivers. Under these circumstances, *y*(*p*) will be used in constraint (1.e) to guarantee the upper bound through the end-to-end weighted connection. By defining the binary variables, the formulation continues in binary mode as:

$$y(p)\left(\sum_{l\in p} w_l\right) \le U^{(t)}, \quad \forall t \in K \tag{1.e}$$

$$y(p) \in \{0,1\}, \ \forall t \in K, \forall p \in P^{(t)} \tag{1.f}$$

Optimal sub-graph selection procedures have functionalities that guarantee the weighted constraints and allow formulation of the cost function as shown in Equation (2). The cost function for the network is assumed to be convergent and will be an incremental relation based on the value of final flow (*Z*) passing through the links of each sub-graph. The final cost of each sub-graph can be obtained using the linear function of Equation (2) based on the sum of flows for all links and the cost coefficient related to the sub-graph links.

$$f(Z) = \sum_{l\in E} e_l Z_l = \sum_{l\in E}\sum_{c=1}^{M} f(z_l^{(c)}) = \sum_{t\in K}\sum_{l\in E}\sum_{c=1}^{M} f(z_l^{(c)(t)}) \tag{2}$$

The relations of the coded-flow optimization problem can be defined based on guaranteed QoS as follows:

$$\min_{\underline{x}} \sum_{l\in E} f(Z_l) \tag{3}$$

where:

$$\begin{cases} Z_l = \sum_{l \in E} \sum_{c=1}^{M} e_l z_l^{(c)} & (3.a) \\ z_l^{(c)} = \max_{t \in T} \left\{ x_l^{(t)(c)} \right\} & (3.b) \\ 0 \le Z_l \le a_l \quad \forall l \in E & (3.c) \end{cases}$$

Based on the introduced framework, Equation (4) always existed:

$$\sum_{p \in p^{(t)}} \delta_l(p) f(p) \le Z_l \qquad \forall l \in E, t \in K \tag{4}$$

Equation (5) denotes the throughput constraint for each session based on the lower bound of throughput for each class.

$$r^{(c)} \ge R^{(c)} \quad c = 1, \dots, M \tag{5}$$

$$\sum_{c=1}^{M} r^{(c)} \le \Re \tag{5.a}$$

Also, Equation (6) shows the delay constraint for each session based on the upper bound of delay relevant to each class.

$$\sum_{l \in E} d_l^{(t)} \delta_l(p). y_p \le D^{(t)} \; ; \quad \forall t \in K \tag{6}$$

For class c, the delay constraints for the set of receivers shown as follow:

$$D(t,c) = \sum_{l \in E} D_l(x_l^{(t)(c)}) \le D^{(c)} ; \qquad \forall t \in K, c = 1, \dots, M \tag{7}$$

$$D(t,c) = \max_{p \in p_l} D_p^{(t)(c)} \le D^{(c)} ; \qquad \forall t \in K, c = 1, \dots, M \tag{8}$$

where:

$$D_l\left(x_l^{(t)(c)}\right) = \begin{cases} d_l & if\ x_l^{(t)(c)} > 0 \\ 0 & if\ Otherwise \end{cases} \tag{9}$$

$$x_l^{(t)(c)} \ge 0 \; ; \quad \forall l \in E, t \in K, c = 1, \dots, M \tag{9.a}$$

To use node splitting technique, according to **Fig. 1** we have:

$$\begin{cases} Delay\ v \sim delay_{dummy\ link}:\ delay_{Input\ Processing}\delta^{-}(v) + delay_{Output\ Processing}\delta^{+}(v) \\ \qquad delay_{s-\delta^{+}(v)} = delay_{s-\delta^{-}(v)} + delay_{dummy\ link} \end{cases}$$

The total end-to-end delay and cost constraints appear to be as follows:

$$D_{total}(Z_l) = \sum_{l \in p^k} delay\ (l) + \sum_{n \in p^k} delay\ (n) \tag{10}$$

$$C_{total}(l, n) = \sum_{l \in p^k} cost\ (l) + \sum_{n \in p^k} cost\ (n) \tag{11}$$

Based on these formulae, $f(Z_l)$ is the cost function of the problem and a decrease in its value will bring the sub-graph solution of the problem closer to an optimal value. Moreover, a_l and e_l denote the capacity of the link and linear cost coefficient for using the link, respectively, and d_l denotes the constraint for total delay of link l. In the formulation for the proposed model, $R^{(c)}$ is the minimum acceptable threshold of the transmission rate and $D^{(c)}$ is the maximum tolerable threshold for the end-to-end delay in class c. $\mathfrak{R}$ refers the network max-flow/min-cut rate and $r^{(c)}$ represents the actual class rate relevant to class c.

Equation $(3.\mathrm{c})$ provides the capacity constraint. This relation will ensure that the total value of the flow passing through a link cannot be higher than the capacity of that link. Equation (5) requires that the data transmission rate for class c must be higher than the minimum rate requested by the user. Equation $(5.\mathrm{a})$ is rooted in the theorem of max flow-min cut and indicates that the total rate of data transmission for all classes will be lower than the max-flow rate of the network. Note that the max-flow rate is the maximum theoretical rate of the network.

Equations (7) and (8) represent the end-to-end delay constraint in which the maximum tolerable delay of class c must be lower than the defined threshold for this class. Equation $(9.a)$ shows the necessity for nonnegative flow constraint. In other words, the direction for the coded flow is always constant and a link is not capable of sending a flow in both directions. Also, equations (10) and (11) refer to the total end-to-end delay and cost constraints. By accounting for the constraints of the problem, solving Equation (3) will result in the solution with the lowest possible cost among the sub-graphs satisfying the requirements of QoS.

In order to solve the formulized problem, the convex optimization methods are used to reduce the complexity of problem using primal-dual techniques and dividing the problem into sub-problems with lower complexity. Using this defined framework, achieving to QoS constraint with minimum cost is possible in the form of coded multicast optimization model. As shown in [18], the separate use of these two techniques: finding optimal flow sub-graphs and using coding techniques, will not affect the optimal nature of the combination methods. So other coding methods can be used to increase the performance of the scheme [19].

The next section extends the proposed approach to wireless networks. For this purpose, two distributed resource allocation algorithms; random power allocation (RPA) and gradient power allocation (GPA) have been introduced. The performance of the proposed method in more practical scenarios will be assessed by considering the limitations and constraints of real wireless networks.

3. Applying the Proposed Algorithm for Realistic Wireless Networks

The evaluation of previous schemes indicates that using optimal routing techniques in centralized networks makes finding the optimal sub-graph possible [20]. However, in distributed wireless networks, implementation of QoS assurance services is more difficult because the lack of centralized control over network entities, limited available resources and the Rician/Rayleigh fading that arise in the propagation environment. Accordingly, to implement the proposed approach, it was extended in a distributed manner taking into account the specific characteristics of wireless networks. Also, resource allocation and power control techniques were employed in this dynamic model in which the capacity of the links varies based on allocated resource.

Our proposed model is composed of two major phases. In the first phase, the optimal transmission power level for each node that maximize the total network capacity are found. Then, in the second phase, to reduce the number of unavailable links in the sub-graph and satisfy the total constraint for the total aggregated transmission power of all nodes in the network, a power control algorithm is used. The transmitters which signal-to-interference-plus noise ratio (SINR) of their outgoing links lies below the acceptable threshold are allowed to increase their transmission power so that the SINR of their link satisfies the minimum acceptable threshold for reliable transmission.

3.1 RPA-based Power Control

As indicated, to use a power control algorithm in a wireless network, a power allocation method is required to allocate resources dynamically based on environmental factors like interference and noise. To enhance the number of available links, RPA is proposed to satisfy the interference constraints and enhance connection stability. In the proposed approach, each node is locally able to individually determine the sub-session rate in its outgoing links [21]. The RPA algorithm determines a set of transmission powers for the nodes included the sub-graph besides of satisfying the total aggregated transmission power constraint.

This could make some previously unavailable links available by increasing their transmission power levels. It is assumed that the wireless network is interference limited, so that the capacity of link (i,j), denoted as C_{ij}, is a nonnegative function of the SINR at the receiver of the link, i.e., $C_{new}ij = f(SINR_{ij})$. It is also assumed that $C(.)$ is increasing, concave, and continuously differentiable. For $(i,j) \in E$, the end point (*j*) determines the value of the SINR according to matrix G by Equation (12):

$$SINR_{ij}(P) = \frac{G_i P_{ij}}{G_{ij} \sum_{n \neq j} P_{in} + \sum_{m \neq i} G_{mj} \sum_n P_{mn} + N_j} \quad (12)$$

In the SINR formula, N_j is the noise power at receiver node j and G_{ij} represents the gain of a link between nodes i and j. Corresponding matrix G can be obtained as:

$$G = \begin{bmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{n1} & \cdots & g_{nn} \end{bmatrix} = [g_{ij}]_{\substack{i=1,\dots,n \\ j=1,\dots,n}}; \quad where: \begin{cases} 0 < g_{ij} = g_{ji} < 1 & ;\ if\ i \neq j \\ g = 0 & ;\ if\ i = j \end{cases} \quad (13)$$

The set of allocated powers to the set of nodes denoted by the vector $P = [p_1, p_2, \dots, p_n]$, so that:

$$P_i = \{P: \sum_j p_{ij} \leq \overline{p}_i \quad \forall i \in V,\ \ p_{ij} \geq 0, \forall i,j \in V\}.$$

It should be noted that each node can assign primary power to its outgoing links by considering its individual power limitation: $\sum_j P_{ij} = P_i \leq \overline{P}_i$, where P_{ij} is the power allocated to the transmitter of link *ij* from node *i* toward node *j*, P_i is the aggregated power of node *i* on all of its outgoing links, and $\overline{P}_i$ is an acceptable threshold for P_i.

To confirm the accuracy of the data received in the receiver, the value of the SINR is considered as the criterion for reliability of data transmission. In this way, by calculating the value of SINR at the end point of each link, the probability of temporary failure in that link can be estimated. Using the calculated SINR, the link new capacities can be determined as shown in Equation (14).

$$New\ C_{ij} = f(SINR_{ij}) \quad (14)$$

For wireless networks to have reliable transmission, calculated SINR in end point of link should be greater than acceptable threshold which this threshold known as the node sensitivity level. This threshold is assumed to be constant for all nodes in the network. Failed or unavailable links assumed which their SINR is below than threshold and such unavailable links will be ignored during the process of finding the solution. So, they will not be included in the potential solutions. It is evident that, by increasing the number of failure links, the probability of finding the optimal solution will decrease. It will also increase the cost of the obtained sub-graph. In RPA, the nodes located at starting point of corrupted links are allowed to increase their transmission power to satisfy the minimum SINR requirements for reliable transmission. Note that the maximum aggregated transmission power for all nodes in the network is considered to be constant.

The main advantage of using the RPA as the power allocation algorithm is its simplicity and ability to reduce the complexity of the proposed solution. Otherwise, its main drawback is its sub-optimality and blind nature of power allocation instead of intelligent allocation. The next section proposes an algorithm which eliminates these drawbacks and enhances the efficiency of the power allocation algorithm using the gradient optimization techniques.

3.2 GPA-based Power Control

Most literatures in network coding based optimization such as [16] have only focused on acquiring an optimum value for applied utilization function in a cross-layer framework. But, such dynamic optimization algorithms are not effectively possible. The main difference between the proposed approach and such other schemes is its efficiently dynamic properties which introduced a gradient-based optimization framework in order to optimal power allocation with low computational complexity.

Determining the optimal values in the proposed model is not purely based on local node variables. This method solves optimization Lagrange equations iteratively; each node calculates the updated Lagrange coefficients and uses them for the next iteration of the power allocation algorithm. In the system model, it is assumed that the system has N users, and the K^{th} user has a data rate equal to R_k bits per symbol. $C_k^{(c)}$ is the number of bits allocated to the c^{th} class-carrier for the K^{th} user. In the transmission channels, different class-carriers will experience different channel gains, denoted by $\alpha_k^{(c)}$, the magnitude of the c^{th} class-carrier seen by the K^{th} user. It is assumed that N_0 is white noise and is equal for all class-carriers and the same for all users. The goal of this approach is the best assignment of $C_k^{(c)}$ which besides of satisfying total power constraint, the capacity of links has been maximized. Note that this problem can be formulated either to minimize the transmission powers besides satisfying the given QoS requirements or to improve the user QoS parameters for a fixed overall transmission power. The formulation for the second approach can be achieved by changing the class-carrier power levels proportionally using the same set of $C_k^{(c)}$. The enhancement in quality can be demonstrated by the increase in total user transmission rate (R) as follows:

$$maximize \sum_{k=1}^{K} \alpha_k R_k \quad subject\ to: R_k = \sum_{c=1}^{M} C_k^{(c)} \tag{15}$$

Replacing $C_k^{(c)}$ with its equivalent in Equation (15) results the formulation of the goal function in Equation (16). The goal of this formula is to maximize the weighted aggregation rate of the network users considering total power constraints.

$$maximize \sum_{k=1}^{K} \sum_{c=1}^{M} \alpha_k \rho_k^{(c)} \log(1 + \frac{p_k^{(c)} \left|h_k^{(c)}\right|^2}{\Gamma . n_k^{(c)}}) \tag{16}$$

$$subject\ to: \sum_{k=1}^{K}\sum_{c=1}^{M} \rho_k p_k^{(c)} \leq P_{tot}\ , \sum_{k=1}^{K} \alpha_k = K$$

where each user is allowed to transmit at a rate associated with α_k which is related user priority and $\rho_k^{(c)}$ a coefficient with values within the interval 0 and 1 associated with the c^{th} class-carrier for $c = 1,2,..,M$. The Lagrange equation in Equation (17) is used to take the problem constraints into account.

The channel coefficient for the k^{th} transmitter node on the c^{th} class is denoted by $h_k^{(c)}$ and includes the path loss, Rayleigh fading factor, and constant coefficients for the transmitter and receiver antenna. Moreover, Γ indicates the SINR-gap which is function of coding and modulation and the bit error ratio (BER). For example, for the modulation of non-coded QAM, the SINR-gap Γ will be shown as $\Gamma = {}^{-\ln(5.BER)}/_{1.5}$ and $n_k^{(c)}$ denotes the noise of class c for the node k.

$$L(P,\lambda) = \sum_{k=1}^{K}\sum_{c=1}^{M} \alpha_k \rho_k^{(c)} \log(1 + \frac{p_k^{(c)} \left|h_k^{(c)}\right|^2}{\Gamma . n_k^{(c)}}) \ - \lambda\left(\sum_{k=1}^{K}\sum_{c=1}^{M} \rho_k^{(c)} p_k^{(c)} - P_{tot}\right) \tag{17}$$

Assuming that the derivation of Equation (17) is equal to zero, the optimal power allocated to each class-carrier can be obtained. For each iteration of the power assignment algorithm, a current power of the class-carrier that does not satisfy the reliability will be modified and the Lagrange coefficient will be updated accordingly. After R iterations, the algorithm provides an optimal transmission power for each class-carrier as demonstrated by $P_k^{(c)*}$.

$$p_k^{(c)*} = \rho_k^{(c)} \left[\frac{\propto_k}{\lambda} - \frac{\Gamma . n_k^{(c)}}{\left|h_k^{(c)}\right|^2}\right]^+ \tag{18}$$

Where λ is a Lagrange coefficient associated with transmission power constraints. Equation (19) can be solved by defining auxiliary variable g as:

$$g_k^{(c)} = \frac{|\,h_k^{(c)}\,|^2}{n_k^{(c)}}$$

$$p_k^{(c)*} = \rho_k^{(c)} \left[\frac{\alpha_k}{\lambda} - \frac{\Gamma}{g_k^{(c)}}\right]^+ \tag{19}$$

The optimal power allocation formulation in Equation (19) is very similar to the common waterfilling problem. In this formulation, power is assigned to nodes based on the difference in their weighting coefficients. Each class-carrier of user k is assigned to a flow level equal to $\frac{\alpha_k}{\lambda}$. After waterfilling, the different users have flow levels that are proportional to their weighting coefficients. The users with the higher weighting factors have higher flow levels and can allocate more power to their class-carriers. The Equation (20) can be used to convert a single level waterfilling to a multilevel.

$$\frac{p_k^{(c)*}}{\alpha_k} = \rho_k^{(c)} \left[\frac{1}{\lambda} - \frac{\Gamma}{\alpha_k . g_k^{(c)}} \right]^+ \quad (20)$$

Starting with small values for initial Lagrange coefficients, the coefficients are modified in each iteration so that the data rate constraints of different users are satisfied in each node. Each node is treated in turn using the new allocated power until the SINR in the output of the links with temporary failure reach at a minimum level of sensitivity. When the link is recovered, the possibility of obtaining a solution sub-graph in the QoSCM algorithm will increase. In this way, the data rate constraint and dedicated power constraint for all nodes are satisfied and the algorithm will converge.

$$g(\lambda) = \max_{p_k^{(c)}} L(P, \lambda) \approx minimize\ g(\lambda)\ ; \lambda \geq 0 \quad (21)$$

The solution to Equation (21) leads to optimal values for maximizing $L(P, \lambda)$. The gradient method can be used to solve the problem to obtain a distributed solution. According to the steepest descent lemma, we have Equation (22):

$$\lambda(t+1) = [\lambda(t) - \gamma . \nabla g(\lambda(t))]^+ \quad (22)$$

where $g > 0$ is the step size and $[Z]^+ = \max\{z, 0\}$. Using Equation (22) and according to dual function derivation, we will have Equation (23).

$$\nabla g(\lambda(t)) = \sum_{k=1}^{K} \sum_{c=1}^{M} \rho_k^{(c)} P_k^{(c)*} - P_{tot} \quad (23)$$

Using the Lagrange coefficient (λ), the class-carrier powers in the next iteration can be calculated as in Equation (24):

$$P_k^{(c)*} = \rho_k^{(c)} \left[\frac{\alpha_k}{\lambda(t)} - \frac{\Gamma}{g_k^{(c)}} \right]^+ = P_k^{(c)}(\lambda(t)) \quad (24)$$

The Lagrange coefficient λ can be updated for the next step in Equation (25) during successive iterations.

$$\lambda(t+1) = \left[\lambda(t) - \gamma\left(\sum_{k=1}^{K}\sum_{n=1}^{N}\rho_k^{(c)}p_k^{(c)}(\lambda(t) - P_{tot})\right)\right]^{+} \tag{25}$$

Based on the algorithm performance, the SINR can be calculated as the reliability of the transmission in the output of each link. By calculating the value of this parameter, the status of the link becomes clear and the temporary failures can be determined with greater confidence. The new SINR can be used to calculate their new capacities as $C_{ij} = C(SINR_{ij})$.

To achieve transmission reliability, the SINR value at the end point of each link is compared with the minimum SINR required for reliable transmission. In fact, the acceptable SINR threshold is the sensitivity level of the node. If the calculated SINR value is lower than the sensitivity, temporary failure is considered to have occurred for that link and the link will be removed from the set of potential sub-graphs. It is evident that, frequent link failure decreases the possibility of finding a solution sub-graph which satisfies the QoS requirements and increases the cost. To prevent link failure, the power control algorithm is used to gradually increase the transmission power of the node located at the beginning point of the failure links. It should be noted that during all the steps of power control, the total power allocated to all nodes must not exceed a pre-determined value.

4. Experimental Results

The performance of our proposed algorithm QoSCM was evaluated under different scenarios. For this purpose, two experiment modes were considered: one mode relates to QoS provision without considering interference constraints, for wired networks and another one considers the interference constraints for wireless environments. The proposed algorithm is compared in both modes with the minimum-cost multicast over coded network (MCNC) [18] and the delay-constrained minimum cost multicasting in traditional packet networks (CMCT) [22]. CMCT presents a solution by finding a minimum cost tree as the optimal solution and satisfying the constraints for the delay parameter without using a network coding mechanism. MCNC finds the minimum cost sub-graph between the source and destination using coding techniques without satisfying any QoS constraint.

The proposed algorithm was first simulated in its basic mode to find an optimal sub-graph to satisfy QoS constraints with minimum possible cost without considering interference. This was then extended to enhanced mode by introducing the effects of interference constraints into the problem. To compare the results of CMCT and MCNC algorithms, a multiple scenario was generated in which source s, which transmits in multicast mode, has three terminals t_1, t_2 and t_3. The QoS classes were determined with maximum transmission rates equal to 2 and 1 assuming a maximum tolerable delay of 4, 10, and 8 units as acceptable thresholds for the three receivers, respectively. To evaluate and also compare the resulted graph with other schemas, we have implemented the method by one QoS class with lower bound 1 for transmission rate.

Each link is associated with two entries (delay, cost). **Fig. 2** shows the delay and cost coefficients respectively, related to each link. The solution obtained by the MCNC algorithm which only considers the links costs shown in **Fig. 3**, although minimum cost multicast over the coded network does not satisfy QoS constraints. The sub-graph extracted by CMCT is shown in **Fig. 4**. Note that this solution is achieved without the use of coding schemes. The solution obtained by QoSCM has been shown in **Fig. 5** in which reducing the end-to-end delay and satisfying the transmission rate requirement for each class are the priorities to achieving the QoS target and minimizing the sub-graph cost.

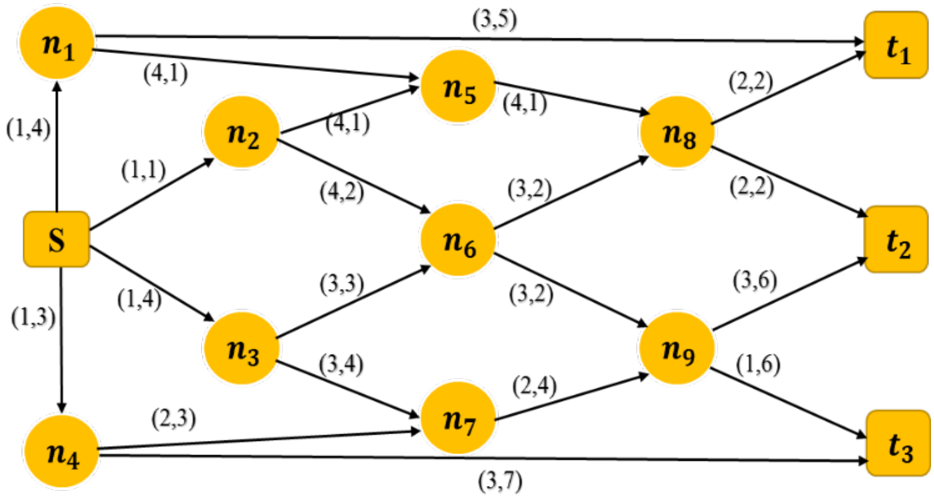

Fig. 2. Delay weight and Cost coefficient of each link

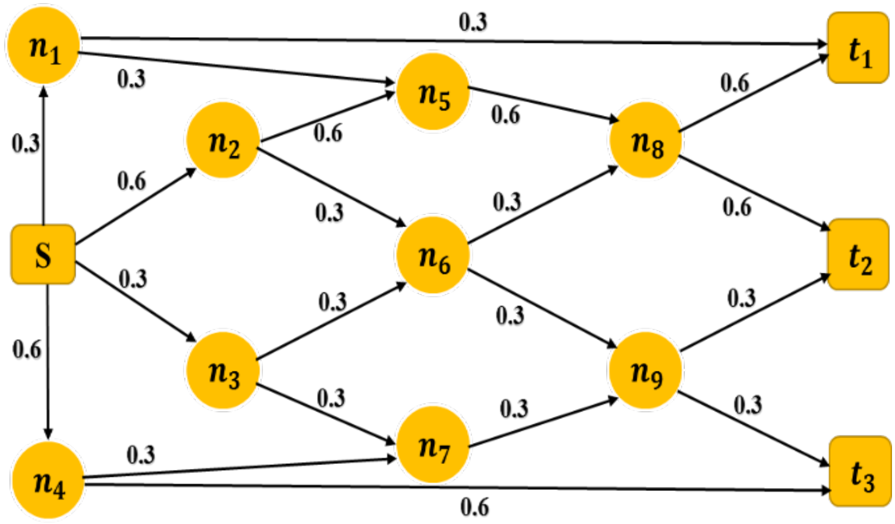

Fig. 3. MCNC algorithm sub-graph solution

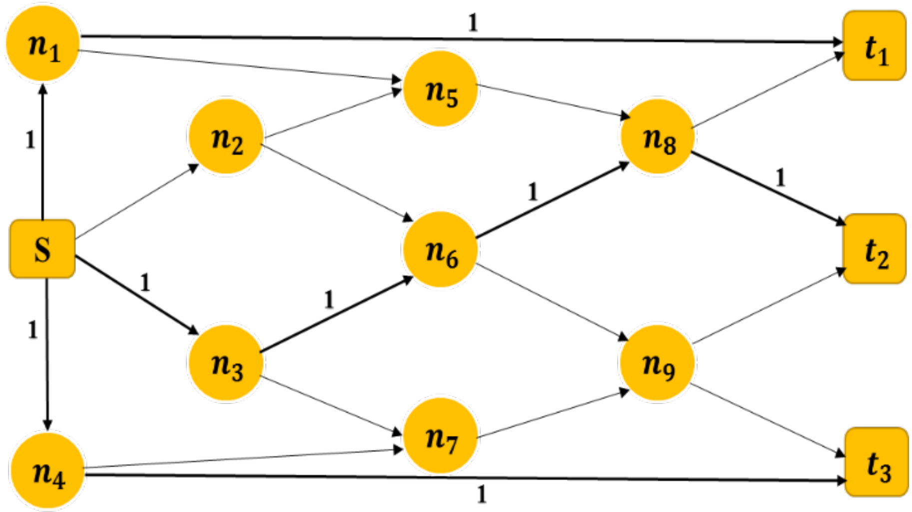

Fig. 4. CMCT algorithm sub-graph solution

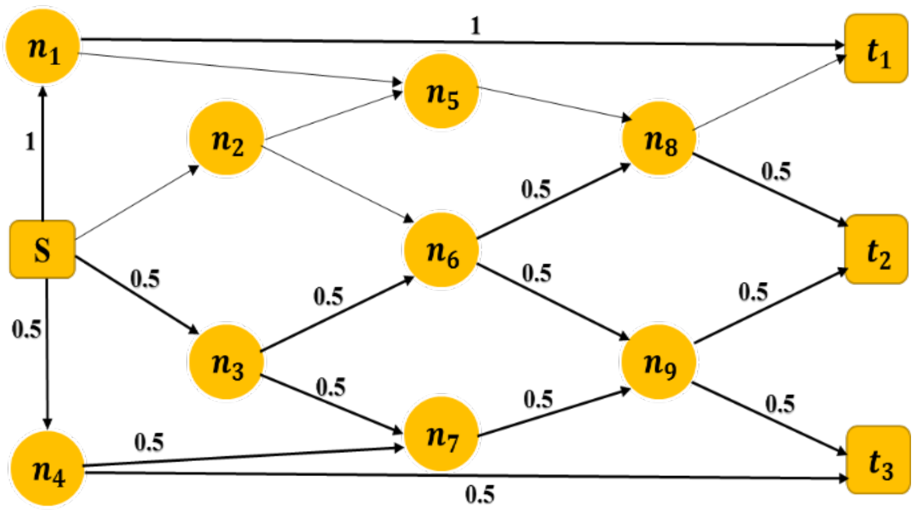

Fig. 5. QoSCM algorithm sub-graph solution

The solutions and cost of finding an optimal sub-graph to satisfy QoS constraints by the QoSCM algorithm were compared with the MCNC algorithm in **Table 1**. It is clear that decreasing the constraint level of the delay reduces the cost of finding an optimal sub-graph. The lower bound of transmission rate (R) is a constant parameter with a value that is twice that of the first class value of the second class. The maximum transmission rate was assumed to be 2 for the first and 1 for second class.

It is evident in the case that the range of the upper bound and lower bound thresholds for the service classes is not very strict, the MCNC algorithm will have a solution similar to the QoSCM algorithm at a lower cost. However, when the threshold becomes stricter, in the sub-graph obtained by MCNC, the delay of some flows will be higher than the threshold for

QoS in classes 1 and 2 (values indicated in bold and italics, respectively). It is clear that the desired level of service quality is not acceptable.

Table 1. Performance evaluation of MCNC and QoSCM

No. of Nodes	Algorithm	Class	Const	Delay (2 sinks)	Delay (4 sinks)	Total Cost (2 sinks)
15	MCNC	1	6	(4,***9***)	(6,5,4,4)	17.2
		2	8	(5,***9***)	(6,7,4,6)	
	QoSCM	1	6	(5,5)	(5,4,4,5)	21.1
		2	8	(6,7)	(7,5,6,6)	
20	MCNC	1	9	(***12***,8)	(***12***,8,7,6)	19.2
		2	11	(***12***,8)	(***12***,8,7,7)	
	QoSCM	1	9	(7,6)	(7,8,7,7)	25.3
		2	11	(10,8)	(11,9,7,9)	
32	MCNC	1	12	(***15***,***13***)	(***13***,***13***,11,8)	27.7
		2	15	(14,12)	(***17***,9,10,12)	
	QoSCM	1	12	(11,12)	(11,9,11,8)	36.1
		2	15	(11,14)	(14,12,9,13)	

To compare the cost of QoSCM with similar schemes, the network is considered to be random. It is assumed that there is an arc with a probability of 0.2 between i and j nodes. Nevertheless, the basis for the presence of a link between two nodes can be interpreted based on the fact that the distance between two nodes is less than a constant value. The linear cost coefficient related to each link is assumed to be a random number from interval [1 5] and the weight coefficient of each link is assumed to be selected randomly from interval [1 10].

In the three scenarios with different number of nodes in **Table 2**, increasing the size of the network increases the cost of finding a solution. It is clear increased cost using the QoSCM algorithm is more than MCNC because the functionality of MCNC is based on minimum cost and it does not satisfy QoS constraints. Also, the QoSCM has less optimal-cost in comparison with CMCT because of its packet coding ability.

Table 2. Cost comparison of MCNC, CMCT and QoSCM

No. of Nodes	Algorithm	Total Cost (2 sinks)	Total Cost (4 sinks)
(15) Nodes	**MCNC**	17.2	33.1
	CMCT	21.3	37.2
	QoSCM	*19.1*	*36.8*

(20) Nodes	**MCNC**	19.2	44.1
	CMCT	28.7	76.5
	QoSCM	*25.3*	*66.7*
(32) Nodes	**MCNC**	27.0	53.1
	CMCT	45.2	83.2
	QoSCM	*36.1*	*67.0*

To evaluate the effects of the proposed power allocation algorithms (RPA and GPA) at different levels of QoS constraint, the probability of finding a solution and the cost of proposed QoS coded multicast (QoSCM) were compared in two different modes. First, the QoSCM has been evaluated individual without a power control feature. Then, evaluation of QoSCM done with power allocation features (random and gradient).

Table 3 shows that for highly strict QoS constraints, neither could find an optimal solution. By gradually reducing the QoS constraint, for the constraint (6, 9) relevant to two receivers, the QoSCM algorithm in combination with GPA found an optimal solution while the two other approaches could not find a solution. These results were predictable because the gradient-based power allocation algorithm is the most efficient. Also, with the looser QoS constraints, other two approaches found a solution at nearly identical costs and number of iterations because despite some links have been corrupted, the majority of potential sub-graphs satisfied the QoS constraints.

Table 3. Optimality evaluation of QoSCM algorithm in combination with RPA and GPA power allocation schemes

Constraints	Algorithm	Performance		
		SUCCESS	**COST**	**ITERATION**
(5,7)	**QoSCM**	*NO*	----	----
	QoSCM+GPA	*NO*	----	----
	QoSCM+RPA	*NO*	----	----
(6,9)	**QoSCM**	*NO*	----	----
	QoSCM+GPA	*YES*	33.57	126
	QoSCM+RPA	*NO*	----	----
(9,11)	**QoSCM**	*NO*	----	----
	QoSCM+GPA	*YES*	27.20	113
	QoSCM+RPA	*NO*	----	----
(10,12)	**QoSCM**	*YES*	38.31	60
	QoSCM+GPA	*YES*	22.1	94
	QoSCM+RPA	*YES*	28.78	64
(14,17)	**QoSCM**	*YES*	24.11	32
	QoSCM+GPA	*YES*	20.89	71
	QoSCM+RPA	*YES*	20.89	36

Table 3 shows that, when all the algorithms arrived at a solution, the optimality of solutions found by the QoSCM algorithm in combination with GPA was much better than that of the solution found by the basic QoSCM algorithm, although the iterations required to achieve a solution was much higher than those required by the QoSCM individual. So, the complexity of the GPA is higher than for the RPA; however, in environments with high interference, the GPA can better satisfy strict QoS requirements of multicast sessions.

Fig. 6 shows an example of the QoSCM convergence. As shown in this figure, the bounded cost converges to an optimal value.

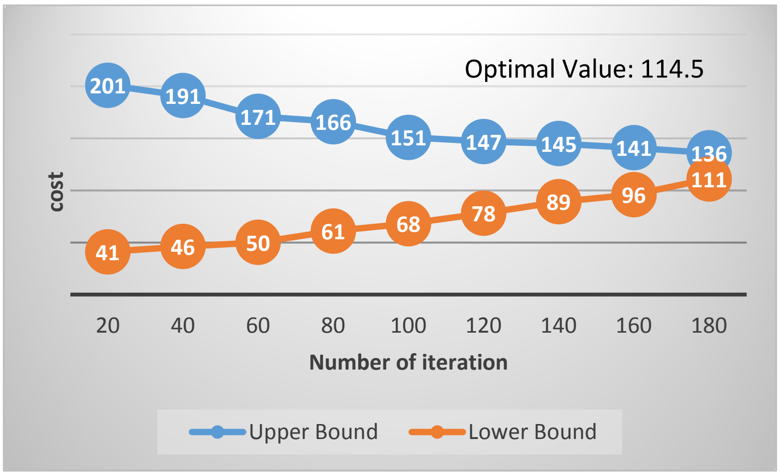

Fig. 6. Bounded algorithm's convergence during iterations with 60 nodes and 5 receivers

5. Conclusion and Future Works

The present paper proposed a flow-based optimization approach for distributed wireless networks in which coded flow multicasting has been utilized for guaranteeing the QoS constraints by finding an optimal sub-graph. In comparison to other schemes, the simulation results reveal that although the cost of the proposed model is somewhat more than MCNC algorithm, however it can better satisfy the QoS requirements of multicast sessions and its cost is lower than CMCT. Also, a discrete power control scheme was also applied for the QoS-aware multicast model to accommodate the effect of transmission power level based on link capacity requirements. we proposed random power allocation (RPA) and gradient power allocation (GPA) algorithms to efficient resource distribution and to increase link resistance to temporal failure caused by interference and noise. The simulation results prove that using introduced gradient power allocation algorithm considerably increases the chance of finding an optimal sub-graph.

A further study with more focus on other requirements of wireless networks are therefore recommended, such as considering different types of fading. Also, further research might investigate determining the propagation model intelligently based on environmental conditions.

References

[1] Liem, Andrew Tanny, I-Shyan Hwang, AliAkbar Nikoukar, Cheng-Zen Yang, Mohammad Syuhaimi Ab-Rahman, and Ching-Hu Lu, "P2P live-streaming application-aware architecture for QoS enhancement in the EPON," *IEEE Systems Journal*, pp. (99): 1-11, 2016. Article (CrossRef Link).

[2] De Pessemier, Toon, Isabelle Stevens, Lieven De Marez, Luc Martens, and Wout Joseph, "Quality assessment and usage behavior of a mobile voice-over-IP service," *Telecommunication Systems 61*, no. 3, 417-432, 2016. Article (CrossRef Link).

[3] Murugeswari, R., S. Radhakrishnan, and D. Devaraj, "A multi-objective evolutionary algorithm based QoS routing in wireless mesh networks," *Applied Soft Computing 40*, 517-525, 2016. Article (CrossRef Link).

[4] Qin, Yang, and Xiaoxiong Zhong, "Network Coding at Network Layer in Multi-hop Wireless Networks," in *Proc. of Network Coding at Different Layers in Wireless Networks*, pp. 59-93. Springer International Publishing, 2016. Article (CrossRef Link).

[5] Kwon, Minhae, and Hyunggon Park, "The Impact of Network Coding Cluster Size on Approximate Decoding Performance," *KSII Transactions on Internet & Information Systems 10*, no. 3, 2016. Article (CrossRef Link).

[6] Ning, Zhaolong, Qingyang Song, Lei Guo, Zhikui Chen, and Abbas Jamalipour, "Integration of scheduling and network coding in multi-rate wireless mesh networks: Optimization models and algorithms," in *Proc. of Ad Hoc Networks 36*, 386-397, 2016. Article (CrossRef Link).

[7] Maheshwar, Shreya, Zongpeng Li, and Baochun Li, "Bounding the coding advantage of combination network coding in undirected networks," *IEEE Transactions on Information Theory 58*, no. 2, 570-584, 2012. Article (CrossRef Link).

[8] Le, Tan, Xing Chen, and Yong Liu, "NCOM: network coding based overlay multicast in wireless networks," *Wireless Networks 21*, no. 1, 187-199, 2015. Article (CrossRef Link).

[9] Yu, Zhanke, Feng Ma, Jingxia Liu, Bingxin Hu, and Zhaodong Zhang, "An efficient approximate algorithm for disjoint QoS routing," *Mathematical Problems in Engineering 2013*, 2013. Article (CrossRef Link).

[10] Gabrel, Virginie, Maude Manouvrier, Kamil Moreau, and Cecile Murat, "QoS-aware Automatic Syntactic Service Composition problem: complexity and resolution," 2015. Article (CrossRef Link).

[11] Waller, S. Travis, David Fajardo, Melissa Duell, and Vinayak Dixit, "Linear programming formulation for strategic dynamic traffic assignment," *Networks and Spatial Economics 13*, no. 4, 427-443, 2013. Article (CrossRef Link).

[12] Aquino, Guilherme Pedro, Dayan Adionel Guimarães, and Marco EGV Cattaneo, "Energy efficient scheme based on simultaneous transmission of the local decisions in cooperative spectrum sensing," *KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS 10*, no. 3, 996-1015, 2016. Article (CrossRef Link).

[13] Jiang, Dingde, Zhengzheng Xu, Wenpan Li, and Zhenhua Chen, "Network coding-based energy-efficient multicast routing algorithm for multi-hop wireless networks," *Journal of Systems and Software 104*, 152-165, 2015. Article (CrossRef Link).

[14] Farooqi, Muhammad Zubair, Salma Malik Tabassum, Mubashir Husain Rehmani, and Yasir Saleem, "A survey on network coding: From traditional wireless networks to emerging cognitive radio networks," *Journal of Network and Computer Applications 46*, 166-181, 2014. Article (CrossRef Link).

[15] Aktas, Tugcan, A. Ozgur Yilmaz, and Emre Aktas, "Practical methods for wireless network coding with multiple unicast transmissions," *IEEE Transactions on Communications 61*, no. 3, 1123-1133, 2013. Article (CrossRef Link).

[16] Xi, Yufang, and Edmund M. Yeh, "Distributed algorithms for minimum cost multicast with network coding," *IEEE/ACM Transactions on Networking 18*, no. 2, 379-392, 2010. Article (CrossRef Link).

[17] Wei, Kaimin, Xiao Liang, and Ke Xu, "A survey of social-aware routing protocols in delay

tolerant networks: applications, taxonomy and design-related issues," *IEEE Communications Surveys & Tutorials 16*, no. 1, 556-578, 2014. Article (CrossRef Link).
[18] Zhao, Fang, Muriel Médard, Asuman Ozdaglar, and Desmond Lun, "Convergence study of decentralized min-cost subgraph algorithms for multicast in coded networks," *IEEE Transactions on Information Theory 60*, no. 1, 410-421, 2014. Article (CrossRef Link).
[19] Bassoli, Riccardo, Hugo Marques, Jonathan Rodriguez, Kenneth W. Shum, and Rahim Tafazolli. "Network coding theory: A survey," IEEE Communications Surveys & Tutorials 15, no. 4, 1950-1978, 2013. Article (CrossRef Link).
[20] Frangioni, Antonio, Laura Galli, and Giovanni Stea, "QoS Routing with worst-case delay constraints: models, algorithms and performance analysis," 2015. Article (CrossRef Link).
[21] Dai, Wenhan, Yuan Shen, and Moe Z. Win, "Distributed power allocation for cooperative wireless network localization," *IEEE Journal on Selected Areas in Communications 33*, no. 1, 28-40, 2015. Article (CrossRef Link).
[22] Oliveira, Carlos AS, and Panos M. Pardalos, "Mathematical aspects of network routing optimization," *Berlin: Springer*, 2011. Article (CrossRef Link).

Amin Mohajer received the BE degree in biomedical engineering from Shahed University, Tehran, Iran, in 2008, and the MS degree in electrical engineering from the Malek-Ashtar University of Technology, Tehran, Iran, in 2011. In August 2012, he joined Iran Telecom Industries where he is currently working in the field of radio frequency planning and network optimization. His research interests include wireless networks, optimization, next generation mobile networks and data analysis. E-mail: mohajer@mut.ac.ir

Morteza Barari was born in Freydoonkenar, Iran. He received the PhD degree from AmirKabir University of Technology in 2003. He is currently a faculty member at the Department of Electrical Engineering of the Malek-Ashtar University of Technology, Tehran, Iran. He published more than 50 papers and 2 books. His research interests are in stochastic signal processing, radar design, satellite communication, and adaptive array processing. E-mail: barari@mut.ac.ir

Houman Zarrabi received his BSc degree in computer hardware in 2003 from Shahid beheshti University, Tehran, Iran. He received his master of science and PhD degrees from Concordia university, Concordia, Canada, in electrical engineering and computer science, in 2006 and 2011, respectively. He is currently faculty member of Iran Telecommunication Research Center (ITRC), the Department of Communications Technology, Tehran, Iran. His research interests are Energy Management, Embedded Systems, Wireless Sensor Networks, Green Computing, HW/SW Co-Design, VLSI. E-mail: h.zarrabi@itrc.ac.ir