Optimum Internet Gateway Selection in Ad Hoc Networks

Felix Hoffmann, Daniel Medina
Institute of Communications and Navigation, German Aerospace Center (DLR), Wessling, Germany
Email: Felix.Hoffmann@DLR.de, Daniel.Medina@DLR.de

***Abstract*—Wireless ad hoc networks are connected to the fixed Internet by means of Internet gateways. Whenever a node within the ad hoc network wishes to communicate with a host in the Internet, it selects a default Internet gateway to relay its traffic from the ad hoc network to the Internet. In this paper, we formulate the problem of selecting the best Internet gateway as a mixed integer linear program minimizing the maximum node utilization in the wireless network. By simulations, we show that the performance that can be achieved by solving this optimization problem is significantly higher than what is achieved by standard gateway selection algorithms based on hop count or gateway load. In particular, these heuristic algorithms fail to adapt to the offered traffic and available capacity in the network.**

I. INTRODUCTION

In recent years, there has been a significant amount of research in the field of Mobile Ad Hoc Networks (MANETs). Whereas most work initially focused on stand-alone MANETs, there has been a growing interest in the connection of MANETs to the fixed Internet. This trend is also reflected by the ongoing activities of the IETF MANET and AUTOCONF Working Groups [2], [3].

Since all traffic that is sent between a MANET node and the Internet must pass through an Internet gateway (IGW) connecting the wireless and wired networks, these nodes present a potential bottleneck. Therefore, careful attention must be paid to selecting the Internet gateways in order to balance the load between them in a reasonable manner.

To be able to send traffic to the Internet, a mobile node within the MANET must first become aware of an IGW. This can be done either *proactively*, meaning that the IGW periodically sends out Internet Gateway Advertisements (IGWADVs) that are flooded through the MANET, or *reactively*, meaning that a node that has traffic to send floods an Internet Gateway Solicitation (IGWSOL) through the MANET. When an IGW receives such a solicitation, it responds with an IGWADV message that is unicasted back to the originating node. In addition, a *hybrid* approach has been suggested by Ratanchandani and Kravets [4] that makes use of proactive IGW discovery only within a certain distance of the IGW, measured in the number of hops. Mobile nodes outside this region must transmit IGWSOL messages on demand. This approach is intended to limit the total overhead due to the flooding of advertisements throughout the MANET.

Regardless of whether the proactive, reactive, or hybrid approach is used, an IGWADV message generally carries some information allowing the mobile node to choose the best gateway. The most common and most simple approach is to use the hop count of the message to choose the IGW that is closest in terms of hops [5]. More refined approaches propose to carry information about the current load of the IGW [6], [7], [8], thereby allowing for more efficient load sharing between IGWs.

In this paper, we consider the problem of optimum Internet gateway selection in heterogeneously connected wireless ad hoc networks, and formulate the gateway selection problem as a mixed integer linear program. By means of simulations, we compare the performance of the optimum allocation of gateways to nodes to the performance that can be achieved by hop count based selection and the Minimum Load Index solution presented by Huang et al. in [6]. To our knowledge, this is the first time that such a mathematical optimization approach has been applied to the gateway selection problem in wireless ad hoc networks. Although such an approach is not applicable to a real world MANET, since it requires complete knowledge of the network topology, traffic demands, etc., we expect that the results of this work will support the evaluation of different practical gateway selection schemes and provide insight into the impact of IGW selection on the network performance.

Compared to the problem of optimal routing in ad hoc networks, there is an additional degree of freedom here, since the destination of the flow is not fixed, but can be chosen from all gateways which a mobile node can reach through the MANET. This provides substantial room for optimization, since changing the destination of a flow may lead to a better spatial separation of flows, and thus less interference, in the network. Here, we separate the gateway selection problem from the routing and assume that the routes from the wireless nodes to the gateways are provided by a routing protocol running independently of the gateway selection.

This paper is structured as follows: First, we discuss several approaches to the gateway selection problem that have been proposed in previous work in Section II. In Section III, we present our network model that allows us to formulate the gateway selection problem as a mixed integer linear program in Section IV. In Section V, we present simulation results to compare the performance of the optimum gateway selection with the previous heuristic solutions. Finally, Section VI

This work is partially funded by the European Commission through the NEWSKY project [1] under contract no. 37160.

concludes this paper and presents steps for further work.

II. INTERNET GATEWAY SELECTION

The simplest method of Internet gateway selection in wireless ad hoc networks is to choose the gateway that is closest to in terms of hops [5]. The distance to the gateway can be inferred from the hop count of the gateway advertisements. The obvious drawback of this approach is that load balancing between gateways is not supported.

In contrast, Huang et al. [6] propose the Minimum Load Index (MLI) solution to select the Internet gateway based on the current load of the gateways. The metric L_i that is advertised by gateway i, referred to as its Load Index, is calculated according to the equation

$$L_i = \frac{\sum_x \rho_{x,i} \times T_x}{C_i}, \tag{1}$$

where C_i is the bandwidth of gateway i, and $\rho_{x,i}$ is the fraction of the total traffic T_x that host x sends to gateway i. However, a gateway cannot calculate its load index directly according to (1), since it is not aware of the hosts' T_x or $\rho_{x,i}$. Therefore, the gateway must estimate its current load. We assume that this is done according to a running average. Every time that a new advertisement is generated, the gateway estimates its current load $\widehat{L_i}[n]$ based on the previous estimate $\widehat{L_i}[n-1]$ and the volume of traffic $T[n]$ that it has processed since the last advertisement:

$$\widehat{L_i}[n] = \frac{1}{C_i}\left(\alpha T[n] + (1-\alpha)\widehat{L_i}[n-1]\right), \tag{2}$$

where $\alpha \in (0,1]$ is a weighting coefficient to choose the impact of the new load measurement $T[n]$.

In principle, each traffic source in the network selects the gateway advertising the lowest load according to (1) as its default gateway. In addition, several rules ensure that hosts do not switch between gateways too rapidly. When a host currently using gateway i receives an advertisement from another gateway j advertising a lower load, it configures j as its new gateway only if

- it as been using gateway i for a minimum duration T_{hold}, and
- after switching from gateway i to gateway j, the Load Index of gateway j will still be at least a predetermined threshold Δ_L less than the Load Index of i.

Even if these two conditions are fulfilled, the node only switches gateways with a certain gateway-switching probability, intended to reduce the amount of fluctuation between gateways. To assess the load balancing capability of the MLI algorithm, Huang et al. define the Load Balance Index (LBI) of the network as

$$LBI = \frac{\max\{L_i\} - \min\{L_i\}}{\max\{L_i\}}. \tag{3}$$

In case the network's traffic load is distributed among the gateways according to the gateways' capacities, the LBI becomes 0. If one gateway handles the entire traffic, the LBI is 1.

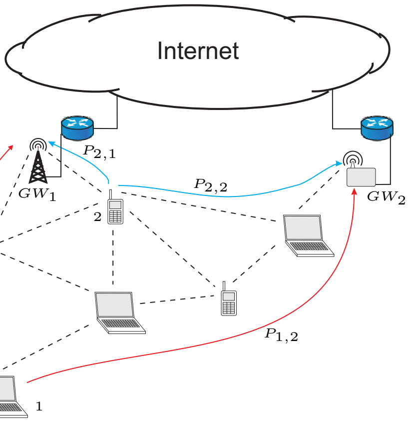

Fig. 1. Ad hoc network that is connected to the Internet via heterogeneous access links. Node 1 and node 2 each have two paths to a gateway.

One weakness of load based approaches such as the MLI solution presented above is that although they reflect the load of the gateway, they do not reflect the situation of the path through the MANET between the mobile host and the gateway. Also, the gateways must have an accurate estimate of their traffic load, which may fluctuate considerably.

To compensate for this weakness, Brännström et al. [12] propose to select the Internet gateway based on the variance of the time that elapses between the reception of two successive gateway advertisements. Intuitively, higher traffic load anywhere along the path between the gateway and the node will cause the advertisements to experience different delay and thus lead to a higher variance of their inter-arrival times. Hence, according to this scheme, a node should always select the gateway associated with the lowest variance.

III. NETWORK MODEL

Since we are considering heterogeneously connected networks, we assume that the wireless nodes are equipped with multiple interfaces: one interface for communication within the ad hoc network itself, as well as one or more interfaces used to connect to an Internet gateway. These different wireless communication systems are assumed to be sufficiently far apart in the frequency domain that interference between them does not need to be considered.

An example of such a heterogeneous network composed of laptop computers and smart phones is shown in Fig. 1. Here, there are two Internet gateways: a cellular base station, labeled as GW_1 and a WiFi access point, labeled as GW_2. Each of these gateways is connected to the Internet via an access router. Node 1 at the bottom of the figure has two potential paths to the two gateways, labeled as $P_{1,1}$ and $P_{1,2}$, respectively. These paths are determined by the routing protocol running within the wireless network.

For each of the systems, our interference model is similar to the *protocol model of interference* for a contention-based MAC presented by Gupta and Kumar [13]. According to this model, two nodes are able to communicate if they are within the maximum reception distance d_{rx}. A transmission potentially poses interference to any node within interference distance $d_{\text{int}} \geq d_{\text{rx}}$ of the sender. The wireless network topology is represented as a network graph, in which two nodes i and j are connected if the distance $d_{i,j}$ between the nodes is less than the reception distance d_{rx}. Nodes that are using multiple wireless interfaces, such as a node that is receiving a packet from another node using a Bluetooth link and forwarding it to a gateway using e.g. IEEE802.11, are conceptually split up into multiple virtual nodes that do not mutually interfere.

To account for interference, the authors in [9] make use of a *conflict graph*, which is constructed such that for every link in the network graph from node i to node j, there is a corresponding vertex $\ell_{i,j}$ in the conflict graph. An edge between two nodes of the conflict graph indicates that these two links cannot be used at the same time due to mutual interference. In IEEE 802.11, a sender only transmits if it senses the channel as free, i.e. if no other node within interference distance d_{int} is currently transmitting, and the transmission is received correctly if no node within interference distance of the receiver is transmitting simultaneously. Thus, two vertices $\ell_{i,j}$ and $\ell_{m,n}$ of the conflict graph are connected by an edge if $d_{i,m} \leq d_{\text{int}}$, $d_{i,n} \leq d_{\text{int}}$, or $d_{m,j} \leq d_{\text{int}}$. In contrast to the protocol model of interference in [9], we also require that the sender is free of interference to more accurately represent the sender's channel sensing. Given this model, finding combinations of links that may be used simultaneously then amounts to finding *maximal independent sets* of the conflict graph [9]. In graph theory, an independent set of a graph is defined as a set of vertices of the graph such that no two vertices are connected by an edge. A maximal independent set is an independent set that is not a subset of any other independent set of the graph.

A vertex of the conflict graph, representing a link of the network graph, may belong to multiple maximal independent sets. Since each independent set i may be used only for a fraction λ_i of the total time, the available data rate of each link is reduced accordingly. If $\mathcal{I}_e$ refers to the set of all maximal independent sets of the conflict graph that contain link e, the usable rate of link e reduces from its nominal rate r_e to $r_e \sum_{i \in \mathcal{I}_e} \lambda_i$.

Since all independent sets must share the same channel, the total utilization may not be greater than one:

$$\sum_i \lambda_i \leq 1. \tag{4}$$

Effectively, finding the factors λ_i amounts to solving the scheduling problem of a contention-based MAC.

In our network model, a total of S nodes act as sources for traffic flows. Each source s generates a constant bit rate flow of volume h_s, measured in bits per second. From each source, a total of P_s paths lead to one of the Internet gateways. To indicate if a link e lies on a certain path p from source s to a gateway, we introduce the binary variable $\delta_{e,s,p} \in \{0,1\}$, which is equal to one if link e lies on path p of source s and equal to zero otherwise.

To formulate the gateway selection problem, we introduce the binary integer variables $u_{s,p} \in \{0,1\}$, which are equal to one if path p is used by source s and zero otherwise. To assure that only one gateway is selected for every demand, we introduce the constraint

$$\sum_{p=1}^{P_s} u_{s,p} = 1 \quad \forall s. \tag{5}$$

Using this notation, the total traffic that is transmitted over link e, denoted as y_e, can be written as

$$y_e = \sum_{s=1}^{S} \sum_{p=1}^{P_s} \delta_{e,s,p} h_s u_{s,p}. \tag{6}$$

Then, the traffic load that is handled by node n is $\sum_{e \in \mathcal{E}_n} y_e$, where $\mathcal{E}_n$ is the set of all links arriving at or leaving node n. Traffic that is forwarded by a node contributes twice to the node's load, since it arrives on one link and is retransmitted on another link.

IV. FORMULATION OF THE GATEWAY SELECTION PROBLEM

In selecting Internet gateways, our optimization goal is to minimize the maximum node utilization within the network, i.e.

$$\min \max \frac{1}{R_n} \sum_{e \in \mathcal{E}_n} y_e, \tag{7}$$

where R_n is the nominal data rate of node n. This will result in a better spatial distribution of traffic within the network, thereby avoiding ”hot spots” prone to congestion or excessive interference, though possibly at the expense of longer paths through the network.

Summarizing the above constraints, the gateway selection problem can be formulated as the Mixed Integer Linear Program (MILP)

$$\min_{\{\lambda_i\},\{u_{s,p}\}} \max_n \frac{1}{R_n} \sum_{e \in \mathcal{E}_n} y_e, \tag{8}$$

$$\text{s.t.} \quad y_e \leq r_e \sum_{i \in \mathcal{I}_e} \lambda_i \quad \forall e, \tag{9}$$

$$\sum_{p=1}^{P_s} u_{s,p} = 1 \quad \forall s, \text{ and} \tag{10}$$

$$\sum_i \lambda_i \leq 1, \tag{11}$$

where the minimization is performed with respect to the allocation of nodes to gateways $u_{s,p}$ and the scheduling of independent sets λ_i. Constraint (9) assures that no link is overloaded, constraint (10) assures that each node only uses one gateway, and (11) assures that the total utilization of the wireless channel is not greater than one. Note that the scheduling of the independent sets and the definition of the link load according to 6) also implicitly guarantee that no node is overloaded. In the following, this gateway selection scheme

number of hosts	50
size of playground	1000 m × 1000 m
number of GWs	2
max. transmission range	250 m
max. interference range	330 m
data rate of MANET nodes	{ 11 Mbps, 2 Mbps}
data rate of GW 1	11 Mbps
data rate of GW 2	5.5 Mbps
packet size	1024 byte
packet rate	40 ms^{-1}
GWADV size	32 byte
GWADV rate	1 s

TABLE I
PARAMETERS USED FOR SIMULATIONS

will be referred to as the MMNU (MinMax Node Utilization) solution.

Although the class of Mixed Integer Linear Programs is known to be NP-complete [14], small to medium sized problems can still be solved efficiently. Relaxing the problem by allowing the $u_{s,p}$ to take on continuous values between 0 and 1 would allow a node to split its traffic arbitrarily between multiple paths to potentially multiple IGWs, thereby opening a further possibility for load balancing between gateways. However, this might result in packets belonging to the same flow experiencing significantly different delay, which is generally not desirable from the end to end point of view.

Finding all maximal independent sets of a graph is also an NP-complete problem. However, it is also possible to approximate the problem by calculating only a subset of the conflict graph's maximal independent sets. In this case, the solution to the optimization above is still a feasible solution to the original problem *with* all maximal independent sets, since the exclusion of a part of the maximal independent sets reduces the effective capacity of the network. The solution found in this case would be a lower bound of the true maximum performance.

The framework used here to model the wireless network is based on a contention based MAC such as IEEE802.11. However, it can readily be applied to other systems using TDMA or FDMA instead. Our search for maximal independent sets in the conflict graph, representing links that can be scheduled simultaneously without causing mutual interference, can also be adapted to the problem of assigning time slots or frequency channels to the hosts.

V. SIMULATION RESULTS

We have analyzed the performance of the MMNU gateway selection according to (8) as well as the MLI scheme according to [6] and simple hop count based selection by means of simulations. All simulations were carried out using the OMNeT++ discrete event simulator [15] and its INET Framework for the simulation of wireless communication networks. To solve the optimization problems, we used the GNU Linear Programming Toolkit (GLPK) [16]. The nodes access the wireless channel using the INET implementation of IEEE 802.11.

For our simulations, we consider a rectangular area in which static nodes are randomly distributed. Two Internet gateways are placed in diagonally opposite corners. The parameters used in our simulations are given in Table I. The number of nodes generating traffic is kept variable, with each source producing a packet of size 1024 bytes every 40 ms, corresponding to a data rate of ca. 200 kbps. The node density is sufficiently high that all sources typically have a path to both gateways. One gateway provides a data rate of 11 Mbps, whereas the other only provides 5.5 Mbps.

We do not make use of a full-fledged routing protocol, since we only require packets to be sent from some of the nodes to the gateways. Instead, we perform shortest path routing by making use of the gateway advertisements that are flooded through the network. Every node simply tracks which of its neighbors is closest to the gateway, based on the hop count of the advertisements, and uses this information for forwarding packets to the gateway. Due to the static topology and the fact that communication only takes place between the gateways and the wireless nodes, this approach is sufficient in our scenario.

We first consider the case where the MANET nodes are interconnected by an 11 Mbps link. In this case, the gateways will pose a bottleneck for the data traffic. The Packet Delivery Ratio (PDR), average delay, and LBI for each of the three gateway selection schemes considered are presented in Fig. 2.

As can be seen from the LBI of 0.5, the hop count scheme divides the traffic evenly between both gateways. This behavior can be expected due to the symmetric placement of the gateways. This leads to congestion at the weaker gateway, resulting in both significantly higher packet loss and delay than the MLI and MMNU solutions. MLI and MMNU achieve very similar PDR, but the delay of the MMNU scheme is much lower. The sharp drops in the LBI of the MMNU solution for three and six sources can be attributed to the fact that the MMNU is able to allocate exactly twice as many flows to the gateway with twice the capacity, leading to an LBI very close to zero. In general, the LBI decreases as the number of sources increases due to the higher flexibility in the allocation of a larger number of flows to the gateways and the smaller effect of the granularity of the flows.

In Fig. 3, the capacities of the gateways are left unchanged, but the nodes within the MANET are able to communicate at only 2 Mbps. All three schemes suffer from this drop in capacity. However, the MANET itself now poses a bottleneck, instead of the gateways as before. The MLI and hop count schemes are both agnostic of this change and assign the traffic flows quite similarly as before. This leads to a significant drop in the PDR of the MLI solution, whereas the hop count solution comes close to the PDR of MNU. MMNU achieves approximately half the delay of both the MLI and hop count schemes.

One weakness of a pure load based approach becomes evident if many packets are already dropped due to congestion in the network before they reach the gateway. In this case, the load measured by the gateway will be reduced, leading to the advertisement of a lower load, in turn attracting even more

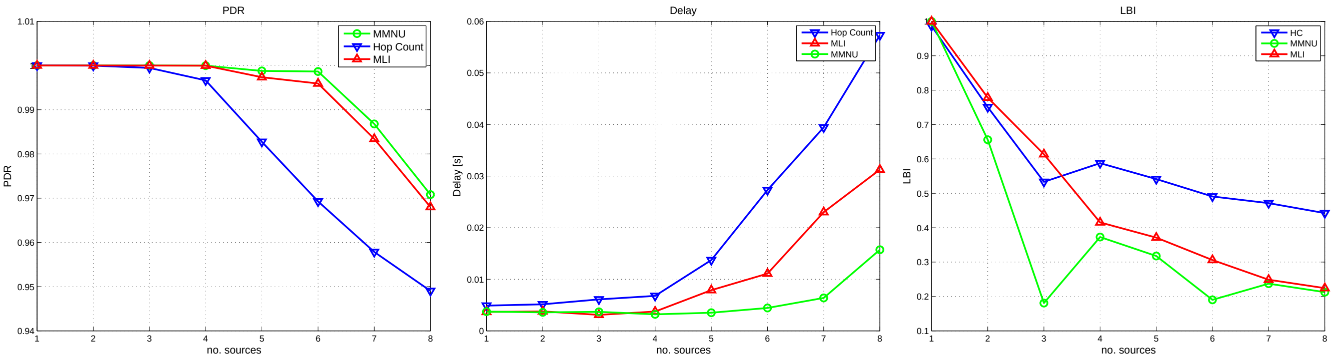

Fig. 2. PDR, delay, and LBI in the case of imbalanced gateways with $R_1 = 11$, $R_2 = 5.5$ Mbps, 11 Mbps within the MANET

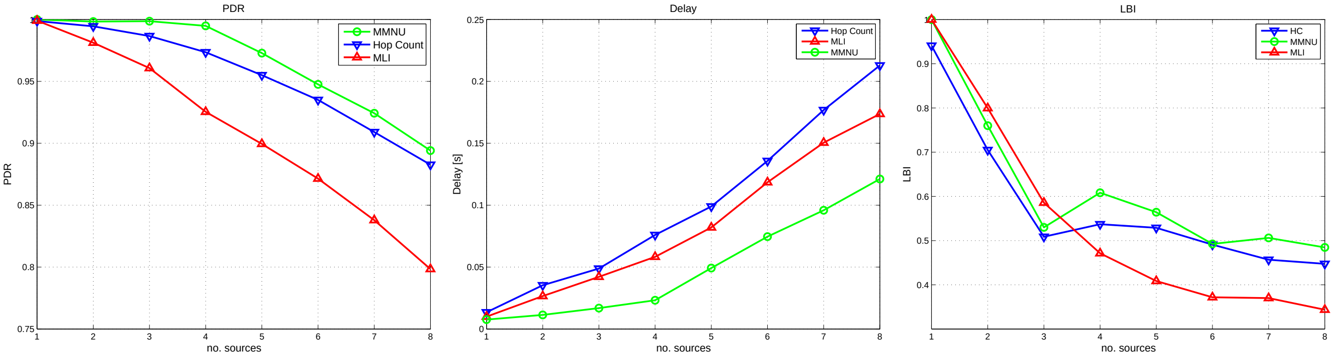

Fig. 3. PDR, delay, and LBI in the case of imbalanced gateways with $R_1 = 11$, $R_2 = 5.5$ Mbps, 2 Mbps within the MANET

traffic that will be lost due to congestion. This effect may cause a slightly congested network to collapse entirely.

VI. CONCLUSION

These results show that common Internet gateway selection algorithms are not flexible enough to adapt to different situations in a MANET, depending on the traffic demands and the capacity that is provided by the MANET and the gateways.

By formulating the gateway selection problem as a linear program, we show that it is possible to increase network performance significantly, regardless whether the MANET or the gateways are limiting the performance.

Since solving this optimization problem is obviously not possible in a real MANET, further work will focus on finding a distributed solution that is better suited to the heterogeneous environment than the heuristic solutions proposed so far, taking into consideration the situation of both the gateways and the path between the mobile node and the gateway. Additional performance improvements may be gained by considering the problem of joint gateway selection and routing.

REFERENCES

[1] M. Schnell and S. Scalise, “NEWSKY - NEtWorking the SKY Concept for Civil Aviation,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 22, no. 5, 2007.

[2] IETF MANET WG, Mobile Ad Hoc Networks (MANET) Charter. Online. http://www.ietf.org/html.charters/manet-charter.html.

[3] IETF AUTOCONF WG, Ad-Hoc Network Autoconfiguration (AUTOCONF) Charter. Online. http://www.ietf.org/html.charters/autoconf-charter.html.

[4] P. Ratanchandani and R. Kravets, “A Hybrid Approach to Internet Connectivity for Mobile Ad Hoc Networks,” in *Proc. IEEE WCNC 2003*, March 2003, new Orleans, USA.

[5] Y. Sun, E. M. Belding-Royer, and C. E. Perkins, “Internet Connectivity for Ad Hoc Mobile Networks,” *International Journal of Wireless Information Networks*, vol. 9, no. 2, April 2002.

[6] C. Huang, H. Lee, and Y. Tseng, “A Two-Tier Heterogeneous Mobile Ad hoc Network Architecture and Its Load Balance Routing Problem,” *ACM Mobile Networks and Applications*, vol. 9, no. 4, pp. 379–391.

[7] S. Ahn, Y. Kim, Y. Lim, and J. Lee, “Load Balancing in MANET with Multiple Internet Gateways, draft-ahn-manet-multigateway-00,” IETF Internet Draft, work in progress, October 2005.

[8] Y. Kim, S. Ahn, and J. Lee, “Load-Balancing Proactive Internet Gateway Selection in MANET, draft-kim-autoconf-gatewaysel-01,” IETF Internet Draft, work in progress, February 2007.

[9] K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, “Impact of Interference on Multi-hop Wireless Network Performance,” in *Proc. ACM MobiCom 2003*, September 2003.

[10] L. Chen, S. Low, M. Chiang, and J. Doyle, “Cross-layer Congestion Control, Routing and Scheduling Design in Ad Hoc Wireless Networks,” in *Proc. IEEE Infocom 2006*, April 2006, barcelona.

[11] D. O’Neill and Y. Li, “Optimal Routes and Flows in Multicasting Over Ad Hoc Networks,” in *Proc. IEEE VTC 2004 Spring*, May 2004, milan.

[12] R. Brännström, C. Ahlund, and A. Zaslavsky, “Maintaining Gateway Connectivity in Multi-hop Ad hoc Networks,” in *Proc. IEEE WLN 2005*, November 2005, tampa, USA.

[13] P. Gupta and P. R. Kumar, “The Capacity of Wireless Networks,” *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[14] M. R. Garey and D. S. Johnson, *Computers and Intractibility*. W. H. Freeman and Co., 1979.

[15] OMNeT++ Discrete Event Simulation System. Online. Available online at http://www.omnetpp.org.

[16] “GNU Linear Programming Toolkit,” Online, available online at http://www.gnu.org/software/glpk/.