Ilmari Juva. 2007. Robust load balancing in wireless networks. In Proceedings of the EuroFGI Workshop on IP QoS and Traffic Control, Lisbon, Portugal, 6-7 December 2007, pages 127-134.

EuroFGI Workshop on IP QoS and Traffic Control 2007
P. Pereira (Eds.)
Lisbon, Portugal, December 6–7, 2007

ROBUST LOAD BALANCING IN WIRELESS NETWORKS

Ilmari Juva

TKK, Helsinki University of Technology
e-mail: ilmari.juva@tkk.fi

Keywords: Robust Load Balancing, Wireless Mesh Networks.

Abstract.
In this paper we study load balancing in a wireless network. We present a robust approach that enables cross-layer optimization of both transmission schedule and routing without the knowledge of traffic matrix. This approach is shown to clearly outperform an approach where only the MAC layer is optimized. Further, we show that even if a traffic matrix estimate is available, it would have to be rather accurate to warrant its use, as optimization done with an inaccurate traffic matrix yields typically worse results than the robust approach.

1 INTRODUCTION

Wireless mesh networks (WMN) [1] are an architecture of wireless multihop networks, consisting of clients and fixed wireless routers which form a mesh network. The network supports multihop routing, enabling cost effective communications between the clients. Typical applications include wireless access networks and neighborhood networks.

Optimization of resource usage in wireless networks has gained interest recently, from throughput maximization [12, 5] to delay minimizations [3] and dimensioning [7]. In this paper our goal is to find a routing and a MAC layer schedule that minimize the maximum link utilization in the network. This problem is studied in [11]. However, that approach, as overwhelming majority of load balancing methods do, requires knowledge of the traffic matrix. But, in reality, the traffic matrix is typically not readily available. Instead, to use these approaches, a possibly inaccurate estimate of the traffic matrix has to be used.

We propose a robust approach which does not require knowledge of the traffic matrix, but only link count measurements and routing information which are typically more easily obtainable. The idea behind the method is to balance the load for the worst case traffic matrix included in a larger polytope of matrices, not just for a single traffic matrix.

This kind of robust approach to traffic engineering has gained interest in recent years in IP networks. Our work in [6] is based on the concept of routing a polytope by Ben-Ameur and Kerivin [2], which Johansson [4] proposed to use for load balancing without a traffic matrix estimate. That approach is not directly applicable to wireless framework, but we draw on these concepts to develop the robust approach for wireless networks.

The rest of the paper is organized as follows. Section 2 introduces the framework. In Section 3.1 the traditional load balancing method for wireless framework is reviewed, while sections 3.2.2 and 3.2.3 formulate the robust approaches. The latter introduces the main contribution of the paper, the cross-layer robust approach. Section 4 presents a simulation study comparing the different methods, and section 5 concludes the paper.

2 MODELING THE WIRELESS NETWORK

The MAC layer is modeled by Spatial TDMA [9], where the transmission resources are divided into time slots. The links that are not interfering each other can transmit in the same time slot. A set of links that are able to transmit simultaneously is called a transmission mode.

The interference model assumes that interference restricts the links that can transmit simultaneously, but does not affect capacity of those that are able to transmit. The communication range of node i is denoted by R_i and the distance between nodes i and j by d_{ij}. Now, if

$$d_{ij} \leq R_i$$

there is a link between nodes i and j. The interference range is denoted by R'_i. For transmission to be successful on link ij, there cannot be a node k transmitting such that

$$d_{kj} \leq R'_k.$$

Based on this criterion we can determine all possible transmission modes. That is, the sets of links that can transmit simultaneously. Finding all feasible modes requires extensive calculations, but maximal transmission modes can be generated by the algorithm proposed in [11]. The M different transmission modes are represented in the $L \times M$ matrix $\boldsymbol{S}$, which has an element $S_{l,m} = b_l$ if link l transmits in mode m and $S_{l,m} = 0$ otherwise.

Let $\boldsymbol{b}$ be a vector of length L denoting the nominal link bandwidths. The elements of the vector are the link bandwidths for each link, with the links indexed by l. The M-vector $\boldsymbol{t}$, whose elements sum to unity, denotes the proportional time that is spent in each transmission mode. The actual capacities $\boldsymbol{C}$ of the links are thus

$$\boldsymbol{C} = \boldsymbol{St}. \tag{1}$$

Further, we define vector $\boldsymbol{q}$ such that

$$q_m = ut_m, \tag{2}$$

where the elements of $\boldsymbol{q}$ sum to u, which represents the percentage of time that is spent in any transmission mode, with the rest of the time the network being idle.

We denote the link loads by $\boldsymbol{y}_0$, which is a vector of length L. The link - OD pair incidence matrix, or the routing matrix, is denoted by $L \times K$ matrix $\boldsymbol{A}$, where K is the number of OD pairs in the network, indexed by k. The element $A_{l,k}$ of the routing matrix gives the proportion of the traffic of the kth OD pair that is routed through link l. If the OD pair does not use the link in question, then $A_{l,k} = 0$. We consider multipath routing where fractions of traffic for the same OD pair may use different routes.

The routing matrix during the link count measurements is assumed to be known and is denoted by $\boldsymbol{A}_0$. The traffic matrix $\boldsymbol{x}$ is a vector of length K which is composed of traffic flows x_k between origin-destination pairs and satisfies the link count relation

$$\boldsymbol{A}_0\boldsymbol{x} = \boldsymbol{y}_0. \tag{3}$$

Finally, let $\boldsymbol{L}$ denote the node-link incidence matrix with element $L_{nl} = +1$ and $L_{n'l} = -1$ if (directed) link l leads from node n to node n', and 0 otherwise. And let $\boldsymbol{R}$ denote the node - OD pair incidence matrix with element $R_{n_o,k} = +1$ and $R_{n_d,k} = -1$ if OD pair k enters the network at node n_o and exits at node n_d, and 0 otherwise.

3 LOAD BALANCING

The goal is to find such a transmission schedule and routing that the traffic load is as evenly balanced over the network as possible. Our performance metric u refers to the relative utilization of the most heavily congested link,

$$u = \max_l \frac{(\boldsymbol{Ax})_l}{C_l} = |\boldsymbol{Ax}/\boldsymbol{St}|, \tag{4}$$

where in the latter form we have introduced the notation $|\boldsymbol{a}/\boldsymbol{b}| = \max_l a_l/b_l$. In the starting situation the utilization is

$$u_{initial} = |\boldsymbol{y}_0/\boldsymbol{C}|, \tag{5}$$

When the transmission schedule is also optimized we can write $u = \boldsymbol{e}^T\boldsymbol{q}$.

The following sections formulate the different approaches to achieve load balancing in the wireless network.

3.1 Load Balancing with Estimated Traffic Matrix

In traditional load balancing approach, the traffic matrix is assumed to be known. In reality, an estimate is needed for the traffic matrix. There are several estimation techniques proposed in literature (see e.g. [8] and references therein).

With the estimated traffic matrix, denoted by $\hat{\boldsymbol{x}}$, we can formulate the load balancing problem as an LP problem [11]

Problem 1 (Traditional Load Balancing Problem)

$$\min_{\boldsymbol{A}\geq\boldsymbol{0}\ ,\ \boldsymbol{q}\geq\boldsymbol{0}} \boldsymbol{e}^T\boldsymbol{q} \tag{6}$$

such that

$$\boldsymbol{S}\,\boldsymbol{q} \geq \boldsymbol{A}\hat{\boldsymbol{x}}, \tag{7}$$

$$\boldsymbol{L}\boldsymbol{A} = \boldsymbol{R}, \tag{8}$$

where (7) is the link capacity constraints. The left hand side term gives the allocated bandwidth with the given transmission schedule. It follows from (1) and (2) that $\boldsymbol{S}\,\boldsymbol{q} = \boldsymbol{S}\,u\,\boldsymbol{t} = u\,\boldsymbol{C}$. This is required to be more than the right hand side term which gives the bandwidth usage for given routing $\boldsymbol{A}$ and traffic matrix $\hat{\boldsymbol{x}}$.

Equation (8) is the flow conservation condition. $\boldsymbol{L}$ is the node-link incidence matrix, telling which links begin/end to any given node. An element of $\boldsymbol{LA}$ tells for a given OD pair the difference of its traffic coming into the node and leaving the node. The condition thus requires routing matrix $\boldsymbol{A}$ to be selected so that the elements coincides with corresponding elements of the node - OD pair incidence matrix $\boldsymbol{R}$, which are zero unless the OD pair in question enters/exits the network through that node.

This solution is optimal only with regard to the maximum link load, and usually is not unique. A second optimization is needed to ensure that traffic is optimally balanced in less loaded links also, not just in the bottleneck link [11, 6]. The resulting link loads are optimal if the traffic matrix used in the optimization is accurate. This is used as a reference value in our simulation study is section 4.

In reality, as estimated traffic matrices are always at least somewhat erroneous, optimally balanced load is not achieved. How far off the optimum the result is, depends on the accuracy of the estimate.

3.2 Robust Load Balancing

In the Robust method, instead of using a fixed traffic matrix estimate, we try to find a routing matrix such that the worst case performance is optimized over all feasible traffic matrices. By finding a transmission schedule and routing matrix which perform well over this whole set of traffic matrices, we avoid having to estimate the traffic matrix.

Each non-negative vector $\boldsymbol{x}$ that satisfies relation (3) is a feasible traffic matrix. Denote the polytope of such traffic matrices by $\mathcal{D}$.

$$\mathcal{D} = \{\boldsymbol{x} \geq 0 : \boldsymbol{A}_0\boldsymbol{x} = \boldsymbol{y}_0\}. \tag{9}$$

3.2.1 MAC Layer Optimization

Optimizing only the MAC layer transmission schedule based on the link counts is a straightforward approach to achieve load balancing in the network without the knowledge of the traffic matrix. In essence, the link capacities are changed to fit the traffic volumes on the links by changing the scheduling. We denote the current routing matrix, for instance the shortest path routing, by $\boldsymbol{A}_0$, and the corresponding link counts are denoted by $\boldsymbol{y}_0$. The optimization problem is

Problem 2 (MAC Layer Optimization)

$$\min_{\boldsymbol{q}\geq\boldsymbol{0}} \boldsymbol{e}^T\boldsymbol{q} \tag{10}$$

such that

$$\boldsymbol{S}\,\boldsymbol{q} \quad \geq \quad \boldsymbol{y}_0. \tag{11}$$

The result gives the optimal schedule for the fixed routing which yielded the link counts $\boldsymbol{y}_0$.

3.2.2 Optimizing the Layers Separately

One load balancing approach would be to first balance the load using the nominal link capacities, and then optimize the transmission schedule. However, as we assume that the traffic matrix is not available, we cannot use the traditional load balancing approach. Instead, we have to use the robust approach [6]. Another drawback is that we would then need to make new measurements to obtain the link counts associated with the new routing matrix in order to optimize the transmission schedule.

The idea behind the robust method is to consider, instead of a single traffic matrix, a polytope of all the traffic matrices that are consistent with the link load measurements. The routing is then optimized for the worst case traffic matrix among the polytope. The problem is

Problem 3 (Robust Load Balancing Problem)

$$\min_{\boldsymbol{A}\geq\boldsymbol{0}} u \tag{12}$$

such that

$$u\,\boldsymbol{b} \quad \geq \quad \boldsymbol{A}\boldsymbol{x}, \qquad \forall \boldsymbol{x} \in \mathcal{D}, \tag{13}$$

$$\boldsymbol{L}\boldsymbol{A} \quad = \quad \boldsymbol{R}, \tag{14}$$

where $\boldsymbol{b}$ is the nominal link capacity vector, and our performance metric u refers to the relative utilization of the most heavily congested link,

$$u = \max_l \frac{(\boldsymbol{A}\boldsymbol{x})_l}{b_l}. \tag{15}$$

This problem is difficult to solve because of the infinite number of constraints in (13). Therefore, in [6] we divide it into two optimization problems following the approach of [2].The Constraint generation problem generates and updates a finite set of constraints D^*, while the Link Load Optimization is a simple LP problem that finds the optimal values to minimize link utilization such that the constraints in D^* hold.

Problem 4 (Link Load Optimization)

$$\min_{\boldsymbol{A}\geq\boldsymbol{0}} u \tag{16}$$

such that

$$u\,\boldsymbol{b}_l \quad \geq \quad (\boldsymbol{A}\boldsymbol{x})_l \qquad \forall (l, \boldsymbol{x}) \in D^* \tag{17}$$

$$\boldsymbol{L}\boldsymbol{A} \quad = \quad \boldsymbol{R}, \tag{18}$$

and such that a secondary objective function is used to obtain optimal balancing throughout the network.

Problem 5 (Constraint Generation) *For each link l solve with current values $\boldsymbol{A}^{(i)}$, $u^{(i)}$*

$$\boldsymbol{x}^* = \arg\max_{x\in\mathcal{D}} (\boldsymbol{A}^{(i)}\boldsymbol{x})_l$$

If

$$u^{(i)}\,\boldsymbol{b}_l < (\boldsymbol{A}^{(i)}\boldsymbol{x}^*)_l$$

then

$$D^* \leftarrow D^* \cup (l, \boldsymbol{x}^*).$$

The set D^* is initially empty. Problem 5 is solved to obtain constraints. The iteration can be started using the initial routing $\boldsymbol{A}_0$. For each link we find the traffic matrix $\boldsymbol{x}^* \in \mathcal{D}$ that would maximize the traffic on that link. If this link utilization is larger than the current value for u, the corresponding constraint

$$u\,\boldsymbol{b}_l \geq (\boldsymbol{A}\boldsymbol{x}^*)_l$$

is added to the set D^* to be used as a constraint in (17) in the next iteration of problem 4. As the constraint can be identified by the pair $(l, \boldsymbol{x}^*)$ we denote

$$D^* \leftarrow D^* \cup (l, \boldsymbol{x}^*).$$

The algorithm then iterates between these two problems until no new constraints are found. The transmission schedule is then optimized as in problem 2 using the routing obtained by the robust approach and the new link loads.

3.2.3 Cross-Layer Robust Load Balancing

To optimize the layers at the same time we have to modify the approach of problem 1 to wireless framework by incorporating the transmission schedule variables $\boldsymbol{q}$ to the optimization problem. The minimization problem is now

Problem 6 (Cross-layer Robust Load Balancing)

$$\min_{\boldsymbol{A}\geq\boldsymbol{0}\ ,\ \boldsymbol{q}\geq\boldsymbol{0}} \boldsymbol{e}^T\boldsymbol{q} \tag{19}$$

such that

$$\boldsymbol{S}\,\boldsymbol{q} \geq \boldsymbol{A}\boldsymbol{x} \qquad \forall \boldsymbol{x} \in \mathcal{D}, \tag{20}$$

$$\boldsymbol{L}\boldsymbol{A} = \boldsymbol{R}. \tag{21}$$

Again, we use the iterative approach shown above for the case of constant link bandwidths. The Link Load Optimization problem and the constraint generation problem for the cross-layer approach can be written as

Problem 7 (Cross-layer Link Load Optimization)

$$\min_{\boldsymbol{A}\geq\boldsymbol{0}\ ,\ \boldsymbol{q}\geq\boldsymbol{0}} \boldsymbol{e}^T\boldsymbol{q} \tag{22}$$

such that

$$(\boldsymbol{S}\boldsymbol{q})_l \geq (\boldsymbol{A}\boldsymbol{x})_l, \qquad \forall (l, \boldsymbol{x}) \in D^*, \tag{23}$$

$$\boldsymbol{L}\boldsymbol{A} = \boldsymbol{R}, \tag{24}$$

and such that a secondary objective function is used to obtain optimal balancing throughout the network.

Problem 8 (Cross-layer Constraint Generation) *For each link l solve with current values $\boldsymbol{A}^{(i)}$, $\boldsymbol{q}^{(i)}$*

$$\boldsymbol{x}^* = \arg\max_{x \in \mathcal{D}} (\boldsymbol{A}^{(i)} \boldsymbol{x})_l$$

If

$$(\boldsymbol{S}\boldsymbol{q}^{(i)})_l < (\boldsymbol{A}^{(i)} \boldsymbol{x}^*)_l \tag{25}$$

then

$$D^* \leftarrow D^* \cup (l, \boldsymbol{x}^*).$$

The iteration is then performed as before to yield the results for the transmission schedule and the routing matrix. While there are more variables in this formulation, the problem is not significantly harder and the running time of the algorithm remains about the same.

4 SIMULATION STUDY

4.1 Simulation Setup

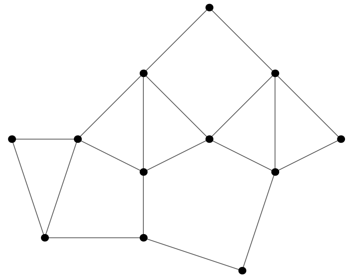
To validate the above approach we performed a simulation study with synthetic traffic, using the topology depicted on the left hand side of Figure 1. The network in question has 12 nodes, each having the same transmission range R. Also, we set

$$R_i = R'_i = R, \qquad \forall i.$$

That is, the range is the same for each node and the transmission range has the same value as the interference range.

The methods considered are: 1. Optimization based on a traffic matrix estimate, 2. MAC layer only optimization with shortest path routing, 3. Two-layer Robust method and 4. Cross-layer Robust method. The first method optimizes both routing and the schedule. If an accurate traffic matrix were available this approach would yield optimal results. However, the traffic matrix is rarely known. Instead, estimates have to be used and they come with estimation error. We study the performance of this method with traffic matrices of various accuracies with mean relative errors from zero to 0.30. The second approach optimizes only the transmission schedule while routing remains fixed. It does not thus require knowledge of the traffic matrix, but only the link load measurements. The two-layer robust approach optimizes the layers separately, first routing and then transmission schedule after new link counts are obtained. The cross-layer Robust method optimizes simultaneously the schedule and the routing.

We generate synthetic traffic such that the traffic volumes of the OD pairs are drawn randomly following the gravity model. For each node we first draw a *mass* coefficient from uniform distribution between 1 and 3. The OD pair traffic volume is then taken to be proportional to



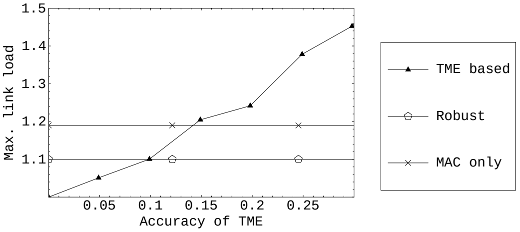


Figure 1: Test topology on the left, performance of traffic matrix optimization approaches on the right.

the product of the coefficients of the origin node and destination node. This yields a traffic distribution where the largest traffic volume is slightly less than ten times the smallest one.

As the methods using estimated traffic matrices are dependent on the accuracy of that estimate, we inserted error to the estimates by adding a random noise term to the OD pair traffic volume. Various size noise terms lead to different size errors in estimates, where by error we mean the mean relative error between the estimated and real OD pair volumes. We studied seven cases with gradually worsening estimation accuracies. For each of the synthetic datasets we then evaluate the link utilization achieved by each of the considered methods. The simulation results, shown in Table 1, are taken as the average values over one hundred simulation rounds.

4.2 Simulation Results

Table 1 shows the results of the methods using the robust optimization approach instead of a traffic matrix estimate. The results are normalized such that 1.00 is the theoretical optimal, which is the value obtained by the cross-layer optimization of section 3.1 using the accurate traffic matrix. The initial utilization is given by using the shortest path routing and a transmission schedule that assigns equal time for each transmission mode.

Table 1: Values of u for different methods, with 1.00 being the optimal solution

Initial	3.17
MAC layer only (shortest path routing)	1.19
Two-layer robust approach	1.18
Cross-layer robust approach	1.10

It can be seen that the cross-layer approach performs clearly best of the robust methods. Although it has to optimize the routing for all traffic matrices in the polytope, there is still enough information about the traffic matrix so that it is beneficial to optimize routing also instead of just the transmission schedule. The two-layer approach is only marginally more accurate than the MAC layer only optimization and thus probably not useful.

If a traffic matrix estimate (TME) is available, the accuracy of an optimization based on this estimate would obviously depend on the accuracy of the estimate. Therefore we evaluated the performance of this method using different estimates. Figure 1 shows how the performance of this approach deteriorates compared to the cross-layer Robust and MAC layer only approaches as the accuracy of the TME gets worse. It can be seen that the approach using estimates quickly becomes worse than the robust approaches. With an estimation error larger than 10% it is already worse than the cross-layer robust approach. That size of error is considered typically rather small in traffic matrix estimation. In fact, in IP networks this is often considered the target value which estimation methods strive for. Only third generation methods that use extensive Netflow measurements achieve typically errors from 5% to 15%, while traditional methods relying on SNMP measurements cannot usually do better than 20% [10].

In light of these results it is highly questionable whether it is worth the trouble to try to estimate the traffic matrix in this situation, as in real networks there is no way of knowing the actual accuracy of the given estimate, and most likely the robust approach outperforms the method using the estimate.

5 CONCLUSION

In this paper we proposed a novel approach for the network load balancing task in wireless networks. Instead of relying on inaccurate traffic matrices, we used a robust approach which uses only link count measurements as input. From the link counts a polytope of plausible traffic matrices is formed and the routing is done with regard to the worst case traffic matrix within the polytope. This robust routing method was used as a part of a two-layer optimization approach.

We then proposed a modification to extend the robust approach to cross-layer optimization. This was achieved by relaxing the link capacities to be free variables that depend on the transmission schedule and adding the vector defining the schedule to the decision variables. In this novel cross-layer robust approach the transmission schedule and routing are optimized simultaneously.

We studied the performance of the methods by a simulation study with synthetic traffic. We found out that in the two-layer sequential optimization approach, the first part of routing optimization is of very little use, as almost as good results where obtained by performing only the MAC layer optimization with shortest path routing. The cross-layer robust method was shown to be the best of the studied methods. It yielded maximal link utilization of only 10% worse than optimal and outperformed the other robust approaches. Further, it was shown that even the traditional load balancing methods using an estimated traffic matrix typically performed worse than the cross-layer robust method. Thus, we believe that in a realistic network setting, the proposed cross-layer robust approach would yield the lowest link loads.

ACKNOWLEDGMENTS

The author would like to thank Prof. Jorma Virtamo, Ph.D. Samuli Aalto and Lic.Sc.(Tech) Riikka Susitaival from TKK for their help. The work for this paper was carried out as a part of project Fancy supported by the Academy of Finland grant 210275.

REFERENCES

[1] I. Akyildiz and X. Wang. A survey on wireless mesh networks. *IEEE Radio Communications*, september 2005.

[2] W. Ben-Ameur and H. Kerivin. Routing of uncertain traffic demands. *Optimization and Engineering*, September 2005.

[3] C. Chen, W. Wu, and Z. Li. Multipath routing modelling in ad hoc networks. In *ICC*, Seoul, Korea, 2005.

[4] M. Johansson. Data driven traffic engineering. Presentation in LSNI Workshop, 2005.

[5] M. Johansson and L. Xiao. Cross-layer optimization of wireless networks using nonlinear column generation. *IEEE Transactions on Wireless Communications*, february 2006.

[6] I. Juva. Robust load balancing. In *Globecom 2007*, Washington D.C., United States, 2007.

[7] P. Lassila, A. Penttinen, and J. Virtamo. Dimensioning of wireless mesh networks. In *ACM PE-WASUN'06*, Malaga, Spain, 2006.

[8] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot. Traffic matrix estimation: Existing techniques and new solutions. In *SIGCOMM'02*, Pittsburg, USA, 2002.

[9] R. Nelson and L. Kleinrock. Spatial-tdma: A collision-free multihop channel access control. *IEEE Transactions on Communications*, 33, 1985.

[10] A. Soule, A. Lakhina, N. Taft, K. Papagiannaki, K. Salamatian, A. Nucci, M. Crovella, and C. Diot. Traffic matrices: Balancing measurements, inference and modeling. In *SIGMETRICS'05*, Banff, Canada, 2005.

[11] R. Susitaival. Load balancing by joint optimization of routing and scheduling in wireless mesh networks. In *ITC-20*, Ottawa, Canada, 2007.

[12] Y. Wu, P. Chou, Q. Zhang, K. Jain, W. Zhu, and S. Kung. Network plannign in wireless ad hoc networks: A cross-layer approach. *IEEE Journal on Selected Areas in Communications*, 23(1), 2005.