ROUTING STRATEGIES FOR COVERAGE EXTENSION IN HETEROGENEOUS WIRELESS NETWORKS

Marco Miozzo, Michele Rossi, Michele Zorzi *

Abstract

The focus of this paper is on routing over heterogeneous networks. We consider a scenario involving both infrastructure and infrastructureless wireless networks, where a set of mobile users are interested in communicating with several access points (APs). Multi-hop routing, possibly over heterogeneous technologies, is exploited to extend the coverage for those users that are not within the transmission range of any APs. We propose a proactive tree-based approach for the dissemination of routing information and the subsequent data forwarding towards the APs. Subsequently, we compare its performance against reactive routing algorithms for ad hoc networks. The network performance is obtained via simulation through careful modeling of the considered radio interfaces. The results indicate the superiority of proactive schemes under moderate/high traffic conditions and motivate further research.

I. Introduction

In the last few years, advances in the hardware for wireless networking, micro–fabrication and embedded microprocessor technologies have made it possible to manufacture a generation of very small radio equipment in a cost-effective way. This allows to integrate different technologies in the same device. However, multi-technology solutions raise many research issues. One of these is to devise good methods to exploit the available technologies in order to improve performance. In fact, multi-technology solutions are expected to provide benefits from both the operator's and the user's point of view. Specifically, the improvements may be in terms of network capacity, coverage and efficiency in the usage of radio resources. Also, for the users it may be possible to get better performance in terms of mobility support and service availability.

The focus of the present paper is on the following network scenario: we consider a set of fixed access points (APs) and a number of mobile wireless users. Users are interested in transmitting data to one of these access points (uplink communication) in order to gain access to the fixed Internet. Both users and APs may own several different radio access technologies (RATs). For illustration purposes, in this paper we focus on IEEE802.11b [1] and UMTS [2]. These are widespread and currently used access techniques which are already integrated in many commercial products. From a technical standpoint, these RATs have very different capabilities. IEEE802.11b typically offers high bit-rates but the service is best-effort (no QoS guarantees). This is mainly due to the channel access scheme, which is based on CSMA and is therefore affected by collisions and consequent throughput reduction as the number of users increases. In IEEE802.11b data delivery typically occurs through multi-hop communication. On the other hand, UMTS offers lower bit-rates but can guarantee a minimum QoS and a better coverage. The problem to be solved is to allow the mobile users to transmit their data flows to one of the APs. The considered environment is heterogeneous. That is, both APs and users may own either all or a subset of the available RATs. We furthermore consider an environment where not all users are within coverage of an AP. For this reason, multi-hop communication may be exploited for coverage extension.

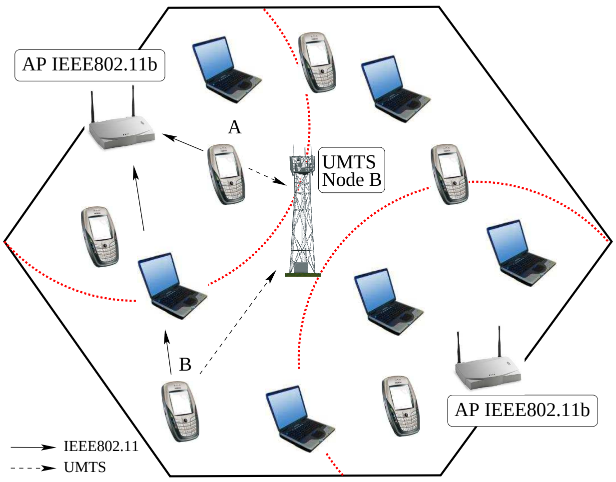

Figure 1: An example of heterogeneous wireless network.

In Fig. 1 we show the considered network scenario. As an example, we assume that a user wants to connect to an Internet service (e.g., web, e-mail). His mobile terminal can be in one of the following three situations: 1) he is within the coverage area of an AP; 2) he is within the coverage area of different APs (possibly having different technologies, see case A in the figure) and 3) the user does not have an AP within range (case B in the figure). In this last case, the user must exploit multi-hop communication to reach an AP. Cases 2 and 3 are the most interesting to our study. In these two cases, a user has to select an access from a candidate set; two issues arise: how the candidate set is discovered and which policy is considered to make routing decisions. In this paper we investigate these two topics.

In the past literature, several papers have been published on these issues. The authors in [3] propose a family of QoS based routing schemes. They consider multiple data flows, with heterogeneous requirements in a possibly multi-technology scenario. Their main focus is on the definition of metrics to be exploited by routing algorithms and on the evaluation of their effectiveness in terms of load balancing. The work is mainly analytical and does not report detailed results considering actual wireless technologies. Reference [4] focuses on a network scenario similar to the one that we consider here. The authors propose a two-tier architecture, where they exploit UMTS base stations (node Bs in the UMTS terminology [2]) and special gateway nodes to access the fixed Internet, whereas they use multi-hop ad hoc routing to deliver data to the gateways. The proposed solutions exploit probabilistic routing, where routing probabilities are calculated based on link costs. The scheme

* All authors are with the Consorzio Ferrara Ricerche (CFR), via Saragat 1, 44100 Ferrara, Italy. M. Rossi and M. Zorzi are also with the University of Padova, Dept. of Engineering (DEI), via Gradenigo 6/B, 35131 Padova, Italy. E-mail: mzzmrc@unife.it, {rossi,zorzi}@dei.unipd.it

1-4244-0330-8/06/$20.00

is effective but some extra overhead is needed to maintain the two-tier structure, which acts as a backbone. In the present paper, we continue the previous work by proposing *online* routing solutions and demonstrating their effectiveness through detailed simulation results. Further, we carefully account for both the PHY and MAC layers of the considered technologies, by implementing them in an event driven simulation tool (see [5] for further details).

The paper is structured as follows. In Section II we discuss possible algorithms to achieve multi-hop routing in our scenario, by emphasizing pros and cons of classical solutions such as the *Dynamic Source Routing* (DSR) [6] protocol. In Section III we introduce our proposal, which we call *Tree Based Routing* (TBR). In Section IV we compare DSR and TBR in the heterogeneous network discussed above by means of simulation. Finally, Section V concludes the paper.

II. Routing in Heterogenous Networks

In the heterogeneous scenario considered in this paper there are several problems to solve to find a route to a given destination (AP). First of all, we must discover a (possibly multi-hop) path to any of the available APs. Second, in the case of multiple available routes, we must select a path that meets the QoS requirements, by cutting the wanted tradeoff between user performance and network capacity. For the coverage extension case (node B in Fig. 1), we may think of exploiting routing algorithms which have been proposed for ad hoc networks. These can find a path to a given destination in environments that change frequently and in an unpredictable manner. However, we should bear in mind that route discovery implies some extra overhead which likely impacts the overall network performance. In general, we should devise solutions that balance the load due to data traffic (routing) and the load due to route discovery (RD). As a reference routing algorithm, in this paper we consider the DSR reactive protocol. A brief description of the algorithm is given next, whereas the reader is referred to [6] for a more precise treatment. In DSR each node starts an RD procedure every time it needs to send data to a destination for which no routing path is known. This involves an initial delay, since routing information has to be acquired before communication can start, as well as an exchange of signaling packets. The RD procedure has to be also initiated when all the discovered paths fail due to changes in the network topology. Hence, in the presence of high network dynamics, the traffic load due to RD may cause a large overhead. Alternatively, we may tackle the routing problem by exploiting proactive solutions. These rely on periodic updates of routing information, which is refreshed even when no data has to be transmitted. The initial delay due to the RD phase is eliminated at the cost of a constant exchange of routing packets. Whether the best option is to exploit proactive or reactive schemes depends on the network dynamics, traffic patterns as well as node mobility. We further observe that once multiple routes to a destination are available at a given node, a further problem arises. In fact, it is necessary to devise good rules to select the path to be actually used. These should account for the delay (which may be estimated according to the number of hops separating the source and the destination), the congestion level of the nodes in the path as well as the node residual energies. In addition, in a multi-technology scenario, further metrics are to be considered. In fact, we may want to route packets towards a non-congested AP or, in general, to an AP which guarantees the desired QoS with a high enough probability. Route selection should be done carefully. For instance, IEEE802.11b APs offer lower QoS guarantees than UMTS APs. However, 802.11 technology may offer higher bit-rates. Also, the considered technologies exploit different physical/MAC techniques which, in turn, translate into different system capacities.

III. Tree Based Routing

In this section, we present a proactive algorithm called Tree Based Routing (TBR), which is based on the DSDV protocol [7]. In particular, TBR accounts for additional features to handle technology diversity and both local (first order neighbors) and global (nodes → APs) flow management according to the imposed QoS requirements. As demonstrated in Section IV the proactive nature of the scheme is justified by its good performance. In fact, in the considered scenario, where the primary interest is to route packets towards an AP, the overhead introduced by reactive solutions (e.g., DSR) may be too large, even when the data traffic is low. In our settings, mobile nodes need to learn routing paths towards the available APs, whose number is usually low. In such a case, it is often more convenient to broadcast some (*common*) routing information to get to the APs rather than having all nodes to independently start RD procedures, as done in [6]. The TBR algorithm works as follows. Each AP periodically sends a message, called *beacon*, including its own identifier (id), a metric indicating its *congestion status*, a sequence number and a *hop count* (HC) number, which is initially set to one. For the UMTS APs, the *congestion status* is estimated by the Radio Network Controller (RNC) [2] as follows. The RNC collects the status of each active user connected to the AP (node B in the UMTS terminology [2]), i.e., by considering its current transmission activity, and calculates the maximum bit-rate that can be allocated for a new data connection to this AP. This calculation follows an estimation procedure which consider the (current) total power level in the cell, the QoS requirement for the new connection and the reaction due to fast power control algorithms after allocating a new channel. This is implemented through standard techniques used in Call admission Control (CaC) solutions [2]. In this paper we consider uplink connections only. The downlink case can be treated similarly and its study is left for future research. For the IEEE802.11b RAT the congestion estimate is derived in a slightly different manner. In particular, the AP multiplies the fraction of time in which it is idle by the maximum effective transmission rate. This rate is calculated by multiplying the maximum physical rate (11 Mbps) by a factor which accounts for the overhead due to the RTS/CTS/ACK handshaking (see the calculations in [8]). As said above, beacons are periodically broadcast by each AP. Such an operation may take several seconds. Upon receiving a beacon from a given radio interface (either IEEE802.11b or UMTS), a node looks for the identifier i of the AP that generated the beacon and for the hop count number carried by the packet $\mathrm{HC}(i)$. Subsequently, it checks in its local routing table whether a beacon from the same AP i with a higher sequence number exists. If so, the beacon is discarded. Otherwise, the node creates a new entry for AP i in its local routing table and the beacon is marked as *valid*. This entry includes the AP identifier i, the congestion status contained in the beacon, its sequence number and its HC number ($\mathrm{HC}(i)$). In addition, if the node owns an IEEE802.11b

RAT, it re-broadcasts over this interface a copy of the beacon by increasing its HC field by one unit. This last operation is performed through the *basic random access mode*. According to this technique, the retransmission of the beacon is scheduled after a time interval which is uniformly distributed in $[0, T]$. T is a function of the average number of nodes within range that own an IEEE802.11b RAT, and is accounted for to reduce the beacons' collision probability. Further, in case a user i receives two or more *valid* beacons from the same AP which carry different HCs, the smallest HC is considered as the actual hop count distance for the node (HC(i)). Moreover, in a wireless network routes may fail due to changes in the topology, which may be due either to user movements or channel conditions (e.g., interference and fading). We therefore account for a refresh mechanism which removes from the routing table all the entries which are considered too old. This allows to track the network dynamics and adapt the broadcast trees to its variations. Finally, in an improved version of the beacon propagation scheme, we additionally account for a threshold based policy which consists of considering only those beacons whose SINR level is above a minimum threshold. The rationale behind this policy is to prune unstable links. This mechanism is referred to here as *link stability* scheme. To sum up, the propagation of beacon messages from each AP allows to create multiple shortest path trees routed at the APs, which can be subsequently used by any node in the network to route packets. In more detail, routing towards the APs is performed as follows. The node first checks among those entries in its routing table whose congestion status can be considered adequate for the reception of the flow in question (i.e., whose congestion level is sufficiently low). Among these selected APs the one with the lowest HC metric is chosen. If two or more APs have the same HC level, the one with the lowest congestion is picked. If i is the identifier of the selected AP and HC$(i) = n$ is the HC associated with this AP at the current node, the next hop is selected among the neighbors having an entry for AP i in their routing tables for which HC$(i) < n$. The selection among these nodes is done according to the state of their queues, i.e., by accounting for the (local) congestion level of the node's neighbors. The path selection policy is a two-step decision making process: step1) the nodes first picks a destination AP according to the APs' congestion level (global metric) and, step2) the next hop towards the selected AP is picked by considering both HCs and local congestion indicators.

IV. Simulation Results

In this section we report accurate simulation results obtained with the multi-technology simulator presented in [5]. The tool has been specifically designed to simulate multi-technology, mobile and wireless communication scenarios. The channel is modeled by accounting for path loss, shadowing and multipath fading and using their product as the link gain which is subsequently associated with each transmission link. Path loss (PL) is implemented according to the Hata model [9]: if P_{tx} is the transmitted power, the PL (in linear units) is calculated as $P_{tx}/P_{rx} = Kd^{\beta}$, where d is the distance separating the two communicating entities, whereas K and β are proper constants. Shadowing is accounted for according to the Gudmunson model and multi-path fading is modeled through the Jakes simulator [9]. The physical modules of the simulated technologies account for multi-user interference and track SINR values

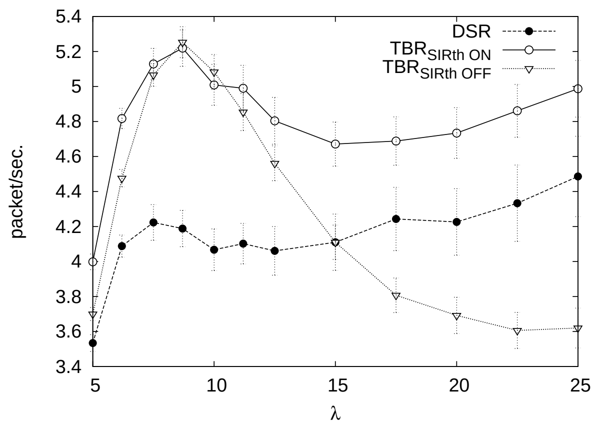

Figure 2: Average throughput as a function of the traffic load.

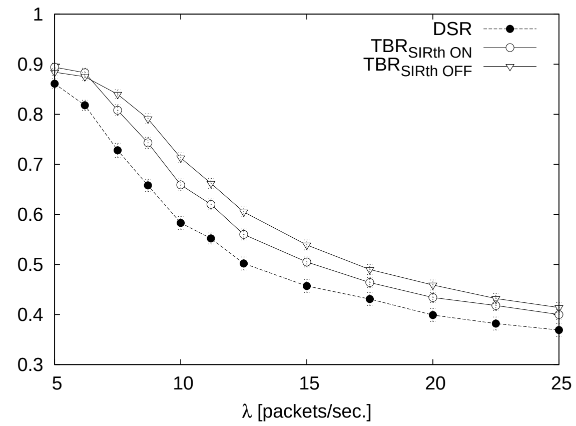

Figure 3: Jain's Fairness index vs. traffic load.

at the symbol level. SINR values are subsequently translated into bit errors according to modulation and coding.

The considered scenario consists of the two radio access technologies introduced in Section I: IEEE802.11b and UMTS. Users are randomly placed at the beginning of the simulation within an area of 240 x 240 m^2. We consider a single UMTS base station, placed in the center of the simulation area (all users are within coverage of such AP). We also consider 4 IEEE802.11b APs randomly placed within the area with the constraint that they must be at least 70 m apart from each other. We account for 35 randomly placed users: 10 of them are equipped with both RATs, while the 25 remaining users are equipped only with IEEE802.11b. This gives an average of 4.6 users in the IEEE802.11b coverage area, which is enough to perform relaying. For the traffic, only 18 nodes (*active users*) generate data traffic, whereas the remaining devices only perform relaying. Active nodes generate data according to a Poisson distribution with intensity λ packets/sec, where all packets have the same size of 512 bytes. All results are obtained averaging over 200 simulations of 120 seconds each, which provide the required statistical confidence. For the UMTS, the spreading factor for each active user is $SF = 16$, which allows for a bit-rate of 240 Kbps. This value is chosen to guarantee a good trade-off between the available rate and the multi-user interference due to simultaneous transmissions. The maximum and minimum uplink transmission powers are set to -16 dBm and -96 dBm, respectively. The step of the fast power control algo-

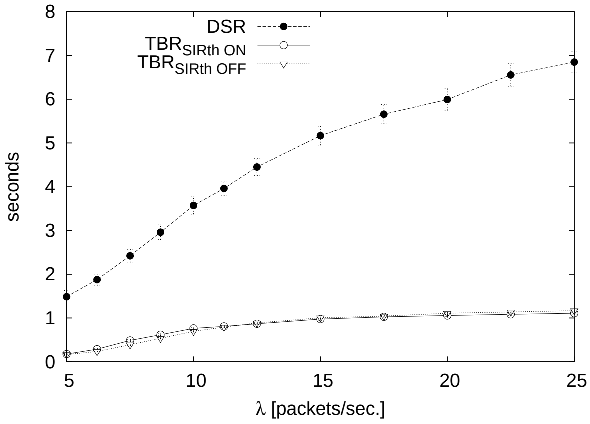

Figure 4: Average end-to-end (nodes → APs) packet delay.

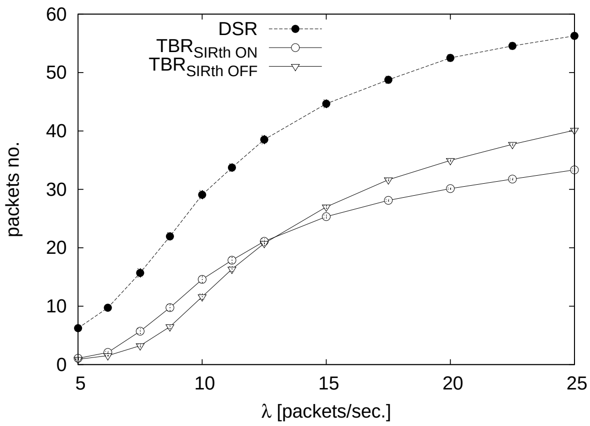

Figure 5: Average number of queued packets at the MAC level.

rithm is set to 0.5 dB. At the UMTS physical layer, a half rate convolutional code is used in each data channel. The packet size at the Radio Link Control (RLC) is set to 360 bits. For the IEEE802.11b RAT, the maximum transmission power is set to 20 dBm. For the MAC, we implemented the Distributed Coordination Function (DCF) mechanism as specified in the standard [1]. For the basic access mode we consider $T = 1$ s, see Section III. In DSR, the time to live (TTL) of Route REQuest (RREQ) packets is set to 15. Finally, for both DSR and TBR routing tables are refreshed every 1.5 seconds. As far as TBR is concerned, we refer to the version of the algorithm with and without the *link stability* mechanism as $\text{TBR}_{\text{SIRth-ON}}$ and $\text{TBR}_{\text{SIRth-OFF}}$, respectively. In $\text{TBR}_{\text{SIRth-ON}}$, only those beacons received with an adequate signal to noise ratio (enough to guarantee an error rate lower than 10^{-2}) are re-broadcast. The SINR threshold is derived from the transmission rate and modulation technique in use. In the results that we show next, 95% confidence intervals are plotted in all graphs by means of vertical dotted lines.

In Fig. 2 we plot the average throughput experienced by the *active users*. At low traffic loads, both $\text{TBR}_{\text{SIRth-ON}}$ and $\text{TBR}_{\text{SIRth-OFF}}$ give similar performance and outperform DSR. At high traffic load, instead, the throughput of the $\text{TBR}_{\text{SIRth-OFF}}$ scheme is substantially reduced. This is due to that fact that if weak links are not pruned from the shortest path trees, not all routes found by TBR are stable. The stability problem is more evident under the high traffic regime, where the multi-user interference is higher. Unstable links are pruned by DSR as well. This is achieved thanks to the two way procedure (RREQ/RREP) used to set up communication paths. According to this procedure control messages must flow in both directions (user → AP and AP → user) to set up a route. This helps in reducing the probability that the selected route is unstable. For this reason, the DSR performance is not reduced as the traffic load increases. Also, from Fig. 2 we observe that the average throughput increases when the traffic load is high for both DSR and $\text{TBR}_{\text{SIRth-ON}}$. This is explained by looking at the fairness among different data flows. Fairness results are reported in Fig. 3. In this figure, we plot Jain's Fairness Index [10] (FI):

$$FI = \frac{\left(\sum_{i \in N_A} \frac{T_i}{B_i}\right)^2}{n \sum_{i \in N_A} \left(\frac{T_i}{B_i}\right)^2} \tag{1}$$

where n is the number of *active users* and N_A is their set, while T_i and B_i are the throughput and the generated data rate (traffic load) associated with user i, respectively. When FI is close to one the service is fair. From Fig. 3, we observe that an increasing traffic load always leads to a decreasing fairness index. This is due to the fact that as the load increases the system interference also increases and only a few paths towards the APs are still stable and offer sufficient quality. In this case, only the very few users on these paths communicate with the APs, whereas the remaining users experience a very low throughput. The fact that the throughput at some point starts increasing with an increasing traffic load (see Fig. 2) is justified by the behavior of the IEEE802.11 MAC. In fact, with the IEEE802.11 RAT a few users sharing the same channel achieve an aggregate throughput which is higher than the one achievable for a larger number of users accessing the wireless medium. In Fig. 4 we plot the end-to-end (users → APs) delay. TBR schemes give substantial improvements for the delay metric. This is mainly due to the following facts. In DSR, in order to discover a route to a given AP, each node needs to independently start a RD procedure. RD impacts the performance for two reasons: 1) RD introduces extra delays, needed to contact the destination and establish the route (two way procedure) and, 2) frequent RDs increase the system interference, therefore reducing the bandwidth available for data transmission. In addition, the paths discovered by DSR are static, i.e., once discovered they are used until a Route FAILure (RFAIL) occurs. Hence, due to network dynamics the discovered routes may become sub-optimal. In contrast, TBR routes data on the fly by means of hop-by-hop routing decisions, by considering both *local* routing tables and the state of the neighbors of the current node (state of their queues). In addition, routing tables are updated proactively by means of beacon messages, by creating shortest path trees to every AP which are subsequently exploited by all user. The delay performance is also reflected in Fig. 5, where we show the queue length at the MAC layer, by averaging this value over all *active users*. The performance degradation due to frequent RDs in the DSR scheme is further illustrated in Fig. 6. In this graph, we report the time trace of the number of data packets transmitted by a terminal having the IEEE802.11b RAT only, which is not within the communication range of any AP. In the DSR case, the instantaneous data rate frequently drops to

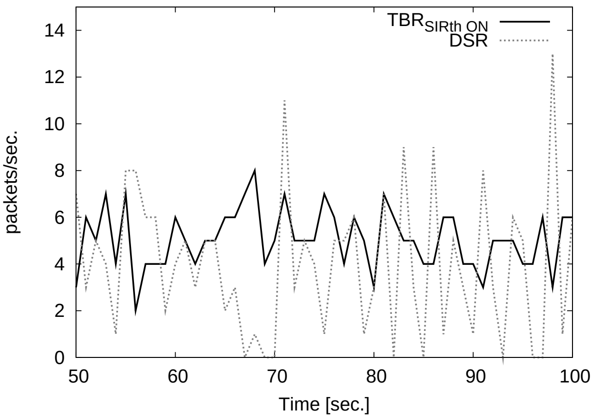

Figure 6: Throughput traces.

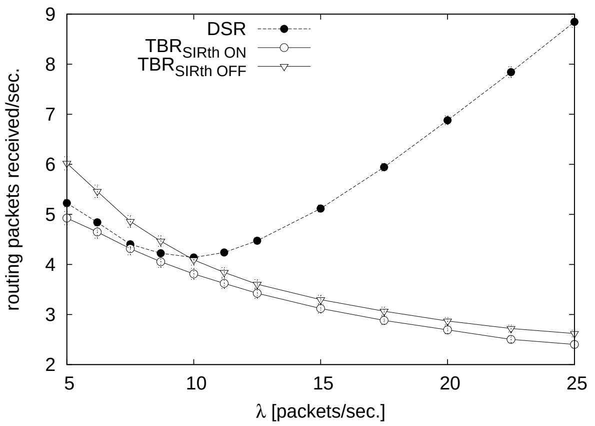

Figure 7: Routing control messages.

zero due to the activation of new route discovery procedures. This clearly reduces the average throughput and also the delay performance in terms of jitter. We note that with TBR the transmission is continuous, leading to a more desirable behavior. Finally, in Fig. 7 we show the number of control messages (RREQs/RREPs/RFAILs and beacons) received per second by a node. This graph demonstrates the impact of the traffic load on the number of routing messages that are pushed into the network. DSR tends to increase the number of RDs when the load increases. This is due to the fact that a higher load leads to a higher interference which, in turn, increases the route congestion probabilities. However, in such a case increasing the number of RDs has a negative impact on the performance, as it further increases the interference. TBR instead relies on periodic transmissions of control packets. This solution is clearly better when the network is congested. We finally observe that when the traffic load is very low or absent reactive solutions such as DSR are the best option. Hybrid schemes may therefore be designed, where the APs estimate the traffic load and switch between proactive (e.g., TBR) and reactive (e.g., DRS) mechanisms to achieve the best performance under any setting. In addition, the technique considered here, through minimal modifications, can also be exploited for downlink data transmission. A detailed study of these issues is left for future research.

V. Conclusions

In this paper we studied the problem of multi-hop data delivery in next generation heterogeneous wireless networks. The proposed algorithms are specifically tailored for routing over a well defined set of access points and are based on the proactive routing paradigm. We demonstrate that, in our settings, such an approach outperforms reactive schemes in moderate and high traffic load situations. Performance results are obtained via simulation by thoroughly modeling the wireless channel behavior and the considered radio access technologies.

Acknowledgment

This work has been carried out in the framework of the Ambient Networks project that is partially funded by the Commission of the European Union. The views expressed in this paper are solely those of the authors and should not be interpreted as necessarily representing the views of their employers, this project or the European Commission.

References

[1] IEEE802 LAN/MAN Standards Commitee, `http://www.ieee802.org/`.

[2] H. Kaaranen, A. Ahtiainen, L.Laitinien, S. Naghian, and V. Niemi, *UMTS Networks.* John Wiley & Sons, 2001.

[3] S. Nelakuditi, Z. L. Zhang, R. P. Tsang, and D. H. C. Du, "Adaptive proportional routing: a localized QoS routing approach," *IEEE/ACM Trans. Networking*, vol. 10, no. 6, pp. 790–804, 2002.

[4] C.-F. Huang, H.-W. Lee, and Y.-C. 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, 2004.

[5] L. Badia, N. Bui, M. Miozzo, M. Rossi, and M. Zorzi, "On the Exploitation of User Aggregation Strategies in Heterogeneous Wireless Networks," in *IEEE CAMAD*, Trento, Italy, Jun 2006.

[6] D. B. Johnson, D. A. Maltz, and J. Broch, *Ad Hoc Networking, Chapter 5: "DSR: The Dynamic Source Routing Protocol for Multihop Wireless Ad Hoc Networks"*. Addison Wesley, 2001.

[7] C. E. Perkins and P. Bhagwat, "Highly dynamic Destination-Sequenced Distance-Vector routing (DSDV) for mobile computers," in *ACM SIGCOMM*, New York, NY, USA, Aug 1994.

[8] S. Basagni, M. Conti, S. Giordano, and I. Stojmenovic, *Mobile Ad Hoc Networking.* Wiley-IEEE Press, 2004.

[9] G. L. Stüber, *Principles of Mobile Communication.* Kluwer Academic Publishers, 2003.

[10] D. M. Chiu and R. Jain, "Analysis of Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks," *Elsevier Computer Networks and ISDN Systems*, vol. 17, no. 1, pp. 1–14, 1989.