

Comparative Analysis of Quality of Service Routing in Wireless Metropolitan Area Networks

Matthias Hollick¹, Tronje Krop¹, Jens Schmitt¹, Hans-Peter Huth² and Ralf Steinmetz¹

¹Multimedia Communications Lab, Faculty of Electrical Engineering and Information Technology
Darmstadt University of Technology, Germany

²Siemens AG, Corporate Technology, Dept. Information and Communications, Munich, Germany
{Matthias.Hollick, Tronje.Krop, Jens.Schmitt, Ralf.Steinmetz}@KOM.tu-darmstadt.de
{Hans-Peter.Huth}@mchp.siemens.de

Abstract

*Currently, we see the evolution of large scale community and metropolitan area networks based on inexpensive wireless local area network technology. We present the results of an experimental analysis which investigates the potential of quality of service routing mechanisms within this challenging environment. Our investigation is based on a model of a radio access network designed to cover a large city center by means of decentralized and distributed routers, which are tightly meshed. The workload is modeled to reflect the estimated usage patterns based on statistical data collection of user mobility and combined with synthetic traffic matrices. We present results for various routing strategies including shortest path routing, delay constrained routing as well as various multipath quality of service routing variants. Moreover, we investigate different traffic distributions. Our findings are, that multipath routing is able to enhance the utility of the network significantly.**

1. Introduction

The concepts of 4G networks and the Wireless Internet embrace a heterogeneous set of access technologies, ranging from UMTS over wireless local area networks to small scale personal area networks. Currently, we see the evolution of large scale community and metropolitan area networks based on inexpensive wireless local area network technology. The resulting radio access networks give rise to several interesting research challenges.

Despite the fact that a huge amount of work is performed in this area, the implications of quality of service mechanisms within the context of large scale metropolitan area networks are not well known. We believe that today's mainly hierarchical radio access network topologies as well as static resource management approaches need reconsideration. Within our study, we

assume a distributed and decentralized network topology which is characterized by small cell sizes. Tightly meshed routers interconnect the cells and allow to follow the packet-switched paradigm for resource management purposes.

Induced by user and device mobility - while tightly coupled to usage and application-specific traffic patterns - these networks will experience heavily varying loads on different timescales. We regard the investigation of the hereby caused problems as crucial to allow for efficient network planning and control of large scale wireless communication networks.

Within this paper, we evaluate and analyze the potential of QoS routing strategies for load balancing. The chosen scenario for this study is a model of a radio access network covering Darmstadt, Germany, a city with around 140.000 inhabitants. Since analytical tractability is infeasible due to the given complexity of the scenario, an experimental analysis is performed. We give a brief description of the employed workload model as well as of the investigated topology.

The major contributions of our investigation are:

- (I) The analysis of various routing strategies in wireless metropolitan area networks with respect to load balancing. We include shortest path routing, delay constrained routing as well as various multipath QoS routing strategies.
- (II) The evaluation of the influence induced by different traffic distributions. This includes centralized traffic only to/from an edge gateway as well as different ratios of external/internal traffic.

The remainder of the paper is organized as follows. In the next section, previous and related work is surveyed. In Section 3 we describe the overall system as well as the goals and services therein. We present the selection of metrics and parameters to study. Section 4 contains the concise description of the experimental design based on the selection of factors and workload. The analysis and interpretation of the results is presented in Section 5. The paper concludes by summarizing the major results and discussing further aspects of future research in Section 6.

* This work is partially funded by Siemens AG, CT IC 2 within the MobQoS project. Additional information may be obtained at <http://www.kom.tu-darmstadt.de/Research/MobQoS>

2. Related work

The area of QoS routing has recently attracted a lot of research. Traditionally, QoS routing is applied to find constrained paths within the context of fine-grained resource reservation. Constraints may be of additive, concave or multiplicative type [1]. The complexity of finding a feasible path through the network depends on the number of constraints and their type as well as on the nature of the algorithm (centralized, decentralized, hierarchical). Constraints may be, for example, delay, capacity, jitter and loss-ratio. [1] showed that the problem of finding feasible paths with two independent types of constraints is NP-complete (note that delay and capacity are not independent and allow for algorithms and heuristics with lower complexity).

Current QoS routing strategies can be divided into source-routing, distributed-routing and hierarchical routing. A good overview to the field of constraint-based routing and QoS routing can be found in [2] and [3]. For an excellent algorithmic treatment of constraint-based routing algorithms see [4], where the authors simulated several algorithms in order to determine their worst case complexity. Last, [5] gives an algebra for QoS path computation and concisely describes hop-by-hop QoS routing mechanisms including multipath algorithms as well.

The use of QoS routing mechanisms for load balancing has been discussed for multiple protocols. However, most of the work was done for source routing algorithms like the Q-OSPF protocol [6]. Since Q-OSPF was designed for QoS routing in the sense of finding hard-QoS constrained paths, the "abuse" for resource management / traffic engineering inherits some restrictive assumptions. This includes the operation on flow-granularity and the need for a surrounding framework for admission and reservation of resources. There are other algorithms such as, for example, "shortest path first" algorithms optimized for computation of minimum congestion paths [7] which are limited by the source routing paradigm, too. In summary, source routing strategies are well-suited within reservation-based or connection-oriented systems. However, they impose unnecessary and unwanted complexity in connectionless systems which rely on hop-by-hop operation.

If we regard connection-oriented networks, it is quite natural to use multiple paths (one per flow) between source-destination pairs and thus follow a multipath paradigm. In our context, namely connectionless networks, the class of multipath routing algorithms is often overlooked, however. The use of multiple paths at the same time allows for dispersion of the traffic on different granularity and thus alternate paths can help improving network performance significantly. Work in the area of traffic dispersion on packet or sub-packet level range from early work [8] to more recent approaches. Examples for multi-

path routing algorithms include [9], [10], [11] and [12]. Hereby the first three are based on a link state protocol and thus operate on global imprecise network state. The last algorithm is a distance vector based algorithm operating on information provided from neighboring nodes.

The OSPF variants ECMP (equal-cost multipath) [9] and OMP (optimized multipath) [10] are heuristics which are targeted to better distribute load over multiple paths. OSPF-ECMP, for example, distributes the load over multiple equal-cost paths using simple round robin mechanisms and thus limits the possible load-balancing to this restricted set of paths. The equal distribution among these paths may result in suboptimal performance, too. OSPF-OMP optimizes this distribution using heuristics to predict which path to use. The necessary load information of the network is hereby distributed using OSPF opaque link state advertisements.

In contrast to the more intuitive heuristics given in [9] and [10] the recent work from Vutukury and Garcia-Luna-Aceves approaches the problem from a more theoretical perspective. Based on optimal routing [13] they formulate the properties a distributed multipath routing protocol must follow. The application for traffic engineering is described, too [14].

For multipath algorithms it is crucial to efficiently deal with routing loops, because the number of feasible paths largely depends on the strategy chosen to handle loops [5]. However, it is not necessary for the paths to be disjoint if the algorithm is carefully designed. The promising analysis of the performance properties of multipath algorithms in [5] and the robustness of such algorithms are of high interest.

What currently misses for all of the above mentioned QoS routing protocols are investigations of their performance within our target scenario: wireless metropolitan area networks. Most of the protocols have been designed and evaluated for use in backbone topologies. However, in our scenario we expect heavily varying loads on different timescales induced by user and device mobility. The network structure is likely to be more regular than in backbone networks while the degree of interconnection will be higher.

We are not aware of comparative investigations of QoS routing algorithms within metropolitan radio access networks. This is not surprising, because so far these have not been routing networks. There exists, on the other hand, some recent work which investigates large scale wireless LAN topologies. In particular, the work of Tang and Baker [15] is able to provide deep insights of user behavior for a metropolitan area wireless network. The work of Kotz and Essien [16] claims to be the largest and most comprehensive real world trace of a production wireless LAN. The results, however, do account for a special campus style network and mainly focus on traffic analysis - the mobility aspect is restricted by the campus setup and thus cannot be

transferred to public networks. While these works address the area of wireless metropolitan area networks, they are only able to survey the current usage patterns within these networks. The investigation of underlying network characteristics or specialized routing strategies are beyond the scope of these works.

Our work addresses the combination of the above outlined research areas: the comparative analysis of QoS routing strategies for use in wireless metropolitan area networks.

3. Scenario

Why do we want to analyze the performance of QoS routing algorithms within wireless metropolitan area networks? We believe that the challenges of mobility should be addressed using novel paradigms in network control and design. We expect access network infrastructures to become more and more heterogeneous, for example by incorporating wireless local area network technology into traditional cellular network infrastructures. With the increasingly important role of Internet technology, there are also new opportunities for optimization of traditional mechanisms. We are particularly interested in the investigation of resource management and load balancing issues within future metropolitan radio access networks.

The system boundaries for our analysis are defined by the MobQoS scenario [17]. The topology we investigate is a tightly meshed metropolitan radio access network infrastructure which is instantiated for the case of the city of Darmstadt. Hereby, we follow a three-tier architecture depicted in Fig. 1. The architecture is derived from existing concepts of the IETF, here especially Mobile IPv4 [18] and IPv6 [19] respectively, QoS architectures [20] as well as traffic engineering work [21], and the concepts of telecommunication, here especially UMTS technology [22]. A crucial point is the distributed and decentralized nature of the proposed topology compared with traditional hierarchical topologies in telecommunication networks like GSM.

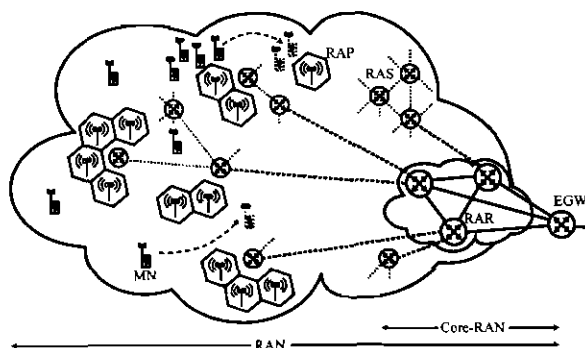


Figure 1: Schematic overview of network topology

The mobile nodes or terminals (MN) of the customers are associated to wireless base stations, the so-called radio access points (RAP) representing the last hop of the provision network. The function of the first tier thus can be described as radio access.

The second tier, the core radio access network (RAN) comprises so called radio access server (RAS), which are used to attach multiple RAPs. Within the radio range covered by all RAPs attached at one RAS mobility is supported by appropriate link layer mechanisms, like, for example, handover. Each RAS has built-in IP-Router functionality. The RAS are meshed with neighboring systems and thus allow the start of resource management at this level. The core of the second tier is built by radio access routers (RAR). Between selected RAS and RAR are uplinks, the RAR are fully meshed.

The transition to the third tier of the architecture, the core provider network respectively the Internet, is performed by a single or multiple edge gateways (EGW). At this point the technology may be mapped to the according mechanisms and strategies of the core provider network or the Internet and vice versa.

We model the presence of a large number of radio access points deployed in decentralized fashion. However, current approaches towards traffic engineering mainly rely on centralized components and build upon explicit mechanisms for signaling and resource admission. Within our investigation we decided to preserve the connectionless nature of the Internet by employing decentralized hop-by-hop QoS routing strategies. The approach should be able to work without explicit signaling and without a need to operate on the flow level. Goals and services include a resource management to ensure resource-efficient network operation in a decentralized fashion while supporting mobile users.

The radio access network for Darmstadt resulting from the assumptions introduced above consists of a large number of micro cells and 83 macro cells (see [17] for a map of Darmstadt and the corresponding macro cells). Each of these macro cells is maintained by a RAS. Since a core idea behind MobQoS is a decentralized RAS structure which allows for Resource Management starting at the RAS level, the RAS are modeled as routers which are interconnected to provide for alternate links. We have chosen a tightly meshed infrastructure as a starting point. See Fig. 2 for the resulting topology.

4. Experiment

Since analytical tractability is out of scope for the given complexity of the problem, we use the methodology of experimental analysis for our investigation [23]. This section comprises the concise description of the experimental design based on the selection of factors and workload.

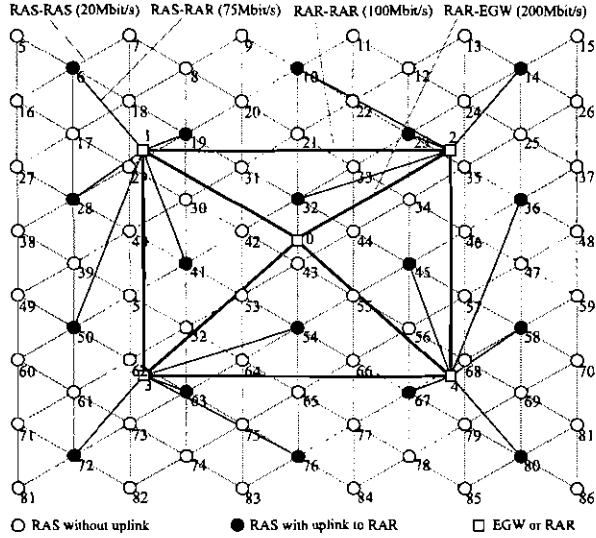


Figure 2: Topology of radio access network

The experiment was implemented using ns-2 [24]. We incorporated major changes to the ns-2 simulation framework. This includes the implementation of a flexible multipath QoS algorithm which can be parameterized to reflect different variants of routing strategies. Moreover, we needed to implement methods to transfer the results of our traffic model into the simulation environment.

4.1. Factors

The parameter of utmost importance is the routing strategy. Other parameters include the influence of user mobility and the influence of the traffic distribution. The routing strategies investigated include combinations out of the sets {static, dynamic}, {singlepath, multipath} and {distance, delay}. In particular we include the following routing algorithms in our investigation:

- Static shortest-distance singlepath algorithm: *dst*.
- Static minimum-delay singlepath algorithm: *dly*.
- Static equal-shortest-distance multipath algorithm: *smp*.
- Dynamic minimum-delay multipath algorithm, multipath scheduling is done using round robin on packet level: *dmp*.
- Dynamic minimum-delay multipath algorithm, multipath scheduling is done using weighted round robin on packet level: *wmp*.

Due to oscillating behavior of the original *dmp* and *wmp* algorithms we also included the damped versions *dmp1* and *wmp1* (averaging of the last two delay measures) and the predicting versions *dmp2* and *wmp2* (based on the trend of the delay measures) in our investigation.

We also investigate the influence of different traffic distributions. The starting point is derived from current

radio access network structures, which are strictly hierarchical. This corresponds to only external traffic for the end-systems to/from a single edge gateway. Moreover, we use different ratios of external to internal traffic (*e/i*-ratio) within our investigation. This reflects the amount of local traffic which may arise for future services. We use the four traffic classes {conversational, streaming, transactional, background} introduced in [22] since they should reflect a good guess of the intended usage of the network in question. The *e/i*-ratio is set for each class independently. See Table 1 for the predictor variables and their corresponding levels. The capacity of the links is shown in Fig. 2, the buffer of the routers equals the equivalent of 5 ms at full speed.

4.2. Workload

We use a workload model which is a hybrid between purely synthetic and statistical models for our investigation. The analytical foundations of the model are described in [25]. An instantiation based on extensive collection of statistical data for the scenario of Darmstadt can be found in [26]. In short, the model combines a statistical mobility part with a synthetic traffic part to allow for flexible but realistic workload generation. The resulting workload for

Table 1: Predictor variables, levels and description

Variable /Factors	Description			
Routing Algorithm				
	<i>Dynamic</i>	<i>Metric</i>	<i>Type</i>	<i>Multipath Scheduling¹</i>
<i>dst</i>	No	Distance	Singlepath	-
<i>dly</i>	No	Delay	Singlepath	-
<i>smp</i>	No	Distance	Multipath	RR
<i>dmp</i>	Yes	Delay	Multipath	RR
<i>wmp</i>	Yes	Delay	Multipath	WRR
Traffic Distribution² (external/internal)				
<i>Class³</i>	<i>Conv.</i>	<i>Stream.</i>	<i>Interact.</i>	<i>Back.</i>
Dist. (0)	100/0	100/0	100/0	100/0
Dist. (1)	75/25	90/10	80/20	85/15
Dist. (2)	50/50	80/20	60/40	70/30
Dist. (3)	25/75	70/30	40/60	55/45
Dist. (4)	0/100	60/40	20/80	40/60

¹ RR: round robin, WRR: weighted round robin

² The traffic distribution is given as external/internal traffic ratio (*e/i*-ratio). For example 80/20 means: 80% of total traffic to/from the edge gateway, 20% of total traffic internal.

³ *Conv.*: conversational, *Stream.*: streaming, *Interact.*: interactive, *Back.*: background

the ns-2 simulation can be summarized as follows. We see low user activity in the early morning. Starting from 4:00h in the morning, the number of active users increases as people get ready to work and on their way to work (including pupils and students as well). From 8:00h on most people are up and active. It is important to keep in mind that the geographical distribution of user density changes significantly over day, due to the different activities pursued by the different species of users. This shift in density can be seen especially from residential areas to workplaces. Special places like the university attract large numbers of user during daytime. The decrease in the evening is not as abrupt as the increase seen in the morning. Nevertheless, the focus is changed from work to residential areas. Fig. 3 illustrates the summarized number of active users over a day for all user groups.

The overall traffic estimate over the day is given in Fig. 4. Because users in different classes produce different load, the peak traffic value does not correspond directly to the peak value of user activity. As we noticed earlier, our traffic distribution is dependent on the geographical location which cannot be seen in Fig. 4.

The summarized traffic density for the modeled day is given in Fig. 5. The snapshots presented cannot show the time fluctuation of traffic. To do so, an animation of the traffic density over the day can be found at [17].

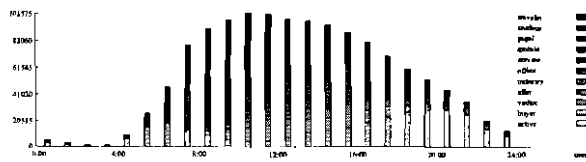


Figure 3: Number of active users over 24h

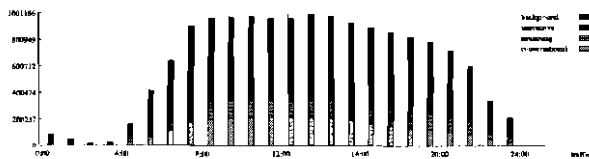


Figure 4: Traffic rate over 24h in kbit/s

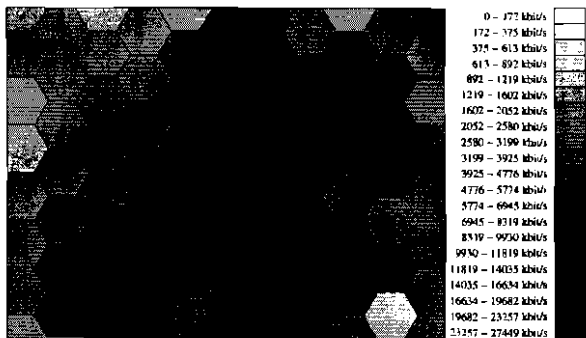


Figure 5: Traffic density over 24h

4.3. Experimental design

We have chosen to implement a full factorial experiment, since we regard only two predictor variables with five levels each [23]. The routing algorithm and the traffic distribution act as predictor variables. There are several factors which are kept constant for all experiments.

The scaling of the experiment to a different timescale is inevitable due to the size and complexity of the modeled scenario. We choose the simulation time to be 960 seconds to match the real time of 24 hours. Thus one second in simulation equals 90 seconds in real time. The amount of traffic is given by our workload model and used to parameterize the traffic agents at regular intervals. The updates of the workload were performed every 2 minutes (real time). The routing updates were performed every 6 minutes (real time). The mobility of users is inherently included within the mobility model and thus reflected in the workload model.

The runtime of each series of experiments is approximately one week while memory usage approaches nearly 800MB of RAM (we used four simulation machines each equipped with Pentium4 2,2GHz, 1GB of RAM in parallel).

5. Results

The main goal of our analysis was the investigation of the feasibility of QoS routing mechanisms to support load balancing within future wireless metropolitan area network topologies. Efficient load balancing translates into maximizing the overall utility of the network, which cannot easily be measured. We investigate various response variables, which are tightly related to the overall utility of the network. Moreover, we analyze the results on different aggregation levels of the obtained data, which are *link level*, *flow level* and *packet level*.

Besides measuring the exact numbers, we give various visualizations to allow for better interpretation of the results. For most parts, we provide results which cover the complete 24 hour simulation period. If necessary, we give the results for the busy hour. Since we obtained a large set of results, we will only give the most important ones within this document due to space limitations. We use the notation *algorithm-0.distribution-varianta* in the figures, therefore the suffix *wmp-0.3-2a* denotes the *wmp* algorithm variant 2, traffic distribution *dist. (3)*.

The results are structured as follows. Results on link level are given in Section 5.1. This includes figures to illustrate *load* and *loss* for RAS-RAS, RAS-RAR and RAR-Backbone level for selected algorithms and traffic distributions as well as a summary of *load* and *loss* on link level for selected algorithms. The results obtained on *flow level* are given in Section 5.2. This includes results for *path length* as well as *delay per hop* on flow level. The

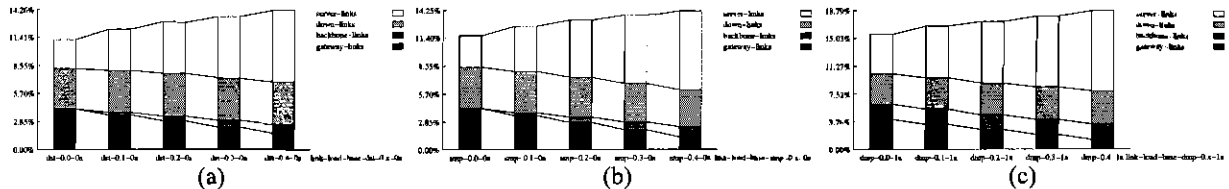


Figure 6: Average load, 24h, all traffic distributions, (a) *dst*, (b) *smp* and (c) *dmp1*

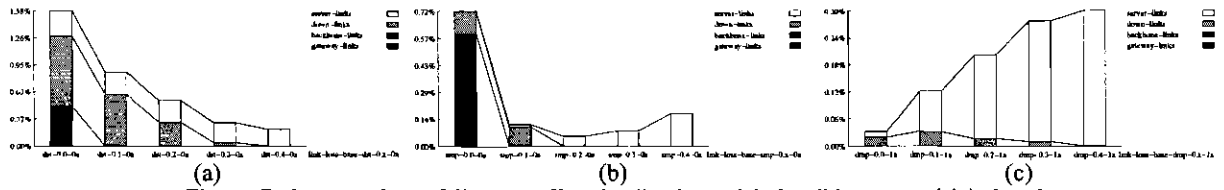


Figure 7: Average loss, 24h, all traffic distributions, (a) *dst*, (b) *smp* and (c) *dmp1*

results for *delay* of individual packets (*packet level*) are presented in Section 5.3. Finally, we give a *summary* of all *application* and *routing* related response variables in Section 5.5.4.

5.1. Results on link level

The results on link level include the mean load and loss for the various levels of the network topology. This accounts for multiple aspects of the algorithms studied:

- Does the algorithm use the interconnections on RAS-RAS level efficiently.
- Which levels of the infrastructure are dimensioned appropriately.
- Is the performance of the algorithm acceptable under normal conditions (investigated over 24 hours) and under heavy load (investigated over the busy hour).

We give some diagrams which allow for comparison of the different algorithms for an *e/i*-ratio of 100/0 for all traffic classes as well as for selected other distributions. Moreover, we provide some selected visualizations of individual algorithms for all traffic distributions investigated.

5.1.1. Load on link level. The load on link level clearly favors the shortest path algorithms including the shortest multipath as well (see Fig. 8 and Fig. 9). The more advanced algorithms try to use unused links at the expense of longer routes. This increases the overall load per level. The absolute load cannot be used as indicator for the quality of the load balancing, though. It allows, however, to see at which levels the algorithms act differently. In our example the increased use of server links is visible for the *dmp* and *wmp* variants. The results also exhibit the possible gain within the modeled network if we assume alternate paths in combination with multipath routing to distribute the load.

The differences between the algorithms remain visible if we introduce local traffic. However, the individual algorithms show different tendencies under these conditions. While the *dst* algorithm improves due to the shift of load from the uplinks, the *dly* algorithm suffers under the same conditions. The *smp* algorithm performs well under all traffic distributions outperforming all single path algorithms. Fig. 6 illustrates this for (a) *dst*, (b) *smp* and (c) *dmp1* for all investigated traffic distributions.

5.1.2. Loss on link level. The loss on link level is an indicator for congested links and areas within the network. Moreover, the hierarchy level at which loss occurs is closely related to the fundamental principles of the underlying algorithms. Our results for the best algorithm *smp* show minor losses on the downlinks (RAS-RAR) and on the gateway links (RAR-EGW). These losses occur because the links are not capable of carrying all traffic to/from the EGW to the source or destination even using multiple paths. The single path algorithms are clearly limited by the capacity of the links (Fig. 10). The results for *dmp* and *wmp* cannot easily be analyzed without additional data, though *wmp* suffers from heavy oscillations. It seems that the load balancing achieved comes at the expense of overloading individual levels unnecessarily.

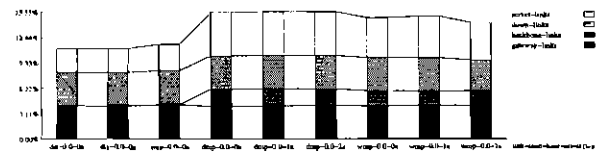


Figure 8: Average load, 24h, all algorithms, dist (0)

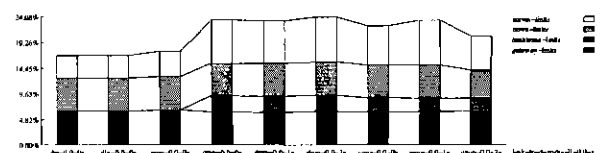


Figure 9: Average load, 1h, all algorithms, dist. (0)

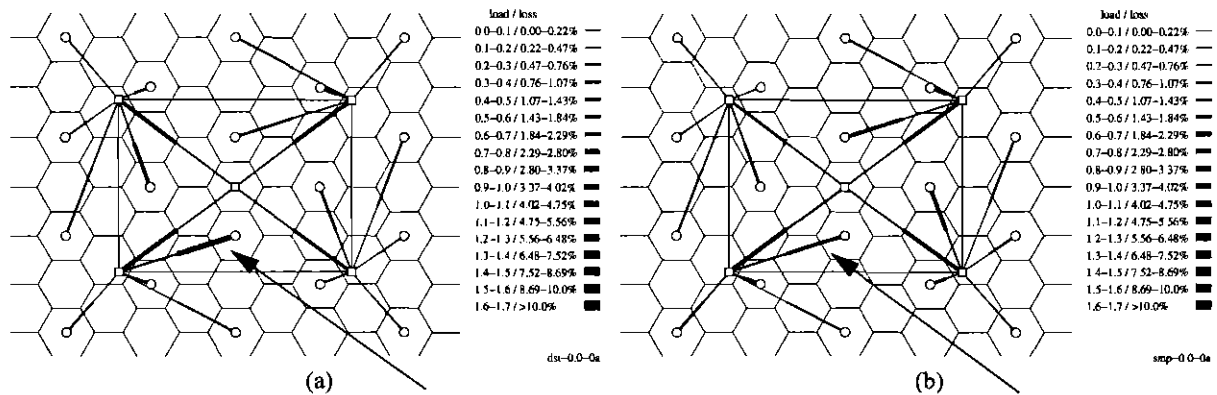


Figure 13: Load / loss visualization, 1h, (a) *dsf* RAR-EGW, (b) *smp* RAR-EGW, dist. (0)

The loss behavior of the algorithms changes significantly if the traffic distribution includes more internal traffic. Fig. 11, and Fig. 12 show a comparison of the average loss for all algorithms for the traffic distributions dist. (2) and dist. (4) during the busy hour. Fig. 7 gives the average loss over 24 hours for the algorithms (a) *dsf*, (b) *smp* and (c) *dmp1* to illustrate the loss characteristics for all investigated traffic distributions.

5.1.3. Visualizations of load and loss. Besides the above mentioned values, we are interested in finding the “problem zones” of our network. It is obvious that the hot spots of our workload model are correlated to these areas. However, we observed interesting effects on how the different routing algorithms deal with these hot spots. The following visualizations have to be interpreted as follows: The width of the edges entering a node depicts the mean percentage of load (in relation to link capacity) entering this node (the traffic direction is important to model asym-

metry, because we estimate more gateway to host traffic than vice versa). The brighter edges mark links where loss occurs. The brighter the gray, the more loss is measured. We give results for the RAR-Backbone level.

The busy hour plot shows different degrees of loss for both algorithms. The hot spots during the busy hour are clearly visible. The losses for the single path *dsf* algorithm are induced from traffic which flows from the gateway to the center of the city (see Fig. 13 (a)). In addition to the large losses on RAR-EGW level and RAS-RAR level there are minor losses at the RAS-RAS level. The *smp* algorithm is able to split the traffic and thus only a small fraction of packets is lost (see Fig. 13 (b)). Moreover, the link utilization (load) is distributed more uniformly. We can conclude that even simple multipath algorithms are able to achieve a better load distribution than standard Internet routing.

5.2. Results on flow level

Some response variables including delay and path length can only be measured for individual flows, thus we give these results on flow level. Delay acts as a response variable describing the ability of the algorithm to balance the network load (the delay decreases or remains low even under heavy load) while path length is of special interest for the quality of the multipath algorithms. These results account directly for the treatment individual flows get throughout the entire network while load and loss on link level (see above) accounts for the treatment of all flows in the corresponding tier of the network.

In the following, we regard the path length and delay per hop on flow level. The increased path length of *dmp* and *wmp* are due to the usage of longer routes for load balancing. On the other hand, the normalized delay per hop is smaller for *dmp* and *wmp* compared to the single path algorithms. In Fig. 14 we show the average path length for all algorithms in hops. Fig. 15 gives the results for the delay per hop in ms. Both figures assume all traffic distributions.

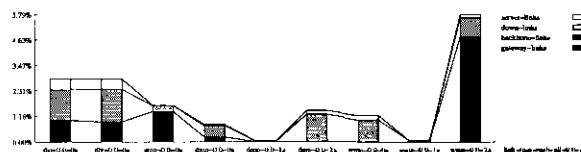


Figure 10: Average loss, 1h, all algorithms, dist. (0)

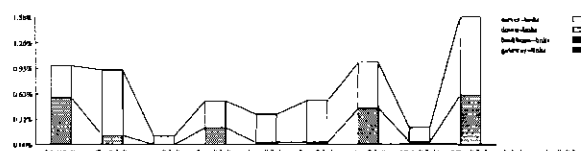


Figure 11: Average loss, 1h, all algorithms, dist. (2)

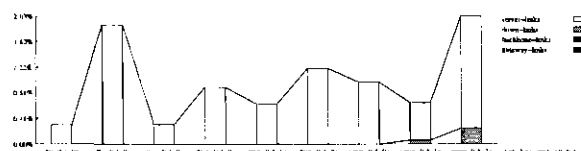


Figure 12: Average loss, 1h, all algorithms, dist. (4)

5.3. Results on packet level

The trace of the treatment of individual packets allows to give results on packet level. We injected some measurement flows into the network, which were closely monitored. The flows were injected in both directions between measurement points to account for asymmetric behavior. From a total of eight measurement flows which were present in each run of the experiment we only provide the most interesting insights. The individual measurement flows have been set up as constant bitrate over UDP, 128kBit/s each.

In the following, we present the comparison of some algorithms using selected measurement flows. We selected a trace from the gateway to a RAS located in the center of the network (EGW-RAS53, see Fig. 2). We expect the results to be of interest, because during the busy hour this area of the network is under heavy load. The second trace selected starts from RAS 83 and is destined to RAS 43. This trace crosses the backbone (shortest path) or tries to find a way around if the backbone is under heavy load conditions (for delay sensitive algorithms). We present these traces for *dst*, *smp*, *dmp* and *wmp2*.

The scatter-plots give the delays for individual packets. If packets are dropped due to congestion, the time delay until their drop is counted. Note that the scale on the y-axis may vary and thus cannot easily be compared between algorithms.

5.3.1. Delay of packets, EGW to RAS53. The figures Fig. 16 to Fig. 19 show a measurement flow from the edge gateway to the city center. All scatter-plots show an increasing delay for the time period between 8:00 and 18:00h. The *dst* algorithm shows some dramatic increases in delay if the path chosen is congested. Please notice the different scale on the y-axis: The delay for the *dst* algorithm rises as high as 19ms while the *smp* algorithm constantly remains below 8ms. The *dmp* variant shows a good performance most of the time, the maximum delay being around 8ms. However, there are some short periods where the algorithm fails and as a result dramatic losses

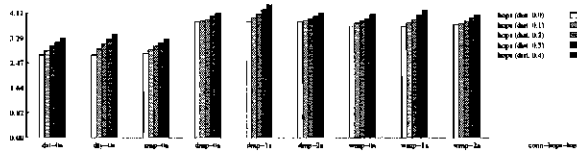


Figure 14: Path length, 24h, all algorithms, all dist.

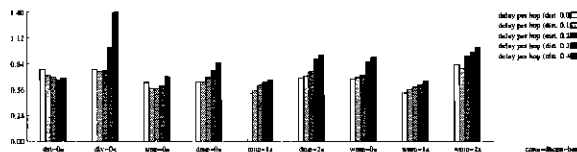


Figure 15: Delay per hop, 24h, all algorithms, all dist.

and an increasing delay up to around 20ms can be observed.

We provide the results for the *smp* algorithm to show the behavior for traffic distribution (4). Fig. 19 clearly shows the smaller delay (reduced load) compared to Fig. 17 due to the shift from Edge Gateway traffic to localized traffic. The oscillating behavior of the *dmp* (and *wmp*) algorithms will be investigated below. Periods of good delay / loss ratio are followed by periods where loss occurs and the achieved delay increases significantly. With the current parametrization of the routing algorithms these variants do not provide a stable performance.

5.3.2. Delay of packets, RAS83 to RAS43. The results for the measurement flow between RAS83 and RAS43 are very interesting. The single path algorithm *dst* “fails” if the backbone is under heavy load. The loss increases and the delay is as high as 21ms. The rise is relatively abrupt and the decline is smoother. This is due to the traffic load which is similarly shaped. The *smp* algorithm has only short periods of time where the increase in delay and loss is heightened (see Fig. 21). The increase in delay starts approximately 2 hours later in time and the decrease around 3 hours earlier in time as observed with the *dst*

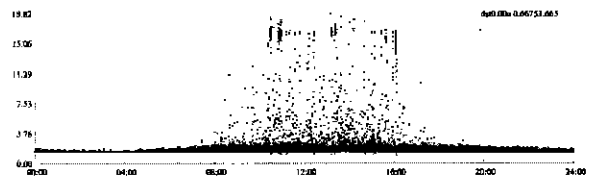


Figure 16: Delay, *dst*, 24h, dist. (0), EGW-RAS53

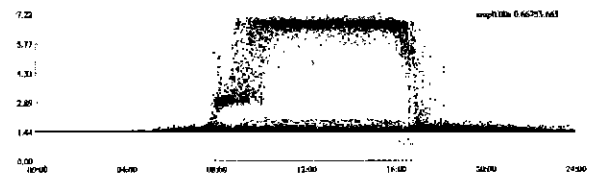


Figure 17: Delay, *smp*, 24h, dist. (0), EGW-RAS53

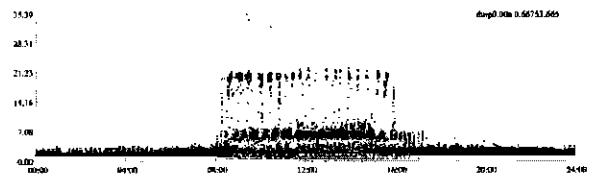


Figure 18: Delay, *dmp*, 24h, dist. (0), EGW-RAS53

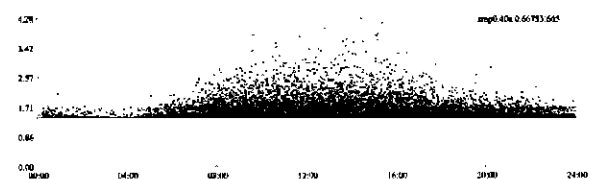


Figure 19: Delay, *smp*, 24h, dist. (4), EGW-RAS53

Table 2: Summary of results

	<i>dst</i>	<i>dly</i>	<i>smp</i>	<i>dmp</i>	<i>dmp1</i>	<i>dmp2</i>	<i>wmp</i>	<i>wmp1</i>	<i>wmp2</i>
<i>Application related</i>									
Average E2E Delay (standard deviation)	2.13ms (0.0027)	2.14ms (0.0028)	1.78ms (0.0013)	2.46ms (0.0022)	1.97ms (0.0010)	2.64ms (0.0026)	2.48ms (0.0026)	1.94ms (0.0010)	3.09ms (0.0030)
Loss	4.24%	4.13%	1.96%	1.26%	0.12%	2.83%	2.46%	0.13%	10.52%
<i>Routing related</i>									
Mean Path Length [hops]	2.73	2.73	2.77	3.82	3.82	3.82	3.66	3.66	3.71
RAS-RAS Level Load, 24h (variance)	4.65% (3.90)	4.68% (3.91)	5.12% (2.30)	8.39% (1.54)	8.49% (1.53)	8.51% (1.81)	7.74 (1.46)	8.07% (1.72)	7.27% (1.45)
RAS-RAR Level Load, 24h (variance)	20.59% (0.97)	20.60% (0.98)	21.05% (0.87)	20.45% (0.65)	20.52% (0.65)	20.62% (0.69)	20.43% (0.63)	20.59% (0.66)	18.99% (0.51)

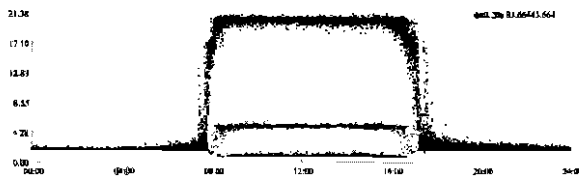


Figure 20: Delay, *dst*, 24h, dist. (2), RAS83-RAS43

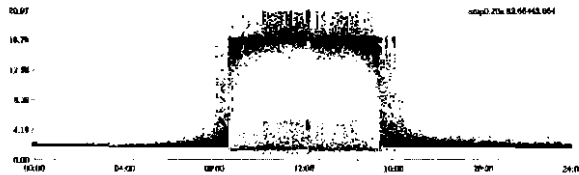


Figure 21: Delay, *smp*, 24h, dist. (2), RAS83-RAS43



Figure 22: Delay, *dmp*, 24h, dist. (2), RAS83-RAS43

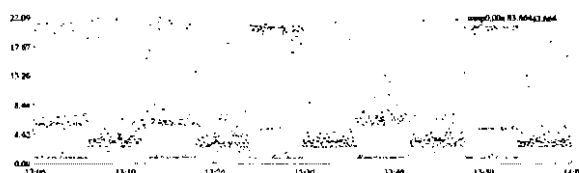


Figure 23: Oscillating behavior of *wmp2*

algorithm (see Fig. 20). This shows the much better handling of congestion using multipath algorithms which are able to split the traffic.

The *dmp* and *wmp* algorithms do not perform as well as intended. We see the influence of multiple paths with different visible steps in the delay. *Dmp* as depicted in Fig. 22 shows heavy oscillations. The increase in delay and loss occurs later than for the *dst* algorithm, although earlier than for the *smp* algorithm. The delay of the

majority of packets for all multipath variants remains low (<10ms) even under heavy load.

To further illustrate the oscillating behavior observed for the dynamic multipath algorithms Fig. 23 shows the busy hour for the *wmp2* algorithm. The oscillations are clearly visible: periods of good performance (no loss, low delay) alternate with periods of bad performance (high loss, higher delay). The boundaries between these periods are determined by our rerouting interval of 6 minutes.

5.4. Summary of Results

Table 2 gives a summary of the overall results obtained for the 100% to/from Edge Gateway traffic distribution. The overall performance of the network can be judged using the application-related parameters of average loss and average end-to-end delay. The routing-related values include the mean path length which is of particular interest for the multipath algorithms. Moreover, the load in combination with the load variation is given for different hierarchy levels. The smaller the variance, the more effective the load balancing achieved.

6. Conclusions

We have discussed the influence of QoS routing algorithms within wireless metropolitan area networks with respect to load balancing. As a first step, we defined the requirements for use within the investigated environments and modeled a corresponding topology.

We have chosen a set of routing strategies to compare. Moreover, we gave a concise description of the experimental design. This included the definition of predictor variables as well as a brief description of the workload model employed. The models served as a basis for the second part of our investigation: the quest for efficient load balancing within wireless metropolitan area networks. The impact of the different algorithms as well as of various traffic distributions on the overall routing performance was traced by means of simulation.

Efficient resource management increases the overall network utility—which cannot be measured easily, however. Thus, a detailed analysis of the results has been performed. We used the loss and load of individual link classes as well as of the entire network to represent this overall utility. Moreover, we investigated the delay and delay variation of individual streams and packets. Our findings are that multipath algorithms exhibit very promising results compared to single path algorithms. However, we have not been able to tap the full potential of dynamic multipath routing algorithms based on the minimum delay metric, because the working point of the algorithm was estimated suboptimally.

The insights presented aid network designers in developing new paradigms in the area of wireless metropolitan area networks. This especially includes resource management scenarios. As future work, we perceive the analysis of the oscillating behavior of the delay-based multipath algorithms.

7. References

- [1] Z. Wang and J. Crowcroft. Quality-of-Service Routing for Supporting Multimedia Applications. *IEEE Journal of Selected Areas in Communications*, 14(7):1228–1234, 1996.
- [2] S. Chen and K. Nahrstedt. An Overview of Quality-of-Service Routing for the Next Generation High-speed Networks: Problems and Solutions. *IEEE Network Magazine, Special Issue on Transmission and Distribution of Digital Video*, 12(6):64–79, November-December 1998.
- [3] P. Paul and S.-V. Raghavan. Survey of QoS routing. In *Proceedings of the Fifteenth International Conference on Computer Communication (ICCC 2002), Mumbai, India*, August 2002.
- [4] F. Kuipers, T. Korkmaz, M. Krunz, and P. V. Mieghem. A review of constraint-based routing algorithms. *IEEE Communications Magazine*, 40(12):50–55, December 2002.
- [5] J. L. Sobrinho. Algebra and Algorithms for QoS Path Computation and Hop-by-Hop Routing in the Internet. In *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'01), Anchorage, AK, USA*, volume 2, pages 727–735, April 2001.
- [6] G. Apostolopoulos, D. Williams, S. Kamat, R. Guerin, A. Orda, and T. Przygienda. QoS Routing Mechanisms and OSPF Extensions. Experimental RFC 2676, September 1999.
- [7] S. Park, Y. Serbest, and S. qi Li. Minimum congestion traffic engineering with multi-resource constraint. In *Proceedings of IEEE International Conference on Computer Communications and Networks, ICCCN 2001, Scottsdale, AZ, USA*, pages 130–137, October 2001.
- [8] N. F. Maxemchuk. Dispersity routing. In *Proceedings of ICC '75, San Francisco, CA, USA*, pages 41.10–41.13, June 1975.
- [9] J. Moy. OSPF Version 2. Proposed Internet Standard, RFC 2328, April 1998.
- [10] C. Villamizar. OSPF optimized multipath (OSPF-OMP). Internet Draft, draft-ietf-ospf-omp-03, January 2002.
- [11] S. Vutukury and J. J. Garcia-Luna-Aceves. A Simple Approximation to Minimum-Delay Routing. In *Proceedings of the annual conference of the Special Interest Group on Data Communication (SIGCOMM)'99, Cambridge, MA, USA*, pages 227–238, August 1999.
- [12] S. Vutukury and J. J. Garcia-Luna-Aceves. MDVA: A Distance-Vector Multipath Routing Protocol. In *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'01, Anchorage, AK, USA)*, pages 557–564, April 2001.
- [13] D. G. Cantor and M. Gerla. Optimal routing in a packet-switched computer network. *IEEE Transactions on Computers*, C-23(10):1062–1069, October 1974.
- [14] S. Vutukury. *Multipath Routing Mechanisms for Traffic Engineering and Quality of Service in the Internet*. PhD thesis, University of California Santa Cruz, March 2001.
- [15] D. Tang and M. Baker. Analysis of a Metropolitan-Area Wireless Network. *Wireless Networks*, 8:107–120, March-May 2002.
- [16] D. Kotz and K. Essien. Analysis of a campus-wide wireless network. In *The Eighth Annual International Conference on Mobile Computing and Networking (ACM MOBICOM)'02, Atlanta, GA, USA*, pages 107–118, September 2002.
- [17] M. Hollick, J. Schmitt, H.-P. Huth, and J. Sokol. MobQoS - Mobility Aware QoS Routing. Available at <http://www.kom.tu-darmstadt.de/Research/MobQoS/MobQoS.html>, August 2003.
- [18] C. Perkins. IP Mobility Support for IPv4. Proposed Standard RFC 3344, August 2002.
- [19] D. B. Johnson, C. E. Perkins, and J. Arkko. IP Mobility Support in IPv6. Internet Draft draft-ietf-mobileip-ipv6-19.txt, October 2002.
- [20] G. Huston. Next Steps for the IP QoS Architecture. Informational RFC 2290, November 2000.
- [21] D. O. Awduche, A. Chiu, A. Elwalid, I. Widjaja, and X. Xiao. Overview and Principles of Internet Traffic Engineering. Informational RFC 3272, May 2002.
- [22] 3GPP. QoS concept and architecture. TS 23.107, Release 5, Version 5.9.0, June 2003.
- [23] R. Jain. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. Wiley-Interscience, New York, NY, USA, May 1991.
- [24] K. Fall and K. Varadhan. The ns Manual, 2002. Software and manual available at <http://www.isi.edu/nsnam/ns/>.
- [25] M. Hollick, T. Krop, and J. Schmitt. Modeling mobility for cellular networks based on statistical data collection. Technical Report TR-KOM-2003-02, Darmstadt University of Technology, Multimedia Communications Lab, 2003. Available at <ftp://ftp.kom.tu-darmstadt.de/pub/TR/TR-KOM-2003-02.pdf>.
- [26] M. Hollick, T. Krop, J. Schmitt, H.-P. Huth, and R. Steinmetz. A hybrid workload model for wireless metropolitan area networks. To appear in *Proceedings of VTCi03, Orlando, FL, USA, 2003*, October 2003. In press.