QoS-Aware Mesh Construction to Enhance Multicast Routing in Mobile Ad Hoc Networks

Harald Tebbe, Andreas J. Kassler, Pedro M. Ruiz

***Abstract*— Mobile Ad-hoc Networks (MANETs) are seen as an essential technology to support future Pervasive Computing Scenarios and 4G networks. In a MANET, efficient support of multipoint communications is essential in order to provide services like group audio and video conferencing, dissemination of data to a set of receivers or collaboration of a group of users. However, most of these interactive services have very strong requirements regarding delay and bandwidth. Controlling the end-to-end delay and maintaining low packet loss rate is vital to support interactive multimedia applications in MANETs. In this paper, we present QAMNet, an approach to improve the Quality of Service (QoS) for multicast communication in MANETs. We extend existing mesh based multicast routing protocols by introducing traffic prioritization, distributed resource probing and admission control mechanisms and adaptive rate control of non-real-time traffic based on Medium Access Control (MAC) layer feedback to maintain low delay and required throughput for real-time multicast flows. Simulation results show that, by reusing control messages already used by those protocols, our approach does not significantly increase control overhead nor state stored at nodes compared while it manages to increase packet delivery ratio and reduce latency for real-time traffic.**

***Keywords* — Quality of Service, Multicast Communication Mobile Communications, Wireless Networks, Ad-hoc Networking, Pervasive Computing**

I. Introduction

Mobile Ad-hoc networking has been considered as one of the most important and essential technologies to support future Pervasive Computing scenarios [1] In fact, MANETs are witnessed as an integral part of 4G networks like the one proposed by the EU IST-DAIDALOS project[1]. A MANET is a collection of mobile nodes (MN) that communicate using wireless links without support from any pre-existing infrastructure network. Packets are delivered from a source to a destination using packet forwarding capabilities of intermediate nodes. Therefore, MNs act as both end hosts and routers. In such an autonomous system, MNs are usually free-to-move without predefined mobility patterns, making the design of efficient and scalable routing protocols a challenging task.

The increasing popularity of collaborative multimedia applications in the mobile market is making the support of advanced services like QoS or multicast a key requirement. Multicasting is therefore an essential technology to efficiently support one-to-many or many-to-many bandwidth-avid applications. In multicast communication, a source is sending only one packet with a group address as a destination. The network will be in charge of replicating that packet only when necessary to make it reach all the destinations, i.e. all the nodes that have joined the group associated with that specific group address. This leads to bandwidth savings and high scalability, which is essential in MANETs. Even with the use of multicast, multimedia applications still require QoS guarantees in terms of bounded delays and low packet loss rates. Providing strong guarantees in a MANET is really hard due to a number of factors (varying link quality and capacity, mobility, etc.) which are not under control of the routing protocol. Therefore, it is essential for routing protocols to do their best to control the congestion and manage resources for real-time multicast services in MANETs providing at least soft-QoS guarantees.

Quality of Service is mainly related to resource allocation and management, where it is necessary to decide how to allocate resources such that QoS requirements of all flows can be satisfied as best as possible. However, in MANETs several problems arise. Node mobility and channel variation lead to dynamic per node bandwidth availability. The properties of the standard 802.11 MAC and the shared medium make it very difficult to provide QoS, even for wireless Local Area Networks (LANs). Dynamic changes in topology caused by node mobility together with energy constraints lead to frequent re-routing. In addition, it is hard to estimate available resources in a multi-hop wireless environment due to node mobility and channel contention. Similarly, it is difficult to reserve resources for a packet flow, as the shared medium requires global reservation coordination. Nevertheless, QoS

Harald Tebbe is with University of Ulm, Department of Media Technology, Oberer Eselsberg, 89069 Ulm, Germany (e-mail: harald.tebbe@informatik.uni-ulm.de).

Andreas J. Kassler is with Karlstad University, Department of Information Technology, Universitetgatan 2, 65188 Karlstad, Sweden (e-mail: kassler@ieee.org).

Pedro M. Ruiz is with University of Murcia, Department of Information and Communication Engineering, Fac. Informatica, Campus de Espinardo s/n. E-30071 Murcia, Spain (e-mail: pedrom@dif.um.es).

[1] http://www.ist-daidalos.org

violations still can occur as bandwidth availability fluctuates due to dynamic contention.

In this paper, we present QAMNet, an enhancement to mesh-based multicast routing protocols to deal with real-time multimedia data. With QAMNet additional control messages are not needed compared to mesh based multicast routing protocols. Instead, intermediate nodes measure resource availability and update additional QoS-related fields that we added to standard messages of multicast routing protocols. Those fields provide QoS awareness to the mesh (or tree) creation and maintenance process. Therefore, our approach can be seen as a resource aware mechanism to build QoS enabled delivery meshes without the need for additional control messages. We distinguish between best-effort and real-time packets, being the latter the ones receiving preferential treatment. To manage resource availability, QAMNet features distributed resource probing and admission control interweaved with multicast mesh creation and maintenance. Adaptive rate control of non-real-time traffic is based on Medium Access Control (MAC) layer feedback and helps to maintain low delay and required throughput for real-time multicast flows. To cope with mobility and transient bandwidth fluctuations, QAMNet resorts to a dynamic regulation mechanism.

This paper is structured as follows. In chapter 2 we provide an overview on related work in the area of Quality of Service management for Mobile Ad-hoc networks and multicasting. In Chapter 3, we introduce our approach QAMNet for providing QoS for multicast in MANET. Chapter four evaluates our approach using simulations and finally chapter five presents some conclusions and future directions.

II. Related Work

A. QoS support in MANETs

Currently, there exist several QoS architecture proposals for MANETs. These architectures are mainly based on previous work carried out in terms of QoS support in infrastructure networks. The main idea is to maintain the fewest QoS states as possible because topology changes due to mobility make it hard to maintain reservation states overtime without high control overhead expenditures. The goal of those architectures is to provide some sort of service differentiation and soft QoS rather than hard guarantees because providing hard QoS guarantees is almost impossible given the dynamics of a MANET. In such an environment, it is a bad idea to pin resource reservations to a specific route because such reservations must be re-established whenever route changes occur.

INSIGNIA [2] is a QoS model that comprises an in-band signaling protocol with support for adaptive reservation-based services in ad-hoc networks. In-band information concerning the required resources for the flow is used by intermediate nodes to establish, maintain and restore per-flow soft-state reservations. Admission control is performed hop by hop and the destination host informs the source node of the result of the reservation using a QoS report mechanism. Reservations are maintained as long as packets associated with a particular flow are periodically received at intermediate nodes to refresh timers. INSIGNIA copes with re-routing by triggering the admission control and resource reservation through the new path while reservation state along the old path times out.

SWAN [3] is a QoS model that tries to keep the network not overloaded. Local rate control is used for best-effort traffic based on MAC delay measurements to set aside resources for real-time data. Distributed admission control is performed for real-time flows by the source, based on the result of an end-to-end request/response probe that senses the available bandwidth through the path from the source to the destination. SWAN resorts to dynamic regulation of real-time sessions when congestion/overload conditions occur (e.g due to node mobility) by sending regulate messages to the source, which might lead to a new probing request being sent to the destination.

FQMM [4] was developed for assuring a certain level of service differentiation in ad-hoc networks. It is a hybrid Integrated Services (IntServ)/Differentiated Services (DiffServ) model and supports hybrid per-flow/per-class provisioning. Source nodes are considered as ingress nodes, which use classification, marking, policing and shaping. Intermediate nodes only apply shaping. Traffic conditioning is carried out according to a traffic profile. A relative adaptive traffic profile is defined as the relative percentage of the effective link capacity, in order to keep the differentiation between classes predictable and consistent under the dynamics of the network. It is used to keep consistent differentiation between sessions (flows/aggregates).

B. Multicasting and QoS in MANET

Several protocols have been proposed for multicast routing in mobile ad hoc networks. They can be classified into tree or mesh-based depending upon the underlying forwarding structure that they use. Tree-based schemes [5, 6, 7, 8, 11] construct a multicast tree from each of the sources to all the receivers using either source based trees or shared trees. Mesh-based approaches [9, 10] compute several paths between senders and destinations. Thus, when the mobility rate increases they are able to tolerate link breaks better than tree-based protocols. Hybrid approaches [12, 13] try to combine the robustness of mesh-based ad hoc routing and the low overhead of tree-based protocols. Finally, there are stateless multicast protocols [14, 15] in which there is no need to maintain a forwarding state on the nodes as for example the set of nodes to traverse is included in the data packets themselves.

As our proposed enhancements are mainly targeted to mesh-based multicast routing protocols, we focus our discussion on the protocols falling within this category, although they could be integrated with tree-based approaches as well. For the reduction of data overhead in mesh-based multicast ad hoc routing protocols, Ruiz et al. [16] demonstrated that the problem of computing the multicast tree with minimal data overhead is NP-complete. They introduced an adaptive mobility-aware mesh construction heuristic, which managed to cut the number of relay nodes in the mesh by 50%, while maintaining nearly the same packet delivery ratio. This lightweight heuristic approximates an Steiner tree and adds additional paths only when they are needed to provide reliability due to the mobility of the network. The same authors demonstrated in [17] that the well-known Steiner tree is not optimal for the problem of minimal data overhead. In fact, for the case of static wireless mesh networks, they presented a better heuristic which outperforms existing Steiner tree heuristics.

Only very few works like [18] support QoS for multicast but those approaches introduce network state and additional signaling. The main challenge for providing QoS in Multicast MANETs is the heterogeneous nature of the branches of a multicast tree or mesh as perceived quality varies among users [19] and depends on the resource availability and mobility of a receiver. A tight integration or interaction of resource management and distributed admission control mechanisms developed for unicast QoS architectures with the multicast protocol seems to be beneficial in order to avoid overhead and minimize network state. Additional signaling packets for reservation protocol should be avoided as this contributes to network congestion, especially in high mobility and traffic-intensive scenarios.

III. QAMNET – PROVIDING QUALITY OF SERVICE FOR MULTICAST IN MANETS

A. Congestion Problem for Mesh based Multicasting

Mesh based multicasting delivery structures have proven to be robust and can effectively cope with topological changes due to mobility of nodes. However, robustness comes at a price of higher overhead due to redundant routes (i.e. a higher number of nodes added to the multicast mesh). This might lead to an early saturation of the network if multicast senders inject high traffic into the network. In order to support real-time multi-party videoconferencing, a multicast source using the H.264 [20] video codec at QCIF or CIF image size creates a multicast stream consuming between 30 Kb/s and 200 Kb/s depending on the video size and quality. However, the performance of multicast protocols in MANETs have been typically evaluated at much lower data rates (~10 kbps) which does not allow drawing any conclusions on suitability for rate intensive real-time multimedia group communication support.

Therefore, we conducted several simulation runs to evaluate how mesh based multicast routing protocols performance scale. We were interested in throughput and end-to-end delay as a function of increased traffic rate. We used the reduced data overhead version of ODMRP (denoted as ST-variant from now on) presented by Ruiz et al. [16] that already reduces the mesh density and thus data overhead. We used 50 randomly distributed stationary nodes and in total created 6 multicast groups, each group having 1 CBR source sending 512 byte packets to 2 receivers.

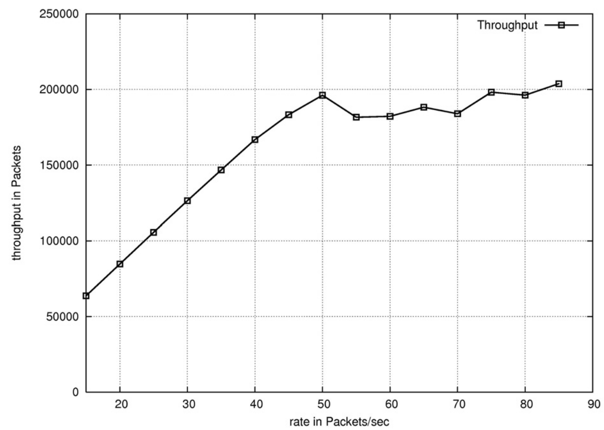

Figure 1: Throughput as function of offered traffic load

Simulation results in Figure 1 and Figure 2 show that there is an inversely proportional correlation between the throughput and the delay. With increasing traffic load first the throughput increases until the load reaches the network capacity. Increasing the load further does not lead to increase in throughput but will lead to increased buffering of packets inside nodes, increased MAC contention which results in higher packet drop. The delay stays nearly constant around 5 ms until the load reaches again the capacity then queues start building up and consequently the delay increases dramatically. The idea is therefore to control the load for real-time multicast traffic according to the QoS requirements of the multicast groups to stay below the network capacity in order to control the end-to-end delay for real time traffic (such as interactive voice and video).

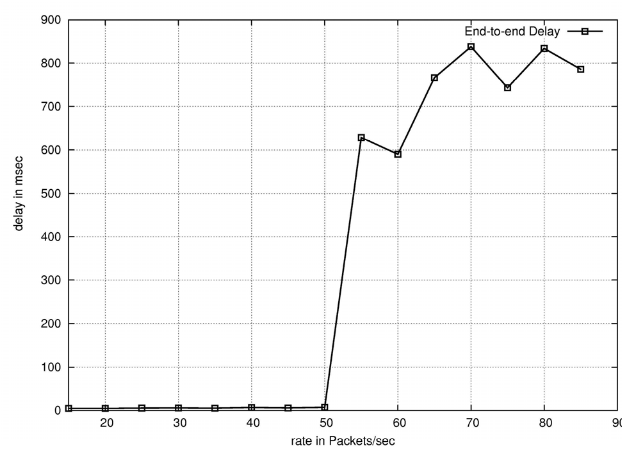

Figure 2: End-to-End Delay as function of traffic load

To do that, we need some kind of traffic differentiation to prioritize real-time flows. Also, we have to design admission control algorithms to limit the amount of real-time traffic in order to stay below the network capacity. However, in a MANET, admission control algorithms should be fully distributed because centralized control elements are not present. Some capacity should then be set aside for real-time group communication while the rest of the network capacity could be absorbed by best effort and elastic traffic like TCP. Distributed management of network resources has to be coordinated in the presence of mobility of nodes. Therefore, introducing additional signaling traffic for resource reservation seems to be harmful. The reason is that doing that, the protocol creates higher overhead in a situation where the network is already close to its capacity limit.. Thus, our approach is to integrate the QoS and resource management into the multicast delivery structure creation and maintenance algorithms in order to reduce additional overhead as much as possible. While in the rest of this paper we focus on multicast delivery meshes, our approach can easily be adapted to multicast trees.

B. *The QAMNET approach*

Our approach to provide QoS for multicast communication in Ad Hoc Networks – QAMNet – is based on a standard 802.11 MAC without the need for QoS support. The idea is to extend existing approaches of mesh based multicasting (like ODMRP [6]) and unicast QoS provisioning (like SWAN [3]) by introducing service differentiation between real-time and best-effort traffic class, distributed resource probing and admission control mechanisms as well as adaptive rate control of non-real-time traffic based on MAC layer feedback. The goal is to maintain low delay and provide the required throughput for real-time multicast flows. As we will see in the evaluation section, our approach is very scalable and does not require more states nor a significant increase in signaling than normal mesh based multicasting protocols. Control packets sent by the standard multicast routing protocols are reused, and only some bytes are added to carry QoS related information. This is a big advantage for memory and power constraint devices compared to related work, e.g. [18].

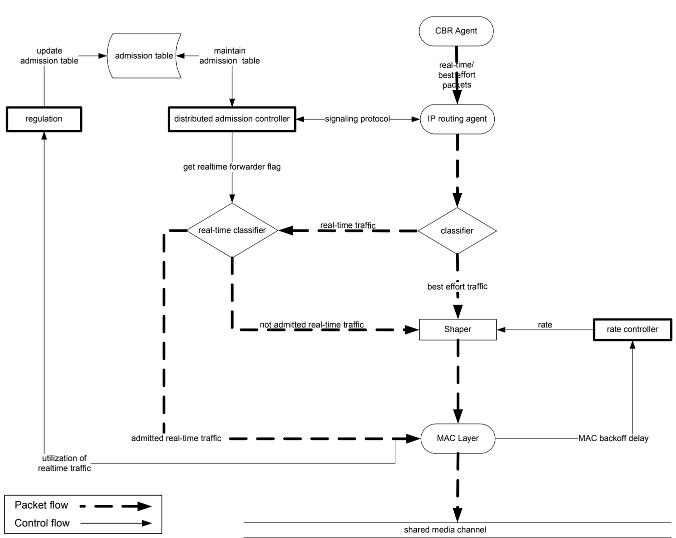

Figure 3: QAMNet Architecture within MANET nodes

The architecture of QAMNet enabled MANET nodes is depicted in Figure 3. When a MANET node has to forward a packet it enters first into a classifier. Admitted real-time packets bypass the local shaper and are directly forwarded to the MAC layer. A local rate controller is used to control the amount of best effort traffic and non-admitted real-time traffic that is injected through a packet shaper into the MAC layer. A distributed Admission Controller is used to maintain reservation information for real-time multicast flows based on an admission table, which is connected to the multicast forwardeing table of the multicast protocol. We distinguish between admitted real-time flows, which are flows for which enough resources are available during mesh creation phase, non-admitted real-time flows, which are flows that belong to the real-time traffic class but not enough resources are available in the network to accommodate such traffic, and best-effort traffic. Packets belonging to non-admitted real-time flows have to traverse the shaper just like the best effort packets. However the admitted real-time packets bypass the shaper. The goal of the shaper is to delay best effort and non-admitted real-time packets in conformance with the rate calculated by the rate controller, based on information on congestion in the vicinity of the node gathered through feedback from the link layer.

A key operation of the admission controller is to efficiently estimate local bandwidth availability. In QAMNet the source initiates a signalling process to initiate the distributed admission control at the beginning of a real-time multicast session, interweaved with the mesh creation process. During that step, the admission control estimates locally available bandwidth and decides whether to admit the request for QoS for the new multicast session or not. If the request can be admitted, an additional flag is used to indicate that all packets for the admitted flow should be treated as admitted real-time packets and thus bypass the shaper. In case of mobility of nodes or long term fading that reduces the capacity, a regulation mechanism and a periodic signalling process is used to re-establish the QoS provisioning along with the mesh maintenance. In the following, mechanisms of the control algorithms and QoS extensions to the mesh creation and maintenance are described in more detail.

C. *QoS enabled mesh creation in QAMNET*

When a QAMNet node in a MANET has real-time traffic to send to a multicast group, it starts with flooding the entire network with a control message to advertise the multicast source to receivers, which carries the first data packet using piggybacking. In contrast to ST-variant of ODMRP, where the control/signaling information is composed of a Join-Query message and a counter of relay nodes, we additionally extend it with a probing request, which contains bottleneck bandwidth (BB) and required bandwidth (RB) fields. We refer to the first data packet as the Join-Probe packet. The source broadcasts a Join-Probe packet and upon reception of the first, non-duplicate, Join-Probe packet, intermediate nodes set pointers towards their upstream nodes and rebroadcast it, after modifying the probing request information. To do so, each intermediate node additionally updates the bottleneck bandwidth field, if the local bandwidth availability at the given node is lower than the current value. Bandwidth availability at the local node is calculated similar to SWAN [3].

Once a Join-Probe packet reaches a multicast receiver, BB indicates the bottleneck bandwidth found along the path. The receiver waits a small time period MAX_JOIN_WAIT_TIME to collect all Join-Probe packets received from other branches of the multicast mesh. The receiver evaluates if the BB with the largest value is greater than RB and if so creates a Join-Reply, piggybacking a Probe-Response which contains the largest BB and the same value in the RB field that it received in the Join-Probe.

The Join-Reply is relayed by the intermediate nodes all the way from the sink to the source following the pointers established during the propagation of Join-Probes. However, each intermediate node on the path from the receiver to the source does not forward a received Probe-Response immediately. Instead, the node waits for a short time MAX_PROBE_WAIT_TIME in order to collect Probe-Response packets from other branches of the multicast mesh. When this timer expires, the node updates the bottleneck bandwidth in the Probe-Reply packet with the maximum value of all received Probe-Response messages when forwarding to the source. It also sets a real-time forwarder flag (RTF_FLAG) for the given multicast group if the forwarded BB value is larger than RB. Note, that at this point the node already has set a state in order to be a multicast forwarder (FG_FLAG) for the group, so we just add one more flag. In this way the forwarding mesh is constructed in a similar way than the ST-variant of ODMRP [16] .

Note, that the Join-Probe does not introduce additional control packet overhead compared to ODMRP as it is piggybacked on ODMRPs Join-Query message, which is flooded throughout the network. Only a few bits together with the RB and BB fields are required. In a similar way, the Probe-Response is piggybacked on ODMRPs Join-Reply messages. Thus, it does not contribute to additional packet overhead. In that sense, Join-Probe and Probe-Response messages are disseminated periodically throughout the network gathering information on the resource availability of individual nodes at the same periodicity as the messages that are responsible for the mesh creation and maintenance (Join-Query and Join-Reply) without creating additional signalling packets. In our current implementation we flood these messages periodically (e.g. at an interval of 3 sec.) but that could change depending on the dynamicity of the network in terms of topology changes, e.g. using mechanisms similar to Motion Adaptive Refresh [9].

Once the Join-Reply reaches the source, it multicasts (real-time) packets with the help of the (real-time) forwarder nodes through the forwarding mesh. For all packets with real-time constraints, the source sets the Type of Service (ToS) bit in the IP-header and sends it via MAC-layer broadcast. Before an intermediate forwarding node re-broadcasts the packets, the classifier of that node checks, if the RTF_FLAG for the given group is set. If it is set, the packet bypasses the nodes' shaping mechanism, remains unregulated and is directly passed to the MAC layer for re-broadcasting. If the RTF_FLAG is not set, the node will set the ToS bit in the header to zero and put the packet into the shaper, if the FG-flag for the given group has been set through the reception of the proper Join-Reply.

D. Regulation of Best-Effort traffic in QAMNET

Each QAMNet node in the MANET regulates independently best effort and rejected real-time traffic. That is, if the BB field is lower than the RB one, RTF_FLAG was not set and all those packets have to enter the shaper in forwarding nodes. To discriminate among them, a traffic regulator is used. Unicast QoS mechanisms for MANETs (like SWAN [3]) that are based on standard 802.11 DCF MAC use MAC layer feedback based on RTS-CTS-DATA-ACK sequence in order to derive the MAC layer utilization and congestion in the neighborhood of the node. This information is then used e.g. in SWAN to regulate the shaping rate of the regulator. Multicast mesh based protocols on the other hand, employ MAC layer broadcasting without the final ACK and RTS/CTS sequence. Instead, a variant of the 802.11 DCF is used that only involves a local binary exponential back-off (BEB) procedure in case where the channel is occupied during the carrier sense process.

The idea for QAMNet is to still reuse MAC layer information and infer MAC layer utilization based on the *local back-off delay* of 802.11 DCF as feedback. In order to verify that this approach is viable, we also analyzed the local back-off delay in the experiment from section III.A. As we can see in Figure 4, the local MAC delay that includes the time spent in the MAC layer for the binary exponential back-off first increases linearly while during network congestion it increases exponentially. Once the rate is increased beyond the capacity, the local MAC delay does not increase more as the BEB bounds the maximum back-off time.

Therefore, in QAMNet, if the MAC-layer back-off increases, we decrease the rate at which BE and rejected real-time traffic enters the MAC-layer at each forwarding node, otherwise we increase it according to an Additive Increase Multiplicative Decrease (AIMD) control algorithm. We introduce AIMD here because we want to achieve fairness and efficiency in allocating resources in a distributed way.

The QAMNet AIMD rate control algorithm is inspired by SWANs rate control mechanism [3] and the only difference is the threshold delay calculation. Every T seconds, each mobile node increases its transmission rate for the BE and rejected real-time packets in the queue gradually (additive increase with increment rate of c Kbps) until the local MAC backoff delay of packets becomes excessive (one or more packets experience MAC backoff delays greater than MAC_BO_DELAY sec). As soon as the rate controller detects excessive delays, it backs off the rate (multiplicative decrease by r percent). The threshold delay

MAC_BO_DELAY is based on the real-time delay requirements of applications in wireless networks, as discussed in [22]. Several simulation runs have suggested to use an effective value of MAC_BO_DELAY=0.002s, while the optimal increment rate is *c*=10 Kbps/s and decrement rate is *r*=50%.

The Rate Controller of each node monitors the local MAC backoff delay and adjusts the shaping rate every T seconds. The period T should be small enough to be responsive to the dynamics of Mobile Ad hoc Networks. If there is a large difference between the shaping rate and the actual transmission rate, then a mobile node is capable of transmitting a burst without much control thus potentially limiting the performance of real-time traffic in the neighborhood. To resolve this conflict, the rate controller monitors the *actual* transmission rate. When the difference between the actual shaping rate and the actual transmission rate is greater than *g* percent of the actual rate, then the rate controller adjusts the shaping rate to be *g* percent above the actual rate. This "gap" allows the best-effort traffic to increase its actual rate gradually [3].

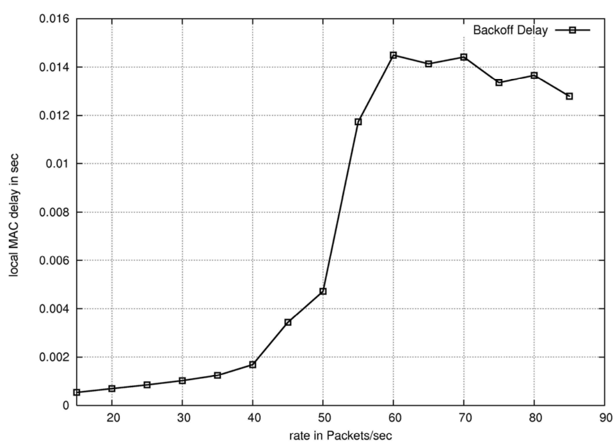

Figure 4: Local nodes MAC backoff delay as function of offered traffic load

We use this distributed regulation mechanism to control the amount of BE and rejected RT traffic a node injects into the MANET based on the load of other nodes that compete for resources locally. We denote it as distributed as the back-off delay is a function of the contention and thus on the amount of traffic in the neighborhood of a node. This also helps to set aside resources for the real-time packets. Although the backoff mechanism employed by 802.11 has not been shown to be effective and fair in a MANET environment, it gives important information on the congestion state in the environment of a node. As the rate-regulation of the traffic regulator is based on that scheme, this could lead to somewhat sub-optimal performance. We are currently evaluating alternative approaches to improve the fairness of our mechanisms.

E. Coping with mobility and transient bandwidth fluctuation in QAMNet

In order to cope with false admission and bandwidth availability fluctuation that might cause flows to be admitted while there are in fact no resources left, we include dynamic regulation of real-time traffic. Each node monitors utilization of its real-time traffic class when estimating available resources not only during the Probe-Request phase but also periodically. When it detects violation, e.g. more resources are actually consumed than have been set aside for the real-time class, it selects randomly one of its real-time flows and sets congestion experienced (CE) bit in all packets belonging to the given flow. Also, RTF_FLAG for that flow is set to zero which results in the next real-time packet of the given flow to enter the shaper. QAMNet uses soft states to maintain the RTF_FLAG and sources periodically transmit Join-Probe packets. A node with an already activated RTF_FLAG copies infinity into the BB-field because the session has already been admitted. Signaling messages are treated with high priority and bypass the local shaping mechanism, too.

These mechanisms also help to cope with mobility of nodes. When nodes are mobile and routes break, the mesh structure leads to a higher probability that alternative paths are available. Also, as nodes periodically flood the Probe-Request (as it is also done with ODMRP), the mesh structure will be updated periodically with the mobility of nodes. That results in new nodes participating in the mesh, which also participate in the processing of the Probe-Request packets. That helps in establishing the reservation states according to the RTF_FLAG for the given real-time packets which will then bypass the shaper at new nodes within the modified mesh thus getting highest priority. However, there might be some time where reservation states at those new nodes are not yet established as we use a 3 sec. interval to maintain the mesh structure. During this time, nodes might experience bandwidth fluctuations due to mobility of nodes. QAMNet deals with this situation by using the regulation mechanisms described above.

IV. Performance Evaluation

We have implemented QAMNet as an extension of the ODMRP protocol[9] and its ST-variant (ODMRP-ST) based on the code from [16]. We have simulated both the original ODMRP and ODMRP-ST as well as our QAMNet extensions using NS-2 [23].

A. Basic Simulation Setup

Throughout all our simulations the channel capacity was fixed to 2 Mbit/s using standard 802.11b DCF MAC layer and the communication range was set to 250 m. Each MN waits MAX_JOIN_WAIT_TIME=0.025 s before sending a Join-Reply to collect other Join-Probes received via other paths and each MN waits MAX_PROBE_WAIT_TIME=0.015s before propagating an updated Probe-Reply towards the source. Join-Probes with piggy-backed probing requests are sent every three seconds by the source, regardless of the mobility of nodes. We set the threshold for the MAC backoff-delay

MAC_BO_DELAY= 0.002 s, which influences the AIMD rate control mechanism for the BE-traffic.

We randomly distributed 50 nodes over an area of 1500m x 300m. The nodes move using a random waypoint mobility model with varying pause times at an average speed of 10m/s. We considered recommendations in the literature to avoid the concentration of nodes in the middle of the simulation area when using this model. At the begining of the simulation, nodes are static for pause time seconds and move then to a random destination inside the simulation area at a speed uniformly distributed between 0 and 20 m/s (mean speed = 10m/s). Once they reach the destination this behavior is repeated until the end of the simulation. We used seven different pause times: 0, 50, 100, 150, 200, 250, and 300 seconds. For each pause time, five different scenarios where simulated. The results were obtained as the mean values over these 5 runs. We used one CBR real time and two non-real time sources each with 15 receivers. Each of the sources generates 330 byte packets at a rate of 45 packets per seconds (118,8kbps), a typical video conferencing data rate. As we used only one RT-group, we did not test the effectiveness of our admission control strategy in detail. This is one of our future work topics.

B. Simulation Results

As we can see in Figure 5, the average delay of real-time packets (QAMNet RT) can be controlled by QAMNet efficiently and is bounded between 15 and 50 msec only marginally increasing with higher mobility. This is the result of controlling the load of the real-time packets at the MAC layer through the usage of distributed admission control and the regulation of best-effort traffic of the shaper. The drawback of our approach is the additional delay for the best-effort packets (QAMNet-BE) as those packets are regulated by the shaper. As there are much more BE packets than RT packets in our simulation, the average delay over all packets is marginally higher when using QAMNet compared to ODMRP (denoted as ODMRP-BE) as the high traffic leads to high delay due to MAC contention and long queues for BE traffic. As we can see, the ST-variant of ODMRP has a lower delay than original ODMRP due to its reduction in the number of forwarding nodes (i.e. contention). However, by not being QoS-aware, it cannot offer the low delay for real-time traffic that we offer with QAMNet.

The delivery ratio of real-time packets (RT) is significantly higher than for BE packets and stays between 68 and 84%. In contrast, the delivery ratio for the BE packets is lower than when using ODMRP or the Steiner tree variant. This is because BE packets are dropped once the shaper queue is full and cannot accept more BE packets. But, as we can see in Figure 6, the average delivery ratio over all packets is higher when using QAMNet than using standard ODMRP. This is also due to the preferential treatment of QAMNet control messages which bypass the shaper and are thus treated as real-time packets. The ST variant of ODMRP has a slightly better overall packet delivery ratio due to its lower data overhead, but it does not manage to offer good conditions to support multimedia communications in the scenario with congestion that we simulate. The good point with QAMNet is that, although its overall packet delivery ratio is a little bit lower than the ST-variant of ODMRP, non-elastic traffic is given enough bandwidth and higher packet delivery ratio, while elastic bandwidth will manage to adapt to remaining bandwidth, decreasing thus its PDR.

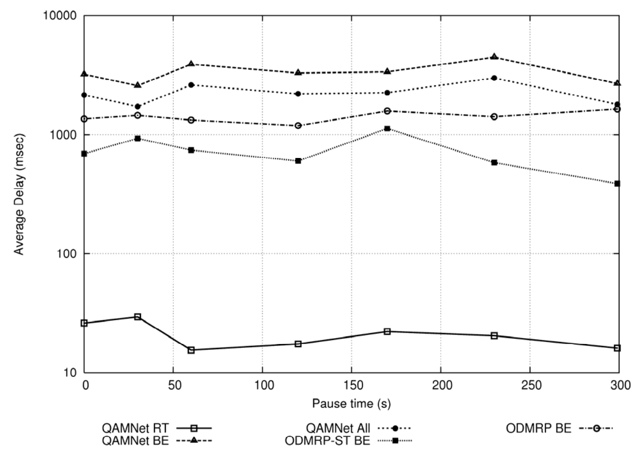

Figure 5. Average Delay in msec as a function of pause time.

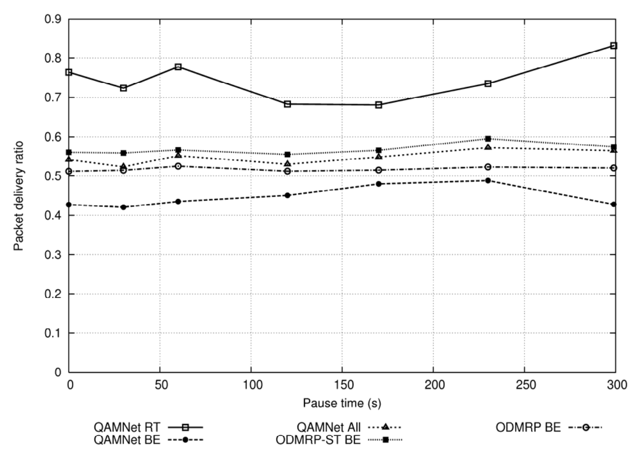

Figure 6. Packet Delivery Ratio as a function of pause time.

The preference given by QAMNet to control traffic is also visible in Figure 7, where we plot the normalized overhead as a function of the pause time. Using QAMNet limits the overhead to stay below one for all scenarios whereas ODMRP shows higher normalized overhead (between 1.4 and 1.7). The overhead of the Steiner Tree variant (which minimizes the number of forwarding nodes in the multicast mesh) is lower than plain ODMRP but still higher than using QAMNet approach. The preferential treatment of control packets in QAMNet leads to higher routing stability and thus to lower re-transmission of control packets which reduces the overhead. In addition, QAMNet does not introduce additional signalling packets compared to ODMRP. This shows the scalability of our approach wrt. mobility, where QAMNet can control the overhead effectively and maintain low delay for real-time traffic.

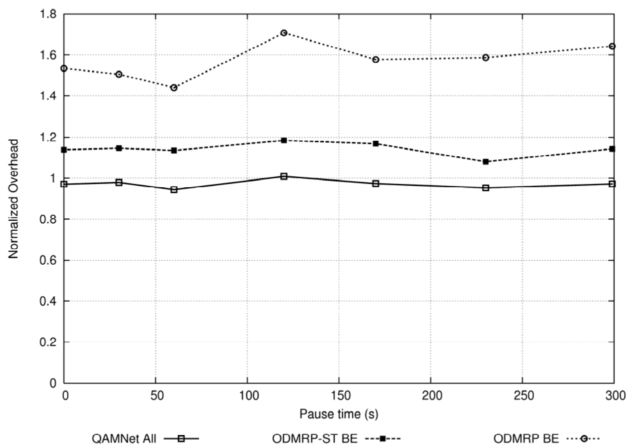

Figure 7. Normalized Overhead as a function of pause time.

V. Summary and Conclusion

We have developed a mechanism to provide Quality of Service for Multicasting in Mobile Ad-hoc networks. Our approach, called QAMNet, extends existing approaches of mesh-based packet forwarding in MANET and adds distributed resource probing and admission control, which is interweaved with the mesh creation process. In that sense, our approach can be categorized as a resource-aware mesh creation scheme. We also use local traffic regulation of non real-time flows in order to control the delay of real-time packets. We cope with mobility of nodes and transient bandwidth fluctuation by resorting to a regulation process. Our approach is scalable as it does not require significantly more state than standard mesh-based multicasting protocols. We do not introduce additional control packets compared to mesh based multicast routing protocols as we re-use signaling packets designed originally for mesh creation and maintenance and piggyback resource requirements and probing information. Our evaluation showed that when using QAMNet the delay and packet loss rate of multicast real-time packets in mobile Ad-hoc networks can be significantly reduced and limited to at most 50 msec over the whole range of mobility of the nodes. In addition, the overall overhead is reduced by 60% at the expense of increased delay and packet loss for best effort traffic. This effect is more visible at higher load within the MANET.

It seems natural to extend our work and combine it with the standard SWAN approach so that we can support QoS for both unicast and multicast communication. That would require a generic interface between the MAC and routing layer to provide feedback on MAC layer utilisation to the routing protocol and to adapt the shaping rate. SWAN works with any MANET unicast routing protocol as the probing messages are separated from the routing protocol. However, in our approach, we integrated the probing with the multicast routing in order to reduce the overhead introduced by probing messages for the multicast scenario. In our future work we will extend the simulation by including other mobility and traffic models. We will also make the probing requests adaptive to change with mobility and traffic patterns. In addition, delay jitter is also an important parameter to assess Quality of Service management mechanisms for e.g. audio streams. We will consider that parameter in future simulation runs. Finally, we plan to implement QAMNet on small PDAs and test it in a real setup. We will also extend our approach to cover hybrid MANETs, i.e. mobile Ad-hoc networks that are connected with the public internet using one or more gateways. Finally, we plan to extend MAODV in a similar way so that we can also support QoS for tree-based multicast routing protocols in a MANET.

Acknowledgment

The work described in this paper is based on results of IST FP6 Integrated Project DAIDALOS, which receives research funding from the European Community's Sixth Framework Programme. Apart from this, the European Commission has no responsibility for the content of this paper. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

The work by Pedro M. Ruiz has been funded by the "Ramon y Cajal" workprogramme of the Spanish Science Ministry.

References

[1] Jun-Zhao Sun, "Mobile Ad-hoc Networking: An Essential Technology for Pervasive Computing." In: Proceedings of International Conferences on Infotech & Infonet, Beijing, China, C:316 – 321

[2] Seoung-Bum Lee, Gahng-Seop Ahn, Xiaowei Zhang, and Andrew T. Capbell, "INSIGNIA: An IP-Based Quality of Service Framework for Mobile Ad-hoc Networks." In: Journal Parallel and Distributed Computing, vol. 60 nº4, Apr. 2000, pp. 374-406.

[3] Gahng-Seop Ahn, et al, "Supporting Service Differentiation for Real-Time and Best-Effort Traffic in Stateless Wireless Ad-hoc Networks (SWAN)." In: IEEE Transactions on Mobile Computing, vol. 1, no. 3, 2002.

[4] H.Xiao, K.Chua, W.Seah and A.Lo, "A Flexible Quality of Service Model for Mobile Ad-hoc Networks." In: Proceedings of Vehicular Technology Conference (VTC), Tokyo, Japan, May 2000, pp. 445-449.

[5] C.-W. Wu, Y.-C. Tay and C.-K. Toh, "Ad Hoc Multicast Routing Protocol Utilizing Increasing idnumberS (AMRIS) Functional Specification," Internet-draft, work in progress, draft-ietf-manet-amrisspec-00.txt, November 1998.

[6] E.-M. Royer and C.-E. Perkins, "Multicast operation of the ad-hoc on-demand distance vector routing protocol," Proceedings of ACM/IEEE MOBICOM'99, Seattle, WA, August 1999, pp. 207–218.

[7] L. Ji and S. Corson, "A Lightweight Adaptive Multicast Algorithm," Proceedings of IEEE GLOBECOM'98, Sydney, Australia, November 1998, pp. 1036–1042.

[8] J.-G. Jetcheva, D.-B. Johnson, ”Adaptive Demand-Driven Multicast Routing in Multi-Hop Wireless Ad Hoc Networks,” Proceedings of ACM MobiHoc’01, Long Beach, CA, October, 2001, pp. 33–44.

[9] S.-J. Lee, W. Su, and M. Gerla, “On-demand multicast routing protocol in multihop wireless mobile networks”, ACM/Kluwer Mobile Networks and Applications, vol. 7, no. 6, pp. 441–452, December 2002.

[10] J.J. Garcia-Luna-Aceves and E.-L. Madruga, ”The Core-Assisted Mesh Protocol,” IEEE Journal on Selected Areas in Communications, vol.17, no.8, August 1999, pp. 1380–1394.

[11] C.-K. Toh, G. Guichala, S. Bunchua, ”ABAM: On-Demand Associativity-Based Multicast Routing for Mobile Ad hoc Networks,” Proceedings of IEEE VTC-2000, Boston, MA,September 2000, pp.987–993.

[12] E. Bommaiah, M. Liu, A. MacAuley and R. Talpade, ”AMRoute: Ad hoc Multicast Routing Protocol,” Internet-draft, work in progress, draft-talpade-manet-amroute-00.txt, August 1998.

[13] P. Sinha, R. Sivakumar and V. Bharghavan, ”MCEDAR: Multicast Core-Extraction Distributed Ad hoc Routing,” Proc. of IEEE Wireless Commun. and Networking Conf., New Orleans, LA, September 1999, pp.1313–1319.

[14] L. Ji, and M.-S. Corson, ”Differential Destination Multicast: A MANET Multicast Routing Protocol of Small Groups,” Proc. of IEEE INFOCOM’01, Anchorage, Alaska, April 2001, pp. 1192-1202.

[15] J. Jetcheva, Y. Hu, D. Maltz, D. Jhonson, “ A simple protocol for multicast and broadcast in mobile ad hoc networks,” Internet-draft, Work in progress, draft-ietf-manet-siimple-mbcast-01.txt, July 2001.

[16] P. Ruiz and A. Gomez-Skarmeta, “Mobility-aware mesh construction algorithm for low data overhead in multicast Ad-hoc routing”. In: Journal of Communications and Networks (JNC), vol. 6, no. 5, Dec. 2004.

[17] Pedro M. Ruiz and Antonio F. Gomez-Skarmeta, “Approximating Optimal Multicast Trees in Wireless Multihop Networks,“ In 10th IEEE Symposium on Computers and Communications, La Manga del Mar Menor, Spain, June 2005, pp. 686-691.

[18] K. Bür, C. Ersoy, “Multicast Routing for Ad-hoc Networks with a Multiclass Scheme for Quality of Service”. LNCS, ISCIS 2004, Antalya, October 2004.

[19] B. Wang and J. C. Hou, “Multicast Routing and Its QoS Extension: Problems, Algorithms, and Protocol”. In: IEEE Network, Vol 14, 01/2001.

[20] T. Wiegand, G. J. Sullivan, G. Bjontegaard, A. Luthra, “Overview of the H.264 / AVC Video Coding Standard”. In: IEEE Transactions on Circuits and Systems for Video Technology, July 2003.

[21] Soon Y. Oh, Joon-Sang Park, Mario Gerla: “E-ODMRP: Enhanced ODMRP with Motion Adaptive Refresh”, in Proceedings of ISWCS (International Symposium on Wireless Communication Systems) 2005, September 5-7, 2005. Siena, Italy.

[22] A. Veres, A.T. Campbell, M. Barry, and L.-H. Sun. “Supporting Service Differentiation in Wireless Packet Networks Using Distributed Control”. IEEE Journal on Selected Areas in Communications, vol. 19, no. 10, October 2001, pp. 2081 - 2093.

[23] K. Fall, K. Varadhan, “ns, Notes and Documentation”. The VINT Project, UC Berkeley, LBL, USC/ISI, and Xerox PARC, November 2003