Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Performance Evaluation 62 (2005) 193–209
PERFORMANCE EVALUATION
An International Journal
www.elsevier.com/locate/peva

Scalability of routing methods in ad hoc networks

Valery Naumov*, Thomas Gross

ETH Zurich, 8093 Zurich, Switzerland

Available online 10 August 2005

Abstract

In an ad hoc network each host (node) participates in routing packets. Ad hoc networks based on 802.11 WLAN technology have been the focus of several prior studies. These investigations were mainly based on simulations of scenarios involving up to 100 nodes (usually 50 nodes) and relaxed (too unrealistic) data traffic conditions. Many routing protocols in such setting offer the same performance, and many potential problems stay undetected. At the same time, an ad hoc network may not want (or be able) to limit the number of hosts involved in the network. As more nodes join an ad hoc network or the data traffic grows, the potential for collisions and contention increases, and protocols face the challenging task to route data packets without creating high administrative load. The investigation of protocol behavior in large scenarios exposes many hidden problems. The understanding of these problems helps not only in improving protocol scalability to large scenarios but also in increasing the throughput and other QoS metrics in small ones. This paper studies on the example of AODV and DSR protocols the influence of the network size (up to 550 nodes), nodes mobility, nodes density, suggested data traffic on protocols performance. In this paper we identify and analyze the reasons for poor absolute performance that both protocols demonstrate in the majority of studied scenarios. We also propose and evaluate restructured protocol stack that helps to improve the performance and scalability of any routing protocol in wireless ad hoc networks.

Keywords: Ad hoc networks; Routing; Protocols; Scalability; Simulation; ARP

* Corresponding author.
E-mail address: naoumov@inf.ethz.ch (V. Naumov).

0166-5316/$ – see front matter
doi:10.1016/j.peva.2005.07.024

1. Introduction

Ad hoc (or self-organizing) networks operate without a predefined fixed (managed) infrastructure. Ad hoc networks using 802.11-based WLAN technology have recently received considerable attention, and a number of routing protocols have been proposed and evaluated. Ad hoc on-demand distance vector (AODV) routing and dynamic source routing (DSR) are two protocols that have been compared in prior studies, using the NS-2 simulator [2,3]. A crucial issue for ad hoc networks is the handling of a large number of nodes (hosts). As more nodes join an ad hoc network, contention is more likely. And the open nature of an ad hoc network makes it important that a network continues to operate even if there are more nodes involved than have been considered in previous studies (many of those consider 50–100 nodes).[1] We do not want to promote here large ad hoc networks. We just want to draw researchers' attention to the fact that, while suggesting/evaluating a new routing method, one should not stop at 50 or 100 nodes (unless there are clear reasons for that), since that may not uncover mistakes or potential problems.

Analytical performance models also point to the importance of network size (i.e., the number of nodes) for network performance. Gupta and Kumar [6] studied the transport capacity of random ad hoc networks. They show that for a network with n identical nodes, the total one-hop capacity is $O(n)$, and the maximum achievable end-to-end throughput is $O(1/\sqrt{n})$. In other words, if the density of nodes is constant, then the throughput decreases with the size of the network as the square root of n. Gastpar and Vetterli [5] consider the relay case for ad hoc networks, i.e., there is only one active traffic source and one destination. All other nodes act as relays assisting the single transmission. Total network capacity grows only as $O(\log n)$ [bits per second] in the relay networks studied.

Since there are a number of parameters that influence the performance of a routing protocol (number of nodes, area respectively density, number of sources, bit rate), simulations offer the best approach to investigate these networks although all simulators provide only a simplified view of reality. Like earlier studies, we use NS-2 [10] since it supports the popular WaveLAN cards to study the performance of AODV and DSR in the areas of 2121 m × 425 m, 3000 m × 600 m, 3675 m × 735 m, 4250 m × 850 m, and 5000 m × 1000 m populated by 100, 200, 300, 400, and 550 mobile nodes, respectively.[2] Each chosen combination of area and number of resident hosts keeps the density of nodes constant. Investigation of scenarios with more than 550 nodes do not appear valuable at this time since the performance of both protocols drops so significantly for larger setups that any further increase in the number of nodes would hardly make any sense. We also analyze the influence of the hosts density on the protocols performance varying the density of nodes from 10 to 110 nodes/km^2 in the areas of 5000 m × 1000 m and 3000 m × 600 m.

In this paper we offer a comparative study of both protocols (Section 3) together with an investigation of the main factors that affect scalability. We present a detailed analysis of the reasons that prevent both protocols from functioning effectively, not only pointing the problems, but also explaining why and how they occur (Section 4). We also show that looking at the traditional metrics does not help to understand the source of the problems a routing protocol suffers from. In Section 4 we suggest additional metrics that help to answer this question. Based on the results of our studies, we also propose and evaluate restructured

---

[1] There are other studies that simulate AODV alone (using other simulators) on large scenarios [11,12]. But there is no comparative analysis presented, and since their investigations of AODV performance not as detailed as presented here, they offer different incites.

[2] NS-2 simulator was validated in [8] and verified in a number of later publications, e.g. [13].

protocol stack that helps to improve the performance and scalability of any routing protocol in wireless ad hoc networks (Section 5).

2. Simulation environment and experimental setup

The goal of our experiments is to study the performance of DSR and AODV in large ad hoc wireless networks with varying number of nodes, different movement and communication models. We choose AODV and DSR from the wireless routing protocols as they are the most attractive protocols according to a study that compared DSDV, TORA, DSR, and AODV [2]. AODV and DSR are both reactive, but perform different kinds of routing: table-driven (AODV) and source routing (DSR). We concentrate on reactive routing protocols since previous studies [2,7] report that these protocols generate less routing overhead than proactive protocols and thus have a higher potential for scalability.

In our experiments we use the last release of NS-2 (ns2.27)—a discrete event simulator widely used in the networking research community [4]. NS-2 simulator was validated in [8] and verified in a number of later publications, e.g. [13]. It contains a detailed model of the physical and link layer behavior of a wireless network based on 802.11 [1] and allows arbitrary movement of nodes within a network area. Some of the recently proposed wireless routing protocols (DSDV, TORA, DSR and AODV) are also integrated into NS-2. Each run of the simulator accepts as input a scenario file that describes the exact motion of each node together with the sequence of packets originated by each node as time progresses. We use the default setups for DSR and AODV routing protocols with enabled optimizations designed for better scalability.

To allow comparison with other experiments [3,2], we use 512 byte data packets and CBR traffic. Each node uses the random waypoint model [2][3] and moves at a speed randomly chosen between 1 and 15 m/s (between human slow walk and average allowed maximum car speed in the city area). We simulate populations of 100, 200, 300, 400, and 550 nodes [4] in areas of 2121 m × 425 m, 3000 m × 600 m, 3675 m × 735 m, 4250 m × 850 m, and 5000 m × 1000 m for 600 s of simulation time. We choose the above combinations of areas and number of nodes involved to work with approximately the same node density and simulation area proportions as the earlier studies [2,3]. This density of nodes is high enough to allow a meaningful comparison of the protocols; a markedly lower density may cause the network to be frequently disconnected, and then an investigation of the efficiency of different routing protocols is even more complicated. To prove it, we study the influence of hosts density by varying the number of simulated nodes in the areas of 3000 m × 600 m and 5000 m × 1000 m to model the densities from 10 to 110 nodes/km^2.

According to our experience and previous studies, in the highly mobile scenarios the difference in the protocols performance is the most pronounced. The higher the mobility rate, the more challenging is the task for a protocol to obtain and maintain routes. In stationary networks we can test the ability of a protocol to quickly discover a route and to effectively deliver data. Since no route breaks occur due to

[3] A node begins the simulation by waiting for *pause time* seconds. Then it selects a random destination within the simulation area and moves there with a speed randomly chosen from the above mentioned range. At the time the node reaches the destination, it pauses again for pause time seconds, chooses another destination, and proceeds the same way for the duration of the simulation.

[4] Different scenario files (generated by the scripts *cbrgen.tcl* and *setdest* that are provided with the NS-2 distribution) contain various traffic loads and movement patterns. Researchers interested in a repetition of our experiments should also use these scripts.

nodes mobility, the administrative load should be minimal and should not cause long delays or losses of data packets. That is why with 100, 200, 300, and 400 node we report only the data for the scenarios with constant motion of nodes (0 pause time) and when the hosts are idle (600 s pause time). With 550 we present the influence of mobility rate on overall protocols performance.

In addition to the results for 10 and 40 CBR traffic sources, we also report the results for the scenarios when 30% of the nodes act as traffic sources (for 550 nodes we then have 160 CBR sources). This number of sources allows us to investigate scalability of protocols when the traffic load depends on the size of the network. Such a setting is more realistic than the settings with a fixed number of sources, no matter how large a network is operating.

All peer-to-peer connections are started at times uniformly distributed between 0 and 50 s. The number of unique traffic sources is 70% of the total number of sources. The chosen sending rate is 2 packets/s. Each data point presented in this paper is an average of five runs, each lasting for 600 s of simulated time. We use the suggested parameters to model 914 MHz Lucent WaveLAN DSSS radio interface at a 2 Mb/s data rate.

3. Simulation results

We present several metrics to capture the performance of the routing protocols. Some of these metrics are well known in evaluation of protocols performance [2,3], but we also use others that are essential for understanding of protocols performance (or non-performance) in classical metrics. Due to space limitations we have to skip some results here, leaving only most essential ones. These results can be found in our technical report [14].

- *Routing overhead in packets* (*ROP*) *and in bytes* (*ROB*)—the total absolute number of routing packets transmitted during the simulation and the total amount of bytes transported for routing (in these routing packets). Again each hop in a multi-hop route is counted separately (for both metrics).
- *Packet delivery fraction* (*PDF*)—the fraction of the data packets originated by an application that each routing protocol delivers.
- *Average end-to-end delay* of data packets—the average difference between the time a data packet is originated by an application and the time this data packet is received at its destination. The time expanded to obtain a route (if there is none in the sender's routing table) is included in this figure.

In addition we present other measurements that shed light on the protocols' performance in large ad hoc networks. We present the experimental results in the above order. The routing load is the key metric for assessing the performance of any routing protocol. We have not yet seen a protocol for ad hoc networks which has a high routing overhead but nevertheless shows excellent data throughput. In large networks the level of contention is so high that an administrative load becomes an important issue and is a key contributor to all problems for routing algorithms.

3.1. Influence of the number of nodes

3.1.1. ROP

Fig. 1 shows the ROP as a function of both number of nodes and network load for different numbers of traffic sources. Each chart presents ROP of both protocols in a highly mobile (0 s pause time) and

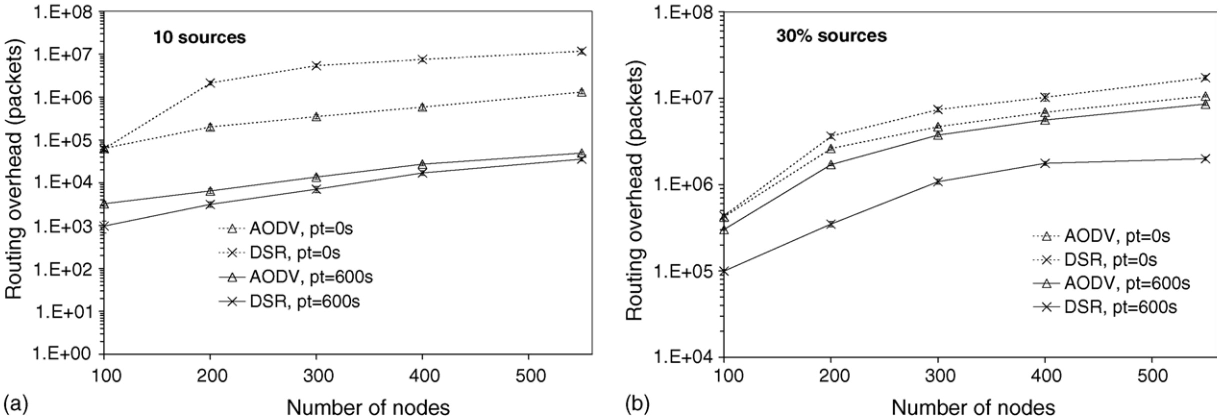

Fig. 1. Routing overhead, in packets (100, 200, 300, 400, 550 nodes): 10 (a) and 30% (b) CBR traffic sources. Log scale.

static (600 s pause time) scenarios. In the static scenarios with 10 CBR sources (Fig. 1a) both protocols generate a moderate routing load that shows little (adequate) dependence on the number of nodes. DSR produces around two times less load than AODV (the difference is more pronounced with low number of nodes). With high mobility rate DSR quickly loses its superiority and already at 200 nodes sends an order of magnitude more administrative packets than AODV. The reasons for such a behavior will be analyzed in Section 4. With 30% of the nodes acting as CBR sources the ROP of both protocols increases by 2 orders of magnitude (more with higher number of nodes). With 0 s pause time ROP for AODV is 50–100% lower than the ROP for DSR; whereas in stationary networks the situation is opposite, DSR produces three to four times less administrative packets than AODV.

To summarize, DSR has a lower ROP than AODV in stationary scenarios (600 s pause time) and a higher ROP in case of constant nodes motion. Results in ROB [14] take away some of DSR superiority in stationary scenarios and increase the gap between the protocols in the dynamic scenarios.

3.1.2. PDF

In our experiments we analyze how the increasing number of nodes influences the performance of both protocols. Fig. 2 depicts the PDF measured with 100, 200, 300, 400, 550 nodes for different traffic loads. In stationary scenarios (600 s pause time) with 10 CBR sources both protocols demonstrate good scalability with respect to the increasing number of nodes (and increasing path length), with AODV having a minor advantage over DSR at 400 and 550 nodes. With 0 s pause time performance of AODV slowly degrades from 96% at 100 nodes to 70% at 550 nodes, whereas DSR falls off markedly at 200 nodes when only 10% of the data packets are delivered. For settings with more than 200 nodes, DSR experiences significant growth in ROP (Fig. 1) which prevents the protocol from functioning effectively.

However, both protocols experience difficulties in terms of PDF when every third node acts as a CBR source (Fig. 2b). Even with 100 nodes and 600 s pause time AODV and DSR show only half the PDF observed in the scenarios with 10 traffic sources. As the number of nodes grows, the protocols PDF falls quickly, and at 550 nodes only 6% of data packets find their way to their destinations. The same picture is observed with 0 s pause time—from initial 40% (AODV) and 28% (DSR) of delivered data packets at 100 nodes performance falls to 1–2% at 550 nodes.

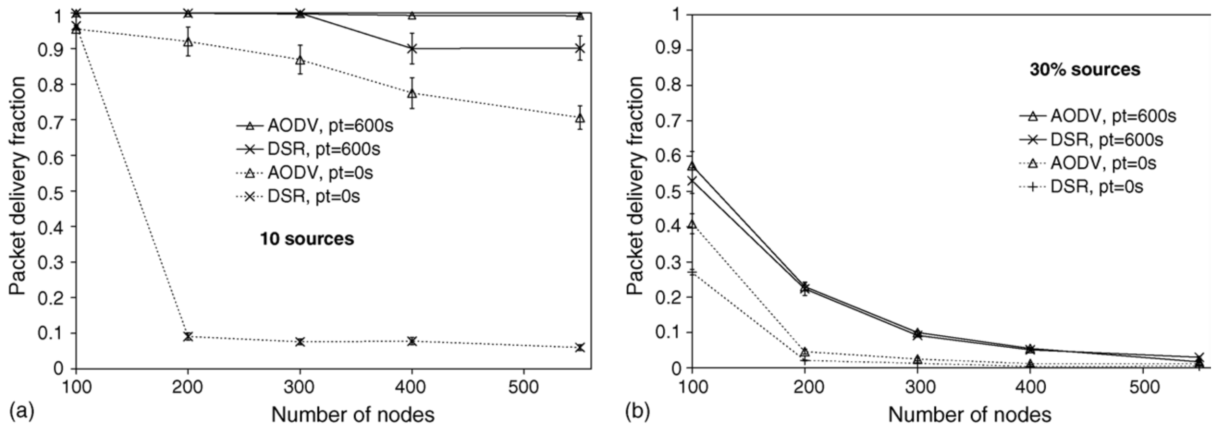

Fig. 2. Fraction of packets sent that is delivered (PDF) (100, 200, 300, 400, 550 nodes): 10 and 30% CBR traffic sources.

DSR caches all known routes; this strategy brings some improvements when the number of sources is high and the mobility rate is low. It is very likely that during route discovery for some destination *D* a route for another node *A* is found, recorded, and later used from the cache. But as the mobility rate increases, the information in caches becomes quickly stale, and this feature of DSR does not contribute significantly. How the mobility rate influences the scalability of both protocols is studied in Section 3.2.

3.1.3. Delay

The best average end-to-end delay is exhibited by both protocols in those scenarios that performed better according to the ROP and PDF metrics. With 10 CBR sources and 600 s pause time the delay for AODV and DSR is below 0.2 s with any number of nodes (Fig. 3). But the two protocols react differently to highly mobile hosts—the latency of AODV jumps up by 3–5 times, whereas DSR's latency increases 8–20 times. These latencies demonstrate more noticeable dependence on the number of nodes than in the

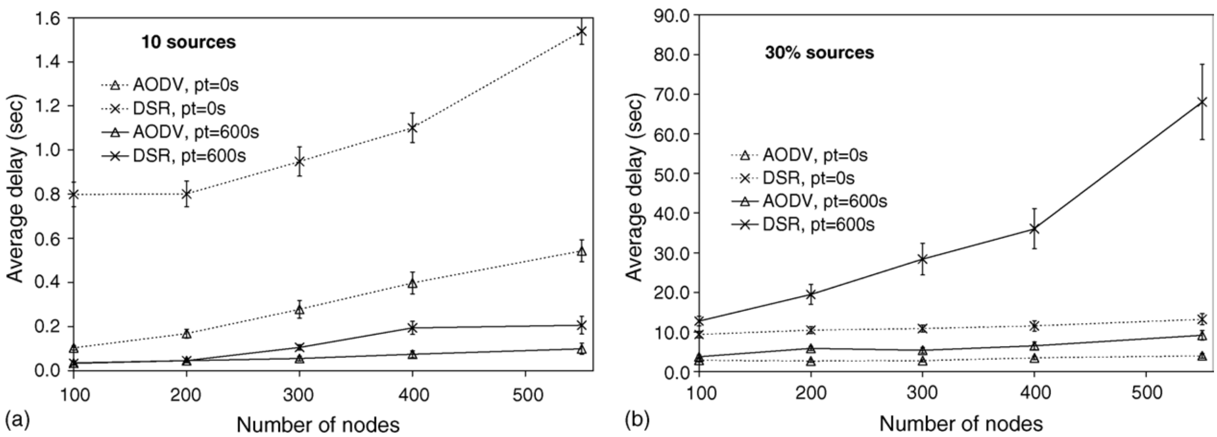

Fig. 3. Average end-to-end data packet delay (100, 200, 300, 400, 550 nodes): 10 and 30% CBR traffic sources. Note different scales for (a) and (b).

stationary case. Since an increasing number of nodes in our experiments means also an increasing path length, this behavior is easy to explain: longer routes takes longer time to discover and to travel along them. They also have a higher probability to break than shorter ones; as a consequence more route repairs are needed, which in turn result in a higher latency of the data packets.

With 30% of nodes acting as CBR sources AODV's delay shows little dependence on the number of nodes and varies from 2.9 to 9.2 s. DSR exhibits similar behavior with 0 s pause time. The delay grows from around 10 s at 100 nodes to 13.3 s at 550. However, the delay of DSR in a stationary case grows dramatically with the increasing number of nodes (from 13.5 s at 100 nodes to 68 s at 550 nodes). At the first glance there is an obvious contradiction with the default timeout value for a data packet (60 s). In reality there is no mistake—this happens because data packets dropped at source nodes by the MAC layer are re-inserted back into the send-buffer. The timeout is reset and the new route discovery is started, leading to such an enormous latency of data packets. The negative influence of this behavior on the overall performance of DSR will be described in Section 4.

3.2. *Influence of mobility rate*

For better understanding of how the rate of mobility and the number of traffic sources affect protocols performance we present the experimental results when the number of nodes in the scenarios is kept constant (550 nodes on 5 km $\times$ 1 km area).

3.2.1. *ROP and ROB*

Fig. 4 depicts ROP as a function of both pause time and network load for different numbers of traffic sources (the graph for ROB is not presented here due to space limits, see [14]).

In the scenarios with 10 CBR sources and 0 s pause time, DSR is afflicted by an ROP that is six times higher than AODV's ROP. For both protocols the load decreases as the pause time grows and the difference between the protocols disappears as the pause time reaches 600 s.

The situation changes when there are more traffic sources, shown in Fig. 4b. AODV has a slight edge over DSR for highly mobile setups, but at a pause time of 200 s for 160 CBR sources and 400 s for 40 CBR sources, the ROP of DSR becomes lower than for AODV. DSR continues to improve as the

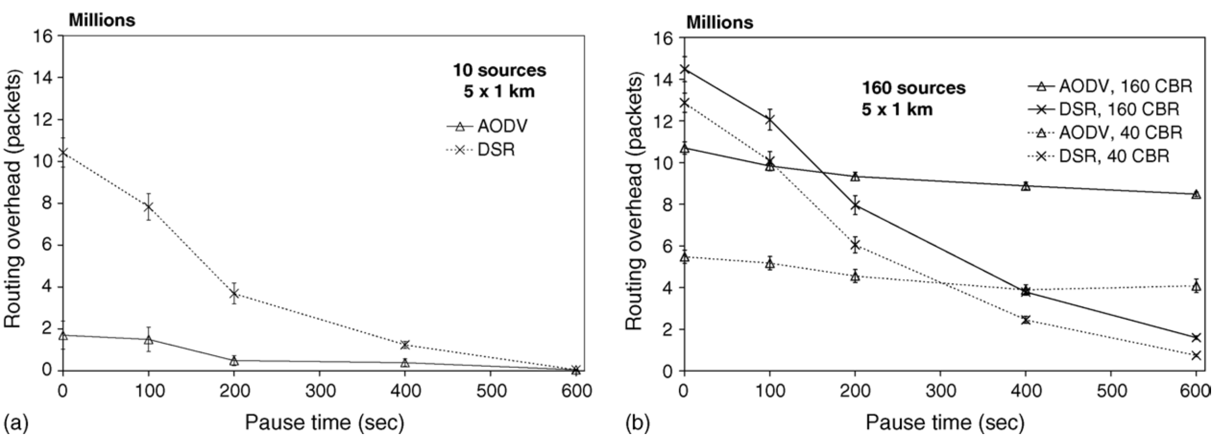

Fig. 4. Routing overhead, in packets (550 nodes): 10, 40, and 160 CBR traffic sources.

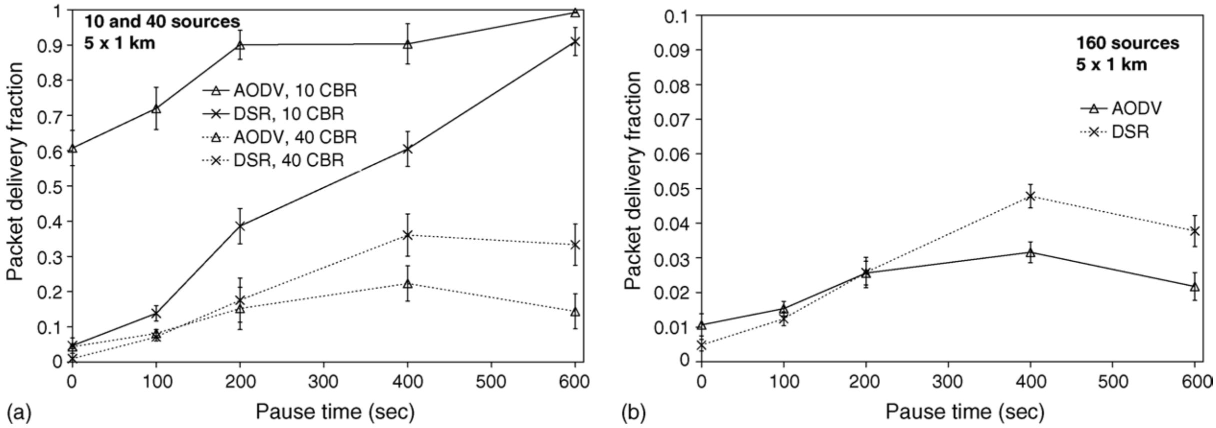

Fig. 5. Fraction of packets sent that is delivered (PDF) (550 nodes): 10, 40 (a), and 160 (b) CBR traffic sources. Note different scale for (a) and (b).

pause time increases, whereas AODV stays more or less constant. We see that the incremental cost of additional traffic sources decreases as we add more sources. AODV demonstrates a higher dependence on the number of CBR sources than DSR. DSR, in turn, is more dependent on the mobility rate.

According to the ROB metric, DSR is always more expensive than AODV, since DSR adds source routes to administrative packets. For example, with 0 s pause time DSR transmits 1930, 2180, and 2300 MB in the scenarios with 10, 40, and 160 traffic sources. AODV sends 4 to 24 times less bytes: 80, 260, and 510 MB in the corresponding scenarios. In the stationary networks (600 s pause time) both protocols transmit similar amount of bytes: 13, 227, and 410 MB.

But although DSR transmits a significantly larger number of bytes than AODV, the number of packets transmitted by DSR is not always higher, and the overall cost does not follow directly the number of bytes. Considering the high cost to access the medium, the cost to send a bigger number of small packets is higher than sending a smaller number of larger packets. Fig. 5a and b (in the next section) revisit this issue. Our data provide additional evidence: for the scenarios with 40 and 160 traffic sources and 600 s pause time, both DSR and AODV have similar ROB, but AODV has a higher ROP. Since in these scenarios DSR has a higher percentage of data packets delivered than AODV, sending the same payload in fewer large packets is indeed costly.

3.2.2. PDF

Fig. 5 depicts the PDF measured with 550 nodes as a function of node mobility, for different numbers of traffic sources. With 10 CBR sources at 0 s pause time, AODV outperforms DSR (60% of the packets are delivered versus 5% for DSR). DSR turns out to be unable to effectivly deliver data in a highly dynamic network populated by a large number of nodes. The reason for this non-performance is the high routing overhead of this protocol, as depicted in Fig. 4 of Section 3.2. The performance of DSR improves more or less linearly as the pause time grows, and DSR manages to deliver 91% of the data packets in stationary networks. AODV, along the same line, also improves and quickly gains a good performance level; AODV delivers 91% of the packets already at 200 s pause time and reaches the level of 99% in stationary networks.

However, the picture changes to the opposite with 40 CBR sources (Fig. 5a). AODV delivers from 5 to 50% *fewer* data packets than DSR for pause times over 200 s; up to 200 s, the protocols performance is about the same. DSR delivers 33% of all packets that originate at a sender with 600 s pause time, whereas AODV delivers only 16%. With 160 CBR sources, DSR also shows better performance than AODV once the pause time is larger than 200 s. But this result is no cause for celebration; the network congestion due to administrative packets is so high that both protocols loose from 97 to 99% of the data packets (Fig. 5b).

As mentioned in Section 3.1.2, DSR caches all known routes; in the experiments with varying mobility rates and different numbers of CBR sources, we see again that this strategy brings to DSR some improvements when the number of sources is high and the mobility rate is low. In the highly mobile scenarios with a big number of nodes a significant amount of information is stored in the nodes caches, but it is very difficultly to profit from this knowledge, since any stale entry only harms performance.

3.2.3. Delay

In all cases discussed so far, DSR has a higher average end-to-end data packet delay than AODV (Fig. 6), although the average number of hops made by a delivered data packet stays lower for DSR than for AODV (exact values for the number of hops can be found in [14]). DSR uses the length of a route as the main criterion for choosing a route from several routes that are either returned by a route discovery or stored in a node's cache. AODV implicitly prefers the least congested routes, since it replies only to the first RREQ, ignoring the route length. That is the reason why AODV exhibits a slightly longer average path, but shorter delays as will be shown in the next paragraphs.

Fig. 6 shows the data for the end-to-end delay. With 10 sources, the difference between the average delay for DSR and AODV is about a factor of 4 (in favor of AODV) at 0 s pause time. AODV and DSR both improve the average delay as the pause time increases. The delay caused by AODV slowly decreases from 0.6 s at constant node motion toward 0.1 s for a stationary network. DSR approaches the delay for AODV in the stationary case, but the delay remains about a factor 2 larger. This behavior corresponds well to performance according to the ROP metric (with 10 CBR traffic sources).

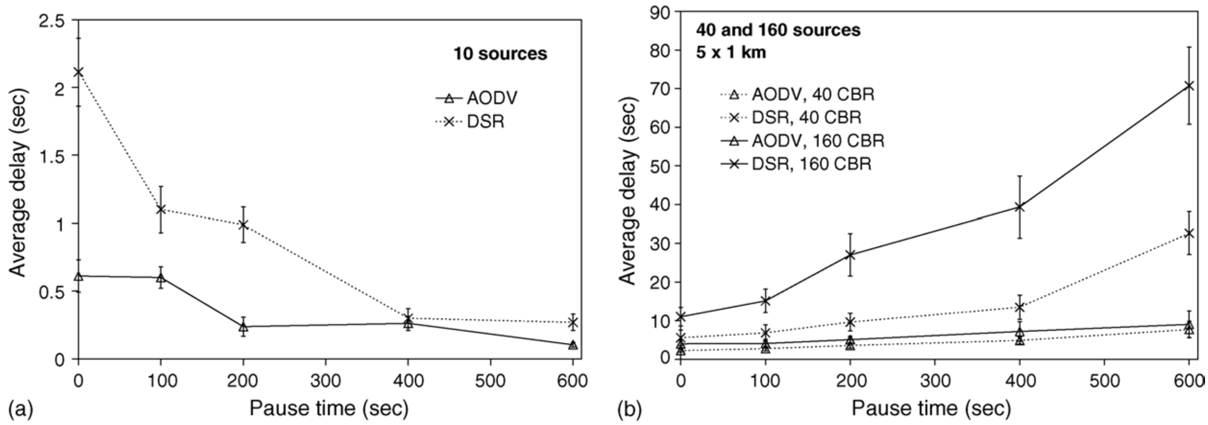

Fig. 6. Average end-to-end data packet delay (550 nodes): 10, 40, and 160 CBR traffic sources. Note different scales for (a) and (b).

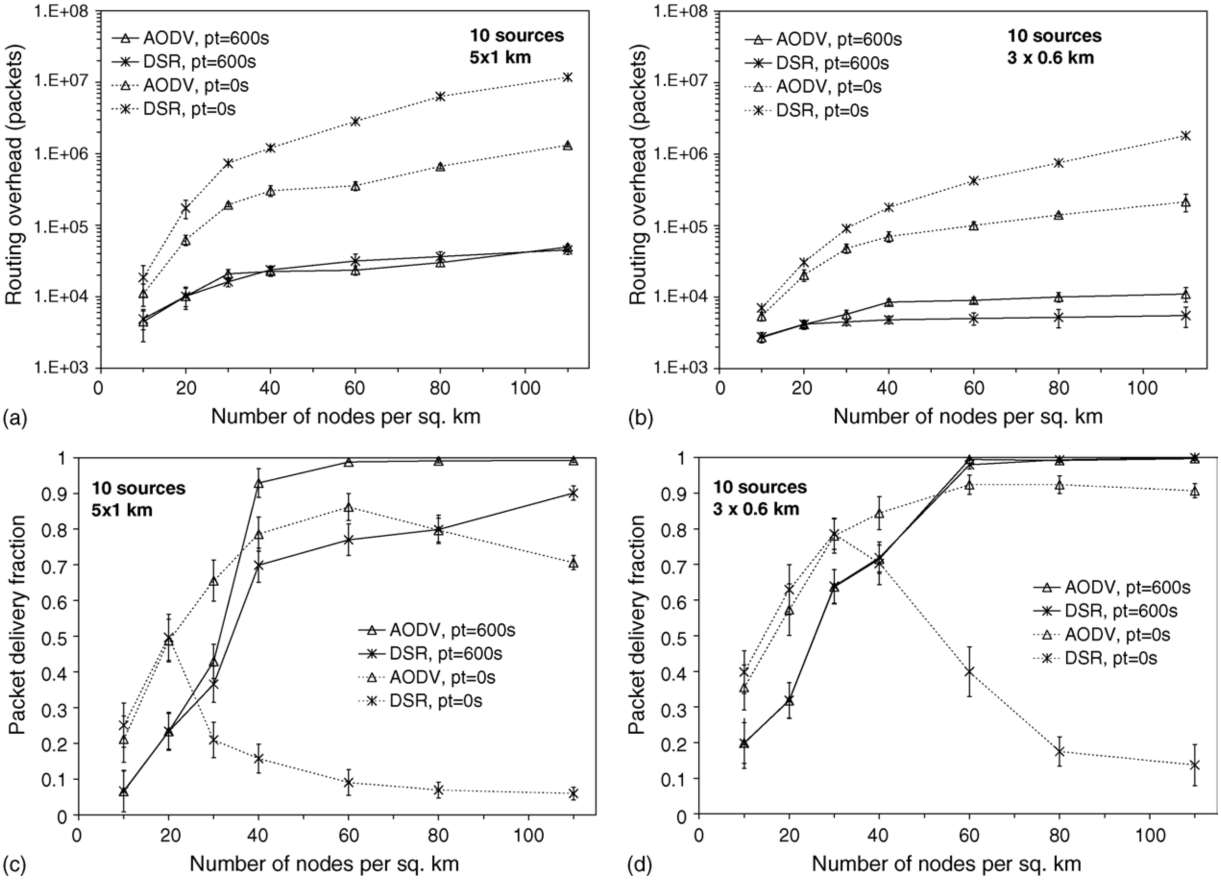

Fig. 7. Influence of nodes' density on the performance of protocols; 5 km × 1 km and 3 km × 0.6 km areas; different nodes densities.

With 40 and 160 CBR sources, we observe a different picture—the delay increases with growing pause time. This behavior can be explained with the contention in heavily loaded networks (see the previous section for figures). Both protocols deliver a higher percentage of data packets at 600 s of pause time than in scenarios with constant node motion, exploiting long paths as well as short ones, whereas in highly mobile scenarios, most successful transmissions involve short paths. The reason is that long paths have a high probability that a route will fail due to some node's movement during the propagation of a data packet. Since a data packet's delay is directly proportional to the number of hops in its route to a destination, the average delay increases when more long (stable!) paths are involved in a simulation. We do not observe this behavior with 10 traffic sources, since the congestion in stationary networks is negligible (around 10–100 times lower), compared to one in highly dynamic environments.

Again, AODV demonstrates a more stable behavior—the average end-to-end delay grows slowly with increasing pause time, whereas delays shown by DSR raise much more quickly and reach the value of 68 s with 160 CBR sources. The enormous increase of the DSR's delay is the result of the resurrection of data packets at source nodes, when the MAC transmission failed, as described in Section 3.1.

3.3. Influence of density of nodes

The density of hosts should have a significant influence on the routing protocols performance. Low density may cause the network to be frequently disconnected. High density increases the contention, resulting in a low per-node throughput. Earlier studies try to find the optimal density of nodes [15,17,16]. Here we show how the density of nodes influences the performance of AODV and DSR in the areas of 5000 km × 1000 km and 3000 m × 600 m. We simulate 10 CBR sources in the population of nodes, varying their density from 10 to 110 nodes/km^2, which corresponds approximately to 2–20 neighbors per node with our hardware parameters.

Fig. 7 presents ROP and PDF metrics for both protocols in two different areas as a function of nodes density. In highly mobile scenarios (0 s pause time) AODV clearly outperforms DSR in both metrics and exhibits good scalability. Both protocols deliver low percentage of data packets in the scenarios with low nodes density—the result of frequently disconnected network. As the density increases, connectivity improves and PDF of both protocols grows. However at around 20–30 nodes/km^2 DSR gets afflicted by the high ROP, which leads to the performance degradation. AODV in turn benefits further from the increasing density of nodes and reaches its performance maximum at around 60–80 nodes/km^2. In the static networks both protocols show similar behavior in ROP and PDF. The difference is more pronounced in the 5000 m × 1000 m area, where the routes are longer. In 3000 m × 600 m area DSR has three to four times lower ROP, but similar ROB. Average end-to-end delay stays higher for DSR in the scenarios with 0s pause time; in the static networks both protocols demonstrate similar delays (the graphs for ROB and average delay are presented in [14]).

In general, AODV shows good scalability to different densities of nodes and area sizes, especially in a highly mobile scenarios, whereas DSR shows good results only in small areas and only with the low mobility of nodes.

4. Analysis of the experimental results

To obtain a better understanding of the causes for the poor performance demonstrated by both protocols, we analyze more details of the network and summarize the problems observed. We also want to stress, that the metrics presented above capture the performance of a protocol. But they do not reveal the problems, that cause protocols poor performance. Thus, a researcher suggesting or evaluating some routing method should look at other metrics that we introduce and discuss in detail in the next subsections.

4.1. Reasons behind low PDF

Our experiments show that with a low traffic load (10 CBR sources) most of AODV data packets are dropped at intermediate nodes (75–80%); the number of nodes and the rate of mobility vary the results significantly. However, DSR has a different tendency, dropping 90% of packets at intermediate nodes with 100 nodes simulated, but from 200 to 550 nodes only 10–18% of the data packets are dropped at intermediate nodes. The mobility rate also has some influence on this factor, but this influence is much less pronounced.

DSR's resurrection of failed data packets at source nodes does not only cause very high data packets delays, it also forces more frequent route discoveries, which load the network. At the same time there is a certain inconsistency in this behavior. Why does a routing agent decide for a higher-level protocol which packets should be retransmitted? There can easily be the case, when such a delayed data packet finally reaches the destination, that it is either no longer needed, or was already retransmitted by the upper-level protocol. When a data packet is dropped at a source node, no RERR packets need to be generated, and it is logical to simply inform the upper-layer protocol about the problem and let it decide whether to retransmit the packet or not. According to our experiments, when the automatic retransmission of a failed data packet is switched off, the PDF of DSR improves in all the scenarios and the average end-to-end delay decreases by up to 60%.

To summarize, in low traffic load scenarios the majority of AODV data packets are dropped at intermediate nodes, although this number is small. DSR drops most of its data packets at intermediate nodes only in the scenario with 100 nodes. The main reasons for both protocols to drop data packets are failed ARP and MAC callback. As the number of traffic sources increases, up to 99% of data packets are dropped at source nodes. The cause for such a high number of drops at source nodes is the heavily congested network, flooded by administrative packets, which have higher priority in queues than data packets.

We observed that address resolution protocol (ARP) induces noticeable load on the network and becomes a reason for dropping different types of packets. In the scenarios with high traffic load up to 22% of the data packets and up to 27% of the administrative packets are dropped by ARP. The question is if ARP is needed at all. In Section 5 we pursue this question in depth.

4.2. Reasons behind high administrative load

To analyze the reasons for high administrative load (especially of DSR) we compare several metrics:

(1) how many neighbors successfully receive a broadcast RREQ;
(2) what is the percentage of successful RREP transmissions;
(3) what is the percentage of successful data packet transmissions;
(4) what makes a difference in protocols performance.

The first three metrics reflect the level of network load and help to understand the reasons for a low PDF. The forth one emphasizes how the different routing strategies of AODV and DSR affect their performance. Here we report the result only when *10 traffic sources* are used, since only in this case the performance of one protocol is good (acceptable). Analysis of this setup allows us to draw conclusions on which routing strategy is better and why. With higher number of sources the performance of both protocols becomes equally poor, which makes it less interesting to compare.

4.2.1. Successful RREQ transmissions

The known problem is that many of the packets sent by a routing layer of a node are consequently dropped due to MAC collisions. RREQ packets are the foundation of route discovery. If a high percentage of a node's neighbors are not able to capture an RREQ broadcast, then there is a high probability that this discovery will fail, waisting network capacity and delaying data packets.

To understand the success of the route discovery of both protocols, we analyze the following metric. Table 1 shows the ratio of the total number of RREQ packets that were successfully received by the

Table 1
Ratio of RREQ packets received to the number of RREQ transmissions, as a function of number of nodes

Protocol	Number of nodes				
	100	200	300	400	550
AODV	16.4	16.8	16.9	17.4	17.4
DSR	17.7	7.4	3.9	3.5	3.14

protocol to the total number of RREQ transmissions, as a function of number of nodes, [5] We see that exactly in the scenarios where DSR has shown poor PDF, only around 3–7 nodes successfully receive a RREQ broadcast packet. In our experiments, the number of a node's neighbors vary from around 10 to around 21, depending on the position of a node and the size of the area used. So most of a node's neighbors cannot capture the RREQ packet because of collisions with some other concurrent transmissions, mostly other broadcasts (hidden terminal problem). AODV drops 5–10% of RREQ transmissions, whereas DSR drops from 1% at 100 nodes to 79% at 550 nodes. With such a drop rate of DSR it is very hard to successfully discover a route. But even if one-way route is found, an RREP packet still has to propagate back to the initiator of the route discovery.

4.2.2. *Successful RREP and data transmissions*

We use the following logic to estimate the success of a unicast transmission. Each send or forward event of a unicast packet should be followed by a receive event, provided the transmission was successful, otherwise the packet is dropped. We count the total number of successfully received unicast packets packets and normalize it by the number of sent and forwarded unicast packets. In the ideal case this ratio must be 1.00 (or 100%), but in reality, it is of course lower. Table 2 reveals the issue. AODV manages to deliver the majority of RREP answers back to sources and, consequently, the majority of data packets to their destinations. However, in the case of DSR only 6–21% of RREP one-hop transmissions are successful, and every second one-hop transmission of a data packets fails.

4.2.3. *Analysis of routing strategy*

AODV and DSR are both reactive protocols that share the same route discovery strategy. However, different approaches in maintaining the routes and in storing routing information lead to different performance results. Here we are trying to analyze the reasons that cause high administrative load.

To establish a route both protocols broadcast RREQ packets that propagate through the network. If a route breaks, the protocols have the possibility to fix a broken route without starting a new route discovery from a source. We analyze how many RREQ are generated by source nodes and other (intermediate) nodes. Intermediate node may generate an RREQ in case of a local route repair (AODV) or of packet salvaging (DSR).

Table 3 shows the number of RREQ packets that were generated on average by each source node and each intermediate node in different scenarios. Switching from the 100 nodes to the 200 nodes scenario, DSR experiences a boost in the number of RREQ generated for each source and intermediate node. Exactly in this scenario DSR starts experiencing performance problems that continue as the number of

[5] For simplicity, in the rest of this paper we use only the number of nodes to represent a scenario. Section 1 contains the complete description of all scenarios.

Table 2
Ratio (%) of RREP and data packets received to the number of RREP and data transmissions, as a function of number of nodes

No. of nodes	AODV		DSR	
	RREP	Data	RREP	Data
100	0.96	0.98	0.76	0.99
200	0.95	0.98	0.21	0.50
300	0.95	0.98	0.10	0.47
400	0.94	0.97	0.08	0.46
550	0.93	0.96	0.06	0.39

Table 3
The number of RREQ packets that an average source and intermediate node sends, as a function of number of nodes

No. of nodes	AODV		DSR	
	Source	Intermediate	Source	Intermediate
100	93	3	20	10
200	132	3	484	465
300	146	2	1635	1586
400	215	2	2124	2198
550	254	2	3179	2797

nodes grows. Intermediate DSR nodes generate approximately as many RREQ packets as source nodes, which speaks for frequent route repair attempts. Physically the routes do not break for DSR more often than for AODV, the high level of congestion makes DSR believe that the next hop node is gone, because it did not respond with the CTS or did not acknowledge the reception of a unicast packet.

A logical continuation of an RREQ is RREP. An RREQ storm of DSR is followed by RREP storm—DSR sends 10 (with 100 nodes) to 100 (with 550 nodes) times more RREP than AODV. DSR replies to all RREQ creating an additional load by RREP packets on the first step. Since in our experiments every node has from 10 to 21 neighbors on average, each route discovery (RD) process will obtain up to 21 different routes to a destination. Of course not all of them will be valid long enough to serve as a backup path for the currently chosen one. If a route breaks, DSR may (and usually does) take another route from the cache. But there is no guarantee a newly chosen route will lead a data packet to the destination. If the route is stale, a new RD is performed, increasing the network congestion. The aggressive use of the cached routing information forces a lot of intermediate nodes to reply to a source of the route discovery. In the end the source node receives a high number of different paths, and consequently it is difficult to choose the best one. As Table 2 reports, the majority of DSR RREP packets are dropped, so they only waste the network capacity. Thus, when the number of highly mobile nodes is large, gathering a lot of routing information may work against a protocol, as we see in the case of DSR.

5. A new look on ARP

As we mentioned in Section 4.2, many routing (unicast) and data packets are dropped due to failed ARP. Here we study this problem and suggest a new revised ARP for wireless ad hoc networks. In computer networking using the internet protocol suite, the Address Resolution Protocol is a method for finding a host's physical address (MAC address), given its IP address. The sender broadcasts an ARP packet containing the Internet address of another host and waits for it (or some other host) to send back its MAC address. Each host maintains a cache of address translations to reduce delay and loading. To meet the needs of different types of local area network, there are separate ARP specifications for Ethernet, Frame Relay, ATM, Fiber Distributed-D ata Interface, HIPPI, etc.

In mobile ad hoc networks with the large number of hosts, a lot of transmissions have to wait until ARP resolves IP to MAC correspondence, since a node's neighborhood changes constantly. In the worst case in the network with n mobile hosts $O(n \cdot (n-1))$ ARP sessions are needed, since for each pair of hosts there must be two ARP exchanges performed. As all wireless transmissions, ARP packets are subject to collisions, which delay data or routing packet transmissions and may (and often do) lead to packet drops. At the same time a host A, before sending a unicast message to another host B, must first receive any packet from B. Thus the ARP-exchange becomes simply obsolete.

There are many ways how we can avoid unnecessary ARP exchanges and still maintain compatibility with systems that rely on ARP. One approach is to use cross-layer feedback. For every packet received, the MAC protocol (Data-link layer) provides IP to MAC address mapping to the ARP protocol (Network layer). The ARP protocol still maintains the cache and adds MAC addresses for outgoing IP packets. If there is no entry in the cache, ARP may use its standard exchange and obtain MAC address of the next hop destination traditionally. Another approach is to add a field for MAC address to IP packet header, e.g., to *Options* field. ARP can be used to add/extract MAC address from IP packets.

To determine the performance gain by omitting the ARP exchange we have conducted several experiments with AODV and GPSR (greedy perimeter stateless routing) protocol [9]. For both protocols we run simulations with standard ARP and a modified ARP, when MAC addresses are added to the IP packet. We took AODV as a representative of broadcast-based routing protocols, since it has shown better performance results than DSR. With DSR it would be difficult to estimate the influence of a new ARP, since DSR suffers more than AODV from a high administrative load imposed by RREQs. GPSR is one of the well-recognized geographic routing protocols, showing good performance [9]. A unicast packet

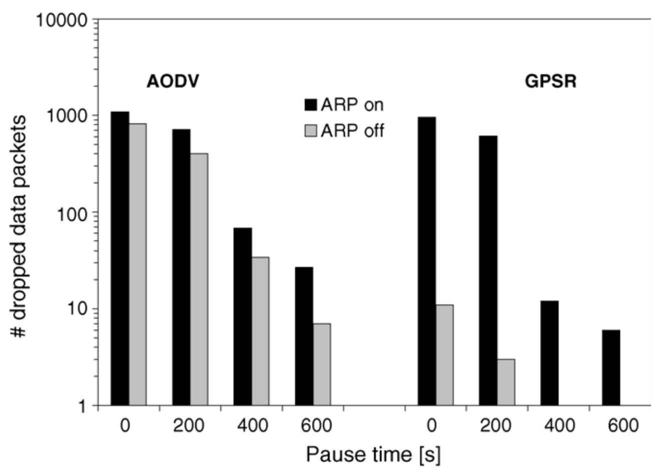

Fig. 8. Number of data packets dropped by AODV and GPSR as a function of pause time. “ARP on”—standard ARP is used. “ARP off”—revised ARP version is used. Ten CBR sources.

needs an ARP exchange if the MAC address of the packet destination is not known. Another reason for choosing GPSR is to test the following assumption: a protocol that operates mostly with unicast packets should benefit more from omitting ARP exchange. Thus, GPSR should reduce the number of drops more significantly than, e.g., AODV.

Fig. 8 shows the number of data packets dropped by both protocols in the scenarios with 10 CBR sources (please note that a logarithmic scale is used). We simulate both standard ARP exchange "ARP on" and revised ARP "ARP off". We see that both protocols drop significantly fewer data packets when no ARP exchange is used. As the mobility of nodes increases (the pause time becomes smaller), a node's neighborhood changes more frequently, requiring more frequent ARP exchanges. As a result many data packets are dropped due to ARP failure. Omitting the ARP exchange saves a lot of them. This phenomena is more pronounced with GPSR, since it relies mostly on unicast transmissions. That result supports our hypothesis that a unicast protocol benefits more from changes to the ARP exchange.

6. Concluding remarks

We looked in detail at different aspects that affect the scalability of two ad hoc routing protocols, AODV and DSR. We analyzed how the performance of AODV and DSR depends on the area, the number of nodes, the density of nodes, the rate of node mobility, and the suggested data traffic load.

In stationary scenarios with a low number of traffic sources both protocols demonstrate good scalability with respect to the number and density of nodes. But as the mobility rate (and/or density of hosts) increases, the routing overhead of DSR precludes this protocol from delivering data packets effectively. The attempts to limit the routing load by aggressive use of the cached routing information results in a higher number of administrative packets. This factor seems to limit the scalability of DSR and there may be an opportunity to change to the way DSR builds up and uses cached information, by adding some optimizations, e.g., like suggested in [18].

The poor results shown by both protocols in the scenarios with high traffic load suggest that effective strategies to eliminate the load of simple RREQ flooding should be adopted by both protocols to improve their scalability. We also point that traditionally considered metrics do not uncover all the sources of protocol performance problems. We suggested and discussed additional metrics that help to understand the issue.

The suggested revised role of ARP in wireless ad hoc network simplifies communication between hosts and improves the performance of any routing protocol, especially in scenarios with high mobility of hosts.

Ad hoc networks show a lot of promise to extend the reach of the world of managed networks. But before ad hoc networks become widely used, it is necessary to improve their scaling behavior.

Acknowledgements

This research was supported, in part, by the NCCR "Mobile Information and Communication Systems", a research program of the Swiss National Science Foundation, and by a gift from the Microprocessor Research Lab (MRL) of Intel Corp.

References

[1] Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, IEEE Standard 802.11, Technical report, 1999.

[2] J. Broch, D. Maltz, D. Johnson, Y. Hu, J. Jetcheva, A performance comparison of multi-hop wireless ad hoc network routing protocols, Proceedings of the MOBICOM'98, 1998, pp. 85–97.

[3] S. Das, Ch. Perkins, E. Royer, Performance comparison of two on-demand routing protocols for ad hoc networks, Proceedings of the INFOCOM'00, 2000, pp. 3–12.

[4] K. Fall, Ns notes and documentation, The VINT Project, 2000.

[5] M. Gastpar, M. Vetterli, On the capacity of wireless networks: the relay case, Proceedings of the INFOCOM'02, 2002.

[6] P. Gupta, P. Kumar, The capacity of wireless networks, IEEE Trans. Inform. Theory 46 (2000) 388–404.

[7] P. Johanson, T. Larsson, N. Hedman, B. Mielczare, M. Degermark, Scenario-based performance analysis of routing protocols for mobile ad-hoc networks, Proceedings of the MOBICOM'99, 1999, pp. 195–206.

[8] D. Johnson, Validation of wireless and mobile network models and simulation, DARPA/NIST NSV Workshop, 1999.

[9] B. Karp, H. Kung, GPSR: greedy perimeter stateless routing for wireless networks, Proceedings of the MOBICOM'00, 2000, pp. 243–254.

[10] V. Naoumov, Th. Gross, Simulation of large ad hoc networks, Proceedings of the MSWiM'03, 2003, pp. 50–57.

[11] Ch. Perkins, E. Royer, Ad-hoc on-demand distance vector routing, Proceedings of the WMCSA'99, 1999, pp. 90–100.

[12] E. Royer, S-J. Lee, Ch. Perkins, Ad hoc on-demand distance-vector routing scalability, ACM MCC Rev. 6 (2002) 94–95.

[13] J. Li, C. Blake, D. De Couto, H. Lee, R. Morris, Capacity of ad hoc wireless networks, Proceedings of the MOBICOM'01, 2001, pp. 61–69.

[14] V. Naumov, Th. Gross, Scalability of routing methods in ad hoc networks, Internal technical report, 2005. http://www.lst.inf.ethz.ch/research/publications.

[15] L. Kleinrock, J. Silvester, Optimum transmission radii for packet radio networks or why six is a magic number, IEEE NTC (1978) 4.3.1–4.3.5.

[16] F. Xue1, P.R. Kumar, The number of neighbors needed for connectivity of wireless networks, Wireless Networks 10 (2) (2004) 169–181.

[17] E. Royer, P. Melliar-Smith, L. Moser, An analysis of the optimum node density for ad hoc mobile networks, IEEE ICC (2001).

[18] M. Marina, S Das, Performance of route caching strategies in dynamic source routing, Proceedings of the ICDCSW, vol. 00, 2001, p. 0425.

Valery Naumov received a Master of Science in Physics/Computer Science from Novosibirsk State University (Russia) in 1999. In July 2000 he started his PhD at Computer Systems Institute of Swiss Federal Institute of Technology in Zurich, Switzerland.

Thomas Gross joined CMU in 1984 after receiving a PhD in Electrical Engineering from Stanford University. In 2000, he became a Full Professor at ETH Zurich, Switzerland. He is now the head of the Computer Systems Institute, from 1999 to 2004 he was the deputy director of the NCCR on "Mobile Information and Communication Systems", a research center funded by the Swiss National Science Foundation. He is also an Adjunct Professor in the School of Computer Science at Carnegie Mellon University.