Ad Hoc & Sensor Wireless Networks, Vol. 1, pp. 159–177
Reprints available directly from the publisher
Photocopying permitted by license only

Adaptive Gateway Discovery Mechanisms to Enhance Internet Connectivity for Mobile Ad Hoc Networks

Pedro M. Ruiz, Antonio F. Gomez-Skarmeta

Fac. Informatica, Campus de Espinardo. E-30071, Espinardo, Murcia, Spain.
E-mail contact: pedrom@dif.um.es

Received: July 26, 2004. In final form: August 2, 2004.

One of the key overhead components affecting the overall performance of hybrid ad hoc networks is the discovery and selection of Internet gateways. We develop an analytical model to evaluate existing proposals and we show that each of them is suited only for a limited range of network conditions. We propose and simulate two new adaptive gateway discovery algorithms (maximal source coverage and maximal benefit coverage) based on the dynamic adjustment of the scope of the gateway advertisement packets. Our results show that the adaptation capabilities of our proposed schemes allow them to outperform existing mechanisms over a variety of scenarios and mobility rates.

Keywords: Hybrid, adaptive, GW discovery, limited advertisements.

1 INTRODUCTION

Mobile ad hoc networks, are extremely flexible, self-configurable and easy-to-deploy wireless networks. These properties are making them a very attractive component in future mobile and wireless network architectures. In addition, with the advent of future wireless systems consisting of an integration of different heterogeneous wireless technologies, the interconnection of MANETs to fixed IP networks is one of the areas which are becoming of paramount importance. In such scenarios, commonly known as hybrid ad hoc networks, mobile nodes are witnessed as an easily deployable extension to the existing infrastructure. Some ad hoc nodes act as “gateways” which can used by mobile terminals to seamlessly communicate with other nodes in the fixed network. The challenge in interconnecting ad hoc networks to Internet, stems from the need to inform ad hoc nodes about available gateways in an extremely changing scenario while making a minimal consumption of the scarce network resources. So, an efficient gateway discovery for ad hoc networks becomes one of the key elements to enable the use of hybrid ad hoc networks in future mobile and wireless networks.

The different proposals to the issue of Internet connectivity for MANETs in the literature have used either a proactive or a reactive gateway discovery. In proactive approaches ([1–3]) the gateways periodically send advertisement messages which are flooded throughout the entire ad hoc network to inform every ad hoc node about available Internet gateways. Although these approaches achieve good connectivity, they have been usually criticized due to the high overhead they require and their limited scalability. In reactive approaches ([4, 5]) those nodes which require connectivity to the Internet reactively find Internet gateways by means of broadcasting some kind of solicitation within the entire ad hoc network. Although this approaches have been considered to require less overhead, we show in the next section that this process of finding gateways can become as costly as the proactive advertisement. In fact, we show that reactive gateway discovery mechanisms scale poorly regarding the number of active sources willing to access the Internet.

There are also some works ([6, 7]) which propose hybrid gateway discovery approaches. In [6], the authors propose an scheme in which advertisements are only propagated up to a fixed number of hops, and those nodes out of that scope will proactively find the gateways. However, as the authors show, there is not a good TTL value which can offer a good performance in different scenarios and network conditions. In [7] the authors proposed a more sophisticated approach in which advertisements are sent out only when changes in the topology are detected. However, they rely on the use of a source-routing protocol, which limits the applicability and scalability of their approach.

In our opinion, existing approaches have neglected the huge overhead that the reactive gateway discovery scheme can have. The overall performance of the fixed approaches proposed so far can vary dramatically as the network conditions change. This is due to the strong performance dependence that they have on the scenarios under consideration (e.g. number of sources, number of nodes, degree of the network, etc.). We propose two different adaptive gateway discovery approaches based on the dynamic tuning of the scope of the gateway advertisements. One key element of our proposal is the simplicity and easy computation without additional overhead of the network parameters under consideration. Just by monitoring data packets, gateways will adaptively select the time to live of their advertisement that best suits the current network conditions. So, even when the network conditions change, the overall network overhead is reduced while still maintaining a good connectivity. In the authors' opinion, the main contributions of this paper are (i) an analytical study of the overhead of different gateway advertisement approaches showing the need for an adaptive scheme, and (ii) two different adaptive gateway discovery approaches for hybrid networks which are shown to outperform existing proposals.

The remainder of the paper is organized as follows: section 2 provides an analytical evaluation of the different approaches and shows the need for adaptive gateway discovery alternatives. In section 3 we describe our proposed adaptive approaches based on the maximal source coverage and the maximal benefit coverage. The results of our simulations are shown in section 4. Finally, section 5 gives some conclusions and draws some future directions.

2 ANALYTICAL EVALUATION OF EXISTING GATEWAY DISCOVERY APPROACHES

In order to compare different gateway discovery approaches under the same conditions, we consider as the baseline scenario an hybrid network using the AODV [8] ad hoc routing protocol with an Internet connectivity approach similar to the reactive one proposed in [4]. In the next subsections, we define how to incorporate the reactive, proactive and hybrid gateway discovery approaches and we evaluate them.

2.1 Operation of existing approaches

The reactive approach is the basic approach described in [4]. RREP and RREQ messages are extended with a new flag ("I") which is used to differentiate control messages used to discover routes to the Internet from usual RREP and RREQs. We refer to the new messages as RREP_I and RREQ_I. A source willing to communicate with a node in the fixed network, will first attempt to contact it within the ad hoc network doing an extended ring search (as described below). If no answer is received after a network-wide search, then the source tries to find a route towards the Internet. So, it broadcasts a RREQ_I to the ALL_MANET_GW_MULTICAST address. Gateways, upon reception of this message will send out a unicast RREP_I message to the source. Then the source will select one of the gateways (based on the hop count) and will send the data towards the fixed node through that gateway.

For the proactive approach we introduce a new message called GWADV ("Gateway Advertisement"). Gateways will periodically broadcast within the ad hoc network these messages in order to inform all the nodes about

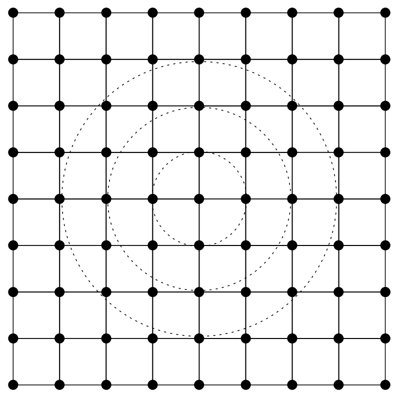


FIGURE 1
Rectangular lattice

the availability of that gateway. Upon reception of a GWADV message, mobile nodes will select their preferred gateway based on the hop count, and they will store a default route entry in their routing table. When a source wants to communicate with a destination, it tries first to find a direct route within the MANET, and if it does not manage to do it, it then uses its default route.

Finally, the hybrid approach we have implemented is basically the one described in [6]. Gateways will periodically flood TTL-limited GWADV messages which will only be forwarded upto a few hops away from the gateway. The sources within that flooding area, upon reception of the GWADV messages, will behave as in the proactive approach. Those nodes beyond that number of hops will find default routes proactively using the same RREQ_I-based reactive scheme described before. So, this approach is somehow a trade-off between the reactive and proactive approaches.

2.2 Analytical Model

We assume that the nodes are uniformly distributed in a rectangular lattice covering a certain area. Each vertex of the lattice is a possible location for a node, but only one node can be at a concrete vertex. An example of such a rectangular lattice is shown in Fig. 1. We start by computing some basic elements which will be later using throughout the model. Given a node n in the lattice, there are $4k$ nodes at a distance of k hops from n. These nodes are placed in the k^{th} concentric ring centered on the node n. It is easy to show that the total number of nodes including n at a distance of k hops is given by (1). We also give the relation between k and N, in which $\lceil x \rceil$ is the standard ceiling operation meaning completion to the next integer. It is used in the expression for obtaining k because the last concentric ring might not be complete. So, given a broadcast message with time to live (TTL) equal to x, $N_r(x)$ will be the number of nodes forwarding that message if $x \leq (\sqrt{2N-1}-1)/2$ and N otherwise.

$$N_r(k) = 1 + \sum_{j=1}^{k} 4j = 1 + 2k(k+1), \qquad k = \lceil(\sqrt{2N-1}-1)/2\rceil \tag{1}$$

Regardless of our gateway discovery mechanism, the approach used in [4] detects that a destination is a fixed node when the source does not receive any answer after a network-wide search. This network wide search is done using an expanding ring search. That is, the first route request message is only sent to the nodes at TTL_START hops. If no answer is received, a new message is sent with the previous TTL plus TTL_INCREMENT. This process is repeated up to a TTL of TTL_THRESHOLD. If no answer is received, then the last request message is sent with a TTL equal to NETWORK_DIAMETER. The typical values defined for these constants in AODV specification are TTL_START=1, TTL_INCR.=2, TTL_THR.=7 and NET_DIAM.=30. Although we think that these values are not appropriate for hybrid networks, we have obeyed the original specification.

Whenever an ad hoc source tries to find a route towards a fixed node it never gets any answer within the ad hoc network. Thus, for each source S, the number of messages associated to realizing that a destination is a fixed node can be calculated according to (2).

$$\Omega_{FN} = \sum_{j \in \{1,3,5,7,30\}} N_r(j) \tag{2}$$

Similarly, whenever a source wants to reactively discover a gateway there is an overhead which is the sum of the number of messages required to do a network-wide distribution of the RREQ_I packet addressed to the ALL_MANET_GATEWAYS multicast address, plus the number of messages required to send an unicast RREP_I reply from every gateway to the source. Assuming that the gateways are in the borders of the lattice, it is easy to demonstrate that the mean path length is $\sqrt{N}-1$. Thus, if we denote the number of gateways by N_{GW}, the overhead of the reactive discovery of the gateways by one source can be computed as it is shown in (3).

$$\Omega_{r-gw} = N + N_{GW} + N_{GW} \cdot (\sqrt{N}-1) = N + N_{GW} \cdot \sqrt{N} \tag{3}$$

Let S be the number of active sources in the hybrid network communicating with fixed nodes, λ_{adv} the rate at which GWADV messages are being sent out by the gateways and t the duration of the time interval under consideration. The overhead of delivering each of this messages to the whole ad hoc network is $N+1$ messages; one forwarding by each of the N nodes (because of the duplicate messages avoidance) plus the message sent out by the gateway itself. In addition, if we take into account that initially all the sources in the network will need to realize that the destination node is a fixed node, then the total overhead in number of messages required by the proactive approach can be obtained using (4).

$$\Omega_p = S \cdot \Omega_{FN} + \lambda_{adv} \cdot t \cdot (N+1) \cdot N_{GW} \tag{4}$$

In the same way, if we denote by R_{dur} the route duration time in AODV[1]. Then, R_{dur} obeys an exponential random distribution with parameter λ_{dur}. Let N_{break} be a random variable representing the number of route expirations during an interval of t units of time. Then, N_{break} follows a Poisson distribution with an arrival rate equal to λ_{dur} so that $P[N_{break} = k] = \frac{e^{-\lambda_{dur}} \cdot \lambda_{dur}^k}{k!}$. So, the mean number of default route expirations per source will be given by $E[N_{break}] = \lambda_{dur} \cdot t$. Accordingly, the total overhead for the proactive route discovery will consist of the initial overhead to make every source aware that their destinations are fixed nodes, plus the overhead associated to the proactive discovery of the gateways whenever their default route expires or breaks. This overhead can be computed according to (5).

$$\Omega_r = \left[\Omega_{FN} + \left(\Omega_{r-gw} \cdot \lambda_{dur} \cdot t\right)\right] \cdot S \tag{5}$$

The hybrid gateway discovery scheme, has an overhead which is a combination of the overheads of the other approaches. For those sources located outside the area covered by GWADV messages, the overhead will be the similar to the overhead of the reactive approach. Thus, in order to asses the overhead of the hybrid approach it is of paramount importance, being able to calculate the mean number of sources which will be within the GWADV range.

Let's assume that the gateways are located in the corners of the lattice as in our simulated scenario. In the hybrid approach it makes no sense sending GWADV at longer distances than $\sqrt{N} - 1$ hops, because other gateways will be covering the area beyond that TTL. Then, its is easy to derive an expression for the number of nodes which are at an scope of s hops from any gateway according to (6), with $s \in [0, \sqrt{N} - 1]$.

$$N_r^{GW_i}(s) = \sum_{j=1}^{s} (j+1) = \frac{s(s+3)}{2} \tag{6}$$

Given a node n from the ad hoc network, the probability that this node will be able to receive a GWADV message from any of the gateways can be computed as shown in (7).

$$P_c(s) = \frac{\sum_{i=1}^{N_{GW}} N_r^{GW_i}(s)}{N - N_{GW}} \tag{7}$$

[1] configured to be 10 seconds unless the route becomes invalid before (e.g. due to mobility)

TABLE 1
Values for analytical evaluation

Constant	N	λ_{adv}	N_{GW}	λ_{dur}	t
Value	25	1/5	2	1/10	900 sec

If we denote N_c as the number of sources being covered by any gateway when using a scope of s hops, then N_c is a random variable obeying a binomial distribution $B \sim (S, P_c(s))$. Thus, the mean number of sources being covered when gateways use a scope of s hops can be computed as $E[N_c] = S \cdot P_c(s)$. So, the overall overhead of the hybrid approach consists of three different parts: the overhead associated to realize that the destinations are fixed nodes, the overhead associated to the propagation of GWADV messages over s hops by each gateway, and the overhead required so that those sources not covered by the GWADV messages can find the gateways and create a default route. An expression for that overhead is shown in (8).

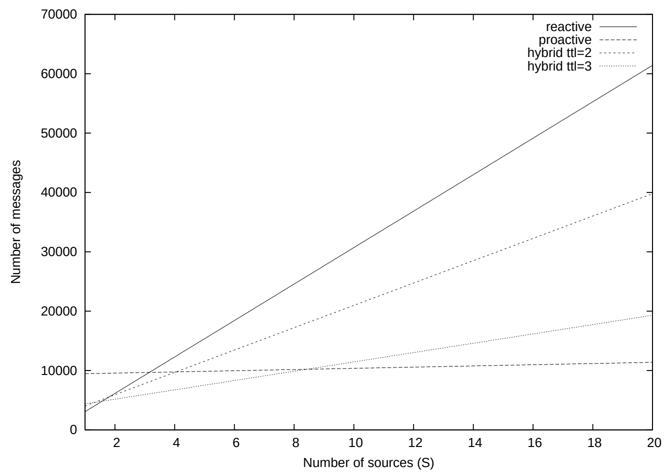
$$\Omega_h = S \cdot \Omega_{FN} + \lambda_{adv} \cdot t \cdot (N_r^{GW}(s) + 1) \cdot N_{GW} + \Omega_{r-gw} \cdot \lambda_{dur} \cdot t \cdot S \cdot (1 - P_c(s)) \quad (8)$$

In the next subsection, we show numerical results from our analytical model to assess the effectiveness of each of the existing alternatives.

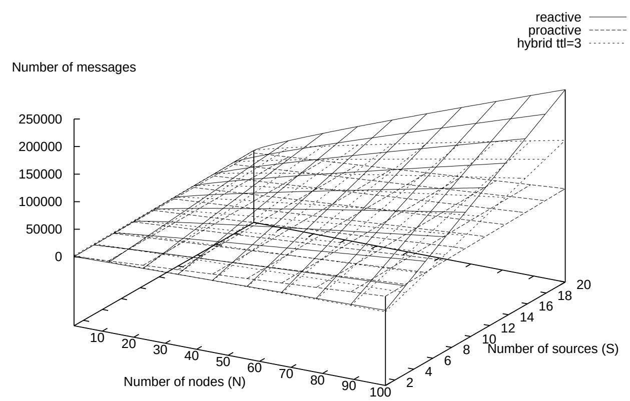
2.3 Analytical Evaluation

To compare the overhead of the different approaches, we have used the figures in Table 1. As it was expected the proactive approach is less scalable regarding the number of nodes in the ad hoc network. This is because the higher the number of nodes, the higher the number of retransmissions which are required to propagate GWADV messages to the whole network. Because of that proactive approaches have been said in the literature to have too much overhead. However, we can also notice that the process of discovering the gateways can be as costly as the process of propagating the GWADV messages. In fact, under certain network conditions the reactive approach can incur in higher overhead than the proactive one. In particular, we have found interesting to stress the poor scalability of the reactive approach as the number of sources connecting to Internet increase. This is supported by the graphs in Fig. 2(a) and Fig. 2(b).

As is it also shown in Fig. 2(a), the hybrid approach is somehow a trade-off between the reactive and the proactive approaches. Different values of TTL lead to different flavors of the hybrid approach. However, as it was also corroborated in [6], the optimal value of TTL is something that strongly varies from one scenario to another. In fact, as depicted in



(a) Overhead vs. number of sources



(b) Overhead vs. number of nodes and number of sources

FIGURE 2
Analytical overhead of the different approaches.

Fig. 2(b), there are situations in which a proactive approach performs better than an hybrid approach and vice versa. Thus, the definition of an universal hybrid gateway discovery approach seems to be unrealistic without adding some degree of adaptability. We describe our proposed adaptive approaches in the next section.

3 PROPOSED ADAPTIVE GATEWAY DISCOVERY MECHANISMS

Given the conclusions of our analytical study, we believe that an adaptive gateway discovery mechanism being able to dynamically change its proactiveness or reactiveness can reduce the overhead of the gateway discovery without jeopardizing the overall network performance. Hence, in this section we propose two different gateway discovery approaches based on the dynamic adjustment of the TTL of GWADV messages. Our previous study has shown that the scope of the advertisements has a strong impact on the proactiveness or the reactiveness of the scheme. Thus, it seems reasonable to use the TTL of the GWADV messages as the parameter to adjust depending on the network conditions. The higher the TTL, the higher the overhead due to the periodic advertisement and the lower the overhead associated to the reactive discovery of the Internet gateways. That is, the higher the TTL the higher the proactiveness of the approach. In fact, a TTL = 0 corresponds to the totally reactive approach whereas a TTL = NETWORK_DIAMETER corresponds to a completely proactive scheme.

There are different criteria to determine when the TTL should be adjusted (i.e. when to perform the adaptation). For instance, the rate at which neighbors change or the mean duration of the links can be an indication of the network mobility. However, these kind of metrics are not usually easy to interpret. In addition, they do not capture one of the key parameters according to our model which is the number of sources.

For a gateway to be aware of the total number of sources communicating with nodes in the Internet it is required some kind of signaling mechanism facilitating such information to the gateway. However, that would incur in extra overhead and it is something which can require changes to the routing protocols. So, we propose to use simpler metrics, being able to convey the required information without any additional overhead, and being able to be locally computed in real scenarios. In our two proposals, each gateway will only know about the sources which are accessing to the Internet through them. This scheme is very convenient because that information is very easy to learn by the gateway provided that it is routing those datagrams that it would receive anyway.

The gateways will keep track (using a structure like the one which is shown in Fig. 3) of the number of hops at which each of its active sources is located. This information is easy to extract by simply looking at the IP header of data packets. This table will be periodically purged so that stale entries do not influence the TTL of the next advertisement.

In the next subsections we propose two different algorithms for the determination of the TTL to be used for the next gateway advertisement. Both of them rely solely on the local information contained in the aforementioned table.

IP address	TTL	Time learnt
	....	

FIGURE 3
Table to store source TTL entries

3.1 Maximal Source Coverage

The "maximal source coverage" algorithm is the most proactive approach from the possible heuristics. Using this algorithm, a gateway will send out the next advertisement with a TTL equal to the minimum number of hops required to reach all of its sources. This simple algorithm has the good property of constraining the flooding of advertisement messages to the smallest TTL that covers the sources. So, this approach combines the advantages of the proactive approach (freshness of routes) and the benefits of reactive schemes (avoid flooding the entire network).

There are other feasible heuristics which can make the gateway advertisement process more reactive. For instance, the TTL can be chosen so that GWADV messages only cover the closest source, the TTL which covers a certain number of sources, or even the TTL which covers a certain percentage of its sources. However, we have selected the maximum source coverage because our main concern is maintaining a high packet delivery ratio as close as possible to the proactive approach even if it costs a little more in terms of overhead. The other heuristics tend to produce in some scenarios less overhead than the selected one, but they achieve a lower packet delivery ratio. As we show in the next section this algorithm is able to obtain a similar throughput than the proactive approach while keeping the overhead at very close values to the hybrid and proactive ones.

The proposed algorithm, while still outperforming existing ones, may offer suboptimal solutions when a small number of sources are at a large number of hops from the gateway. The worse case is the one in which only one of those sources is at a very large distance. In that case, it may be more cost-effective not to expand the advertisement area and let that source to find the gateway on its own. These suboptimal scenarios occur because this algorithm does not have a notion of cost in terms of overhead. In the next subsection we introduce a new algorithm to compute the TTL which takes into account the cost of covering active sources.

3.2 Maximal Benefit Coverage

To avoid the suboptimal scenarios described before, we propose a more sophisticated algorithm called "maximal benefit coverage". In this schema, the gateways will select the TTL of their advertisements so that the overhead savings are maximized. That is, the gateway will select a TTL t so that the overhead of flooding GWADV messages up to t hops plus the overhead associated to the discovery of gateways by sources at distances longer than t hops is minimized.

In order to accurately compute that optimum, the gateway would require complete information of the topology, number of nodes, etc. However, the overhead required to make all this information available to the GW can be higher than the benefit it can obtain. So, following our philosophy of simplicity and use of local information, in our case the gateway will use an heuristic to approximate the benefit even if there is some uncertainty. This heuristic approach manages to outperform the other schemes without the need for any extra information compared to the maximal source coverage scheme.

To compute the benefit of using $TTL = t$ for GWADV messages, each gateway will use the expression shown in (9). In this equation N represents the cost in messages of a network-wide flooding and $S(t)$ is a function which computes the number of active sources for that gateway at a distance less or equal to t. The numerator of (9) represents the cost in terms of number of messages associated to not covering the sources (a network-wide flood for each of them), and the denominator represents the cost of flooding up to t hops (estimated according to our model in section 2). Note that the function $S(t)$ is accurately known by the source because it is stored in the aforementioned 'Source TTL' table.

$$\beta(t) = \frac{N \cdot S(t)}{t(t+3)} \tag{9}$$

Unlike the maximal source coverage algorithm, this approach considers the additional flooding cost of covering a source when selecting the TTL. This is clearly shown in (9) in which the benefit decreases as a function of the number of nodes required to propagate the GWADV messages and it increases as the number of sources at that number of hops is higher.

So, the problem of finding the most appropriate TTL for the next advertisement can be formulated as finding $t \in [1..t_{max}]$ so that $\beta(t) = \max_{1 \le x \le t_{max}} \beta(x)$, where t_{max} is the TTL of the source which is furthest away from the gateway. This problem is a simplified dynamic programming problem, which can be easily solved in $O(S)$ (a single pass on the table with S sources).

Fig. 4 shows the benefit for different TTLs of some combinations of sources. The combinations in the legend of the graph which are in the form $[s_1, s_2, \ldots, s_n]$ represent a particular case in which there are s_1 sources at TTL=1, s_2 sources at TTL=2, and so on. As the figure depicts, the benefit function is not prone to those suboptimal cases of the previous algorithm.

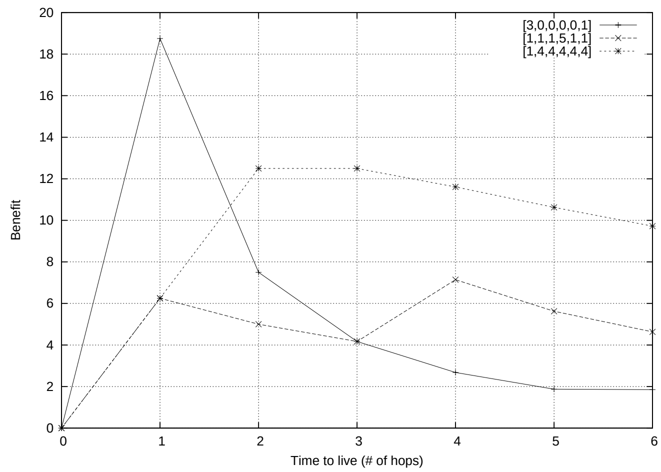


FIGURE 4
Examples of Benefit functions for different combinations.

In addition, it is also shown how the higher the TTL, the bigger should be the number of sources at that TTL so that covering them is cost-effective in terms of overhead.

In the next section, we will show through simulations that the proposed schemes are able to outperform existing ones. Furthermore, we will show that the maximal benefit approach performs better than the rest of approaches.

4 PERFORMANCE EVALUATION

In this section, we present the evaluation of the proposed approaches and we compare them with the reactive, proactive and hybrid ones. For this evaluation we have conducted extensive simulations of the different schemes under a variety of networking scenarios.

4.1 Simulation Environment and Scenarios

All of the gateway discovery mechanisms have been implemented and simulated in the NS-2 [9] network simulator. The simulated scenario consists of 25 mobile hosts randomly distributed over an area of 1200x500 m. The radio channel capacity for each mobile node is 2Mb/s, using the IEEE 802.11b DCF MAC layer and a communication range of 250 m. In addition, there are two gateways; one located at the coordinates (50, 450) and (1150, 50) respectively. In the hybrid approach both of them use a TTL = 2 for their advertisements as it is recommended in [6] for the scenarios under simulation. Each of the gateways is connected to a router and the routers are connected one to each other. Additionally, each router has a fixed node connected to it. All the fixed links have a bandwidth of 10Mb/s, which is enough to accommodate all the traffic coming from the mobile nodes.

Each of the approaches has been evaluated over the same pre-generated set of 840 scenarios with varying movement patterns and traffic loads. Mobile nodes move using a random waypoint model with changing pause times. Nodes start the simulation being static for *pause time* seconds. Then they pick up a random destination inside the simulation area and start moving to the destination at a speed uniformly distributed between 0 and 20 m/s (mean speed = 10m/s). After reaching its destination this behavior is repeated until the end of the simulation. Seven different pause times were used: 0, 30, 60, 120, 300, 600, and 900 seconds. A pause time of 0 seconds corresponds to a continuous motion whereas a pause time of 900 seconds corresponds to a static scenario. For each of these pause times 10 different scenarios where simulated. The results were obtained as the mean values over these 10 runs to guarantee a fair comparison among the alternatives.

Four different traffic loads where tested consisting of 5, 10, 15, and 20 different CBR sources communicating with nodes in the fixed network. Each of these CBR sources start sending data at an uniformly distributed time within the first 10 seconds of the simulation. Each of the sources generates 512 bytes data packets at a rate of 5 packets per second (20Kb/s).

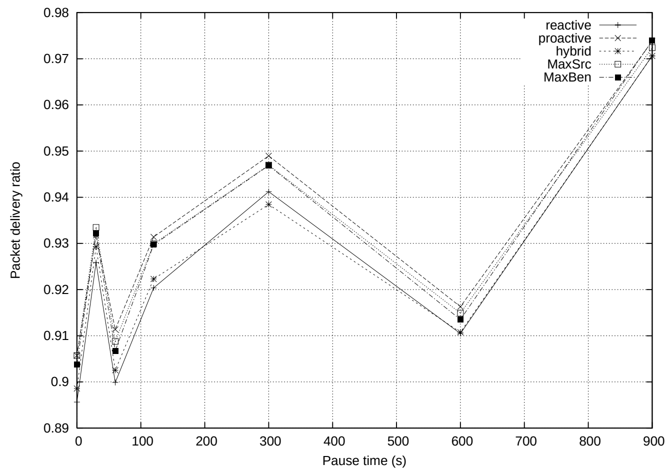
4.2 Performance metrics

To assess the effectiveness of the different gateway discovery mechanisms, we have used the following performance metrics:

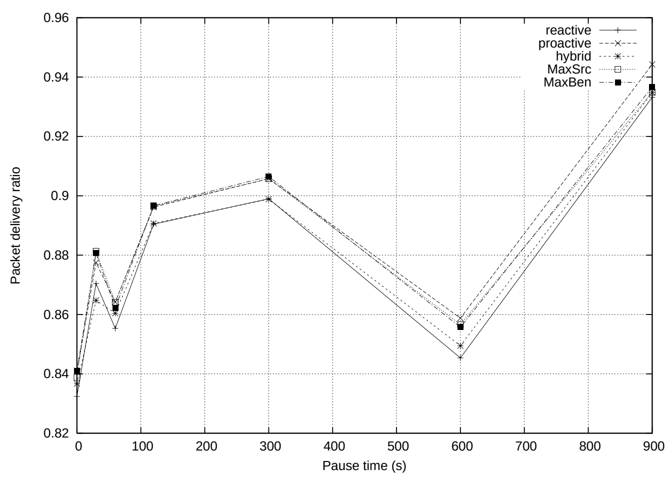
- Packet delivery ratio. Defined as the number of data packets successfully delivered over the number of data packets generated by the sources.
- Routing overhead. Defined as the total number of control packets, including gateway discovery, sent out during the simulation time.
- Normalized effectiveness. Defined as the number of packets successfully delivered minus the (weighted) number of control packets required divided by the total number of data packets generated by the sources. This metric gives a value of the overall performance by taking into account not only the packet delivery ratio but the overhead. The maximum value of 1 would only be achieved in the ideal case in which all data packets are delivered without any overhead.

4.3 Simulation Results

The simulation results show that our proposed approaches are able to offer a packet delivery ratio as higher as the proactive approach at a slightly higher overhead than the reactive and hybrid approaches. This is clearly shown in the case of 10 and 15 sources in figures 5(a) and 5(b) respectively. This differences in overhead are due to the fact that sometimes during the



(a) 10 sources



(b) 15 sources

FIGURE 5
Packet delivery ratio for different number of sources.

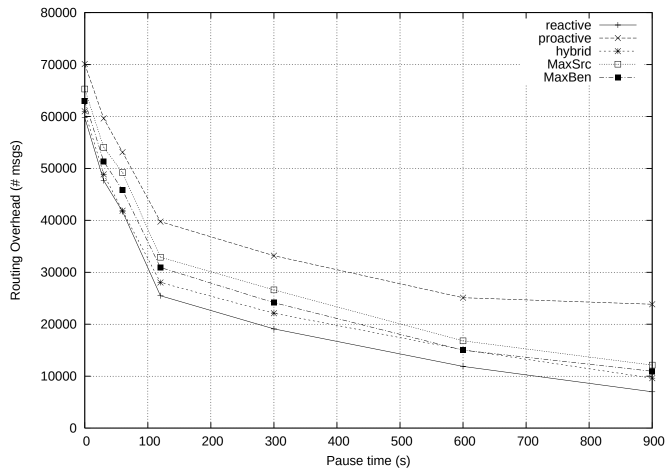
simulation it is required to use higher TTLs than the hybrid approach so that the GWADV messages can reach the sources.

As shown when comparing figures 5(a) and 5(b), the higher the number of sources, the best perform the proposed adaptive schemes compared to the others. In addition, the higher the mobility of the nodes, the best the performance of the adaptive approach. For 10 sources both of the proposed algorithms are almost obtaining the same packet delivery ratio than the proactive scheme and much better than the hybrid and reactive ones. For 15 sources the proposed approaches outperform all the others. The reason is that with 15 sources the reactive and hybrid approaches require too much overhead due to the need for the sources to reactively discover the gateways. The proactive approach also starts working worse because its high control packet load does not leave enough resources to carry all the data packets generated by the sources. However, the proposed approaches by having a lower overhead are able to find a good trade-off between the signaling overhead and the proactivity of the protocol. The small differences in the packet delivery ratio, are explained by the fact that we are using the same routing protocol in all the cases, thus, we only see the impact of the gateway discovery mechanism, which is not much. However, if we look at the routing overhead, we see how there are differences of several thousands of messages between different approaches.

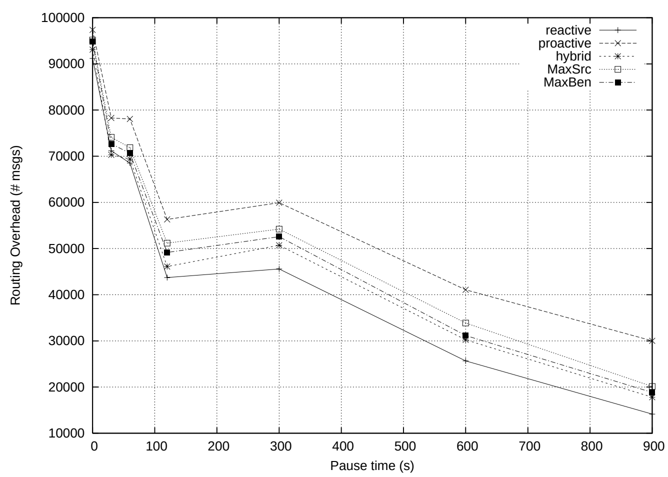
Regarding the routing overhead, a similar trend is observed. As it is depicted in figures 6(a) and 6(b), the proposed approaches has a lower overhead than the proactive approach and a little bit more than the reactive and hybrid ones. The differences in overhead are also lesser as the number of sources increase. As explained in our analytical model, this is due to the cost required in the reactive approach in which the sources are required to perform a network-wide search of the gateways. It is worth mentioning that the "maximal benefit coverage" algorithm outperforms all of the other approaches including the adaptive one based on the maximal source coverage. Key to this is the ability of the maximal benefit algorithm to limit the flooding of GWADV messages only to those sources to which it is cost effective. This is shown in figures 6(a) and 6(b) by a clearly lower overhead of the maximal benefit algorithm compared to the maximal source coverage one.

So, from the graphs in Fig. 5 and Fig. 6, it is clearly shown that the proposed algorithms are able to deliver almost the same packets than the proactive approach at the cost of an overhead close to the one of the reactive approach.

The packet delivery ratio is a good metric to evaluate the performance of the protocol from the outside. That is, what is the performance obtained by the applications. Conversely, the routing overhead is a good internal metric of how much network resources does the protocol need to do its work. So, in order to evaluate the overall performance of the different solutions we need a metric taking into account both internal and external performance metrics. As explained in the previous subsection, the metric that we will use is the normalized effectiveness. Of course, we adjust the weighting factor to give more importance to achieving a good packet delivery ratio, but there is a penalization for not doing it at a low overhead.



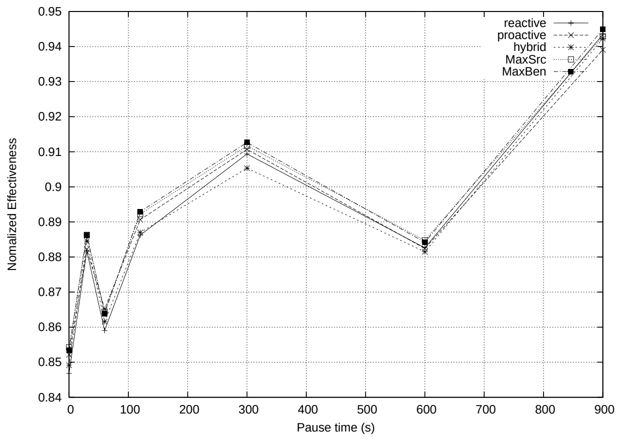
(a) 10 sources



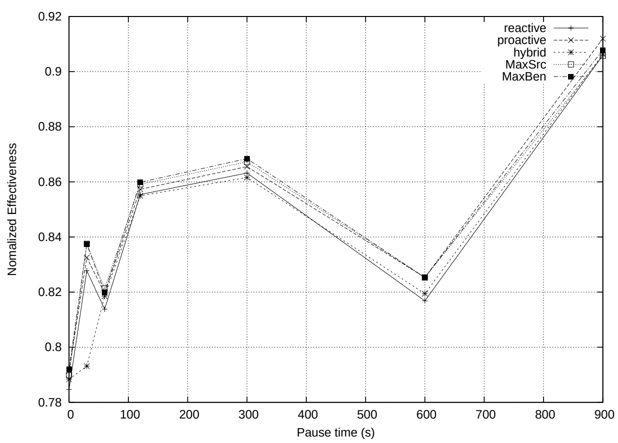
(b) 15 sources

FIGURE 6
Overhead for different number of sources.

In Fig. 7 we show the normalized effectiveness for different number of sources and different mobility rates. As the mobility of the nodes decrease (higher pause times) the performance of all the approaches improve. The cause is that the number of link breaks decreases and so does the control overhead required to re-establish the routes to to the gateways. It is also worth mentioning that the low performance of all the protocols at a pause



(a) 10 sources



(b) 15 sources

FIGURE 7
Normalized effectiveness for increasing number of sources.

time of 600 seconds is due to the fact that the random scenario generator produced several scenarios with unconnected sources and by no means that performance is related with the mobility of the network.

However, the most important result is that, as our analytical model predicted, adaptive approaches can obtain a good trade-off between the efficiency of the protocol in terms of packet delivery ratio and the signaling overhead. As our model also anticipated, the performance of the approaches is highly dependent upon the number of sources. In fact, the adaptive approach has shown to be the one which is less affected by the increase of the number of active sources compared to the others. Whereas for 5 sources most of the protocols obtain a high effectiveness, in the rest of the experiments the adaptive approach outperforms the other approaches. In addition, the proposed schemes also tends to be better than the others as the mobility of the nodes increase, which is precisely when the conditions are more demanding. Finally, the maximum benefit algorithm, by incorporating the notion of cost in terms of overhead has been shown to outperform all of the other schemes.

5 CONCLUSIONS

We have analytically modeled existing alternatives for gateway discovery in hybrid ad hoc networks. The evaluation has shown that previous approaches do not behave well as the number of sources increase and are not able to offer a good performance in the full range of possible scenarios. We have proposed two adaptive approaches being able to dynamically adjust the scope of GWADV messages either to reach the maximal number of active sources (maximal source coverage) or to obtain the maximal benefit in terms of overhead savings (maximal benefit coverage). These adaptive approaches rely solely on local information and they do not incur in any additional data overhead. We have shown through simulation that the proposed algorithms outperform existing schemes. In addition, as our model anticipated, we have shown that the proposed approach is more scalable in terms of mobility of the nodes and number of active sources connecting to the Internet than the other approaches.

As a future work we are considering the evaluation of the impact of the gateway discovery on other types of data sources (e.g. TCP traffic) and the integration of these approaches with other Internet connectivity proposals.

ACKNOWLEDGMENT

Part of this work has been funded by Spanish MCYT by means of the "Ramon y Cajal" workprogramme, the ICSI Call for Spanish Technologist and the SAM (MCYT, TIC2002-04531-C04-03) project.

REFERENCES

[1] U Jonsson, F Alriksson, T Larsson, P Johansson and G Maguire, Jr. (Aug 1999) "MIPMANET - Mobile IP for Mobile Ad hoc Networks," *in Proceedings of IEEE/ACM Workshop on Mobile and Ad Hoc Networking and Computing,* Boston, MA, USA.

[2] Y Sun, E M Belding-Royer and C E Perkins (2002) "Internet Connectivity for Ad Hoc Mobile Networks." *International Journal of Wireless Information Networks,* **9(2)**, 75–88.

[3] C Jelger, T Noel and A Frey (Oct 2003) "Gateway and address autoconfiguration for IPv6 adhoc networks," Internet-Draft, draft-jelger-manet-gateway-autoconf-v6-01.txt, Work in progress.

[4] R Wakikawa, J Malinen, C Perkins, A Nilsson and A Tuominen (Nov 2002) "Internet Connectivity for Mobile Ad hoc networks," Internet-Draft, draft-wakikawa-manet-globalv6-02.txt, Work in progress.

[5] J Broch, D Maltz and D Johnson (June 23-25) "Supporting Hierarchy and Heterogeneous Interfaces in Multi-hop Wireless Ad Hoc Networks," *in Proceedings of the IEEE International Symposium on Parallel Architectures, Algorithms, and Networks*, Perth, Australia, 370–375.

[6] P Ratanchandani and R Kravets (Mar 2003) "A Hybrid Approach to Internet Connectivity for Mobile Ad hoc Networks," *in Proc. of the IEEE WCNC 2003*, New Orleans, USA, **3**, 1522–1527.

[7] J Lee, D Kim, J J Garcia-Luna-Aceves, Y Choi, J Choi and S Nam (Apr 2003) "Hybrid Gateway Advertisement Scheme for Connecting Mobile Ad Hoc Networks to The Internet," *in Proc. of the 57th IEEE VTC 2003*, Jeju, Korea, **1**, 191–195.

[8] C Perkins and E M Belding-Royer (Feb 1999) "Ad hoc On-Demand Distance Vector (AODV) Routing," *in Proc. of the 2nd IEEE Workshop on Mobile Computing Systems and Applications*, 90–100.

[9] K Fall, K Varadhan (Nov 2003) "ns, Notes and Documentation", The VINT Project, UC Berkeley, LBL, USC/ISI, and Xerox PARC.