A Strategy for Traffic Load Balancing in Hybrid Ad Hoc Networks with Mobile Gateways

A. TRIVIÑO-CABRERA, E. CASILARI, A. ARIZA, D. BARTOLOMÉ
Departamento de Tecnología Electrónica, University of Málaga,
ETSIT, Campus de Teatinos, 29071 Málaga,
SPAIN

Abstract:- Ad hoc networks were initially conceived as stand-alone entities without connection to any external infrastructure. A possible solution to give access to ad hoc terminals to the Internet trough conventional WLAN networks is described in the mobile multi-gateway support. This scheme proposes to utilize a mobile node within the transmission range of the WLAN access point as a connecting gateway between the MANET and the Internet. As the gateway can move, its functionality can be transferred to another node present in the cover area of the access point acting as a router. In any case, all the traffic flowing to/from Internet passes through the same node, which can impose a bottleneck in the system. In this paper, a modification is suggested to distribute the Internet traffic among different gateways.

Key-Words: - Hybrid MANET, Internet Gateway, load balancing, AODV, NDP

1 Introduction

The original conception of Mobile Ad hoc NETworks (MANETs) implied a generic scenario in which a number of mobile devices could intercommunicate without the use of any pre-existent infrastructure. However, in most practical applications of ad hoc networks a connection to the Internet is required. Therefore, a crucial aspect in the development of these networks is the incorporation of new mechanisms that give access to the Internet in multi-hop wireless networks.

The integration with the conventional Internet requires to add to the MANET architecture an Access Router (AR) behaving as an Internet Access Point, that is, an interface between the wireless medium and the wired network. However the introduction of a router is not sufficient as a technique for connecting MANET terminals to the exterior must address two basic problems. Firstly, the mobile nodes require a global IPv6 address in order to be accessible from external networks. To guarantee the Internet routing hierarchy, an entity must be responsible for providing the necessary information to enable nodes to get an appropriate IPv6 address. This functionality could be associated to a dedicated entity, as in DHCP, for which a special server assigns the IPv6 global address to all the terminals within the domain under its control [1]. Although this strategy has been employed in different works, it requires an additional central equipment, which restricts the scenarios where MANET could operate, violating the self-configuring philosophy of MANETs.

An alternative to suppress this equipment is based on the address auto-configuration. In this proposal the Access Router cyclically sends Router Advertisements (RA messages) containing a suitable IPv6 address prefix [2]. In order to generate a global IPv6 address, all the terminals requiring a connection to Internet concatenate a random value, for example the EUI-64 MAC identifier, to this received prefix.

The reception of RA messages is also necessary for the nodes to detect the presence of the access router. So, a node perceiving changes in the RA messages will assume to be moved to another domain. In that case, the node triggers the corresponding procedure to carry on the on-going Internet connections. In the IPv6 context, the Mobile IPv6 technology specifies the procedures to communicate with the Internet if changes in the point of attachment take place [3]. However, the RA messages are formatted according to the Neighbor Discovery protocol (NDP). NDP is sufficient to provide connectivity to nodes of an infrastructure WLAN as NDP packets are restricted [2] to one hop and should not be forwarded. In multi-hop wireless LAN, where mobile nodes could be several hops away from the Access Router, this one-hop restriction cannot be admitted. To avoid this limitation, an extra element has to be included in the MANET: the gateway. A gateway is a MANET node in charge of broadcasting the information from the access point by disseminating an altered version of the RA packets: the so called Modified Router Advertisements (MRA messages), which are not one hop restricted [4]. The gateway will perform the ad hoc routing functionalities that are not present in the Access Router.

In most solutions that have been proposed to integrate MANETs into the Internet, the Gateway is defined as a dedicated entity that must be present in those scenarios where ad hoc networks are going to be deployed [4] [5] [6]. Under this consideration, the software of the gateway must be directly loaded in the router or located in a fixed node within the router transmission area. This assumption does not follow the MANET principle, i.e. to operate wherever and whenever without any special hierarchy among the nodes. To overcome this restriction, Mobile Multi-Gateway support could be applied [7]. In this solution, gateways' functions are transferred to a MANET node that is one hop away from the Access Router. As the gateway is now enabled to be a mobile node, the draft in [7] proposes a strategy to switch the role of gateway between the different nodes that receive the RA messages. As all both outgoing and incoming Internet traffic will be traversing the gateway, this node can become a bottleneck for the network performance. In this paper we propose a simple modification in the election of the gateway to distribute the traffic among the different candidate nodes as a function of their traffic load.

The rest of this paper is organized as follows. Section 2 summarizes the principal characteristics of the Mobile Multi-gateway Support. Section 3 describes a proposal for a proper load balancing. Section 4 presents the results of different simulations aimed at evaluating the benefits of traffic balancing. Finally Section 5 summarizes the main conclusions of the paper.

2. Mobile Multi-Gateway Support

This mechanism [7] was created to simplify the restrictions associated to heterogeneous environments where multi-hop ad hoc networks may coexist with fixed networks. By the transfer of Gateway functionalities to a MANET node, the introduction of a dedicated Gateway is suppressed. Therefore, the number of situations and applications where MANETs can perform increases. The resulting architecture could be useful in quite different scenarios where a rapid, robust and auto-configurable connection to Internet is required. Examples of this could be emergency scenarios (e.g.: natural disasters, terrorist attacks, etc) where the ad hoc nodes (in this case the medical equipments) could retransmit via Internet to a hospital the patients' biosignals received from other nodes. In these situations the gateways have a complete freedom of movements. The method to select the node that will operate as the Gateway as well as the treatment associated to its mobility are the main concerns of the Mobile Multi-Gateway Support.

Among all the devices that are in the coverage area of the Access Router, one of them is temporally selected to behave as the Internet Gateway. This node is called the Default Gateway (DG) whereas the rest of the terminals sited in this transmission range are referred as the Candidate Gateways (CG). So, at a certain instant of time, one DG and none or multiple CG may coexist.

All the Gateways (Default or Candidate) hear the RA messages generated by the Access Router but only the DG propagates this information to the network. For this purpose, the DG periodically sends unsolicited MRA (Modified Router Advertisement) messages, which can be forwarded through multiple hops through the ad hoc networks. MRA messages can be typically implemented by adapting NDP Router Advertisement packets as well as by modifying the messages of route announcement in the ad hoc routing protocol. For example, [4] proposes a slight change in the format of AODV Route Replay (RREP) messages to broadcast the identity of the DG.

By the reception of a MRA message, a mobile device can build up its global IPv6 address and update the corresponding route to the DG. When a terminal generates traffic to an external (Internet) host, it will forward the packets to the DG using this route. If the node does not know the identity (IP address) of the DG (e.g.: because the MRA packet was lost), it can initiate a procedure to determine a route to the Internet. This procedure to discover the path to the DG is performed by broadcasting a special message to the network. As in the case of the MRA, this message can be put into operation by altering the typical Router Solicitation that is defined by NDP, generating the so-called Modified Router Solicitation (MRS) messages. Similarly, it could be implemented [4] with a modified version of the Route Request (RREQ) packets of AODV. In any case, as soon as one of these broadcasted packets reaches the DG, a unicast MRA message is backpropagated to the soliciting node. As it will be described in next section, this method could be also utilised to detect the presence of alternative routes to Internet through the use of CGs. On the other side, when the Access Router receives any packet whose destination is a MANET component, it will forward it to the DG, which will perform the conventional ad hoc routing to the destination ad hoc node.

The information conveyed in the MRA messages is considered to be valid during a certain time interval (*modified advertisement interval*). The MRA packets inform nodes about this interval (which must be longer than the period between two consecutive RAs) by means of a lifetime field. Before this period (plus a 'guard' time) expires, the mobile nodes expect to receive a new MRA message from the DG. While the DG resides in the coverage area of the Access Router, the nodes in the MANET will periodically receive its broadcasted MRA messages. On the contrary, if the DG stops receiving RA packets from the Access Router, it will assume that it has escaped from its transmission area and will stop sending MRA messages as its gateway functionalities have ceased.

As no MRA will be received before the expiration of the corresponding modified advertisement interval, all the network nodes (including CGs) will consider there is no active DG. At that moment, the CGs should start the gateway selection procedure. The interchange of messages under this scheme for providing Internet connectivity has been depicted in Figure 1.

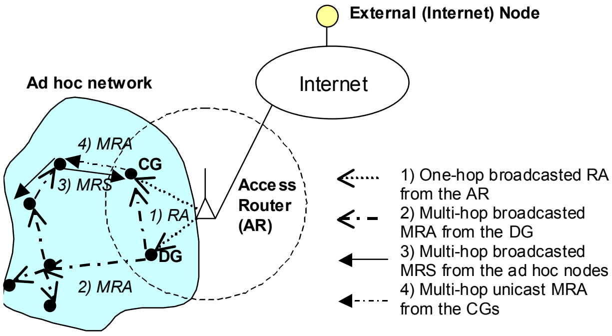

Fig. 1. Messages for gateway configuration under Mobile multiple gateway support

This procedure is a distributed algorithm that respects the MANET nature as it does not need any centralised entity to select the future DG. The mechanism consists of triggering a random internal timer. Every CG verifies if any MRA message is detected during this random period. From the reception of a MRA it would infer the existence of a new configured DG. In other case, the CG will automatically self-configure itself as the new DG by broadcasting its own MRA message immediately. The Internet draft does not contemplate the case of receiving multiple MRAs from different CGs (e.g.: because a MRA packet loss occurs or two CGs keep waiting during very similar random intervals). In that case, we propose a simple election based on the IPv6 address: candidates (including any recently self-configured DG) will give priority to the DG with a lower IPv6 address.

3. A proposal for load balancing

The issue of load balancing is not new in the area of MANETs. In this sense, most studies in the literature have focused on evaluating the benefits of distributing the traffic from the ad hoc nodes by means of multi-path routing.

Due to the wireless nature of transmissions in ad hoc networking, the main problem when using multi-path routing is that alternate routes between two terminals can interfere even if they traverse a completely disjoint set of routing nodes. The study in [8] analytically confirmed that the possible advantages of multipath are strongly related to the number of simultaneous utilised paths (with an optimum in 3) and the average path length. In [9], Pearlman et al. showed that the improvements of multipath routing are insignificant for single-channel networks because of the coupling between the paths. Similarly, in [10], Ganjali et al. demonstrated that a multipath policy does not necessarily imply a smoother traffic distribution within the network. These authors suggest that the benefits of multipath could be achieved only if the alternative routes push the traffic from the "center" of the network, that is to say, where the node (and interference) density is higher. This could be put into practice if nodes implement location based routing. Unfortunately, this routing demands the knowledge of the physical position of the nodes, which requires a special hardware (e.g.: GPS or other positioning system) in each node. In this line, Jones et al. [11] formulate an algorithm that minimises the interference 'correlation' between the paths basing on the network connectivity. However, the algorithm requires a perfect and updated knowledge of the network topology in the nodes, which is only possible if a link-state routing protocol is utilised (not being the case of the reactive protocols). Moreover, multi-path techniques are not natively supported in some ad hoc routing protocols such as AODV.

Apart from these studies on multi-path routing, other works base load balancing strategies on slight modifications in the treatment of Route Request (RREQ) Messages of reactive routing protocols. So, authors in [12] suggest that overloaded nodes should not propagate RREQ messages while the study in [13] proposes that all the nodes retransmitting a RREQ packet should add information to the message concerning their queue occupancy. As a result, the destination node, after receiving different RREQs from different routes, it will backpropagate the Route Reply choosing the route with less queue occupancy.

However, no much research efforts have been devoted to load balancing for the particular case of hybrid networks. In this type of networks, the problem of traffic concentration is specially critical as all the Internet flows (which could correspond to a high percentage of the global traffic in the network for most practical applications) is routed to the same node (the DG). In [11], for example, it is suggested that network efficiency could be dramatically improved by means of load balancing in the scenario where a stub multi-hop network connects to Internet through different gateways.

Aiming at minimizing the congestion around the default gateway, we propose to distribute the Internet traffic among the different CGs basing on the traffic load. In this new architecture, all the nodes initiating an Internet connection will send a Router Solicitation (by emitting a MRS message) independently if they know the existence of the DG. As CGs are allowed to respond (in a unicast way) to MRS messages, nodes will be able to select a particular CG for their packets to Internet.

The election of the gateway node to utilise is performed at the nodes depending on the overall traffic that the gateways are supporting. For this purpose, all the gateways have to regularly compute the bandwidth they are consuming. This computation should be updated in an internal variable for observation windows of a fixed duration. Although different traffic estimators could be employed, we initially consider a classic first order autoregressive filter:

$$X(i) = \alpha \cdot X(i-1) + \beta \cdot BC(i) \qquad (1)$$

being *X(i)* and *BC(i)* the estimator of the bandwidth and the byte count of the routed packets for the *i*-th time window, while α and β are two coefficients describing the relative importance of the previous and the present traffic measurements. Consequently, the computation in the nodes is basically limited to a simple byte count of the packets they are transmitting/receiving to/from the AR.

When a candidate gateway has to respond to a MRS message from a node, it will include information on this measurement adding a field in the MRA packet (in case of using modified NDP messages, the measurement should be incorporated in the options field [2]). As the DG will include its own updated bandwidth estimator in its broadcasted MRA packets, all the ad hoc nodes (including the CGs) will know at any time the load that the DG is supporting. CGs that receive a MRS from a node willing to connect to the Internet will compare their traffic load with that of the DG. Thus, they will only send a MRA answer in case of being less saturated that the DG, which will reduce the overhead provoked by MRA messages. Nodes that have sent MRS will wait for a short time interval to receive the MRAs from the gateways. In case of receiving several answers, the node will elect the gateway with the lowest value of the estimated bandwidth.

In order to avoid sudden and spurious re-distribution of traffic within the network, nodes will keep using the selected gateway until the route breaks. So nodes will start the gateway solicitation just when they have no configured gateway or when the route to the most recent configured gateway is lost. As the behaviour of the DG is respected and its MRA are periodically broadcasted, nodes that do not implement this algorithm could maintain their connections to the Internet through the DG. Thus, this strategy for load balancing could be gradually incorporated to the nodes. On the other hand, a problem is derived from the utilization of distinct gateways in both uplink and downlink communications. If no special routing information is provided within the packets, all the ad hoc nodes that are routing packets to Internet will utilise the route to the default gateway. Similarly, all the incoming traffic will be diverted by the Access Router to the DG. So, every ad hoc terminal should inform both the ad hoc network and its correspondent Internet node about the Candidate Gateway to employ. This action could be accomplished by the inclusion of a proxy routing header in the packets [14].

4. Simulations and Results

In order to evaluate the performance of the proposed mechanism, a software module to include mobile multi-gateway support was developed and integrated into the Network Simulator 2 (Ns-2). Ns-2 is by far the most extended simulation software in the studies on ad hoc networking. Similarly, the proposed algorithm for load balancing between the gateways was incorporated to the software.

Simulations considered a typical 1500 m x 300 m bi-dimensional space, placing the access router in the center of the area. The ad hoc network was formed by 50 simulated mobile nodes. Two opposite scenarios have been evaluated: one with no pause time (for which the nodes are continuously in motion) and a static scenario (without mobility).

In the first scenario, which could represent a typical application in which the mobility of the nodes is determined by human actions (e.g.: laptops during a conference), the movements were based on the popular Modified Random WayPoint model. For our experiments, we considered diverse types of mobility changing the constant velocity of the nodes from 1 m/s (habitual speed of a walking man) to 20 m/s (case of a scenario with motorized vehicles). In the second scenario, which could correspond to a network of fixed sensors, as there is no mobility, the most determining parameter is the node distribution. Up to 10 different random topologies were simulated to estimate the mean network performance.

Due to its popularity, AODV was chosen as the ad hoc protocol (similar results were obtained with DSR). As it respects to the traffic pattern, 10 Constant Bit Rate (CBR) data connections were considered. The connections were designed to simultaneously generate both uplink and downlink traffic between 10 randomly chosen nodes and a generic exterior (Internet) host. For the simulations, all the traffic is transmitted between the mobile node and the Access Router. So, the analysis just computed the packets losses and delays which took place in the hops within the ad hoc network (Internet losses and delays are not modeled). The rest of simulation parameters are summarized in Table 1.

The performance of the network is measured by means of the following metrics:

-PDR (Packet Delivery Ratio), estimated as the ratio between the packets received at the destination and the packets generated by the sources.
-Mean End-to-End (E2E) packet delay.
-Normalized overhead, defined as the percentage of control packets with respect to the received data packets. Each hop of any control packets is computed as a new control packet.

Table 1. Other simulation parameters

Initial Node distribution	Uniform (for both fixed and mobile scenarios)
Traffic pattern (uplink /downlink)	10 CBR sources. Rate = 8 packets/s. Packet size= 512 Bytes
Simulation Time	5000 s
Transmission Range	250 m
Number of runs/point	3
Link Level Layer	802.11a (RTS/ CTS enabled)
Internal Node Queue	64 packets
RA Advertisement intervals	For RA: 2 s For MRA: 4 s
Parameters for traffic Measurements in the gateways	Coefficients for the filter: α=0, β=1 (α=0 as CBR and constantly emitting sources are employed). Observation Window: 10 s

Figures 2-4 show these estimated metrics for the case of the first scenario. Figures 2 and 3 (for PDR and delay respectively) clearly indicate that load balancing techniques are appropriate for situations of low-medium mobility (with node velocities under 10 m/s) although it augments the protocol overhead, especially due to the broadcast of router solicitation (MRS). For low velocities, load balancing can even reduce in a 50% the packet delay just at the cost of a very slight increase in the overhead (see Figure 4). On the contrary, in situations of very high mobility (with typical velocities of motored vehicles, greater than 10 m/s), the behaviour of load balancing clearly underperforms that of the case in which all the traffic is routed through the same node (the DG). This could be explained by the fact that, as the velocity grows, the route duration rapidly decays. In that case, the process of gateway re-configuration (through broadcasted MRS messages) that the load balancing requires after a route break may become highly inefficient.

Effectively, in the case without load balancing, a node can recover from the loss of the route to the DG through the reception of the periodic unsolicited MRA from the DG. In the scheme with balancing, every link break implies the broadcasting of new MRS solicitations. In addition, in this scenario of high mobility it is also more probable that a node selects a CG which is eventually traversing the coverage area of the AR at an elevated speed. As soon as this node escapes from the transmissions of the AR, the process for gateway solicitation is triggered.

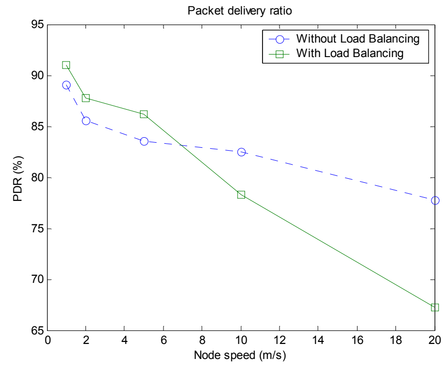

Fig. 2. Packet delivery ratio as a function of the node velocity

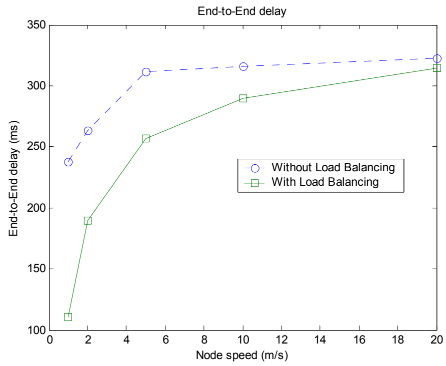

Fig. 3. End-to-end packet delay as a function of the node velocity

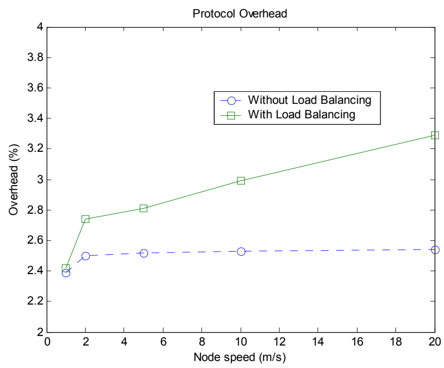

Fig. 4. Normalized Overhead as a function of the node velocity

On the other hand, the case of a static scenario can be regarded as the optimal situation for the utilisation of load balancing. Table 2 condenses the mean results of the performance metrics for ten arbitrary distributions of the nodes within the simulation area. In practically all the tested topologies, the routing with load balancing outperformed the architecture that just employs the DG for supporting the connections to the Internet. The mean statistics (those indicated in Table 2) clearly show a dramatic improvement of the PDR and, specially, the packet end-to-end delay for a practically identical overhead. This implies that, for the case of a sensor network (which could be an evident example of this type of static ad hoc network), the benefits of load balancing could be achieved with a negligible increase of the overhead traffic and consequently, with no extra consumption of the battery power in the nodes (a critical aspect in the management of a sensor network).

Table 2. Results for the static scenario

	Without load balancing	With load balancing
PDR	78.18 %	87.02 %
e2e delay	1214.6 ms	354.79 ms
Overhead	2.94%	2.93%

5. Conclusions

This paper has evaluated the benefits of load balancing in scenarios of hybrid ad hoc networks in which the MANET nodes access to the Internet by means of mobile multi-gateway support. Different simulations in both static and dynamic scenarios show that the use of load balancing notably improves the network performance (measured in terms of Packet Delivery Ratio and End to End packet delay) at the cost of a moderated increase in the protocol overhead. Only in an environment where the nodes exhibit high mobility, the load balancing technique is showed to be more inefficient than the simple utilisation of a single (default) gateway.

For a node wishing to communicate with an external Internet terminal, the proposed architecture suggests a criterion to select the mobile gateway that will provide access to a conventional infrastructure WLAN Access Point. In particular, the election is based on the traffic load that the gateways are retransmitting. The implementation of the proposed method is very simple as long as it just requires that the gateway nodes include information about their supported traffic in the advertisement messages that they emit to inform about their status. Further studies should contemplate the possibility of Variable Bit Rate sources as well as reactive traffic sources (such as TCP data sources).

Acknowledgments

This work was supported by the Research National Project No.TIC2003-07953-C02-01.

References

[1] Alexander, S. et al. ”DHCP Options and BOOTP Vendor Extensions”, IETF RFC 2132, 1997.

[2] Narten, T. et al., “Neighbor Discovery for IP version 6”, IETF RFC 2461, Dec. 1998.

[3] Johnson, D. et al. “Mobility Support in IPv6”, IETF RFC 3775, June 2004.

[4] Wakikawa, R. et al., “Global Connectivity for IPv6 Mobile Ad hoc Networks”, IETF Internet Draft, January 2005.

[5] Jelger, C. et al., “Gateway and address autoconfiguration for IPv6 adhoc networks”, IETF Internet Draft, October 2003.

[6] Cha, H.W. et al.,“Extended Support for Global Connectivity for IPv6 Mobile Ad Hoc Networks”, IETF Internet Draft, October 2003.

[7] Singh, S. et al., “Mobile multi-gateway support for IPv6 mobile ad hoc networks”, IETF Internet Draft, June 2004.

[8] Pham, P. P. et al., “Performance Analysis of Reactive Shortest Path and Multi-Path Routing Mechanism with load Balance”, in Proc. of the IEEE INFOCOM’03, San Francisco, March 2003.

[9] Pearlman, M. R. et al., “On the impact of alternate path routing for load balancing in mobile ad hoc networks”, Proc of ACM MobiHOC, Boston, (USA), August 2000.

[10] Ganjali Y. et al., “Load balancing in ad hoc networks: Single-path routing vs. multi-path routing”, In Proc. of IEEE INFOCOM'04, Hong Kong, March 2004.

[11] Jones, E.P.C. et al., “Multipath load balancing in multi-hop wireless networks”, Pooc. of IEEE WiMob'2005, Montreal (Canada), August 2005.

[12] Lee, Y.L. et al., “A Workload-based Adaptive Load Balancing Technique for Mobile Ad Hoc Networks”, Proc. of WCNC 2005, March 2005.

[13] Lee, S., et al., “Dynamic load-aware routing in ad hoc networks”, Proc. of IEEE ICC, Helsinki (Finland), June 2001.

[14] Johnson, D. et al., “Mobility Support in IPv6”, IETF RFC 3775, June 2004.