Providing Internet connectivity and mobility management for MANETs

Quan Le-Trung[1], Paal E. Engelstad[2], Vinh Pham[3],
Tor Skeie[4], Amirhosein Taherkordi[1], and Frank Eliassen[1]
[1]Department of Informatics, University of Oslo [IFI/UiO], Oslo, Norway
[2]Telenor Research & Innovation [R&I], Fornebu, Norway
[3]University Graduate Center [UniK], Kjeller, Norway
[4]Simula Research Laboratory, Lysaker, Norway

Abstract

Purpose – The purpose of this paper is to describe the required functionalities on providing Internet connectivity and mobility management for mobile ad-hoc networks (MANETs), present discovered problems such as inconsistent contexts, and provide the corresponding solutions. It also provides a hybrid metric for the load-balance of intra/inter-MANET traffic over multiple Internet gateways.

Design/methodology/approach – The paper uses both mathematical analyses and simulations to discover the required functionalities and problems on providing Internet connectivity and mobility management for MANETs. The proposed hybrid metric for Internet gateway (IGW) selection is a replacement of the shortest hop-count metric, and consider three factors: hop-count distance, intra-MANET traffic, and inter-MANET traffic.

Findings - Simulation results show that ad-hoc routing protocols, using our proposed metric, get better performance in terms of packet delivery ratio and transmission delay, at the cost of slightly increased signalling overhead.

Research limitations/implications – In the assessment, simulation results are taken from two mobility scenarios, and the hybrid metric is integrated into only re-active ad-hoc routing. Thus, more case studies need to be carried out to demonstrate the outcomes of our proposed metric compared with others.

Practical implications – This paper provides the needed functionalities for broadening the richness of MANET applications to Internet users, and vice verse.

Originality/value – This paper contributes to the research on internetworking and mobility management between mobile ad-hoc networks (MANETs) and the Internet.

Keyworks Algorithms, Management, Measurement, Performance, Design

Paper type Research paper

1. Introduction

With the development of mobile communications and Internet technology, there is a strong need to provide connectivity for roaming devices to communicate continuously with other devices on the Internet. However, the mobility of Internet hosts is usually within the same broadcast domain where the Internet gateway is located, referred to as 1-hop Internet mobility management. Technology advances have taken to the use of mobile ad-hoc networks (MANETs) as the access networks for the Internet, where MANETs are used to either cover the empty areas or extend the access networks from 1-hop to multi-hop in the current access technologies such as wireless LANs or cellular networks (Abduljalil and Bodhe, 2007; Le

Based on "Load-Balance of Intra/Inter-MANET Traffic over Multiple Internet Gateways", by Le-Trung, Q., Engelstad, P. E., Skeie, T., and Taherkordi, A., which appeared in the Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia, MoMM 2008, November 24–26, 2008, Linz, Austria. (c) 2008 ACM 978-1-60558-269-6/08/0011.

et al., 2006). Typically, the connection between a MANET node and an Internet gateway (IGW) is multi-hop. Therefore, there is normally no direct wireless link from this MANET node to the IGW. Instead, they are connected via other intermediate nodes. Thus, different problems, e.g., inconsistent context, and cascading effect, can happen during the mobility of ad-hoc nodes within a MANET domain if multiple IGWs exist (Le-Trung and Kotsis, 2008; Engelstad and Egeland, 2004; Engelstad *et al.*, 2004).

Since a MANET might be used for both direct communication between MANET nodes and for Internet connectivity, it might be useful to make a distinction between the intra-MANET traffic, which is the traffic constrained within a MANET, and the inter-MANET traffic, which is the traffic between the MANET and the Internet. (In fact, inter-MANET traffic might also include traffic between two different MANET domains, or between a MANET domain and another type of external network, such as a local wired LAN. However, this paper assumes for simplicity that all inter-MANET traffic is traffic between a MANET and the Internet). Research has been in-progress for the load-balancing of intra-MANET traffic within a MANET domain (Royer *et al.*, 2001; Chiu and Gen-Huey, 2003), and that of inter-MANET traffic over multiple IGWs (Hsu *et al.*, 2004). However, intra/inter-MANET traffic are considered separately. Moreover, the load-balancing of inter-MANET traffic over multiple IGWs does not consider many realistic problems like inconsistent context problems. In this paper, we want to control together these types of traffic. For this purpose, a hybrid metric for the load-balancing of intra/inter-MANET traffic among multiple IGWs, and alternative solutions to reduce realistic problems in the implementation, are proposed and evaluted through the simulation.

This paper is structured as follows. Section 2 gives a full description of required functions and related work on providing Internet connectivity and mobility management for MANETs. In Section 3, a hybrid IGW selection metric is proposed as a replacement of the shortest hop-count. It is used for the load-balancing of intra/inter-MANET traffic in situations where there are multiple IGWs on the same MANET domain. An ns-2[1] implementation of the ad-hoc on-demand distance-vector (AODV) routing protocol (Perkins *et al.*, 2003) that uses the above metric is presented in Section 4. It also uses mobile IP (MIP) for the different IGW selection strategies, and the implementation is developed from the AODV and MIP package[2]. The implementation also includes techniques, such as half-tunneling, to reduce the inconsistent context and cascading effect problems. Section 5 presents the simulation scenario for testing the load-balancing of both intra-MANET traffic, i.e., constant bit rate (CBR) traffic, and inter-MANET traffic, i.e., TCP traffic. This scenario comprises multiple IGWs, and a set of fixed and mobile MANET nodes with different sources of MANET traffic, together with Internet hosts. The performance parameters of AODV using our proposed metric, in terms of packet delivery ratio, average packet transmission delay, and signalling overhead, are compared with those of AODV using the shortest hop-count metric. Finally, conclusions and directions for future work are given in Section 6.

2. MANET-Internet connectivity and mobility management

In this section, the required functions of providing Internet access and mobility management for MANET nodes are described first. They include: 1) MANET node location determination, 2) IGW discovery, 3) IGW selection, 4) IGW forwarding strategy, 5) address auto-configuration, and 6) handoff-style. Next, the related work is discussed following the descriptions of above functions. Figure 1 illustrates a typical mobility scenario of a MANET node (MN) (1) while connecting to the Internet. This is a scenario where these functions are needed. In the figure, MIPv4[3] is used for the macro-mobility management, i.e., between MANET domains, while ad-hoc routing[4] is used for the micro-mobility management, i.e., within each MANET domain. This is a popular solution for mobility management in MANETs (Le-Trung and Kotsis, 2008).

[1] http://www.isi.edu/nsnam/ns/

[2] http://core.it.uu.se/core/index.php/AODV-UU_and_Mobile_IP_for_ns-2

[3] http://www.ietf.org/html.charters/mip4-charter.html

[4] http://www.ietf.org/html.charters/manet-charter.html

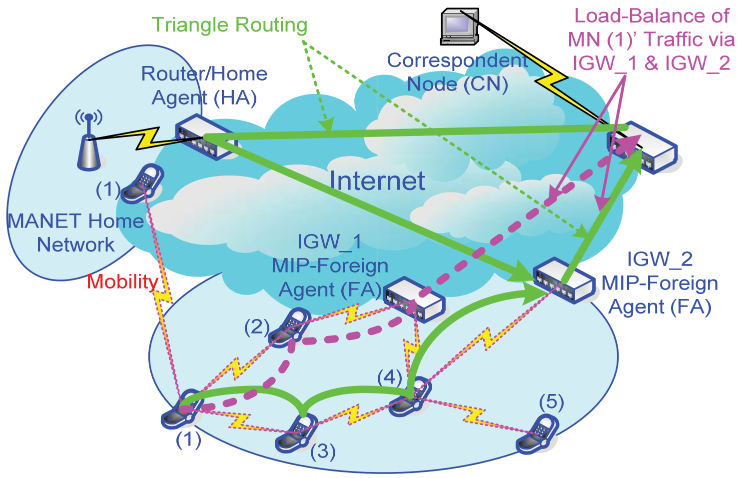

Figure 1. A scenario of MANET mobility management

2.1. Required functions

MANET node location determination is the function that allows a source MANET node to determine whether a destination node is located within the same MANET domain as the source node or outside the MANET domain (e.g., on an external network, such as an Internet host). The function can be implemented by one of the following methods:

Network prefix. All MANET nodes share the same network prefix. With this method, each MANET node must be assigned a global unicast IP address, both home address (HAddr) and care-of address (CoA), i.e., the MANET node address is topologically correct (Perkins *et al.*, 2002).

Routing table in MANET proactive ad-hoc routing protocols (Clausen *et al.*, 2006; Perkins and Bhagwat, 1994). If an entry for the destination is in the routing table of the source MANET node, the destination is either in the same MANET domain, or an Internet host reachable via the IGW. Otherwise, the destination is unreachable.

Flooding the route request (RREQ) and waiting for the route reply (RREP) in reactive ad-hoc routing protocols (Perkins *et al.*, 2003; Perkins *et al.*, 2000). If a host route is returned, the destination is in the same MANET domain. If a default route is returned, the destination is either unreachable or an Internet host reachable via the IGW.

Internet gateway. The IGW responds to a RREQ, sending a proxy RREP to signal that it can route to the requested destination, i.e., analogous to functionality of a proxy ARP, but over the multi-hops. To do this, an IGW must determine that the destination is not in the same MANET by keeping a list of currently known active nodes (called visitor list), or by pinging the destination on the IGW network interface attached to Internet, or by flooding the whole MANET with a new RREQ (Abduljalil *et al.*, 2007; Perkins *et al.*, 2002).

Internet gateway discovery is the function that allows a MANET node to discover an IGW to which traffic bound for the Internet can be forwarded, and from which traffic returned from the Internet can be received. The different discovery mechanisms can be classified into three sub-classes: proactive, reactive, or hybrid (Jin and Christian, 2002). In the proactive approach, each IGW broadcasts periodically an advertisement, while in the reactive approach a MANET node sends a solicitation and waits for a reply from the IGW. The former requires much overhead traffic on the MANET, while the latter entails the longer discovery delay. The hybrid approach compromises with the balance, in which each IGW periodically broadcasts the advertisement within the radius of n-hops. MANET nodes that are located further than n-hops away from the IGW, must use the reactive approach to discover the IGW (Ruiz and Gomez-Skarmeta, 2005).

Internet gateway selection is a function used when a MANET node discovers multiple IGWs for accessing the Internet. A metric is normally needed in order to select the right one. Different metrics can be used:

Shortest hop-count. To the nearest IGW (Perkins *et al.*, 2002).

Load-balancing. For intra-MANET traffic, choose different immediate relays node to destination MANET nodes within the same MANET domain (Royer *et al.*, 2001; Chiu and Gen-Huey, 2003), while for inter-MANET traffic, choosing different IGWs for forwarding traffic from MANET to Internet and vice verse (Hsu *et al.*, 2004).

Service class. Depending on the service classes and data caches provided and managed by each IGW, respectively (Chand *et al.*, 2007), as well as the wireless link quality among MANET nodes and IGWs (Natsheh and Wan, 2008).

Euclidean distance. Spatial distance between the MANET node and the IGW (Ammari and Rewini, 2004).

Hybrid. A combination of some of the above metrics (Ammari and Rewini, 2004).

Internet gateway forwarding strategies is a function that takes the responsibility to forward traffic within the MANET, out of the MANET to the Internet, or from the Internet into the MANET. Typically, it can be classified into *inter-MANET* and *intra-MANET forwarding* strategies. The *inter-MANET forwarding* strategies uses different approaches as follows:

Default routes. Representing the default next-hop to send packets to that do not match any other explicit entry in a MANET node's routing table. Usually, the default route is used to forwards packets towards an IGW, where packets are further forwarded towards the destination in the Internet (Perkins *et al.*, 2002; Benzaid *et al.*, 2004).

Tunneling (or encapsulation). Usually, the IP-in-IP encapsulation technique is used to get traffic into and out of the MANET. The outer IP header is for the tunneling connection between the source MANET node and the IGW, while the inner IP header is for the connection between the source MANET node and the destination (Jönsson *et al.*, 2000).

Half-tunneling. Traffic to the Internet from the MANET domain uses tunneling, while traffic from the Internet to the MANET domain uses ad-hoc forwarding without tunneling (Jönsson *et al.*, 2000).

Source routing. A list of all intermediate nodes between the source MANET node and the IGW are added into the IP header. At the IGW, the source routing header is removed and the packet is forwarded further to the Internet as a normal packet (Broch *et al.*, 1999).

Spanning tree rooted at the IGW. A tree rooted at the IGW is built and maintained using the agent advertisements broadcasted periodically by the corresponding IGW (Ergen and Puri, 2002).

The *intra-MANET forwarding* strategies, on the other hand, is based entirely on the operation of ad-hoc routing protocols[5]. These can be classified as proactive, or reactive, or hybrid. In the proactive approach, each node continuously maintains up-to-date routing information to reach every other node in the network. Routing table updates are periodically transmitted throughout the network in order to maintain table consistency. Thus, the route is quickly established without any delay. However, for a highly dynamic network topology, the proactive schemes require a significant amount of resources to keep routing information up-to-date and reliable. In the reactive approach, a node initiates a route discovery throughout the network, only when it wants to send packets to its destination. Thus, nodes maintain the routes to only active destinations. A route search is needed for every new destination. Therefore, the communication overhead is reduced at the expense of delay due to the route discovery. Finally, in the hybrid approach, each node maintains both topology information within its zone via the proactive approach, and the information regarding neighbor zones via the reactive approach.

Address auto-configuration. In order to enable a MANET to support IP services and the internetworking with the Internet, a MANET address space based on IPv4/IPv6 is required. Moreover, the MANET addressing schemes must be auto-configured and distributed to support for the self-organized and dynamic

[5] http://www.ietf.org/html.charters/manet-charter.html

characteristics of MANETs. Numerous addressing schemes for MANETs based on IP address auto-configuration have been proposed in the literature. They can be classified into two approaches: *conflict-detection allocation* and *conflict-free allocation* (Weniger and Zitterbart, 2004).

Conflict-detection allocation mechanisms are based on picking an IP address from a pool of available addresses, configuring it as tentative address and asking the rest of the nodes of the network, checking the address uniqueness and requesting for approval from all the nodes of the network. In case of conflict, e.g., the address has been already configured by another node, the node should pick a new address and repeat the procedure (as a sort-of "trial and error" method). This process is called duplicate address detection (DAD). *Conflict-free allocation* mechanisms, on the other hand, assume that the addresses are delegated uniquely, and that they are therefore not being used by any other node in the network. This can be achieved by ensuring that the nodes, that delegate the addresses, have disjointed address pools. In this way, there is no need of performing the DAD procedure.

Research has also been in-progress to apply IP address auto-configuration scheme for the addressing of MANETs. However, only stateless mechanism is suitable for MANETs (Bernardos and Calderon, 2005). This is because the stateful mechanism requires a centralized server to maintain a common address pool, while the stateless mechanism allows the node to construct its own address and is suitable for self.organized MANETs. However, to use the IP address stateless auto-configuration scheme for MANET addressing, i.e., a *conflict-detection allocation* approach, a DAD mechanism is required to assure the uniqueness of the address with multi-hop distance, especially to support for MANET merging and partitioning.

Finally, the address allocation space is important. It must be large enough to cover the large-scale MANETs and reduce the probability of address conflicts. The following IPv4 and IPv6 addressing spaces have been proposed for MANETs (Perkins *et al.*, 2002): *169.254.0.0/16* for IPv4, and *FEC0:0:0:FFFF::/64 (MANET_PREFIX)* for IPv6.

Handoff-style. A node performs a handoff if it changes its IGW while communicating with a correspondent node (CN) in the Internet. In conventional mobile networks, e.g., WLANs, the quality of the wireless link between a mobile node and the neighboring access points (APs) determines when to handoff from one AP to another. The performance of these types of handoffs depends on the mobility management protocol in the access network. In MANETs, on the other hand, the situation is more complicated. In general, some nodes do not have a direct wireless link to an AP, but they are connected via other intermediate nodes. Thus, they cannot initiate handoffs that are based on the link quality to the AP. Rather, the complete multi-hop path to the AP, which serves the current IGW, must be taken into consideration. A handoff can occur if an ad-hoc node itself or any of the intermediate relay ad-hoc nodes moves and breaks the active path. In general, if the path between an ad-hoc node and the IGW breaks and there is no other path to the same IGW, the ad-hoc node has to perform IGW discovery to establish a new path to another IGW (Mona *et al.*, 2004).

The IGW discovery scheme and the ad hoc routing protocol both have huge influence on the multi-hop handoff performance. Multi-hop handoff schemes can be classified into *forced handoff* and *route optimization-based handoff*. The *former* occurs whenever the path between the source/destination mobile node and the IGW is disrupted during data transmission due to, e.g., the movement of the MANET node. Therefore, a new path to the Internet has to be set up. The following IGW discovery process may result in the detection of a new IGW, which will consequently result in a handoff. The *latter* is a handoff that results from route optimization. If the source/destination MANET node detects that a shorter path to the Internet becomes available while communicating with a corresponding node, the active path will be optimized. In case the shorter path goes via a different IGW, a *route optimization-based handoff* occurs.

2.2. *Related work*

Figure 2 summarizes the comparison of different approaches for providing Internet connectivity and mobility management for MANETs, based on the description of the required functions.

The following acronyms are used in Figure 2 for the comparison of different mechanisms to provide the Internet connectivity and the mobility management for MANETs:

- TBBR: tree based bidirectional routing
- OLSR: optimized link-state routing
- AODV: ad-hoc on-demand distance-vector routing
- DSDV: destination-sequenced distance-vector routing
- DSR: dynamic source routing
- NAT: network address translation
- DHCP: dynamic host control protocol
- TD/RD: table-driven/root-driven routing
- MEWLANA: Mobile IP enriched wireless local area network architecture

Globalv4 (Belding-Royer *et al.*, 2001) describes mobile IPv4 (MIPv4) extensions for AODV (Perkins *et al.*, 2003). Destinations are first searched for in the MANET. If none is found, a host route is setup to the IGW. This solution suffers from long route discovery delays and lack of the same route aggregation that half-tunneling provides. Similarly, Globalv6 (Wakikawa *et al.*, 2001) can also work with MIPv6 (Johnson *et al.*, 2004), but it is not mandatory. A node may acquire a network prefix from an IGW, and construct a globally routable IP address through IPv6 stateless address auto-configuration. Globalv6 employs a similar technique as Globalv4 to determine the locality of destinations. Routing towards the IGW is done on a hop-by-hop basis using a default route of AODVv6 (Perkins *et al.*, 2000). Cascading effects (Le-Trung and Kotsis, 2008), i.e., all immediate MANET nodes on the chain from the source MANET node to the IGW needs to flood RREQ to determine whether the destination is located in the same MANET, are avoided by requiring intermediate nodes to configure host route entries for Internet destinations, with the downside of losing route aggregation. A summary of Globalv4 and Globalv6 is also presented in (Perkins *et al.*, 2002).

Index	Mechanism	Location determination	IGW discovery	IGW selection metrics	IGW forwarding	Addressing	Handoff style
1	MIPv6+ AODVv6	Network prefix	Proactive & reactive	Not specified, implicitly shortest hop-count	Default route & AODVv6	Deriving from IPv6 stateless auto-configuration	Both
2	MIPMANET	Flooding RREQ	Proactive	Shortest hop-count	Half-tunneling & AODV	Not specified, but Home Address must be IP global unicast	Route optimization-based
3	MIP+DSR	Using IGW	Reactive	Not specified, implicitly shortest hop-count	Source routing & DSR	Home Address must be IP global unicast	Not specified, implicitly route optimization-based
4	MIP+OLSR	Using routing table	Proactive	Not specified, implicitly shortest hop-count	Default route & OLSR	Not specified	Forced (when a prefix change)
5	MEWLANA TD/RD	Using routing table (DSDV) or TBBR Tree	Proactive	Shortest hop-count	Default route & DSDV or TBBR	Not specified	Forced (when a route change or node leave)
6	Two-tier MANET	Using routing table	Reactive	Load-balancing	Tunneling uses extra UDP/IP header & DSDV	Private address & NAT, allocating using DHCP	Not specified, implicitly route optimization-based
7	Hybrid MANET	Using routing table	Reactive	Hybrid: Euclidean distance & load-balancing	Default route & DSDV	Not specified	Forced (using automatic mode-detection and switching)
8	WLAN & MANET	Using routing table	Proactive	Not specified, implicitly shortest hop-count	Default route & OLSR	IPv6 stateless auto-configuration	Forced (using automatic mode-detection and switching)

Figure 2. Different mechanisms for MANET-Internet interworking and mobility management

MIPMANET (Jönsson *et al.*, 2000) studies the integration of mobile IP in the MANET. Tunneling from ad hoc nodes to the foreign agent (FA) is proposed as a way to achieve default route like behavior. This is the half-tunneling approach, where the outbound traffic to the Internet from the MANET uses tunneling and the inbound traffic from the Internet to the MANET is delivered to the corresponding destination

MANET node via the host route, using ad-hoc on-demand distance-vector routing (AODV) (Perkins *et al.*, 2003). However, this work does not explore the benefits of using tunneling, but studies different approaches to disseminate MIP information in the MANET instead.

A technique is described in (Broch *et al.*, 1999) to integrate MANETs with the Internet and to use MIP to support the migration of nodes between MANET and the Internet. Local delivery within a MANET subnet is accomplished using the dynamic source routing (DSR) protocol (Johnson *et al.*, 2007), while standard IP routing mechanisms decide which packets should enter and leave the subnet. For sending packets from the MANET subnet to the Internet, a source MANET node uses the source routing header of DSR to forward the packet to the IGW, where the source routing header will be removed. The packet is then forwarded to the Internet. However, this technique requires that each MANET node selects a single IP address (its home address) from the ones assigned to it, and that it uses only that address when participating in the DSR protocol.

The management of universal mobility, including both large-scale macro-mobility and local scale micro-mobility, is the focus of (Benzaid *et al.*, 2004). A hierarchical architecture is proposed, including: (1) extending micro-mobility management of a wireless access network to a MANET, (2) connecting this MANET to the Internet, and (3) integrating MIP and OLSR (Clausen *et al.*, 2006) to manage the universal mobility. In addition, it uses the optimal default route via the IGW to reach a host outside the MANET (i.e., the Internet host). The traffic to and from the Internet is distributed between base stations (APs) of the local network.

In addition to on-demand and table driven routing protocols in MANETs, a novel ad hoc routing type called root driven routing is introduced (Ergen and Puri, 2002). Using this protocol type for networks where the intensity of inside traffic is negligible makes protocol efficient by eliminating the routing overhead. Main idea of this routing type is formation of a tree whose root is the foreign agent (FA) and branches are mobile nodes, and periodical initiation of this tree formation procedure by the root. To take into account these different cases, two protocols called MEWLANA-TD, which uses table driven routing type, and MEWLANA-RD, which uses root driven routing type, are designed. MEWLANA-TD uses the DSDV routing protocol (Perkins *et al.*, 1994), in which there is a trigger updating either periodically or when there is a change in the routing table. DSDV enables each node have an entry in their routing table for all other nodes. MEWLANA-RD uses tree based bidirectional routing (TBBR) as the routing protocol. TBBR is a special routing protocol designed only by using MIP entities, introducing low overhead at the expense of performance degradation.

Two-tier MANET (Hsu *et al.*, 2004) shows a seamless roaming and load-balancing routing capability for providing Internet connectivity to the MANET. It modifies MIP to make traversing private networks, i.e. using network address translation (NAT). It also proposes a load-balancing routing protocol to improve the Internet access quality by allowing mobile node dynamically changing their IGWs, thus relieving the bottleneck problem.

Hybrid MANET (Ammari and Rewini, 2004) proposes an architecture to provide MANET nodes with Internet access using fixed IGWs and exploiting the mobility capability of additional mobile nodes (mobile IGWs). Since the Internet access for MANET nodes is provided via mobile IGWs, the quality of such service depends on the selection procedure used by MANET nodes to choose the most convenient mobile IGWs and register with. This work suggests using a hybrid criterion based on the weighted sum of the Euclidean distance between MANET nodes and mobile IGWs, and the load of mobile IGWs measured as the number of MANET nodes currently registered with them. MIP and DSDV (Perkins *et al.*, 1994) are extended to integrate the suggested hybrid criterion.

WLAN & MANET (Lamont *et al.*, 2003) presents a novel approach to integrate WLAN and MANET to the Internet (IPv6). The OLSR routing protocol is used within the MANET, and IGWs are used to connect the MANET to the Internet. In addition, the automatic mode-detection and the switching capability is also introduced in each MANET node to facilitate handoffs between WLAN and MANET.

Mobility management across WLAN and MANET is achieved through MIPv6 (Johnson *et al.*, 2004), which is integrated into the extended functionality of OLSR.

2.3. Discussions

A proactive approach to provide Internet connectivity to a MANET relies on ensuring that all nodes are registered with a foreign agent at all times. Mobile IP uses on link layer broadcasts to provide foreign agent information to interested nodes. However, these broadcasts can prove to be extremely expensive in a MANET where a broadcast means that packets are being flooded throughout the network.

In contrast, in a purely reactive approach, mobile nodes obtain foreign agent information by sending out agent solicitations only when data needs to be sent to a node outside the MANET. To limit the amount of flooding, these solicitations might be piggybacked on RREQ packets. An expanding ring search might also be used. In addition, intermediate nodes are allowed to reply with a route to the foreign agent, which reduces the overhead further.

The hybrid approach. provides Internet access to MANETs while attempting to balance the proactive and reactive approaches, and it has many benefits. A proactive solution allows mobile nodes to find the foreign agent closest to them and enables better handoffs, which in turn leads to lower delay. Periodic registrations in such a proactive scheme help foreign agents track the mobility of the mobile node. However, if not all the nodes in the MANET require connectivity, the repeated broadcasting of agent advertisements and solicitations can have a negative impact on the MANET due to excessive flooding overhead. A hybrid approach combines the advantages of both approaches so that the required information is received in a timely fashion and the MANET's scarce resources are not further burdened with the MIP overhead.

Two strategies, default routes and tunneling, are usually used to integrate gateway forwarding with ad-hoc routing protocols. It is found that default routes that are adopted from traditional LAN settings need modifications to work in a multi-hop ad hoc environment. Despite these additions, default routes have problems with multiple gateways and inconsistent routing state. Establishing tunnels to the gateways (either half-tunneling or tunneling), on the other hand, provides an architecturally appealing solution and works well with multiple Internet gateways (Le-Trung and Kotsis, 2008; Engelstad and Egeland, 2004; Engelstad et al., 2004).

3. A hybrid load-balancing metric

In this section, we propose a new metric for IGW selection to balance the inter/intra-MANET traffic load over multiple IGWs. It consists of three components. The first component is the shortest Euclidean distance (in terms of hop-count) between the MANET node and the selected IGW. Second comes the inter-MANET traffic load via each IGW, which is represented as the number of registered (both local and visiting) MANET nodes sending/receiving traffic to/from Internet via that IGW. The final component is the intra-MANET traffic load within the network topology managed by each IGW, which is related to the optimal node density to delivery traffic successfully.

The network model is described in Figure 3, where there are multiple IGWs [IGW_1, IGW_2,…, IGW_n] in a foreign MANET domain, and each IGW_j manages a network topology $(l_j.w_j)$, which can be overlapped with those managed by other IGWs. Each IGW_j attaches to its agent advertisement the following information $[l_j, w_j, n_{Reg}(j), n_j]$. Note that (l_j,w_j) is the managed topology size of IGW_j, $n_{Reg}(j)$ is the number of registered (both local and visiting) MANET nodes with IGW_j for the inbound/outbound traffic from/to the Internet, and n_j is the total (both local and visiting) MANET nodes in the managed topology of IGW_j. This agent advertisement is then broadcasted (for the proactive IGW discovery) periodically, or sent directly (for the reactive IGW discovery) to the source MANET node upon receiving its agent solicitation.

For the MANET *proactive routing* protocols, each IGW_j can determine [$n_{Reg}(j)$, n_j] by looking into its routing table, where $n_{Reg}(j)$ is the total valid routing entries, of which destinations are marked as Internet hosts and next-hop nodes are either "default routes" or "IGW_j", depending on what IGW forwarding strategy is used, see the required functions in Section 2. The value of n_j is equal to the total valid routing entries in the proactive routing table. For the MANET *reactive routing* protocols, the same rule is applied. However, it takes longer convergence time for each IGW_j to determine $[n_{Reg}(j), n_j]$ since it can only learn these values through the operation of routing protocol, e.g., by the periodic hello packet exchange of the neighbor discovery process, or by the on-demand RREQ/RREP packet exchange of the route discovery process.

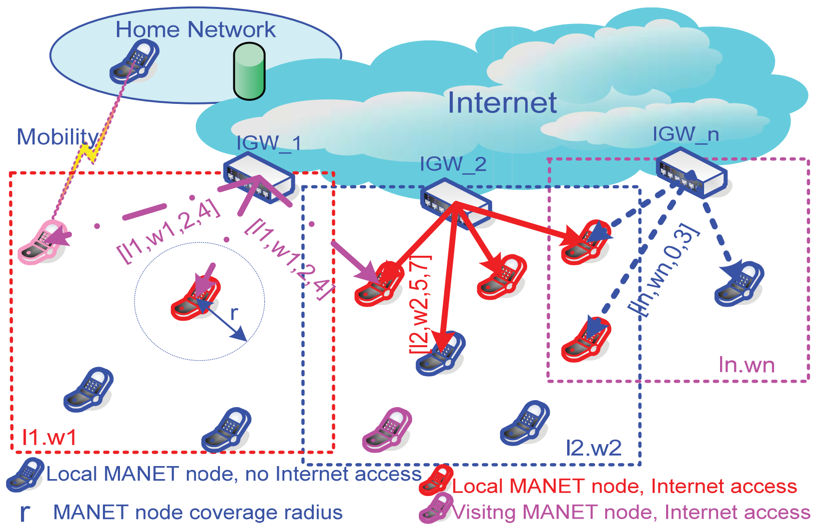

Figure 3. A hybrid load-balancing metric

For *proactive IGW discovery*, each IGW_j will attach $[l_j, w_j, n_{Reg}(j), n_j]$ to the agent advertisements and broadcast periodically. For *reactive IGW discovery*, these information will be attached by each IGW into either the agent advertisements or the proxy RREPs.

Whenever a visited or a local MANET node, which requires the Internet connectivity, receives these agent advertisement or proxy RREP packets from multiple IGWs in the same MANET domain, e.g., these IGWs use the same autonomous system (AS) number or network prefix, it uses the following formulas to choose the best IGW, i.e., the one with the lowest weight, to register:

$$Min\,\{w(i, j)\}_{j \in V_{IGW}} \tag{1}$$

$$w(i, j) = \alpha_1.D(i, j) + \alpha_2.LB_{Internet}(j) + \alpha_3.LB_{MANET}(i, j) \tag{2}$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1 \tag{3}$$

$$LB_{Internet}(j) = n_{\mathrm{Reg}}(j) \tag{4}$$

$$LB_{MANET}(i, j) = \begin{cases} \dfrac{1}{AvgDeg(i, j) \bmod \mathrm{K}} & if\ (AvgDeg(i, j) < (\mathrm{K} - \psi)) \\ \dfrac{1}{AvgDeg(i, j) \bmod \mathrm{K}} + \delta_1 & \text{if } ((\mathrm{K} - \psi) \le AvgDeg(i, j) \le (\mathrm{K} + \psi)) \\ \dfrac{1}{AvgDeg(i, j) \bmod \mathrm{K}} + \delta_2 & if\ (\mathrm{K} + \psi) < AvgDeg(i, j) < 2.K \\ +\infty & otherwise \end{cases} \tag{5}$$

$$AvgDeg(i, j) = \begin{cases} \left(l_j.w_j - \dfrac{r.(l_j + w_j)}{2} + r^2\right).\dfrac{n_j.\pi.r^2}{(l_j.w_j)^2} \text{(a)} \\ \left(l_j.w_j - \dfrac{r.(l_j + w_j)}{2} + r^2\right).\dfrac{(n_j + 1).\pi.r^2}{(l_j.w_j)^2} \text{(b)} \end{cases} \tag{6}$$

Each MANET node i, upon requesting Internet connectivity, register to one of the IGWs discovered. The objective is to select an IGW_j with the lowest weight $w(i,j)$ as described in equations (1) and (2), where α_i, $i \in [1,3]$, is the constant to represent the contribution of each component into the metric. Thus, the sum of these constants in equation (3) is one. First component $D(i,j)$ is the shortest distance in terms of hop-count from the MANET i to the IGW_j. It is determined the MANET node i using either the received IGW discovery packets (agent advertisement/solicitation) or by the corresponding MANET routing protocol (routing table, RREQ packet, or RREP packet). The second component $LB_{Internet}(j)$ is the *inter-MANET traffic load* via IGW_j in the number of current registered MANET nodes $n_{Reg}(j)$[6] at IGW_j that require Internet connectivity, see equation (4). This information is extracted by the MANET node i from either agent advertisement packets (broadcasting periodically in proactive IGW discovery, or upon receiving an agent solicitation from MANET node i in reactive IGW discovery) or proxy RREPs sent by IGW_j (only in reactive IGW discovery). Finally, the third component $LB_{MANET}(i,j)$ is the *intra-MANET traffic load* in the network topology (l_j,w_j) managed by IGW_j. It is determined based on the optimal node density K, and the average node degree $AvgDeg(i,j)$. Work in (Royer *et al.*, 2001) shows that $K=7$ is an appropriate setting for a MANET node speed of *0-1m/s*, $K=15\text{-}20$ is good for a node speed of *5m/s*, while $K=20\text{-}25$ is suitable for a node speed of *10m/s*.

The average node degree *(AvgDeg)* is presented in (Chiu and Gen-Huey, 2003). However, in equation (6), the average node degree is different for a local MANET node (equation 6a) and a visiting MANET node (equation 6b). This is because IGW_j does not know the existence of a visiting MANET node i in its managed network topology until a registration occurs. Figure 4 shows the average node degree specified in equation (6a) of different network topology size $(r=250m)$.

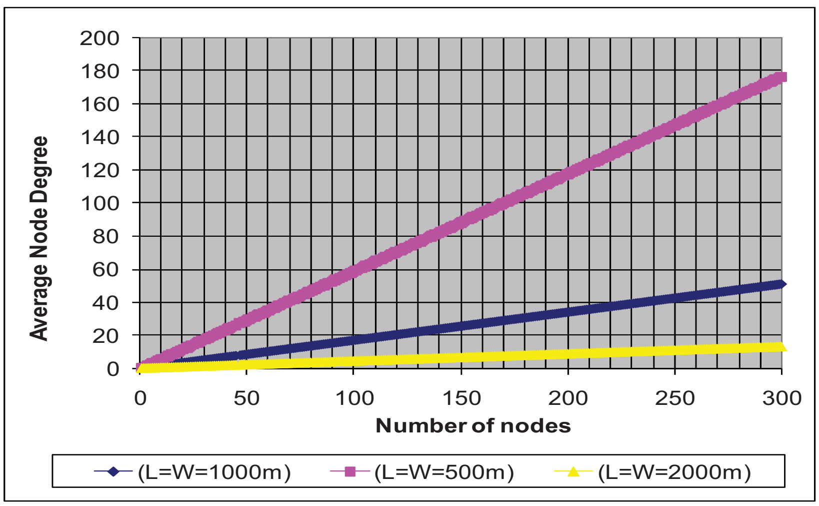

Figure 4. Average node degree

Equation (5) is used to determine $LB_{MANET}(i,j)$. This is because the packet delivery ratio of the intra-MANET traffic increases when the $AvgDeg(i,j)$ is increased, taking maximum value when $AvgDeg(i,j)$ is equal to K (optimal node density), then decreasing even if the $AvgDeg(i,j)$ continues to increase due to the congestion (Royer *et al.*, 2001). The key idea to determine $LB_{MANET}(i,j)$ is to prevent further MANET nodes to join into the already optimal or over-load intra-MANET clusters, while encouraging MANET nodes to join into under-load clusters. This is achieved by adding higher values, i.e., $(+\delta_1,+\delta_2,+\infty)$, for either optimal or lightly over-load or heavily over-load intra-MANET clusters, respectively. Figure 5

[6] A more precise method to determine inter-MANET traffic is to use each IGW to detect the active inter-MANET connections, together with traffic load (e.g., number of transmitting, receiving, forwarding data/control packets). For the simplicity of implementation, we use $n_{Reg}(j)$, and only one TCP connection are setup in the simulation between any pair of MANET/Internet hosts, at the same settings and packet size. Thus, $n_{Reg}(j)$ also represents the inter-MANET traffic at IGW_j.

shows the values of $LB_{MANET}(i,j)$ corresponding to the network topoly size $(L=W=1000m)$, $K\in\{7,20\}$, $\Psi=1$, $\delta_1=5$, $\delta_2=10$, $(+\infty)\sim(+15)$, and the range of MANET nodes *[1..300]*, which is corresponding to the range of $AvgDeg\in[0.18;50.88]$, see Figure 4. We estimate that these settings will represents realistic scenarios of MANET deployment.

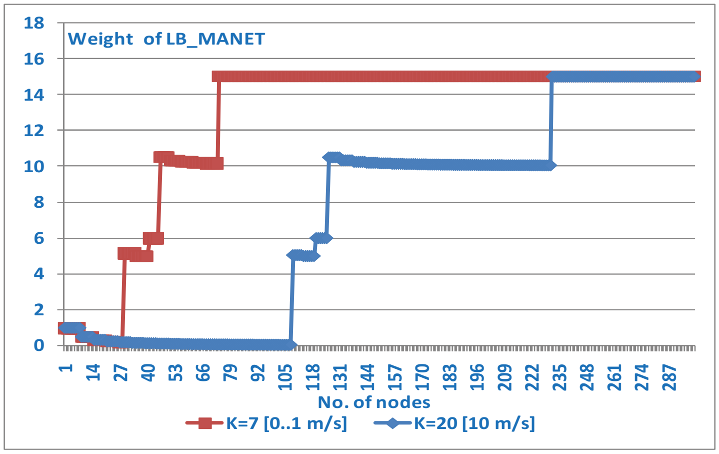

Figure 5. The load-balancing weight of intra-MANET traffic

Another point is that it is better to use the tunneling instead of the default route in forwarding inter-MANET traffic to avoid the *inconsistent context* problems (Le-Trung and Kotsis, 2008; Engelstad and Egeland, 2004; Engelstad et al., 2004), which is defined as the use of different IGWs for inbound/outbound traffic from/to Internet on each connection between a MANET node and an Internet host. These problems have adverse effects on two-way traffic, e.g., TCP, which can terminate the 2-way connection. Moreover, a tunneling solution [3-5,10,19-20] has the potential to exploit efficiently multiple IGWs for the benefit of multi-homing or for performing soft handovers. Figure 6 illustrates different inconsistent context problems, and the corresponding solutions to reduce these problems either partly using default routes, or completely using (half)-tunneling (Le-Trung and Kotsis, 2008).

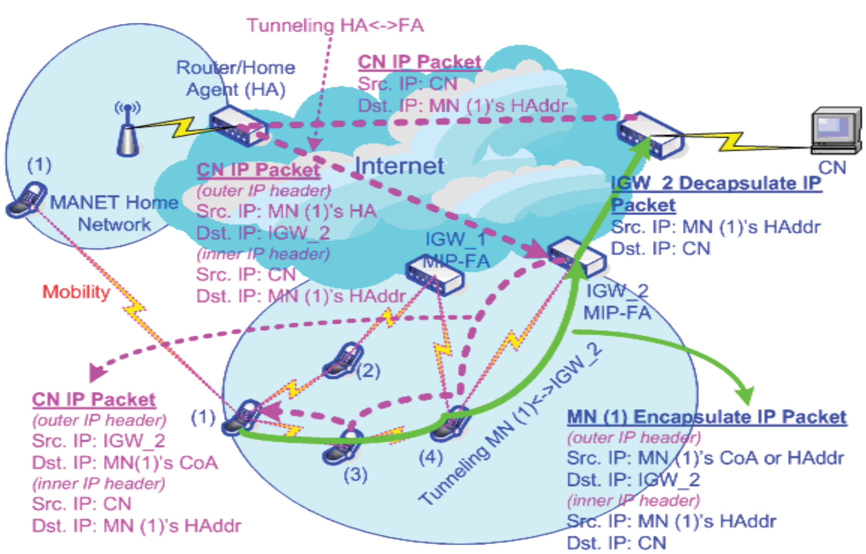

Figure 6. Inconsistent context problems and solutions

Note that our proposed metric will reduce into the shortest hop-count metric by setting the values of $[\alpha_1, \alpha_2, \alpha_3]$ as $[1.0, 0.0, 0.0]$.

4. Simulation implementation

To validate that the proposed metric for IGW selection achieves the load-balancing of intra/intra-MANET traffic, it has been integrated into AODV, and implemented into ns-2[7][8]. Following the specification of the required functions in providing Internet connectivity and mobility management for MANET nodes in Section 2, below is our approach in the implementation:

- The half-tunneling technique (Jönsson *et al.*, 2000) is used, i.e., outbound traffic from source MANET nodes to the Internet hosts uses tunneling to avoid the inconsistent context problems (Le-Trung and Kotsis, 2008), while inbound traffic delivered to destination MANET nodes uses AODV (without tunneling) to reduce the overhead of adding additional IP header for the tunneling.
- Whenever a MANET node moves into a new domain, it uses the address of the corresponding IGW in that domain (selecting the best one using our proposed metric if there exists multiple ones) to register with its home agent. Thus, MIPv4 foreign agent (FA) care-of address (CoA)[9] is used in this implementation.
- Intermediate MANET nodes are not allowed to send a proxy route reply (RREP). This reduces the probability that a route to the destination MANET node in the same domain via the IGW (a not-optimal route) is returned instead of a host route (the optimal one).
- Intermediate MANET nodes are not allowed to forward a proxy RREP without updating from it. This reduces the inconsistent context problems (Le-Trung and Kotsis, 2008).
- Neighbor discovery uses the link-layer feedback (layer 2) instead of the hello packet exchanges (layer 3). In case there are link changes, either a link broken or a new link becoming available, the corresponding active entries in the routing table are updated. Again, if a better route to another IGW is found, or if a new route to an IGW is found while the old route to the registered one is broken, the corresponding MANET node will update all its current routes to destination Internet hosts via the new IGW. Of course, this update is carried out only after this MANET node has registered this new IGW to its home agent as its new MIPv4 FA CoA.
- The information $[l_j, w_j, n_{Reg}(j), n_j]$ is attached into the proxy RREP sent by the corresponding IGW_j.
- Only reactive IGW discovery is implemented, i.e., proxy RREPs, are sent back to the source MANET node by any IGWs, which are reachable to the destination Internet host. Note that if the source MANET node receives multiple proxy RREPs from different IGWs, it uses our new proposed metric to select the best one for connecting to the destination Internet host.
- If a connection from a MANET node to its registered Internet gateway is invalid, either due to a link being broken or due to the lifetime expiry in the routing table, another available route will be chosen as the alternative. All connections from this MANET node to any Internet host via the failed IGW will be updated via the new available IGW.
- Multiple IGWs detected via proxy RREPs will be kept in the source MANET node generating the route request (RREQ). However, this MANET node only uses one IGW (the best selected by our proposed metric and after registering this IGW address with its home agent as its MIP FA CoA) for forwarding traffic to destination Internet hosts. Other IGWs are used as the backup.
- In this implementation, a MANET can update to the better IGW if and only if it has registered this new IGW address (new MIP FA CoA) to its home agent to replace for the old one.

5. Simulation settings and results

The following parameters are used to compare the performance and overhead in providing Internet connectivity for MANET nodes, which uses AODV applying either our proposed metric or the shortest hop-count metric.

---

[7] http://www.isi.edu/nsnam/ns/.

[8] http://core.it.uu.se/core/index.php/AODV-UU_and_Mobile_IP_for_ns-2

[9] http://www.ietf.org/html.charters/mip4-charter.html

Packet delivery ratio. The ratio between the total data packets sent by the sources and the total data packet received correctly by the corresponding destinations.

Normalized signaling overhead. The ratio between the total number of control packets carrying signaling information (including the ad hoc routing, the IGW discovery, and the MIP registration) and the total number of data packets. Each sending or forwarding of packet (data or control) to the next-hop neighbor is counted as one.

Average packet transmission delay. The average time of sending data packets from particular ad hoc sources to their associated IGWs, which can be changed due to the mobility of the ad hoc sources or the broken links. Its unit is second [sec].

Two simulation scenarios have been designed to evaluate our new proposed metric for Internet gateway selection strategies, which are shown in the Figure 7 and Figure 11, respectively. The first scenario is used to test the correctness of our implementation for two-way connections under mobility, avoiding the inconsistent context problems. Moreover, it is also used to compare the performance and overhead of our new proposed metric with the origin shortest hop-count metric for Internet gateway selection. The second scenario is used to demonstrate the outcome of our proposed load-balancing metric over the shortest hop-count metric in multi-homed MANET domain.

In the first scenario, see Figure 7, two Internet gateways (IGW_0, IGW_1) and two Internet hosts $(host_0, host_1)$ are connected to each other using wired links (bandwidth: *5Mbps*, propagation delay: *2ms*) creating a ring. Four MANET nodes (M_3, M_4, M_5, M_6) are located at positions shown in Fig. 7, with the network topology *(l=600m, w=600m)*. Each Internet gateway has two interfaces, one for connecting to the wired network, while another for connecting to the MANET. For IGW_0, its information $[l_j, w_j, n_{Reg}(j), n_j]$ is *[600, 600, 1, 4]*. For IGW_1, its information $[l_j, w_j, n_{Reg}(j), n_j]$ is *[600, 600, 1, 4]*. One TCP connection is setup between M_6 [TCP source] and Internet $host_0$ [TCP destination] for the FTP application. MANET node mobility is setup as follows:

- At *0.5s*, M_5 starts moving to position *[450, 400]* at speed *10m/s*, while M_6 starts moving to position *[450, 550]* at speed *10m/s*.
- At *45.0s*, M_5 continues moving to position *[200, 250]* at speed *10m/s*.
- At *50.0s*, M_5 continues moving to position *[250, 250]* at speed *10m/s*.
- Finally, at *90.0s*, M_6 moves to position *[250, 500]* at speed *10m/s*.

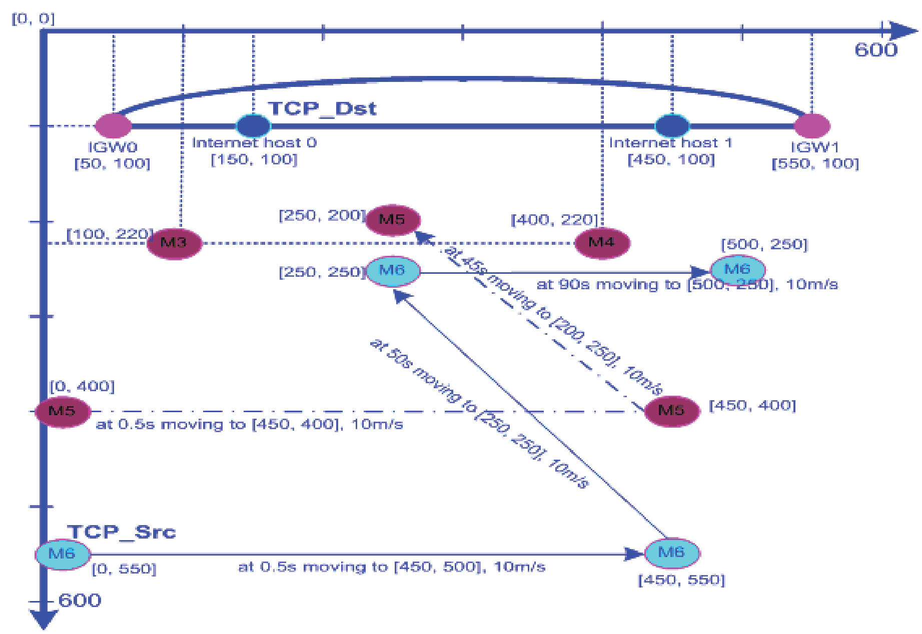

Figure 7. The first simulation scenario

The total simulation time is *200s*. Starting and stopping time for TCP connection is *10.0s* and *200.0s*, respectively. $[\alpha_1, \alpha_2, \alpha_3]$ is setup as either *[1.0, 0.0, 0.0]* (for the shortest hop-count metric) or *[0.8, 0.1, 0.1]* (for the load-balancing metric). Additionally, $K=20$, $\delta_1=5$, $\delta_2=10$, $\Psi=1$, $+\infty$ is setting to *15*. For MANET communications, *MAC 802.11* distributed coordination function (DCF) (bandwidth: *2Mbps*) is used, with MANET node coverage radius *(r=250m)*. Radio propagation uses the two-ray ground model. For TCP connection, the packet size is *512 bytes*. For packet encapsulation using in tunneling, an additional packet header of *62 bytes* is added, so total packet size is *580 bytes*.

In the first scenario, there is only one Internet gateway IGW_0 in the home network of MANET source node M_6, and only another Internet gateway IGW_1 in the foreign network, where M_6 moves into later. Thus, the effect of our proposed metric for Internet gateway selection on the performance, overhead, and delay is the same compared with the origin shortest hop-count metric. Figure 8 shows that the packet delivery ratio of TCP connection $[M_6 \rightarrow host_0]$ using hop-count metric and load-balancing metric (our proposed metric) is the same, i.e., over 90%. The high performance results also show that our implementation reduces the inconsistent context problems (Le-Trung and Kotsis, 2008).

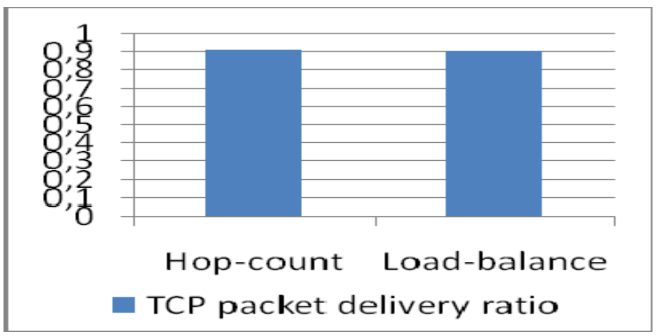

Figure 8. TCP packet delivery ratio in first scenario

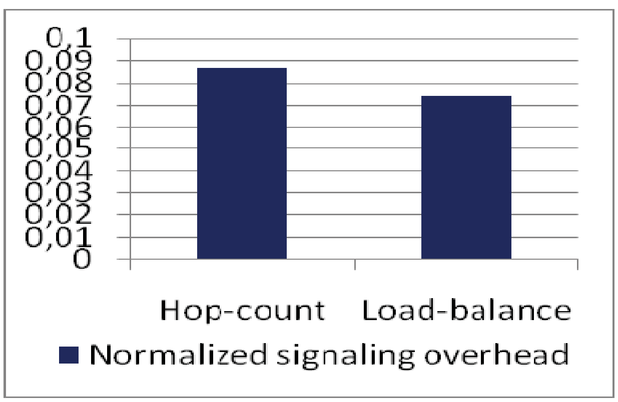

Figure 9. Normalized signaling overhead in first scenario

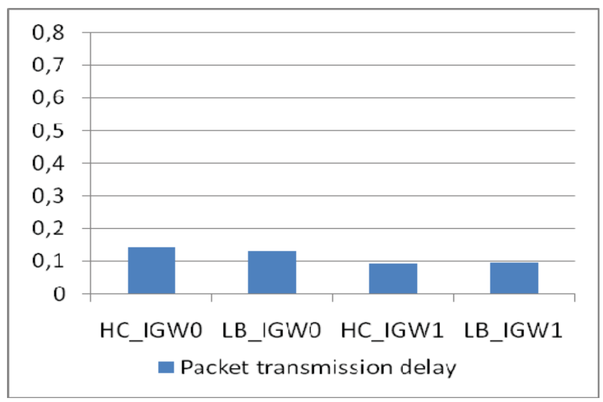

Figure 10. Average packet transmission delay

Additionally, Figure 9 shows that the signalling overhead of using load-balancing metric is slightly lower than that of using hop-count, and the average delay to IGW_0 of using load-balancing metric is slightly lower than that of using hop-count metric, while the average delay to IGW_1 of using load-balancing metric is slightly larger than that of using hop-count metric (see Figure 10). Tracing the routing table of MANET node M_6 in both cases, there are more Internet gateway changes in using load-balancing metric compared with using hop-count around the simulation duration $[103s \rightarrow 145s]$. Within this duration, MANET node M_6 is moving from position *[250, 250]* to position *[500, 250]*. In the implementation, a change in the Internet gateway requires the re-registration of this new Internet gateway to MANET node M_6 home agent as its new MIPv4 FA CoA, which creates more signalling overhead. However, these registration packets also refresh lifetime of the corresponding routing entries in routing table of other MANET nodes on the way to the re-registered Internet gateway, reducing the overhead in the MANET routing, i.e., AODV. Therefore, signaling overhead results in Figure 9 of using hop-count and load-balancing metrics show that the benefit of reducing overhead in AODV routing overcomes the cost of increasing signalling overhead in MIP re-registration, in the first simulation scenario. This factor is further proved by the tracing of the number of dropping packets, which are the same in both cases, either dropped by the congestion in MAC layer or dropped by the fact that no route is found due to the MANET node $[M_5, M_6]$ mobility.

Since there are more Internet gateway changes, i.e., more re-registrations [three times] back to IGW0, in case using load-balancing metric compared with using hop-count metric in the duration $[103s \rightarrow 145s]$, the distances $M_6 \rightarrow IGW_0$ and $M_6 \rightarrow IGW_1$ are nearer during this duration compared with within duration $[0s \rightarrow 50s]$, see Figure 10. Thus, more data packet transmissions via IGW_0 will reduce the average delay $M_6 \rightarrow IGW_0$, while lower data packet transmissions via IGW_1 will increase the average delay $M_6 \rightarrow IGW_1$. This factor explains the simulation results of using hop-count metric and load-balancing metric, as showed in Figure 10.

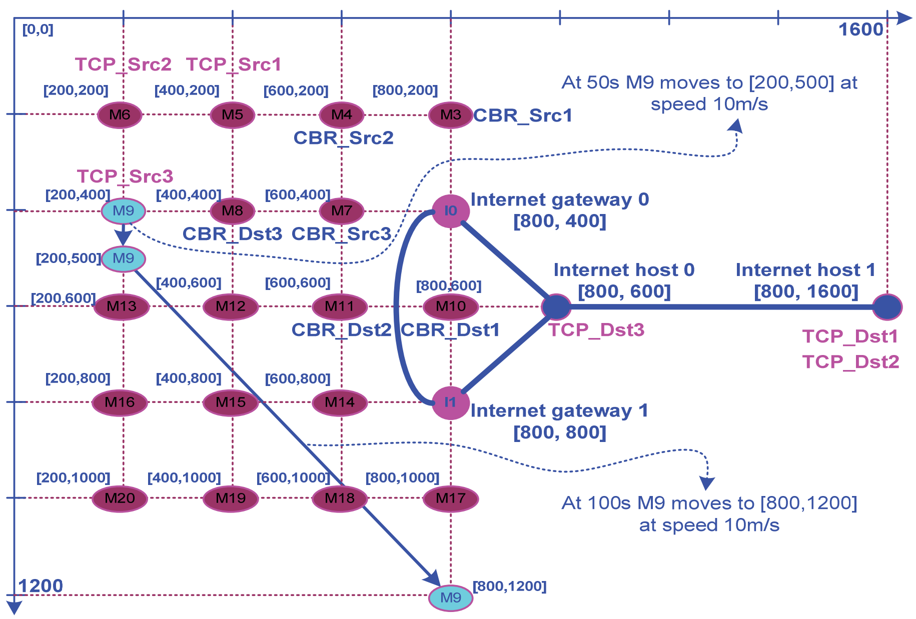

Figure 11. The second simulation scenario

The second simulation scenario is shown Figure 11. It consists of two Internet gateways (IGW_0, IGW_1) and two Internet hosts $(host_0, host_1)$, which are connected to each other using wired links (bandwidth: *5Mbps*, propagation delay: *2ms*), creating a connected wired network. Eighteen MANET nodes $(M_3 \rightarrow M_{20})$ are located at positions shown in Figure 11, with the network topology $(l=1200m, w=1600m)$. Each IGW has two interfaces, one for connecting to the wired network, while another for connecting to the MANET. The network topology is initially partitioned into two sub-areas managed by IGW_0 $[l_0=600m, w_0=1000m, n_{Reg}(0)=3, n_0=11]$ and IGW_1 $[l_1=400m, w_1=1000m, n_{Reg}(1)=0, n_1=7]$. Eleven MANET nodes, $M_3 \rightarrow M_{13}$, are initially in the sub-area managed by IGW_0, while seven MANET nodes, $M_{14} \rightarrow M_{20}$, are initially in the sub-area managed by IGW_1 . Note that the values of $[n_{Reg}(0), n_0, n_{Reg}(1), n_1]$ will be later changed to $[n_{Reg}(0)=2, n_0=10, n_{Reg}(1)=1, n_1=8]$ through the simulation depending on the mobility of MANET node $[M_9]$. This information will be attached into proxy RREPs sent by each IGW upon receiving a RREQ for the destination Internet hosts. For the FTP applications, three TCP connections (for inter-MANET traffic) are set up, including TCP_1 [$M_5 \rightarrow host_1$, starting at *6.0s*, stopping at *150.0s*], TCP_2 [$M_6 \rightarrow host_1$, starting at *11.0s*, stopping *150.0s*] and TCP_3 [$M_9 \rightarrow host_0$, starting at *16.0s*, stopping at *150.0s*]. Three CBR connections (intra-MANET traffic) are also within the MANET domain, including CBR_1 [$M_3 \rightarrow M_{10}$, starting at *5.0s*, stopping at *150.0s*], CBR_2 [$M_4 \rightarrow M_{11}$, starting at *10.0s*, stopping at *150.0s*], and CBR_3 [$M_7 \rightarrow M_8$, starting at *15.0s*, stopping at *150.0s*]. The MANET node mobility is set up as follows:

At *50.0s*, M_9 starts moving to position *[200, 500]* at speed *10m/s*.
At *100.0s*, M_9 continues moving to position *[800, 1200]* at speed *10m/s*.

For MANET communications, *MAC 802.11* distributed coordination function (DCF) (bandwidth: *2Mbps*) is used, with MANET node coverage radius *(r=250m)*. Radio propagation uses the two-ray ground model. For both CBR/TCP connections, the packet size is *512 bytes*. For packet encapsulation using in tunneling, an additional packet header of *62 bytes* is added, so total packet size is *580 bytes*. The packet rate for CBR connection is 4 packets/s. The total simulation time is *150s*. Other parameter settings consist of:

- An appropriate node degree $K=20$ was set, corresponding to the the node speed of *10m/s*.
- $[\alpha_1,, \alpha_2, \alpha_3]$ are set as *[1.0, 0.0, 0.0]*, i.e., the shortest hop-count (HC) metric, and *[0.2, 0.5, 0.3]*, i.e., the load-balancing (LB) metric, respectively.
- $\delta_1=5$, $\delta_2=10$, $\Psi=1$, $+\infty$ is setting to *15*.

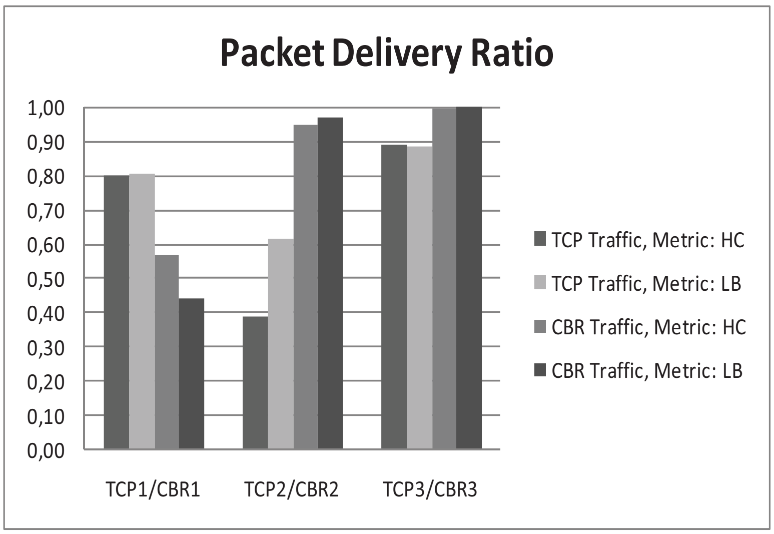

Figure 12. Packet delivery ratio

Figure 12 shows the comparison of packet delivery ratio of three TCP and three CBR connections, using AODV routing protocol with our proposed metric vs. the shortest hop-count metric. The source MANET nodes $[M_5, M_6]$ of TCP_1 and TCP_2 are fixed, while that of TCP_3 is in movement. Since the source node of TCP_1 $[M_5]$ is nearer to IGW_0 than that of TCP_2 $[M_6]$ and both are fixed, the packet delivery ratio of TCP_1 will be higher than that of TCP_2, see Figure 10. However, the packet delivery ratio of TCP_3 is the highest. This is due to the mobility of the source node of TCP_3 $[M_9]$ to the sub-area managed by IGW_1, which is both lower traffic [no intra/inter-MANET traffic] and shorter hop-count to IGW_1.

For intra-MANET traffic within sub-area managed by IGW_0, the packet delivery ratio of CBR_3 is the highest *(nearly 100%)* since the source MANET node $[M_7]$ and the destination MANET node $[M_8]$ are direct neighbor each other. The packet delivery ratio of CBR_2 is slightly lower compared with that of CBR_3 since the distance between the source $[M_4]$ and the destination $[M_{11}]$ is longer. The average length of the route from source $[M_3]$ and destination $[M_{10}]$ of CBR_1 is the same as that of CBR_2, see Figure 11, but the packet delivery ratio is much lower due to the congestion created by inter-MANET traffic of TCP_1, TCP_2 and partly by TCP_3 around *1-hop* vicinity of IGW_0.

Due to the mobility of $[M_9]$, there are larger differences for the packet delivery ratio of AODV using our load-balancing metric vs. that of AODV using the shortest hop-count metric, in TCP_2 and CBR_1 connections, see Figure 12. This is because more TCP_2 traffic, in case AODV is used with our load-balancing metric, is forwarded to the destination Internet $host_1$ via IGW_1, which has lower both inter-MANET traffic (TCP) and intra-MANET traffic (CBR) compared with IGW_0. The load-balancing of TCP_2 traffic also increases the rate of ACK (acknowledgement) packet feedback to the source of TCP_2 *[M6]*, which allows more TCP traffic (*30%* higher compared with the shortest hop-count), reducing packet dropping.

However, the load-balancing of TCP_2 traffic via IGW_1 also creates the side effect, i.e., more traffic on the *1-hop* vicinity of CBR_1 connection $[M_3 \rightarrow M_{10}]$, increasing the congestion on the MAC *802.11 DCF* and causing the CBR_1 packet dropping. This is why the packet delivery ratio of CBR_1, in case AODV is used with our load-balancing metric, is lower compared with the shortest hop-count, see Figure 12.

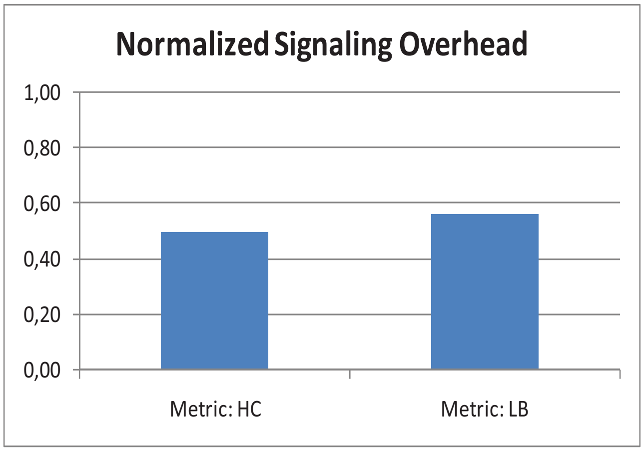

Figure 13. Normalized signaling overhead

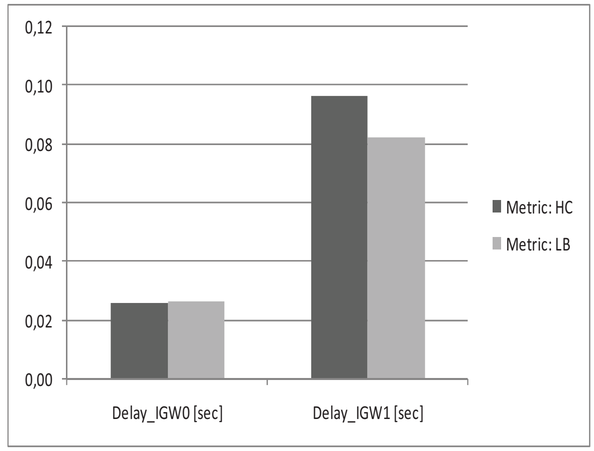

Figure 14. Average packet transmission delay

Since more data packets are delivered for TCP_2, more signaling overhead (including routing and MIP re-registration control packets) are also introduced compared with the signaling overhead under the shortest hop-count. Figure 13 shows that the signaling overhead, in case AODV using our load-balancing metric, is slightly higher (about *10%*).

Figure 14 shows that the average packet transmission delay of TCP data packets from $[M_5, M_6, M_9] \rightarrow IGW_0$ is almost the same under two cases, but lower for that from $[M_5, M_6, M_9] \rightarrow IGW_1$, in case AODV using load-balancing metric. This is due to below factors:

- More inter-MANET traffic (TCP_2), i.e., about *30%* higher, are successfully transmitted in case AODV using our load-balancing metric compared with the shortest hop-count.
- For AODV using the shortest hop-count metric, most $TCP_{1,2}$ traffic is forwarded in/out via IGW_0, while TCP_3 traffic is forwarded between IGW_0 and IGW_1, depending on $[M_9]$ mobility.
- For AODV using our load-balancing metric, most TCP_1 traffic is forwarded in/out via IGW_0. Part of TCP_2 traffic approximately equal to the amount TCP_2 traffic succesfully forwarded in case AODV using the shortest hop-count is forwarded in/out IGW_0, and the rest of TCP_2 traffic (about *30%* that of AODV using the shortest hop-count) is forwarded in/out IGW_1. Finally, TCP_3 traffic is forwarded between IGW_0 and IGW_1, depending on $[M_9]$ mobility.
- Due to the mobility of $[M_9]$, i.e., source of TCP_3, the average transmission delay of sending data packets from $[M_6 \rightarrow IGW_1]$ on TCP_2 is shorter than that of sending data packets from $[M_9 \rightarrow IGW_1]$ on TCP_3.
- Data packets on TCP connections are also used to refresh the lifetime of the corresponding routing entries in routing tables of intermediate nodes on the path.
- Thus, the frequent and higher transmissions of TCP_2 data packets reduce the average delay of sending TCP_2 traffic, shortening the average transmission delay of all TCP connections to IGW_1, in case AODV using our load-balancing metric instead of the shortest hop-count.

6. Conclusions and future work

This paper first shows a deep review on the required functions on providing Internet connectivity and mobility management for MANETs. It then proposes a hybrid metric for IGW selection to balance the intra/inter-MANET traffic load among multiple IGWs on the same MANET domain. Three components are considered: the Euclidean distance (in terms of hop-count), the load-balancing of inter-MANET traffic (TCP), and the load-balancing of intra-MANET traffic (CBR). Two simulation scenarios have been designed: while the first is mostly used to validate the correctness of implementation (avoiding inconsistent context problems), the second is used to compare the packet delivery ratio, signaling overhead, and average packet transmission delay, of AODV using our proposed metric compared with the shortest hop-count metric for multiple IGW selection. Simulation results show the effect of our proposed metric on performance parameters is better, i.e., load-balancing of inter-MANET traffic via multiple IGWs increases the packet delivery ratio, reducing the average delay at the cost of slightly increasing the signaling overhead, e.g., more re-registration packets for changing IGWs. There are also more points that need to be developed:

- Simulation results in this paper are taken from two mobility scenarios. Thus, more case studies need to be carried out to demonstrate the outcomes of our proposed metric compared with others.
- In this paper, the proposed metric is integrated into AODV (Perkins *et al.*, 2003), i.e., a reactive MANET routing protocol. Another point is to integrate the proposed metric into any proactive MANET routing protocol, e.g., optimized link-state routing (OLSR) (Clausen *et al.*, 2006), and compare with those in this paper.
- The setting thresholds of $[\alpha_1, \alpha_2, \alpha_3]$, δ_1, δ_2, Ψ, are important. They are determined based on the traffic patterns, mobility patterns, and the network topology. Up to this point, how to determine these thresholds are still open questions.
- The determination of n_j in this paper is based only on the operation of the corresponding routing protocol. Thus, we assume that a MANET node i will be in the network topology $[l_j, w_j]$ managed by IGW_j if it receives either agent advertisements or proxy RREPs sent by this IGW. Future works will consider the location of MANET nodes, together with the use of location-based ad-hoc routing, e.g., GPSR (Karp and Kung, 2000), for traffic forwarding.

References

Abduljalil, F. M. and Bodhe, S. K. (2007), "A survey of integrating IP mobility protocols and mobile ad hoc networks", *IEEE Commu. Surveys and Tutorials*, Vol.9, No.1, pp. 14-30.

Ammari, H. and Rewini, H. E. (2004), "Using hybrid selection schemes to support QoS when providing multihop wireless Internet access to mobile ad hoc networks", *QSHINE'04*, pp.148-155.

Belding-Royer, E. M., Sun, Y., Barbara, S. and Perkins, C. E. (2001), "Global connectivity for IPv4 mobile ad hoc networks", *Internet Draft draft-royer-manet-globalv4-00.txt*.

Benzaid, M., Minet, P., Agha, K. A., Adjih, C., and Allard, G. (2004), "Integration of Mobile-IP and OLSR for a universal mobility", *Wireless Networks*, 10, 2004, pp.377-388.

Bernardos, C. and Calderon, M. (2005), "Survey of IP address autoconfiguration mechanisms for MANET", *Internet draft-bernardos-manet-autoconf-survey-00.txt*.

Broch, J., Maltz, D. A. and Johnson, D. B. (1999), "Supporting hierarchy and heterogeneous interfaces in multi-hop wireless ad hoc networks", *ISPAN'99*, pp.370-375.

Chand, N., Joshi, R. C., and Misra, M. (2007), "Cooperative caching in mobile ad-hoc networks based on data utility", *Mobile Information Systems*, 3(1): 19-37.

Chiu, C. Y. and Gen-Huey, C. (2003), "A stability aware cluster routing protocol for mobile ad hoc networks", *Wireless Communications and Mobile Computing*; 3:503-515.

Clausen, T., Dearlove, C. and Jacket, P. (2006), "The optimized link-state routing protocol version 2", *IETF Internet draft*, http://www.ietf.org/internet-drafts/draft-ietf-manet-olsrv2-02.txt.

Engelstad, P. E., and Egeland, G. (2004), "NAT-based Internet connectivity for on-demand ad hoc networks", *WONS'04, LNCS 2928*, Italy, pp. 342-356.

Engelstad, P. E., Tonnesen, A., Hafslund, A. and Egeland, G. (2004), "Internet connectivity for multi-homed proactive ad hoc networks", *IEEE ICC'04*, pp.4050-4056.

Ergen, M., and Puri, A. (2002), "MEWLANA-mobile IP enriched wireless local area network architecture", *Proceedings of VTC2002-Fall*, pp.2449- 2453, Vol. 4.

Hsu, Y. Y., Tseng, Y. C., Tseng, C. C., Huang, C. F., Fan, J. H. and Wu, H. L. (2004), "Design and implementation of two-tier mobile ad hoc networks with seamless roaming and load-balancing routing capability", *IEEE QSHINE'04*, pp.52-58.

Jin, X. and Christian, B. (2002), "Wireless multihop Internet access: gateway discovery, routing, and addressing", *Proceedings of 3GWireless'02*, San Francisco, CA, USA.

Johnson, D., Perkins, C., and Arkko, J. (2004), "Mobility support in IPv6", *Internet rfc3775.txt*, http//www.ietf.org/rfc/rfc3775.txt.

Johnson, D., Hu, Y. and Maltz, D., (2007), "The dynamic source routing (DSR) protocol for mobile ad hoc networks for IPv4", *Internet rfc4728.txt*, http://www.ietf.org/rfc/rfc4728.txt.

Jönsson, U., Alriksson, F., Larsson, T., Johansson, P. and Maguire, G. Q. (2000), "MIPMANET – Mobile IP for mobile ad hoc networks", *MobiHoc'00*, pp.75-85, Boston, Massachusetts.

Karp, B. and Kung, H. T. (2000), "GPSR: Greedy perimeter stateless routing for wireless networks", *MobiCom'00*, Boston, MA.

Lamont, L., Wang, M., Villasenor, L., Randhawa, T., and Hardy, S. (2003), "Integrating WLANs & MANETs to the IPv6 based Internet", *ICC2003*, Anchorage, Alaska, USA, May 11-15, 2003.

Le, D., Fu, X. and Hogrefe, D. (2006), "A review of mobility support paradigms for the Internet", *IEEE Commu. Surveys and Tutorials*, Vol.8, Issue 1, pp. 38-51, 1st Quarter 2006.

Le-Trung, Q. and Kotsis, G. (2008), "Reducing problems in providing Internet connectivity for mobile ad hoc networks", *EuroFGI'08, LNCS Vol. 5122*, Barcelona, Spain, pp.113-127.

Levkowetz, H. and Vaarala, S. (2007), "Mobile IP NAT/NAPT traversal using UDP tunnelling", *Internet Draft draft-ietf-mobileip-nat-traversal-07.txt*, http://tools.ietf.org/id/draft-ietf-mobileip-nat-traversal-07.txt.

Mona, G., Philipp, H., Christian, P., Vasilis, F. and Hamid, A. (2004), "Performance analysis of Internet gateway discovery protocols in ad hoc networks", *WCNC'04*, pp.120-125, Atlanta, GA, USA.

Natsheh, E., and Wan, T. C.(2008), "Adaptive and fuzzy approaches for nodes affinity management in wireless ad-hoc networks", *Mobile Information Systems*, 4(4): 273-295.

Perkins, C., Belding-Royer, E. and Das, S. (2003), "Ad hoc on-demand distance vector (AODV) routing", *Internet rfc3561.txt*, http://www.ietf.org/rfc/rfc3561.txt.

Perkins, C., Belding-Royer E., and Das, S. (2000), "Ad hoc on-demand distance vector (AODV) routing for IP version 6", *Internet draft-perkins-manet-aodv6-01.txt*.

Perkins, C., and Bhagwat, P. (1994), "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers", *Proceedings of the SIGCOMM '94 Conference on Communications Architectures, Protocols and Applications*, August 1994, pp. 234–44.

Perkins, C., Malinen, J. T., Wakikawa, R., Nilsson, A. and Tuominen, A. J. (2002), "Internet connectivity for mobile ad hoc networks", *Wireless Communications and Mobile Computing*, 2:465-482.

Royer, E. M., Melliar-Smith, P. M. and Moser, L. E. (2001), “An analysis of the optimal node density for ad hoc mobile networks”, *Proceedings of IEEE ICC*, Helsinki, Finland, pp. 857-861.

Ruiz, P. M. and Gomez-Skarmeta, A. F. (2005), “Adaptive gateway discovery mechanisms to enhance Internet connectivity for mobile ad hoc networks”, *Ad Hoc & Sensor Wireless Networks*, pp.159-177, Vol.1.

Wakikawa, R., Malinen, J. T., Perkins, C. E., Nilsson, A., and Tuominen, A. J. (2001), “Global connectivity for IPv6 mobile ad hoc networks”, *Internet Draft draft-wakikawa-manet-globalv6-00.txt*.

Weniger, K. and Zitterbart, M. (2004) ”Address autoconfiguration in mobile ad hoc networks: current approaches and future directions”, *IEEE Network*, 2004, Vol. 18, No. 4, pp.6-11.

Web site addresses

AODV-UU/Mobile-IP in ns-2, http://core.it.uu.se/core/index.php/AODV-UU_and_Mobile_IP_for_ns-2.
IETF MANET WG Charter, http://www.ietf.org/html.charters/manet-charter.html.
IETF Mobility for IPv4 Charter, http://www.ietf.org/html.charters/mip4-charter.html.
Network Simulator ns-2, http://www.isi.edu/nsnam/ns/.