Load-Balance of Intra/Inter-MANET Traffic over Multiple Internet Gateways

Quan Le-Trung[1], Paal E. Engelstad[2], Tor Skeie[1], and Amirhosein Taherkordi[1]
[1]Department of Informatics, University of Oslo [IFI/UiO]
P.O. Box 1080 Blindern, NO-0316 Oslo, Norway
[2]Telenor Research & Innovation [R&I], 1331 Fornebu, Norway

E-mail: {quanle, amirhost}@ifi.uio.no, Paal.Engelstad@telenor.com, tskeie@simula.no

ABSTRACT

The paper first gives an overview of the required functions for providing Internet connectivity and mobility management for mobile ad-hoc networks (MANETs). Internet gateway selection is one of these functions. Since multiple Internet gateways might exist on the same MANET domain, a hybrid metric for Internet gateway selection is proposed as a replacement of the shortest hop-count metric. The hybrid metric provides load-balancing of intra/inter-MANET traffic. 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.

Categories and Subject Descriptors

C.2.6 [**Computer-Communication Networks**]: Internetworking – *protocol architecture, routing protocols, routers, standards.*

General Terms

Algorithms, Management, Measurement, Performance, Design.

Keywords

Ad-hoc networks, Internet gateway, load-balance, metrics.

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 multihop in the current access technologies such as wireless LANs or cellular networks [1-2]. Typically, the connection between a MANET node and an Internet gateway (IGW) is multihop. 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, cascading effect, can happen during the mobility of ad-hoc nodes within a MANET domain if multiple IGWs exist [3-5].

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-balance of intra-MANET traffic within a MANET domain [25-26], and that of inter-MANET traffic over multiple IGWs [12]. However, intra/inter-MANET traffic are considered separately. Moreover, the load-balance of inter-MANET traffic over multiple IGWs does not consider many realistic problems like inconsistent context [3-5]. In this paper, we want to control together these types of traffic. For this purpose, a hybrid metric for the load-balance of intra/inter-MANET traffic among multiple IGWs, and alternative solutions to reduce realistic problems [3-5] 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 in providing Internet connectivity for MANETs and mobility management. These are described in context of related work. In Section 3, a hybrid IGW selection metric is proposed in replacement of the shortest hop-count. It is used for load-balancing of intra/inter-MANET traffic in situations where there are multiple IGWs on the same MANET domain. An ns-2 [24] implementation of the ad-hoc on-demand distance-vector (AODV) routing protocol [21] that uses the above metric is presented in Section 4. It also uses mobile IP (MIP) [6] for the different IGW selection strategies, and the implementation is developed from the AODV and MIP package [23]. Section 5 presents the simulation scenario for testing the load-balance 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

MoMM 2008, November 24–26, 2008, Linz, Austria.

those of AODV using the shortest hop-count metric. Finally, conclusions and directions for future work are given in Section 6.

2. REQUIRED FUNCTIONS

In this section, the required functions of providing Internet access and mobility management for MANET nodes are described. 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. Related work is also discussed following the descriptions of above functions. Fig. 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 is used for the macro-mobility management, i.e., between MANET domains, while ad-hoc routing is used for the micro-mobility management, i.e., within each MANET domain. This is also the current trend in MANET mobility management [3].

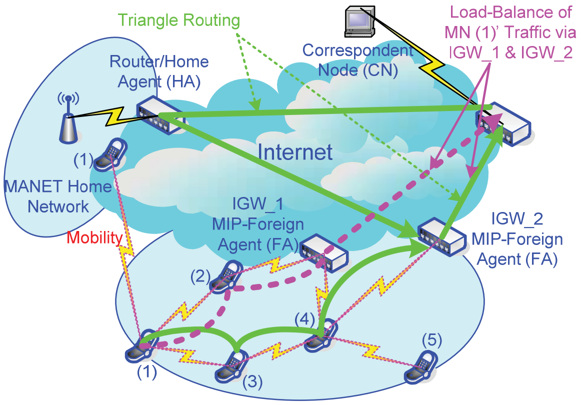

Figure 1. A Scenario of MANET mobility management.

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 [14].
- Routing table in MANET proactive ad-hoc routing protocols. 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 route request (RREQ) and waiting for route reply (RREP) in reactive ad-hoc routing protocols. If a host route is returned, the destination is in the same MANET domain. If a default route is returned, the destination is either an Internet host via the IGW, or unreachable [8].
- 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 [1], [14].

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 [10]. 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 [11].

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 [14].
- Load-balancing. For intra-MANET traffic, choose different immediate relays node to destination MANET nodes within the same MANET domain [25-26], while for inter-MANET traffic, choosing different IGWs for forwarding traffic from MANET to Internet and vice verse [12].
- Service class. Depending on the service classes provided and supported by each IGW.
- Euclidean distance. Spatial distance between the MANET node and the IGW [13].
- Hybrid. A combination of some of above metrics [13].

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 send packets to the IGW, where packets are forwarded to the destination in the Internet [14]-[15].
- Tunneling (or encapsulation). Usually, IP-in-IP encapsulation technique is used for traffic into and out of the Internet. 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 [7].
- 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 [7].

- 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 as a normal packet [16].

- 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 [17].

The intra-MANET forwarding strategies, on the other hand, is entirely based on the operation of ad-hoc routing protocols. These can be classified as proactive, or reactive, or hybrid [8]. 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 only 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 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 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* [19].

In the *former*, 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). In the *latter*, mechanisms assume that the addresses that are delegated are not being used by any node in the network. This can be achieved, for by ensuring that the nodes that participate in the delegation have disjointed address pools. In this way, there is no need of performing the DAD procedure.

Although an IP-based address auto-configuration scheme is preferred in self-organizing the MANETs for their fast deployments, only stateless mechanism is suitable for MANETs [18]. 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 MANETs. However, in the *conflict-detection allocation*, a DAD mechanism is required to assure the uniqueness of the address, 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 [14]: *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., WLAN, 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, 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 the IGW discovery to establish a new path to another IGW [20].

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 is via a different IGW, a *route optimization-based handoff* occurs.

3. A HYBRID LOAD-BALANCE 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 Fig. 2, 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.

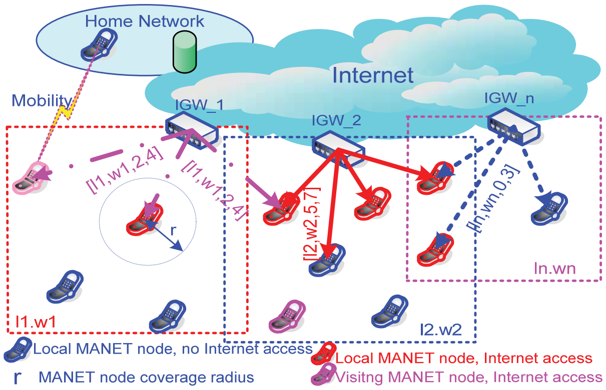

Figure 2.A hybrid load-balance metric.

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 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.

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:

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

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

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

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

$$LB_{MANET}(i,j) = \begin{cases} +\infty \quad \text{if} \ \left([AvgDeg(i,j)] \bmod \mathrm{K}\right) = 0 \\ \dfrac{1}{\left([AvgDeg(i,j)] \bmod \mathrm{K}\right)}, \ \text{otherwise} \end{cases} \tag{5}$$

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

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 Eqs. (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 Eq. (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 $n_{Reg}(j)$ at IGW_j that require Internet connectivity, see Eq. (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 [25] 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 [26]. However, in Eq. (6) average node degree is different for a local MANET node (Eq. 6a) and a visiting MANET node (Eq. 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. Eq. (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* is increased, taking maximum value when *AvgDeg* is equal to K (optimal node density), then decreasing even if the *AvgDeg* continues to increase [25].

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 [3-5], 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, 7, 10-11] has the potential to exploit efficiently multiple IGWs for the benefit of multi-homing or for performing soft handovers.

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 [23-24]. 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 [3-5, 7] is used, i.e., outbound traffic from source MANET nodes to the Internet hosts uses tunneling to avoid the inconsistent context problem [3], 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) [6] 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 problem [3].

- 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 [6].

- The information [lj, wj, nReg(j), nj] is attached into the proxy RREP sent by the corresponding IGWj.

- 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 [6]) 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 [6]) to its home agent to replace for the old one.

5. SIMULATION SETTINGS & 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.

- 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 its associated IGW, which can be changed due to the mobility of the ad hoc sources or the broken links. Its unit is second [sec].

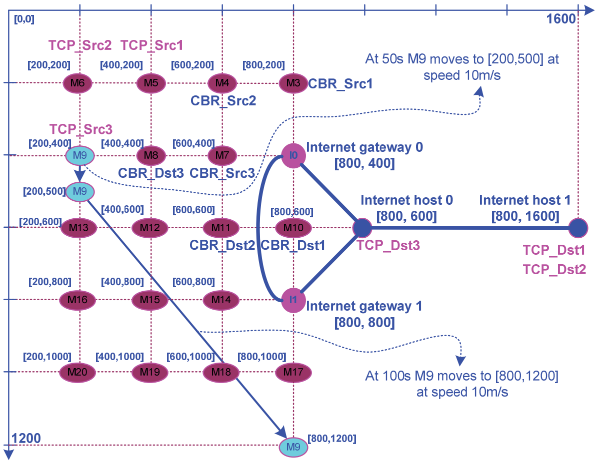

Figure 3.A Simulation scenario.

The simulation scenario is shown Fig. 3. It consists of two Internet gateways (IGW0, IGW1) and two Internet hosts (host0, host1), which are connected to each other using wired links (bandwidth: 5Mbps, propagation delay: 2ms), creating a connected wired network. Eighteen MANET nodes (M3→M20) are located at positions shown in Fig. 3, 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 IGW0 [l0=600m, w0=1000m, nReg(0)=3, n0=11] and IGW1 [l1=400m, w1=1000m, nReg(1)=0, n1=7]. Eleven MANET nodes, M3→M13, are initially in the sub-area managed by IGW0, while seven MANET nodes, M14→M20, are initially in the sub-area managed by IGW1 . Note that the values of [nReg(0), n0, nReg(1), n1] will be later changed to [nReg(0)=2, n0=10, nReg(1)=1, n1=8] through the simulation depending on the mobility of MANET node [M9]. 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 TCP1 [M5→host1, starting at 6.0s, stopping at 150.0s], TCP2 [M6→host1, starting at 11.0s, stopping 150.0s] and TCP3 [M9→host0, starting at 16.0s, stopping at 150.0s]. Three CBR connections (intra-MANET traffic) are also within the MANET domain, including CBR1 [M3→M10, starting at 5.0s, stopping at 150.0s], CBR2 [M4→M11, starting at 10.0s, stopping at 150.0s], and CBR3 [M7→M8, starting at 15.0s, stopping at 150.0s]. The MANET node mobility is set up as follows:

- At 50.0s, M9 starts moving to position [200, 500] at speed 10m/s.
- At 100.0s, M9 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 and 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.
- [α1,, α2, α3] 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-balance (LB) metric, respectively.

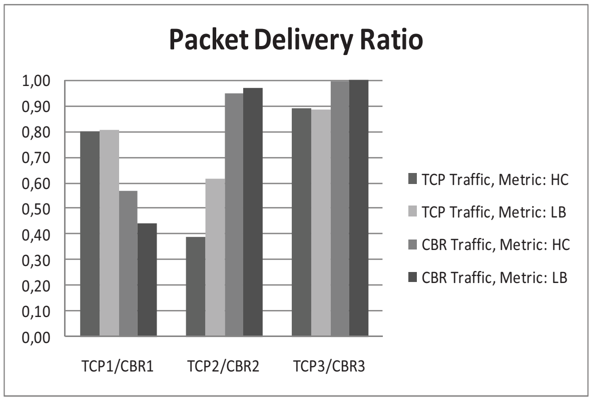

Figure 4.Packet delivery ratio.

Fig. 4 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 Fig. 3. 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 Fig. 3, 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-balance metric vs. that of AODV using the shortest hop-count metric, in TCP$_2$ and CBR$_1$ connections, see Fig. 4. This is because more TCP$_2$ traffic, in case AODV is used with our load-balance 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-balance 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-balance 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→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-balance metric, is lower compared with the shortest hop-count, see Fig. 4.

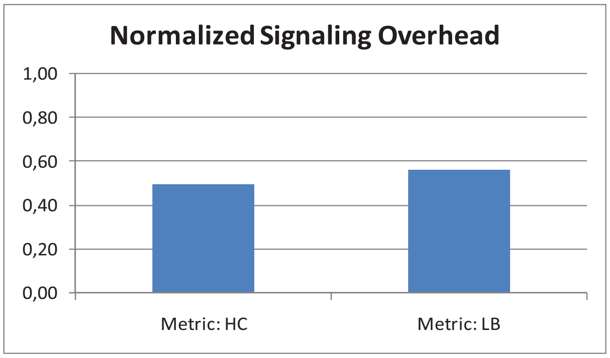

Figure 5.Signaling overhead.

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. Fig. 5 shows that the signaling overhead, in case AODV using our load-balance metric, is slightly higher (about *10%,* see Fig. 5).

Fig. 6 shows that the average packet transmission delay of TCP data packets from [M5, M6, M9]→IGW0 is almost the same under two cases, but lower for that from [M5, M6, M9]→IGW1, in case AODV using our load-balance metric. This is due to the below factors:

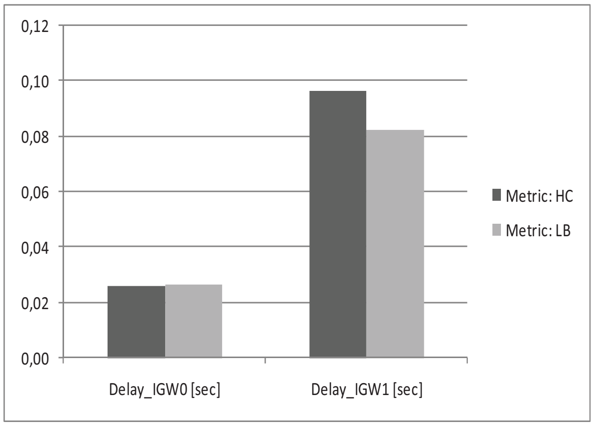

Figure 6.Average packet transmission delay.

- More inter-MANET traffic (TCP2), i.e., about 30% higher, are successfully transmitted in case AODV using our load-balance metric compared with the shortest hop-count.
- For AODV using the shortest hop-count metric, most TCP1,2 traffic is forwarded in/out via IGW0, while TCP3 traffic is forwarded between IGW0 and IGW1, depending on [M9] mobility.
- For AODV using our load-balance metric, most TCP1 traffic is forwarded in/out via IGW0. Part of TCP2 traffic approximately equal to the amount TCP2 traffic succesfully forwarded in case AODV using the shortest hop-count is forwarded in/out IGW0, and the rest of TCP2 traffic (about 30% that of AODV using the shortest hop-count) is forwarded in/out IGW1. Finally, TCP3 traffic is forwarded between IGW0 and IGW1, depending on [M9] mobility.
- Due to the mobility of [M9], i.e., source of TCP3, the average transmission delay of sending data packets from [M6→IGW1] on TCP2 is shorter than that of sending data packets from [M9→IGW1] on TCP3.
- 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 TCP2 data packets reduce the average delay of sending TCP2 traffic, shortening the average transmission delay of all TCP connections to IGW1, in case AODV using our load-balance metric instead of the shortest hop-count.

6. CONCLUSIONS & FUTURE WORK

This paper proposes a hybrid metric for IGW selection to balance the intra/inter-MANET traffic load among multiple IGWs on the same MANET domain. It considers three components: the Euclidean distance (in terms of hop-count), the load-balance of inter-MANET traffic (TCP), and the load-balance of intra-MANET traffic (CBR). A simulation scenario has been designed 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-balance 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 one mobility scenario. 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 [21], 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) [27], and compare with those in this paper.
- The setting thresholds of [α1, α2, α3] 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 nj 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 [lj, wj] managed by IGWj 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 [22], for traffic forwarding.

7. REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] Engelstad, P. and Egeland, G. 2004. NAT-based Internet connectivity for on-demand ad hoc networks. WONS'04, LNCS 2928, Italy, pp. 342-356.

[6] IETF Mobility for IPv4 Charter, http://www.ietf.org/html.charters/mip4-charter.html.

[7] 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.

[8] IETF MANET WG Charter, http://www.ietf.org/html.charters/manet-charter.html.

[9] 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.

[10] Jin, X. and Christian, B. 2002. Wireless Multihop Internet Access: Gateway Discovery, Routing, and Addressing. In Proceedings of 3GWireless'02, San Francisco, CA, USA.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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, pp.377-388.

[16] 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.

[17] Ergen, M. and Puri, A. 2002. MEWLANA-mobile IP enriched wireless local area network architecture. In Proceedings of VTC2002-Fall, pp.2449- 2453, Vol. 4.

[18] Bernardos, C. and Calderon, M. 2005. Survey of IP address autoconfiguration mechanisms for MANET. Internet draft-bernardos-manet-autoconf-survey-00.txt.

[19] Weniger, K. and Zitterbart, M. 2004. Address autoconfiguration in mobile ad hoc networks: current approaches and future directions. IEEE Network.

[20] 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.

[21] 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.

[22] Karp, B. and Kung, H. T. 2000. GPSR: Greedy perimeter stateless routing for wireless networks. MobiCom'00, Boston, MA.

[23] AODV-UU and Mobile IP in ns-2. http://core.it.uu.se/core/index.php/AODV-UU_and_Mobile_IP_for_ns-2.

[24] The Network Simulator ns-2. http://www.isi.edu/nsnam/ns/.

[25] Royer, E. M., Melliar-Smith, P. M. and Moser, L. E. 2001. An analysis of the optimal node density for ad hoc mobile networks. In Proceedings of IEEE ICC, Helsinki, Finland, pp. 857-861.

[26] 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.

[27] 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.