A Novel Scheme for Seamless Hand-off in WMNs

Hung Quoc Vo, Daesun Kim, Choong Seon Hong, Sungwon Lee and Eui-Nam Huh
Dept. of Computer Engineering, Kyung Hee University
1 Seocheon, Giheung, Yongin, Gyeonggi, Korea, 449-701
Phone:+82-31-219-2443, Fax:+82-31-219-1614
[e-mail: vqhung@networking.khu.ac.kr, {dskim, cshong, drsungwon, johnhuh}@khu.ac.kr]
*Corresponding author: Choong Seon Hong

Received May 27, 2009; revised July 14, 2009; accepted July 21, 2009
published August 31, 2009

Abstract

Although current wireless mesh network (WMN) applications experience a perceptually uninterrupted hand-off, their throughput after the hand-off event may be significantly degraded due to the low available bandwidth of the mobile client's new master. In this paper, we propose a novel mobility management scheme for 802.11-based WMNs that enables both seamless hand-off for transparent communications, and bandwidth awareness for stable application performance after the hand-off process. To facilitate this, we (i) present a new buffer moment in support of the fast Layer-2 hand-off mechanism to cut the packet loss incurred in the hand-off process to zero and (ii) design a dynamic admission control to grant joining accepts to mesh clients. We evaluate the benefits and drawbacks of the proposal for both UDP and TCP traffic, as well as the fairness of the proposal. Our results show that the new scheme can not only minimize hand-off latency, but also maintain the current application rates of roaming users by choosing an appropriate new master for joining.

Keywords: Wireless mesh network, hand-off, mobility, admission control

This work was supported by the IT R&D program of MKE/IITA [2009-F-016-01, CASFI] and the MKE, the Korean government, under the ITRC support program supervised by the IITA (IITA-2009-(C1090-0902-0002)).

DOI: 10.3837/tiis.2009.04.005

1. Introduction

Wireless mesh networks (WMNs) are distinguished from other wireless technology by their high reliability, self-configuration capability, self-healing, scalability, and low cost [1][2]. Two primary devices of a typical WMN are a wireless mesh router and wireless mesh client. For the sake of simplicity, we refer to these as the mesh router (MR) and mesh client (MC). MRs are connected in an *ad hoc* manner to form a network backbone. They also function as access points in WLANs that serve MCs. A subset of MRs are required to be attached to a wired network and to operate as gateways to the Internet (Internet gateways-IGW). MCs, either stationary or mobile, can gain network access via their serving MR.

Due to the mobility of WMNs, an MC is able to move out of its current MR's coverage and associate with a new MR with the minimum hand-off time to reduce packet loss as much as possible, because packet loss can degrade the performance of current applications significantly. In general, the hand-off process in WMNs is similar to that of WLANs. It includes two phases: Layer 2 hand-off (similar to roaming in the same domain or the same subnet in WLANs) and Layer 3 hand-off (similar to roaming between different domains or different subnets in WLANs) [3].

Several studies have attempted to reduce Layer-2 and Layer-3 hand-off latency in WLANs. However, most of these studies can not be applied directly to WMNs due to differences in the characteristics of the two kinds of network. APs in WLAN are considered to be Layer-2 devices while MRs in WMN work like real routers, maintaining routing tables and running routing protocols. Furthermore, although minimization of hand-off time can guarantee that current applications are maintained, the performance of the current applications can be significantly degraded due to the low available bandwidth of the new serving MR. These issues are discussed in greater detail in the following sections.

1.1 Issue Description

As described above, there are two major issues that should be take into consideration to design a complete hand-off scheme for MCs. First, the hand-off latency should be small enough to keep the connectivity smooth. Second, the current MC's applications should be able to perform in an acceptable range.

The former issue is related to the hand-off latency in WMNs, which depends on both Layer-2 hand-off latency and Layer-3 hand-off latency. The Layer-2 or MAC layer hand-off process has attracted interest because it is by nature a latent process [4][5][6]. Two factors responsible for this are the large number of channels probed by the MC and the waiting time required for the responses from the available MRs. The first factor is proportional to the number of non-overlapping channels in 802.11 standards (3 with 802.11b and 12 with 802.11a). The second factor depends on the network configuration. These problems can be addressed if the MC knows in advance how many channels it should scan and how many responses it should wait for on a specific channel.

As for the Layer 3 hand-off, the signaling hand-off trigger used in mobile IP and its extensions cannot be applied directly due to the fact that these have long delays, and therefore fail to support real-time traffic. Using a Layer-2 event to trigger the Layer-3 hand-off can reduce the total hand-off latency. In our paper, we consider another factor that supports seamless hand-off as opposed to fast hand-off by asking the question: When is the correct moment to buffer the data during the hand-off process? In current practice [6], the moment when the old MR receives the re-association event from the MC through the new MR is when the data is buffered. However, the problem is that MC cannot receive data from the currently associated MR while it is on channel scanning status for both active and passive modes. If the scanning time is long, the current applications will experience throughput degradation. In the worst case, they will be dropped.

The latter issue affects fairness among MRs and the maintenance of current applications after the hand-off process. In the conventional approach, an MC selects the best MR in terms of maximum RSSI for attachment [8][9][10]. However, maximum RSSI does not mean the highest available bandwidth. This approach may cause an MC to select the worst MR among available ones. In addition, there may be many stations connected to only a few MRs while others are idle. We therefore present a load-balancing mechanism using dynamic admission control that can be efficiently applied in WMNs. In our design, once an MR receives a request for association, it makes a decision as to whether to admit the connection or not, and also decides how much bandwidth to set aside for the new connection for its entire transmission duration.

1.2 Contributions

The key contributions of this paper are three-fold: (i) We propose a timely update mechanism to support the construction of dynamic NGs that significantly limits the number of responses on each scanned channel by the scanning MC, (ii) We present a new moment to buffer data to assure a seamless hand-off process (including Layer-2 and Layer-3) and (iii) We apply a dynamic admission control at each MR to ensure load balancing among them in terms of both the number of clients and the bandwidth consumed.

The rest of this paper is organized as follows: In Section 2, we present the network models and assumptions necessary for our analysis. Section 3 is dedicated to a description of the basic characteristics of the IEEE 802.11 hand-off process. Our proposed scheme for seamless hand-off in WMNs is presented in Section 4. Section 5 includes a performance analysis under various scenarios using network simulation tools. In Section 6, we discuss related studies, and in Section 7, we present our conclusions.

2. Related Work

Whenever a MC moves out of its current MR, it has to scan for new available MRs to join. A previous study [4] demonstrated that WLANs could not complete their Layer 2 hand-off process within 100 ms, even though the maximum end-to-end delay for voice applications is required to be lower than 50 ms. It has also been reported that the probing procedure used in Layer-2 hand-off is responsible for the high latency of the hand-off; approximately 90% of the whole process. In [11], M. Shin et al. presented a novel discovery method that uses a NG to facilitate learning of the neighbors' context, thereby limiting the number of probing channels. A non-overlap graph (NOG) for each channel has also been proposed to indicate the number of responses for which the station has to wait. Although the theory behind NG is lengthy and somewhat complicated, NG can be evaluated empirically by observing the actual movement of the MSs during operation. However, NG implementations have slow updates because AP-to-AP communication via the distributed system is not defined.

As for Layer 3 hand-off, Mobile IP [12][13] and its extensions, HAWAII [14] and Cellular IP [15], are applicable in both WLANs and WMNs. However, all of these protocols experience hand-off delays that cannot support time-sensitive applications. The main drawback of these protocols is the use of periodic signaling messages to trigger the Layer-3 hand-over. Although this type of design is suitable for wired networks, it is not suitable for wireless networks where the hand-off process includes both Layer 2 and Layer 3.

Several studies have discussed the load balancing problem of WLANs. In [8], probe response frames were re-formatted to perform load balancing among APs. Based on the information piggybacked on the probe response, such as the number of associated stations, RSSI value, and mean RSSI value, the MS can determine which AP to associate with to balance the number of MSs among APs. In [9], a three-level load balancing algorithm was proposed to provide a better balanced distribution of the number of STAs to the APs and to improve the overall network performance. However, in [8] and [9], the authors attempted to achieve load balancing by distributing all MSs to all APs in different ways, and did not consider the traffic load of APs. In [10], the authors found that though the number of users associated with each AP was the same, the offered load in terms of bandwidth varied considerably between APs. This indicates that the offered load is highly affected by individual user bandwidth requirements rather than just the number of users.

2. Network Models and Assumptions

2.1 Multi-radio Multi-channel WMNs

Each MR is equipped with multiple radios or multiple wireless cards. Each radio uses a non-overlapping channel. A static channel assignment [16] is used to lessen potential co-channel interference and increase the throughput of the networks. In **Fig. 1**, MR A and MR B are running in *ad hoc* mode on channel 1 while using channels 6 and 11 for their master modes, respectively. Notice that channel 1, channel 6, and channel 11 are three non-overlapping channels in 802.11b.

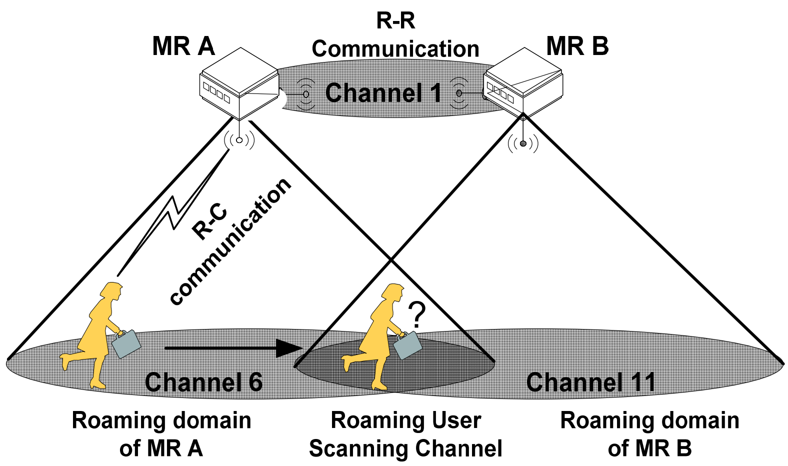


Fig. 1. A typical multi-radio multi-channel WMN.

There are two types of communication in a typical WMN: router-to-router (R-R) communication and router-to-client (R-C) communication. The former permits the exchange of information between two adjacent MRs, while the latter is for clients and their serving MR.

Each MR will manage a subnet or a domain and roaming between two MRs is considered inter-domain roaming where the IP should be re-configured at each client. There are a couple of circumstances in which the network can be expanded in terms of coverage and capacity without introducing additional subnets. The two additional configurations are coverage-oriented and capacity-oriented configurations.

2.2 Coverage-oriented Configuration

In a coverage-oriented configuration, an additional wireless Layer-2 (such as an access point) is installed to help the MR expand its coverage [3]. Some characteristics of coverage-oriented deployment include the following:

- Low packet application type, such as barcode scanning and database queries.
- Low bandwidth requirements, allowing data rate scaling down to lower data rates such as 1 or 2 Mbps.

In **Fig. 2**, A, B, and C are three MRs while D is a repeater of A. The overlapping range between A and D is greater than 50% of their individual transmission ranges.

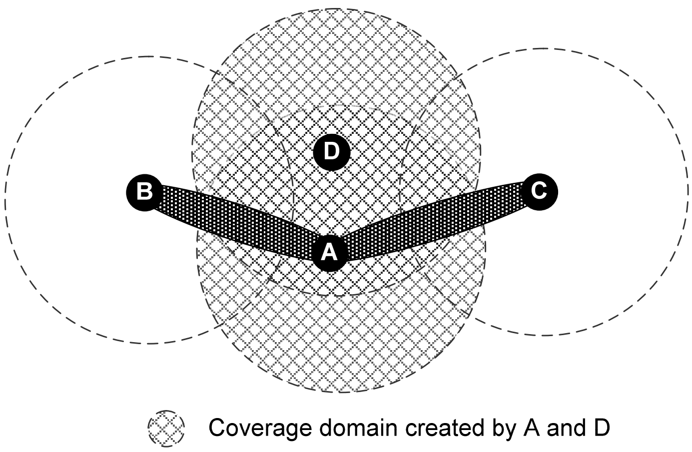


Fig. 2. Coverage-oriented configuration to expand coverage.

2.3 Capacity-oriented Configuration

A capacity-oriented configuration aims at providing high throughput and high packet rates for each client [3]. Capacity-oriented deployments are required for areas that have the following characteristics:

- High packet rate applications.
- Latency-sensitive applications.
- Dense client populations.

In **Fig. 3**, D, E, and X are three main MRs. Node X is composed of three nodes connected through a switch.

2.4 Traffic Models

Most WMN applications require hand-off support except those that have short lives such as domain name service (DNS). We model the traffic as controlled load service (CLS), which requires a minimum data rate (as for best effort, the minimum data rate is zero). Due to the properties of CLS, the required bandwidth of a CLS flow may vary within the range of $[b_{CLS}^{\min}, b_{CLS}^{\max}]$, where $b_{CLS}^{\min}$ and $b_{CLS}^{\max}$ denote the maximum and minimum bandwidth required for the CLS flow, respectively. The application can choose a data rate value b_{CLS}^{i} according to the agreement between the sender and receiver (flow control mechanism) such that $b_{CLS}^{\min} \le b_{CLS}^{i} \le b_{CLS}^{\max}$.

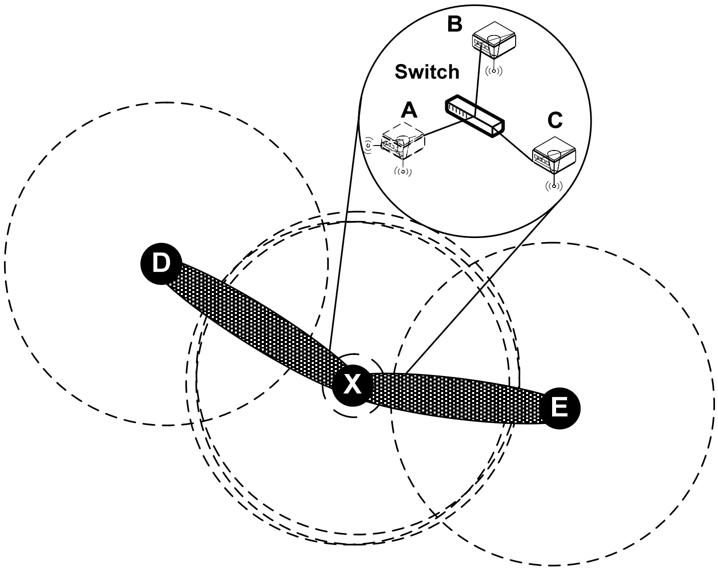


Fig. 3. Capacity-oriented configuration to increase capacity.

2.5 Bandwidth Estimation

In a wireless network environment, estimating the remaining bandwidth at a given time is not trivial due to the shared nature of nodes and their mobility. However, this can be performed for WMN for the following reasons. First, the MRs are absolutely static and the wireless environment is considered to be very stable. Second, our goal is to estimate the available bandwidth of serving channels at each MR instead of those forming the backbone.

Each MR can estimate its available bandwidth based on the ratio of idle time to busy time for a given period of time [17]. The MR is in business if any of the following criteria are met: (i) transmitting or receiving in progress, (ii) the network allocation vector (NAV) is being set and (iii) the MR is sensing a busy carrier with a signal strength larger than a certain threshold (however, the content of message cannot be interpreted). Accordingly, the available bandwidth can be calculated as follows:

$$B_{available} = B_{channel} \times T_{idle} / T_p \tag{1}$$

where T_{idle} is the amount of idle channel time during the period of time T_p and $B_{channel}$ is the channel capacity. The total available bandwidth or available capacity $B_{totoal\_available}$ of an MR is easily kept track of using the above-described technique.

3. Characteristics of 802.11 Hand-off

A hand-off process in WMNs involves both Layer 2 and Layer 3 and is carried out by a sequence of messages exchanged between an AP and a roaming user.

3.1 Layer-2 Hand-off Process

The complete Layer-2 handoff process typically comprises a sequence of decisions [3][18]:

1) Decision on when to roam - How to determine a good moment to initiate a hand-off is not defined in the IEEE 802.11 standard. Vendors implement roaming algorithms for their own products. Some factors such as the signal-to-noise ratio (SNR), frame acknowledgement, and missed beacons, are usually taken into consideration in the algorithm.

2) Decision on where to roam - The MS must determine a suitable AP, and break its connection with the current AP before associating with the new one. This is referred to as "break before make". Active scanning is the most thorough mechanism used to find an AP. As the name suggests, the MS sends probe requests in each channel and receives responses from available APs. The detailed active scanning procedure is as follows [4]:

- The MS broadcasts a probe request frame, sets a probe timer, and waits for probe responses.
- If no response has been received within $MinChanTime$, the next channel is scanned.
- If one or more responses are received within $Min-\ ChanTime$, the timer is extended to $Max-\ ChanTime$ obtain all possible probe responses.
- The above steps are repeated for the next channel. $CS\,\&\,T$ (Channel Switching and Transmission Over-head) is considered to be the channel switching time.

The probe delay T can therefore be approximated as follows:

$$N \times MinChanTime \leq T \leq N \times MaxChanTime$$

where N is the number of channels available (**Fig. 4**).

3) Decision to initiate a roam - The station uses 802.11 re-association frames to associate with a new AP. The procedure includes re-authentication and re-association. Authentication is a process in which the AP either accepts or rejects the identity of the MS. Once a successful authentication has been accomplished, the MS can send a re-association request frame to the new AP, which then replies with a re-association response frame containing an acceptance or rejection notice.

3.2 Layer-3 Hand-off Process

An 802.11-enabled client has to perform a Layer 2 roam before it can begin a Layer 3 roam. The Layer 3 roaming process focuses on issues surrounding Layer 3 routing, specifically the IP protocol and mobile IP extensions [14][15]. Roaming in a mobile IP-aware network involves the following steps, illustrated in **Fig. 5**:

- An MC communicates with its correspondent node (CN) while it is roaming from MR1 to MR2, which are located in different domains (or subnets).
- Whenever the MC detects the presence of a foreign agent (FA), it registers with the FA to obtain the care of address (CoA) from it.
- The FA then negotiates with the home agent (HA) to establish a tunnel between them.

- Packets destined to MC are forwarded to the HA, move through the tunnel to FA, and reach their destination (see the dashed line numbered 1).
- The route from MS to CN will be decided by FA. It can use the old route by employing the "reverse tunnel" mode or set up a new route directly to the CN (see the dashed line numbered 2).

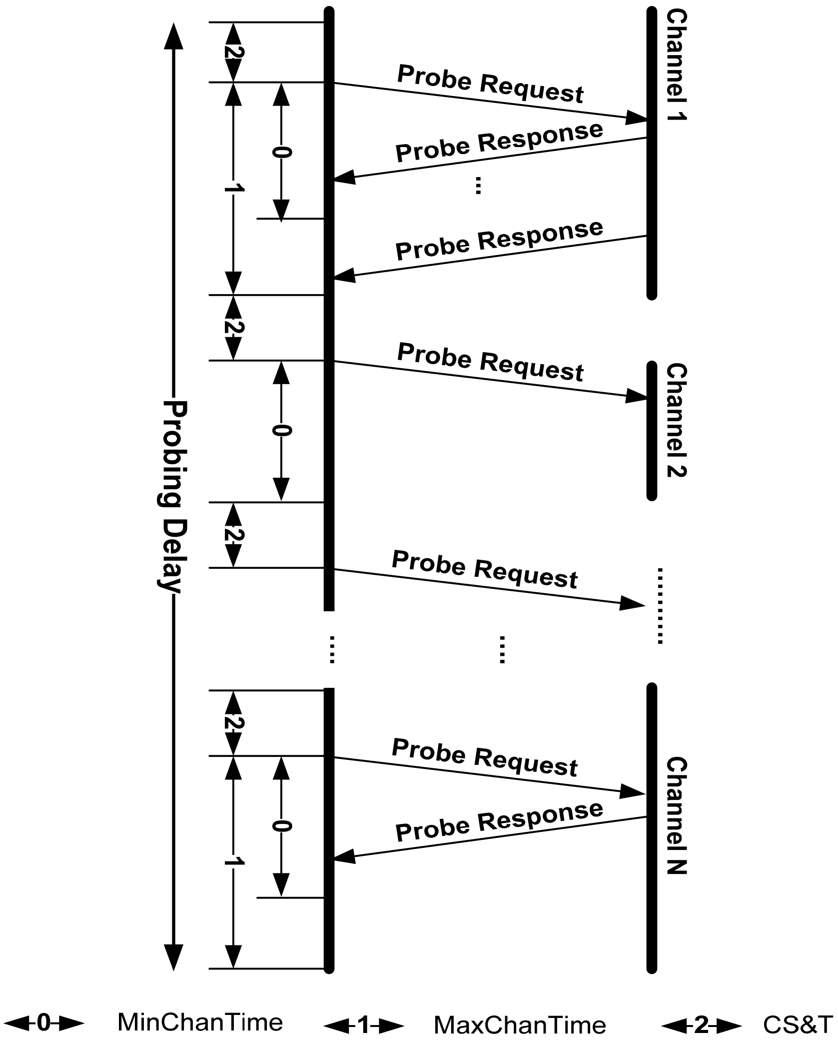


Fig. 4. The probing process in the IEEE Standard 802.11.

Tunneling in mobile IP is synonymous with encapsulation. Tunneling allows two disparate networks to connect directly to one another when they normally would not or when they are physically disjointed. This capability is key for mobile IP because tunneling is what allows the HA to bypass normal routing rules and forward packets to the MN.

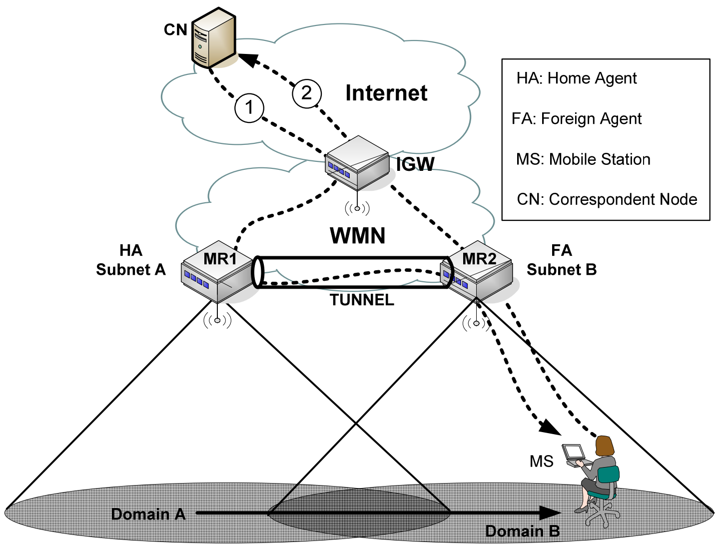


Fig. 5. Typical terms for the Layer-3 roaming process.

4. A Complete Hand-off Scheme for WMNs

4.1 Fast Layer-2 Hand-off

As described earlier, backbone routers in WMNs can exchange their information using peer-to-peer communication. This is also referred to as R-R communication. This facilitates timely updates of the network context among MRs. Accordingly, each MR will maintain a list of its physical neighbors, called the neighbor context table (NCT). The NCT of an MR contains the following information:

- IP and MAC addresses of its physically associated MRs and their serving channel (the channel for serving MCs).
- The supplemental configurations associated with each neighbor MR. Two popular types of supplemental configuration are capacity-oriented network configuration and coverage-oriented network configuration.

Whenever a new MC joins the MR, it will receive the NCT from its serving MR using R-C communication. NCT is updated by periodic local broadcast hello messages or event-driven triggers. Events include changes in neighbors' serving channel, neighbor adding or deleting actions, and neighbors' configuration adjustments. Algorithm 1 describes the process whereby an MC moves out of its current MR.

Fig. 6 illustrates part of a WMN with four main backbone MRs, namely A (channel k), B (channel n), D (channel j), and F (channel l). Thus, there are three rows corresponding to three neighbors in the NCT of MR A (as shown in **Table 1**). Router C is configured as a redundant router located next to MR B to double the network capacity (capacity-oriented network configuration). Router E is router D's repeater, which supports coverage expansion (coverage-oriented network configuration). These supplemental configurations are also indicated in the configuration field in each row in NCT. Based on the channel field in NCT, the MC knows that the maximum number of channels to scan is three (channel n, channel j, and channel l) when it moves out of its current MR, MR A. The MC also knows how many responses it should wait for on each listed channel. When it receives the response from router

D on channel j, it has to wait for another from router E, because E uses the same channel as D. The MC also knows how many additional channels it should scan besides the three official ones. When the roaming user receives the response from router B on channel n, it continues sending a probing message on channel m to scan router C.

Algorithm 1

1: **for all** channels *i* in NCT do
2: Broadcast *probe request* on channel *i*
3: Star probe timer
4: **while** True **do**
5: Read probe responses
6: **if** medium is idle until *MinChanTime*, expires **then**
7: **Break**
8: **else if** there is C-configuration* in the response **then**
9: Repeat the step 2 with channel *j*
10: **else if** there is R-configuration** in the response **then**
11: wait for another response
12: **else if** *MaxChanTime* expires **then**
13: **Break**
14: **end if**
15: **end while**
16: **end for**

*R-Configuration: coverage-oriented configuration
**C-Configuration: capacity-oriented configuration

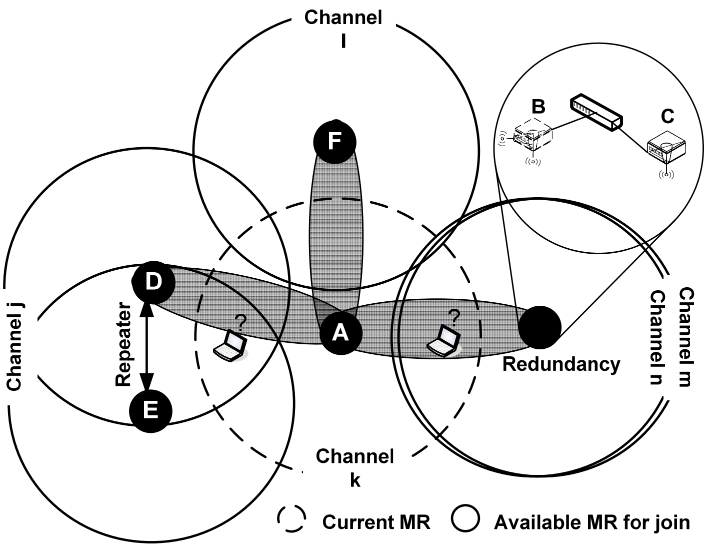


Fig. 6. Supplemental configurations in WMNs.

Table 1. Neighbor context table for mesh routers in WMNs.

Neighbor Context Table				
IP Address	MAC Address	Channel	Configuration	
IP-B	MAC-B	n	Redandency	IP-C, MAC- C, m
IP-D	MAC-D	j	Repeater	IP-E, MAC- E
IP- F	MAC-F	l	x	x

4.2 Buffer Moment for Seamless Hand-off

A MAC layer associate event can trigger the Layer-3 hand-off process more efficiently than conventional approaches such as Mobile IP, HAWAII, or Cellular IP that use periodic signalling messages for this purpose. A Layer-2 trigger is more accurate but consumes no extra bandwidth. However, the general terms used for mobile IP are for the most part kept in this study.

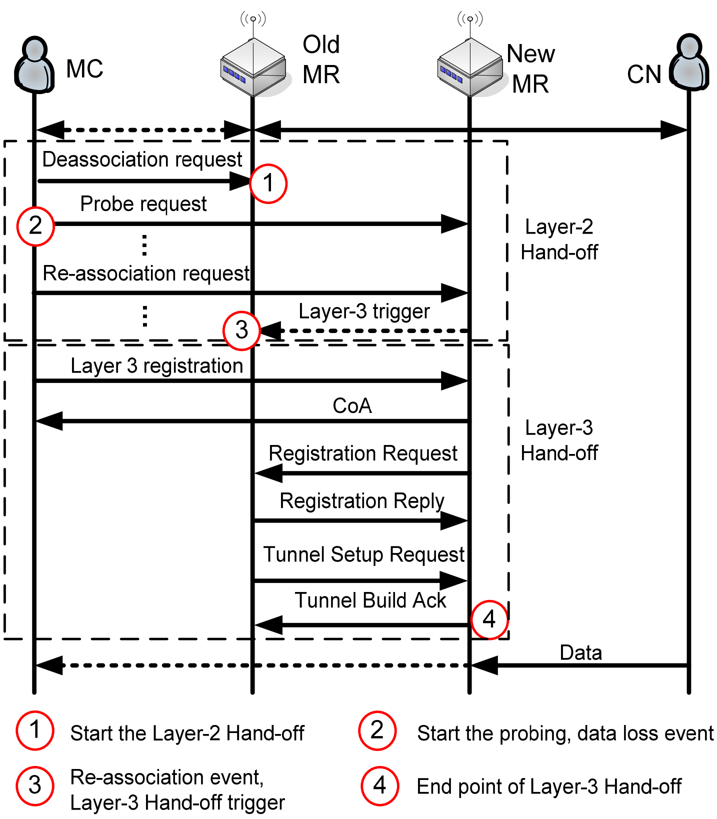


Fig. 7. The hand-off process with different buffering data moments.

Fig. 7 shows four important events in the entire hand-off process in an 802.11-based WMN. A de-association request (number 1) is generated by the MC to inform the MR that it is moving out. A probe request (number 2) is also sent from the MC to scan the available neighboring MRs. During the probing event, the MC cannot receive incoming data. A re-association event (3) is transmitted from the new MR to the old MR to finish the Layer-2 hand-off and trigger the Layer-3 hand-off. A tunnel build ack reply (number 4) is sent from the new MR to create a new ready tunnel through which to send destination packets.

The Layer-3 hand-off takes place immediately after the Layer-2 hand-off. Thus, using a Layer-2 event to trigger the Layer-3 hand-off process is optimal. The re-association event that the MC initiates to leave its HA should also be taken into account.

In current protocols, the re-association event is used as the signal to start buffering the data stream from the CN to the MC. This means that data will be dropped in the scanning phase because MC cannot both scan available channels and receive data packets. This results in TCP traffic degradation due to the slow start phase, which seriously decreases the current rate.

We propose that data should be buffered when the MC sends a de-association event to its current MR to break the Layer-2 association and start the Layer-2 hand-off process. This will ensure that the hand-off latency is zero, albeit at the cost of higher memory usage. With a low Layer-2 hand-off latency, however, memory usage can be reduced significantly to a value acceptable for MRs.

4.3 Bandwidth-Aware Roaming with Dynamic Admission Control

MC should join a new MR that can meet its bandwidth requirements to keep the performance of current applications almost unchanged after the hand-off process. In this study, we only consider the circumstances under which the MC has to choose the best available MR to associate with among several MRs in the capacity-oriented configuration (for coverage-oriented configuration, the MC will join the MR with the highest RSSI).

To avoid joining available MRs with low bandwidth, the MC should piggyback the bandwidth requirement in its probe request. The MRs, on receiving the request, evaluate their current potential residual bandwidth ΔW to decide whether to accept the connection or not.

$$\Delta W = B_{total\_available} - \sum_{i}^{N} b_{CLS}^{\min} . \tag{2}$$

Acceptance will be granted if the potential residual bandwidth is greater than the minimum bandwidth requirement of the roaming MC.

$$\Delta W > b_{CLS}^{\min} . \tag{3}$$

Among several acceptances from available MRs, MC should choose the best one, which it assumes to have the highest potential residual bandwidth.

We define another bandwidth term: the actual available bandwidth. The actual available bandwidth ΔB is the amount of bandwidth that can be granted to new connections without impacting the existing connections.

$$\Delta B = B_{total\_available} - \sum_{i}^{N} b_{CLS}^{i} . \tag{4}$$

The information about the aforementioned bandwidth is easily calculated by using the bandwidth estimation technique described in Section 2 and is piggybacked on the response message to be delivered to the MC in the scanning phase.

If a new connection gets acceptance from MR_x, the following question can be asked: How much bandwidth will the new connection be granted? To answer this question, we consider the following three cases:

- $b_{CLS}^{\max} \le \Delta B$, the new connection will be granted the amount of bandwidth $b_{CLS}^{\max}$ while existing connections maintain their current rates.
- $b_{CLS}^{\min} < \Delta B < b_{CLS}^{\max}$, the new connection will get ΔB bandwith while existing connections maintain their current rates.

$\Delta B < b_{CLS}^{\min}$, the new connection is granted a band-width equal to $b_{CLS}^{\min}$. However, to achieve this, existing connections will have their current rate degraded by the amount of $k \times \delta$ where δ is the degradation step (it is a predefined, tunable value) and k is a positive integer. The value of $k \times \delta$ is chosen such that the new bandwidth values of current applications are not less than their minimum bandwidth requirements. $b_{CLS}^{i} - k \times \delta \ge b_{CLS}^{i,\min}$. Algorithm 2 describes the process that the MR uses to evaluate the joining request.

Algorithm 2	
1:	**for** a probe request from MC i do
2:	**if** $b_{CLS}^{\min} > \Delta W$ **then**
3:	refuse to send a probe response
4:	**Break**
5:	**else if** $b_{CLS}^{\min} < \Delta W$
6:	send *probe response* to accept the joining
7:	**end if**
8:	**if** $b_{CLS}^{\max} \le \Delta B$ **then**
9:	set aside $b_{CLS}^{\max}$ for the new connection
10:	**else if** $b_{CLS}^{\min} < \Delta B < b_{CLS}^{\max}$ **then**
11:	set aside ΔB for the new connection
12:	**else if** $\Delta B < b_{CLS}^{\min}$ **then**
13:	$b_{CLS}^{i} = b_{CLS}^{i} - k \times \delta$

14: set aside $b_{CLS}^{\min}$
15: **end if**
16: **end for**

b_{CLS}^{i}, $b_{CLS}^{\min}$, $b_{CLS}^{\max}$ are defined in section 2.4

We consider an example to clarify our preceding analysis. One MR with two $Mbps$ serving channels is managing eight MCs. Only one application is currently running on each MC. The required minimum bandwidth and actual consumed bandwidth of each application are listed in **Table 2**. From the table, we can calculate the potential residual bandwidth and actual available bandwidth, $\Delta W = 500kbps$ and $\Delta B = 100kbps$ respectively. A new connection with $b_{CLS}^{\min} = 300kbps$ will be accepted with a degradation of 3 or $4 \times \delta$ $(\delta = 10kbps)$ required for each current MC. **Fig. 8** illustrates the degradation $4 \times \delta$ of each application in the presence of the new joining compared to the original values.

Table 2. Bandwidth characteristics of mesh clients.

MC	$b_{CLS}^{\min}$	b_{CLS}
1	250 kbps	300 kbps
2	250 kbps	300 kbps
3	250 kbps	300 kbps
4	200 kbps	250 kbps
5	200 kbps	250 kbps
6	150 kbps	200 kbps
7	150 kbps	200 kbps
8	50 kbps	100 kbps

A bandwidth estimator should be added in the operation architecture of an MR to measure the bandwidth. The accuracy of available bandwidth estimation and admission control can be measured by a metric called the right admissions, β, defined as:

$$\beta = \frac{\text{Number of right admissions}}{\text{Number of flows in the network}} \quad (5)$$

We performed a simulation to investigate β. The topology is presented in **Fig. 1**. MR B is connected to the Internet. A CN is the sender, and MCs that are moving randomly in the roaming domain are receivers. The bandwidth requirements of each connection are uniformly distributed between [100,300] kbps. **Fig. 9** presents the relation between the value β and the number of clients in the mesh network. When the number of MCs is added up, β decreases. However, when the number of clients hit 50, the value of the right admissions metric is still high compared to that of an *ad hoc* network [19]. The static natures of MRs in a WMN will not affect this estimation technique, and the results are expected to be stable and reliable.

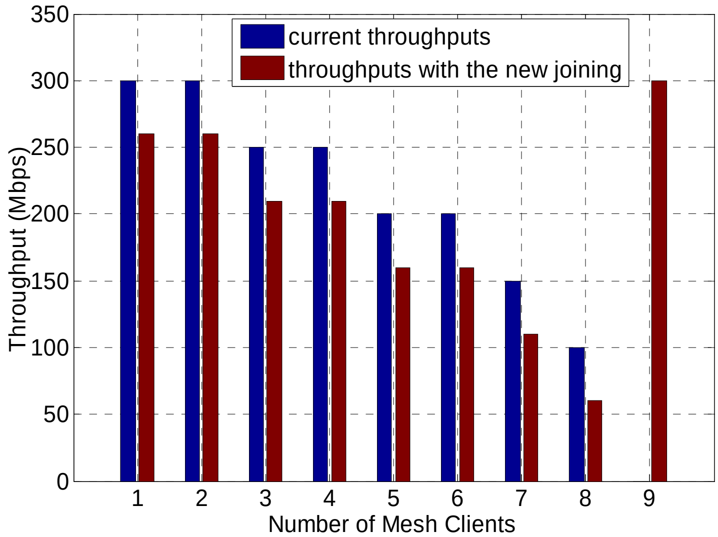


Fig. 8. The degradation of each application in the presence of the new joining.

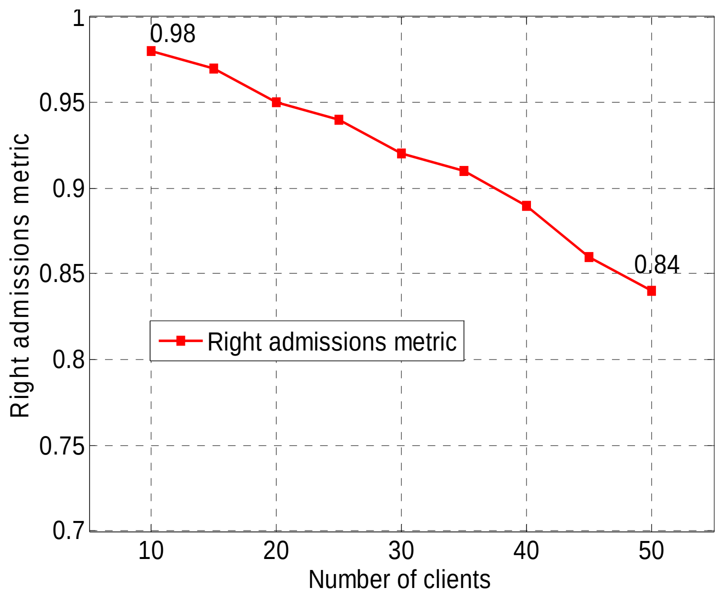


Fig. 9. Admissions right metric versus the increase of MCs.

5. Performance Analysis

The overall goal of our simulations was to measure the hand-off latency and its impact on UDP and TCP traffic, the load balancing ability of multiple MRs, the fairness of multiple MRs, and the increase in overall throughput.

5.1 Performance Metrics

We considered three performance metrics to evaluate our proposed method.

- **Packet Delivery Ratio**: The packet delivery ratio (PDR) is defined as the ratio of the data packets delivered to the destination to those generated by the CBR sources. It describes the loss rate that will happen due to hand-off latency. PDR is used to evaluate UDP traffic.
- **Connection throughput**: The connection throughput of a network is defined as the average transmission rate of a connection in the network. Note that the higher the connection throughput is, the less delay packets experience during transmission. Therefore, the connection throughput is a good indicator of the average end-to-end delay in the network. This metric is used to evaluate TCP traffic throughput in this paper.
- **Fairness evaluation**: We use Jain's fairness index, $F = (\sum_{i=1}^{N} th_i)^2 / (N \sum_{i=1}^{N} th_i^2)$, where th_i denotes the throughput achieved by the ith station and N is the number of stations. In the ideal case where all the stations have the same throughput, F becomes one.

5.2 Simulation Environment and Implementation

In this section, we report the results of several simulations we conducted to evaluate the hand-off latency, the impact of our protocol on both UDP and TCP traffic, and the fairness of our proposed scheme. The simulation platforms are based on ns-2.31 extensions.

The proposed protocol implementations include a new hello message that contains the ID, serving channel, and additional configurations of the neighboring node. The probing procedure follows Algorithm 1 in Section 4.1. An event update is triggered whenever a change in the network occurs. A buffer event starts right after the MR receives the de-association event from the MC. This buffer is discarded when the hand-off event is detected to have failed. The bandwidth estimator based on the NAV value is added for dynamic admission control. Joining acceptance is granted based on Algorithm 2 in Section 4.3.

5.3 Hand-off Latency Measurement

Each MR is equipped with two 802.11b wireless network interface cards. One card operates in *ad hoc* mode to form the mesh network backbone. The other card operates in master mode to serve its mobile users or MCs. We considered the topology presented in **Fig. 10**. This network topology comprises four MRs, namely A, B, C and D, running in *ad hoc* mode on the same channel (channel 11). Every MR has two neighbors. Two neighboring MRs use non-overlapping channels (channels 1 and 6) for their master modes. MR C is physically connected to a server that is considered to be in CN node. The CN is running a voice application with an MC moving in the roaming domain formed by the four MRs.

An MC takes three rounds (red dashed line in **Fig. 10**) in the roaming domain to measure hand-off latency. Each MR maintains a NCT that contains the information of its two neighbors. Initially, a static NCT is issued to the MC immediately after it joins MR D.

In our evaluation, we consider a G711-encoded VoIP application with a data rate of $64 kbps$ and a packet size of 200 bytes (40 byte header and 160 byte payload). Additional parameters are given in **Table 3**.

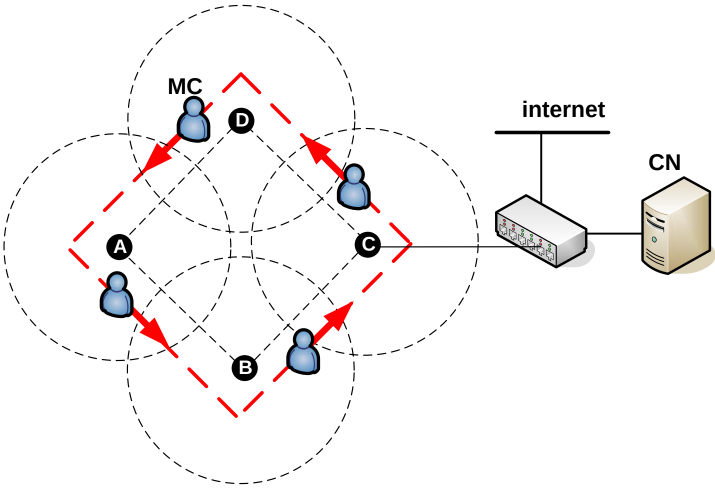


Fig. 10. The mesh network testbed topology.

Table 3. Simulation parameters.

Parameters	Value
Total area	400 m x 400 m
Number of Nodes	4
Transmission Power	Unlimited
Receive Signal Threshold	3.152e-20 W
Transmission Range	250 m
Path Loss Factor	1.6-1.8
Simulation Time	100 s

Fig. 11 shows the latency for various pairs of hand-offs: from D to A, A to B, B to C and C to D. The average values are 38.83 *ms*, 45.76 *ms*, 36.2 *ms*, and 31.33 *ms*, , respectively. The measurement takes place at the beginning of the Layer-2 hand-off latency, which is marked with a probe request. The end of the measurement is specified by the exchange of a registration reply message between the new and old MRs. The A-B pair has the longest hand-off time, 45.76 *ms*, , among all MRs because the distance between these two MRs is the longest path (two hops) from the new master to the original one (MR D).

5.4 Evaluation of UDP Traffic

To evaluate the impact of hand-off on UDP traffic, the network configuration is kept the same as in the previous section, and the packet delivery ratio is measured.

Fig. 12 shows the packet delivery ratios obtained useing two protocols: the fast hand-off protocol proposed in this paper, and mobile IP extension. A successful rate is observed at both MC (when CN is the sender) and CN (when MC is the sender). With fast hand-off, the delivery ratio is up to more than 95% for MC as the receiver and 92.36% for MC as the sender. This difference in delivery ratios between MC as the receiver and MC as the sender is because when MC is the sender, the traffic stream is dropped during the hand-off time because the MR cannot buffer the data from MC. For Mobile IP, the PDRs were 64.1% and 54.28%, respectively, indicating that this protocol cannot support real-time traffic, despite the fact that the mobile IP protocol that we considered has been modified to work well in WMNs.

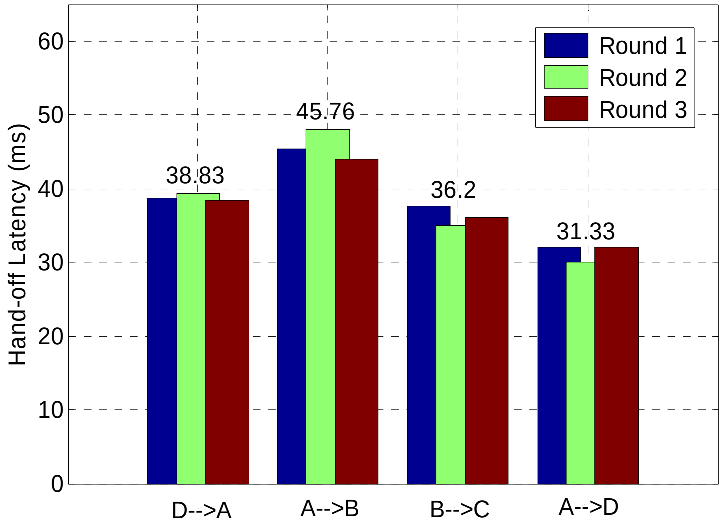


Fig. 11. Hand-off latencies for various pairs of mesh routers.

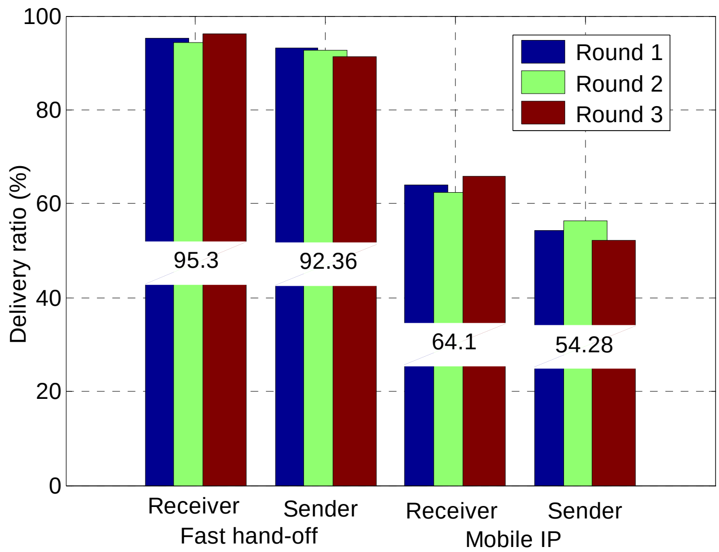


Fig. 12. PDRs seen at both the receiver and sender.

5.5 Evaluation of TCP Traffic

To evaluate the impact of our proposed protocol on TCP traffic and make comparisons with existing schemes, an FTP application that runs between the MH and the CN was evaluated using three different scenarios: seamless hand-off, fast hand-off, and mobile IP extension.

The first two scenarios employ Layer-2 hand-off with context-awareness using dynamic NCT. The difference between them is the moment when data buffering is triggered: our protocol uses a de-association event trigger while the other protocol [7] uses a re-association event trigger. The third scenario employs typical mobile IP that uses a signalling message trigger.

As can be seen in **Fig. 13**, our protocol facilitates real-time seamless hand-over. In other words, packet loss decreases to zero during hand-off events. Fast hand-off can also reduce packet loss, but the throughput still decreases due to the slow start mechanism of TCP when packet loss occurs. In the third scheme (hand-off with typical mobile IP), the throughput reaches zero during hand-off time due to large latency (up to a few seconds) caused by the Layer-3 hand-off signalling trigger. It should be noted, however, that our seamless hand-off scheme comes at the cost of increased memory usage.

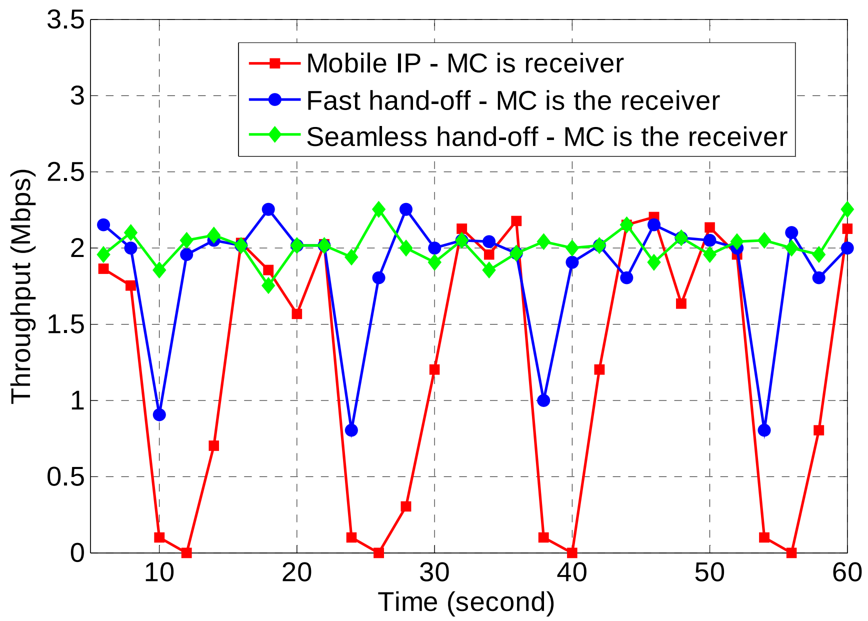


Fig. 13. TCP throughput observed at the CN node.

5.6 Evaluation of Load-balancing

To evaluate the effect of our protocol on dynamic admission control to balance not only the number of clients but also the offered load, we conducted a simulation using ns-2. We consider a topology with two MRs working on the same channel (MR1 and MR2), a CN node, and six mobile nodes called MCs (**Fig. 14**). MR1 is configured to be the redundant router for MR2 to double network capacity (capacity-oriented configuration). Both MR1 and MR2 connect to the CN node through a 100 Mbps cable. The MCs send traffic to the CN node through MRs and the throughputs are observed at the CN node.

Each MC takes turns (every 10 s) to move into the overlapping coverage formed by the two MRs. **Fig. 15** shows that each MR can effectively share the load. Three MCs, namely MC1, MC4 and MC6, join MR1 at 10th s, 40th s and 60th s, respectively while the other three MCs, MC2, MC3, and MC5, associate with MR2 at 20th s, 30th s and 50th s. The bandwidth consumed at each MR is balanced with the even number of clients.

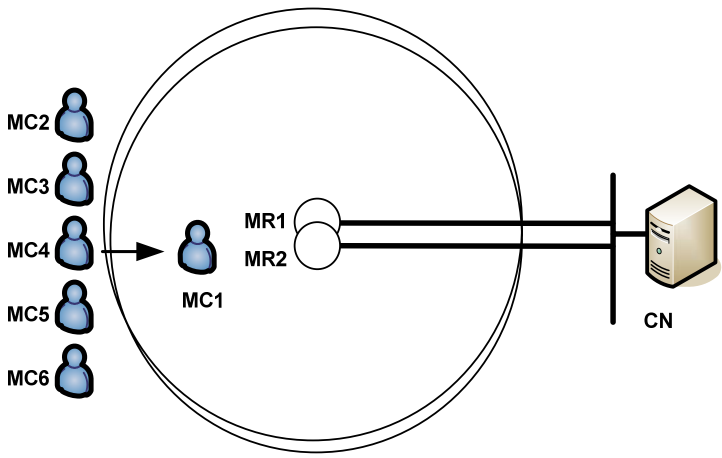


Fig. 14. Topology for the load balancing simulation.

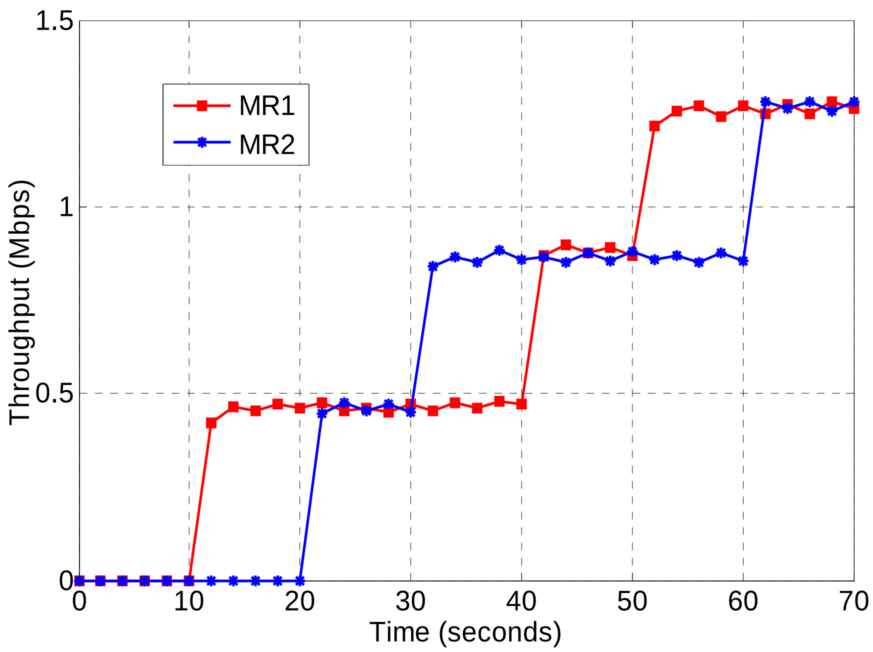


Fig. 15. Load-balancing between two MRs.

We increased the number of clients and measured Jain's fairness index for two scenarios: (i) the MCs decide to join solely based on RSSI and (ii) acceptance is granted due to the current load status of MR in addition to RSSI. For scenario (i), Jain's fairness index decreased when the number of clients increased. In the latter, the index maintained a stable value near 1. The difference in plots shown in **Fig. 16** has been discussed previously. Using RSSI as the sole criterion cannot guarantee fairness among multiple access points.

To evaluate the performance of our protocol for a general topology, we used the scenario depicted in **Fig. 17**. There are four primary MRs: C, D, X, and Y. X and Y use supplemental configurations to improve the network capacity. D is connected to the Internet, and therefore has the role of Internet gateway. We increased the number of MCs to measure the overall throughput in four cases: (i) Fast hand-off without load balancing, (ii) Fast hand-off with load balancing, (iii) Seamless hand-off without load balancing and (iv) Seamless hand-off with load balancing.

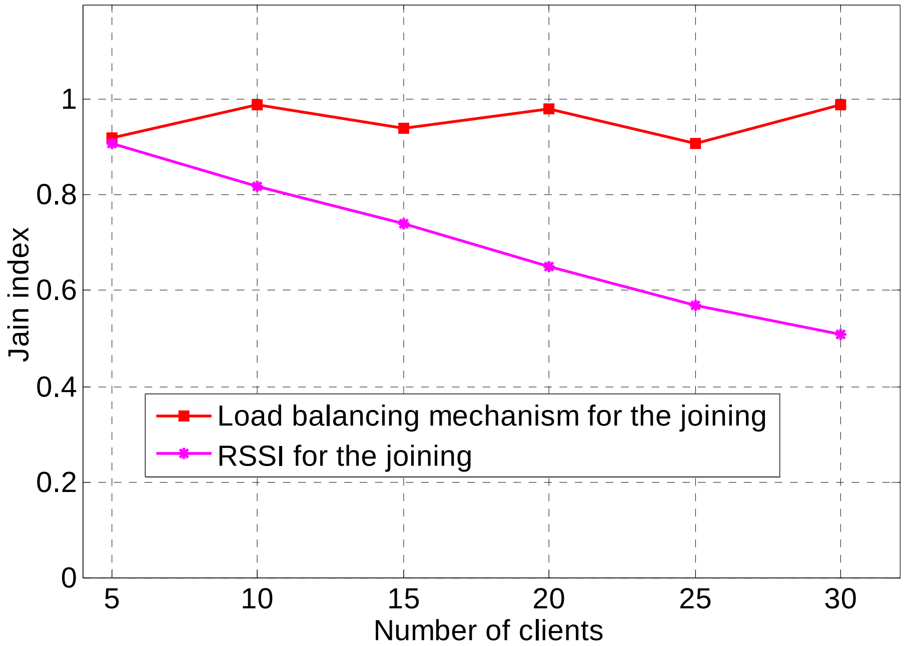


Fig. 16. Load-balancing between two MRs.

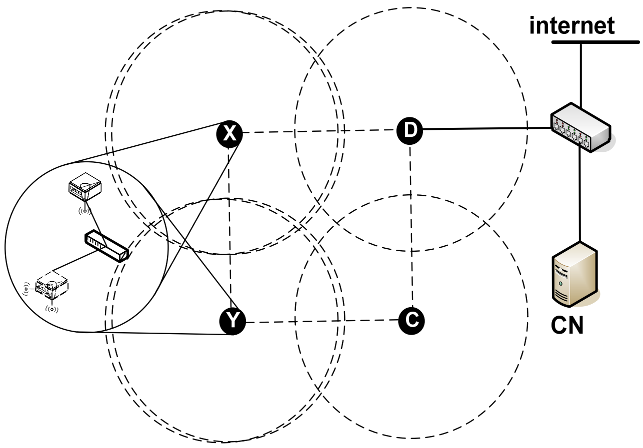


Fig. 17. Simulation topology for throughput measurements.

The results of the simulation are shown in **Fig. 18**. Fast hand-off without load-balancing had the lowest performance among all scenarios, especially when the number of clients exceeded 30. Seamless hand-off with load balancing achieved the best throughput when the traffic load was high. Fast hand-off with load balancing showed superior performance to seamless hand-off without load balancing when the number of clients reached 32. These results demonstrate that a load balancing mechanism improves throughput significantly when the load is high.

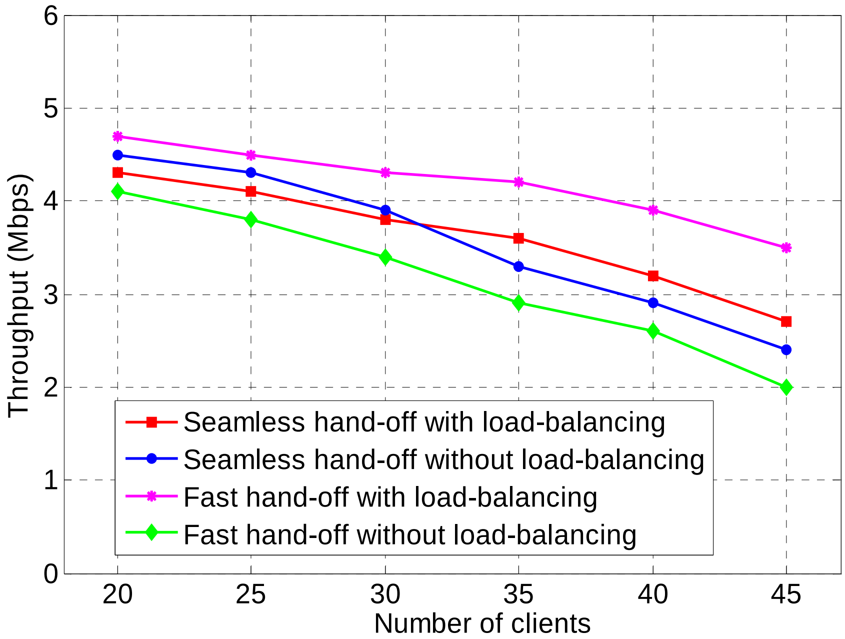


Fig. 18. Load-balancing between two MRs.

7. Conclusions and Future Work

In this paper, we described and evaluated a complete scheme for seamless hand-off in WMNs. The scheme involves necessary enhancements in both Layer-2 and Layer-3 hand-off processes along with the needed bandwidth criteria to assure continuous operation of current applications running at the time of hand-off. We proposed a protocol to construct dynamic NGs for context transfer among MRs to obtain the minimum number of scanned channels that can possibly support a low hand-off time for Layer-2. We also chose a new moment to buffer data to minimize packet loss during Layer-3 hand-off. To maintain the performance of on-going applications after the hand-off process, we recommend using a dynamic admission control for each MR to balance the load and help the MC choose the right master.

Based on our experimental and simulation results, we conclude that our scheme is highly effective and achieves all of our proposed research goals. We speculate that it will perform best when there are a large number of non-overlapping channels available such as in 802.11a; this hypothesis will be investigated in follow-up studies.

References

[1] I.F. Akyildiz and X. Wang, “A survey on Wireless Mesh Networks,” *IEEE Communication Magazine*, vol. 43, no. 9, pp. S23-S30, 2005.

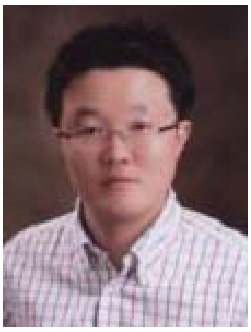
[2] N. Nandiraju, D. Nandiraju, L. Santhanam, Bing He, Junfang Wang, and D.P. Agrawal, “Wireless Mesh Networks: Current Challenges and Future Directions of Web-In-The-Sky,” *IEEE Wireless Communications*, vol. 14, pp. 79-89, Aug. 2007.

[3] Jonathan Leary, “802.11 Wireless LAN Fundamentals,” Cisco Press, 2008.

[4] A. Mishra, M. H. Shin, and W. Albaugh, “An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process,” *ACM SIGCOMM Computer Communication Review*, vol. 3 pp. 93-102, Apr.

2003.
[5] S. Shin et al., "Reducing MAC Layer Handoff Latency in IEEE 802.11 Wireless LANs," *Proc. ACM MobiWac 2004*, Oct. 2004.
[6] S. Pack, J. Choi, T. Kwon, and Y. Choi, "Fast-Handoff Support in IEEE 802.11 Wireless Networks," *IEEE Communications Surveys & Tutorials*, vol. 9, First Quarter 2007, pp. 2-12, 2007.
[7] H. Wang, Q. Huang, Y. Xia, Y. Wu, and Y. Yuan, "A Network-Based Local Mobility Management Scheme for Wireless Mesh Networks," *Proceedings of WCNC*, Hong Kong, pp. 3792-3797, Mar. 2007.
[8] H. Gong and J. Kim, "Dynamic load balancing through association control of mobile users in WiFi networks," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 342-348, May 2008.
[9] I. Papanikos and M. Logothetis, "A Study on Dynamic Load Balance for IEEE 802.11b Wireless LAN," *Proc. 8th International Conference on Advances in Communication & Control*, *COMCON 8*, Rethymna, Crete, Jun. 2001.
[10] A. Balachance, G. Voelker, P. Bahl, and V. Rangan, "Characterizing User Behavior and Network Performance in a Public Wireless LAN," *Proceedings of ACM SIGMETRICS 2002*, Marina Del Rey, California Jun. 2002.
[11] M. Shin, A. Mishra, and W.A. Arbaugh, "Improving the Latency of 802.11 Infrastructure Networks," *Proc. ACM Mobisys 2004*, Jun. 2004.
[12] R. Malekian, "The Study of Handover in Mobile IP Networks," *Proc. of BROADCOM 2008*, pp. 181-185, Nov. 2008.
[13] D. Johnson, C. Perkins, et al., "Mobility Support in IPv6," IETF RFC 3775, Jun. 2004.
[14] R. Gunasundari, S. Shanmugavel, and Bakyalakshmi, "Fault Tolerant HAWAII Protocol for IP services in Wireless Networks," Proc. of AINAW 2008, pp. 457-462, Mar. 2008.
[15] A. Campbell, J. Gomez, S. Kim, A. Valko, and C. Wan, "Design, Implementation and Evaluation of Cellular IP," *IEEE Personal Communications*, Jun.-Jul. 2000.
[16] K.N. Ramachandran, E.M. Belding, K.C. Almeroth, and M.M. Buddhikot, "Interferce-aware channel assignment in multi-radio wireless mesh networks," *Proc. of INFOCOM*, Barcelona,Spain, Apr. 2006.
[17] Y. Yang and R. Kravets, "Contention-Aware Admission Control for Ad Hoc Networks," *IEEE Transactions on Mobile Computing*, vol. 4, no. 4, pp. 363-377, July-Aug. 2005.
[18] IEEE, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," IEEE Standard 802.11, 1999.
[19] C. Sarr, C. Chaudet, G.e Chelius and I.G. Lassous, "Bandwidth Estimation for IEEE 802.11-based Ad Hoc Networks," *IEEE transactions on Mobile Computing*, vol. 7, no. 10, pp. 1228-1241, Oct. 2008.

Hung Quoc Vo received a B.Eng. degree in Electrical Engineering from Hochiminh University of Technology in 2005. During 2005-2006, he worked in Nortel Networks Laboratory (SER 5500), located at TMA Solutions Company, Vietnam. His research interests include wireless and mobile networking and performance analysis of networking systems. He is currently a Master's student in the Department of Computer Engineering, Kyung Hee University, Korea.



Dae Sun Kim received his M.S. and Ph.D. degrees in Computer Engineering from Kyung Hee University, Seoul, Korea, in 2002 and 2008, respectively. His research interests include mobile networking based on multi-homing environments and performance analysis of wireless networking. He is currrently a postdoc in the Department of Computer Engineering, Kyung Hee University, Korea.

Choong Seon Hong received his B.S. and M.S. degrees in Electrical Engineering from Kyung Hee University, Seoul, Korea, in 1983 and 1985, respectively. In 1988 he joined KT, where he worked on Broadband Networks as a member of the technical staff. In September 1993, he joined Keio University, Japan. He received his Ph.D. degree at Keio University in March 1997. He worked for the Telecommunications Network Lab, KT, as a senior member of the technical staff and as a director of the networking research team until August 1999. Since September 1999, he has been working as a professor of the School of Electronics and Information, Kyung Hee University. He has served as a Program Committee Member and an Organizing Committee Member for International conferences such as NOMS, IM, APNOMS, E2EMON, CCNC, ADSN, ICPP, DIM, WISA, BcN, ICOIN and TINA. His research interests include ad hoc networks, network security, and network management. He is a member of IEEE, IPSJ, KIPS, KICS, and KISS.

Sungwon Lee received the Ph.D. degree from Kyung Hee University, Korea. He is a professor of the Computer Engineering Departments at Kyung Hee University, Korea. Dr. Lee was a senior engineer in the Telecommunications and Networks Division at Samsung Electronics Inc. from 1999 to 2008. He is a editor of the Journal of Korean Institute of Information Scientists and Engineers: Computing Practices and Letters.

Eui-Nam Huh earned a B.S. degree from Busan National University in Korea, a Master's degree in Computer Science from the University of Texas, USA in 1995, and a Ph.D. degree from the Ohio University, USA in 2002. He was a director of the Computer Information Center and Assistant Professor at Sahmyook University, South Korea during the 2001 and 2002 academic years. He also served as the program chair of the WPDRTS/IPDPS community in 2003. He is an editor of the Journal of the Korean Society for Internet Information and he has been the Korean Grid Standard group chair since 2002. He was also an Assistant Professor at Seoul Women's University, South Korea. He is now a Professor in the Department of Computer Engineering, Kyung Hee University, South Korea. His research interests include high performance networks, sensor networks, distributed real-time systems, grid middleware, monitoring, network security, and cloud computing.