DE GRUYTER OPEN

Journal of ELECTRICAL ENGINEERING, VOL. 66, NO. 2, 2015, 70–78

A REINFORCEMENT ROUTING ALGORITHM WITH ACCESS SELECTION IN THE MULTI–HOP MULTI–INTERFACE NETWORKS

Amir Hosein Jafari — Hadi Shahriar Shahhoseini *

In this paper, a routing algorithm is proposed for access selection in a network to find the optimal paths among intermediate nodes with multiple interfaces. Markov Decision Process is applied in each node to find optimal policy and select proper paths to the best access point in a dynamic environment. A reward function is defined as environment feedback to optimize and adapt routing behavior of nodes based on the local information. Selection metrics in each node are interface load, link quality and destination condition. It is shown, by using the proposed algorithm, there are better management in the node which decreases interference and collision and selects links with better quality toward the best possible destination. The performance of the method is exemplified and it is shown how the throughput and average delay of the network with more interface in its nodes, improved while packet loss degrades. As an example a two-interface and a one-interface network are studied. It is shown when network load is increased, interface management will improve the throughput, in the network with two-interface nodes. Also, by considering the link quality factor in the reward function, packet dropping becomes less but average delay increases.

K e y w o r d s: routing, multi-destination, load balancing, reinforcement learning

1 INTRODUCTION

Rapid development of technologies in the field of wireless networking resulted in making multi interface devices. Todays, almost all of the network nodes are equipped to different technologies such as cellular networks and WLAN which can use different types of service [1–3]. In this environment, selecting the proper interface for connection is an important challenge. Simple problem has been widely studied when the node can see some access point (APs) related to its interfaces. This problem called network selection. More complex problem is changing the interface and consequently network AP during the routing in a multi-hop multi-interface network. We focus on this problem in this paper. Algorithms for network selection in a single hop network which includes multi base station/access point (BS/AP) have been divided into five categories: cost-function based, utility based, statistical based *eg* Markov Decision Process (MDP), intelligent based and policy-based algorithms [4].

There are similar challenges in these networks, that a part of their solution can be used for our problem. The first one is coverage extension in a multi-hop and heterogeneous network. Multi-hop communication can be used for improving the network capacity and extending the coverage in the case that a user cannot connect to any APs. In this case, mobile stations (MSs) communicate with the AP via intermediate nodes. Several approaches have been proposed for coverage extension by combing multi-hop with other cellular networks. In [5] Yang *et al* presented a policy-based quality of service (QoS) supporting system and a QoS-aware routing algorithm based on ad-hoc overlay networks to improve performance in the cellular networks. The algorithm allows the networks to adjust their policies such as interface selection of intermediate nodes to meet network requirement. By this, the users would be diverted from congested cells to neighbor cells and bandwidth would be reserved for the diverted call in intermediate nodes.

The second problem is load balancing within the network and/or over destinations. There are several approaches that try to do this by suitable routing. Among them some have used learning and switching from one interface to another one [6, 7]. Load balancing between AP [8, 9] which belongs to Radio Resource Management (RRM) and call diverting have been used in the centralized and distributed manners during recent researches. It uses resource sharing in the heterogeneous networks to improve the system capacity and service delivery for users [10, 11]. In [12], Hu *et al* presented dynamic load balancing implementing a routing protocol to use cellular networks and ad hoc technology with fixed relays, simultaneously. They located a number of fixed relays called integrated cellular and ad hoc relaying (iCAR) to divert traffic from one congested network to another non-congested ones. This improves the received QoS by the users and provides a cost-effective way to overcome the congestion problem by dynamically balancing the traffic between different access networks. Authors in [13] performed resource utilization in overlapping heterogeneous networks by cooperative load balancing by packet routing based on ad hoc networks which are used in the different applications [14]. In [15] Arun used mobile nodes as

* Electrical Engineering Department, Iran University of Science and Technology, Tehran, Iran hshsh@iust.ac.ir

DOI: 10.1515/jee-2015-0011, Print ISSN 1335-3632, On-line ISSN 1339-309X © 2015 FEI STU

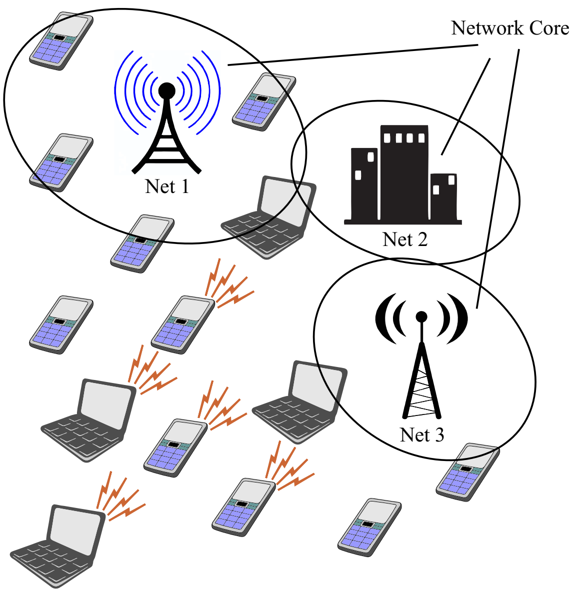

Fig. 1. A heterogeneous network with different technologies

routers to support seamless mobility for users in the network results in load balancing between APs. Route overhead in studied environment due to maintaining route was calculated by Hu for generated control messages, but the simulation for overhead was not presented. Cavalcanti *et al*, in [16] introduced an Integrated Routing Protocol (IRP) used in the wireless heterogeneous networks for integration cellular and WLAN results in traffic distribution between BS/AP. It exploited link quality, number of hops as metrics and topology information in routing mechanism. However, the problem of limitation capacity of intermediate nodes for transferring traffic has not been considered. In [17], Sharif *et al* studied a hybrid routing protocol in two active and proactive regions for load balancing based on multi-hop transmission. In proactive region, APs send advertisement messages contained load information to nodes of the network. In the reactive region, mobile nodes should find nodes in the proactive region to communicate with APs through them. However, due to their algorithm was based on flooding, total overhead and route discovery load increase when traffic of sources in the reactive region increase.

The last related issue in the previous works which its situation may be used in multi interface node routing is gateway selection in the heterogeneous wireless networks. In [18], a multi-metric gateway selection mechanism have been proposed in the architecture integrating ad-hoc with UMTS networks by Manoharan. Gateways in integrated structure have been used to forward data packets of nodes. Their node selection algorithm uses simple additive weighting for residual energy, signal strength and mobility speed metrics utilized in the dual interface nodes to sustain connectivity of the integrated network.

In this paper, multi-destination routing through different communication interfaces of network nodes is studied. We considered multi-interface nodes which can connect a source node to the best destination while destinations and intermediate links may have dynamic conditions. MDP is used for decision making in network nodes to adapt and optimize routing behavior in the dynamic environment by using feedback based on reward function. Cross layer information including MAC and physical layer conditions in terms of link quality, node delay and load as well as access characteristics are considered for reward calculation. In the proposed algorithm, each node has been modeled as an agent which maps its observations into actions related to environment feedback and browses the optimal route path through a set of policies and evaluates them by trial through interaction with the environment. The local information has been used in the proposed algorithm instead of the total network information and route selection is done in a distributed manner.

2 NETWORK CONSIDERATION

It is assumed that the network consists three main components (1) mobile stations (MSs), (2) BS/APs as different possible destination to connect MSs to IP nework and (3) a core IP network, which connects APs/BSs to a requested server by MSs (Fig. 1).

The MS has equipped by multiple radio access and can simultaneously receive the offered services through assumed access networks (Fig. 2). In fact each interface of nodes makes a plane; we called it interface plane, and depicts in Fig. 3. Each MS can connect to one of the APs existed in interface planes in two ways: single-hop and multi-hop mode. In the single-hop mode, it can connect to an AP/BS directly, and in the multi-hop mode, it uses the routing mechanism to find the best path among MSs to reach a MS which can connect to AP/BSs in single hop mode. We call this last MS in the path as a gateway

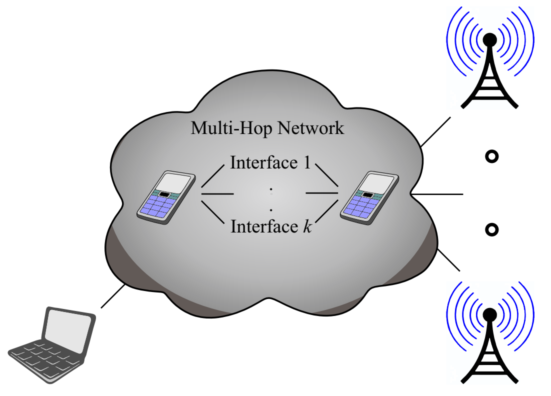

Fig. 2. A multi-destination access network with multi-interface nodes

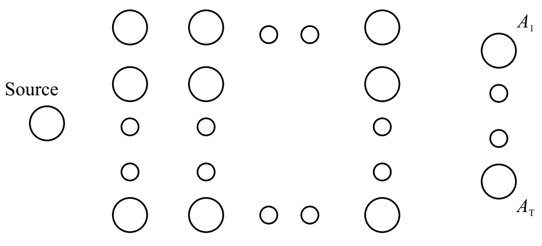

Fig. 3. A typical interface plane

(GW). In the considered network, there are N nodes. Each node has and T interfaces as shown in Fig. 2. There are A_i APs/BSs ($i = 1, \dots, T$) in each interface plane, to support the connection to radio networks i (Fig. 3). The nodes which are in coverage each AP/BS network i can connect to that AP/BS in single hop manner and also act as a gateway to that AP/BS for other MSs. The number of gateway nodes for access network (AP/BS) i is G_i. Generally when a MS wants to send its packets to the core network, it tries to find a GW and consequently an AP through intermediate nodes by a routing mechanism. The routing mechanism tries to find the best path through any T interfaces of each node. We model the network state by MDP approach and try to find the optimal policy in choosing the best path.

3 THE PROPOSED MULTI INTERFACE ROUTING

In some cases, the users are not covered by a network, or because of non-uniform distribution of traffic for the AP, a percentage of users may miss the service due to the congestion of APs; while other AP (for the same interface or another interface) may have enough bandwidth. In this regards, users' packets can be transferred to other AP which has lower load through other MSs. This mechanism, we call it multi-destination routing, will also result in load balancing between APs [16, 17].

This section presents a MDP-based framework for each MS to select next node and its related appropriate interface. Because of bandwidth limitation of the intermediate nodes, they may be congested by arrived packets for forwarding to APs. So, an intermediate node may be bottleneck in delivering the arrived packet to destinations. To overcome the bandwidth limitation of intermediate nodes and dynamic link conditions, the channel conditions, and the impact of existing load at each intermediate node including queue delay in each node and link quality of node's interfaces has been considered. The interfaces in the investigated multi-hop-multi-radio network are assumed to be different in the data rate.

3.1 The Proposed Routing Algorithm

The routing algorithm has been presented in this section. As mentioned earlier, the considered network contains nodes cooperate to connect MSs to APs in the following cases: (1) – the users are not under the coverage of the APs, but want to connect to the best APs through GWs, (2) – the existing AP has been congested while the other APs have enough capacity to serve the clients, and (3) – the users need services, but the AP in same interface plane are not accessible on appropriate time. In each of the above cases, by using the considered algorithm, the user's packets can reach to an appropriate AP in the multi-destination multi-interface network. The objective of the proposed routing algorithm is to transmit the users' packets to the best AP in the network while considering the QoS requirements of users. Since the environment is dynamic and the conditions of networks and links vary, MDP approach is employed for selecting the proper interface and the next node which the nodes have no knowledge about network dynamics.

As shown in Fig. 2, each intermediate node n has T choices (interfaces) to forward packets, and each interface have different characteristics and impose QoS constraints on that interface. Clearly, if the routing algorithm selects the best interface, the network resources will be better utilized.

Intermediate nodes in the network update their route table according periodically, as a proactive approach. By this, mobile hosts also learn where destinations are. To keep the costs of routing data updated between nodes in the routing tables, each AP alternately sends an update packet (UP) including the ID of the relevant AP and the characteristics of its related load along with the services it supports.

They are periodically broadcast from APs towards intermediate nodes. Each intermediate node periodically forward UPs that contain the least transmission cost to its neighbors. UPs are used to update network status including the cost information associated with the neighbors relative to the destinations. Jointing a new node as source to send packets is done based on a dynamic address resolution mechanism to obtain an address. The intermediate nodes decide to forward source packets to the best APs as the destination. The routing cost in terms of different metrics as reward from one node to another is calculated based on the information received from the UPs, link conditions and the length of queue of one-hop node interface described in the next section. Decision making is performed according to the local information of each node. We utilized MDP to help intermediate nodes to select the best interface and finding the next node as next hop in reaching to the selected APs.

3.2 MDP Formulation

To obtain the optimum route from source to the best destination among the existing ones as a decision problem, each node should choose the appropriate interface and consequently, the next hop node. So, each intermediate node is considered as an agent and network states are modeled by MDP.

By using the MDP and modeling the behavior of nodes and their interactions with network conditions, each node employs a trial-and-error manner for finding its better candidate action. The action outcome would be observed as a reinforcement that, leads an adaptation to the optimal policy including interface and node selection with regards to the experiences and information related to environment dynamics and status of destination in a distributed way. Each node tries to improve the overall reward of the network by using its local information from neighboring agents. Through the acquired learning, each agent takes an action based on adopted policy (Fig. 4).

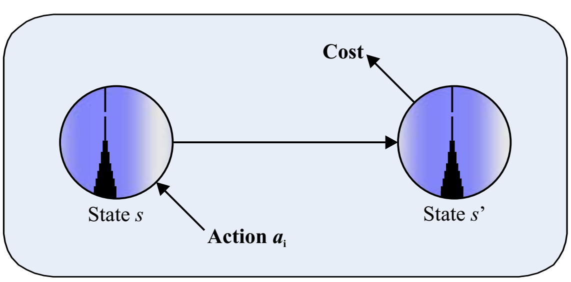

Fig. 4. Action, state and cost

The components of an MDP model are:
S – The state space $s \in S$,
A – The set of actions $a \in A$,
r – Reward function $r\colon S \times A \times S' \to r(s,a,s')$

A '*state*' is the packet position that should be sent from node l to node m in next hop $(l, m \in \{1,\dots,N\})$ through interface k.

When an action is chosen for node l and packet $p_l \in K_{S,t}$ (where $K_{S,t}$ is the set of packets of state S at time t) is transmitted from interface k of node l to node m, the probability (Pr) where packet $p_l \in K_{S,t}$ or $p_l \in K_{S',t}$ can be obtained as

$$\begin{cases} Pr_l^m(s' \mid s,a)\,, & p_l \in K_{S',t}\,, \\ 1 - Pr_l^m(s' \mid s,a)\,, & p_l \in K_{S,t}\,. \end{cases} \tag{1}$$

By choosing an action (a), a packet in state s goes to the next state (s') with transition probability of $P^a_{ss'}$ and receives a reward r_{t+1}. The decision variable $a \in A(s)$ is an action that determines an interface and selects a next hop node m. As shown in Fig. 4, an action triggers a transition from state s to s'. The transition probability for action (a) is obtained as follows

$$P^a_{ss'} = pr\{s_{t+1} = s' \mid s_t = s,\ a_t = a\}\,. \tag{2}$$

When an agent chooses action (a) at time t with regards to system state s, it receives the reward $R^a_{ss'}$ in proportion to transition probability $P^a_{ss'}$ according to

$$R^a_{ss'} = \{r_{t+1} \mid s_t = s,\ a_t = a,\ s_{t+1} = s'\}\,. \tag{3}$$

The goal of every agent is to apply *policy* π in such a way that the obtained reward in time t (R_t) is maximized with regards to the sum of long-term rewards.

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{4}$$

where γ is the discount factor, which is added to the above equation for converging (4). The future reward is due to action a on time t.

Reward function has been defined based on network conditions and performance metrics. Here, link quality for different node interfaces and load of nodes as well as destination conditions have been considered for reward calculation which are described more thoroughly in the next section.

In our proposed algorithm, the probability of transition between the states is assumed to be one, initially, regarding to the quality of all links. So the packets will go to the next hop nodes from each interface with equal chance. With the elapse of time, to have a safe transmission of packets from the node l to the node m in the next hop through link L_{lm}, each link can be monitored by nodes at time intervals of τ seconds. For calculating the probability of state transition, the percentage of safely-arrived packets in link L_{lm} has been calculated. For this purpose, we consider the ratio of sent packets in a link to received packets as the probability of safe transmission of packets along link L_{lm}. If only channel conditions prevailed, we could use $P_{err} = 1 - (BER)$ where BER is bit error rate of link. But now, since congestion in each link may be the cause of packet losses, the probability of safe arrival of packets in link L_{lm} is calculated based on the statistics obtained from the probability of safe arrival of packets along link L_{lm} via interface k that is different depending on the route length, [19]. The transmission probability from node l to m through interface k can be calculated from

$$P^k_{lm}(s' \mid s,a) = \frac{Pk^k_s}{Pk^k_t}\,, \tag{5}$$

$$Pk^k_s = a_1\,Ack^k_s + a_2\,CTS^k_s + a_3\,RTS^k_s + a_4\,Data^k_s + a_5\,Upd\_pk^k_s\,, \tag{6}$$

$$Pk^k_t = a_1\,Ack^k_t + a_2\,CTS^k_t + a_3\,RTS^k_t + a_4\,Data^k_t + a_5\,Upd\_pk^k_t \tag{7}$$

where, the parameters Ack_x^k , CTS_x^k, RTS_x^k, $Data_x^k$ and $Upd\_pk_x^k$ for $x \in \{s,t\}$, respectively, indicate the number of Ack, CTS, RTS, data and UPs which has been sent successfully from interface k, and also parameters s and t indicate the number of successful and total packet sent. Since the error probability of a packet in a transmission channel depends on the length of that packet, coefficients a_d $(d = 1 \dots 4)$ in (5) can be proportionate to the ratio of different packet lengths over the mean length of packets. Parameter 'Pk_s^k' indicates the number of packets delivered to node m, and parameter 'Pk_t^k' expresses the total number of packets sent over transmission link L during transmission in the physical layer of the interface k of node l.

To obtain the optimal policy (deciding on the selection of the next interface and node) in the presented MDP model, each agent as a node should maximize of the obtained reward. Policies which are used by each agent are based on local information in a distributed manner. For this purpose, the '*value* v' array should include real values and the '*policy* π' array should include the actions associated with *state* s. Finally policy π can be obtained as follows

$$\pi(s) = \arg\max_a \Big\{ \sum_{s'} P_a(s,s')\big(R_a(s,s') + \gamma V(s')\big) \Big\},$$
$$V(s) = \sum_{s'} P_{\pi(s)}(s,s')\big(R_{\pi(s)}(s,s') + \gamma V^t(s')\big). \tag{8}$$

As was previously stated, the nodes of a network cooperate with each other as agents, and their overall objective is to obtain the most reward for the network. For cases that the agents in a network cooperate in a distributed manner, according to [7, 20], the general value function can be approximated as the sum of value-action functions of agents

$$Q(S,a) = \sum_{i=1}^{n} Q_i(s_i, a_i). \tag{9}$$

For finding the optimal $\pi(s)$, the Variable Elimination Algorithm (VEA) [21] is used in order to obtain the highest value for $V(s)$. In VEA, the agents are eliminated one by one. Before an agent (node) be eliminated, it first sums up the values of all the payoff functions corresponding to its edges. The agent then returns the conditional payoff function (which expresses the highest value it can obtain in actions corresponding to its neighbors) and the best strategy related to that maximum value.

Let us assume that there is a network consisting of H hops (Fig. 3), and the amount of value function from source to destination is equal to U. According to (9), the value function can be written as the sum of value functions obtained from each hop by

$$\max_{a\in A}(U) = \max_{a_1,a_2,\dots,a_h} \{f_{12} + f_{23} + \dots + f_{h-1\,h}\} \tag{10}$$

where U expresses the global value of the network, and each f_{pq} function indicates a local payoff function between the agents of hops p and q. (10) can be written as

$$\max_{a\in A}(U) = \max_{a_1,a_2,\dots,a_H} \{f_{12} + f_{23} + \dots + f_{H-1\ H}\}. \tag{11}$$

The nodes of the h-th hop can be eliminated and their effects can be considered in the form of function $f_{h\,H}$ (11), provided that actions 'a' lead to the maximization of U.

$$\max_{a\in A}(U) = \max_{a_1,a_2,\dots,a_{h-1}} \{f_{12} + f_{23} + \dots + f_{h-1\ h} + \max_{a_h,a_{h+1},\dots,a_H} (f_{h-1\,h} + \dots + f_{H-1\,H})\} \tag{12}$$

As can be observed in Fig. 3 and (10), at each time interval, the nodes of hop h obtain their minimum payoff value (P_h) from each destination (by means of (11)), which this amount only depends on the payoffs of hops $h+1$ to H (P_{h+1} to P_H). This procedure continues until the best strategies for each node in the h-th hop are obtained so that U (total network payoff) has its highest value. By doing this jointly in a distributed way ($a \in A$), the best route (best policy) for each node will be calculated.

3.3 Reward Calculation

This section deals with the manner of calculation of reward which is used in a node for making decisions. It has been utilized in the routing algorithm for selecting the node interface and next intermediate node. When a node chooses action a, it receives a reward $R(s,a)$. The reward function is related to a utility function, which will be calculated based on the queue length of node's interface, link load and quality. Since, the quality of the links between nodes is different and dynamic, therefore in this paper, the reward for the selected node as the next hop is considered in terms of reward associated with link quality and packet delay due to congestion of the intermediate nodes (loads of links can affect the number of retransmissions and packet losses).

Link state as a metric has been used for the reward used in the model. Because quality of transmission link between nodes is one of the important criteria that influences the network efficiency and packet transmission delay. This criterion indicates the speed and probability of packets' arrival along the selected link between nodes. This measure is shown as follows

$$f(s,a)_{ETX_{lm}} = \frac{ETX_{la} - ETX_{\min}}{\arg\max_{q\in N}(ETX_{lq} - ETX_{\min})} \tag{13}$$

$$\text{subject to } ETX_{lm} = \frac{1}{d_f d_b}.$$

Expected transmission count (ETX) shows link quality and used for modeling number of tries for packet transmitting in link between node l and m. Parameters d_f and d_b indicate delivery ration forward and backward for link of node l and m, respectively. To calculate ETX,

Table 1. Network Parameters

Simulation Parameters	
BS/AP Type	WLan AP
No. of Nodes	26
Interface Link rate	1Mb/s and 2Mb/s
Channel Characteristics	Propagation model with a shadowing effect
MAC	802.11, RTS/CTS/DATA/ACK
Number of interfaces in nodes	$T = 2$
Number of BS/AP per each interface	$\forall i = 1, 2 \ \ A_i = 1$
Gateway Number	$\forall i = 1, 2 \ \ G_i = 3 - 4$
Network size	$1500 * 1500\,\mathrm{m}^2$
Packet size	512 bytes
Simulation Time	140 seconds

each node should broadcast probe packets for its neighbors and calculate link quality [22]. In MAC 802.11, there is no retransmission for broadcasting packet. Another metric is neighbors' node activity. It have been used in selecting the proper nodes for forwarding packet to the destination for each interface. It is an effective strategy in reducing the effect of link load. The cost function for link load l of each deciding node n can be obtained by

$$f(s,a)_{LI_l} = \frac{LI_l - LI_{\min}}{\max_{q\in N}(LI_q - LI_{\min})} \tag{14}$$

where, LI_l can be obtained by

$$LI_l = \Big(\frac{\sum_{t=1}^{\theta_1} a_t^1\, LI_t^1}{\theta_1} + \frac{\sum_{t=1}^{\theta_2} a_t^2\, LI_t^2}{\theta_2} + \cdots + \frac{\sum_{t=1}^{\theta_{T_n}} a_t^h\, LI_t^h}{\theta_h}\Big) \quad \text{for } l = 1, 2, \ldots, T \tag{15}$$

where T is the number of interfaces of node n and θ_h is the number of active neighbors in hop h from the interference distance from the deciding node n that collision may be occurred. Coefficient vector a_t^h shows the effect of interference in the deciding link from the $t-th$ node in the hop h from the deciding node n. LI_t expresses the amount of time the link is occupied for the transmission of packets and indicates the load of link t with bandwidth BW_t and neighbor's number NB_t which obtained as

$$LI_t = \frac{NB_t}{BW_t}\,. \tag{16}$$

Another important metric in the interface and node selection is the amount of traffic in transmitting node; in other words, the amount of delay imposed on the packet for success transmission. For this purpose, the cost of queue length in node n, due to busy channel, is considered and calculated as

$$f(s,a)_{QL_{la}} = \frac{QL_{la}}{\max_{q\in\theta_l(QL_{ql})}}\,. \tag{17}$$

In the above equation, QL_{la} denotes the queue length in node l for interface and next node from action a. θ_l is the set of all nodes neighboring node l. Total reward is related to the normalized cost functions, reversely. The total cost function can be obtained according to

$$F(s,a) = w1 * f(s,a)_{ETX_{lm}} + w2 * f(s,a)_{LI_l} + w_3 * f(s,a)_{QL_{la}} \quad s\in S,\ a\in A \tag{18}$$

subject to $w_1 + w_2 + w_3 = 1$.

4 SIMULATION RESULTS

In this section we describe a set of simulation to exemplify and evaluate the proposed multi-destination routing method in a network with multi-interface nodes. In these simulations, we have considered a network with 26 nodes. Each node has two interfaces with 1 Mbps and 2 Mbps bit rate. Node transmission range is 200^m. Constant bit rate (CBR) traffic model is considered for traffic generated by intermediate nodes. Simulation time is 140 seconds.

The size of data packet has been set to 512 bytes and the packet arrival rate has been considered variable related to each case in the simulation. The network nodes have been assumed stationary which packets are generated and routed from source to destination node and by hopping on intermediate nodes. Network parameters and simulation conditions are summarized in Table 1.

The following evaluation metrics are determined as the simulation results; Packet delivery rate which is ratio of the received packets in destination to the sent packets from source, and average end to end delay which is overall delay ratio of generated packet between source and destination.

In the first case, the proposed route selection algorithm is studied when there is only one destination. It means there is one AP for the first interface and the second one has not active AP during simulation time. Node 1 and node 20 are considered as source and destination, respectively. In this case we examine the system operation when there is only our source sending out the packet to the only existed destination. As shown in Fig. 5, both networks (the networks contain one-interface and two-interface nodes) have the same throughput in low loads. But when the traffic load increases, better performance

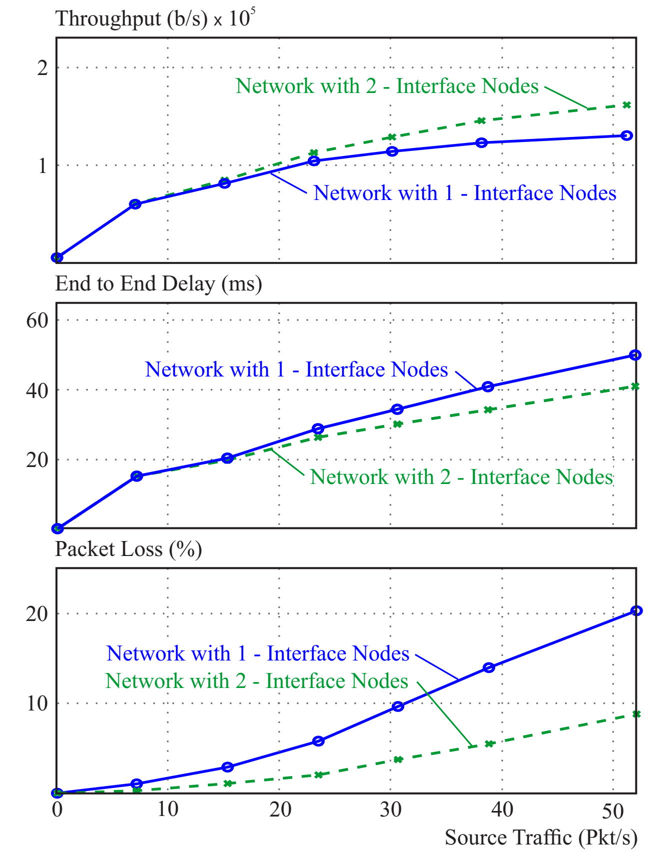

Fig. 5. Performance metrics in networks with one AP in the first interface plane (a) — throughput, (b) — end to end delay and (c) — packet loss

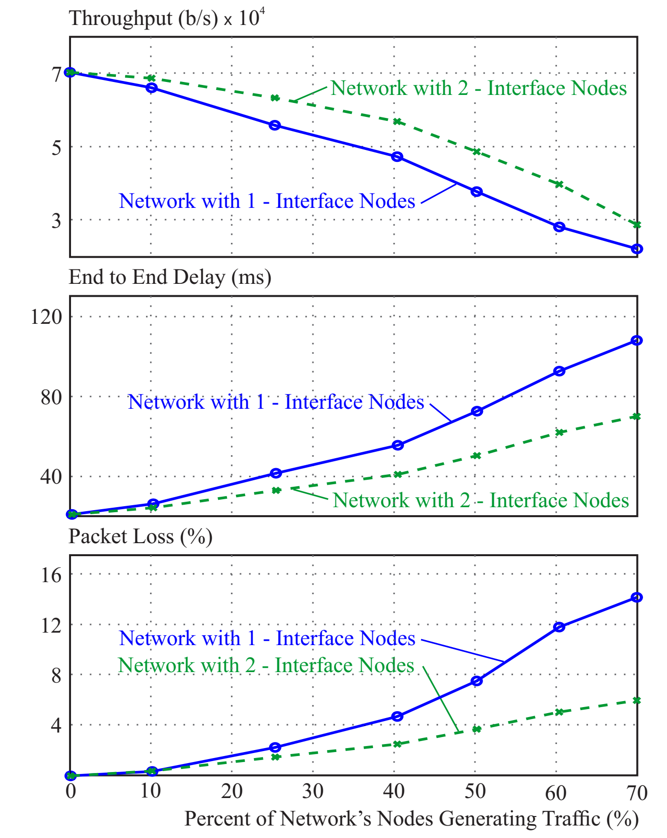

Fig. 6. Performance metrics of the network *vs* network load for different traffic flows (a) — throughput, (b) — end to end delay and (c) — packet loss

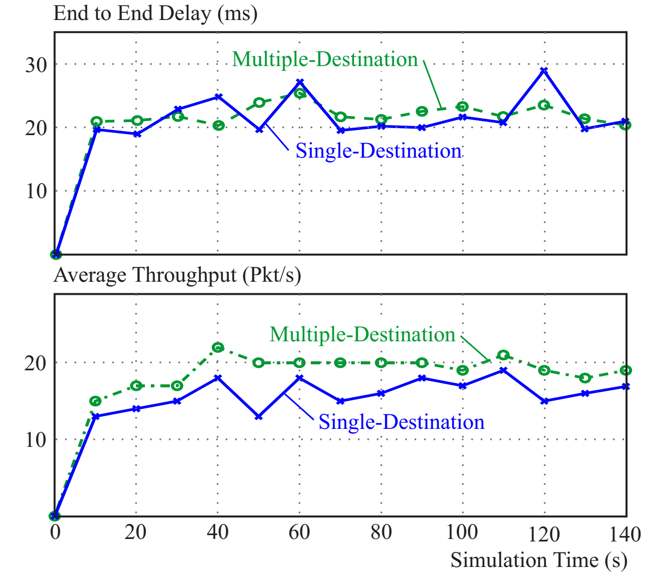

Fig. 7. Performance metrics of the network for single and multiple APs networks (a) — end to end delay (b) — average throughput

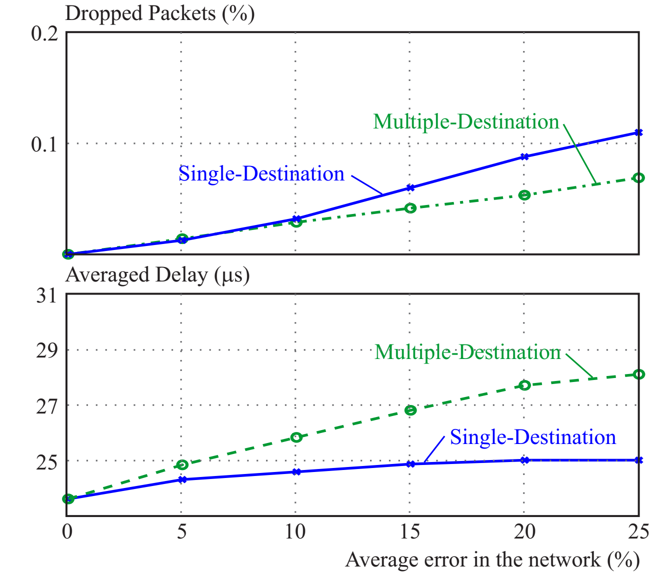

Fig. 8. The effect of link error on the network's performance (a) — dropped packets and (b) — average delay

in terms of lower packet loss and delay will be obtained for two-interface network.

When the rate of generated traffic from source is 50 Pkt/s, the throughput of the network with two interfaces will be better up to 20 %. This is due to the fact that when network nodes have only one interface, there is less resource for forwarding packets in the network with one interface.

Also, it can be seen that by increasing the load, packet loss increases slower for network with two-interface nodes.

This is due to the fact that, during routing in the network, nodes are aware of the network condition and link status by using reward function. So, they can manage their interfaces to select better ones in the routing process. When nodes have only one interface, a route for packets should be selected using that interface. In the network with multiple interface nodes, the intermediate nodes within the selected route have multiple choices to send the arrived packet on different interfaces. For example, in our case when the interface 1 of a node is overloaded, the node can switch to other interface for finding better path, by utilizing interface management.

As a result, when generating traffic in the source node is not high, utilizing one-interface will be more efficient. But, when source generates high traffic, if packet loss be important, utilizing two-interface nodes will be more efficient; this is due to the fact that packet loss is less for the network with two-interface nodes but there is not a significant difference for delay and throughput of the network with one-interface nodes and the network with two-interface nodes.

To have more practical simulation, the simulation is repeated when intermediate nodes generate traffic. We change the number of nodes who generate traffic from 0 to 70 % of total number of the network nodes, to view the effect of the interflow interference load. Each source sends packet with CBR flow that generate 10 packets of size 512 bytes per second.

As shown in Fig. 6, for low network load, there is not considerable difference between the performance metrics of networks with one-interface nodes and two-interface nodes. When the network load increases, flow interference will increase more in the network with one-interface nodes and it will be congested sooner. Throughput in the network with two-interface nodes is more up to 25 % when the percentage of intermediate nodes generating traffic is 70 %. This is due to the fact that in the network with one-interface nodes, collision will be higher specifically for higher traffic. We have utilized MDP as decision maker in each node to manage interface selection considering a reward from link load. For network load greater than 40 %, the delay and packet loss in the network with one-interface nodes are 31 % and 49 % more than those in the network with two-interface nodes. As a result, in high traffic, the network with two-interface nodes is better for flows which are sensitive to delay and packet loss.

In the next case, the proposed multi interface multi destination routing algorithm is evaluated in terms of selecting the best AP in the network. We consider two APs, each one can communicate with one of the interfaces of the nodes, with bandwidth of 11 Mb/s and 2 Mb/s called AP1 and AP2, respectively. In this case, network traffic is due to CBR traffic generated by 30 % of network nodes, selected randomly, with rate 12 packets per second. The generated packets may be routed to any AP through any interface plane of nodes.

AP1 and AP2 have been located in different positions in the network. So, from the viewpoint of source, the path to AP1 and AP2 are not the same and packets which are routed to AP1 and AP2 will experience different delays. When AP1 has been overloaded, decision maker in the intermediate node decides to forward packets to AP2 based on reward function. In this case, as shown in Fig. 7, the average delay will increase about 10 % and average throughput in the network increases about 25 %. This result can be affected by AP positions and their loads. But, in general, there is a tradeoff between old destination and new one which depends on reward function. The rate of updating reward used for decision making in the nodes would change the adaptation time.

As the last example, we will study the behaviour of the proposed algorithm when the links in the network are faulty. As described before, the link quality metric is considered in reward function to forward a unicast packet on the high quality link. To evaluate the effect of link error on the performance of the multi-destination multi-interface routing algorithm, the percent of errors in the network links will change from 0 to 25 %. As shown in Fig. 8, when packets allow to be routed to other APs, links with better quality will be selected so packet drop decreases. Since the link selection is based on different metrics, there is a trade off between different aspects in the routing process. In this case, dropped packets decreases but the average delay will be increase, slightly.

5 CONCLUSION

In this paper, we proposed a routing algorithm based on MDP for route selection when the nodes of network have multiple interfaces. Our algorithm tries to find the optimal policy according to the feedback that gathered from dynamic network environment. The adaptation behavior of the proposed algorithm helps to address issues related to dynamic network conditions. The proposed algorithm uses rewards for route selection which is a function of link load, link quality and node load as well as destination conditions. This reward function helps to employ the tradeoff between performance metrics for reaching the best performance in different situations of the network. Considering destination conditions in terms of load and delay helps to switch to better APs during packet forwarding. It is shown that interface management in the network nodes increases the network's throughput. In the high load situation, the packet loss and delay are less for network with two-interface nodes which make it suitable for cases which data flows are sensitive to packet loss and delay. The simulation results show the ability of selection of access network causes our algorithm improves throughput about 25 % but average delay will increase about 10 %. Also, by considering link quality factor, the paths with lower error will be selected and packet dropping will be decreased.

The future work can be done in considering heterogeneity in the network interface for nodes. Another area for future study is how employ multi interfaces for providing more reliability in the network.

References

[1] TZENG, S. S.—LU, H.: Qos Provisioning for Multiple Non-Real-Time Services in Cellular Wireless Networks, Wireless Communications and Mobile Computing **11** No. 10 (2011), 1392–1405.

[2] SENDREI, L.—VALISKA, J.—MARCHEVSKÝ, S.: H.264 Video Transmission over Wlan in Opnet Modeller, Journal of Electrical Engineering **64** No. 2 (2013), 112–117.

[3] BOJKOVIC, Z.—TURAN, J.—OVSENIK, L.: Towards to Multimedia Across Wireless, Journal of Electrical Engineering **56** No. 1–2 (2005), 9–14.

[4] LUSHENG, W.—KUO, G.: Mathematical Modeling for Network Selection in Heterogeneous Wireless Networks — A Tutorial, IEEE Communications Surveys & Tutorials,.

[5] YANG, K.—ZHANG, J.—CHEN, H.: A Flexible QoS-Aware Service Gateway for Heterogeneous Wireless Networks, Journal of Network **21** No. 2 (2007), 6–12.

[6] MENG, X.—TAN, K.—ZHANG, Q.: Joint Routing and Channel Assignment in Multi-Radio Wireless Mesh Networks, in Proceedings of the international Conference on Communications, 2006, pp. 3596–3601.

[7] ZHICHU, L.—SCHAAR, M.: Autonomic and Distributed Joint Routing and Power Control for Delay-Sensitive Applications in Multi-Hop Wireless Networks, IEEE Transactions on Wireless Communications **10** No. 1 (2011), 102–113.

[8] THINH, Q.—VUONG, N.—AGOULMINE, N.: Efficient Load Balancing Algorithm over Heterogeneous Wireless Packet Networks, REV Journal on Electronics and Communications **1** No. 1 (2011), 53–61.

[9] SHI, Z.—ZHU, Q.—ZHU, X.: Performance Analysis of an Improved Common Radio Resource Management Algorithm Based on Load Balance and Service Characteristics, Wireless Personal Communications **71** No. 1 (2013), 317–338.

[10] HYUKMIN, S.—SANGHOON, L.—SOO-CHANG, K.—YEON-SEUNG, S.: Soft Load Balancing over Heterogeneous Wireless Networks, IEEE Transactions on Vehicular Technology **57** No. 4 (2008), 2632–2638.

[11] FUJIWARA, S.—YAMASHITA, Y.—YAMAMOTO, M.: Performance Evaluation of Load-Balancing Gateway Selection Method in Multi-hop Wireless Networks, in Proceedings of the International Conference on Communications, Computation, Networks and Technologies, 2012, pp. 19–24.

[12] HU, W.—QIAO, C.—DE, S.—TONGUZ, O.: Integrated Cellular and Ad Hoc Relaying Systems: iCAR, IEEE Journal on Selected Areas in Communications **19** (2001), 2105–2115.

[13] PEI, X.—ZHU, G.—WANG, Q.—LIU, G.—YUAN, W.: Cooperative Load-Balancing Strategy Based on N-Hops Routing Algorithm of Ad-Hoc Network in Heterogeneous Overlapping Networks, Annals of Telecommunications **63** (2008), 597–606.

[14] RASTEGARNIA, A.—TINATI, M.—MOZAFFARI, B.—KHALILI, A.: A Localized Recursive Estimation Algorithm for Vector Parameter Estimation in Ad Hoc Wireless Sensor Networks, Journal of Electrical Engineering **61** No. 3 (2010), 171–176.

[15] ARUN, P.—RAJESH, V.—RAJEEV, T.—KSHIRASAGAR, N.: Multiple Mobile Routers Based Seamless Handover Scheme for Next Generation Heterogeneous Networks, in Proceedings of the First International Conference on Networks and Communications, NETCOM 2009, pp. 72–77.

[16] CAVALCANTI, D.—CORDEIRO, C.—KUMAR, A.—AGRAWAL, D.: A New Routing Mechanism for Integrating Cellular Networks, WLAN Hot Spots and MANETS, Proceedings of the 16th International Symposium on Personal, Indoor and Mobile Radio Communications, 2005, pp. 1414–1418.

[17] SHARIF, K.—CAO, L.—WANG, Y.—DAHLBERG, T.: A Hybrid Anycast Routing Protocol for Load Balancing in Heterogeneous Access Networks, in Proceedings of the 17th International Conference on Computer Communications and Networks,, 2008, pp. 1–6.

[18] MANOHARAN, R.—ILAVARASAN, E.—RAJARAJAN, S.: Adaptive Gateway Management in Heterogeneous Wireless Networks, Research and Reviews in Ad Hoc Networks (IJRRAN) **1** No. 2 (2011).

[19] DOWLING, J.—CURRAN, E.—CUNNINGHAM, R.—CAHILL, V.: Using Feedback in Collaborative Reinforcement Learning to Adaptively Optimize MANET Routing, IEEE Transactions on Systems, Man, and Cybernetics – Part a: Systems and Humans **35** No. 3 (2005), 360–372.

[20] KOK, J. R.—VLASSIS, N. A.: Collaborative Multiagent Reinforcement Learning by Payoff Propagation, Journal of Machine Learning Research, **7,** (2006), 1789-1828.

[21] GUESTRIN, C.—LAGOUDAKIS, M.—PARR, R.: Coordinated Reinforcement Learning, in Proceedings of the International Conference on Machine Learning (ICML), 2002, pp. 227–234.

[22] COUTO, D. S. J.—AGUAYO, D.—BICKET, J.—MORRIS, R.: A High-Throughput Path Metric for Multi-Hop Wireless Routing, in Proceedings of the Annual International Conference on Mobile Computing and Networking, 2003, pp. 419–434.

Received 28 October 2013

Amir Hosein Jafari received the BSc degree in electrical engineering in 2005 from Kashan University, and MSc degree in electrical engineering in 2008 from Shiraz University of Technology, Shiraz. He is currently pursuing the PhD degree at Iran University of Science and Technology. His research interests include data communication networking and optical communications.

Hadi Shahriar Shahhoseini received BSc degree in electrical engineering from University of Tehran, in 1990, MSc degree in electrical engineering from Azad University of Tehran in 1994, and PhD degree in electrical engineering from Iran University of Science and Technology, in 1999. He is an associate professor of the electrical engineering department in Iran University of Science and Technology. His areas of research include networking, supercomputing and reconfigurable computing. More than 150 papers have been published from his research works in scientific journals and conference proceedings. He is an executive committee member of IEEE TCSC and serves IEEE TCSC as regional coordinator in Middle-East countries.