Layered Sequential Decision Policies for Cross-Layer Design of Wireless Ad-Hoc Networks

Zhenzhen Ye and Alhussein A. Abouzeid

***Abstract*—Efficient transmission control (e.g., power/rate control) in the physical layer, link scheduling in the link layer and routing in the network layer are critical design issues of wireless ad hoc networks. By considering both the quality-of-service and the utilization of resources under the temporally correlated uncertainties of the network conditions, a sequential decision framework is proposed for protocol design and layering. By observing that there are usually different performance concerns in the network layer versus lower layers, two correlated sequential decision models for the operations at different layers are proposed. With the decision model in the link and physical layers, scheduling and transmission control are jointly optimized. By adding a decision model in the network layer, the benefit of cross-layer information is quantified and the optimal routing/forwarding strategy is characterized. Practical algorithms based on the proposed decision models are also developed to achieve optimal/near-optimal performance.**

I. Introduction

The concept of wireless ad hoc network (WANET) has gained popularity in network research and engineering, due to its flexibility in supporting a variety of new applications in military and civilian settings. In such a multihop wireless transmission environment, regardless of the application, efficient designs of transmission control (e.g., power/rate control) in the physical (PHY) layer, link scheduling in the link layer and routing in the network layer are always critical. The design should be concerned with both the Quality-of-Service (QoS) of information transport in the network (e.g., delay) and the utilization of resources (e.g., energy consumption) under network dynamics (e.g., dynamics on wireless link qualities, multi-access channel availability, buffer occupation status). Since the network dynamics are usually *temporally correlated* and also interacting with network operations [1], [2], *the optimization of network operations should be carried out not only across layers but also over time*. This fact motivates us to consider a *sequential* decision framework in optimizing operations in various layers. Furthermore, one should note that there are usually different performance concerns in the network layer versus lower layers in WANETs. In the network layer, routing is concerned with the end-to-end performance of a *specific* session. In contrast, in the lower layers, since a node in the network not only might be a source or destination of a communication session but also usually needs to act as a wireless *relay* (i.e., router) for other sessions and different sessions can have different end-to-end performance concerns, the optimal scheduling and transmission strategies should take into consideration the performance of packets from *all* sessions going through it, not just for one specific session. Also the time-scale of operations/dynamics in the lower layers can be much smaller than that in the network layer. Such natural separation between routing and operations in lower layers motivates us to consider different but correlated sequential decision models for the operations at different layers. Figure 1 illustrates an architecture of joint design of transmission control, link scheduling and routing. The routing problem for a specific session in the network layer is formulated with a decision model (i.e., MDP_R model) which makes optimal routing/forwarding decisions with the cross-layer information from lower layers. In lower layers, the problem of transmission control and link scheduling at a node is formulated with another decision model (i.e., MDP_{TC-LS} model), which jointly optimizes both operations when the forwarding decisions of sessions going through the node are given.

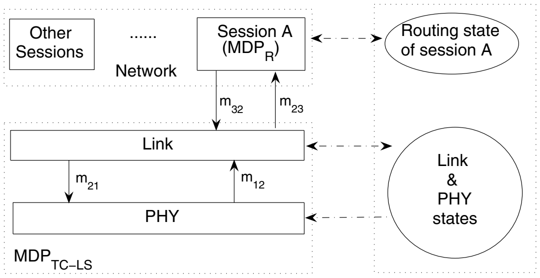

Fig. 1. An architecture for cross-layer control of operations in the network, link and PHY layers.

Extensive research effort has been carried out in cross-layer design in recent years. Regarding the *multihop* transmission scenario, almost all existing works follow the well-known network utility maximization (NUM) framework (see [3] and the references therein). Our work here is different from the NUM type framework in two major aspects. First, we explicitly consider the *temporal correlation* of network conditions and the *interaction* between network operations and network conditions which are hard to model with the NUM type framework. Second, the proposed sequential decision framework can handle QoS metrics (i.e., utilities) other than average source rate, for example, end-to-end information latency, which are hard to be included in the NUM type framework. In the network layer, when the cross-layer information such as instantaneous performance of links is conveyed from lower layers, the conventional *deterministic* shortest path routing is not optimal in general. Compared to the existing works on route design for stochastic networks (e.g., [4], [5], [6]), instead of formulating the routing problem as the route design in a connected graph with random link costs, we include the cross-layer information such as the instantaneous performance of links at any node as a

Z. Ye is with R&D Division at iBasis, Inc., Burlington, MA 01803; A. A. Abouzeid is with the Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180-3590, USA (email: zhenzhen.ye@ieee.org, abouzeid@ecse.rpi.edu).

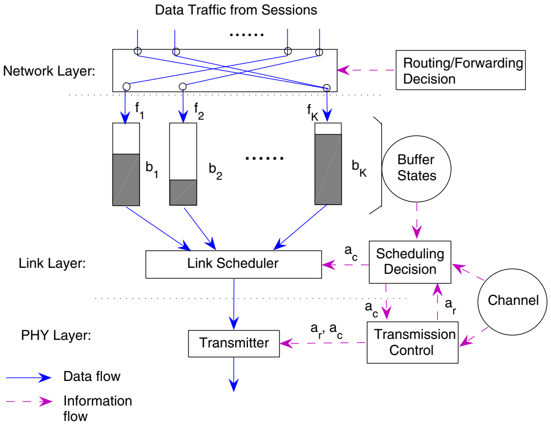

Fig. 2. The system model of scheduling and transmission control at the link and PHY layers with given forwarding decisions at the network layer; a_c is the scheduling action specified by the link layer and a_r is the transmission control action given by the PHY layer.

part of the nodal state in the decision model. This formulation accurately characterizes the optimal routing/forwarding strategy in a dynamic network and meanwhile has a *controllable* complexity to allow designing practical routing algorithms. A similar application of stochastic decision model for adaptive opportunistic routing has been presented in [7].

II. Joint Scheduling and Transmission Control

The information conveyed by the network layer to lower layers (i.e, the link and PHY layers) is the forwarding decisions of all sessions going through a node, where the node works as a relay of the traffic of these sessions. Given this information, the objective of joint scheduling and transmission control at the node is naturally to optimize the nodal performance averaged over *all* packets forwarded by the node.

A. System Model

Consider an arbitrary node in the network and assume that there are K wireless links associated with the node, where $K \geq 1$. The data traffic from all sessions, except the one destined to the node, will be sent out via these links. For each link, there is a data buffer with a finite size B, in the unit of packet, where $B > 0$. The buffer state of a link i is denoted as b_i, which is in a set $\mathcal{B} \triangleq \{0, 1, ..., B\}$. The vector $b \triangleq (b_1, ..., b_K)$ is the joint buffer states of all links at the node. As each link associates with a wireless channel, the channel state (i.e., quality) of a link i is denoted as h_i, which is in a finite set $\mathcal{H} \triangleq \{\xi_1, ..., \xi_H\}$, where $H > 0$. The vector $h \triangleq (h_1, ..., h_K)$ is the joint channel states of all links at the node. The (random) channel states of different wireless links are independent and the evolution of the channel states of a link is assumed to follow a finite-state Markov chain (FSMC) model [1].

Time is slotted. If a packet is to be transmitted at a slot, the transmission starts at the beginning of the slot. A random access MAC model is assumed to be used in the network. At the considered node, the interval between two available transmission slots is a random variable $z \in \{1, 2, ...\}$, in the unit of slot. The probability that the interval is m slots is $P_z(m) \triangleq P(z = m)$. The random interval z is under the control of random access MAC and assumed to be independent of the buffer and channel states as well as the scheduling and transmission actions at the current transmission slot. Let $\bar{F}_z(m) \triangleq \sum_{j=m+1}^{\infty} P_z(j)$. The transceiver of the node works in a half-duplex mode and a slot that channel is not available for transmission is called the *non-transmission slot*.

When the forwarding decisions of sessions going through the node are fixed, the data traffic from a session will be sent into the buffer of the link specified by the forwarding decision of the session. As illustrated in Fig. 2, the overall traffic into the buffer of a link can be composed of the traffic from several different sessions and also the traffic of any specific session can only go into one buffer (i.e., link) once the forwarding decision of the session is fixed. Let the traffic flow into a buffer i in a non-transmission slot be $f_i, i = 1, ..., K$, which arrives at the buffer at the end of the time slot. By noting that the node is half-duplex, the traffic flows into the node's buffers in a transmission slot are zero. As both the transmission rate of a wireless link and the number of links associated with a node are finite, it is natural to set $f_i, \forall i$, to be in a finite set $\mathcal{F} \triangleq \{0, 1, , ..., F\}$, in the unit of packet, where $F > 0$. Furthermore, we assume that the (random) vector $f = (f_1, ..., f_K)$ is independently identically distributed (i.i.d.) in non-transmission slots. This assumption approximately holds under a random access MAC control, where the temporal dependence of the traffic across slots is greatly reduced as different neighboring nodes might access the multi-access channel for transmission in consecutive slots and their transmission rates are independent of each other.

For the nodal performance metric in the optimization, we consider a weighted sum of both energy related cost and buffer occupation related cost at the node, where the positive weight on the energy related cost is λ, given that the weight on the buffer occupation related cost is normalized to be one. One example energy related cost is the energy per data unit (e.g., per packet) which characterizes how much energy of the node spends for a successful forwarding of a data unit. The example metrics of the buffer occupation related cost include the expected delay of a packet experienced at the node as well as the packet loss rate due to buffer overflows. The weighted sum type of cost naturally characterizes the concern on both QoS of information transport and resource utilization.

B. MDP_{TC-LS}: A Markov Decision Process Model for Joint Scheduling and Transmission Control

Under the given setting, the problem of joint scheduling and transmission control can be formulated with a Markov decision process (MDP) model, we call it MDP_{TC-LS} model. An MDP model is composed of a 4-tuple $\{S, A, P(\cdot|s, a), r(s, a)\}$, where S is the state space, A is the action set, $P(\cdot|s, a)$ is a set of state- and action-dependent state transition probabilities and $r(s, a)$ is a set of state- and action- dependent instant costs [8]. We define the components of the MDP_{TC-LS} model as follows.

1) State: $s = (b, h)$, i.e., the observed states of buffers b and wireless channels h. By defining $s_i = (b_i, h_i)$ as the state associated with link i, we also have $s = (s_1, ..., s_K)$.

2) Action: The action includes both scheduling action a_c and transmission control action a_r. The scheduling action $a_c \in \{0, 1, ..., K\}$, where $a_c = 0$ represents that no link is scheduled for transmission in a slot and $a_c = i\ (\neq 0)$ represents that link i is scheduled. Transmission control action $a_r = (a_{r,1}, ..., a_{r,K})$

is defined as the transmission rate vector in a slot, where $a_{r,i}$ is the transmission rate of link i, constrained by the channel state of the link. As there is no transmission in a non-transmission slot, *an action is only carried out in a transmission slot.*

3) Costs: The nonnegative energy related cost depends on the channel state of the scheduled link and the assigned transmission rate. When no link is scheduled, the energy related cost is zero. Let $c_p(s,a)$ be the energy related cost, we have $c_p(s,a) = c_p(h_i, a_{r,i}) \geq 0$ if $a_c = i \neq 0$ and $c_p(s,a) = 0$ if $a_c = 0$. The nonnegative buffer occupation related cost only depends on the buffer states of links. Specifically, let $c_{b,i}(b_i)$ be a nonnegative buffer occupation related cost of link i. The overall buffer occupation related cost given the buffer state $b = (b_1, ..., b_K)$ is $c_b(b) \triangleq \sum_{i=1}^{K} \beta_i c_{b,i}(b_i)$, where β_i is a positive weight constant of link i. Let $c_t(s,a)$ be the overall cost in a transmission slot. We have $c_t(s,a) = \lambda c_p(s,a) + c_b(b)$, where $s = (b,h)$. In a non-transmission slot, as there is no transmission and scheduling action, let $c_{nt}(s)$ be the overall cost in a non-transmission slot, then $c_{nt}(s) = c_b(b)$.

4) State Transitions: As actions are only carried out in transmission slots, we are concerned with the state transitions in consecutive transmission slots. Since the interval between two consecutive transmission slots is random due to the random access MAC control, a complete description of state transition should include this randomness. Given the state-action pair (s,a) at the current transmission slot, we define the probability that the next available transmission slot is m (≥ 1) slots later and the corresponding state is s' as $P(s', m|s,a)$. Specifically,

$$P(s', m|s,a) = P(s^{(m)} = s'|s^{(0)} = s, a)P_z(m), \qquad (1)$$

where $s^{(0)}$ $(s^{(m)})$ is the state at current (next) transmission slot.

5) Objective and Optimality Equations: To jointly optimize scheduling and transmission control performance at the node, we consider the expected total discounted cost criterion [8], i.e., minimize the expected overall costs in transmission control and scheduling in an infinite horizon with a discount $\gamma \in (0,1)$ on future costs. One reason that we adopt this optimization criterion is that it is natural to put more weight or concern on the cost of the current slot in the overall cost by discounting since the forwarding decisions are likely to be changed in the future. By defining a *decision rule* ϱ_n as the collection of actions at all states at n-th transmission slot, the optimization targets to find a *policy* $\pi = \{\varrho_n\}, n = 0, 1, ...$, to minimize the expected total discounted cost.

Consider a stationary, deterministic policy $\pi = \{\varrho, \varrho, ...\}$ and any initial state $s = (b,h)$ at the current transmission slot, with some algebraic manipulation, the achievable cost of the policy is given by [9]

$$v^{\pi}(s) = r(s, \varrho(s)) + \sum_{s'} q_{ss'}^{\varrho(s)}(\gamma) v^{\pi}(s'), \quad \forall s, \qquad (2)$$

where $q_{ss'}^{a}(\gamma) \triangleq \sum_{m=1}^{\infty} \gamma^m P_z(m) P(s^{(m)} = s'|s^{(0)} = s, a)$ and

$$r(s,a) \triangleq \sum_{s''} \left[\sum_{i=1}^{\infty} \gamma^i \bar{F}_z(i) P(s^{(i)} = s''|s^{(0)} = s, a) \right] \times c_{nt}(s'') + c_t(s,a). \qquad (3)$$

We note that $\sum_{s'} q_{ss'}^{a}(\gamma) = \sum_{m=1}^{\infty} \gamma^m P_z(m) \leq \gamma < 1$ and $r(s,a)$ is bounded since $r(s,a)$ is finite for any state-action pair (s,a) and the state-action space is finite. Therefore, from the analogy of Prop. $1.2.1 - 1.2.3$ in [10], it is straightforward to show that the optimal cost vector v is the *unique* solution of optimality equations:

$$v(s) = \min_{a} \left\{ r(s,a) + \sum_{s'} q_{ss'}^{a}(\gamma) v(s') \right\}, \quad \forall s; \qquad (4)$$

Furthermore, there exists an optimal stationary deterministic policy $\pi = \{\varrho, \varrho, ...\}$, where $\varrho(s)$ is the action minimizing the righthand-side (RHS) of (4). Thus, the optimal strategy (i.e., π) on joint link scheduling and transmission control at a node can be obtained by solving (4). In addition, the MDP_{TC-LS} model satisfies the following property [9].

Theorem 2.1: For any state $s = (b,h)$, the optimal cost $v(s)$ is nondecreasing with the current buffer occupation state b_i of any link i $(i \in \{1, ..., K\})$.

We further consider a case that the energy related cost $c_p(s,a)$ is *nonincreasing* with the transmission rate $a_{r,i}$ of the scheduled link i. For example, when the transmission power of a node is constant and the energy cost per data unit (e.g., bit, byte or packet) is concerned, a higher transmission rate indicates a lower energy related cost. Let

$$Q(s,a) \triangleq r(s,a) + \sum_{s'} q_{ss'}^{a}(\gamma) v(s'), \forall (s,a).$$

In this case, Theorem 2.1 implies that

Corollary 2.2: For any state s and action $a = (a_c, a_r)$ where $a_c = i, i \in \{1, ..., K\}$, $Q(s,a)$ is nonincreasing with the rate assignment action $a_{r,i}$.

With Corollary 2.2 and note that

$$v(s) = \min_{a_c} \left\{ \min_{a_r} \{Q(s,a)\} \right\}, \qquad (5)$$

we obtain the optimal rate assignment strategy as follows.

Proposition 2.3 (Transmission Control): For a given scheduled link i $(\neq 0)$ at current (transmission) slot, if the energy related cost is nonincreasing with the transmission rate $a_{r,i}$, a greedy assignment of the transmission rate with respect to the current channel condition h_i is optimal.

This proposition implies a natural separation of transmission control and scheduling decisions where, as long as the scheduling decision in the link layer is given, the transmission control decisions in the PHY layer only depends on the channel state of scheduled link which is observable in the PHY layer.

C. Index-based Scheduling: the Rule and Algorithm

Once the optimal transmission control strategy is specified, the remaining problem is link scheduling. From (5), with the given transmission rates of links, it is essentially a joint channel-aware and buffer-aware scheduling problem [11], [12]. To make the solution practically useful, we here develop a low-complexity suboptimal scheduling rule.

We first consider a link i. The part of the state associated with this link is (b_i, h_i), denoted as s_i, and the given transmission rate is $a_{r,i}$. As for scheduling, the part of the action at the link is either "scheduled" or "not scheduled", denoted as e_i. The part of the cost associated with the link is denoted as $r_i(s_i, e_i; a_{r,i})$, where $r(s, a) = \sum_{i=1}^{K} r_i(s_i, e_i; a_{r,i})$. For the simplicity of notation, we drop $a_{r,i}$ in the following analysis, since it has been specified for each state. Thus, if the scheduling decision at a link i can be independently made without considering the decisions at other links, an MDP model can be formulated specifically for a link to optimize the expected total discounted cost at the link. We call it $MDP_{l,i}$ model. Let the optimal expected total discounted cost at link i be v_i, and let

$$Q_i(s_i, e_i) \triangleq r_i(s_i, e_i) + \sum_{s_i'} q^{e_i}_{s_i s_i'}(\gamma) v_i(s_i'). \tag{6}$$

where $q^{e_i}_{s_i s_i'}(\gamma) \triangleq \sum_{m=1}^{\infty} \gamma^m P_z(m) P(s_i^{(m)} = s_i' | s_i^{(0)} = s_i, e_i)$.

As there are K links associated with the node and the optimal cost $v(s)$ is the overall cost of all links, we approximate the optimal cost as

$$v(s) \approx \sum_{i=1}^{K} v_i(s_i). \tag{7}$$

Furthermore, let

$$\chi_i(s_i) \triangleq Q_i(s_i, e_i = 1) - Q_i(s_i, e_i = 0), i = 1, ..., K. \tag{8}$$

From (5), (6), (7) and (8), we can show that, if the link i is scheduled, i.e., $a_c = i, i \in \{1, ..., K\}$,

$$r(s, a_c) + \sum_{s'} q^{a_c}_{ss'}(\gamma) v(s') \approx \chi_i(s_i) + \sum_{j=1}^{K} Q_j(s_j, e_j = 0), \tag{9}$$

otherwise, if no link is scheduled, i.e., $a_c = 0$,

$$r(s, a_c) + \sum_{s'} q^{a_c}_{ss'}(\gamma) v(s') \approx \sum_{j=1}^{K} Q_j(s_j, e_j = 0). \tag{10}$$

Thus the minimization over a_c in (5) is approximated by comparing the values in (9) and (10). The approximation of the optimal scheduling rule is given by

$$\tilde{\varrho}_c(s) = \begin{cases} \arg\min_i \{\chi_i(s_i)\}, & \min_i \{\chi_i(s_i)\} < 0, \\ 0, & \text{otherwise} \end{cases}, \tag{11}$$

and the corresponding stationary scheduling policy is $\tilde{\pi}_c \triangleq \{\tilde{\varrho}_c, \tilde{\varrho}_c, ...\}$. From (8), we observe that $\chi_i(s_i)$ essentially characterizes the cost reduction of scheduling link i for transmission. Intuitively, $\chi_i(s_i)$ might be seen as a performance *index* of scheduling the link i at state s_i and the decision rule (11) is thus an index-based scheduling rule.

Proposition 2.4 (Index-based Scheduling): In a transmission slot with a state s, if there is a cost reduction by scheduling a link (i.e., at least one *negative* index), schedule the link which achieves the largest cost reduction (i.e., the smallest index); otherwise no link is scheduled for transmission.

The key to implement the index-based scheduling rule is to obtain the performance indices of links. In turn, with (8), it is to find Q-values of all links. This requires to solve K parallel MDP models (i.e., $\{MDP_{l,i}\}$), where each MDP model is for an individual link. Note that the state space size of the MDP model for a link is only $(B+1)H$, which is much smaller than $(B+1)^K H^K$, the state space size of the original scheduling problem, the complexity of solving these K MDP models is thus greatly reduced than directly solving (5).

TABLE I
PARALLEL LEAST-SQUARES POLICY ITERATION (P-LSPI) ALGORITHM

```
Select basis function sets $\{\phi_i(s_i, e_i)\}, i = 1, ..., K$, and probability
sequence $\{p_1, p_2, ...\}$ for exploration; set stopping criterion $\epsilon$;
Initialize weight vector sets $\{w_i^{(0)}\}$ and sample sets $\{\mathcal{D}_i^{(0)}\}$;
$n = 1$;
Repeat
  Policy evaluation at links: $w_i^{(n)} = \text{LSPI}_i(\mathcal{D}_i^{(n-1)}, w_i^{(n-1)})$
  Scheduling: at current state $s = (s_1, ..., s_K)$, calculate indices
    of links $\chi_j(s_j) = [\phi_j(s_j, e_j = 1) - \phi_j(s_j, e_j = 0)]^T w_j^{(n)}$;
    scheduling according to the index-based rule (11) w.p. $p_n$;
    otherwise, w.p. $1 - p_n$, a randomly selected link is scheduled;
  Update sample sets with actual state transitions:
  $\mathcal{D}_i^{(n)} = \mathcal{D}_i^{(n-1)} \cup (s_i, e_i, s_i', z, \{c_{b,i}(b_i^{(j)})\}_{j=1}^{z}), i = 1, ..., K$
     where $e_i = 1$ if the link is scheduled, otherwise 0;
Until $\sum_i ||w_i^{(n)} - w_i^{(n-1)}|| < \epsilon$
Return $\{w_i^{(n)}\}, i = 1, ..., K$
```

In practice, the Q-values are usually solved via reinforcement learning approaches. Here, for the $MDP_{l,i}$ model of any link i, we develop an algorithm which is adapted from the well-known least-squares policy iteration (LSPI) algorithm [13], by extending its applicable scenario to the MDP model with a random state transition interval. Since the adapted LSPI is carried out in parallel at all K links, the algorithm is called Parallel LSPI (P-LSPI), which is shown in Table I. We evaluate the index-based scheduling rule and the P-LSPI algorithm at a node which has $K = 2$ neighboring nodes (i.e., two links). The transmission power of the node is 1mW and the length of a time slot is $T_s = $ 1ms. The length of a data packet is 1000 bits and the target packet loss rate is 10^{-2}. An adaptive uncoded M-PSK modulation is used in transmission control, where M is up to 16. The concerned energy related cost in the simulation is set as the transmission energy cost per packet. The buffer occupation related cost is the sum of the expected packet loss due to a buffer overflow and a linear cost associated with the buffer occupation. The second row of Table II shows the relative difference in costs (in percentage, with 95% confidence interval) between the index-based scheduling rule obtained from (11) and the optimal scheduling rule, under different scenarios where (d_1, d_2) represents the distances (in meter) of two neighboring nodes. We see that the costs achieved as well as the scheduling actions adopted by the index-based scheduling rule are very close to the optimal ones at *all* states. Meanwhile, the proposed P-LSPI algorithm can also achieve a near-optimal performance (as shown in the last row of Table II).

III. ROUTING

We now consider routing between an arbitrary source-destination pair in a WANET with $N+1$ nodes. The set of node indices is $\{1, 2, ..., N, d\}$ where node d is the destination node and node 1 is the source node. The network is connected. A forwarding action of a data packet over a wireless link between two nodes, say from node i to node j, incurs a *positive* forwarding cost $c_{(ij)}(x_{(ij)})$ where $x_{(ij)}$ is defined as the observed state of the link from node i to node j. *The link state is an abstraction of the states observed at lower layers*, for example, the status of the associated buffer occupation, queuing delay and/or supported data transmission rate of the link, depending

TABLE II
PERFORMANCE OF INDEX-BASED SCHEDULING RULE AND P-LSPI ALGORITHM, COMPARED TO OPTIMAL SCHEDULING RULE

Setting: (d_1, d_2) (meter)	$(3.0, 7.0)$	$(10.0, 10.0)$	$(4.0, 8.0)$	$(6.0, 10.0)$	$(2.0, 9.0)$
Cost Difference (%)	0.3330 ± 0.0124	0.0315 ± 0.0058	0.2546 ± 0.0107	0.0147 ± 0.0029	0.3260 ± 0.0099
Action differences (%)	6.6406	2.5391	3.5156	0.9766	0.7813
Parallel LSPI (%)	0.7866 ± 0.0298	0.1700 ± 0.0089	0.9806 ± 0.0397	0.6078 ± 0.0295	0.7601 ± 0.0368

on which information conveyed by lower layers to the network layer. As the state space observed in lower layers (i.e., the state space of MDP_{TC-LS}) is finite, $x_{(ij)}$ is also in a finite set. The forwarding cost could be any performance concern of routing, such as delay, energy, throughput or any combination of single performance metrics, as long as the end-to-end routing performance metric can be transformed/decomposed into the sum of link performance metrics of the route. We target to find routes with the minimum *expected* total cost from the source node to the destination node.

A. MDP_R: A Markov Decision Process Model for Routing

This optimal routing problem can be modeled with an MDP model, denoted as MDP_R. Its components are as follows.

1) State: $s = [i, x] \in S$, where i is the node index and x is a (vector) variable to represent the states of links associated with the node and we call it the *nodal state* of the node. For example, suppose node i has K neighboring nodes (i.e., K wireless links) and let x_k be the link state of link $k, k = 1, ..., K$, then $x = (x_1, ..., x_K)$, representing the states of all links observed at node i. Obviously the state space S is finite. For the destination node d, we set its nodal state $x \equiv 0$ and define an *absorbing state* as $s_d = [d, 0]$.

2) Action: the forwarding action at a node $i (\neq d)$ is defined as $a = j \in A_i$, where A_i is the set of neighboring node indices of node i. As no forwarding action needed at the destination node d, A_d is an empty set.

3) State Transitions: the state transition probability under a state $s = [i, x], i \neq d$ and a forwarding action a is given by $P(s'|s, a) = P([j, y]|[i, x], a)$, where $s' = [j, y]$. It can be further shown that $P(s'|s, a) = P(y|j)$ if $a = j \in A_i$ otherwise zero. Specially, the destination node d is always in the absorbing state s_d.

4) Cost: the forwarding cost at a state $s = [i, x], i \neq d$ with a forwarding action a is given by $c(s, a) \triangleq c_{(ia)}(x_{(ia)})$, where $x_{(ia)}$ is a component of x and $c(s, a) > 0$ for $s \neq s_d$. For the destination node d, as there is no further forwarding, $c(s_d, a) \equiv 0, \forall a \in A_d$.

Then the objective of the routing problem becomes to find a policy to minimize the expected total cost from the source node to the destination node of the session. It can be shown that, for any state $s = [i, x] \neq s_d$, the optimal expected routing cost $v(s)$ satisfies the optimality equations

$$v(s) = \min_{a \in A_i} \left\{ c(s, a) + \sum_{s'} P(s'|s, a) v(s') \right\}. \qquad (12)$$

TABLE III
Q-VALUE BASED DISTRIBUTED STOCHASTIC ROUTING ALGORITHM (Q-DSR) AT NODE i

```
The set of routing/forwarding instants: $\mathcal{T}_i$;
Initialize $Q_i^{(0)}$ and counters $\{N_k(x_k), N_k\}, \forall x_k, k = 1, ..., K$
to zeros; Set initial learning rate $\alpha$, $\tau$ and probability sequence
$\{p_0, p_1, ...\}$ for exploration;
Repeat {
    At each $t_n \in \mathcal{T}_i$, given the state at node $i$ is $x = (x_1, ..., x_K)$;
    Routing/forwarding action:
      w.p. $p_n$, next-hop node is $j = \arg\min_k Q_i^{(n)}(x_k, k)$,
      otherwise, w.p. $1 - p_n$, the next-hop node $j$ is randomly
      selected from neighboring nodes of $i$;
    Communicate with node $j$ for message $\varpi_j$;
    Q-value update:
    $Q_i^{(n+1)}(x_j, j) = Q_i^{(n)}(x_j, j) + \frac{\alpha\tau}{\tau + N_j(x_j) + 1} \left[ c^{(n)}([i, x], j) \right.$
                    $\left. + \varpi_j - Q_i^{(n)}(x_j, j) \right]$,
    $Q_i^{(n+1)}(x_j, j') = Q_i^{(n)}(x_{j'}, j'), j' \neq j$;
    Message update:
    $\varpi_i \leftarrow \frac{1}{\prod_k (N_k + 1)} \left[ \varpi_i \left( \prod_k N_k \right) + \prod_k \left( N_k(x_k) + 1 \right) \right.$
        $\left. \times \min_l Q_i^{(n+1)}(x_l, l) - \prod_k N_k(x_k) \min_l Q_i^{(n)}(x_l, l) \right]$;
    Update counters: for $k = 1, ..., K$
    $N_k(x_k) \leftarrow N_k(x_k) + 1$,
    $N_k \leftarrow N_k + 1$; }
```

In a connected network, since there always exists a route from any node to the destination node and the forwarding cost $c(s, a) > 0$ for $s \neq s_d$, it can be shown that [9]: the stationary policy $\pi = \{\varrho, \varrho, ...\}$, where $\varrho(s)$ is the action minimizing the RHS of (12), is optimal and with this policy, a packet in the network has a *positive* probability to reach the destination node after at most $|S| - 1$ forwarding steps, where $|S|$ is the dimension of the state space.

B. Implications of the Nodal State in Cross-layer Design

The conventional deterministic shortest path (DSP) routing can be seen as an extreme case of the proposed MDP formulation where any state $s = [i, x]$ is degenerated to i (i.e., the node index only). Under this degeneration, the state transition of MDP_R becomes deterministic. The optimality equations of this degenerated MDP_R model are

$$u_{DSP}(i) = \min_{j \in A_i} \{ \bar{c}(i, j) + u_{DSP}(j) \}, \forall i \neq d, \qquad (13)$$

where $\bar{c}(i, a) \triangleq \mathbb{E}_x \{ c([i, x], a) \}$. That is, the (deterministic) dynamic programming formulation of conventional DSP problem. We can naturally interpret the degeneration of a nodal state as the loss of information at nodes for routing/forwarding decisions. In the other words, accurate nodal performance metrics (i.e., $c([i, x], a)$) are unobservable at the network layer. In a cross-layer design, with the help of information exchange between the network layer and lower layers, a more accurate nodal performance model is available and the routing performance can thus be improved.

C. An On-line Distributed Routing Algorithm

To develop practically useful routing algorithms from the proposed MDP_R model, we first note that, in the optimality

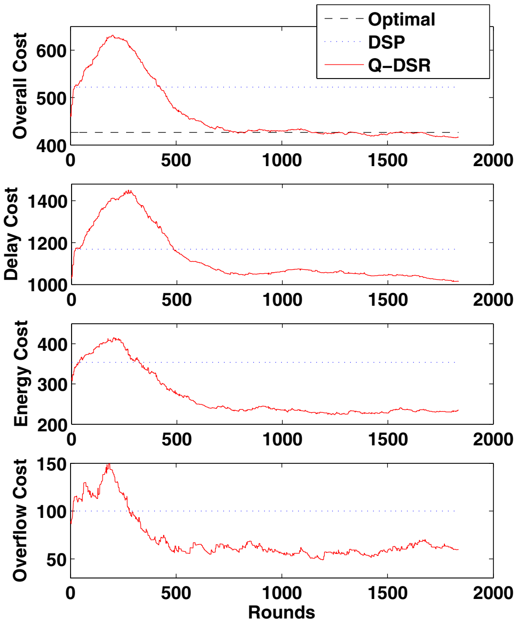

Fig. 3. The end-to-end expected overall routing cost and individual metrics (i.e., delay, energy and overflow costs) achieved by Q-DSR routing algorithm with quantized cross-layer information, compared to the conventional DSP algorithm, in a WANET with 25 nodes in a 40m×30m area.

equations (12), for any state $s = [i, x], i \neq d$, forwarding action $a = j \in A_i$ and the next state $s' = [j, y]$, $\sum_{s'} P(s'|s, a)v(s') = \sum_y P(y|j)v([j, y])$, which only depends on node index j. Thus it can be treated as a "message" from node j, denoted as ϖ_j. Furthermore, by observing that the forwarding cost $c(s, a) = c([i, x], a) = c_{(ia)}(x_{(ia)})$, only depending on the state of the link (i, a) (i.e., $x_{(ia)}$), we can define Q-values: $Q_i(x_{(ij)}, j) \triangleq c([i, x], j) + \varpi_j$, where $x_{(ij)}$ is a component of x. Note that $v([i, x]) = \min_j \{Q_i(x_{(ij)}, j)\}$, the optimal routing costs can be obtained by solving Q-values. As the computation of Q-values can be easily implemented in a distributed fashion where neighboring nodes exchange messages (i.e., $\{\varpi_i\}$) and use received messages and (random) forwarding costs to update the corresponding Q-values, we develop an online distributed routing algorithm which essentially learns Q-values from actual routing/forwarding actions and meanwhile improves routing/forwarding decisions with the updated Q-values. The part of the proposed Q-value based distributed stochastic routing (Q-DSR) at an arbitrary node $i(\neq d)$ is shown in Table III. The instants of Q-value update at node i are specified by a set $\mathcal{T}_i$ (line 1), where update instants can be asynchronous among nodes.

We evaluate Q-DSR routing algorithm in a two-dimensional network where 25 nodes are randomly uniformly deployed in a 40m×30m plane area. The transmission range of a node is 10m. A link state information conveyed from lower layers consists of the abstractions of the buffer state (e.g., "light" or "heavy") and the wireless channel state of the link (e.g., "poor" or "good"). We consider a routing performance metric consisting of delay, energy as well as the packet-loss due to buffer overflows. The top plot in Fig. 3 illustrates the end-to-end performance of Q-DSR, which achieves the analytical optimal performance after about 700 learning rounds. Here a learning round can be interpreted as the forwarding operation of one or more packets. The remaining plots in Fig. 3 further demonstrate the performance improvements of the proposed Q-DSR algorithm compared to conventional DSP algorithm, in terms of end-to-end delay cost, energy cost and packet loss cost due to buffer overflows.

IV. Conclusion

We have developed a layered sequential decision framework for the design of transmission control, link scheduling and routing in a multihop wireless ad hoc network that takes into account the temporal correlations of network conditions. The optimal scheduling and transmission control operations at a node are achieved by a joint design in general. In some cases (e.g., the transmission power of a node is constant), the optimal transmission control is uniquely determined by the channel conditions of wireless links. For scheduling, an index-based rule can achieve near-optimal scheduling performance and it has been implemented with a practical online algorithm (i.e., P-LSPI). In the network layer, the proposed decision model has allowed us to characterize the benefit of the cross-layer information in routing, as well as provided a theoretical foundation on designing efficient routing algorithms. The proposed routing algorithm Q-DSR, based on the routing decision model, has been demonstrated to be able to achieve a significant performance improvement over the conventional deterministic shortest path routing.

References

[1] Q. Zhang and S. A. Kassam. Finite-state markov model for rayleigh fading channels. *IEEE Transactions on Communications*, 47(11):1688–1692, November 1999.

[2] T. Holliday and A. Goldsmith. Optimal power control and source-channel coding for delay constrained traffic over wireless channels. In *ICC '02: Proceedings of IEEE International Conference on Communications*, pages 831–835, New York, NY, USA, May 2002.

[3] M. Chiang, S. H. Low, R. A. Calderbank, and J. C. Doyle. Layering as optimization decomposition. *Proceedings of IEEE*, 95(1):255–312, January 2007.

[4] R. P. Loui. Optimal paths in graphs with stochastic or multidimensional weights. *Communications of the ACM*, 26(9):670–676, September 1983.

[5] R. A. Guérin and A. Orda. Qos routing in networks with inaccurate information: Theory and algorithms. *IEEE/ACM Transactions on Networking*, 7(3):350–364, June 1999.

[6] T. Korkmaz and M. Krunz. Bandwidth-delay constrained path selection under inaccurate state information. *IEEE/ACM Transactions on Networking*, 11(3):384–398, June 2003.

[7] A. Bhorkar, M. Naghshvar, T. Javidi, and B. Rao. An adaptive opportunistic routing scheme for wireless ad-hoc networks. In *ISIT '09: Proceedings of IEEE International Symposium on Information Theory*, pages 2838–2842, Seoul, Korea, June 2009.

[8] M. L. Puterman. *Markov Decision Processes—Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1994.

[9] Z. Ye and A. A. Abouzeid. *Layered Sequential Decision Policies for Cross-layer Design of Wireless Ad Hoc Networks*. Technical Report, Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, 2009.

[10] D. P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 2. Athena Scientific, Cambridge, MA, 2001.

[11] P. Lassila and S. Aalto. Combining opportunistic and size-based scheduling in wireless systems. In *MSWiM '08: Proceedings of ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 323–332, Vancouver, BC, Canada, October 2008.

[12] M. Hu, J. Zhang, and J. Sadowsky. Traffic aided opportunistic scheduling for wireless networks: Algorithms and performance bounds. *Computer Networks*, 46(4):505–518, November 2004.

[13] M. G. Lagoudakis, R. Parr, and L. Bartlett. Least-squares policy iteration. *Journal of Machine Learning Research*, 4:1107–1149, December 2003.