Fabian Hohmann, Andrea Ortiz and Anja Klein, ”Optimal Resource Allocation Policy for Multi-Rate Opportunistic Forwarding,” in *Proc. of the IEEE Wireless Communications and Networking Conference (WCNC 2019)*, Marrakech, Morocco, April 2019.

Optimal Resource Allocation Policy for Multi-Rate Opportunistic Forwarding

Fabian Hohmann, Andrea Ortiz and Anja Klein
Communications Engineering Lab, Technische Universität Darmstadt, 64283 Darmstadt, Germany
Email:{f.hohmann, a.ortiz, a.klein}@nt.tu-darmstadt.de

***Abstract*—Many opportunistic routing protocols for wireless multi-hop networks rely on a fixed channel rate and a fixed priority order to manage the access to the channel by the involved nodes. Thereby, the actual channel capacities in the network are not considered and the diversity of links is not fully exploited. Furthermore, the data buffer of the nodes is not taken into account. In this work, we consider a wireless multi-hop scenario consisting of multiple cooperative nodes within each hop that share channel resources and adapt their channel rates based on local channel knowledge. A Markov Decision Process (MDP) model is used to derive an optimal resource allocation policy that minimizes the number of required time slots to forward all data packets to the next hop. Furthermore, we propose a state approximation technique that limits the required number of states, but captures the most important features of the problem. Simulation results demonstrate that the proposed policy achieves throughput gains of up to 25% compared to a fixed order transmission policy and up to 49% compared to a unipath approach.**

***Index Terms*—Opportunistic routing, wireless multi-hop networks, Markov decision process**

I. Introduction

Traditional multi-hop routing techniques like Dynamic Source Routing (DSR) [1] or Ad-hoc On Distance Vector (AODV) Routing [2] forward data along a fixed unipath. In wireless networks, these techniques struggle with the dynamic nature of channels due to the lack of alternative options in the path. Opportunistic routing protocols, such as ExOR [3], exploit the broadcast nature of wireless transmissions to overcome the limitations of a single unreliable channel. Instead of selecting a fixed next forwarding node, a set of candidates is considered as possible forwarders which leads to higher reliability and a lower amount of retransmissions. However, to avoid channel estimation, most of the proposed opportunistic routing protocols are based on a fixed single channel rate and do not adapt the channel rate to the actual channel conditions [4]. In [5], it is shown that opportunistic routing operating with multiple rates can achieve higher throughput than a single rate approach. Moreover, channel resources are usually assigned to the nodes according to a prioritization order of the forwarders. Each node waits for its turn to transmit based on, for instance, the expected transmission cost [3] or its distance to the destination [6]. Again, the actual channel conditions of the potential transmitters are not taken into account and therefore, the available diversity of channels is not fully exploited. Furthermore, the data buffer of the nodes is usually not taken into account. Buffer-aware opportunistic routing [7] combines the location and data buffer level of forwarding nodes to prioritize the selection of forwarders. Thereby, negative effects on the network throughput caused by accumulation of data packets at certain forwarding nodes are reduced.

In [8], short term channel fading statistics are incorporated to optimize the long term slot assignment, routing and scheduling in meshed networks. In [9] and [10], corridor-based routing is proposed which is an opportunistic strategy based on local channel knowledge and cooperation among the forwarding nodes. Channel resources are assigned according to current channel conditions and channel rates are adapted accordingly to fully exploit actual channel capacities. However, for the resource allocation, only suboptimal or heuristic approaches are used.

In this work, we consider a wireless multi-hop scenario with multiple cooperative forwarders in each hop that share the available channel resources and adapt their transmission rates according to the current channel states. We propose a resource allocation policy based on a Markov Decision Process (MDP) model. An MDP is also used in [11] to model a whole end-to-end forwarding process and the selection of forwarding nodes in opportunistic routing. However, the proposed strategy is limited to the transmission of a single packet and it struggles with increased number of nodes in the network. In this work, we use an MDP model that incorporates local channel statistics as well as the data buffers of the nodes to find a local optimal policy for resource allocation that minimizes the expected number of required time slots to forward data packets to the next hop. Furthermore, an approximation technique is proposed that limits the required number of states in the model and makes the problem feasible even in case of a large number of data packets or potential forwarding nodes.

The rest of the paper is organized as follows. The system model is explained in Section II. In Section III, the MDP model, the derivation of the optimal policy and the approximation technique are presented. In Section IV, the performance is evaluated and Section V concludes the paper.

II. System Model

In this work, we consider a multi-hop transmission from one source S to one destination D based on a support structure

This work was supported by the LOEWE initiative within the NICER project and by the German Research Foundation (DFG) within the Collaborative Research Center (CRC) 1053 - MAKI.

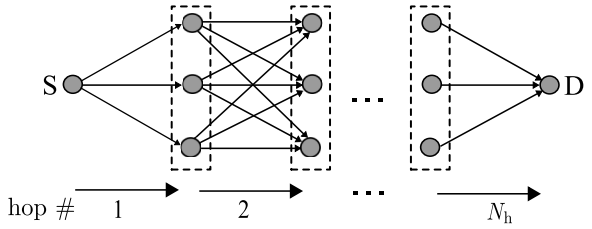

Fig. 1. Multi-hop network with $N_\mathrm{f} = 3$ forwarding nodes.

as shown in Figure 1. The support structure consists of N_h hops and N_f potential forwarding nodes from the second hop on. For simplicity reasons, we assume the number N_f to be constant within the network but the proposed algorithms can also handle varying numbers of forwarding nodes in the hops. It is assumed that the nodes are fully connected within each hop as shown in Figure 1. Nodes that are further away are assumed to be out of each other's transmission range. Rayleigh fading is assumed between the nodes. The channel between transmitter i and receiver j is described by the channel transfer factor $h_{i,j}$ and is modeled as a complex Gaussian distributed random process. The channels are considered to be constant within one time slot. All transmitters use the same fixed transmit power and all receivers have the same noise power. The transmit power and noise power are included in the channel transfer factor such that the Signal-to-Noise Ratio (SNR) of the channel is given by

$$\gamma_{i,j} = |h_{i,j}|^2, \tag{1}$$

Local channel knowledge is assumed to be available, i.e., the SNR of the strongest channel of each transmitter of a certain hop needs to be available at all transmitters of the corresponding hop. Each transmitter i needs to determine the SNR to receiver j that provides the currently highest SNR $\gamma_{i,\max} = \max_j \gamma_{i,j}$ and needs to share this value within the group of transmitters. In addition, the nodes need to know the average SNR $\bar{\gamma}_{i,j}$ of the channels within their hop which is the expected value of $\gamma_{i,j}$. To this end, hello messages can be used on a regular basis to measure and share this information in larger time intervals. In order to provide the current maximum SNR knowledge, channel estimation as well as one hop feedback is required at the beginning of each time slot as long as data is transmitted in the corresponding hop. We assume perfect channel estimation and do not consider channel estimation errors. In some cases, such estimation errors do not cause any consequences. In other cases, a channel estimation error could either be compensated by an error correcting channel coding or it could lead to a packet error such that a retransmission of the corresponding data packet is required. However, it would equally affect all considered transmission strategies and therefore, it is not crucial for the investigations. In our setting, the source wants to transmit $N_\mathrm{p,total}$ packets to the destination. The nodes can adapt their transmission rate stepwise according to the current channel state which results in a certain number N_p of transmitted packets. To model this rate adaptation, a mapping as shown in Table I is used. According to Shannon capacity $C = \log_2(1 + \mathrm{SNR})$, an SNR of 4.8 dB leads to a channel capacity of $C = 2$ bits/s/Hz. Without loss of generality we do not take assumptions on the available bandwidth, channel coding, packet size or time slot duration and take 2 bits/s/Hz as the minimum required capacity to transmit one data packet in one time slot. Therefore, to transmit two packets within a time slot, a channel capacity of 4 bits/s/Hz and an SNR of 11.8 dB is required and so on. This model captures the general functioning of an adaptive modulation and coding mechanism used in most communication protocols. It could be easily extended to a more detailed mapping.

TABLE I
RATE ADAPTATION

SNR	capacity	N_p / time slot
$<$ 4.8 dB	$<$ 2 bit/s/Hz	0
4.8 - 11.8 dB	2 bits/s/Hz	1
11.8 - 18 dB	4 bits/s/Hz	2
$>$ 18 dB	6 bits/s/Hz	3

The data packets are transmitted hop-by-hop using a decode-and-forward protocol, where nodes of a certain hop start to forward data packets only after the nodes of the previous hop have finished forwarding all $N_\mathrm{p,total}$ data packets. There is always only one node transmitting at a time, so there is no interference.

III. Resource Allocation Problem

In the first hop of the network, the source node transmits data packets to the next hop forwarders. In each time slot, data packets are transmitted to only one receiver that provides the highest channel capacity in this time slot. The number of transmitted data packets is adapted to the corresponding channel capacity. In case that multiple receivers have the same channel capacity in a time slot, the source node transmits to the receiver with less data packets in its data buffer. The proposed policy could be easily extended to handle and to benefit from an availability of data packets at multiple forwarders, but for simplicity reasons, we do not consider duplicates of packets that are available at multiple forwarders.

In the following hops, the data packets are distributed among the multiple forwarding nodes and therefore, channel resources need to be assigned to the nodes until all packets are forwarded to the next group of nodes. The forwarding nodes have a different amount of data packets in their data buffer and the aim is to find an allocation policy that minimizes the required number of time slots needed to forward all data packets to the next hop. The solution to this problem is not straightforward. A greedy policy that always assigns the channel to the transmitter with the highest capacity would maximize the achievable throughput, but only until the data buffer of one transmitter is empty. After that, the channels from this transmitter could not be used anymore and the diversity of available links to choose from is reduced. In order to minimize the required

number of time slots for transmission, an allocation policy is required that takes the data buffer levels as well as the channel conditions into account. To solve this problem, we take the following steps. Firstly, we model this local resource allocation problem as an MDP. Secondly, we derive an optimal policy using a policy iteration algorithm and thirdly, we present a state approximation technique to keep the required number of states in the model manageable.

A. Markov Decision Process Model

To find an optimal policy for the resource allocation problem within a certain hop, a dynamic programming algorithm can be used. These algorithms require a perfect model of the environment as an MDP. This means that all relevant deterministic variables need to be known as well as the probability distributions of the relevant stochastic processes. The MDP consists of a finite state set $\mathcal{S}$ and an action set $\mathcal{A}$. Each state $s \in \mathcal{S}$ is a function of the data buffer level B_i^t of each forwarding node i in the current time slot t. In addition, it is a function of the maximum SNR $\gamma_{i,\max} = \max_j \gamma_{i,j}$ of each transmitter i considering all available receivers j. The action corresponds to the assignment of the channel to a certain transmitter. The dynamics of the process are described by the transition probabilities $\mathcal{P}_{ss'}^a = \Pr\{s_{t+1} = s' | s_t = s, a_t = a\}$ where s_t denotes the state in time slot t and a_t denotes the action taken in time slot t. The immediate reward $\mathcal{R}_{ss'}^a$ incurred by the action a chosen in state s and leading to state s' is given by the number of transmitted packets in the corresponding time slot.
To complete the MDP model, the transition probabilities between the states are required. The current data buffer levels B_i^t in time slot t are known and by choosing an action a, i.e., allocating the channel in the current time slot to a certain transmitter, the data buffer levels of the nodes in time slot $t+1$ are also known. Therefore, the state s_{t+1} only depends on the channel states which are known to follow a Rayleigh distribution. This means that the envelope of a channel $h_{i,j}$ follows the probability density function (pdf)

$$P(|h_{i,j}| = x) = \frac{2x}{\bar{\gamma}_{i,j}} e^{-\frac{x^2}{\bar{\gamma}_{i,j}}}, \quad \text{for } x \geq 0. \tag{2}$$

The cumulative distribution function (cdf) describes the probability that the channel envelope is below a certain value and is given by

$$P(|h_{i,j}| \leq x) = 1 - e^{-\frac{x^2}{\bar{\gamma}_{i,j}}}, \quad \text{for } x \geq 0. \tag{3}$$

Since the state only depends on the strongest outgoing channel of each potential forwarding node, we need to consider the pdf

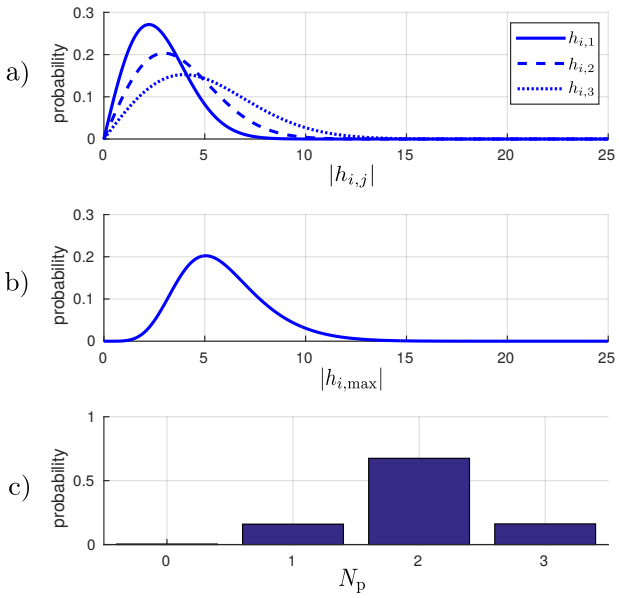

Fig. 2. a) The pdf of three Rayleigh channels with 10/12.5/15 dB average SNR. b) The pdf of the strongest channel out of the three channels. c) Resulting pmf for the number of transmittable packets

of $|h_i^{\max}| = \max_j |h_{i,j}|$ which is given by

$$\begin{aligned} P(|h_i^{\max}| = x) &= \sum_{j_1=1}^{N_\mathrm{f}} P(|h_{i,j_1}| = x) \prod_{\substack{j_2=1,\\ j_2 \neq j_1}}^{N_\mathrm{f}} P(|h_{i,j_2}| \leq x) \\ &= \sum_{j_1=1}^{N_\mathrm{f}} \frac{2x}{\bar{\gamma}_{i,j_1}} e^{-\frac{x^2}{\bar{\gamma}_{i,j_1}}} \prod_{\substack{j_2=1,\\ j_2 \neq j_1}}^{N_\mathrm{f}} \left(1 - e^{-\frac{x^2}{\bar{\gamma}_{i,j_2}}}\right). \end{aligned} \tag{4}$$

In Figure 2 a), the pdfs of the envelope of three Rayleigh fading channels $h_{i,1}, h_{i,2}$ and $h_{i,3}$ are shown with average SNR of $\bar{\gamma}_{i,1} = 10$ dB, $\bar{\gamma}_{i,2} = 12.5$ dB and $\bar{\gamma}_{i,3} = 15$ dB, respectively. The resulting pdf of $|h_i^{\max}| = \max_j |h_{i,j}|$ is shown in Figure 2 b). In Figure 2 c) the resulting probability mass function (pmf) of the number N_p of transmittable packets based on these three channels is shown. The pmf of each forwarding node is used to determine the state transition probabilities $\mathcal{P}_{ss'}^a$ in the MDP model.

B. Optimal Resource Allocation Policy

A policy π associates an action a to each state s. To find an optimal resource allocation policy π^* that minimizes the number of required time slots to forward all data packets to the next hop, policy iteration [12] can be used according to Algorithm 1. The process can be started with an arbitrary policy π'. In our case, we start with a greedy policy, i.e., in each state the action which provides the highest reward is selected. As a first step, we need to evaluate the current policy. To this end, the state value function $V(s)$ is used. Initially,

$V(s) = 0,\ \forall s$. In each iteration of the algorithm, the state value function is determined by

$$V(s) = \sum_{s'} \mathcal{P}^a_{ss'}(\mathcal{R}^a_{ss'} + \gamma V(s')), \tag{5}$$

where the action a is given by the current policy π'. This function gives the expected reward under the current policy. While the reward indicates the immediate value of an action, the value function evaluates the usefulness of an action in the long run. The value of a state gives the expected accumulated reward starting from this state. The so-called discount factor γ can be used to weight the importance of future rewards compared to the present reward. We want to maximize the expected reward until a final state is reached in which all data buffers are empty. Therefore, $\gamma = 1$ since the present reward has the same importance as future rewards. Next, policy improvement takes place, using the action-value function

$$Q(s, a) = \mathcal{R}^a + \gamma \sum_{s'} \mathcal{P}^a_{ss'} \cdot V(s'). \tag{6}$$

This function is used to update the policy and to make it greedy with respect to the value function. This means that the updated policy is not focused on the immediate reward, but on the expected reward over the entire procedure. The policy evaluation and policy improvement steps are repeated until the optimal policy π^* that provides the optimal action a for each possible state s has been found.

Algorithm 1 Policy iteration algorithm

Require: inital value function V ($V(s) = 0, \forall s \in \mathcal{S}$, initial policy π', $\Delta = 0$)
  while $\Delta > \epsilon = 0.0001$ **do**
    for each $s \in \mathcal{S}$ **do**
      $v \leftarrow V(s)$
      $V(s) = \sum_{s'} \mathcal{P}^a_{ss'}(\mathcal{R}^a_{ss'} + \gamma V(s'))$
      $\Delta = \max(\Delta, v - V(s))$
    end for
  end while
  for each $a \in \mathcal{A}(s)$ **do**
    for each $s \in \mathcal{S}$ **do**
      $Q(s, a) = \mathcal{R}^a + \gamma \sum_{s'} \mathcal{P}^a_{ss'} \cdot V(s')$
    end for
  end for
  set $\pi = \pi'$
  $\pi' := \arg\max_a(Q(s, a))$ (policy improvement)
  Repeat until $\pi = \pi'$

C. State Approximation

The computation of the optimal policy can be done offline beforehand. Then, the policy can be used as a look-up table consisting of the optimal action for each state. Since the required information is available at all forwarding nodes, there is no need to determine the optimal action in a centralized instance, but the optimal policy can be determined by each node in a distributed manner. However, large state spaces can be problematic for computation and storage reasons. The number of required states to model one hop of the network is given by $|\mathcal{S}| = (N_{\text{rates}} \cdot (B_{\max} + 1))^{N_{\text{f}}}$, where N_{rates} is the number of possible channel rates (including 0) and $B_{\max}$ is the data buffer size, i.e., the maximum number of data packets in a data buffer. For instance, considering 4 possible channel rates, a maximum of 10 data packets in a data buffer and a network with $N_{\text{f}} = 3$ forwarding nodes, already leads to $|\mathcal{S}| = 85184$ states.

In order to allow for a larger number of packets and nodes and still capture the relevant information of the problem, we propose the following approximation method. We found that the actual number of data packets is not critical for the optimal assignment of the channel resource. Instead, it is the ratio between the numbers of data packets in the data buffers of the nodes that is important. To capture this information with a fixed amount of states in the model, we proceed as follows. Let the maximum buffer size covered by the MDP model be $B_{\max}$ and let the actual maximum buffer level of the forwarding nodes be $B^{\text{actual},t}_{\max} = \max_i B^t_i > B_{\max}$. Then, we can break down the actual buffer levels of each node i to approximated values $B^{\text{approx.}}_i = \text{round}\left(\frac{B^t_i \cdot B_{\max}}{B^{\text{actual},t}_{\max}}\right)$ that are used in the MDP model, where $\text{round}(x)$ rounds x to the closest integer. Thereby, the maximum approximated buffer level equals the maximum buffer captured by the MDP model. The real buffer levels are not considered in the model anymore, but the ratio between the approximated buffer levels $B^{\text{approx.}}_i$ is approaching the real ratio. This ratio is the relevant information within the model. Using this approximation, the number of states in the MDP model can be fixed to any desired number and each actual buffer state can be mapped to an existing state within the MDP model and used to find the corresponding action.

IV. Performance Evaluation

In this section, numerical results based on MATLAB simulations are presented to evaluate the performance of the proposed policy. The simulations are restricted to three-hop networks since a higher number of hops would not provide any more insights in the performance of the proposed scheme. The proposed algorithms work only on local buffer and channel information and additional hops would provide the same problem as in the second hop, i.e., multiple transmitters with different data buffer levels forward to multiple receivers. Only the last hop provides a different situation because there is only one receiver available. Therefore, the proposed policy is evaluated in the second and third hop separately. We consider 2000 independent end-to-end transmissions of a data batch, each consisting of N_{packets} for each result. The average SNR of each channel is chosen randomly between 10 dB and 15 dB to model different distances between the nodes. In the following, the proposed policy is compared to a fixed order policy which emulates the operation of most other opportunistic forwarding strategies [4]. This means that one node is selected to first transmit until its data buffer is empty.

Then the next node is selected and so on until all packets are forwarded. The nodes adapt their transmission rate to the strongest receiver as in the proposed policy. Furthermore, unipath forwarding is considered for comparison where in each hop only one randomly chosen node is selected to forward all data packets. Again, rate adaptation is applied.

Figure 3 shows the average required number of time slots in hop 2 and hop 3 for a network with $N_{\mathrm{f}} = 2$ forwarding nodes per hop. In this case, the source transmits $N_{\mathrm{packets}} = 15$ packets per simulation run which is completely captured by the MDP model without any approximation. This means that $B_{\max} = 15$ and the proposed policy is optimal in this case. As can be seen, the proposed policy outperforms the fixed order policy and the unipath forwarding in both hops. In the third hop, the average number of required time slots is higher than in the second hop for both the proposed and the fixed order policy since there is only one available receiver and therefore, there is less channel diversity for the proposed policy and no channel diversity for the fixed order policy. Of course, the performance of unipath routing remains the same in both hops, since there is always only one transmitter and one receiver considered in each hop. In the second hop, the proposed policy requires on average 23% less time slots compared to the unipath approach which corresponds to a throughput gain of 30%. In the third hop, the proposed policy saves 11% compared to the fixed order policy. The fixed order policy performs slightly worse compared to the unipath approach in the third hop. This is caused by the case when the channel capacity exceeds the number of remaining data packets in the data buffer of a node which can happen only once per hop using the unipath approach but multiple times per hop using the fixed order policy.

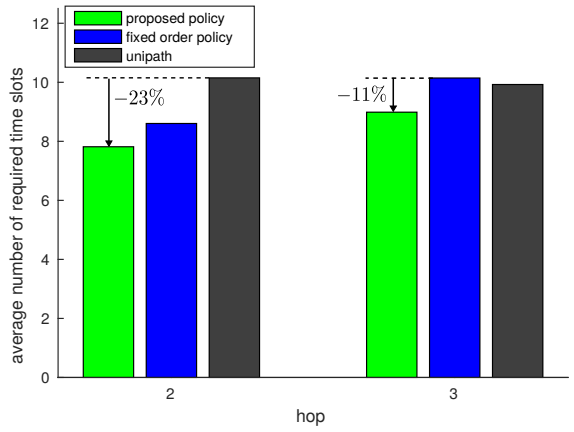

Fig. 3. Average required number of time slots to forward $N_{\mathrm{packets}} = 15$ packets with $N_{\mathrm{f}} = 2$ forwarders per hop.

In Figure 4, again $N_{\mathrm{f}} = 2$ forwarding nodes per hop are considered, but this time $N_{\mathrm{packets}} = 100$ packets are transmitted in each simulation run utilizing the approximation technique proposed in Section III C. It can be seen that the number of required time slots increases due to the higher number of packets to be transmitted compared to Figure 3. However, the achievable gain of the proposed scheme remains despite the use of the approximation technique. In fact, the gain is increased compared to the case with $N_{\mathrm{packets}} = 15$, due to a lower impact of an edge effect, when for residual packets that are left at only one node the policies perform the same. Of course, this effect has lower impact in case of a higher amount of packets. However, the state approximation does not seem to have a significant negative impact on the performance of the proposed policy.

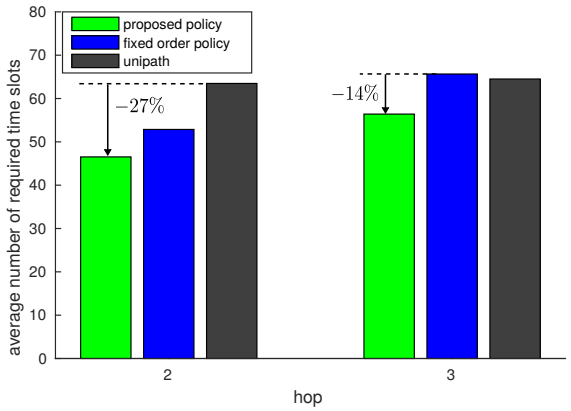

Fig. 4. Average required number of time slots to forward $N_{\mathrm{packets}} = 100$ packets with $N_{\mathrm{f}} = 2$ forwarders per hop using state approximation.

In Figure 5, a network with $N_{\mathrm{f}} = 3$ forwarding nodes is considered and $N_{\mathrm{packets}} = 100$ packets are transmitted in each simulation run. Again, the proposed approximation is used. It can be seen that with more forwarding nodes in each hop, the achievable gain increases. In the second hop, the proposed policy requires on average 33% less time slots compared to the unipath approach. This equals a throughput gain of 49%. In the third hop, the proposed policy on average saves 20% of time compared to the fixed order policy which equals a throughput gain of 25%.

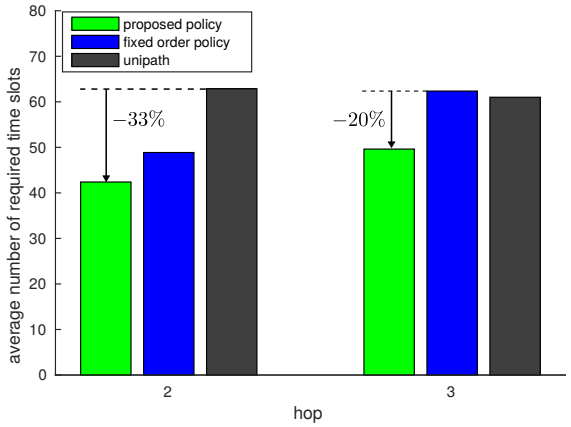

Fig. 5. Average required number of time slots to forward $N_{\mathrm{packets}} = 100$ packets with $N_{\mathrm{f}} = 3$ forwarders per hop.

V. Conclusion

In this work, we considered opportunistic forwarding in multi-hop networks based on multiple cooperating nodes in each hop. An optimal resource allocation policy is proposed that minimizes the expected number of required time slots to forward data packets to the next hop under use of local channel knowledge. The resource allocation is modeled by an MDP taking into account the data buffer levels of the nodes and multiple possible channel rates for transmission. To handle large state spaces, we propose a state approximation technique that limits the required number of states while it captures the most important state information. Numerical results show that the proposed policy outperforms a fixed order transmission policy by up to 25% and unipath transmissions by up to 49% in terms of average throughput.

References

[1] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*, 1996, pp. 153–181.

[2] C. Perkins and E. Royer, "Ad-hoc on-demand distance vector routing," in *Proc. IEEE Workshop on Mobile Computing Systems and Applications*, 1999.

[3] S. Biswas and R. Morris, "Exor: Opportunistic multi-hop routing for wireless networks," *SIGCOMM Comput. Commun. Rev.*, vol. 35, no. 4, pp. 133–144, Aug. 2005.

[4] N. Chakchouk, "A survey on opportunistic routing in wireless communication networks," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2214–2241, Fourthquarter 2015.

[5] K. Zeng, W. Lou, and H. Zhai, "On end-to-end throughput of opportunistic routing in multirate and multihop wireless networks," in *IEEE INFOCOM 2008 - The 27th Conference on Computer Communications*, April 2008, pp. 816–824.

[6] S. Yang, F. Zhong, C. K. Yeo, B. S. Lee, and J. Boleng, "Position based opportunistic routing for robust data delivery in MANETs," in *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference*, Nov 2009, pp. 1–6.

[7] W. Cui, Y. Yao, and L. Song, "Buffer-aware opportunistic routing for wireless sensor networks," in *2017 14th IEEE Annual Consumer Communications Networking Conference (CCNC)*, Jan 2017, pp. 268–271.

[8] W. Chen, C. Lea, S. He, and Z. XuanYuan, "Opportunistic routing and scheduling for wireless networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 320–331, Jan 2017.

[9] A. Kuehne, A. Klein, A. Loch, and M. Hollick, "Opportunistic forwarding in multi-hop OFDMA networks with local CSI," in *OFDM 2012; 17th International OFDM Workshop 2012 (InOWo'12)*, Aug 2012, pp. 1–8.

[10] A. Loch, M. Hollick, A. Kuehne, and A. Klein, "Corridor-based routing: Opening doors to PHY-layer advances for wireless multihop networks," in *Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014*, June 2014, pp. 1–3.

[11] J. Hao, X. Jia, Z. Han, B. Yang, and D. Peng, "Design of opportunistic routing based on markov decision process," in *2017 36th Chinese Control Conference (CCC)*, July 2017, pp. 8976–8981.

[12] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 1998.