Interference and Mobility-Aware Routing Scheme Based on Reinforcement Learning

Jae-Joon Lee and Jinsuk Kang

***Abstract*—Autonomous systems with communication capability are being widely used in diverse fields as the technology rapidly advances. One of key requirements for these systems is to provide reliable data communication in the face of mobility and interference that can significantly deteriorate network performance and service. To deal with this problem, we propose an interference and mobility-aware routing scheme based on reinforcement learning algorithm. The proposed scheme utilizes the mobility and interference characteristics of the network as a reward from the environment. Then, routes are selected based on learning from the network. Simulation results show that our scheme can achieve reliable data delivery paths in the face of interference and mobility.**

***Index Terms*—Interference, mobility, reinforcement learning, and routing.**

I. Introduction

As the technology in diverse fields including autonomous systems and communications rapidly advances, unmanned aerial vehicles (UAV) and autonomous systems including cars attract global attention in many areas. In commercial and military areas, the usage of UAVs including drones has been widely discussed and explored for practical applications. Especially, military has used UAVs to increase operational efficiency with reduction of damages.

Communication among autonomous systems as well as any mobile communication devices through wireless links continues to increase. As autonomous operation of the unmanned systems is important, autonomous construction and operation of reliable communication paths are also basic requirements in the advanced commercial and military operations [1]-[3].

Besides the autonomous operation of communication devices, the use of diverse mobile communication devices incurs interference among communication nodes that use the same frequency band [4]. In addition, intentional jamming especially in the military field significantly affects data communication through wireless mobile networks. There have been many studies on jamming attacks and defense strategies on wireless networks [5]. But, mobility and interference or jamming in autonomous systems were rarely studied together, while both issues can significantly deteriorate the network performance and reliability of wireless mobile networks.

Manuscript received March 8, 2016; revised April 13, 2016. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2009569) and (2014R1A2A1A11049469).

The authors are with Ajou University, Korea (e-mail: jjnlee@ajou.ac.kr).

In this paper, we discuss these two major issues in wireless mobile networks and propose a scheme that can overcome the effect of interference and mobility on reliable data delivery in wireless mobile networks. One of the main difficulties is to ensure the stable data connectivity in spite of mobile nature of the network. The other is to provide interference-aware networking scheme that can avoid data loss and service interruptions.

Our proposed scheme, interference and mobility-aware routing (IMR) scheme, is based on reinforcement learning, which is one of machine learning algorithms. Reinforcement learning has been widely examined in artificial intelligent applications as well as communications [6]. By adaptively learning network dynamics from the environments including the other nodes and links, the proposed scheme identifies the better route to deliver the data to the destination in order to reduce data loss and service interruption in a distributed way.

The rest of the paper is organized as follows. Related work is presented in Section II. Section III describes the reinforcement learning framework and Section IV presents the proposed scheme. Section V shows the simulation results and Section V concludes the paper.

II. Related Work

Recently, UAVs as a relay network has been studied to enhance the network coverage and operation. Disconnection problem due to failures in wireless sensor networks can be overcome by using UAVs as a relay node [7]. The coverage of UAV and characteristics of UAV communication links has been examined in [8]. Physical link characteristics and throughput of air to ground communication with high mobility is investigated in [9].

Interference and jamming are significant problems in wireless communications. Effect of jamming and methods to detect jamming occurrence are studied in [5], [10]. Corruption in packet data and received signal strength are good indicators of jamming occurrence [5]. Impact of interference on communication devices that use the same frequency band is also examined [4].

Reinforcement learning has been also studied to address the routing problem in wireless networks. Wireless networks have dynamic network topology caused by node mobility and wireless links' unstableness. Reinforcement learning-based adaptive routing protocol for underwater sensor networks is discussed in [11]. QELAR scheme in this work addresses the routing issues including energy efficiency and lifetime extension by utilizing Q-learning algorithm. In order to avoid slow convergence issue of model-free Q-learning that requires sufficient visits every state, they adopt model-based

doi: 10.18178/ijfcc.2016.5.2.451

Q-learning. QoS aware route selection algorithm is presented in [12]. This scheme aims to satisfy industrial application requirements such as reliable data delivery with limited memory and computational power. Reinforcing probabilistic routing algorithm to ensure QoS requirements is presented in [13]. In this work, N best optimal path Q-routing algorithm is proposed. This scheme optimizes cumulative cost path and end-to-end delay for route selection.

III. Reinforcement Learning Framework

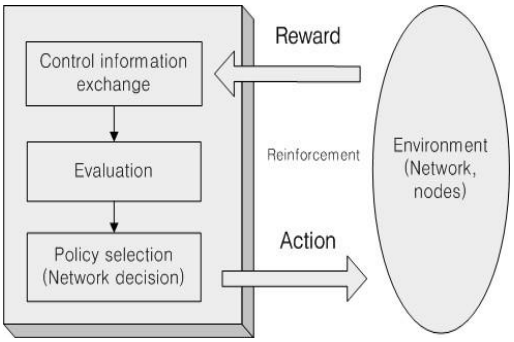

Fig. 1. Reinforcement learning framework.

A. Reinforcement Learning

One of widely used artificial intelligence techniques in networks is Reinforcement Learning (RL) [6]. RL satisfies the property of a Markov Decision Process (MDP). The Markov property implies that action outcomes depend only on the current state. To define a finite MDP, we need the followings: a set of states $s \in S$, a set of actions $a \in A$, a model $T(s,a,s')$ defined by the transition probabilities $Pr(s'|s,a)$, and a reward function $R(s,a,s')$. A policy π (s) gives selection of an action for state s. A MDP looks for theoptimal policy π *(s), which is the optimal action that maximizes the expected utility. The expected utility is the accumulated (discounted) reward from state s. $V_{\pi}(s)$ is the state value which represents the expected discounted reward from state s and $V^*(s)$ is the highest state value with the optimal policy π *(s) [6].

$$V^*(s) = \max_a Q^*(s,a)$$

where $Q^*(s, a)$ is the expected discounted reward with the action a from state s and acting optimally thereafter.

$$Q^*(s,a) = \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')], \quad (1)$$

where γ is the discount factor, $0 \le \gamma \le 1$, which determines the effect of future rewards on the current value.

The relation between policy and Q-value is the following.

$$\pi^*(s) = \arg\max_a Q^*(s,a)$$

From evaluation of Q-value, we can obtain the optimal policy.

RL is a model-free and unsupervised machine learning algorithm that does not construct a state transition probability matrix, a model $T(s, a, s')$ with an external supervision. To construct a model representing state transition in MDP, an offline processing is required. Instead of relying on a model T, RL approach utilizes the feedback from the environment to find an optimal policy based on experience without prior knowledge of the environment. The RL agent improves performance by exploring its operating environment. To learn the best action that achieves the maximum expected accumulated reward, the RL agent selects an action at each state and time step, and observes state and reward based on its selection of an action by trial and error approach.

B. Q-Learning

Q-learning is a well-known technique of RL approach. Q-learning utilizes Q-value to find optimal policy. Q-value is updated by averaging over previous Q-values and direct reward from environment with discounted accumulated reward.

Since RL approach does not rely on the pre-constructed state transition model, $T(s,a,s')$ with an external supervision, the expected value is obtained by sample-based running average with exponential moving average method. Thus, from (1), Q-value can be updated online with new feedback from environment without relying on the pre-constructed state model [6].

$$Q_{t+1}(s,a) = (1-\alpha)Q_t(s,a) + \alpha(r + \gamma \max Q_t(s',a')), \quad (2)$$

where α is the learning rate, $0 \le \alpha \le 1$, which determines the weight between the present and the past value, and r is the reward, which is the immediate feedback from the environment. The reward can be also regarded as the cost that is incurred by the chosen action. This is a sample-based approximation.

We can apply this RL method to network operation as shown in Fig. 1. At time t, an agent observes state s and can select an action a, which can be a next-hop node for relay of data. Through continuous actions and feedback from its operating environment, an agent learns an action that can improve the target utility. Feedback from the operating environment is reward. Whenever an agent acts, it receives a reward which indicates the performance or cost of the action result.

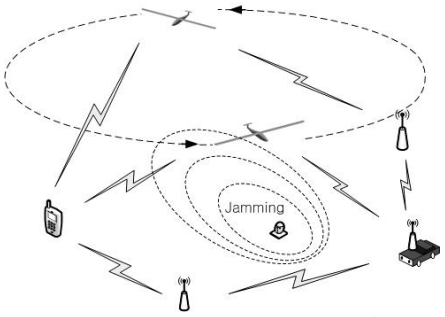

Fig. 2. Wireless mobile network with interference.

IV. Interference and Mobility

To provide reliable data communication in mobile networks with interference as shown in Fig. 2 by utilizing Q-learning, the feedback from the networking environment

should be incorporated into the Q-learning model. The reward function applies the effect of interference and mobility in the network environment to the model.

A. The Effect of Interference and Error

Interference affects the communication quality and even hinders the proper communication. Interference includes the effect of other communication using the same channel and intentional jamming signals generated by malicious attackers.

Communication devices that use the same frequency band can interfere with each other. Scarcity of frequency resource leads to use of the same frequency band [4]. Increased use ofinternet of things (IoT) devices is makingthe wireless medium further crowded.

To indicate the interference and packet error effect, the cost c_E is defined as follows:

$$c_E = PER_{ij}$$

PER_{ij} is the packet error rate between nodei and j. Packet error rate increases as the interference increases or link status is not good due to any hindrance to wireless link.

B. Mobility Awareness

UAVs are rapidly increasing and communication through and among UAVs are of increasing interest. One of main differences from the other types of communication devices is mobility and autonomous operations of the system.

In our scheme, each node measures the mobility by using the distance change based on the location information. If the distance between two nodes increases, the cost will increase because the possibility of disconnection goes up. However, when the distance between two nodes is less than the average distance of all neighbors, the possibility of disconnection will not be significant compared to the other neighbors.

$$c_M = \frac{d_{ij}(t) - d_{ij}(t-\delta)}{\alpha_1 \cdot \delta \cdot \overline{d_n}},$$

where $d_{ij}(t)$ is the distance between node i and j at time t. $\overline{d_n}$ is the average distance between the node and its neighbor nodes. α_1 is the weight parameter. δ is the time gap between adjacent measures of distance between node *i* and *j*. The result indicates the relative speed of the possible relay node. If the distance between two nodes increases from the previous time, then the cost will increase.

C. Interference and Mobility Prediction

In addition, the distance from the interference region is incorporated into the cost evaluation. In this case, nodes preemptively avoid the imminent interference region in one or multiple hops away. When the interference region approaches a node, that node has high possibility of experiencing interference in near future compared to the other nodes.

$$c_I = (\min_{j \in F} v_{ij}(t))^{-\beta_1} + (v_{ij}(t-\delta) - v_{ij}(t))/(\beta_2 v_{ij}(t)) \quad (3)$$

where F is the set of nodes that are affected by the severe interference or jamming. v_{ij} is the minimum distance between the node i and a node in the set F. Each node in the interference region notifies the existence of interference to the other nodes outside the interference region by sending notification message as described in [5]. In addition, the change of distances from the interference region is incorporated into this cost function. When the distance from the interference region becomes closer, the cost should increase.C_I indicates the cost that increases as the minimum distance from the interference region decreases. β_1 and β_2 are the weight parameters and greater than or equal to one.

Reward function consists of the above three cost functions.

$$r = -w_1 c_E - w_2 c_M - w_3 c_I \quad (4)$$

where w_1, w_2, w_3 are the weight parameters. Reward is a negative value and incorporated into (2) for update of Q-values.

V. Network Operation

Network operation of our scheme consists of three phases: network information exchange, reward evaluation and route selection based on Q-learning as shown in Fig. 1.

A. Control Information Exchange

First, each node participating in the network periodically exchanges control message among one-hop neighbors, which is common in most network protocols. Hello message or beacon is an example of popular control messages, which indicates the existence to its neighbors and provides basic control information required for network operation. Control information can be also included in the head of data packet.

In our scheme, the following information should be included in the control message by each node: location information, maximum Q-value and distance from interference region. Location information can be obtained through diverse localization schemes such as GPS or Wi-Fi positioning technique. Many networking schemes adopt location information of nodes in the networks. In addition, as discussed in [5], interference notification from the node inside the jammed or interference region is utilized to calculate the minimum distance from the interference region. Each node includes the above information in the periodic control message.

B. Reward Evaluation

When receiving the periodic control information, each node examines the interference degree of neighbors, mobility estimation of neighbors and interference prediction based on the computation in the previous section. By utilizing information from network environments and neighbors, Q-values are computed and updated.

Packet error rate can be also obtained from normal data transmission with acknowledgements and from physical layer. Periodic exchange of location information provides the mobility status of neighbors. Interference and the effect of mobility are examined and the cost is computed with (3). Based on computation results, reward of each action, which is

route selection, is computed as in (4).The obtained reward for each action is provided to (2) for Q-value update.

C. Route Selection

After evaluating and updating Q-value, each node takes the optimal policy by selecting the neighbor node with the highest Q-value. This selection procedure is applied to neighbor nodes that can reach the destination and become a relay node. A mobile network incurs dynamic network topology changes, which result in changes of nodes' locations or interference region due to mobile interference source. Thus, periodic exchange of control information, including location of nodes, Q-value from neighbors, interference and mobility awareness information, can expedite the processing of Q-learning algorithm.

Conventional q-learning requires overhead of sufficient exploration period to check undiscovered states for convergence, which can incur data delivery failure in dynamic mobile networks. However, because our scheme utilizes periodic control information exchanged among neighbors similar to [11], we can focus on exploitation of the gathered information to search more stable and protected network route to a destination. By immediately utilizing the network information, routing decision can more quickly adapt to the dynamics of mobile networks. Thus, each node selects a relay node in a greedy way by choosing the neighbor node with highest q-value.

VI. Simulation

To verify our scheme, the interference and mobility aware routing (IMR), we conducted computer simulations with the other two different schemes, which can react to the interference or jamming incidents and reconstruct route to the destination. The first routing scheme is based on the interference-reactive minimum hop routing (IMHR), which adaptively adjusts the routing path when route failure occurs due to severe interference or mobility. The second routing scheme to be compared predicts future risk of being under interference by examining the distance change from the interference region with mobility, which is called the jamming-aware routing (JAR) scheme. In this heuristic scheme, if the approaching speed of the interference source increases, the cost of using that node as a relay node increases. Then, jamming aware link cost of that node is updated and propagated to nodes in the network following conventional routing protocols based on Dijkstra's algorithm, which can incur control message overhead. Mobility effect is also similarly incorporated into this scheme.

In the simulation comparison, total 25 nodes were deployed in a network. Mobile nodes moved randomly with a certain speed and direction bound. Each simulation ran 100s and 50 runs were averaged per each different average speed of mobile nodes setting. Interference source is located in the middle of the network deployment, which generates jamming signals to its neighboring area.

Fig. 3 shows the number of route losses due to severe interference or disconnection in mobile environment. Route loss occurs when any node in a current data delivery path from the source to the destination experiences severe interference or disconnection from neighbors that consist of the path. As the average speed of mobile nodes increases, the existing route to the destination has high possibility to be interfered if preemptive avoidance scheme is not applied to route selection as in IMHR. JAR and IMR schemes adopt preemptive approach to reduce the risk of disconnection due to either interference or mobility. When the average speed of mobile nodes is slow, IMR shows more reliable preemptive route construction than JAR. However, as the speed increases, loss occurrence of IMR becomes close to that of JAR due to convergence time.

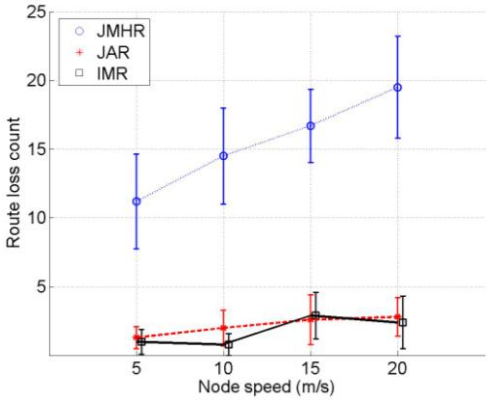

Fig. 3. Route loss occurrences.

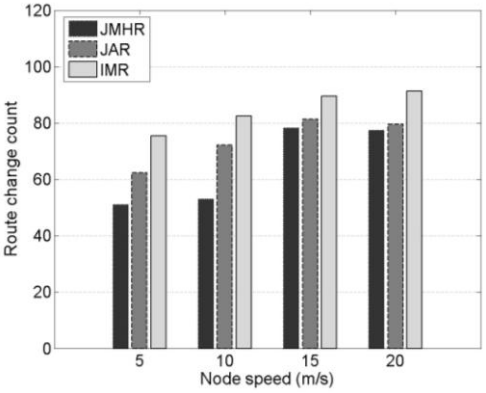

Fig. 4. Route changes occurrences.

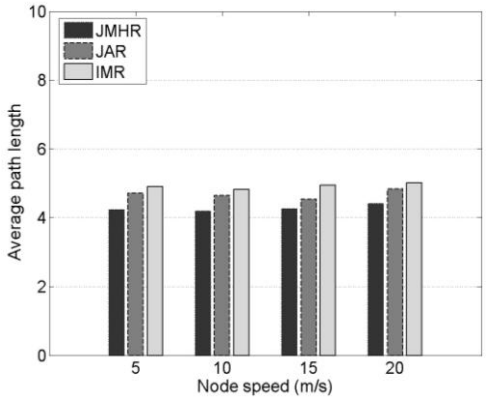

Fig. 5. Average path length.

Fig. 4 presents the number of route changes with respect to node speed. Since preemptive approaches adaptively change data delivery path based on the network environments, more route changes were made. Dynamic change of data delivery paths can avoid interference and disconnection due to mobility. In the case of IMR, there is a trade-off between more adaptive approach and stable learning. If the learning rate α is

close to 1, then the route selection scheme more adaptively responds to the dynamics of network environment, which can incur more route changes. In our simulation, α is set to 0.9 to more promptly react to the mobile network.

Fig. 5 shows the average number of hops in the data delivery paths from the source to the destination during the simulations. The path length of IMR is slightly higher than the other schemes when we use conservative weight against risk of interference and mobility.

Simulation results show that our proposed scheme IMR provides reliable data delivery path in the network with mobility and interference. Especially, when the mobility is not significantly high, IMR shows the better reliability than the other heuristic preemptive scheme.

VII. Conclusion

We have examined the use of reinforcement learning in mobile networks with interference. Reinforcement learning method can achieve reliable data delivery path in a preemptive way so that data loss and service disruption can be prevented. Reinforcement learning can reduce the control overhead of network flooding by utilizing neighbor information and feedback from nodes with direct links. We showed that the proposed scheme can achieve reliable data path in the face of dynamic networks with mobility and interference.

By utilizing our preemptive interference and mobility aware networking scheme, any autonomous systems with communication ability, such as UAVs, can avoid data loss and service interruptions due to interference in mobile environments. Especially, military operation that requires reliable communication among autonomous systems in the hostile situation, including jamming, can benefit from our scheme.

In the future work, we can further investigate the effect of learning rate and other parameters as well as convergence time of the reinforcement learning scheme on the network performance. In addition, benefits of function approximation scheme in the mobile environments can be also examined.

References

[1] B. Ilker, O. K. Sahingoz, and Ş. Temel, "Flying ad-hoc networks (FANETs): A survey," *Ad Hoc Networks*, vol. 11, no. 3, 2013, pp. 1254-1270.

[2] H. Dan, H. T. Kung, and B. Suter, "Field experimentation of cots-based UAV networking," presented at Military Communications Conference, 2006.

[3] F. W. Eric and T. X. Brown, "Networking issues for small unmanned aircraft systems," *Journal of Intelligent and Robotic Systems*, vol. 54, no. 1-3, 2009, pp. 21-37.

[4] S. Y. Shin, H. S. Park, S. Choi, and W. H. Kwon, "Packet error rate analysis of zigbee under WLAN and bluetooth interferences," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, 2007, pp. 2825-2830.

[5] W. Xu, K. Ma, W. Trappe, and Y. Zhang, "Jamming sensor networks: Attack and defense strategies," *Network*, vol. 20, no. 3, 2006, pp. 41-47.

[6] S. Richard and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 1998.

[7] E. P. Freitas *et al*., "UAV relay network to support WSN connectivity," presented at 2010 International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), IEEE, 2010.

[8] C. Carmen, "An analysis of unmanned airborne vehicle relay coverage in urban environments," presented at Military Communications Conference, 2007.

[9] Y. Evsen, R. Kuschnig, and C. Bettstetter, "Achieving air-ground communications in 802.11 networks with three-dimensional aerial mobility," in *Proc. INFOCOM*, 2013.

[10] A. D. Wood and J. A. Stankovic, "Denial of service in sensor networks," *IEEE Computer*, vol. 35, no. 10, pp. 54–62, 2002.

[11] H. Tiansi and Y. Fei, "QELAR: A machine-learning-based adaptive routing protocol for energy-efficient and lifetime-extended underwater sensor networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 6, 2010, pp. 796-809.

[12] Villaverde, B. Carballido, S. Rea, and D. Pesch, "In rout — A QoS aware route selection algorithm for industrial wireless sensor networks," *Ad Hoc Networks*, vol. 10, no. 3, 2012, pp. 458-478.

[13] M. Abdelhamid, S. Hoceini, and M. Cheurfa, "Reinforcing probabilistic selective quality of service routes in dynamic irregular networks," *Computer Communications*, vol. 31, no. 11, 2008, pp. 2706-2715.

Jae-Joon Lee received the Ph.D degree in computer engineering from the University of Southern California (USC) in 2007. Till January 2010, he served as a senior engineer at Samsung Electronics and participated in research and development of a linux-based mobile platform. In February 2010, he joined Ajou University for mobile wireless network research as a research professor. His research interests are modeling, analysis and design of mobile, wireless embedded systems, ad hoc and sensor networks including internet of things as well as data analysis based on machine learning. He has led several projects including wireless tactical networks and context-aware preemptive schemes in mobile environments.

Jin-Suk Kang received his B.S. degree in information engineering from Cheju National University, Jeju, Korea, in 1999 and his M.S. and Ph.D. degrees in computer engineering from Cheju National University, Jeju, Korea, in 2001 and 2005, respectively. From 2006 to 2009, he was with the University of Incheon, Korea, as a research professor. From February 2009 to March 2010, he worked with Chungbuk National University, Korea, as a visiting professor. Since March 2010, he has been with the Jangwee Research Institute for National Defence, Ajou University, Suwon, Korea, where he is currently a research professor. His research interests include the areas of multimedia, computer vision, human-computer interaction, mobile computing and embedded system, etc.