Indonesian Journal of Electrical Engineering and Computer Science
Vol. 5, No. 2, February 2017, pp. 383 ~ 390
DOI: 10.11591/ijeecs.v5.i2.pp383-390

Prioritized Sweeping Reinforcement Learning Based Routing for MANETs

Rahul Desai*[1], B. P. Patil[2]
[1]Sinhgad College of Engineering, Army Institute of Technology, Pune, Maharashtra, India
[2]E&TC Department, Army Institute of Technology, Pune, Maharashtra, India
Corresponding author, e-mail: desaimrahul@yahoo.com*[1], bp_patil@rediffmail.com[2]

Abstract

In this paper, prioritized sweeping confidence based dual reinforcement learning based adaptive network routing is investigated. Shortest Path routing is always not suitable for any wireless mobile network as in high traffic conditions, shortest path will always select the shortest path which is in terms of number of hops, between source and destination thus generating more congestion. In prioritized sweeping reinforcement learning method, optimization is carried out over confidence based dual reinforcement routing on mobile ad hoc network and path is selected based on the actual traffic present on the network at real time. Thus they guarantee the least delivery time to reach the packets to the destination. Analysis is done on 50 Nodes Mobile ad hoc networks with random mobility. Various performance parameters such as Interval and number of nodes are used for judging the network. Packet delivery ratio, dropping ratio and delay shows optimum results using the prioritized sweeping reinforcement learning method.

Keywords*: Confidence Based Routing, Dual Reinforcement Q Routing, Q Routing, Prioritized sweeping*

1. Introduction

The most simplest and effective policy used in the network is the shortest path routing. In shortest path routing the path with minimum number of hops is selected to deliver the packet from source to the destination. In shortest path routing, cost table and neighbor tables are present to store the appropriate information and tables are exchanged frequently for adaptation purpose. The shortest path routing policy is good and found effective for less number of nodes and less traffic present on the network. But this policy is not always good as there are some intermediate nodes present in the network that are always get flooded with huge number of packets. Such routes are referred as popular routes. In such cases, it is always better to select the alternate path for transmitting the packets. This path may not be shortest in terms of number of hops, but this path definitely results in minimum delivery time to reach the packets to the destination because of less traffic on those routes. Such routes are dynamically selected in real time based on the actual traffic present on the network. Hence when the more traffic is present on some popular routes, some un-popular routes must be selected for delivering the packets. This is the main motivating factor for designing and implementing various adaptive routing algorithms on a network.

Ad Hoc networks are infrastructure less networks. These are consisting of mobiles nodes which are moving randomly [1]. Routing protocols for an ad hoc network [2] are generally classified into two types - Proactive and On Demand. Proactive protocols which are table driven routing protocols which attempt to maintain consistent, up to date routing information from each node to every other node in the network. These protocols require each node to maintain one or more tables to store routing information and they respond to changes in network topology by exchanging updates throughout the network. Destination Sequenced Distance Vector (DSDV) is proactive routing protocols. Dynamic Source Routing (DSR) and Ad Hoc On Demand Distance Vector (AODV) [3-4] are on demand routing protocols for an ad hoc network. DSDV is based on distance vector routing protocol and uses destination sequence numbers to avoid count to infinity problem. DSR is characterized by the use of source routing. That is, the sender knows the complete hop-by-hop route to the destination. These routes are stored in a route cache. AODV uses traditional routing tables, one entry per destination. Ad Hoc On Demand Multipath

Received September 2, 2016; Revised December 25, 2016; Accepted January 11, 2017

Distance Vector (AOMDV) is an another optimized version of AODV where multiple entries are stored in cache. A comparison of various existing routing protocols are specified in [5]

2. Reinforcement Learning

Reinforcement learning is learning where the mapping between situations to actions is carried out so as to maximize a numerical reward signal [6, 7]. Figure 1 shows agent's interaction with the system. An agent checks the current state of system, chooses one action from those available in that state, observes the outcome and receives some reinforcement signal [8-10].

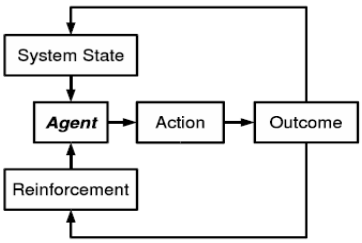

Figure 1. Reinforcement Learning Approach

Q Routing is one of the best reinforcement based learning algorithm. In this, each node contains reinforcement learning module which dynamically determines the optimum path for every destination [11-13]. Let Qx(y, d) be the time that a node x estimates it takes to deliver a packet P to the destination node d through neighbor node y including the time that packet would have to spend in node x's queue. Upon sending packet to y, x gets back y's estimate for the time remaining in the trip. Upon receiving this estimate, node x computes the new estimate [14-15].

In another optimized form, Confidence Based Q Routing (CBQ), each Q value is associated with confidence value (real number between 0 and 1). This value essentially specifies the reliability of Q values All Intermediate nodes along with Q value, also transmits C values which will updated in confidence table [14-16].

Dual reinforcement Q Routing (DRQ) is another optimized version of the Q Routing, where learning occurs in both ways. Performance of DRQ routing almost doubles as learning occurs in both directions. The various optimizations on Q routing are also studies in [17-20].

2. Prioritized Sweeping Reinforcement Learning

Mostly, a packet has multiple possible routes to reach to its destination. The decision of selecting best route is very important in order to reach the packets to the destination having a least amount of time and without packet loss. This selection has three main challenges, first, coordination and proper communication among nodes in a network is always required. Second, link and node failure cases should be handled gently. Third and very most important in dynamic environment, routes must be able to change dynamically according to the state of the network [18]. The shortest path routing policy is good and found effective for less number of nodes and less traffic present on the network. But this policy is not always good as there are some intermediate nodes present in the network that are always get flooded with huge number of packets. Such routes are referred as popular routes. In such cases, it is always better to select alternate path for transmitting the packets.

For example, in order to demonstrate limitation of shortest path algorithms (Figure 2), consider that Node 0, Node 9 and Node 15 are simultaneously transferring data to Node 20. Route Node 15-16-17-18-19-20 gets flooded with huge number of packets and then it starts dropping the packets. Thus shortest path routing is non-adaptive routing algorithm that does not take care of traffic present on some popular routes of the network. Learning such effective policy for deciding routes online is major challenge, as the decision of selecting routes must be taken in real time and packets are diverted on some unpopular routes. The main goal is to

optimize the delivery time for the packets to reach to the destination and preventing the network to go into the congestion. There is no training signal available for deciding optimum policy at run time, instead decision must be taken when the packets are routed and packets reaches to the destination on popular routes [19].

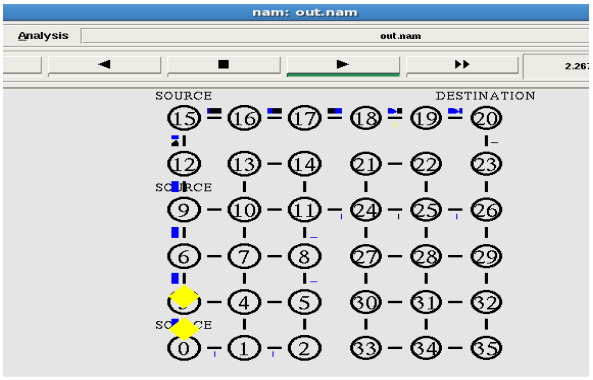

Figure 2. Limitation of Shortest Path Algorithms

Prioritized sweeping is a method that requires a model of the environment. The prioritized sweeping technique makes sweeps through the state of spaces, generating for each state the distribution of possible transactions [20]. It uses all previous experiences both to prioritize important dynamic programming sweeps and to guide the exploration of the state space. In the Q-Routing framework, the state was a packet finds itself in, is defined by the node that has the packet in its waiting queue and by the destination the packet is destined to. The actions available in that state are represented by sending the packet to one of the node's neighbors. When a node n selects greedy its best action A' for a particular packet P(S, D) , it forwards the packet P(S, D) to node N' the neighbor-node for which node n believes that it has the best estimate for delivering packet P to its final destination D. In order that prioritized sweeping can give a high priority to the preceding states of a changed state, node N' needs to send a control message M to all the neighbor nodes n that can make a transition to node N'. The control message M takes along with it, the destination D, its own node-id id, and the priority P. A node n receiving such a control message looks in its routing table if node N's best estimate for delivering a packet P(S, D) to destination D would use node id. In order that this preceding state can be updated node N places the tupel (d, id) in its priority queue with priority P, if this is not the case the packet is simply discarded [20].

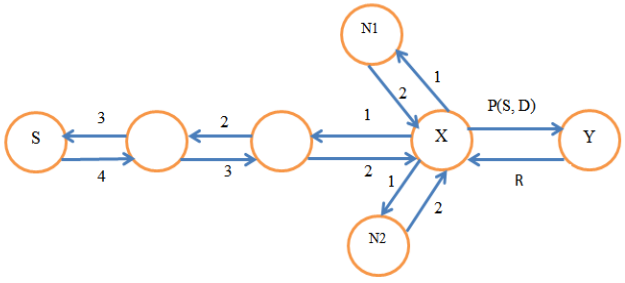

Figure 3. Prioritized sweeping technique for the CDRQ Routing framework

Figure 3 shows prioritized sweeping technique for the CDRQ Routing Framework. When node X sends a packet P(S, D) to node Y, it immediately gets back node Y's best estimate R for delivering the packet to the destination. Node X updates its model and computes

the absolute difference, if this is larger than small threshold θ, it places the tupel (D, Y) in its priority queue with priority P. Node X will make such N state transitions, for each state transition, it pops a state action pair (S, A) from its priority queue, control message M is sent to all the neighbors of the node (labeled as 1) [20].

When node N receives a control message M, it extracts the state S, action id and the reward R. if the absolute difference is bigger than the threshold θ and node N's best estimate for delivering the packet with destination s uses the neighbor node id then the tupel (S, id) is placed in node N's priority queue with priority P, thus each time when absolute difference is greater than the threshold θ, the state change is propagated further throughout the network (labeled as 2,3 and 4) [20].

4. Results and Analysis

All experiments are performed using standard network simulator NS-2.34. This experiment is carried on 50 Nodes MANET with random mobility of nodes. Default packet size is 512 bytes. Interval varies from 0.004 to 0.008. Thus around 125 to 250 packets are transmitted per second. Simulation is carried out for 200 seconds. Figure 4 refers to interval versus PDR, while Figure 5 presents interval versus dropping ratio. Prioritized sweeping reinforcement learning based method is compared with DSDV, AODV and DSR protocols.

Also the performance parameters - delay is also studied using prioritized sweeping reinforcement learning method. Prioritized sweeping reinforcement learning method provides very low delay for all packets reaching to the destination. Figure 6 refers to interval versus delay for 50 nodes MANET.

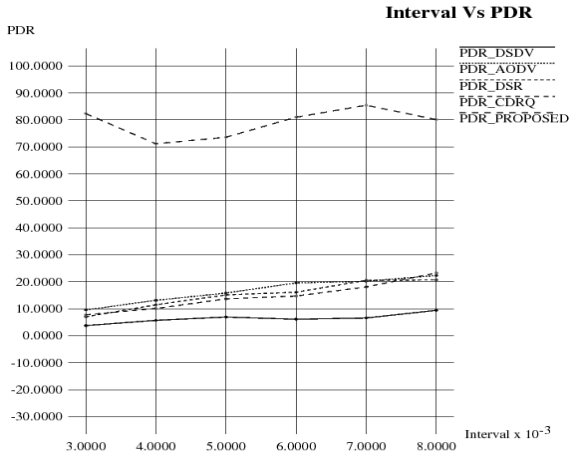

Figure 4. Interval vs. PDR

Table 1. Interval vs. PDR for 50 Nodes Mobile Ad Hoc Network with Random Mobility

Interval	0.003	0.004	0.005	0.006	0.007	0.008
AODV	9.57	13.09	15.82	19.56	20.21	22.32
DSDV	3.74	5.68	6.91	6.09	6.56	9.38
DSR	6.99	11.44	15.13	16.10	20.55	20.69
CDRQ	7.72	10.08	13.64	14.69	18.08	23.25
PSRL	82.34	71.16	73.60	81.01	85.45	80.14

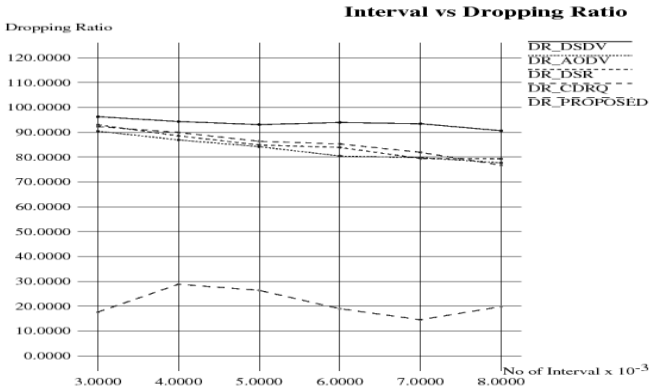

Figure 5. Interval vs. Dropping Ratio

Table 2. Interval vs. Dropping Ratio for 50 Nodes Mobile Ad Hoc Network with Random Mobility

Interval	0.003	0.004	0.005	0.006	0.007	0.008
AODV	90.42	86.90	84.17	80.43	79.78	77.67
DSDV	95.25	94.31	93.08	93.90	93.43	90.61
DSR	93.00	88.55	84.83	83.89	79.44	79.30
CDRQ	92.27	89.91	86.35	85.30	81.91	76.74
PSRL	17.65	28.84	26.39	18.98	14.54	19.85

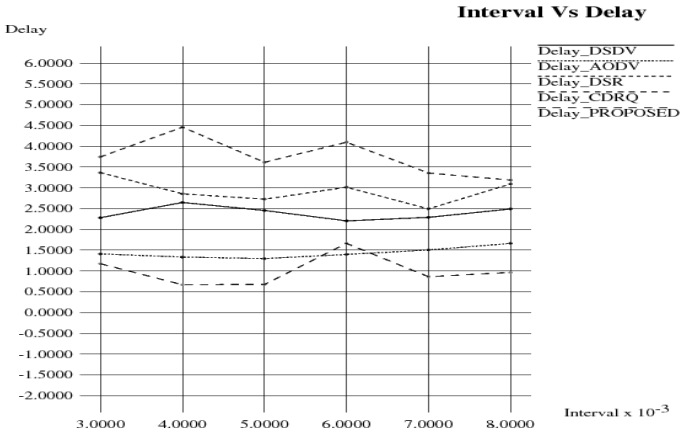

Figure 6. Interval vs. Delay

Table 3. Interval vs. Delay for 50 Nodes Mobile Ad Hoc Network with Random Mobility

Interval	0.003	0.004	0.005	0.006	0.007	0.008
AODV	1.41	1.33	1.29	1.39	1.50	1.66
DSDV	2.28	2.64	2.45	2.20	2.28	2.49
DSR	3.36	2.85	2.72	3.01	2.49	3.09
CDRQ	3.74	4.45	3.61	4.09	3.35	3.18
PSRL	1.17	0.66	0.67	1.66	0.86	0.96

Experiment is carried out 50 Nodes MANET by changing number of nodes. Number of nodes varies from 10 to 100. Packet rate is 250 packets per second and size of packet is 512 bytes. The results obtained are shown in Figure 7 to Figure 9. It is found that PDR is in range of 60% to 100% throughout the network for prioritized sweeping reinforcement learning method. The dropping ratio is high in case of CDRQ method. Also as we increase the number of nodes, delay increases in CDRQ method, but prioritized sweeping reinforcement learning method maintains constant delay irrespective of increasing the number of nodes. The numbers of control packets generated are almost constant.

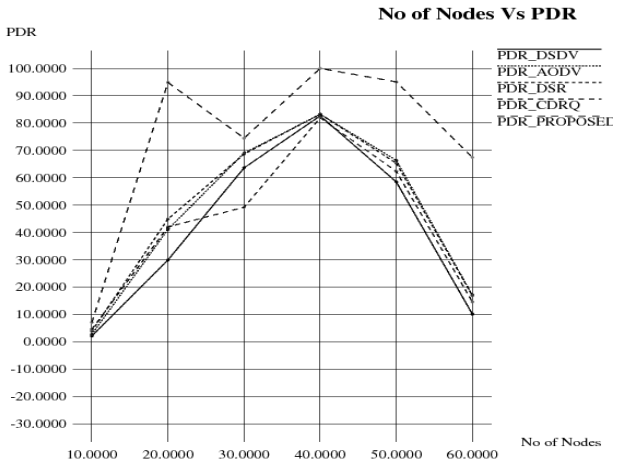

Figure 7. No of Nodes vs. PDR

Table 4. No of Nodes vs. PDR for MANET with Random Mobility

No of Nodes	10	20	30	40	50	60
AODV	2.74	41.11	69.01	83.27	66.38	17.12
DSDV	2.14	29.92	63.64	82.65	58.49	10.13
DSR	3.94	44.91	68.67	83.24	65.35	16.89
CDRQ	4.58	42.06	49.23	81.93	62.32	14.65
PSRL	7.10	94.87	74.49	100	95.06	67.43

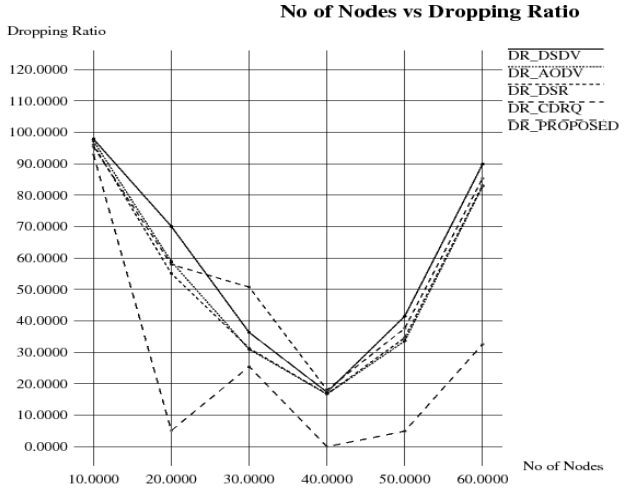

Figure 8. No of Nodes vs. Dropping Ratio

Table 5. No of Nodes vs. Dropping Ratio for MANET with Random Mobility

No of Nodes	10	20	30	40	50	60
AODV	97.25	58.88	30.98	16.72	33.61	82.87
DSDV	97.85	70.07	36.35	17.34	41.50	89.86
DSR	96.05	55.08	31.32	16.75	34.64	83.10
CDRQ	95.41	57.93	50.76	18.06	37.67	85.34
PSRL	92.89	5.12	25.50	0	4.93	32.56

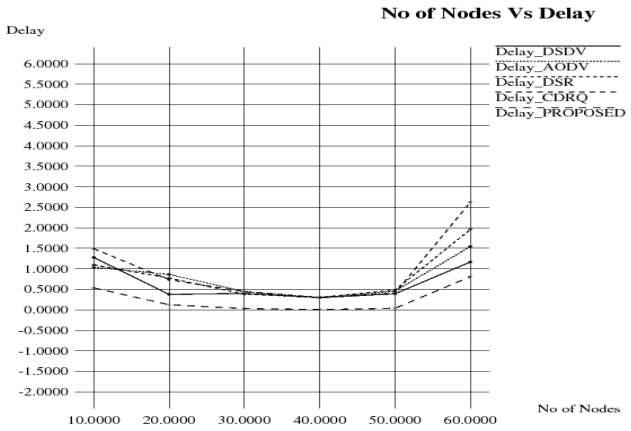

Figure 9. No of Nodes vs. Delay

Table 6. No of Nodes vs. Delay for MANET with Random Mobility

No of Nodes	10	20	30	40	50	60
AODV	1.038	0.864	0.446	0.296	0.455	1.544
DSDV	1.276	0.378	0.397	0.298	0.392	1.167
DSR	1.096	0.775	0.384	0.302	0.484	1.958
CDRQ	1.491	0.736	0.443	0.301	0.406	2.623
PSRL	0.536	0.123	0.032	0.007	0.038	0.808

5. Conclusion

In this paper, various reinforcement learning algorithms were presented. Prioritized Sweeping Reinforcement learning method is compared with existing routing protocols such as DSDV, AODV, and DSR and also compared with CDRQ protocol. PSRL method shows prominent results as compared with shortest path routing for medium and high load conditions. At high loads, PSRL method performs more than twice as fast as CDRQ Routing. Packet delivery ratio and average packet delivery time are used to decide the reliability of PSRL method. In our simulation environment PDR and delay in PSRL method outperforms AODV and DSR routing protocols with almost 90-95% without packet loss with lower delay.

References

[1] Jogendra kumar. Broadcasting Traffic Load Performance Analysis of 80211 MAC in Mobile Ad hoc Networks MANET Using Random Waypoint Model RWM. *International Journal of Information and Network Security (IJINS)*. 2012; 1(3): 223-227.

[2] Yang Shengju, Shi Shaoting, Zhao Xinhui. Research on Security of Routing Protocols Against Wormhole Attack in the Ad hoc Networks. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2014; 12(3): 2110-2117.

[3] Deni Parlindungan Lumbantoruan. Performance Evaluation of AODV Routing Protocol by Simulation and Testbed Implementation. *TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2015; 16(1).

[4] G vetrichelvi, G Mohankumar. Performance Analysis of Load Minimization In AODV and FSR. *International Journal of Information and Network Security (IJINS).* 2012; 1(3): 152-157.

[5] Anand Prakash. A Comparison of Routing Protocol for WSNs: Redundancy Based Approach A Comparison of Routing Protocol for WSNs: Redundancy Based Approach. *Indonesian Journal of Electrical Engineering and Infromatics*. 2014; 2(1).

[6] Fahimeh Farahnakian. *Q-learning based congestion-aware routing algorithm for onchip network.* 2011 IEEE 2nd International Conference on Networked Embedded Systems for Enterprise Applications. 2011.

[7] Parag Kulkarni. Introduction to Reinforcement and Systemic Machine Learning in Reinforcement and Systemic Machine Learning for Decision Making. 1st edition. Wiley-IEEE Press. 2012: 1-21.

[8] S Nuuman, D Grace, T Clarke. *A quantum inspired reinforcement learning technique for beyond next generation wireless networks.* 2015 IEEE Wireless Communications and Networking Conference Workshops (WCNCW). New Orleans, LA. 2015: 271-275.

[9] MI Khan, B Rinner. *Resource coordination in wireless sensor networks by cooperative reinforcement learning.* IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012. Lugano. 2012: 895-900.
[10] MN ul Islam, A Mitschele-Thiel. *Reinforcement learning strategies for self-organized coverage and capacity optimization.* 2012 IEEE Wireless Communications and Networking Conference (WCNC). Shanghai. 2012: 2818-2823.
[11] Oussama Souihli, Mounir Frikha, Mahmoud Ben Hamouda. Load-balancing in MANET shortest-path routing protocols. *Ad Hoc Networks*. 2009; 7(2): 431-442.
[12] Ouzecki D, Jevtic D. *Reinforcement learning as adaptive network routing of mobile agents.* MIPRO, 2010 Proceedings of the 33rd International Convention. 2010: 479-484.
[13] Ramzi A Haraty, Badieh Traboulsi. *MANET with the Q-Routing Protocol.* ICN 2012: The Eleventh International Conference on Networks. 2012.
[14] S Kumar. *Confidence based Dual Reinforcement Q Routing: An on line Adaptive Network Routing Algorithm*. University of Texas. Technical Report. 1998.
[15] Kumar S. Confidence based Dual Reinforcement Q-Routing: An On-line Adaptive Network Routing Algorithm. Master's Thesis. Austin: Department of Computer Sciences, The University of Texas; 1998.
[16] Kumar S, Miikkulainen R. *Dual Reinforcement Q-Routing: An On-line Adaptive Routing Algorithm.* Proc. Proceedings of the Artificial Neural Networks in Engineering Conference. 1997.
[17] Shalabh Bhatnagar, K Mohan Babu. New Algorithms of the Q-learning type. *Science Direct Automatica 44*. 2008: 1111- 1119.
[18] Soon Teck Yap, Mohamed Othman. An Adaptive Routing Algorithm: Enhanced Confidence Based Q Routing Algorithms in Network Traffic. *Malaysian Journal of Computer*. 2004; 17(2): 21-29.
[19] Rahul Desai, BP Patil. Analysis of Reinforcement Based Adaptive Routing in MANET. *Indonesian Journal of Electrical Engineering and Computer Science*. 2016; 18(2): 684-694.
[20] Moore AW, Atkeson CG. Prioritized Sweeping: Reinforcement Learning With Less data and Less Time. *Machine Learning*. 1993; 13.