Volume: 09 Issue: 02 Pages: 3391-3395 (2017) ISSN: 0975-0290

Maximizing Throughput using Adaptive Routing Based on Reinforcement Learning

Rahul Desai
Research Scholar, Sinhgad College of Engineering, Asst Professor, Information Tech. Department
Army Institute of Technology, Savitribai Phule Pune University, India
Email: desaimrahul@yahoo.com
Dr. B P Patil
Principal, Professor, E&TC Department, Army Institute of Technology, Savitribai Phule Pune University, India
Email: bp_patil@rediffmail.com

ABSTRACT

In this paper, prioritized sweeping confidence based dual reinforcement learning based adaptive routing is studied. Routing is an emerging research area in wireless networks and needs more research due to emerging technologies such as wireless sensor network, ad hoc networks and network on chip. In addition, mobile ad hoc network suffers from various network issues such as dynamicity, mobility, data packets delay, high dropping ratio, large routing overhead, less throughput and so on. Conventional routing protocols based on distance vector or link state routing is not much suitable for mobile ad hoc network. All existing conventional routing protocols are based on shortest path routing, where the route having minimum number of hops is selected. Shortest path routing is non-adaptive routing algorithm that does not take care of traffic present on some popular routes of the network. In high traffic networks, route selection decision must be taken in real time and packets must be diverted on some alternate routes. In Prioritized sweeping 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 MANET with random mobility and 50 nodes fixed grid network. Throughput is used to judge the performance of network. Analysis is done by varying the interval between the successive packets.

Keywords – **DSDV, AODV, DSR, Q Routing, CBQ Routing, DRQ Routing, CDRQ Routing**

Date of Submission: Aug 05, 2017 Date of Acceptance: Sep 30, 2017

I. INTRODUCTION

Information is transmitted in the network in form of packets. Routing is the process of transmitting these packets from one network to another. Different routing algorithms such as shortest path routing, bellman ford algorithms are used. The most simplest and effective policy used is the shortest path routing. 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.

Ad Hoc networks are infrastructure less networks. These are consisting of mobiles nodes which are moving randomly. Routing protocols for an ad hoc network 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 one of the earliest pro-active routing protocol developed for an ad hoc networks[1]. DSDV is the extension of Bellman-Ford algorithm[2]. This protocol uses sequence number to avoid count-to-infinity problem. Every node maintains sequence number in increasing order. In addition, it maintains highest sequence number for every destination in the network. This distance information along with destination sequence numbers are exchanged using routing update message among all neighbor nodes. Ad Hoc on Demand Distance Vector (AODV) routing protocol is on-demand routing protocol. Here the routing tables are used to store routing entries. It uses route discovery process to find the shortest route to the destination [3]. The destination node replies with route response message. Thus, the shortest path is stored in routing tables. There will be a single entry of route available in routing tables. Ad hoc On Demand Distance Vector Multipath (AOMDV) routing

Volume: 09 Issue: 02 Pages: 3391-3395 (2017) ISSN: 0975-0290

protocol is just extension of AODV protocol where multiple entries are stored in routing tables such that if one path fails, another path will be available in routing tables [4]. Dynamic source routing is on-demand routing protocol. Here instead of routing tables, routing caches are used to store routing tables. It also uses route discovery process to find the optimum route to the destination. All intermediate nodes only broadcasts this requests. Only the destination node replies with the response message. Thus the shortest route is stored in routing caches [5].

II. 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]. Fig 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 [8-9].

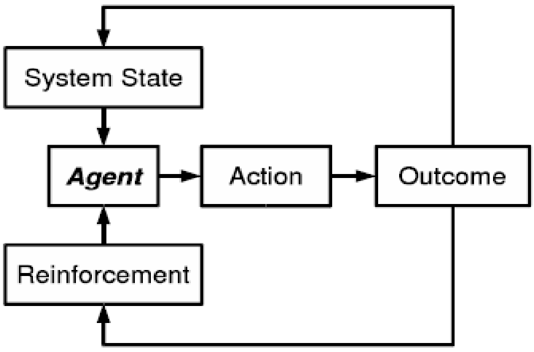

Fig 1: Reinforcement Learning Approach

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 [10-12]. 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 [13-15]. In Q routing, there is no way to specify the reliability of Q values. 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. [14-15]

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 [14-16].

III. 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 [17].

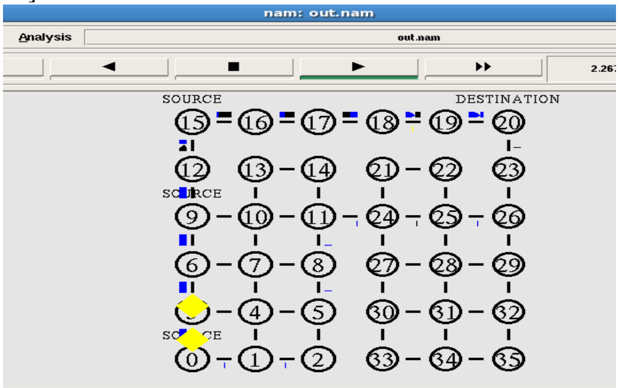

Fig 2: Limitation of Shortest Path Algorithms

For example, in order to demonstrate limitation of shortest path algorithms (fig 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[18,19].

Prioritized sweeping is a method that requires a model of the environment. Model of the environment specifies that agent can use to predict how the environment will respond to its action. This technique is suited for efficient prediction and control of stochastic Markov systems. Agents are used to predict how the environment will respond to its actions. The prioritized sweeping technique makes sweeps through the state of spaces, generating for each state the distribution of possible transactions. It uses all previous experiences both to prioritize important dynamic programming sweeps and to guide the exploration of the state space [19].

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

Volume: 09 Issue: 02 Pages: 3391-3395 (2017) ISSN: 0975-0290

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 [19]. The Q values of the form Qx(*, y) and Qy(*, x) are given a value close to zero when the link R is restored. This causes certain packets to be routed in the wrong direction for a short period of time after a new link becomes available, but more important, the new link will be explored and the routing policy will revert to the optimal routing policy for the new established network state[19]. The fig 3 shows a proposed optimization on CDRQ method.

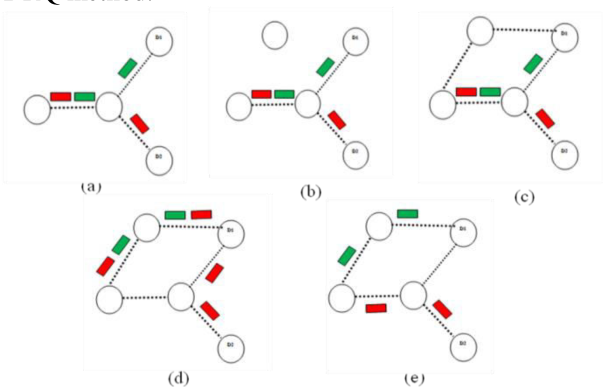

Fig 3: Optimization for CDRQ Routing framework

Fig 4 shows prioritized sweeping technique (PSRL) 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) [19].

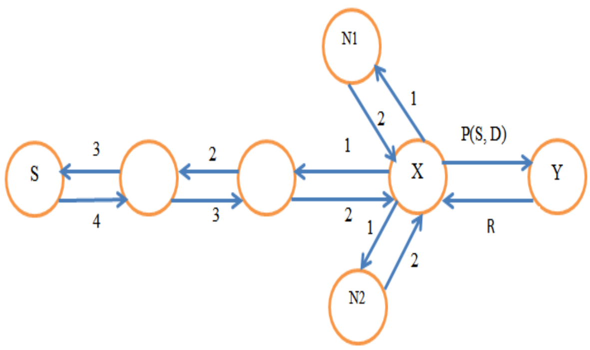

Fig 4: Prioritized sweeping technique for the CDRQ Routing framework

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. [19].

IV. PERFORMANCE ANALYSIS

Simulation always helps in analyzing the design and performance of networks before implementing it in the real application. The various network simulators are available whose output goes as close as possible to real time implementation. In this work, we use the discrete-event simulator NS2 (version 2.34) and the performance analysis is done using AWK script. This experiment is carried on 50 Nodes MANET with random mobility of nodes as shown in Fig 5. The default packet size is 512 bytes. The interval between successive packets varies from 0.1 to 0.2 second. The simulation is carried out for 200 seconds. The various performance parameters are used to judge the quality of network such as packet delivery ratio, dropping ratio, delay and throughput. Throughput is one of most important parameter used to judge the quality of a network. In general terms, throughput is the maximum rate of production or the maximum rate at which something can be processed. In communication terms, network throughput is the rate of successful message delivery over a communication channel. Throughput is the rate at which data is traversing a link while Goodput is the rate at which useful data traverses a link. Fig. 6 refers to interval versus Throughput. Prioritized sweeping CDRQ method is compared with DSDV, AODV, DSR and CDRQ protocols. Table 1 specifies throughput values for different intervals.

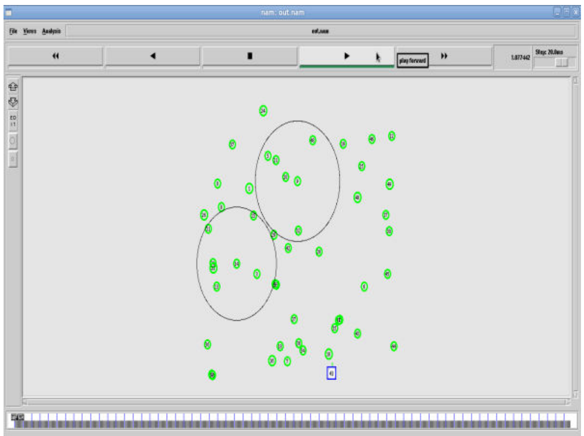

Fig. 5: 50 Nodes Mobile Ad Hoc Network with Mobility

Volume: 09 Issue: 02 Pages: 3391-3395 (2017) ISSN: 0975-0290

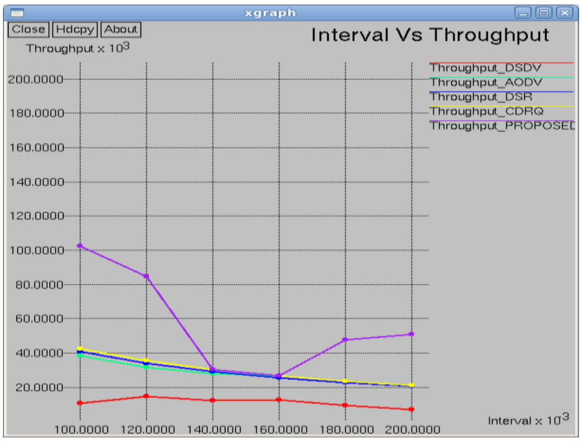

Fig. 6: Interval vs. Throughput for 50 Nodes MANET with Random Mobility

Table 1: Interval (s) vs. Throughput (bps) for 50 nodes MANET

Interval vs. Throughput for 50 Nodes Mobile Ad Hoc Network with Random Mobility						
Interval	**0.1**	**0.12**	**0.14**	**0.16**	**0.18**	**0.20**
AODV	38325	31445	27854	26578	23580	21237
DSDV	10437	14739	12216	12455	9458	7058
DSR	40890	33992	29177	25459	22656	20444
CDRQ	42495	35466	30378	26621	23344	21237
PSRL	102299	84568	30378	26621	47678	50965

The experiment is also carried on 50 nodes fixed grid network with no mobility as shown in Fig 7. The default packet size is 512 bytes. The interval varies from 0.1 to 0.2 second. The simulation is carried out for 200 seconds. Fig. 8 refers to interval versus Throughput. Prioritized sweeping method is compared with DSDV, AODV, DSR and CDRQ protocols. Table 2 specifies throughput values for different intervals.

Fig. 7: 50 Nodes Fixed Grid with No Mobility

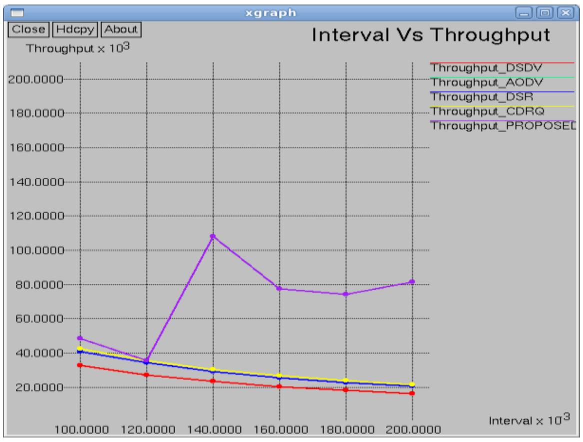

Fig. 8: Interval vs. Throughput for 50 Nodes Fixed Grid

Table 2: Interval (s) vs. Throughput (bps) for 50 nodes Fixed Grid

Interval vs. Throughput for 50 Nodes Fixed Grid						
Interval	**0.1**	**0.12**	**0.14**	**0.16**	**0.18**	**0.20**
AODV	42560	35488	30421	26621	23665	21280
DSDV	32593	27167	23319	20373	18146	16296
DSR	40960	34153	29277	25620	22776	20480
CDRQ	42560	35488	30421	26621	23665	21280
PSRL	48275	35488	107929	77424	74091	81491

V. CONCLUSION

In this paper, various reinforcement learning algorithms were presented. Prioritized Sweeping Confidence Based Dual Reinforcement Learning method is compared with existing routing protocols such as DSDV, AODV, and DSR and also compared with CDRQ protocol. Prioritized Sweeping Confidence Based Dual Reinforcement Learning method shows prominent results as compared with shortest path routing for medium and high load conditions. Throughput is analyzed by varying the interval between successive packets. It is observed that throughput is highly increased in the proposed method as compared with existing routing protocols such as DSDV, AODV and DSR.

REFERENCES

[1]M. Imran and M. A. Qadeer, "Evaluation Study of Performance Comparison of Topology Based Routing Protocol, AODV and DSDV in MANET," *2016 International Conference on Micro-Electronics and Telecommunication Engineering*, Ghaziabad, 2016, pp. 207-211.

[2]C. Cheng, R. Riley and S.P.R. Kumar, “A loop-free extended Bellman–Ford routing protocol without bouncing effect” , *Proc. of ACM SIGCOMM Conf.* , 1989, pp. 224–236.

[3]M. K. Marina and S. R. Das, “Ad-hoc on-demand multi-path distance vector routing,” *Wireless Communication. Mobile Computing*, vol. 6, no. 7, 2006, pp. 969–988

[4]C. E. Perkins, E. M. Royer, and S. Das, “Ad hoc on-demand distance vector routing," document RFC 3561, IETF, Oct. 2003

[5]C. Liu, Y. Shu, and Y. Zhou, et al., “A comparison of DSR, MSR and BSR in wireless ad-hoc networks,” *SPIE*, vol. 6011, 2005, pp. 601–610.

[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, 12/2011

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

[8]S. Nuuman, D. Grace and 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, pp. 271-275.

[9]M. N. ul Islam and A. Mitschele-Thiel, "Reinforcement learning strategies for self-organized coverage and capacity optimization," 2012 IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, 2012, pp. 2818-2823.

[10]Oussama Souihli, Mounir Frikha, Mahmoud Ben Hamouda, "Load-balancing in MANET shortest-path routing protocols", Ad Hoc Networks, Volume 7, Issue 2, March 2009, Pages 431-442

[11]Ouzecki, D.; Jevtic, D., "Reinforcement learning as adaptive network routing of mobile agents," MIPRO, 2010 Proceedings of the 33rd International Convention , pp.479,484, 24-28 May 2010

[12]Ramzi A. Haraty and Badieh Traboulsi “MANET with the Q-Routing Protocol” ICN 2012 : The Eleventh International Conference on Networks

[13]S Kumar, Confidence based Dual Reinforcement Q Routing : An on line Adaptive Network Routing Algorithm. Technical Report, University of Texas, Austin 1998.

[14]Kumar, S., 1998, “Confidence based Dual Reinforcement Q-Routing: An On-line Adaptive Network Routing Algorithm, “Master’s thesis, Department of Computer Sciences, The University of Texas at Austin, TX-78712, USA Tech. Report AI98-267.

[15]Kumar, S., Miikkulainen, R., 1997, “Dual Reinforcement Q-Routing: An On-line Adaptive Routing Algorithm,’’ Proc. Proceedings of the Malaysian Journal of Computer, Vol. 17 No. 2, December 2004, pp.21-29

[16]Artificial Neural Networks in Engineering Conference.

[17]Shalabh Bhatnagar, K. Mohan Babu “ New Algorithms of the Q-learning type” Science Direct Automatica 44 (2008} 1111-1119.
Website: www.sciencedirect.com

[18]Soon Teck Yap and Mohamed Othman, “An Adaptive Routing Algorithm: Enhanced Confidence Based Q Routing Algorithm in Network Traffic.

[19]Rahul Desai, B P Patil, “Analysis of Reinforcement Based Adaptive Routing in MANET”, Indonesian Journal of Electrical Engineering and Computer Science Vol. 2, No.3, June 2016, pp.684-694

[20]Moore, A.W., Atkeson, C.G., Prioritized Sweeping: Reinforcement Learning with less data and less time. Machine Learning, Vol. 13, 1993

Author Biography

Rahul Desai received his Bachelor of Engineering degree and Masters in engineering degree from Pune university. He is currently pursuing Ph.D. from Pune University, Sinhgad College of Engineering as a research center. Presently working as Asst Professor, Dept. of Information Technology in Army Institute of Technology, Pune, India. He has published 25 plus research papers in various international and national referred journals and conferences.

Dr. B.P. Patil received received Ph.D. in Electronics Technology from Guru Nanak Dev Univ., Amritsar India in year 2000. Presently working as Professor, Dept. of E&TC Army Institute of Technology, Pune, India. He has published 120 plus research papers in various international and national referred journals and conferences. He is having 25 years of teaching and industry experience.