Research of QoS Routing Algorithm in Ad Hoc Networks based on Reinforcement Learning

Yuchen Fu[1,2], Quan Liu[2]
[1]*School of Computer Science and Technology, Soochow University, Suzhou 215006, P.R. China*
[2]*Department of Information Technology, Suzhou Industrial Park Institute of Services Outsourcing, Suzhou 215123, P.R. China*
yuchenfu@suda.edu.cn

***Abstract*—With the prevalence of multimedia application, it has become a research focus to provide QoS in ad hoc mobile network. According to the features of recent routing algorithms over ad hoc network, such as the discrete, bimodal model for links between nodes, a new routing algorithm called SNLQ is proposed based on continuous link model with reinforcement learning literature. More concretely, it moves the method of calculating Q-values onto link-values with an eye to a combination of the fixed-time retransmissions mechanism in 802.11MAC and the continuous state-values and Q-values in reinforcement learning. Different scenario-based performance evaluations of the protocol in NS-2 are presented. The results show that our algorithm effectively improves the link table and considerably increases the packet delivery ratio which is superior to AODV and DSR in the congested wireless networks.**

***Index Terms*—QoS, ad hoc networks, routing protocols, routing algorithms.**

I. INTRODUCTION

Ad hoc mobile network is a multi-hop temporarily autonomous system of mobile nodes with wireless transmitters and receivers without any network infrastructure. Through wireless connection, these mobile nodes as routers can constitute any network topology which can work independently and also can connect with the Internet or cellular wireless network. With the development of technology and society, Ad hoc network has attracted considerable attentions in many application fields, including both military and civilian areas. Many routing protocols have been proposed without taking Qos of generating paths into consideration yet. In 2006, Ku-Lan Kao et al. made a complete comparison among existing routing protocols for video transmission over ad-hoc networks. Findings include: DSR can only meet general requirement without ensuring consistency and fluency of the video data, DSDV is not suited for the requirements of multimedia services, and AODV with small delay jitter and short convergence time is particularly in favor of Real-time video data transmission. None of abovementioned protocols can effectively ensure the demand on real-time transmission, as their average delayed time exceeds the range in QoS [1]–[9]. Qos extensions for existing routing have been developed, which can divided into three categories: QoS Extensions for AODV [10]; QoS Extensions for OLSR [10], [11]; QoS Extensions for DSR [11].

Recently some studies have begun to pay attention to provide QoS in ad hoc mobile network and a number of QoS routing protocols over ad-hoc networks have been put forward [1]–[4]. Paper [1] proposed an ant-colony-optimization based multi-QoS constraint routing algorithm for mobile ad hoc networks which introduced the concept of entropy scale as a characterization of routing stability in global information updated policy. Then paper [2] proposed a path stability based QoS routing protocol which introduced path stability factor PSF to consider the stability problem of a feasible path.

Through analyzing technical details of link model based reinforcement learning routing algorithm for Ad Hoc, we propose a statistical metric based link mode against the discrete bimodal link model of existing routing algorithm, and map Ad Hoc routing decision problem to reinforcement learning routing algorithm by Markov decision process(SNLQ). Simulation results show that our algorithm is superior to AODV and DSR in the congested wireless networks.

II. ROUTING MODEL

QoS routing is a discovery process to find a path of minimum cost from the source to the destination under certain constraints. In order to facilitate the analysis, there are general definitions[3]:

Definition 2.1. A network topology is represented as a weighted graph $G = (V, E)$, where V denotes a finite nonempty node set and E denotes a set of two-way links connecting the nodes in V. Assuming that the effective emission among nodes in Ad Hoc Networks is equidistant, we call two nodes neighbors and interlink them if they are in each other's launch range. For any link $e \in E$, $e=(V_i,V_j)$. Meanwhile, each node exchanges the Hello message periodically to maintain the dynamic network topology. If

Manuscript received March 19, 2012; accepted May 22, 2012.
This work was supported in part by a grant from The National Natural Science Foundation of China (61070122); Jiangsu Provincial Key Laboratory for Computer Information Processing Technology (kjs1024); JiangSu Province Support Software Engineering R&D Center for Modern Information Technology Application in Enterprise (SX200902).

the node is out of coverage, the link to it is deleted. Then routing table is updated using the information of neighbor nodes, and a new link is added to maintain the network topology.

Definition 2.2. For any link $e \in E$, the QoS characteristic value is represented as a forth tuples (delay(e, bandwidth(e), jitter (e), cost(e)), where delay( ), bandwidth( ), jitter( ) and cost( ) represent the delay function, bandwidth function, jitter function and cost function respectively.

Definition 2.3. If p(s, d) denotes a path from the source node s to the destination node q and e is a link on the path, there is following relations:

$$Delay(p(s,d)) = \sum_{e \in p(s,d)} delay(e) + \sum_{n \in p(s,d)} delay(n), \quad (1)$$

$$Bandwidth(p(s,d)) = \min\{b(e)\}, \quad (2)$$

$$Delayjitte\ r(p(s,d)) = \sum_{e \in p(s,d)} delayjitte\ r(e) + \sum_{n \in p(s,d)} delayjitte\ r(n), \quad (3)$$

$$Delayjitte\ r(p(s,d)) = \sum_{e \in p(s,d)} delayjitte\ r(e) + \sum_{n \in p(s,d)} delayjitter(n), \quad (4)$$

$$Cost(p,d) = \sum_{e \in p(s,d)} cost(e) + \sum_{m \in p(s,d)} cost(m), \quad (5)$$

where (1) denotes that the path delay is the total delay of all nodes and links on the path, (2) denotes that the path bandwidth is the shortest link bandwidth on the path, (3) denotes that path delay jitter is equal to the total delay jitter of both nodes and links on the path, (4) denotes that the path cost is equal to the total cost of both nodes and lins on the path[4].

QoS routing is to discover a path p(s, d) satisfying the following constraints:

1) Delay: constraint delay (p(s, d))<<D, where D denotes the maximum delay.
2) Bandwidth constraint: bandwidth (p(s, d))>>B, where B denotes the bottleneck bandwidth.
3) Delay jitter constraint: delay_jitter (p(s, d))<<J, where J denotes the maximum delay jitter.
4) Packet loss constraint: packet_loss (p(s, d))<<L, where L denotes the most large package loss rate under certain constraints with smallest cost(p(s, d)).

The QoS route is the path which satisfies the above four constraints and has the smallest value of cost($p(s, d)$), where B, D, L and J respectively denotes the band width, the delay, the packet loss rate and the delay jitter on the path $p(s, d)$[1], [2].

As to the design and implementation of outlier mining, we have put forward the algorithm QOD(quick outlier detection) based on the definition of outliers [6].When using outlier mining method based on the distribution to process the data stream, we have to implement it through an iterative manner. So does QOD. Thus it seems to be slow. Another defect may be the inability to run efficiently in limited memory, so when the amount of data is too huge, we probably have not enough memory. Aiming at solving these problems, this paper optimizes QOD from two aspects.

III. SNLQ: QoS Routing Algorithm with Reinforcement Learning

A. Description of SNLQ protocol

This section lists the state, action, transfer and reinforced value in system model and models the Ad Hoc routing protocols as a reinforcement learning problem.

Statistical link model in Ad Hoc network. It's difficult to use the initial link value of to represent the dynamic link quality in the Ad Hoc network. Some literatures also show that no direct relationship between the signal strength and link quality is found. To estimate the link quality, we take advantage of transmission data between nodes to estimate the link quality in certain time windows. In NS2 it's easy to obtain the following data: the number of data packets to be sent NA, the number of packets successfully received NS, the number of data packets to be failed sent NP, the number of forward packets received NR, and the number of broadcast packets received NB. The ratio of these nodes information is used to estimate the probability of successfully sending a data packet through a link

$$E\left(\frac{N_S}{N_A}\right) = \frac{N_S + \alpha\beta\,(N_R + N_B + N_P)}{N_A + \beta\,(N_R + N_B + N_P)}. \quad (6)$$

Parameter α denotes the reliability of the probability of successfully transmitting data packet, and parameter β denotes the ratio of received and forwarding data packets . In the following experiments, these two parameters are 0.5 and 0.2 respectively.

Statistical link based Ad Hoc network link model. Return function reflects the state transition which depends on the system's action and state. In the reinforcement learning based Ad Hoc routing model, the choice of action is to decide which node to transmit data packets.

Ad Hoc routing protocols SNLQ is based on 802.11MAC protocol, so the node W knows whether the transmission is success or not in the unicast between *W-P* nodes. Our routing model uses U (W, P) to represent the unicast packet action from node P to node W.

If unicast is successful, the transfer from W to P occurs and the system model RS is used to mark the return value. Otherwise, the packet is still in the state W and Rf is used.

However, a packet may be forwarded or discarded when it is accepted on a node. In the reinforcement learning model the packet has not been modeled loss as an action in the system, that is to say, the packet be will discarded only when its TTL is 0. The packet transmission takes some network resources regardless of whether it is successful or not in the wireless data transmission network, so signal channels must be preserved to the data transmission and receiving nodes must transmit packet to affirm the acceptation. These transmission processes must complete channels with other nodes in the network. Failed transmissions will occupy lots of channels, because the retransmission mechanism in 802.11 will be activated for the maximum number of retransmissions.

The return value of the system model denotes the time required to transmit data packet. The negative return value

represents the cost value. The allocation of fixed return values reflects the packet transmission success or failure. In 802.11, the data packets can be retransmitted seven times, otherwise the data transmission is defined to be failure, then our model distributes -70 to return value.

There is a discount factor in our cumulative return model when calculating the value of cumulative return function

$$V(\mathrm{s}) = \mathrm{E}(\sum_{\mathrm{t}}^{\infty} \mathrm{r_t}). \tag{7}$$

Given the estimate model $T(s, a, s')$ and $R(s, a)$ to routing optimization problem, the optimal value function can be obtained by solving the Bellman equation

$$V(\mathrm{s}) = \max_{\mathrm{a}}[\sum_{s'} T(s,a,s') \cdot (R(s,a,s') + V(s'))] = \\ = \max_{a} Q(s,a). \tag{8}$$

The transfer process only two results, success and failure, so the calculation for the Q value is very simple, for example the next hop is P, then Q value is

$$Q(\mathrm{N},\mathrm{P}) = p_s[r_s + V(P)] + p_F[r_F + V(N)], \tag{9}$$

where Ps is the probability of successful transmission to P and Pf when fail. Due to

$$V(N) = \max_{a}[p_s(r_s + V(p)) + p_F \mathrm{r}_F + p_F \mathrm{V(N)}]. \tag{10}$$

The following formula can be given

$$V(\mathrm{N}) = \max_{\mathrm{a}}[\frac{p_s(r_s + V(P)) + p_F r_F}{1 - p_F}] = \\ = \max_{a}[V(P) + r_s + \frac{\mathrm{p}_F}{\mathrm{p_s}} r_{\mathrm{F}}]. \tag{11}$$

Reinforcement learning model in SNLQ. Reinforcement learning method is a solution for Markov processes. According to characteristics of Ad Hoc networks, we make the following changes on the basic elements in Markov process [5]:

1) Set of states: each node has a state set s_i= {*W*, *U*, *D*}, where W represents that a data packet is waiting to be sent, *U* represents that a packet has been successfully sent to the destination node, *D* represents that a data packet has been forwarded to a node;

2) Set of actions: the collection of actions for node n_i, A_i = A_{ni}∪{deliver}∪{broadcast}. A_{ni}=$UN_i(v_1)$,…, $UN_i(v_m)$, where m denotes the number of neighbors of node ni and $UN_i(v_j)$ denotes the unicast from node ni to node nj. Actions in {deliver} represent transferring data to the current node, and actions in {broadcast} are to discover new neighbors;

3) Return function: return value indicates the wireless link transmission interval. Negative return value indicates the time cost in transferring data packets. We give the return value with fixed value whether the transmission is success or not. Because 802.11MAC protocol can not distinguish whether the failure transmission is caused by getting out of the transmission scope or due to interference, the protocol will confirm the failure transmission until retransmitting seven times. The return value is set to -10 when transmission successes, and to -70 when fail.

4) State transition function: the state transition for the successful transmission is: $p_{ijs} = P_i(P \mid B, a_j)$, and the state transition for the failing transmission is: $p_{ijf} = P_i(B \mid B, a_j) = 1 - P_i(P \mid B, a_j)$. In the forwarding process, if the current node is the destination node, the state transfer function is: $P_i(D \mid B, deliver) = 1$. The transfer probability of other states is: $P_i(S' \mid S, a) = 0$.

Algorithm(Fig. 1).

```
Calculate state V-value()
{
State value = max Q-value in (next-hop node set) where {Q-value = Calculate Q-value (next-hop node); }
Return state-value;
}
Calculate action Q-value(next-hop node)
{
Estimate value = time of recession (lastAdvertised of next-hop node, lastAdvertisedTime of next-hop node);
Probability of success = obtain Probability Ofsuccess (next-hop node);
Q-value := estimate value + unicastSuccessfulReturn + unicastFaithfulReturn × (1− probabilityOf success) / probabilityOffail ;
if (Q-value < minimum threshold)
Q-value := minimum threshold;
return Q-value;
}
```

Fig. 1. Algorithm 1 Calculate V-value and Q-value in reinforcement learning model.

B. SNLQ routing algorithm implementation

Packet Format. The format of each packet Transmission in the protocol is shown in Table 1, some of them may only contain header without data content yet [8].

Exploration action in routing. In the reinforcement learning system, whether the strategy is best or not depends largely on whether the system is explored with a reasonable action. On the other hand, the strategy is decided on the quality of the system model. Therefore, adequate sampling must be selected to build a good model.

In the system model, the balance between exploring and exploiting strategies is the key problem in reinforcement learning. In the routing model, SNLQ protocol does not always choose the best action, but also the second-best action according to the probability in order to achieve strategy exploration.

In the 802.11 wireless networks, a unicast packet needs a destination address. Due to the mobility of Ad Hoc network where the neighbor nodes are constantly changing, so it's not possible to make the MAC address of the entire network space as the possible action for each state. In SNLQ protocol, we introduce a special action called explore action into each state. This action is realized by a broadcast packet.

Because the Q-value represents the exploiting strategies, this action has not been calculated by valued function.

TABLE I. DATA PACKET FORMAT.

Fields	Content
OriginAddr	Source address of data packet
DestinationAddr	Destination address of data packet
SequenceNum	Identifier generated at the source
SourceValue	$V_{SRC}(N)$
DestinationValue	$V_{DEST}(N)$
HadError	True if previous transmission failed
TTL	Time to live
IpPacket	Packet content

SNLQ chooses an action according to Boltzmann [7] action selection probability, which is to endow the exploration action with a return value by adding a fixed value to current node V. When a node has no other actions except the selected exploration action, it will generate a flood behavior to find neighbor nodes. Then if the neighbor node has a route to the destination node, it is very likely to be selected.

Action feedback mechanism. In the 802.11 wireless networks, a packet transmission goes through three stages, including conflict detection, recognition and retransmission. the cost of data packets retransmission is much greater than increasing the packet size. We increase the size of the packet to carry the routing information. Each time a node transmits a data packet, the data packet carries the estimated optimal value V of the source and destination node. The Q value and V value of the source node S and destination node can be obtained when any neighbor receives the packet node. This process makes full use of each transfer process.

When one action is selected in the transfer process without feedback information, SNLQ will give the action a negative feedback.

Due to the mobility of Ad Hoc network, system state is changing over time, so we need a mechanism to change the static information of the model. Therefore SNLQ adopts a exponential decay mechanism which is similar to the information decay mechanism in ant colony algorithm. Beginning with latest transmission for the of packets, each node determines the Q values of adjacent points by the rate of this decay. So the nodes don't transfer data to notify the decline of the value of adjacent points which are assumed to follow this exponential decay mechanism

IV. EXPERIMENTS

The statistical network link model based Reinforcement Learning routing protocol SNLQ is implemented in the NS-2 simulator. The simulation arena of 670m x 670m with 50 mobile nodes is used, with the transmission power of the radio interfaces set to 250m. Random way-point mobility model is used, with a maximum speed of 20 m/s and varying pause times. Constant bit rate traffic of 64 byte packets, 4 packets a second, with 10 flows between random pairs of nodes is used.

Performance is measured according to following metrics:

1) Packet Delivery Ratio: This is the ratio of destination node which are successfully received data and the source node.

2) Packet Delivery Cost. This is the ratio of the number of packet transmissions made to the number of packets delivered.

The first experiment verifies the impact on protocol performance by different parameters. we select the following two parameters:

– TEMPERATURE parameter, used to control boltzmann action selection;

– EXPLORATIONUTILITY parameter, used to impact the selection of the exploration action.

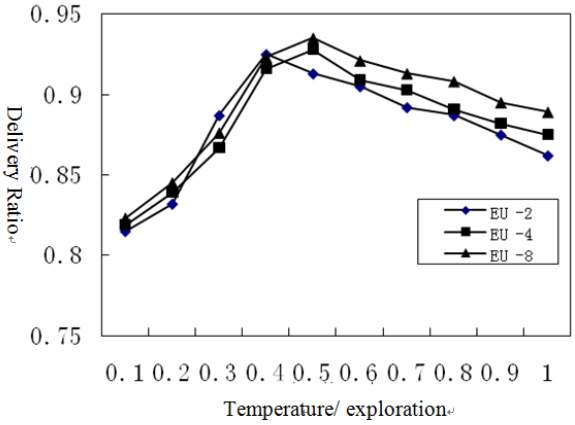

Fig. 2. Delivery Ratio.

Fig. 2 shows that the exploration action value and the Boltzmann action selection is closely related to each other. More negative values of exploration utility will cause the exploration action to be taken less often. Larger values for temperature will cause the exploration action to be taken more often. However, increasing values of temperature will also increase how often other sub-optimal actions will be taken. As can be seen in Fig. 2, the lower the values of T/u is, the worse the protocol performance is, as exploration is perform not enough in the environment. Without exploration, agents do not discover new or better routes until their old ones break. Conversely, with high values of T/u the routing agents do not sufficiently exploit their routing information as they spend too much effort on exploration.

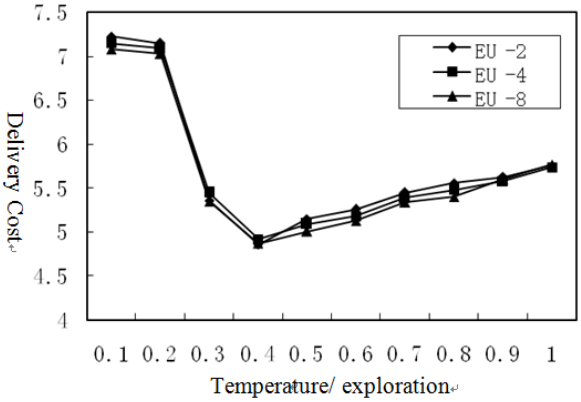

Fig. 3. Delivery cost.

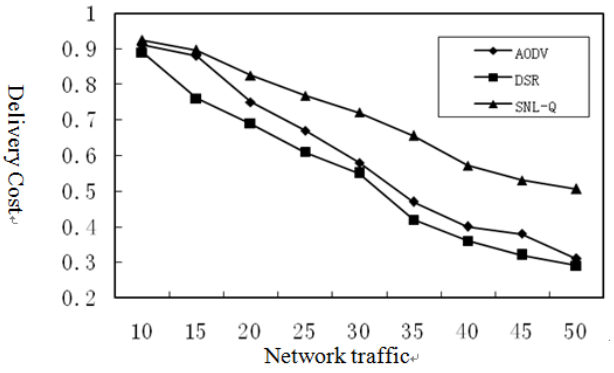

Fig. 4. SNLQ versus AODV&DSR.

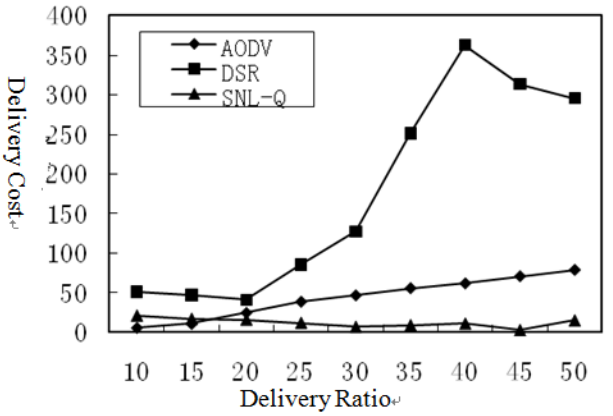

Fig. 5. Delivery cost.

In the second experiment, the performance of the SNL-Q protocol is analyzed by comparison with AODV and DSR. These experiments are carried out using NS version 2.31. The versions of AODV and DSR used were those supplied with NS2. We compare the transmission rate and transmission costs by changing the flow of the node, and change the traffic of source node by changing the package size and CBR. As shown in Fig. 3 and Fig. 4, the perform of SNLQ is better than AODV and DSR in the congested network.

V. CONCLUSIONS

In this paper, we proposed a new routing algorithm called SNLQ based on continuous link model with reinforcement learning literature. More concretely, it moves the method of calculating Q-values onto link-values with an eye to a combination of the fixed-time retransmissions mechanism in 802.11MAC and the continuous state-values and Q-values in reinforcement learning. Different scenario-based performance evaluations of the protocol in NS-2 are presented. The results show that our algorithm effectively improves the link table and considerably increases the packet delivery ratio which is superior to AODV and DSR in the congested wireless networks.

REFERENCES

[1] Xiang Yang, Li Layuan, Cheng Chuanhui, "Application Research Based Ant Colony optimization for MANET", in *Proc. of the Wireless Communications, Networking and Mobile* (WiCOM 2006) Conf., 2006.

[2] Peng Ge-Gang, Yang Jiang-Hu, Gao Chuan-Shan, "A QoS Routing Protocol Based on Path Stability in Mobile Ad Hoc Network", *Journal of Computer Research and Development,* vol. 41, no. 6, pp. 916–922, 2004.

[3] Liu Yong-Qiang, Yan Wei, Dai Ya-Fei, "A QoS-Aware Adaptive Multipath Routing Protocol in Ad Hoc Network", *Chinese Journal of Computers,* vol. 29, no. 5, pp. 681–689, 2006.

[4] Guo Xiao-Feng, Chen Yue-Quan, Chen Gui-Hai, "An Aggregated Multipath Routing Scheme for Ad Hoc Networks", *Journal of Software,* vol. 15, no. 4, pp. 594–603, 2004.

[5] E. Curran, J. Dowling, "SAMPLE Statistical Network Link Modeling in an On-Demand Probabilistic Routing Protocol for Ad Hoc Networks", in *Proc. of the Second Annual Conference on Wireless On-demand Network Systems and Services (WONS 05),* 2005, pp. 200–205. [Online]. Available: http://dx.doi.org/10.1109/WONS.2005.30

[6] E. Curran, "SWARM Cooperative Reinforcement Learning for Routing in Ad-hoc Networks", M.S. thesis, University of Dublin, Trinity College.

[7] S. Ziane, A. Melouk. "A Swarm Intelligent Scheme for Routing in Mobile Ad hoc Networks", in *Proc. of the 2005 Systems Communications (ICW 05),* 2005, pp. 2–6. [Online]. Available: http://dx.doi.org/10.1109/ICW.2005.18

[8] Sun Qiang, Li LaYuan Chen NianSheng, "A Power Control Algorithm Based on Game Theory in Ad Hoc Networks", *Chinese Journal of Computers,* vol. 32, no. 1, 2009.

[9] Ku-Lan Kao, Chih-Heng Ke, Ce-Kuen Shieh, "Video Transmission Performance Evaluation of Ad Hoc Routing Protocols", in *IEEE Proc. of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 06),* 2006.

[10] M. Dassr, "On-demand multipath distance vector routing in Ad Hoc networks", in *Proc. of the ICNP,* Washington, DC: IEEE Computer Society, 2001, pp. 14–23.

[11] Zhang Xu, Cheng Sheng, Feng Mei-Yu, et al., "Fuzzy logic QoS dynamic source routing for mobile Ad Hoc networks", in *Proc. of the 4th International Conference on Computer and Information Technology,* IEEE, 2004, pp. 652–657.