社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2005－EVA－12（3）
2005／3／23

アドホックネットワークのための Q-routing Protocolの性能評価

Tang Tao　田頭 茂明　藤田 聡
広島大学大学院 工学研究科

本稿では、アドホックネットワークのための新しいルーティングプロトコルを提案する。提案手法は、Q-Routingを基にしたプロアクティブ型のプロトコルで、アドホックネットワークが持つホストの移動性と動的なネットワーク形状の変化に自律的に適応する。具体的には、現在のネットワークにおける最適なルーティングポリシーを教師なし学習により学習し、その学習結果として経路表を各ホスト上に作成する。その経路表に従ってルーティングを行うことにより、パケットは現在のネットワークに適切なルートにより配送されることになる。本稿では、提案手法をNS-2シミュレータ上に実装し、実験により評価する。

Performance Evaluation of Q-routing Protocol for Mobile Ad-hoc Networks

Tang Tao　Shigeaki Tagashira　Satoshi Fujita
Graduate School of Engineering, Hiroshima University

In this paper, we proposed a new routing protocol for mobile ad-hoc wireless networks. The proposed protocol is a proactive routing protocol based on Q-routing for autonomously adapting the routing to mobility and topological change inherent in mobile ad-hoc wireless networks. It learns the task of finding an optimal routing policy on the underlying network while unsupervised and constructs a routing table on each host as the learning result. By using the routing table, the delivery of packets can be realized through the appropriate route according to the current network state. Furthermore, we implement the proposed protocol on NS-2 simulator and show the result of several experiments by using it.

1 Introduction

Mobile multiple Ad-hoc Network (MANET) is a wireless network consisting entirely of host computers that can communicate with each other without the aid of any fixed infrastructure such as wired networks and base stations, even while they continuously change their physical locations that may change logical configuration of the overall network. Since it does not rely on any static infrastructure as in cellular phone systems (i.e., since it is constructed in an *ad hoc* manner), MANET has attracted considerable attentions as a tool to realize flexible and robust communications in many application fields such as private communication among participants to unplanned meetings, cooperated disaster recovery by rescue robots, secret operation in battle fields, and so on.

In MANET, a mobile host can directly send a message to the other host if they are "visible" with each other in the sense that the distance between them is smaller than the transmission radius of the equipped wireless communication devices. If they are not visible with each other, on the other hand, a message sent out to the destination host must be routed via several intermediate hosts. In other words, each host in MANET operates not only as a client end-system but also as a router of messages. How to find an appropriate route connecting two communicating mobile hosts is a major problem in MANET, that has been investigated extensively during the past decade [1][2][3].

In this paper, we propose a new routing protocol for MANET. The proposed protocol is a combination of Destination-Sequenced Distance-Vector (DSDV) [2] that is known as a typical proactive routing protocol for MANET and the Q-routing [6] that is known as an adaptive routing scheme based on the notion of reinforcement learning. In DSDV, each host propagates routing updates periodically. A remarkable advantage of DSDV over

traditional vector protocols is that it could avoid the loop of advertisements. In addition, it could adaptively acquire knowledge from the environment that may change dynamically in general MANET's. More concretely, we focus on the Q-value that represents the estimated time that a packet would take to reach its destination mobile host D instead of hop counts that have been considered in traditional protocols to decide the shortest path for a routing request [2]. Moreover, aiming at the mobility of the hosts in MANET, we proposed a scaling scheme for the stability (denoted by *Lifetime*) of the mobile host to obtain steadier route for the routing request preferentially. We implement the proposed protocol on NS-2 simulator [5] and analyze the performance of it.

2 Preliminaries

2.1 DSDV

Let $V = \{1, 2, \ldots, n\}$ denote a set of mobile hosts. Conventional routing protocols for MANET can be classified into three categories, i.e., proactive, reactive, and hybrid protocols [4]. In DSDV [2], that is known as a typical proactive routing protocol, every mobile host maintains its own routing table to store the "next step" for each destination. More concretely, a message destined for host $d \in V$ will be forwarded to the (neighboring) host indicated by the d^{th} entry of the routing table; i.e., the d^{th} entry of the table contains the (neighboring) host on an appropriate route to the destination from the current host. Routing tables are periodically exchanged among nearby hosts, and in addition, it is transmitted to its neighbors when a significant change occurs after the last transmission. More concretely, each host "advertises" routing updates to its neighbors. After receiving the advertisement, the neighbors update their routing tables and retransmit the advertisements to their neighbors. The process will be repeated until all the hosts.

In DSDV, each route is labeled with a sequence number which is originated by the destination host. The main reason of providing a sequence number to each route is the avoidance of possible loop structures. Sequence numbers are used in the following manner: If a host receives multiple advertisements for a same destination, the advertisement associated with the largest sequence number will be preferred (a tie is broken by using an appropriate metric such as the distance and the communication delay). The other advertisements will be discarded.

2.2 Q-Learning

Reinforcement learning (RL) is an unsupervised learning scheme based on the notion of reward received from the environment; i.e., the designer of the scheme could set up rewards to navigate the agent (i.e., learner) to an appropriate goal, where typically, a reward is given to the agent only when it arrives at a certain goal state. In general RL schemes, each agent repeats the following procedure:

1. Based on the observation of the current state, it selects an action from a predetermined set of actions, and takes it.

2. The action causes a state transition of the environment, and the agent receives a reward from it, if any.

In each time step $t \in \{1, 2, \ldots\}$ of the learning process, each agent tries to maximize an expected return R_t received from the environment, that is formally defined as follows:

$$R_t \stackrel{\text{def}}{=} \sum_{k=1}^{T} r_{t+k}, \tag{1}$$

where T is the final time step of the learning process, that is generally assumed to be a finite value, and r_{t+k} denotes an **estimated reward** received from the environment at time $t+k$, where the estimation is carried out at time t. Such an estimation of future rewards would generally be done with the notion of **discount rate** γ (≤ 1). More concretely, r_{t+k} is estimated as being γ^{k-1} times of the reward that could be obtained if it were received immediately at time $t+1$.

Let S denote a finite set of states of the environment, and A a finite set of actions that could be taken by each agent. The Q-learning is an RL scheme based on the notion of Q-values, that represents the "appropriateness" of action a at the current state s. When it takes action a at state s and receives reward r from the environment, Q-value, $Q(s, a)$, is updated as

$$\begin{aligned} Q(s,a) \quad \leftarrow \quad & (1-\alpha)Q(s,a) + \\ & \alpha\{r + \gamma \max_{a'} Q(s', a')\}, \end{aligned} \tag{2}$$

where α is a parameter called the learning rate. In general, Q-values are maintained in the form of a table called **Q-table** with size $|S| \times |A|$, where S is the set of states and A is the set of actions.

2.3 Q-Routing

Q-routing [7] is an adaptive routing scheme inspired by the Q-learning. The basic idea of the Q-routing is to use Q-table instead of routing tables used in conventional routing protocols. More concretely, we associate Q-value $Q(s, a)$ with an appropriateness of the selection "a" of a neighbor as the next step of a route connecting to a given destination "s" from the current host. The appropriateness

of selection "a" is measured in terms of the expected routing time to the destination through the selected neighbor, that could be represented by a sum of the waiting time in FIFO queues at a receiver host and the transmission delay between two neighboring hosts. It should be worth noting that such a routing table is maintained by each host in a distributed manner similar to conventional routing schemes.

Let Q_x be the Q-value maintained by host x. In the following, let q_x denote the waiting time in the queue at host x, and δ denote the transmission delay between two neighboring hosts. By using those notions, we have the following inequality:

$$Q_x(y,d) \ \leq \ q_y + \delta + Q_y(z,d),$$

where y is a neighbor of x and z is a neighbor of y. When a host x receives a packet P destined for d, it selects a neighbor y such that $Q_x(y,d)$ is minimum over its all neighbors, and sends it out to the selected neighbor y. Upon receiving the packet from x, host y sends back its best estimate $Q_y(z,d)$ for the destination d to host x. By receiving the estimated value, host x computes its new estimate for $Q_x(y,d)$ as

$$Q_x(y,d)^{est} \ = \ Q_y(z,d) + q_y + \delta,$$

and by using the estimated value, it updates its Q-table as follows:

$$\begin{aligned} Q_x(y,d)^{new} \ &\leftarrow \ Q_x(y,d)^{old} + \\ &\quad (Q_x(y,d)^{est} - Q_x(y,d)^{old}). \end{aligned}$$

By using the learning rate η_f,

$$\begin{aligned} Q_x(y,d) \ &\leftarrow \ Q_x(y,d) + \\ &\quad \eta_f(Q_y(z,d) + q_y + \delta - Q_x(y,d)). \end{aligned}$$

3 LQ-Routing for Ad-Hoc Networks

3.1 Overview

In this paper, we propose a new routing scheme for MANET. The proposed scheme is a combination of DSDV and the Q-routing, in the sense that it adopts the notion of sequence numbers to avoid possible loop structures as in DSDV, and it adopts Q-value as the metric for selecting the next hop as in the Q-routing. It should be worth noting that although the notion of Q-routing could directly be applied to MANET, it must be followed by an appropriate mechanism to increase the adaptivity of the scheme, since frequent connection and disconnection of a link would significantly degrade the convergence speed of the underlying learning processes. More concretely, although the change of network configurations could ultimately be informed to all hosts in the system by repeating the exchange of advertisement packets, in the framework of the Q-routing, it exchanges expected rewards among nearby hosts that is more sensitive to the accuracy of information compared with direct metrics used in DSDV; i.e., we have to introduce additional mechanisms to explicitly take into account the mobility of hosts in MANET.

To resolve such problems, the proposed scheme introduces the notion of lifetime to represent the stability of connections such as links and routes. We define two kinds of lifetime in the scheme; i.e., hop-lifetime and path-lifetime, that will be abbreviated as HLT and PLT, respectively. HLT is the traffic of a hop until now and PLT is defined as the minimum one of all the HLT's on that path.

In our routing protocol, we combine the notions of the Q-value and the lifetime, and introduce a new metric called LQ-value to make the actual routing decision, ulteriorly, for a routing request, the route of all optional ones, with less LQ-value will be preferred and the current source host will send the packet to the next hop of such route.

Table 1: Structure of the routing table.

Dest	Next	Hops	Seq. No	PLT	Q-Value
x	x	0	308	250	0.001
y	y	1	312	120	0.001
...	...	...	...	...	...
d	y	3	326	120	0.003
d	a	3	334	50	0.004

3.2 Routing Table

In the proposed method, the structure of the routing table is shown as Table 1. The first entry of the routing table is the destination. The second one is the next hop with an estimated number of hops to the destination via the designated next hop. The third one is the estimated number of hops. This entry is used for only keeping the compatibility of DSDV. When a link to the next hop is disconnected, every route passing through the hop is marked for unreachable; i.e., assigned ∞ hops and with an updated sequence number. And building information to describe broken links is the only situation when the sequence number is generated by any host other than the destination. The next entry of the routing table is the sequence number. It could be maintained in the similar manner of DSDV to avoid possible loop structures. Moreover, PLT and Q-value could be maintained in the routing table,

3.2.1 PLT

For each neighbor y of x, HLT of link (x,y), denoted by $H_x(y)$, is defined as the number of successful advertisement transmissions from y to x; i.e., it is initialized to zero when y is disconnected (realized by detecting a predetermined timeout of adver-

tisement transmissions) and is incremented by one for each successful transmission, where by technical reasons, we bound the maximum lifetime by 250.

PLT is used to describe how stable the specific path is. The PLT of a path from x to destination d is computed as follows:

$$PL_x(y,d) \;=\; \min\{PL_y(z,d), H_x(y)\}, \qquad (3)$$

where y is a next hop of x, z is a next hop of y. $PL_d(d,d)$ is fixed to be the maximum lifetime (i.e. 250). $PL_x(y,d)$ could be maintained in the routing table of host x. It would be updated by receiving $PL_y(z,d)$ from y through advertisements.

3.2.2 Q-Value

The estimated packet delivery time (Q-values) described in Section 2.3 can be maintained in the routing table. For applying the Q-routing scheme to ad-hoc networks, it is extended as follows: Let Q_x be the Q-value maintained by host x, q_x^{ave} denote the average waiting time in the queue at host x between two consecutive transmissions of advertisements packets, and δ denote the transmission delay between two neighboring hosts. The Q-value of a path from x to destination d is defined as follows;

$$\begin{aligned} Q_x(y,d) \;\leftarrow\; & Q_x(y,d) + \\ & \eta_f(Q_y(z,d) + q_y^{ave} + \delta - Q_x(y,d)), \end{aligned}$$

where y is a neighbor of x, z is a neighbor of y on the path, and η_f is the learning rate. Host y advertises $Q_y(z,d)$ and q_y^{ave} to all neighboring hosts. After receiving the advertisement, host x updates $Q_x(y,d)$ in the routing table. When a link to the next hop is disconnected, every route passing through the hop is marked for unreachable; i.e., assigned ∞ value and with an updated sequence number.

3.3 Routing Decision

Each host x makes routing decision for an arrival packet with destination $d(\neq x)$ as follows: First, it computes $LQ_x(y,d)$ for each neighbors y as LQ-values and then the packet is forwarded to the neighbor host of the smallest LQ-value. More concretely, let $W_x(y,d)$ be the value defined as follows:

$$W_x(y,d) \;=\; \frac{PL_x(y,d)}{\max_{y'}\{PL_x(y',d)\}}, \qquad (4)$$

where the denominator is the maximum over all paths connecting to the destination d. Then, $LQ_x(y,d)$ is defined as follows:

$$LQ_x(y,d) \;=\; \frac{Q_x(y,d)}{W_x(y,d)}; \qquad (5)$$

If $PL_x(y,d)$ is equal to zero, parameter $W_x(y,d)$ is set to enough small value.

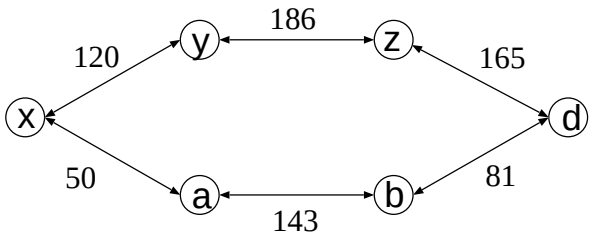

Figure 1: An ad-hoc network.

Table 2: Routing table of x after host a leaves the network.

Dest	Next	Hops	Seq. No	PLT	Q-Value
x	x	0	308	250	0.001
y	y	1	312	120	0.001
a	a	∞	311	0	∞
...	...	...	...	...	...
d	y	3	326	120	0.003
d	a	∞	335	0	∞

3.4 Example of Updating Routing Table

In our proposed routing protocol, any change of the routing table will cause the advertisement to its neighboring hosts. The advertisement packet includes only changed entries of the routing table. For instance, given an ad-hoc network as shown in Figure 1, if host a leaves the network, after detecting a timeout period, the neighboring hosts x and b will update their routing tables and advertise the fact to their neighboring hosts. Table 2 shows the routing table of host x. The sequence number of the route from x to a is increased by one, i.e., from the previous 310 to an odd number 311. The Q-value becomes ∞ and the number of hops is ∞. Host x advertises the changed entries to the neighbors. Furthermore, every route passing through the hop will be updated in the every routing table.

Figure 2 shows a case where a new host c joins in the network. A new host c advertises the fact to the neighboring hosts (i.e., x and b). After receiving the advertisements, the neighboring hosts x and b will create the new entry of host c in their routing tables and initialize it as shown in Table 3. If the entry of host c already exists in the table (i.e., the host rejoins in the network.), the entry is reused and initialized.

4 Simulation

4.1 Simulation

In this section, we evaluate the performance of the proposed schemes by simulation. In particular, we measure the packet loss rate and the average packet delay time for communicating between hosts. The

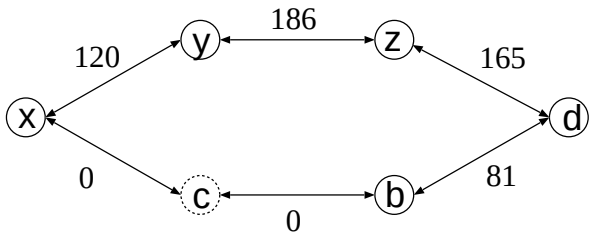

Figure 2: Host c joins in the network.

Table 3: Routing table of x as host c joins in the network.

Dest	Next	Hops	Seq. No	PLT	Q-Value
x	x	0	308	250	0.001
y	y	1	312	120	0.001
a	a	∞	311	0	∞
...	...	...	...	...	...
d	y	3	326	120	0.003
d	a	∞	335	0	∞
c	c	1	2	1	0.000

main objective of simulation is to examine the impact of the mobility , which is connatural characteristic of mobile hosts in the simulative wireless ad-hoc networks, to the performance of congestion of the routing path.

4.1.1 Simulation Environment

Simulation was conducted by using NS-2[5]. A set of hosts are connected by a wireless link for constructing an ad-hoc network. The simulated ad-hoc network uses an area of 670 meter $\times$ 670 meter and each host has a power range of 250 meter. The distributed coordination function (DCF) of IEEE 802.11 for wireless LANs is used as the MAC layer.

Flows are generated dynamically in the network and they are controlled on the UDP protocol. CBR (constant bit-rate) is used as traffic sources and only 512 byte data packets are used. The packet sending rate on each host is fixed to 3 packets/sec. A request of flows arrives at a source hosts according to a Poisson distribution with mean λ, that varies from 100 sec to 1,000 sec in order to change the offered load in the network. The destination of the request is chosen randomly from the set of all hosts except the source vertex. The holding time of a request is exponentially distributed with mean $1/\mu$, that is fixed to $1/\mu = 100$ sec, in our simulation.

The period between advertisements is fixed to 15 sec in the proposed method and the DSDV method. The total simulation time is 1000 sec.

4.1.2 Congestion

We compare the performance of three different versions of LQR (LQR(r) for r = 1.0, 0.85, 0.7, where r

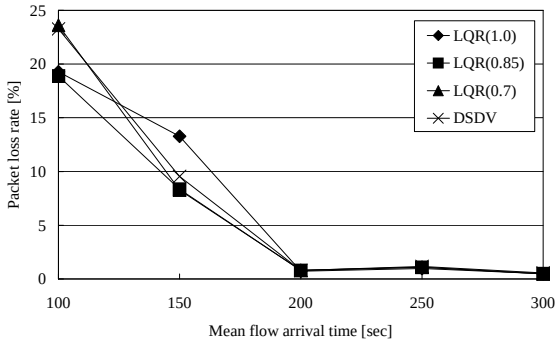

Figure 3: Impact of congestion without mobility to packet loss rate.

denotes the learning rate in Q-Learning and DSDV (traditional technique). The number of hosts is fixed to 50 and all hosts do not move in the simulated area. Figure 3 shows the packet loss rate of the simulation. The horizontal axis of the figure represents mean flow arrival time and the vertical axis represents packet loss rate. Note that more shorter arrival time means more heavily loaded environment. From this figure, we can observe that LQR(0.85) can improve the better performance than the traditional routing technique DSDV. The loss rate could be reduced to about 18% by using LQR(0.85) when mean flow arrival time is equal to 100 sec.

Figure 4 shows the average packet delay time of the simulation. From this figure, we could find that LQR(0.85) can achieve minimum delay time compared to the other techniques. Especially, when mean flow arrival time is equal to 100 sec, the delay time of LQR(0.85) is 1.1 sec, whereas that of DSDV is 2.4 sec.

The main reason of the above phenomena is the effect of Q-routing adopted in the proposed method; i.e, the proposed method will autonomously select an alternate route when the load of the shortest route (selected by DSDV) becomes heavy. Although the alternate route is not the shortest, the state of the route is relatively idle. As a result, the deliver of packets will be realized with low loss rate and short delay.

4.1.3 Impact of Mobility

We evaluate the impact of the proposed methods and DSDV. The number of hosts is fixed to 50 and 5 hosts of them move at average speed 5 meter/sec in the simulated area and the other hosts do not move. Figure 5 shows the packet loss rate and Figure 5 shows the average packet delay. In more locomotive environment, the proposed protocol with the best Learning Rate, 0.85 (see Figure 2), can result in much more better performance relative to the traditional routing protocol DSDV. The reason

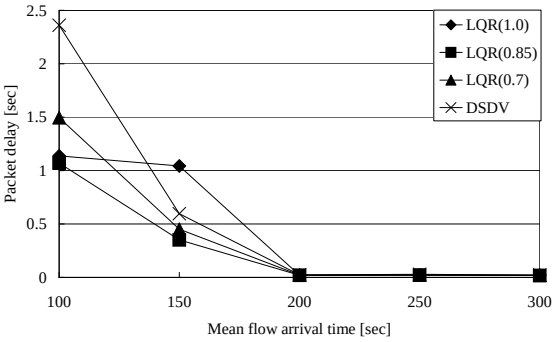

Figure 4: Impact of congestion without mobility to average packet delay.

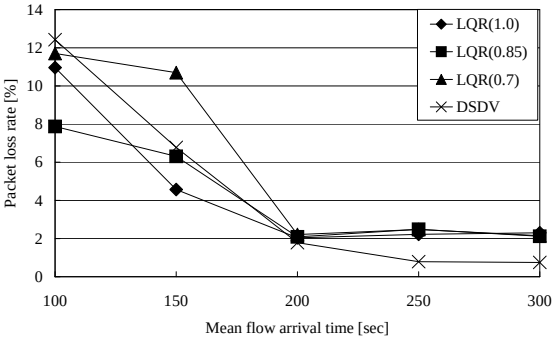

Figure 5: Impact of congestion with mobility to packet loss rate.

is to adopt the stability in the routing decision, in addition to the effect of Q-routing as was described in Section 4.1.2.

5 Conclusions

In this paper, we propose a routing protocol which based on the traditional protocol DSDV in mobile wireless ad-hoc networks and Q-routing protocol. In our proposed routing protocol, for avoiding loop information, we use the sequence number which is used in DSDV. And in the process of routing, mobile hosts make routing decision by using LQ-values, the estimated delivery time, instead of the hops used by DSDV, and use Lifetime to guarantee selecting much more stable route. Aiming at the problem of packet loss and packet delay time which mostly induced by congestion, we simulate the proposed LQ-Routing protocol on NS-2, from the result, we can see that, when the learning rate is fixed to 0.85, the proposed LQ-Routing protocol does get lower packet loss rate and shorter packet delay time even though at a high locomotive environment.

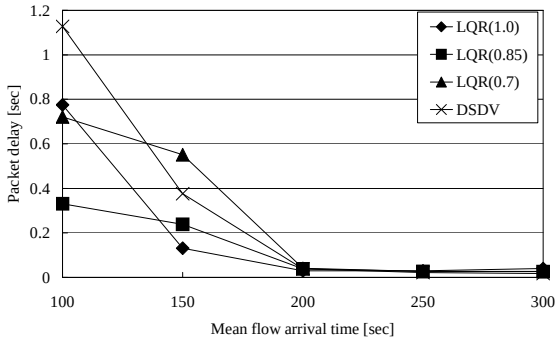

Figure 6: Impact of congestion with mobility to average packet delay.

References

[1] S. Ni, Y. Tseng, Y. Chen, and J. Sheu: The Broadcast Storm Problem in a Mobile Ad Hoc Netowork. In *Proc. MOBICOM*, pp.151-162, 1999.

[2] C. E. Perkins and P. Bhagwat. Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers. In *Proc. ACM SIGCOMM*, pp. 212–225, 1994.

[3] R. Castañeda and S. R. Das. Query Localization Techniques for On-demand Routing Protocols in Ad Hoc Netoworks. In *Proc. MOBICOM*, pp.186-194, 1999.

[4] Z. J. Haas, M. R. Pearlman, and P. Samar. The zone routing protocol (zrp) (draft-ietf-manet-zone-zrp-04.txt). IETF MANET Draft, 2002.

[5] *NS-2 Simulator*, http://www.isi.edu/nsnam/.

[6] M. L. Littman and J. Boyan. A distributed reinforcement learning scheme for network routing. In *Proc. of the First Internation Workshop on Applications of Neural Networks to Telecommunications*, pp.45–51, 1993.

[7] J. Boyan and M. L. Littman. Packet routing in dynamically changing networks: A reinforcement learning approach. Advances in Neural Information Processing Systems 6, MIT Press, 1994.