# A Q-learning Based Routing Protocol for Body Area Networks

Md. Golam Rabiul Alam, Choong Seon Hong
Department of Computer Engineering, Kyung Hee University
robi@networking.khu.ac.kr, cshong@khu.ac.kr

**Abstract**

Body area networks (BANs) have some unique characteristics like severe attenuation, miniaturized size of in-vivo sensors, high reliability and energy efficiency requirement, and higher throughput and low latency requirement etc. Due to meet these inimitable requirements, it is challenging to design network protocols for body area networks (BANs). In this paper, we focused on energy-efficient and minimum hop routing for body sensor networks. Here we used Q-learning approach for exploring and exploiting route from sensor nodes to sink node of body area networks. The proposed Q-learning based routing protocol (QRP) efficiently selects the optimal route with maintaining nearly balanced energy consumption of sensor nodes of BAN. We found normalized energy consumption and minimum hop count in our simulation study in comparison to existing routing protocols of BAN.

## 1. Introduction

Wireless body area networks (WBAN) are an emerging wireless networking technology. It has enormous applications including healthcare, athletic training, workplace safety, consumer electronics, and secure authentication and safeguarding of uniformed personnel [1]. The sensor nodes of BANs communicate with sink nodes using single-hop or multi-hop. Transmission of body sensor node covers a very short range. Although WBANs are deployed in a compact spatial region (along the human body), multi-hop communication rather than single-hop is their main communication pattern [2]. But it is not feasible to design the body area network following the single-hop communication architecture. Thus only the efficient routing strategy can save valuable energy of the small batteries of body sensors.

Conventional routing methods of wireless sensor networks are not suitable for such resource constrained networks. So, we need some estimation method that can produce near-optimal policies without complex computation and underlying model like Q-learning. We used Q-learning approach to learn the working environment as state-action pairs by using reinforcement learning strategy because it learns optimal policy without knowing the probability model. Then based on the learning i.e based on the Q-table it can perform best action to the environment by using optimal policy.

## 2. Related Works

In reference [2], the authors of this paper proposed a distributed prediction based secure and reliable routing framework (PSR) for wireless body area networks. They presented PSR with autoregressive model for prediction. The proposed approach is reliable and impressive in respect to dropping rate and hop count per node. But in this approach the consumed energy per node is mostly imbalanced and as a result operating lifetime of the network is shortened.

An energy-balanced rate assignment and routing protocol (EBRAR) is presented by the authors of reference [3]. They select routes based on residual energy and routing data through an energy-efficient fixed path. They achieved good energy balanced output. But, the data packets of some sensor nodes experience longer path (more hops) to reach at the sink node.

## 3. Description of the Proposed Routing Protocol

In wireless body area networks, we consider two types of sensor nodes as shown in Figure 1. The sensor nodes are connected each other according to their communication range. We also considered that the nodes can dynamically adapt their transmission energy as their communication requirement within the limit.

There are $n$ numbers of sensor nodes in the network. The set of sensor nodes are considered as $X$ = {A,B,C,D,…..P}, where $P$ is the sink node and others are different sensor nodes. The set of nodes which can communicate with each other's is based on the received signal strength (RSS) and defined by an adjacency matrix *Mat*. To deploy Q-learning approach for routing of data packets in our body area network, we consider each of the sensor nodes as a state. Thus the set of states are $\{x_1, x_2, \cdots x_n\} = \{A, B, \cdots P\}$. In each state the node can perform packet forwarding action or search for another path to forward the packets. The data packets may successfully forward or packet failure may occur in that case the node retransmit the packet either in the same path or in different suitable route. So, we consider set of actions as $\{a_1, a_2, \cdots a_n\}$.

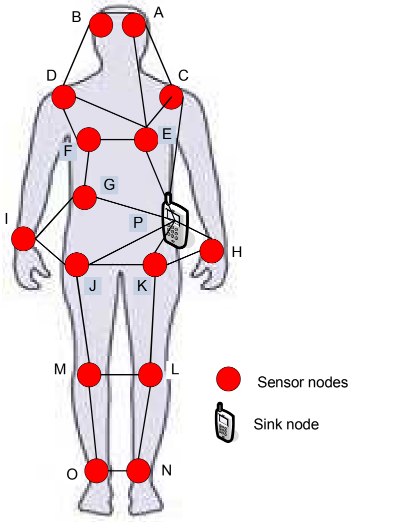


Figure 1. Wireless body area networks

Q-learning worked as state-action pairs $Q(x, a)$. We define the value of taking action $a$ in state $x$ under a policy $\pi$, as the expected return from taking such an action and thereafter following policy $\pi$ [4].

$$Q^{\pi}(x,a) = E_{\pi}\{R_t | x_t = x, a_t = a\} \quad (1)$$

Where $R_t$ is the expected reward at time $t$. By following the optimal policy we can find near optimal values, which can be approximated by iterations [4]:

$$Q(x_t,a_t) \leftarrow Q(1-\beta)Q(x_t,a_t) + \beta[R_t + \gamma \max_a Q(x_{t+1},a)] \quad (2)$$

Where $\beta \in (0,1]$ is the convergence rate.

For energy-aware and efficient routing policies in body area networks, we define states, actions and rewards as follows.

State: $X = \{x_i$ , the sensor nodes of body area networks; where $i = (1,2,3 \cdots\cdots n)\}$

Action: $A = \{a(x_i|x_j)$, packet forwarding actions between sensor node $x_i$ and $x_j$; where $x_j$ is in the transmission range of $x_i$ and $x_i, x_j \in X\}$

Reward function:

$$R_t = P_a^s(x_i,x_j) R_a^s(x_i,x_j) + \left(1 - P_a^s(x_i,x_j)\right) R_a^u(x_i,x_j) \quad (3)$$

Where $P_a^s(x_i,x_j)$ is the probability of successful packet transmission using action $a$ from node $x_i$ to $x_j$, where $x_i$ is adjacent to node $x_j$.

However, $R_a^s(x_i,x_j)$ is the reward of successful transmission, which can be defined as follows:

$$R_a^s(x_i,x_j) = -\alpha_1[C_i(residual_{energy}) + C_j(residual_{energy})] + \alpha_2 D_i(signal_{strength}) + \alpha_3 E_{j,p}(heuristic_{distance}) \quad (4)$$

Where $\alpha_1$, $\alpha_2$ and $\alpha_3$ are weight factors of residual energy, signal strength and heuristic distance respectively. Generally, summation of their values is 1.

$C(residual_{energy})$ is the cost function of residual energy, which depends on initial energy of a sensor node $i$ and the present residual energy of that node. In case of successful packet transmission both sender and receiver nodes ($x_i$ and $x_j$ ) respectively are losses their energy.

$$C(residual_{energy}) = \frac{1}{E_{init}}(E_{init} - E_{res}) \quad (5)$$

$D(Signal_{strength})$ is the reward function of signal strength received by sender node $x_i$ from its neighboring nodes $x_j$. The kurtosis of the received signal strength returns the height and sharpness of the peak relative to the rest of the signals. So, it is better to select that node which has largest peak value as a forwarder of data packets.

$$D(Signal_{strength}) = Kurtosis(RSS(x_i,x_j)) = \frac{\Sigma(s_j-\bar{s})^4/m}{\left(\Sigma(s_j-\bar{s})^2/m\right)^2} \quad (6)$$

$E(Heuristic_{distance})$ is the reward function of the admissible heuristic value of straight line distance between sink node $x_p$ and expected receiver node $x_j$. The receiver node with minimum heuristic distance is more suitable for data packet forwarding. This reward function leads to route through the optimal path i.e with minimum hop as well as minimum delay.

$$E(Heuristic_{distance}) = 1 - \left[\min_{dist} h_{dist}(x_p,x_j) / \max_{dist} h_{dist}(x_p,x_i)\right] \quad (7)$$

$R_a^u(x_i,x_j)$ is the reward function for unsuccessful packet transmission. In that case receiver node doesn't losses any energy. As it is an packet transmission error, so such action should be punished with a constant value $\kappa$. It resists the sender node to select those forwarding node which are in sleep mode or having low reliability of packet transmission.

$$R_a^u(x_i,x_j) = -\kappa - \alpha_1[C_i(residual_{energy})] + \alpha_2 D_i(signal_{strength}) + \alpha_3 E_{j,p}(heuristic_{distance}) \quad (8)$$

## 4. Performance Evaluation

We used MATLAB to study performance of the proposed QRP. We considered wireless channel capacity as 2 Mbps and link capacity as 1 Mbps. The maximum and minimum acceptable data rates are 128Kbps and 512Kbps respectively. There are total 16 nodes in the proposed BAN architecture. Each of the sensor nodes has 2 Jules of battery energy. We run simulation for 2000 seconds. The used constant values are assigned as follows $\alpha$=0.5, $\alpha_1$=0.4, $\alpha_2$=0.3, $\alpha_3$=0.3, $\beta$=0.5 and $\kappa$=1.

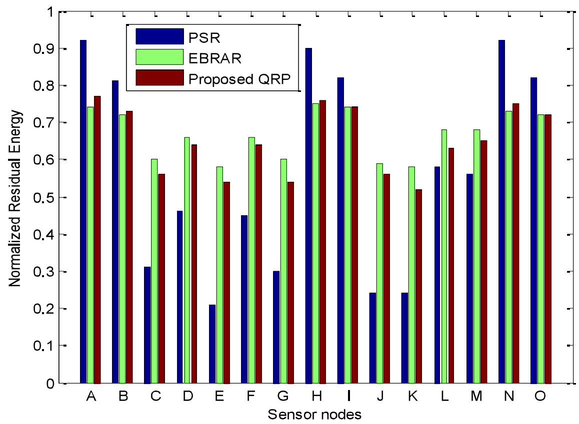


Figure 2. Residual energy distribution

The normalized residual energy per node is shown in Fig. 2. Here we have seen that normalized energy varies from approximately 0.2 to 0.92 for PSR routing strategy, 0.58 to 0.74 for EBRAR routing strategy and 0.52 to 0.76 for proposed QRP strategy. So, energy becomes more balanced in EBRP, imbalanced in PSR and moderately balanced in QRP.

Fig. 3 shows the average hop count per node to send packets to sink node. Here we have observed that the average hop count of most of the sensor nodes is below 2.5 for PSR, below 3.0 for QRP but above 3.0 for EBRAR. So, data packets of EBRAR travel longer path then other two approaches.

Both the PSR and EBRAR routing strategy follows extreme approach. The PSR fully concentrates on minimization of path length as a result its energy consumption of different nodes mostly imbalanced. So it is pruned to network partition and its consequence is lower network lifetime. Conversely, EBRAR fully concentrates on even energy-consumption of different nodes as a result most of the nodes experiences huge path cost. So, it takes more time to send packets to sink node. Surprisingly, the QRP made the trade-offs between path cost and energy consumption balancing by realizing reward and punishment function of Q-learning.

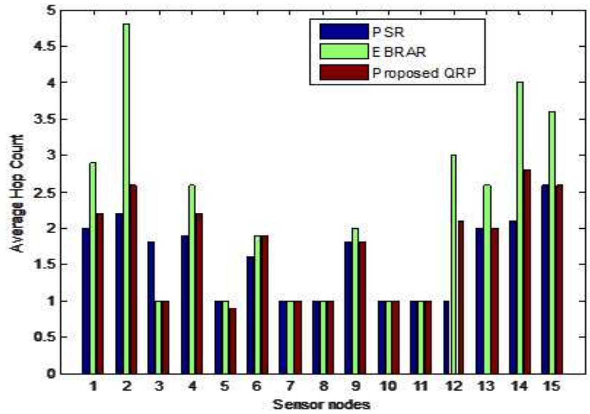


Figure 3. Hop count per node

## 3. Conclusion

In this proposed QRP routing framework, firstly each sensor node learns the nearby shortest route of sink node by updating the Q-values. Then uses that documented paths for transferring current and future data packets. The reward function for optimal Q-values contemplates residual energy, RSS and admissible heuristic. Hence, QRP shows moderately balanced energy consumption of different sensor nodes and nearly optimal path cost (hop count) for data packet transmission to sink node.

### Acknowledgement

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2011-0020518) *Dr. CS Hong is corresponding author.

## 4. References

[1] H. Cao et al., “Enabling Technologies for Wireless Body Area Networks: A Survey and Outlook”, IEEE Communications Magazine, pp. 84-93, Dec. 2009.

[2] X. Liangt et al. “Exploiting Prediction to Enable Secure and Reliable Routing in Wireless Body Area Networks”, in Proc. IEEE INFOCOM, pp. 388-396, 2012

[3]Nedal Ababneh et al.”Energy Balanced Rate Assignment and Routing Protocol for Body Area Networks”, in Proc. 26th International Conference on Advanced Information Networking and Applications, pp. 466-471, 2012.

[4] R.S. Sutton and A.G. Barto, “Reinforcement Learning: An Introduction”, The MIT Press, March 1998.