

# MABRP: A Multi-Armed Bandit Problem-Based Energy-Aware Routing Protocol for Wireless Sensor Network

Mahmoud Parvin

Computer Engineering Department

Islamic Azad University of Qazvin

Qazvin, Iran

Mahmoud.parvin@yahoo.com

Mohammad Reza Meybodi

Computer Engineering Department

Amirkabir University of Technology

Tehran, Iran

meybodi@ce.aut.ac.ir

**Abstract**— Wireless Sensor Networks consist of tiny sensors which have limitations such as energy. These networks are utilized in a variety of applications which aids in boosting the network lifetime hence great amount of researches have been devoted to them. One of the major problems to reach this aim is to focus on routing for utilizing balanced and optimum energy consumption in sensor networks. In this paper we have presented an energy-aware routing protocol namely MABRP which endeavors to select a path to transmit the data to the sink node thus balanced and optimal energy consumption in node is achieved and in so doing it gets help from multi-armed bandit problem. To assess its efficiency, the proposed protocol has been simulated using ns2 simulator and the results obtained have been compared to that of two important routing protocols namely EAR and PGR.

**Keywords:** *Wireless sensor networks, energy-aware routing, multi-armed bandit, network lifetime*

## I. INTRODUCTION

Sensor networks are an especial type of computer networks that due to their high potential and vast applications have had a considerable growth. Such networks consist of thousands of tiny sensor nodes which are high potential, cost effective, limited processing and computing resources, low power which enables them to sense special attributes like damp, temperature, pressure etc. from the environment and transmit it to the Sink.

So, on the whole, these networks are noticeable in two ways:

- sensing especial parameters from the environment and
- communicating so as to pass their received packets to the Sink.

Sensor networks have many applications some of which include intelligent surveillance on expressways and areas which are difficult to reach, environmental monitoring and

target tracking. These applications entail sensor networks to be deployed as wireless [2].

Basically these networks have static nodes or nodes with a limited movement and a Sink to which any node transmits its data. Since the direct relationship between the nodes and the Sink, for the most part, needs high power consumption, so these networks should be designed in a way that each node communicates with the Sink through its neighbors and in an indirect way passing its data received by sensing the environment. Then finding the best route considering the minimum power consumption necessitates the nodes to be aware of the routes status toward the Sink [1][2].

This paper presents a routing protocol based on Multi-Armed Bandit problem for wireless sensor networks which through optimal and balanced energy consumption could enhance the network lifetime.

The remainder of the paper is organized as follows. Section II presents the survey of related works. Section III introduces the Multi-Armed Bandit in brief. In section IV the proposed protocol is discussed. Simulation results are explained in section V and section VI deals with conclusion.

## II. RELATED WORKS

Considering the importance routing has in increasing the capacity and improvement of sensor network, several routing protocols have been presented for such networks which could be categorized in different respects. Taking the view of network structure, one could divide routing protocols in sensor networks into two groups [3][4]:

- Flat Protocol: like EAR, PGR, SPIN, Rumor, Direct Diffusion methods.
- Hierarchical: like LEACH, TEEN, and MECN.

In hierarchical protocols, nodes are organized into clusters with each cluster having a cluster head which enjoys more energy compared with other members and through which the

nodes communicate with the Sink whereas in Flat protocols all of the nodes enjoy the same level that enables them to sense the environment parameters and be used as an intermediate node for routing. Based on the functions they perform, routing protocols could be categorized as Data-Centric Protocols, Location-Based Protocols, and Energy-Aware Protocols etc.

Increased research on Energy-Aware Routing in sensor networks can obviously depict the weighty importance and attention given by researchers to this problem aiming to enhance the network lifetime. From among the most important Energy-Aware Protocols is EAR.

In phase 1 the EAR transmits the expected packets and floods them locally and by so doing recognizes all probable paths to the destination and puts them into the forwarding table. In phase 2 based on the consumption energy and the distance to the next node, protocol attributes a probability to each corresponding path in the forwarding table. So when transmitting a packet by a node, that node with respect to the attributed probabilities to each of its paths selects a route and floods the data packet to the destination node. In this way instead of utilizing a fixed path for transmitting the data packets, several paths are deployed whereby the network lifetime is enhanced [17].

Probabilistic Geographic Routing (PGR) protocol is a Location-Based Protocol which uses the situational information and energy of the nodes to identify an optimal path to the destination. It is assumed that in a PGR each node knows about its destination node. In phase 1 any node exchanges its information with its neighbors for recognizing their location and communication power. Completing this phase, each source node which aims to transmit the data to the destination creates an entry of neighbors which are in  $\theta$  angle in relation to it and the Sink as the destination node. Then, based upon the amount of energy and potentials of the nodes, attributes some probability to them [6].

In transmission phase, any source node selects one node regarding the probabilities attributed to each node. The selected node determines the immediate neighbor in the same way and this process is followed until the packet reaches the destination node, namely the Sink.

One of the approaches toward the enhancement of the wireless sensor networks lifetime is to ensure balanced energy utilization in the nodes of the network. This objective could be accomplished through the selection of the optimum paths for transmitting the data.

In case low-energy path is selected for transmitting the data, the energy of the nodes en-route is dissipated; the death of these nodes would be the subsequent result. Such an incident can lead to the death of the network. The end of a sensor network is defined in a couple of ways.

Following are some of the definitions most referred to in papers. Lifetime is the period of time from the point a network initiates its function to the point where:

- the energy is terminated in one or more nodes and the sensor network turns to a disconnected one.
- the energy of a particular percentage of the nodes is terminated.
- the energy of the first node in the network is terminated.

Regarding the importance that has been given to the balanced utilization of the energy among nodes, this paper applies the third definition. This paper presents a Flat and Energy-Aware protocol based on a Multi-Armed Bandit called MABRP. This protocol which has a multi-path notion endeavors to get use of the Multi-Armed Bandit Problem and find a path in which the utilization of the energy among the nodes is balanced, thus enhancing the lifetime of the network.

In this Energy-Aware Protocol the forwarding table is created using the overall flooding of the packet. The selection of the path by any node in any time is done by the use of Multi-Armed Bandit running on it. This MAB selects the proper paths based on the Gittins Index and performs the routing operation.

### III. MULTI-ARMED BANDIT (MAB)

The classical Multi-Armed Bandit Problem is a random, probabilistic decision-making problem which in any phase of its function there is  $K$  possible actions or functions to select. Selecting the function  $j$  makes observing the results possible. These observations could be considered as reward for the expected function and provide useful information about adopting a function in the future. An MAB is defined as a process of Multi-Armed Bandit involving a controller and some arms [8].

MAB controller in any instant of time  $t$  based on an idea named Gittins Index adopts an arm for action and does it. In this way active arm state changes according to its performance and other inactive arms remain frozen. Each arm  $i$ ,  $i = 1, 2, \dots, k$ , is described as

$$\{ (S_i(t)), R(S_i(t)) \} \quad (1)$$

where  $S_i(t)$  is the state of  $i$  arm of the machine at time  $t$  and  $R(S_i(t))$  is its reward in case it is chosen to perform in  $S_i(t)$  state.

Since the controller can operate on exactly one arm at each instant time, the control action  $U(t)$  takes values in  $\{e_1, e_2, \dots, e_k\}$ , where  $e_j = (0, 0, \dots, 0, 1, 0, \dots, 0, 0)$  is a unit  $k$ -vector with 1 at the  $j^{\text{th}}$  position. So the modifications in the states of machines is determined as follows:

$$S_i(t+1) = \begin{cases} f_i(S_i(t)) & \text{if } U_i(t) = 1 \\ S_i(t) & \text{if } U_i(t) = 0, \end{cases} \quad (2)$$

arm  $i$  generates a reward only when it is operated by the controller and in this way modifies the arm status. Let  $R_i(t)$  denote the reward generated by arm  $i$  at time  $t$ ; then

$$\begin{cases} R_i(S_i(t)) & \text{if } U_i(t) = 1, \\ 0 & \text{if } U_i(t) = 0, \end{cases} \quad (3)$$

A Scheduling policy is one in the form of  $\pi=(\pi_1, \pi_2, \dots)$  and consist of some decision making rules that in any instant of time  $t$  sets a value to a controlling word  $U(t)$  in the set  $\{e_1, e_2, \dots, e_k\}$  according to

$$U(t) = \pi_t(S_1(t), \dots, S_k(t), U(t-1)) \quad (4)$$

The goal of MAB Problem is to represent an optimal Scheduling policy  $\pi$  in order to maximize the obtained reward. This theorem is formulated as follows :

$$J^\pi = E \left[ \sum_{t=0}^{\infty} \beta^t \sum_{i=1}^k R(S_i(t), U_i(t)) | S(0) \right] \quad (5)$$

where  $\beta$  called the discount factor where  $0 \leq \beta \leq 1$  and is used to give importance to the obtained rewards in early phase. Applying the Bellman's equation which shows the relationship between the value of the state of  $S$  and the immediate occurred state, the expected discount reward is calculated as following:

$$V_{s_i}(s_i(0)) = \max_{\tau > 0} \frac{E \left[ \sum_{t=0}^{\tau-1} \beta^t R_i(S_i(t)) | S_i(0) \right]}{E \left[ \sum_{t=0}^{\tau-1} \beta^t | S_i(0) \right]} \quad (6)$$

Here  $V_{s_i}(S_i(0))$  is the maximum expected discounted reward per unit of expected discounted time that can be obtained at the first stage of a forward induction policy that continues the arm  $i$  with initial state  $S_i(0)$ . The corresponding stopping time  $\tau_i(S_i(0))$  is the first time at which the expected discounted reward per unit of expected discounted time equals  $V_{s_i}(S_i(0))$ . Consequently, at  $t = 0$  an optimal forward induction policy continues an arm  $j$  such that:

$$V_{s_j}(s_j(0)) = \max_i V_{s_i}(s_i(0)) \quad (7)$$

Arm  $j$  is continued until  $\tau_j(s_j(0)) - 1$ . This constitutes the first stage of an optimal forward induction policy, and can be summarized as follows:

Step1: Determine  $V_{s_i}(S_i(0))$ ,  $i = 1, 2, \dots, k$ .

Step2 : Select an arm  $j$  such that

$$j = \arg \max_i V_{s_i}(s_i(0)) \quad (8)$$

Continue operating arm  $j$  until the minimum time that achieves the maximum in the right hand side of the above calculation.

#### IV. PROPOSED PROTOCOL

Some existing routing protocols suffer the problem of finding the lowest energy consumption path through which all of the communication between the source and destination are implemented. Since the continuous utilization of these patch, causes the nodes associated with them to deplete and cuts

down on the network lifetime, this view couldn't be an appropriate one as far as energy is concerned. The goal of the proposed protocol (MABRP) is to solve the problem posed via creating uniformity in energy consumption among the nodes of the network for enhancing the network lifetime. This objective is fulfilled using a multi-Armed Bandit for each node. Proposed protocol operates in this way that in each phase any node using its MAB Selects a patch aiming to maximize its own obtained terminating reward that is more and more packets transmitted and enhancing the network lifetime. One could put the performance of this method into 2 phases: forwarding Table Construction Phase, Indexing and Routing phase.

##### A. Forwarding Table Construction Phase:

This phase initiates with creation and transmission of CR Message From the sink toward its neighbor nodes to flood it throughout the network. This packet contains three fields namely the number of the node from which the packet has been received the amount of energy and hopcount from the sender node to the sink. The sink then gives a number to the sender node as its own, the amount of energy of the sender node as itself and the hopcounts from the sender node to the sink the number zero, and by passing the message through local flooding to all of its neighbors initiates the creation of forwarding table phase. Any other nodes in this network upon receiving the packet adds a record into its forwarding table and then checks whether or not has received such message before.

In case it is the first time i.e. the first record added into the forwarding table, after saving the data, the node updates the received message and via overall flooding routes it to all of its neighbors in the network, otherwise omits it. To update, a node changes the number of the sender node and the amount of the energy changes the hopcount to the sink with a single unit increase.

Consequently any node realizes all of its neighbors and preserves them for future use in the next phase. At the end of this phase there will be bandit automation on each node of the network which has some arms equal to the number of records or routs in its forwarding table that corresponds to the number of the neighboring nodes. The selection of any of these routes in routing phase depends on the Gittins Index obtained for its corresponding arm.

##### B. Routing and Indexing Phase:

This phase pertains to the transmission of data sender by each node towards the sink. Since the consistent use of a path between the sender node and the sink gradually decreases the energy of the nodes belonging to that path and finally divides the network which in turn leads to the network termination. Therefore, in this phase to refrain this problem, the selection of the path is done based on the index obtained by the amount of energy of neighboring by the amount energy of the neighboring nodes as the intermediate destination and their hopcounts to the sink.

In this phase, the node which has sensed the expected

parameter from the environment and has a packet for transmission initiates to route the data packet DP based on its own bandit automation. The controller or MAB selects one of the data to its corresponding neighbor. The controlling word itself is created based on scheduling policy which aims to optimize the terminating reward obtained by the machine, so is our indexing that is the optimal network lifetime it should be noticed that in this paper a network terminates if the energy of the first node is finished hence, the complete use of the energy of the network and in turn enhancing its lifetime is applicable if energy consumption is uniformed and optimized. Gittins index which has been used for preferring the arms for routing in this protocol enjoys two criterion namely the amount of the energy of the neighboring node and its hopcounts to the sink where by one could identify  $\gamma_i$  as Gittins index for the neighboring node  $i$  as in the following calculation:

$$\gamma_i = c \frac{e_i}{e_s} + (1 - c) \left(1 - \frac{h_i}{h_s}\right) \quad (9)$$

In the above calculation  $e_i$  is the amount of energy of the corresponding neighbor node with  $i$  arm and  $e_s$  is the sum of the energy of all of the neighbors of the sender node,  $h_i$  is the number of hopcounts of the  $i$  node to the sink node and  $h_s$  is the sum of the hopcounts any neighbor to the sink node. The field which depicted the amount of the energy was used to increase the preference of selecting the nodes with maximum energy thus the energy of the nodes is uniformed. The hopcount field was used to give a preference to the shortest path and by so doing the consumption of energy of the nodes is optimum.

With the help of the  $c$  coefficient the effectiveness of each criterion is realized. After indexing, the sender node places a value on packet fields DP and transmits it to the corresponding route with the help of the selected arm on its Bandit automaton which is the arm with the maximum index. Intermediate node, too, upon receiving the data packet DP with the help of its multi-armed bandit automaton and based on its forwarding table routes to the sink. This trend is repeated until the destination node is reached.

Any node after receiving and transmitting a DP constructs a RP (Reply Packet) message which contains its own number and the amount of the energy, and routes it to the node of the sender of the DP message whereby updates its amount of the energy in the table of the sender node. The sender node then omits the RP after receiving it and updating.

With regard to what was said, any node by selecting an arm and its corresponding neighbor node as an intermediate destination, allows that node to take part in the transmission of the data packet (DP) and it turn consumes its energy. So, the field that shows the amount of the energy which is thought to be the arm of the bandit is modified and updated in the record corresponding to that node in the forwarding table whereas the record corresponding to other arm of the bandit remain frozen and unchanged.

## V. SIMULATION CONCLUSIONS

The protocol presented in this paper is simulated and assessed with the help of the network environment simulator software called ns2. It is assessed using criterion like the number of the controlling and data packets transmitted and received which could serve as a criterion for measuring the network lifetime. Besides its results have been compared to that of the routing protocols like EAR and PGR.

To fulfill this goal, the results are simulated in several regions which have areas of 1500\*1500, 2000\*2000, 2500\*2500, and 3000\*3000 meters having 100, 200, 300, and 400 nodes respectively. The Sink node and the sensor nodes are assumed to be fixed during the simulation and the radius of the transmission of sender nodes are fixed during simulation and equals 25. The initial energy of all nodes is the same and equals 100 Joules and the consumption energy needed to transmit any kind of packet is 660 MilliJoules and the consumption energy for the reception of any kind of packet equals 395 MilliJoules. For the simulation to reach the most similar environment, the location of the nodes is determined randomly either.

Since controlling messages cause the depletion of the energy of the nodes which in turn decreases the network lifetime hence it is a good criterion for assessing and comparing the protocols. The proposed protocol MABRP utilizes the CRP controlling packets to construct the forwarding table. But it enjoys the advantage that it is transmitted as overall flooding thus it is optimal when compared with the EAR which uses the node to node flooding packet method. To greet the nodes in order for gathering information about the position and the potentials of the neighboring nodes, the PGR Protocol needs a high amount of controlling message.

Figure 1 depicts the number of the controlling packets Exchanged by nodes for three protocols i.e. MABRP, EAR, and PGR in different areas of the network.

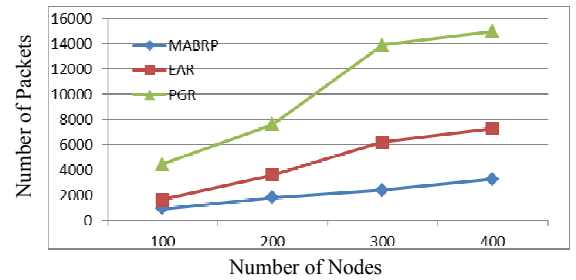


Figure 1 : number of the controlling packets Exchanged for three protocols i.e. MABRP, EAR, and PGR.

Since the proposed protocol aims to boost the network lifetime hence it could be examined by one of the most important criterions for realizing the network lifetime namely the number of the data packets received by the Sink.

Figure 2 depicts the number of the data packets received by the Sink for three protocols i.e. MABRP, EAR, and PGR in different areas of the network.

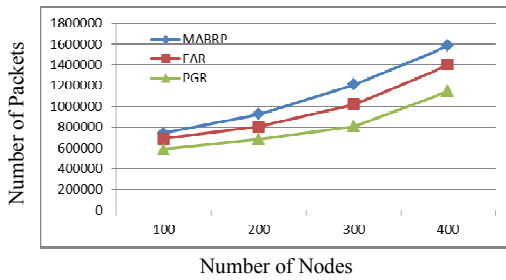


Figure 1 : the number of the data packets received by the Sink for three protocols i.e. MABRP, EAR, and PGR.

The observations imply the superiority of the proposed protocol over EAR and PGR. It is superior for it has such more precise function balancing the energy consumption among the nodes of the network as compared with EAR, PGR.

As EAR protocol utilizes all of the routes to the destination, therefore it receives more data in the Sink node when compared with PGR which uses the closest routes to the destination. In the protocol proposed here, the network lifetime is boosted because no fixed route between the source node and the Sink is utilized continuously, but the routes are selected by alternation regarding the length of the route and the energy of its nodes. This point, however, is worth noting that in the protocol MABRP the maximum area of the transmission for each node is used to communicate between the nodes, whereas in protocols like PGR and EAR the communication between two nodes is in relation with the distance between those two nodes.

Therefore in case the density of the nodes is increased in the network, the efficiency of the protocol MABRP to boost the network lifetime is then no longer better than that of EAR and PGR.

## VI. CONCLUSION

In this paper a novel routing protocol based on multi-armed bandit problem for sensor networks is presented. The proposed protocol holding the notion of utilizing balanced and optimum energy consumption endeavors to select the best path for the transmission of packet data from the nodes of the network. Simulation results show that as far as the nation of uniform and optimum energy consumption among the nodes which in turn leads to the enhancement of the network lifetime is concerned the proposed protocol is superior to protocols like EAR and PGR.

## REFERENCES

[1] J. Yick, b. Mukherjee and d. Ghosal, "Wireless Sensor Network Survey", Journal of Elsevier on Computer Networks, Vol. 52, 2008.

[2] F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "A Survey on Sensor Networks", Of IEEE Communications Magazine, Vol. 40, 2002.

[3] K. Akkaya and M. Younis, "A Survey on Routing Protocols for Wireless Sensor Networks", In Proc. Of Elsevier on Ad Hoc Networks, Vol. 3, 2005.

[4] J. N. Al-Karaki and A. E. Kamal. "Routing techniques in wireless sensor networks: a survey". In IEEE Wireless Communications, Volume 11, pp. 6-28, 2004.

[5] S. Madiraju, C. Mallanda, R. Kannan, A. Duresi, and S. S. Iyengar, "EBRP: Energy Band based Routing Protocol for Wireless Sensor Networks", Department of Computer Science, Louisiana State University 298 Coates Hall LA 70803, USA, 2004.

[6] T. Roosta, "Probabilistic geographic routing protocol for adhoc and sensor networks", Proceeding of Wireless Networks and Emerging Technologies, 2005.

[7] Y. Yao, G. B. Giannakis, "Energy-Efficient Scheduling for Wireless Sensor Networks", IEEE Transactions on Communications, VOL. 53, NO. 8, August 2005.

[8] E. Frostig and G. Weiss, "Four proofs of Gittins' multi-armed bandit theorem. Technical report", The University of Haifa, Mount Carmel, 31905, USA, November 1999.

[9] J. C. Gittins, "Multi-Armed Bandit Allocation Indices", John Wiley and Sons, New York, NY, 1989.

[10] V. Krishnamurthy and J. Mickova, "Finite Dimensional Algorithms for The Hidden Markov Model Multi-Armed Bandit Problem", In Proc. Of IEEE on Acoustics, Speech and Signal Processing, Phoenix, AZ, USA, Vol. 5, 1999.

[11] V. Krishnamurthy, "Emission Management For Low Probability Intercept Sensors In Network Centric Warfare", of IEEE Transactions on Aerospace and Electronic Systems, Vol. 41, 2005.

[12] C. Yunxia, Z. Qing, V. Krishnamurthy and D. Djonin, "Transmission Scheduling for Optimizing Sensor Network Lifetime: A Stochastic Shortest Path Approach", Of IEEE Transactions on Signal Processing, Vol. 55, 2007.

[13] J. Mersereau, P. Rusmevichientong and J. N. Tsitsiklis, "A Structured Multiarmed Bandit Problem and the Greedy Policy", In Proc. Of 47th IEEE Conference on Decision and Control, Mexico, 2008.

[14] V. Phan, Y. Park, H. H. Choi, J. Cho and J. G. Kim, "An Energy-Efficient Transmission Strategy For Wireless Sensor Networks", Of IEEE Transactions On Consumer Electronics, Vol. 56, 2010.

[15] R. Srivastava and C. E. Koksal, "Energy Optimal Transmission Scheduling in Wireless Sensor Networks", Of IEEE Transactions On Wireless Communications, Vol. 9, 2010.

[16] H. Luo, Y. Liu and S. K. Das, "Routing Correlated Data in Wireless Sensor Networks : A Survey", Of IEEE Network, Vol. 21, 2007.

[17] R. Shah, J. Rabaey, "Energy aware routing for low energy ad hoc sensor networks", in: Proceedings of the IEEE Wireless Communications and Networking Conference(WCNC), Orlando, FL, March 2002.

[18] Y. R. Tsai, "Coverage-preserving routing protocols for randomly distributed wireless sensor networks", Of IEEE Transactions On Wireless Communications, Vol. 6, 2007.

[19] <http://www.isi.edu/nsnam/ns/>.

[20] D. Ganesan, R. Govidan, S. Shenker and D. Estrin, "highly-Resilient, Energy-Efficient Multipath Routing in Wireless Sensor Networks", ACM Mobile Computing and Communication review, vol.5, No.4, october 2001, pp.10-24.