ALOHA and Q-Learning Based Medium Access Control for Wireless Sensor Networks

Yi Chu, Paul D. Mitchell and David Grace

Department of Electronics, The University of York
York, United Kingdom
yc541@york.ac.uk
pdm106@york.ac.uk
dg@ohm.york.ac.uk

Abstract— **This paper presents a novel Medium Access Control (MAC) protocol named ALOHA and Q-Learning based MAC with Informed Receiving (ALOHA-QIR) for Wireless Sensor Networks (WSNs). This scheme is implemented by combining frame based Slotted ALOHA and Q-Learning. ALOHA based schemes have the benefits of simplicity, low computation and overheads, but suffer collisions from the blind transmission strategy. By considering the previous transmission experience, Q-Learning offers nodes certain intelligence to access slots with a lower probability of collision. During data transmissions, receivers are informed by transmitters about their future slot selections so that the receivers can turn off their radio in other slots to save energy. Compared with basic Slotted ALOHA, ALOHA-QIR obtains significant improvements in throughput, delay and energy efficiency.**

I. INTRODUCTION

Owing to the rapid development of sensor motes, WSN has become an emerging technology with a wide range of applications in industry, for the military and for environmental monitoring [1]. A WSN is usually a self-organised network consisting of a large number of nodes which are capable of sensing and processing data via multihop communication.

In a WSN, a node usually has to share the medium with other nodes within communication range, which makes Medium Access Control (MAC) an important issue to ensure the network is operational. Meanwhile, the lifetime of a WSN usually depends on the energy efficiency because of the difficulties in recharging or replacing node. In a sensor node, energy is mostly consumed by the radio transceiver. Except for the useful throughput, the unnecessary energy consumption arises from collisions, retransmissions, control packet overheads, idle listening and over hearing [2]. A well designed MAC protocol needs to ensure the successful delivery of the data while keeping the energy consumption to a minimum.

Researchers have proposed many MAC protocols for WSNs. SMAC [3] periodically switches nodes between sleep and active modes to save energy. T-MAC [4] and DS-MAC [5] apply on adaptive duty cycle to SMAC based on current performance to improve the energy efficiency. DW-MAC [6] dynamically schedules the wake up times of the nodes to improve the effective channel capacity. LEACH [7] periodically builds clusters and schedules in-cluster transmission by Time Division Multiple Access (TDMA). TRAMA [8] avoids collisions and the hidden terminal problem by exchanging scheduling information.

The work stated above has made certain achievements on channel performance and energy efficiency. However as the research proceeds, the MAC protocols are becoming more and more complicated and the control packet overheads consume more channel resource (e.g. some protocols require periodic updates of the scheduling information of one-hop or even two-hop neighbours), simpler solution is a promising alternative.

ALOHA based schemes have the benefits of simplicity, low computation and overheads, but suffer collisions from the blind transmission strategy. Q-Learning is selected as the Reinforcement Learning algorithm to avoid collisions and retransmissions, thereby improving throughput and benefiting energy efficiency. Reinforcement Learning is a technique which learns behaviour through trial-and-error interactions from a dynamic environment, and determines the actions by the experience built up from rewards and punishments [9]. It has been widely used in the research and developments of Artificial Intelligence (AI), and introduced into MAC protocols in recent years. In [10] and [11], several Reinforcement Learning based MAC protocols for wireless networks are proposed, and they control the channel access based on the learning experience.

The rest of the paper is organised as follows. Section II introduces the details of the protocol design of ALOHA-QIR, including Q-Learning, Informed Receiving (IR) and ping packets. Section III outlines some assumptions made for this scheme. Section IV describes the simulations and analyses the performance results. Section V concludes the work in this paper.

II. ALOHA-QIR PROTOCOL DESIGN

A. *Apply Q-Learning to Slotted ALOHA*

To apply Q-Learning, repeating frames are incorporated into a conventional Slotted ALOHA. Each frame contains N slots and node i is allowed to use n_i slots per frame. N is the same for all the nodes in the network but n_i is decided based on the number of sources (including itself) the node needs to relay information for. For example, if node i relays the packets from two other nodes and itself, n_i will be 3.

Stateless Q-Learning [12] is used in this scheme to obtain the learning experience. Each node has individual Q values for every slot, and they represent the preference of the slot selections. Q values are denoted by $Q(i, k)$ and they represent that node i takes an action on slot k. The previous Q values

978-1-4673-0762-8/12/$31.00 ©2012 Crown

and the current reward all contribute to the Q value update. The Q value is updated as below after the reward is returned.

$$Q_{t+1}(i, k)=Q_t(i, k)+\alpha(r-Q_t(i, k)) \quad (1)$$

Where α is the learning rate (usually set to less than 1) and r is the current reward. If the transmission succeeds, the current reward will be +1 and if the transmission fails, the current reward will be -1. The slots with higher Q values will always be preferred and if multiple slots have the same Q value, the node will randomly select one (or more) of them. All the Q values are initialised to 0 at the very beginning. Figure 1 shows an example of the frame structure and Q-Learning algorithm for a single node. In this example, N equals 3 and n_i equals 2.

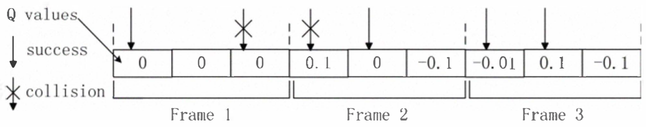

Fig. 1 Example of Q values and repeated frames

After a packet is generated by the upper layers or arrives from the previous hop, it will be put into a first-in-first-out queue and only the packet at the head can be transmitted. A packet needs to wait until the next preferred slot to be sent after its arrival or collided. A maximum retry limit M = 6 is given according to the IEEE 802.11 specification [13], and a packet will be discarded if it exceeds the retry limit.

After the learning process, each node will find unique slot (slots) to transmit packets if N is set appropriately. N is an important parameter in this scheme and its optimum value is related to the network topology, the density of the nodes and the number of source nodes. Generally, N needs to be set higher when the network has a greater number of source nodes and a higher density. But if N is too large, some slots might be wasted and the delay will increase.

B. *Informed Receiving (IR)*

During the learning process the nodes keep hopping to different slots to find the optimum, and they need full-listening to avoid missing any information from the previous hops. To accommodate the energy issue in WSNs, IR is applied to appropriately switch the nodes to a sleep mode to avoid idle listening and overhearing.

During packet reception, it is efficient for a receiver to only listen to the current preferred slot (slots) of its transmitter (transmitters). The transmitters may have various preferred slots during different time periods due to the dynamics of the network. IR is applied to provide the receiver with recent information from its transmitters, thereby making the listening more efficient.

When a node switches on, it listens all the time and waits for packets from potential transmitters for a certain time period. After this time period, the receiver decides which slot (slots) to listen to based on the information obtained from the transmitters. The transmitter piggybacks a number *(m)* on each data packet to inform the receiver its transmission pattern. *m* represents the number of future frames the transmitter will keep using the current preferred slot *i*, and it is calculated based on the iterations of (1). An example is given below to clearly demonstrate the calculation of *m*.

Suppose N = 4, $n_1 = 2$ and the Q values of node 1 are $Q(1, 1) = 0.5$, $Q(1, 2) = 0.1$, $Q(1, 3) = -0.5$ and $Q(1, 4) = -0.2$ Slot 1 and 2 are the current preferred slots for node 1 according to the Q values. Considering the worst case where all future transmissions will fail, the returned reward r will always be -1. $Q(1, 2)$ is the smallest value of the current preferred slots, and $Q(1, 4)$ is the largest value of the slots which are currently not preferred. Suppose that after 5 iterations of (1) (with the reward of -1) $Q(1, 2)$ becomes less than $Q(1, 4)$, then the current value of m will be 5. It represents the number of future frames that the node will keep using the same preferred slots. m is calculated before every data packet is transmitted. To avoid unnecessary calculations, m is given a maximum value of 20 which means the transmitter will keep with the current preferred slots for quite a few frames.

The receiver has a group of timers T:

$$T=\{t_1, t_2, ..., t_N\} \quad (2)$$

T has the same number of timers as the number of slots per frame N, and the timers represent the number of future frames the associate transmitters of different slots will use (one node may receive data packets from multiple nodes). The timers are updated at the beginning of each frame (all minus 1) and after the arrival of the data packets in the associate slots (set to m of the transmitter plus 1 to accommodate the offset caused by the other update). Each node knows the number of sources it relays, then it needs to listen to the same number of slots per frame n_i (if the node is a source the number will be $n_i - 1$). If the n_ith largest timer is larger than 1, the node will only listen to the n_i slots with larger timers. However, if the n_ith largest timer is equal to or lower than 1, the node will listen to all the slots to see if the transmitters will switch to other slots.

Based on the previous example, suppose node 3 is the next hop of node 1, and node 2 also selects node 3 as its next hop. Suppose $n_2 = 1$ and node 3 is not a source node, so $n_3 = 3$ and node 3 listens 3 slots per frame. Suppose the preferred slot of node 2 is slot 4 and m of node 2 is 3, and the preferred slots of node 1 are slot 1 and 2 and m is 5, so the timers of node 3 are $t_1 = 5$, $t_2 = 5$, $t_3 = 0$ and $t_4 = 3$. Node 3 will only listen to slot 1, 2 and 4 until anyone of t_1, t_3 and t_4 falls below 2.

C. *The ping packet*

IR switches on the radio in the preferred slots of the associated transmitters and switches it off in the others. By applying IR, the nodes can follow the pattern of the associated transmitters to avoid idle listening and over hearing, thereby improving the energy efficiency. However, when the traffic is very low, one node may keep silent for several frames between two packet transmissions. The low traffic causes problems because the receiver timer T will not be updated over a long period. The timer may timeout frequently and the receiver will switch to full-listening mode more often, which will lead to unnecessary energy consumption caused by idle listening and over hearing. If the receiver receives nothing when listening to a slot, it has no idea whether the reason is that the packet has collided or the transmitter has no data to send. The ping packet is applied to frequently provide the receiver correct transmission information and appropriately switch off the radio under low traffic conditions.

When a node reaches its preferred slot and has no data to send, it transmits a ping packet to its next hop instead. A ping packet contains the same information of the transmitter as described in the previous section. The node which receives a ping packet will send back an acknowledgement packet immediately, then update the associated timer and switch off its radio until the next listening or transmitting slot. If a ping packet is not successfully acknowledged, the node will receive a -1 reward and update the associated Q value. The Q value will not be updated if the ping packet is acknowledged, because a ping packet is tiny and its success does not indicate that the data packet can succeed in the same slot.

The overheads caused by ping packets are inversely proportional to the traffic level of the data. The ping packet consumes more energy and channel resource with low traffic, but it avoids idle listening in blank slots and prevents the node from switching to full listening mode when the transmitter is not going to hop to other preferred slots.

III. Assumptions and Scenario

A. The Basic Scenario

ALOHA-QIR is simulated and evaluated in OPNET. 50 nodes are randomly deployed in a $50m \times 50m$ square area and all nodes share the same channel (all nodes transmit and receive on the same channel). All nodes sense the environmental data, fragment the data into packets with the same length and forward them to the sink. The scenario has one sink node that is deployed at the centre of the sensing area. Most Forward Routing (MFR) [14] is used as the routing protocol in the simulation, and all the schemes in the simulations are assumed to be synchronised to slots. The learning rate is set to 0.001 in the simulation.

B. Traffic

We assume the sensing data is generated according to a Poisson arrival model [15]. All the nodes are given the same average packet inter-arrival times which is calculated from the normalised overall generated traffic (Erlangs) given by the simulation. 1 Erlang represents an amount of traffic corresponding to a node transmitting or receiving constantly. The average packet inter-arrival time is obtained as below,

$$\tau = \frac{L \times S}{G \times R} \quad (3)$$

Where τ denotes the average packet inter-arrival times, L denotes the length of a data packet, S represents the number of source nodes in the network, G is the generated traffic of the network and R denotes the data rate of the channel. Each node can store maximum 200 packets in the buffer, and more incoming packets will be discarded directly.

C. Radio

Considering the complexity of implementing the full interference model, we provide some definitions to the radio transceiver. During packet reception, any partial or entire overlap of the packet will be considered as failed. The propagation delay is ignored because the nodes are close to each other and the propagation delay is tiny compared with the waiting time a packet spends in the queue. Only the packets within a certain radius of a node can be correctly received, but the transmissions within a larger radius of the node can cause interference and collisions. Some radio transceiver parameters are obtained from IRIS motes datasheet [16].

D. Scheme for Comparison

Slotted ALOHA with Binary Exponential Backoff (ALOHA-BEB) [17] is used to compare the performance with ALOHA-QIR. BEB is a retransmission strategy for Slotted ALOHA. After a packet has collided, a retransmission will be scheduled in a random slot within an initial contention window W. If the retransmission fails, the contention window will be doubled either until the packet is received or reached the maximum retry limit ($M = 6$) is reached.

E. Simulation Parameters

Details of the simulation parameters are given in Table I.

TABLE I
SIMULATION PARAMETERS

Parameters	Values
Channel bit rate	250 Kbits/s
Data packet length	1044 bits
ACK packet length	20 bits
Ping packet length	28 bits
Slot length	1100 bits
Transmit power	51 mW
Receive power	48 mW
Interference range	30 m
Receive range	15 m
Simulation length	5×10^6 slots

IV. Performance Evaluation

A. Metrics for Analysis

In the simulations, we assume the nodes know the physical position of all the other nodes so that the routing protocol MFR can calculate the routes, and all nodes determine their routes to the sink during the initialisation of the simulation by MFR. The collection of the simulation data starts after the simulation has run for 500,000 slots to make sure the system is in the steady state.

Five metrics are calculated to compare the performance between ALOHA-QIR and ALOHA-BEB. Normalised throughput is the amount of data traffic that successfully arrives at the sink. End to end delay is the time from when a packet is generated until it completes reception at the sink. The energy efficiency is demonstrated by three metrics:

energy cost per bit throughput which shows the average cost of delivering data, energy cost per second which indicates the lifetime of the network, and proportion of energy spent by transmitting/receiving data which describes and compares energy cost by data and overheads to show if the energy is efficiently used.

B. Performance Evaluation

Figure 2 shows the throughput performance of ALOHA-QIR with different frame sizes N and ALOHA-BEB with varying traffic levels. The throughput of the schemes grow linearly with the traffic (which indicates that all generated data is delivered) until it reaches certain limits. ALOHA-QIR (N =100) is the optimum and has a maximum throughput of approximately 0.47 Erlangs, which is more than twice Slotted ALOHA (0.22 Erlangs). Frame size N is an important parameter for ALOHA-QIR which affects the performance significantly. According to the learning algorithm each node i will own n_i unique slots of its next hop in steady state to avoid collisions. However if N is smaller than the optimum, some nodes may find their unique slots while others have to share, which causes collisions. If N is larger than the optimum, the system can reach a steady state with no collisions but some slots may never be used, which wastes channel capacity. Therefore ALOHA-QIR with N = 80 and 120 experience lower throughput than 0.47 Erlangs, the throughput of ALOHA-QIR with N=120 reaches a limit of about 0.39 Erlang and the throughput of ALOHA-QIR with N=80 falls below the generated traffic much earlier.

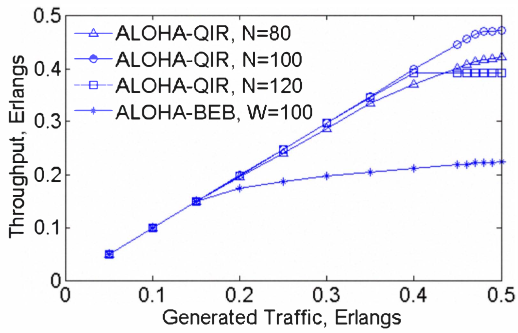

Fig. 2 Throughput under different traffic levels

Figure 3 shows the end-to-end delay performance. ALOHA-QIR with N =100 maintains a delay of less than 1 second under most traffic levels. ALOHA-QIR with N =120 has slightly higher delay because of the larger frame size and N =80 has much higher delay because of more collisions and retransmissions. ALOHA-BEB has the lowest delay when the traffic is very low because it immediately sends the packet in next slot after the packet is generated or received and does not have to wait until the next preferred slot.

Figures 4 to 6 demonstrate the energy efficiency. In figure 4, ALOHA-QIR with N = 100 and 120 experience similar energy cost per bit because they almost completely avoid the collisions and achieve perfect scheduling. The slightly higher energy cost at low traffic levels is caused by the ping packets, because nodes transmit many ping packets instead of the data packets under low traffic. It is shown more clearly in figure 6 that the network consumes 70% of its energy on data under low traffic conditions and over 95% when the traffic is high. ALOHA-QIR with N = 80 has higher energy cost because it experiences more collisions and retransmissions which also leads to more energy consumption through listening. ALOHA-BEB has the highest energy cost because it consumes a large amount of energy for listening. Figure 6 shows that only a tiny proportion of energy is consumed on data. Figure 5 shows the energy cost per second of the four schemes. The energy cost of ALOHA-QIR with N = 100 and 120 is dominated by data so that their energy cost per second increases linearly with the traffic. When N = 80, listening consumes more energy under low traffic and as the traffic increases the energy consumed by data becomes a greater part. However, when the traffic is to high, collisions and retransmissions dominate the energy cost and cause more energy waste. Moreover, when the traffic is 0.5 Erlangs, ALOHA-QIR with N = 100 consumes about 95 J/sec and ALOHA-BEB 2476 J/sec, which means that the network can survive at least 25 times longer.

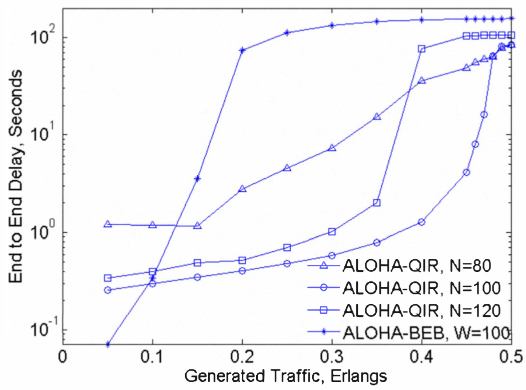

Fig. 3 Delay under different traffic levels

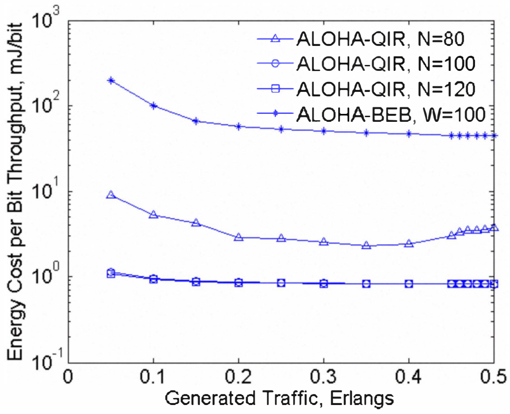

Fig. 4 Energy cost of the network per bit throughput

All the performance results above are obtained under steady state, but we also pay attention to the learning process before steady state. Figure 7 shows the Cumulative Distribution Function (CDF) of number of frames the simulations take before converging to the optimum solution under different traffic levels. All the simulations converge within 3000 frames so that we start the simulation data collection after 500,000 slots to make sure the system is under steady state condition. Under medium traffic (0.2 Erlang) about 90% of simulations converge within 500 frames (about 220 seconds).

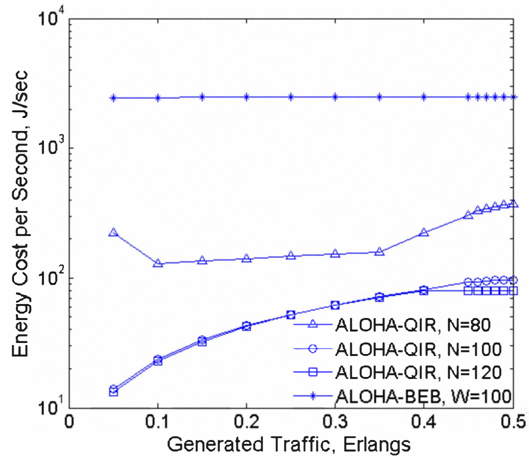

Fig. 5 Energy cost of the network per second

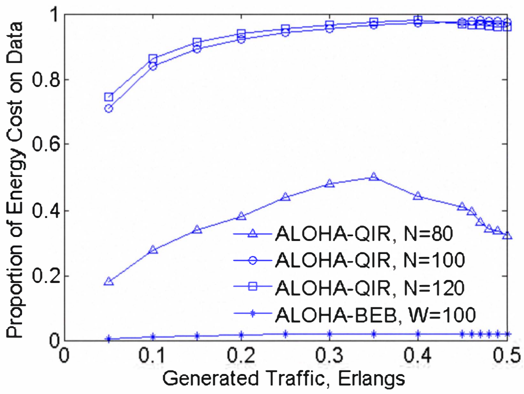

Fig. 6 Proportion of energy cost on data transmission and reception

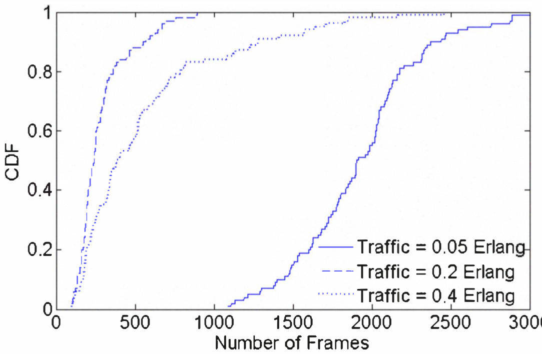

Fig. 7 CDF of the converging speed with different traffic levels

V. Conclusions

This paper presents ALOHA-QIR, an energy efficient MAC protocol for WSNs. According to the general behaviour of ALOHA, this scheme benefits from low complexity and control overheads, but by applying Q-Learning ALOHA-QIR achieves almost perfect scheduling with minimal overheads under steady state to avoid collisions and improve the channel utilisation performance. IR and ping packets set the nodes to sleep mode as long as possible when necessary to improve the energy efficiency, especially in low traffic operation. The simulations show that ALOHA-QIR achieves over twice the maximum throughput of Slotted ALOHA. End-to-end delay and energy efficiency are significantly improved as well.

References

[1] I. F. Akyildiz, Y. Sankarasubramaniam, E. Cayirci and W. Su, *Wireless sensor networks: A survey*, Computer Networks, vol.40, pp. 393-422, 2002.

[2] I. Demirkol, C. Ersoy and F. Alagoz, *MAC protocols for wireless sensor networks: A survey*, Communications Magazine, vol.44, pp. 115-121, 2006.

[3] W. Ye, J. Heidemann and D. Estrin, *An energy-efficient mac protocol for wireless sensor networks*, IEEE INFOCOM, pp. 1567-1576, 2002.

[4] T. v. Dam and K. Langendoen, *An adaptive energy-efficient mac protocol for wireless sensor networks*, SenSys'03, Los Angeles, California, USA, pp. 171-180, 2003.

[5] P. Lin, C. Qiao and X. Wang, *Medium access control with a dynamic duty cycle for sensor networks*, WCNC 2004 / IEEE Communications Society, pp. 1534-1539, 2004.

[6] Yanjun Sun, Shu Du, Omer G. and David Johnson, *DW-MAC: A Low Latency, Energy Efficient Demand-Wakeup MAC Protocol for Wireless Sensor Networks*, MobiHoc' 08, pp.53-62, Hong Kong, 2008.

[7] W. Heinzelman, A. Chandrakasan and H. Balakrishnan, *Energy-efficient communication protocol for wireless microsensor networks*, Proc. 33rd Hawaii Int'l. Conf. Sys. Sci., Japan, 2000.

[8] V. Rajendran, K. Obraczka and J. J. Garcia-Luna-Aceves, *Energy-efficient, collision-free medium access control for wireless sensor networks*, Wireless Networks, vol.12, 2006.

[9] L. P. Kaelbling, M. L. Littman and A. W. Moore, *Reinforcement learning: A survey*, Artificial Intelligence Research vol.4, pp. 237-285, 1996.

[10] Haibin Li; Grace, D.; Mitchell, P.D., *Cognitive radio multiple access control for unlicensed and open spectrum with reduced spectrum sensing requirements*, Wireless Communication Systems (ISWCS), 2010 7th International Symposium on , pp.1046-1050, 2010.

[11] Yi Tang; Grace, D.; Clarke, T.; Jibo Wei; , *Multichannel non-persistent CSMA MAC schemes with reinforcement learning for cognitive radio networks*, Communications and Information Technologies (ISCIT), 2011 11th International Symposium on, pp.502-506, 2011.

[12] Sutton, R. S., and Barto, A. G., *Reinforcement learning: An introduction*, Cambridge, MA: MIT Press, 1998.

[13] *IEEE Computer Society LAN MAN Standards Committee. IEEE std 802.11-1999m Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, 1999.

[14] N. AL-KARAKI and A. E. KAMAL, *Routing techniques in sensor networks: A survey*, IEEE Wireless Communications December, pp. 6-28, 2004.

[15] Moran, P.A.P., *An Introduction to Probability Theory*. 1984: Oxford University Press, 524.

[16] Datasheet for IRIS nodes available at [Online] http://www.memsic.com/products/wireless-sensor-networks/wireless-modules.

[17] Byung-Jae, Nah-Oak Song, *Performance analysis of exponential backoff* , Networking, IEEE/ACM Transactions on , vol.13, no.2, pp. 343- 355, April 2005.