Engineering Applications of Artificial Intelligence 55 (2016) 146–154

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Use of Q-learning approaches for practical medium access control in wireless sensor networks

Selahattin Kosunalp [a,b,*], Yi Chu [b], Paul D. Mitchell [b], David Grace [b], Tim Clarke [b]

[a] Department of Electricity and Energy, University of Bayburt, Bayburt, Turkey
[b] Communications and Signal Processing Research Group, Department of Electronics, University of York, Heslington, York YO10 5DD, UK

ARTICLE INFO

Article history:
Received 2 August 2015
Received in revised form
5 May 2016
Accepted 19 June 2016

Keywords:
Q-Learning
ALOHA
Medium access control
Wireless sensor networks

ABSTRACT

This paper studies the potential of a novel approach to ensure more efficient and intelligent assignment of capacity through medium access control (MAC) in practical wireless sensor networks. Q-Learning is employed as an intelligent transmission strategy. We review the existing MAC protocols in the context of Q-learning. A recently-proposed, ALOHA and Q-Learning based MAC scheme, ALOHA-Q, is considered which improves the channel performance significantly with a key benefit of simplicity. Practical implementation issues of ALOHA-Q are studied. We demonstrate the performance of the ALOHA-Q through extensive simulations and evaluations in various testbeds. A new *exploration/exploitation* method is proposed to strengthen the merits of the ALOHA-Q against dynamic the channel and environment conditions.

1. Introduction

Wireless sensor networks (WSNs) have gained ever increasing interest given their significant potential in various applications, ranging from environmental monitoring to industry, military and health applications (Akyildiz et al., 2002). Such a network is normally expected to include a large number of small, inexpensive and battery-operated sensor nodes which are typically deployed in an ad-hoc fashion to perform a common task. A common and unique characteristic of a WSN after deployment is the inaccessibility of sensor nodes which makes recharging or replacing exhausted batteries challenging or impractical. This inherent feature brings about the necessity to design robust operation of the networks without external intervention and access. A typical WSN employs a medium access control (MAC) protocol which regulates user access to the shared radio communications medium, affording significant impact on the overall network performance. Designing power efficient MAC protocols is therefore of paramount importance.

In the last two decades, researchers have focused on developing a wide range of MAC protocols for WSNs in an energy-efficient manner. Current MAC protocols can be broadly divided into *contention-based* and *schedule-based*. The majority of the proposed schemes are contention-based and inherently distributed, but they introduce energy waste through *overhearing, collisions, idle-listening and re-transmissions* (Demirkol et al., 2006). Schedule-based protocols can alleviate these sources of energy waste by dynamically assigning transmission schedules, but these benefits are incurred at the expense of higher complexity and overheads. On the other hand, there are a number of MAC schemes which combine the features of both contention-based and schedule-based approaches, called *hybrid* protocols. The main drawback of the hybrid schemes is their complexity which may make them suitable for only a limited number of applications.

Reinforcement learning (RL) has been recently applied to design new MAC protocols in WSNs. Many developed RL-based schemes aim to adaptively adjust the duty cycle of the nodes which is best illustrated by S-MAC (Ye et al., 2002). These protocols provide the nodes with an intelligent way of predicting other nodes' wake up times based upon the transmission history of the network. RL-based protocols significantly reduce the energy consumption due to both idle listening and overhearing in the context of duty cycling. ALOHA and Q-Learning have been integrated to establish a new MAC protocol, namely ALOHA-Q (Chu et al., 2012). ALOHA-based techniques are important for certain categories of Wireless Personal Networks (WPNs) and WSNs such as those based on Radio Frequency Identification (RFID) systems which have limited memory and power capabilities. The ALOHA-Q scheme inherits the merits of contention-based and schedule-based approaches while offsetting their drawbacks. ALOHA-Q uses slotted-ALOHA as the baseline protocol with a key benefit of simplicity. It allows users to find unique transmission slots in a

* Corresponding author at: Department of Electricity and Energy, University of Bayburt, Bayburt, Turkey.
E-mail addresses: skosunalp@bayburt.edu.tr (S. Kosunalp), yi.chu@york.ac.uk (Y. Chu), paul.mitchell@york.ac.uk (P.D. Mitchell), david.grace@york.ac.uk (D. Grace), tim.clarke@york.ac.uk (T. Clarke).

http://dx.doi.org/10.1016/j.engappai.2016.06.012
0952-1976/

fully distributed manner, resulting in a scheduled outcome. ALOHA-Q aims to reach an optimal steady state where each user within an interference range has unique transmission slots. Therefore, ALOHA-Q works like a contention-based scheme gradually transforming into a schedule-based scheme, which can be fully achieved in the steady-state without the need for centralised control and scheduling information exchange in steady state conditions. In order to demonstrate the effectiveness of this approach, performance evaluations in both single-hop and multi-hop communication scenarios have been carried out (Chu et al., 2012; Yan et al., 2013, 2014). These studies show that ALOHA-Q can achieve a high channel utilisation while minimising the energy consumption. However, these evaluations are restricted to simulation-based evaluation, where the practical implementation issues can be avoided.

This paper extends the implementation of ALOHA-Q in both simulation and realistic testbeds, to linear-chain, grid and random topologies. Practical implementation issues of ALOHA-Q are studied based upon hardware limitations and constraints. In order to strengthen the merits of the ALOHA-Q against dynamic the channel and environment conditions, the epsilon-greedy strategy is integrated. We compare the performance of the ALOHA-Q with a well-known MAC protocol Z-MAC (Rhee et al., 2008) and a self-learning protocol SSA (Niu and Deng, 2013). Section 2 presents the related work highlighting the main features of existing RL-based MAC protocols for WSNs. Section 3 introduces a brief description of ALOHA-Q and *exploration/exploitation* methods proposed. The performance of ALOHA-Q in simulation in comparison to ZMAC and SSA is validated with practical results in Section 4. We evaluate experimentally the performance of ALOHA-Q in two real-world events: (1) packet losses in steady state and (2) participation of new nodes to the network. ALOHA-Q with *exploration/exploitation* is implemented to provide better performance in these two events. Finally, Section V concludes the paper.

2. Related work

A wide range of MAC protocols for WSNs have been proposed with the primary objectives of improving energy-efficiency and channel throughput. The majority of the existing schemes are contention-based and employ CSMA as the baseline approach. In this section, our focus is on RL-based MAC schemes proposed in the literature.

S-MAC is a well-known RTS-CTS based MAC protocol that introduced the concept of a duty-cycle in which the nodes in the network periodically sleep, wake up, listen for potential transmissions, and return to sleep (Ye et al., 2002). A large number of existing protocols have used the theme of S-MAC in order to further schedule the nodes' sleep and active periods as the duty-cycle period in S-MAC is of a fixed duration. The nodes form virtual clusters to determine a common schedule among the neighbouring nodes. A small SYNC packet is exchanged among neighbours at the beginning of each active period to assure that they wake up concurrently to reduce the control overheads. After the SYNC exchange, the data packets can be transmitted using the RTS-CTS mechanism until the end of the active period. S-MAC adopts a *message passing* technique to reduce contention latency. A long message is fragmented into many small fragments which are sent in bursts. The main drawback of S-MAC is the predefined constant listen-sleep periods which can cause unnecessary energy consumption, particularly when some nodes those are located near the sink, may require a higher duty-cycle in order to serve the traffic passing through them.

RL-MAC is a RL-based protocol that adaptively adjusts the sleeping schedule based on local and neighbouring observations (Liu and Elhanany, 2006). The key property of this scheme is that the nodes can infer the state of other nodes using a Markov Decision Process (MDP). For local observation, each node records the number of successfully transmitted and received packets during the active period to be a part of the determination of a duty cycle. As for neighbouring observations, the number of failed attempts is added to the header to inform the receiver, which saves energy by minimizing the number of missed packets (early sleeping).

A decentralised RL approach is proposed to schedule the duty-cycle of sensor nodes based only on local observations (Mihaylov and Elhanany, 2012). The main point of this study is to shift active periods in time based on transmission paths and ranges along a route. Similar to that of S-MAC, the wake-up periods of the nodes which need to communicate with each other are synchronised, whereas the schedules of nodes on neighbouring branches are desynchronised to avoid interference and packet losses. The active periods are further divided into time slots and the nodes are allowed to wake for a predefined number of slots in each period. The slots where a node will wake up are decided by the learning process. Each slot within a frame is given a quality value which shows the efficiency of the slots. The quality values of slots are updated by rewarding successful transmissions and penalising the negative interactions. As a result of the learning process, the quality values of one group of slots become strictly higher than the rest.

Another similar approach combining the Slotted-ALOHA and Q-Learning algorithms which achieves high channel utilisation while mitigating energy consumption is ALOHA-Q (Chu et al., 2012). Each node repeats a frame structure which includes a certain number of time slots. The packets are transmitted in these time slots. Each node stores a Q-value (equivalent to the quality value above) for each time slot within a frame. The Q-value of a slot is updated individually when a transmission happens and this is used to explore each slot more frequently. Successful transmissions are denoted by small acknowledgement packets which are immediately sent upon the correct reception of the packets. The nodes generate a positive outcome (reward) when the acknowledgement packet is received, otherwise a punishment is applied to update the Q-value. Consequently, the slots with higher Q-values are preferred for data communication and this behaviour repeats the same actions. This process continually returns rewards which serve to decrease the probability of unsuccessful transmission. Eventually, the learning process leads the network to a steady state condition where unique slots are assigned to the nodes. Although this approach appears to be a schedule-based network, it does not require any scheduling information exchange. However, it is critical to set a sufficient number of slots within a frame. Redundant slots in a frame will result in achieving lower channel throughput. None of the work described above has addressed optimum frame size assignment problem. Instead, Reference (Yan et al., 2013) presents a distributed frame size selection algorithm for ALOHA-Q in a single-hop scenario. Furthermore, the selection of the frame size has been discussed for a multi-hop wireless sensor networking (Yan et al., 2014).

QL-MAC, Q-Learning-based MAC, is proposed to deal with the issue of finding an efficient way of scheduling the duty-cycle of sensor nodes (Stefano et al., 2013). Each node considers its own traffic load and the network load of its neighbours. The basic underlying concept of QL-MAC resembles that of a decentralised RL approach, whereby time is divided into time slots (*frames*) which are further divided into smaller time units (slots). Every node using the Q-Learning algorithm individually limits the total number of slots in which the node wakes up. The frame length and the number of slots constituting the frame remain constant.

All of the protocols introduced in this section have only been evaluated through simulations. The feasibility of the schemes for