Ad Hoc Networks 34 (2015) 105–120

Contents lists available at ScienceDirect

Ad Hoc Networks

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

Utilizing kinematics and selective sweeping in reinforcement learning-based routing algorithms for underwater networks

R. Plate *, C. Wakayama

Maritime Systems Division, SPAWAR Systems Center Pacific, Code 56490, 53560 Hull St., San Diego, CA 92152-5001, USA

ARTICLE INFO

Article history:
Received 7 February 2014
Received in revised form 8 September 2014
Accepted 19 September 2014
Available online 6 October 2014

Keywords:
Reinforcement learning
Routing protocol
Acoustic communication
Underwater sensor network

ABSTRACT

Effective utilization of mobile ad hoc underwater distributed networks is challenging due to high system costs and the harsh environment characterized by low bandwidth, large latency, high energy consumption, and node mobility. This work addresses the routing issue, which is critical in successfully establishing and utilizing an underwater network. In particular, it focuses on reinforcement learning (RL)-based routing algorithms, which possess the ability to explore the network environment and adapt routing decisions to the constantly changing topology of the network due to node mobility and energy usage. This paper presents a routing algorithm based on *Q*-learning, one of the RL approaches, with additional Kinematic and Sweeping features, therefore referred to as QKS. These two additional features are introduced to address the potential slow convergence associated with pure RL algorithms. The results of a detailed packet-level simulation have been obtained using the NS-2 open-source network simulator with underwater modeling additions. The energy efficiency, convergence, and delivery performance of QKS are compared with two other routing protocols for underwater networks, a basic flooding approach (ICRP (Liang, 2007)) and a basic *Q*-learning implementation (QELAR (Hu, 2010)), using simulations of networks with both fixed and mobile nodes.

1. Introduction and related work

Given the advances in underwater transmission capabilities, underwater wireless networking has emerged as an active research area with applications in a wide range of scientific and military domains [1–3]. The physical limitations of the underwater acoustic environment characterized by long propagation delay, low bandwidth, high energy consumption and node mobility prevent the direct use of conventional routing protocols developed for radio frequency (RF) ad hoc networks. These factors must be carefully taken into account in designing wireless underwater acoustic networks.

Conventional routing protocols developed for RF ad hoc networks can be classified into three basic categories: table-based proactive, on-demand reactive, and geographical [4]. Each of these has inherent weaknesses when applied in underwater acoustic environments. Proactive routing has high maintenance costs for storing information about neighbor nodes every time there is a topology change, while reactive routing has high route discovery costs to gather neighbor information as it is needed. Geographic routing relies on position and/or velocity information, which can be difficult to obtain or inaccurate depending on node capabilities. Combining the strengths of these three can produce a more robust routing approach adequate for underwater applications.

Many existing routing protocols for underwater networks address particular underwater challenges and

* Corresponding author. Tel.: +1 (619) 553 2076.
E-mail addresses: rplate@spawar.navy.mil (R. Plate), wakayama@spawar.navy.mil (C. Wakayama).

http://dx.doi.org/10.1016/j.adhoc.2014.09.012
1570-8705/

depend on limiting application scenarios. A recent survey on routing protocols for underwater acoustic wireless sensor networks can be found in [5]; we present a subset of the existing protocols according to their special features addressing node mobility and energy efficiency. Vector-based-forwarding (VBF), depth-based-routing (DBR), and mobicast routing (MR) protocols address the challenges due to node mobility. The VBF protocol [6] establishes a routing pipe between the source and destination, and only the nodes in the routing pipe can participate in message forwarding. The positions of the source and the destination are assumed available in defining a routing vector between the source and the destination. In DBR [7], each node accounts for its depth and the depth of the previous node, and uses a greedy algorithm to forward packets only along the upward directions. The MR [8] protocol uses predefined routes for autonomous underwater vehicles (AUVs) for gathering data in a predefined sensor network. A dedicated AUV is controlled to follow a specified route to sequentially service nodes that are allocated into three-dimensional zones of reference (3D-ZOR); nodes in adjacent 3D-ZOR must be woken up to be queried while trying to avoid topology holes due to node drift by the ocean current.

Limited battery energy is a fundamental issue for underwater nodes and many routing efforts focus on energy efficiency and network lifetime. Focused beam routing (FBR) [9] is a cross-layer approach in which the routing protocol, the medium access control and the physical layer functionalities are coupled by power control to establish energy efficient routes dynamically. In FBR, the positions of both the source and destinations must be known. Energy-efficient routing protocol (EUROP) described in [10] is designed to reduce broadcasting hello messages by equipping each node with pressure sensing capability. Each node calculates its depth, determines its corresponding layer, and selects the next node by applying a deep to shallow routing rule; it is assumed that the destination node is located on the water surface. Another Energy-efficient Routing Protocol based on Physical distance and Residual energy (ERP^2R) [11] takes into account the physical distances of the nodes towards the destination and the residual energy of the nodes to determine the next forwarder in order to extend the network lifetime. The information-carrying-based routing protocol (ICRP) [12] is a reactive protocol designed to provide energy efficient, real-time and scalable routing based on a limited flooding approach. Due to the high overhead of finding new routes as they change, ICRP may not be appropriate for underwater networks consisting of high mobility nodes, yet is attractive due to its simplicity.

The underwater challenges and limited applications of existing underwater routing protocols demand robust, adaptive, and energy-efficient routing approaches which require minimum a priori knowledge about the network and restrictions on network architecture. These requirements motivate the development of underwater routing protocols based on machine learning techniques. Multi-agent machine learning techniques, such as Multi-Agent Reinforcement Learning (MARL), have been applied successfully in many problem domains involving distributed decision making in which each agent must learn by interacting with its dynamic environment and attempt to discover optimal actions on its own [13]. A fully distributed underwater routing problem can be seen as a collaborative multi-agent system. In [14], a commonly used machine learning technique known as *Q*-learning is implemented in a routing algorithm called QELAR to address a fully distributed architecture. Nodes compute their own routing decisions by storing routing information (*Q*-values) of their direct neighbor nodes. In QELAR, the *Q*-value estimates consider the energy consumption of sensor nodes and residual energy distribution among neighboring nodes to optimize total energy consumption and network lifetime. The environment is learned as the estimated *Q*-values converge to reflect the network topology. However, convergence could occur slowly for certain network configurations such as those comprised of a large number of nodes and/or nodes with high mobility, resulting in excess resource usage. A multi-level routing protocol (MURAO) [15] is a two-layer *Q*-learning based routing protocol which is designed for dense, hybrid acoustic-optical underwater sensor networks. In MURAO, a cluster-based two layer network is formed, where neighboring clusters of nodes share at least one common gateway node. A cluster head is associated with each cluster and directs inter-cluster routing, while the individual nodes of the clusters perform intra-cluster routing. The slow convergence of *Q*-learning associated with a dense network is partly addressed by allowing routing in different clusters to take place in parallel.

In this paper, we investigate two additional features that can improve the convergence rate of a basic *Q*-learning algorithm and its ability to track changes in the network while balancing the routing overhead. The first is the use of kinematic information to add a geographic component to the RL approach. Nodes transmit their own position and velocity estimates and store those of their neighbors to enable more accurate estimation of successful transmission probability, resulting in fewer failed transmissions. We assume that a node can estimate its own position and velocity (either independently or collaboratively) but is unaware of the destination location. The second is the addition of selective backward exploration (sweeping) to the forward exploration of QELAR, such that nodes actively propagate significant changes in their *Q*-values back toward the source. Although overhead cost is increased with the addition of each feature, the improved convergence rate and tracking of network changes results in an overall improvement in energy consumption and/or latency as compared to the baseline *Q*-routing approach. We will refer to our algorithm incorporating these features as QKS (*Q*-learning utilizing Kinematics and Sweeping). Our objective is to investigate the performance of RL techniques as applied to routing in underwater networks and evaluate how the two additional features developed can improve performance and address certain weaknesses in a more basic *Q*-learning implementation. Detailed packet-level simulations are run to understand the trade-offs achievable when employing these algorithms in different network scenarios so that the best features/parameter values can be chosen for maximal efficiency.

This paper proceeds as follows. In Section 2, we introduce *Q*-learning and a baseline algorithm, QELAR. In