Computer Networks 146 (2018) 1–11

Contents lists available at ScienceDirect

Computer Networks

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

A data sample algorithm applied to wireless sensor network with disruptive connections

Israel L.C. Vasconcelos[a], Ivan C. Martins[a], Carlos M.S. Figueiredo[b], Andre L.L. Aquino[a,*]

[a] Computer Institute, Federal University of Alagoas, Maceió, AL, Brazil
[b] Department of Computer Engineering, State University of Amazonas, Manaus, AM, Brazil

ARTICLE INFO

Article history:
Received 7 June 2018
Revised 9 August 2018
Accepted 5 September 2018
Available online 11 September 2018

Keywords:
Wireless sensor networks
Data sample algorithms
Application modeling

ABSTRACT

This paper presents a data sample algorithm applied to wireless sensor network applications with disruptive connections. Additionally, it defines a model for delay-tolerant sensor network where drop strategies are applied to improve the phenomenon coverage in an application that monitors the forest temperature incidence for wildlife observation. The environmental application model comprises: i) Phenomenon generation based on a Gaussian random field along with the Matern covariance model; ii) Sensing nodes deployment based on simple sequential inhibition process with a mobile sink node following a random walk process; iii) Data collection and processing based on a data-aware drop strategy; and iv) Phenomenon reconstruction based on simple kriging interpolation. This research employed the data-aware drop strategy and compared it with the others, reported in the literature. Besides the satisfactory application of this model, the results show that the performance of data-aware drop strategy is twice better than conventional ones in all evaluated scenarios.

1. Introduction

The world around us has a variety of phenomena that can be monitored by devices provided with sensing, processing and communication capabilities. While cooperatively working in an area of interest, such devices comprise a wireless sensor network (WSN) [1,2]. The primary task of these networks is sending information, such as temperature, humidity, and pressure, to an outside observer in a multi-hop manner.

In general, to achieve a continuous communication among the sensor nodes, in applications with a sparse area of interest, is a complex task. The strategy applicable to collect data, in this case, is the adoption of the concept of delay-tolerant networks (DTNs) [3,4]. The main idea is to use mobile nodes that navigate through the monitored area and opportunistically collect data from different nodes that communicate with each other. This combination of WSNs and DTNs defines the concept of Intermittently Connected Delay-Tolerant WSNs (ICDT-WSNs) [5,6]. These networks may be useful for different applications, such as wildlife tracking [7], water quality monitoring [8], urban monitoring [9], and surveillance or disaster-recovery infrastructures [6].

For instance, this study case considers an ICDT-WSN application of monitoring the forest temperature incidence for wildlife observation. However, any other generic ICDT-WSN application can inherit from the chosen scenario with all targeted settings from this study. This monitoring solution comprises sensor nodes that are deployed over an area of interest and acquire measurement data. This application should deal with long distance gaps observed in sparse environments by turning off the communication between the sensing nodes and becoming the sink to be mobile. The above-cited mobile sink travels across the area collecting packets from these sensor nodes and that processes their data. Finally, an observer that receives all the data from the mobile sink node and reconstructs an overall view of the network.

To ensure accurate functioning, the nodes in these networks should sense and store data until a communication opportunity is present. However, the network size, the amount of sensed data, and the frequency of mobile contacts overload the nodes' storage capacity. The above issue encourages the use of packet drop strategies [10]. Besides, applying some drop strategy is considerably useful for energy sparing of the battery-operated nodes due to the less transmitted data as well as for transmitting the most critical data when the opportunistic contact is quick. The problem with the traditional work in the literature [11] is that it considers dropping techniques as a network layer problem, i.e., the techniques are not data-aware as a WSN application is expected to be. Thus,

* Corresponding author.
E-mail addresses: israel.vasconcelos@laccan.ufal.br (I.L.C. Vasconcelos), ivan.martins@laccan.ufal.br (I.C. Martins), cfigueiredo@uea.edu.br (C.M.S. Figueiredo), alla@laccan.ufal.br (A.L.L. Aquino).

https://doi.org/10.1016/j.comnet.2018.09.006
1389-1286/

these conventional solutions may seriously affect the quality of the overall network observation.

To deal with the constraints mentioned above, it applies an algorithm based on histogram data sampling to provide a the data-aware drop solution. This solution considers that every time a mobile sink node acquires data packets from other sensor nodes in the environment, it analyzes the collected and stored data to apply a sampling algorithm for maintaining only part of the overall data. The kernel of data sampling algorithm is based on the previous research of Aquino and Nakamura [12] to deal with real-time communication constraints. The functions of this algorithm include sorting the data in ascending order, grouping them in equal-ranged histogram classes, and sampling the central elements (the medians) of each histogram class. As a result, from the entire data set, only the most representative data is stored. The mobile sink nodes successively employ this method using the monitored area, and at the end of the data acquisition process, an observer reconstructs the entire network vision by using a simple kriging interpolation.

Besides the model proposed, an additional contribution of this research is the proposed delay-tolerant sensor network model where drop solutions are applied to improve the phenomenon coverage. The proposed solution modeling comprises of: i) Phenomenon generation based on Gaussian random field [13] along with the Matern covariance model [14]; ii) Sensing nodes deployment based on simple sequential inhibition process [15] with a mobile sink following a random walk process [16]; iii) Data collection and processing through a data-aware drop strategy based on histogram sampling algorithm [12]; and iv) Phenomenon reconstruction based on simple kriging interpolation [17]. Among the above steps, only the phenomenon generation is specific to this study case application that monitors the forest temperature incidence for wildlife observation. Due to the specificity of each application, other ones must consider a specific data generation model. The other steps can be applied directly in general applications where drop solutions are adopted.

The proposed model is applied satisfactorily, and the data-aware drop solution based on sampling algorithm is evaluated from simulated scenarios wherein different events can occur in an environment. The phenomena are characterized by possible events that alter the distribution of data values over the sensed area. It compared this solution with those reported in the literature, and the principal evaluation is the quality of reconstruction of the observed phenomena. The results show that the performance of the proposed solution is twice better than the common drop strategies in all evaluated scenarios. Besides, this solution can cover, for the most part, an area three times wider than others. The use of a data-aware drop strategy in delay-tolerant scenarios represents an additional and relevant contribution to this field.

The organization of this study is as follows. Section 2 discusses the related work. Section 3 presents the ICDT-WSN application model. Section 4 discusses the results of this research. Finally, Section 5 concludes the study and suggests some future directions.

2. Related work

As previously mentioned, the research compares the strategy developed with the general strategies used in DTNs. These strategies, as presented by Jain and Chawla [11], are discussed as follows:

- **Drop the last packet:** This is an intuitive easy-to-deploy policy which eliminates the last received message. This policy suggests that old data are more important than new data, because the newer data may provide communication opportunities at a later stage.
- **Drop the first packet:** This is also a quite intuitive policy that eliminates the older message first. This policy suggests that old data do not represent the monitoring needs.
- **Drop the packet randomly:** This policy eliminates the packet randomly. The policy suggests that all received data have the same importance.
- **Drop based on message size:** This policy discards information based on the buffer size. It discards the more considerable size information or information of a given size. The policy focuses on the utilization rate of the buffer to decide which message it needs to drop.
- **Drop based on the number of forwards:** This policy eliminates the messages forwarded several times. The information in these messages has the highest probability to have a copy somewhere in the network or to already have arrived at its destination. The forward occurs when the sensor nodes use, in an ad-hoc fashion, other nodes to achieve the sink node, i.e., it does not need direct contact between the sink and sensor nodes. The designer is responsible for defining the maximum number of forwarding and retransmission depending on project restrictions. Deploying this policy requires the tracking of the number of forwarded messages in the network.
- **Drop based on lifetime:** This policy discards the messages that have been running through the network for a longer time, i.e. with a shorter remaining lifetime. This policy suggests that a message with a short life has already been existing in the network for a considerable amount of time and it is more likely that it arrives at its destination.
- **Drop based on global network coverage:** This is a policy that performs the data analysis on the data stream, discarding packets following criteria which allow the application to cover a wider area, i.e., nodes neighborhood, intersections, size of packets, among others.

A specific strategy used in ICDT-WSNs is the drop based on network coverage CBDP [18]. This strategy requires data to calculate the global coverage and then choose which messages to discard. The strategy aims to minimize the impact of drops on the monitoring application. Aoun et al. [19] propose another strategy that uses optimization methods to identify when and how multiple packets must be discarded to achieve the application's real-time deadline.

Vasconcelos et al. [20] evaluate some mentioned strategies. This paper showed that the random packet drop strategy was the best-behaved technique among the traditional ones. For this reason, some experiments presented in Section 4 only consider the strategy in comparison to the algorithm proposed in this study. Additionally, this research performs a complete evaluation in a simulated delay tolerant WSN environment applied to monitor the forest temperature incidence where a more accurate sink mobility model was used based on random walking.

There are proposals about buffer management algorithm in ICDT-WSNs. Benhamida et al. [21] survey, in application level, different solutions to support new emerging network-based applications, in which DTN techniques can be used to offer delay tolerant support, for instance, Internet of Things (IoT), and consequently the WSN infrastructure used on this applications. Söderman et al. [22] propose, in a routing level, a strategy that continuously measures a parameter and the value of a single packet is the estimation of its importance in the recreation of the original signal. Maranhão et al. [23] present, in a radio level, a scheme based on multi-hop cooperative multiple/single input multiple/single output transmission for ICDT-WSN to achieve longer communication ranges and consequently reduce the message delivery time to the sink, maximizing the energy efficiency.

The main problem with the above strategies is that they only consider packet-related aspects in the network layer. Thus, they