Ad Hoc Networks 49 (2016) 70–89

Contents lists available at ScienceDirect

Ad Hoc Networks

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

Feature selection for performance characterization in multi-hop wireless sensor networks

Athanasia Panousopoulou[a,*], Mikel Azkune[b], Panagiotis Tsakalides[a,c]

[a] Institute of Computer Science, Foundation of Research and Technology, Hellas, Heraklion, Crete, Greece
[b] Applied Photonics Group,Faculty of Engineering, University of the Basque Country, Bizkaia, Spain
[c] Department of Computer Science, University of Crete, Heraklion, Crete, Greece

ARTICLE INFO

Article history:
Received 8 October 2015
Revised 28 April 2016
Accepted 20 June 2016
Available online 21 June 2016

Keywords:
Wireless sensor networks
Unsupervised learning
Network measurement and analysis
Testbeds and experimental evaluation

ABSTRACT

Current trends in Wireless Sensor Networks are faced with the challenge of shifting from testbeds in controlled environments to real-life deployments, characterized by unattended and long-term operation. The network performance in such settings depends on various factors, ranging from the operational space, the behavior of the protocol stack, the intra-network dynamics, and the status of each individual node. As such, characterizing the network's high-level performance based exclusively on link-quality estimation, can yield episodic snapshots on the performance of specific, point-to-point links. The objective of this work is to provide an integrated framework for the unsupervised selection of the dominant features that have crucial impact on the performance of end-to-end links, established over a multi-hop topology. Our focus is on compressing the original feature vector of network parameters, by eliminating redundant network attributes with predictable behavior. The proposed approach is implemented alongside different cases of protocol stacks and evaluated on data collected from real-life deployments in rural and industrial environments. Discussions on the efficacy of the proposed scheme, and the dominant network characteristics per deployment are offered.

1. Introduction

Over the past years, Wireless Sensor Networks (WSN) have been closing the gap between theory and application in real-life scenarios, thereby gaining prominence as the key enabling technology for addressing significant societal challenges [1–3]. Exploiting WSN-based schemes for solving modern engineering problems, intensifies the necessity of transiting from episodic sampling to truly pervasive paradigms relying on resilient, long-term, and unattended operation. As such, monitoring and characterizing the performance of the network in realistic deployments is gaining increasing interest [4], as a process influenced by multiple factors. Recent works [5,6] emphasize the necessity for providing systematic tools, capable of capturing a variety of different aspects of radio transmission and wireless network deployments. High level requirements, such as application-driven positioning and scale, can impact the network performance [7]. The behavior of multi-hop links is dominated by the dynamics of wireless connectivity and power autonomy, even when the sensor nodes are in fixed positions [8]. The combination of the operational space and the hardware characteristics become key factors. Finally, from the perspective of application-driven deployments, guaranteeing the desired Quality of Service is considered more important than the low-level details of sophisticated protocol stacks.

Addressing the aforementioned challenges can be accelerated by employing passive monitoring mechanisms that observe the performance of user-designated end-to-end links. By the term "end-to-end", we refer to network links, which are built over a multi-hop network topology and are responsible for the application-driven data flow. Opposed to the well-studied point-to-point links that are formulated at the Physical layer, and are capable of link quality estimation between 1-hop neighbors, end-to-end links expand towards two different directions: (a) across different sides of the network, exceeding the constrained limits of 1-hop neighborhoods, (b) across different layers of a fully functional protocol stack, ranging from the Physical to the Transport and Application layers. As such, end-to-end links convey a larger volume of information than the one captured by point-to-point, low-level links. Thus, the systematic study of their performance could provide the means for understanding the multi-dimensional behavior of the entire network.

Enabling the systematic collection and process of sufficient amounts of data for characterizing the performance of end-to-

* Corresponding author.
E-mail address: apanouso@ics.forth.gr (A. Panousopoulou).

http://dx.doi.org/10.1016/j.adhoc.2016.06.011

end links can be a difficult task, especially in real-life WSN deployments. A significant challenge is related to the discovery of correlations in the network measurements for understanding the network performance. Towards this direction, two key aspects should be simultaneously addressed, namely: (a) identify and discard the attributes of the data that provide no significant correlations between network measurements and network performance, and (b) calculate and retain the attributes that convey the essential information for extracting the required correlations on the network measurements. To tackle these issues, Machine Learning techniques [9] in general, and feature-level fusion in particular, can be employed for selecting the *dominant* set of features, which have the greater impact on inferring the network performance of multi-hop, end-to-end links, based on historical data patterns.

Notably, Machine Learning algorithms, have recently been proposed for providing solutions to various WSN aspects, namely routing, localization and tracking, intrusion detection, and hierarchical data aggregation [10]. Nevertheless, in this work, our objective is to examine the efficacy of feature-level fusion on capturing and understanding the dynamics of end-to-end links in real-life deployments. We propose an integrated framework for the systematic collection of the essential network measurements, the extraction of the network features, and the design of a novel algorithm for unsupervised feature selection. We integrate dimensionality reduction with information compression towards feature selection techniques that fit the WSN paradigm, in terms of lightweight implementation on memory-constrained operational devices. The resulting framework has been implemented and deployed in real-life multi-hop WSN deployments in rural and industrial environments, while remaining independent from the technical details of the implemented protocol stacks. The proposed approach applies on elongated and unattended operation of the network, resulting to traffic that exceeds 680,000 instances recorded at the Application layer. The efficacy of our feature selection technique is evaluated against widely known supervised and unsupervised feature selection algorithms, and the results indicate the superiority of the herein proposed method in machine learning terms. In addition, the results justify the importance of extending the characteristics of the information available in the feature vector beyond parameters and metrics captured at the Physical layer.

In a nutshell, our contributions are summarized as follows:

- A design of a framework for applying unsupervised learning techniques over multi-hop WSN topologies that considers a collection of diverse network parameters in a passive fashion, which introduces to the network neither additional, nor dedicated traffic;
- The combination of network metrics collected from different sides of the network, and corresponding to different layers of the protocol stack to a feature-level fusion mechanism for delivering high-level inference on the dominant network features;
- The synthesis of a thorough learning model for characterizing the performance of a multi-hop WSN, that covers the formulation of the classification problem and the engineering of network features;
- The application of the proposed framework on real-life deployments and the explanation of the findings within the WSN context.

The remainder of the paper is organized as follows: in Section 2, the current state of the art in link performance estimation is outlined. In Section 3, the problem is formulated, accompanied by the proposed feature selection framework in Section 4. Evaluation methodology and experimental results are presented in Section 5, while conclusions are drawn in Section 6.

2. Related work

Current trends on experimentally characterizing the network performance concentrate exclusively on link quality aspects of point-to-point links resulting to either empirical studies [11–13] or behavior analysis based on tools adopted by machine learning [14–17]. The authors in [11] have performed experimental studies on WSNs in controlled environments and the implications of common assumptions on the packet delivery performance of WSN by using commercial transceivers. The emphasis was on how observed quantities, such as the Received Signal Strength Indicator (RSSI), the Link Quality Indicator (LQI), the Signal-to-Noise Ratio (SNR), and the Acknowledgment Reception Ratio (ARR) can be used for explaining the observed link behavior. Their key finding was that the spatial and temporal correlation, along with link asymmetries are the dominant, qualitative characteristics of point-to-point link behavior. In addition, they observed that the statistical attributes of LQI per packet offer a better correlation with Packet Reception Ratio per link, than the one provided by RSSI. Along the same lines, in [12] Baccour et al. surveyed the experimental studies for link quality estimation in controlled environments, highlighting the fact that different experimental conditions yield different results. According to the authors this is due to (a) the lack of standardization in terms of evaluation metrics, assumptions, and approach; (b) the asymmetry of the hardware employed introducing antennae irregularities, dependency of radio transceivers on temperature and humidity, and radio hardware inaccuracy. While this inconsistency intensifies when using independently LQI or SNR in order to characterize links with moderate performance, non-linear combinations of link-layer quality metrics can yield a fast and reliable assessment of point-to-point link quality [13].

In parallel to the empirical characterization of network performance in link-quality terms, exploiting learning techniques [18] for performance estimation has been gaining an increasing interest during the past few years. The efficacy of supervised learning, involving two primary phases, namely offline training and online classification, has been evaluated in [14] in the context of point-to-point links. The methodology adopted emphasizes on classification and the conclusions derived highlight the benefits of Decision Tree Learners [18] for estimating the link quality performance, in terms of computational complexity and accuracy. In [15] a distributed online protocol is introduced in order to estimate wireless link quality based on supervised incremental learning methods. The approach adopted combines Locally Weighted Projection [19] and locally available measures of direct links, such as SNR and traffic rate, towards building regression maps between the local network configuration and the expected link quality. In addition, the authors in [16] combine the value of Packet Reception Ratio, and the levels of RSSI, LQI, and SNR with logistic regression classifiers. By the means of a three-step procedure which involves (a) data collection from point-to-point links, (b) off-line training, and (c) on-line prediction, the proposed approach yields routing metrics capable of predicting the success probability of the next packet. In a similar fashion, Stochastic Gradient Descent is employed in [17] in order to address aspects related to the estimation of links with moderate performance. The resulting on-line and unsupervised schemes are integrated with low power listening protocols towards adaptive schemes for link-quality prediction.

A common characteristic of the aforementioned approaches, is that the analysis of the network behavior relies on point-to-point non-competitive links, which are part of well-defined testbeds [20–22] in controlled environments. Shifting towards realistic WSN deployments, recent works examine the performance of WSN in RF-harsh environments, and in particularly in applications associated to Smart Grids [23,24]. The evaluations conducted on a simulation basis, emphasize on the link quality estimation with