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
1570-8705/

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

respect to the RF-harshness of smart grid environments. The observations conclude to the fact that estimators that consider the link asymmetry [23], while extending to metrics associated to the MAC layer (e.g., packet delivery and retransmissions) and channel quality [24], yield the optimal performance.

The discussion thus far highlights the literature gap on extending the network performance characterization well beyond link quality estimation, and addressing the behavior of end-to-end links that are built upon a multi-hop topology. As such, realistic factors that span across different sides of the network, and different layers of the protocol stack should be unified and taken into account. Towards this direction, our approach differs from the related bibliography in the following ways:

- We extend the problem of network performance characterization to multi-hop network topologies, while we additionally consider a wider range of network parameters that span across all layers of the protocol stack, going well beyond traditional link quality metrics, i.e., RSSI, LQI, SNR;
- Opposed to the current state of the art, in our approach we do not attempt to apply classification techniques for the network performance. We concentrate on the problem of feature selection, which is in principle a search problem with the objective of reducing the dimensionality of a search space in typical data mining applications. As such, appropriate feature selection can improve the quality of classification, and we leverage on this potential for compressing redundant attributes as well as characterizing the dominant factors that impact the behavior of multi-hop links. In addition, motivated by the WSN paradigm, we propose a novel unsupervised feature selection technique that combines features clustering with uncertainty, and can be applied to different application domains;
- Instead of introducing a methodology that considers the generation of dedicated traffic for the collection and the calculation of the network attributes, we adopt a structured, passive approach for capturing network parameters available at different layers of the protocol stack, while each data packet travels from its source to its destination;
- The current trend on empirical performance characterization considers point-to-point networks in well-controlled environments, wherein power limitations, ambient conditions, and concurrent transmissions are not taken into account. By contrast, we apply our framework on application-driven multi-hop deployments, characterized by the expected accompanying challenges, such as competitive protocols for accessing the medium, battery-operated sensor nodes, and increased exposure to the harshness of the surrounding environment.

3. Problem formulation

We consider a WSN comprised of energy-autonomous, power-constrained, and IEEE 802.15.4-compliant N nodes, which are operating over long periods of time in an unattended fashion. Each node implements a full protocol stack, covering communication aspects that range from the Physical to the Application Layer. Without loss of generality, the lifetime of each sensor node is dictated by the adopted network policy, the hardware characteristics of the transceiver chip, and the input voltage supply (Appendix A).

The in-network operation is dictated by the establishment of *end-to-end links*, expressing the unicast connections established between different sensor nodes at the Application Layer. Each end-to-end link $i \rightarrow j$ between two sensor nodes i and j is constructed over a network path $P_{ij} = \{i, \ldots, k, \ldots, j\}$, where i is the transmitter, j is the receiver, and each kth node is a relay node between i and j. An example is shown in Fig. 1, highlighting the link $i \rightarrow j$ and the path P_{ij}, over a multi-hop WSN.

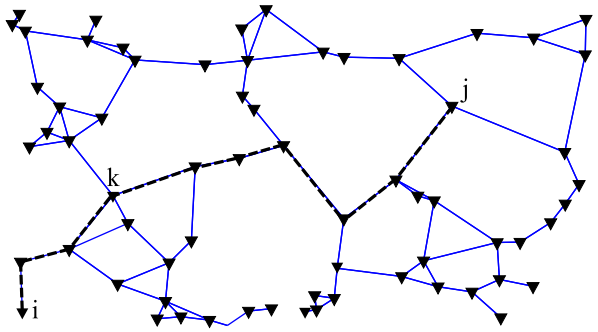

Fig. 1. An example of an end-to-end link $i \rightarrow j$ and the corresponding path P_{ij} (dashed line), established over a multi-hop network topology (solid line).

The key assumption made is that during the normal operation of the network, each node $k \in P_{ij}$ can monitor simple network metrics, which are related both to its own functionality, as well as the quality of the link $i \rightarrow j$. These metrics span across the protocol stack, while remaining independent of the specific protocol solution adopted by each layer. Representative parameters are the RSSI, LQI per received packet, the noise floor (NF), the unicast network activity at the MAC layer, the battery level, and the on-board temperature, and humidity.

Except for these network heuristics, the performance of the link $i \rightarrow j$ is periodically monitored at the side of the jth node. Without loss of generality, the performance of the $i \rightarrow j$ link is translated in the Packet Reception Ratio $PRR_{ij} \in [0, 1]$, which is defined as the ratio between the packets received by the jth node and the packets transmitted by the ith node. The value of PRR_{ij} can be classified into discrete, user-defined labels l_{ij} [12,13]. Typically, three thresholds are defined, α_E, α_G, α_P, $\in [0, 1]$, where $\alpha_E \geq \alpha_G \geq \alpha_P$. Based on these thresholds the performance of the $i \rightarrow j$ link is categorized as follows:

$$l_{ij} = \begin{cases} \text{Excellent}, & \text{if } \alpha_E \leq PRR_{ij} \leq 1 \\ \text{Good}, & \text{if } \alpha_G \leq PRR_{ij} < \alpha_E \\ \text{Problematic}, & \text{if } \alpha_P \leq PRR_{ij} < \alpha_G \\ \text{Poor}, & \text{if } 0 \leq PRR_{ij} < \alpha_P \end{cases} \quad (1)$$

The problem at hand is the automated calculation of the network factors that are responsible for classifying the performance of the link $i \rightarrow j$ to different values of l_{ij}. In well-controlled environments, characterized by single-hop links and congestion-free protocols, such associations are simplified in characterizing the quality of the link in terms of physical layer metrics. As we shift towards more complex experimental settings, PRR_{ij} can be affected by different factors that vary with respect to the operational space, the ambient conditions, and intra-network behavior. For instance, nodes may have irregular transmission patterns due to the operational environment [25], and the on-board temperature [26,27], while suffering from interference, and compete with their neighbors for accessing the transmission medium [28]. In addition, their performance is subject to their lifetime, and thus severely affected by the adopted energy replenishment policy [29,30] (e.g., replacement of conventional batteries, renewable energy resources, energy transfer) and related network operations, such as the adopted radio duty cycle policy, and the transmission/reception data rates at the MAC and Network layers.

With these considerations in mind, the acquisition of different network metrics available at different layers and different sides of P_{ij} can provide a set of M features and form the feature vector $\mathbf{f}_{ij}$ for each link $i \rightarrow j$:

$$\mathbf{f}_{ij} \rightarrow \left(PRX^*_{ij}, LQI^*_{ij}, NF^*_{ij}, f^{tx}_i, f^{rx}_i, \left|P_{ij}\right|, \widehat{PRR}_{ij}, T_i, H_i, V_i\right). \quad (2)$$

All related metrics are summarized in Table 1, while Table 2 presents the notation used throughout work.

Table 1
The network metrics employed for forming the feature vector $\mathbf{f}_{ij}$, defined in Eq. (2).

Network metric	Description
PRX^*_{ij}	Receiver Power over path P_{ij} (dBm)
LQI^*_{ij}	Link Quality Indicator over path P_{ij}
NF^*_{ij}	Noise Floor over path P_{ij} (dBm)
f_i^{tx}	Transmission rate at the MAC layer, excl. neighborhood discovery (bpm)
f_i^{rx}	Reception rate at the MAC layer, excl. neighborhood discovery (bpm)
$\|P_{ij}\|$	Length of path P_{ij}
$\overline{PRR}_{ij}$	The Windowed Mean Exponential Moving Average (WMEWMA) of PRR_{ij} [31]
T_i	On-board temperature of the ith node (^oC)
H_i	Percentage of on-board humidity for the ith node
V_i	Input power level for the ith node (Volt)

Table 2
The mathematical notation.

Description	Symbol
Network and node parameters (Section 3)	
Network size	N
Sensor nodes	i, j, k
End-to-end link between i and j	$i \rightarrow j$
Multi-hop routing path for establishing $i \rightarrow j$	P_{ij}
Lifetime and power consumption on the ith node	τ_i, P_i^{cons}
Packet reception rate for link $i \rightarrow j$	PRR_{ij}
User-defined label for classifying PRR_{ij}	l_{ij}
Thresholds for categorizing network performance (Eq. (1))	α_E, α_G, α_P
Vector of features related to network metrics	$\mathbf{f}_{ij}$
Length of $\mathbf{f}_{ij}$	M
Network metrics used for extracting $\mathbf{f}_{ij}$	Listed in Table 1
Reduced vector containing the dominant features	$\mathbf{f}^*_{ij}$
Length of $\mathbf{f}^*_{ij}$	R
Feature extraction and selection (Section 4)	
Mean value operator	$\mu(.)$
Standard deviation operator	$\sigma(.)$
Batch on subsequent NM measurements for the feature extraction	W
Length of fixed windows of measurements for extracting $\mathbf{f}_{ij}$	w
Features matrix corresponding to $\mathbf{f}_{ij}(D \times M)$	$\mathbf{A}$
Features matrix corresponding to $\mathbf{f}^*_{ij}$ $(D \times R)$	$\mathbf{A}^*$
The mth eigenvalue of the covariance matrix of $\mathbf{A}$	λ_m
Normalized value of λ_m with respect to the sum of all eigenvalues	$\tilde{\lambda}_m$
Representation entropy over $\mathbf{A}$	$H_{\mathbf{A}}$
Features set corresponding to the mth column of $\mathbf{A}$	$\mathbf{A}_m$
Representation entropy over $\mathbf{A}\backslash\mathbf{A}_m$	$H_{\mathbf{A}\backslash\mathbf{A}_m}$
Difference between $H_{\mathbf{A}}$ and $H_{\mathbf{A}\backslash\mathbf{A}_m}$	dH_m
Top-ranked feature that maximizes dH_m	m^*
Representation entropy between two features m, m^*	$H_{\mathbf{A}_{m^*,m}}$
Representation entropy between feature m^* and its kth nearest neighbor	$h_{m*}(k)$
Cluster of features centered at m^*	$\mathbf{c}_{m^*}$
Upper threshold for $h_{m*}(k)$	ε
Evaluation metrics (Section 5)	
κ-Nearest neighbor cross validation accuracy	CV
Normalized value of representation entropy over $\mathbf{A}^*$	$\overline{H}$
Fuzzy feature evaluation index	FFEI
Mean square error	MSE
Compression ratio	CR

These metrics allow the capture of insights from the on-node behavior at different levels of the implemented stack. The question raised is whether the resulting vector $\mathbf{f}_{ij}$ conveys the dominating attributes that can characterize the network performance of the $i \rightarrow j$ link, expressed in terms of a class label l_{ij}. Characterizing a few operational and network performance-related attributes as "dominant" implies that they have sufficient information for inferring the value of class l_{ij}. These dominant variables can be later on used for constructing a predictive model for calculating the value of l_{ij}, while the remaining attributes can be eliminated since they contribute limited or redundant information. Thus, the objective becomes to exploit the contents of $\mathbf{f}_{ij}$ in order to automatically calculate the subset of R features $(R \leq M)$ $\mathbf{f}^*_{ij} \subseteq \mathbf{f}_{ij}$ that are most relevant to inferring the performance of each link $i \rightarrow j$. For network example depicted in Fig. 1, the problem is graphically presented in Fig. 2, while, the formal definition of the performance characterization of end-to-end links in multi-hop WSN topologies is given as follows:

Problem 1. Consider a WSN comprised of N sensor nodes. Each end-to-end link $i \rightarrow j$ between two nodes i and j is constructed over a multi-hop path P_{ij}. The performance of $i \rightarrow j$ is characterized by the value of PRR_{ij}, expressed in terms of a label l_{ij} (Eq. (1)) and its behavior along P_{ij} is described by the feature vector $\mathbf{f}_{ij}$ (Eq. (2)). Based on the contents of $\mathbf{f}_{ij}$ extract the subset of R features $(R \leq M)$ $\mathbf{f}^*_{ij} \subseteq \mathbf{f}_{ij}$, that are most relevant to inferring the network label l_{ij}, $\forall\ i \rightarrow j$.

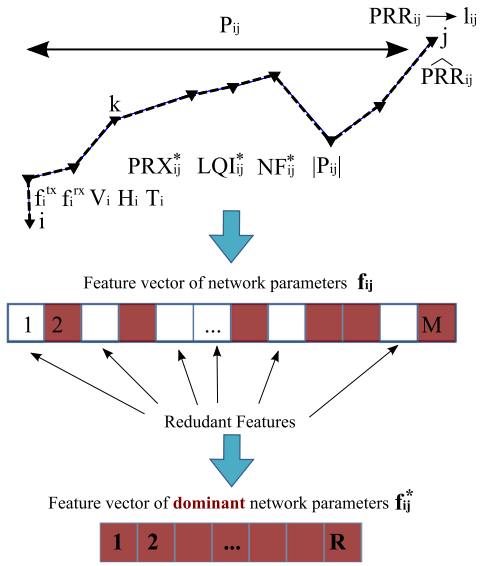

Fig. 2. The definition of the feature selection problem for end-to-end link $i \rightarrow j$ presented at Fig. 1.

4. The feature selection framework for multi-hop WSN topologies

Problem 1 is essentially a feature selection problem, with the objective of reducing the search space by including and excluding attributes present in the data, without changing them. Feature selection aims to provide a better understanding of the underlying process that generated the data, and it is considered an important step of filtering redundant information, prior to applying any classification technique in typical data mining applications. Shifting to the WSN paradigm, the importance of feature selection is amplified, due to the inherited compactness it introduces. Nevertheless, numerous technical challenges arise. First, the construction of the initial feature vector $\mathbf{f}_{ij}$, relying on the network parameters across path P_{ij} should neither affect the nominal network operation, nor increase the network traffic. Second, a-priori knowledge of the labels l_{ij} is not available and the existence of training periods is considered an unrealistic assumption, especially for unreachable and unattended WSN deployments. In addition, both the construction of $\mathbf{f}_{ij}$ as well as the calculation of $\mathbf{f}^*_{ij}$ should be lightweight, for efficiently handling both increased volumes of measurements, as well as co-existing multiple sources - single destination end-to-end links. Finally, achieving a highly compressible set of dominant features $\mathbf{f}^*_{ij}$ that can efficiently infer the network label l_{ij} is an important aspect for allowing energy-efficient and distributed implementations over computationally constrained sensor nodes.

With these considerations in mind, in this work we propose an integrated framework for enabling feature selection for performance characterization over multi-hop WSN topologies. Our approach considers three steps (cf. Fig. 3) namely: (a) the collection of network measurements, based on passive monitoring and protocol-independent mechanisms, (b) the construction of the initial feature vector $\mathbf{f}_{ij}$, and (c) the feature selection technique for the calculation of $\mathbf{f}^*_{ij}$, while taking into account both the lack of a-priori known labels, as well as the importance of compressing redundant information into the dominant features. In the following subsections, each of these steps is analytically described.

4.1. Collection of network measurements

We propose passive monitoring techniques for collecting and transferring network measurements along path P_{ij}, without generating additional network traffic. Specifically, the functionality of each node is enhanced by a network monitoring mechanism, which is located alongside the protocol stack as a horizontal plane, henceforth referred to as the Network Monitoring (NM) plane. The NM plane interacts with the vertical layers of the protocol stack, as shown in Fig. 4(a), for processing and updating network measurements. These network measurements are either triggered by the traffic generated at the Application layer between end-to-end links $i \rightarrow j$, or periodically monitored on the kth node. As such, at each node $k \in P_{ij}$ the NM plane collects the PHY layer parameters, which are related to both the RSSI and the LQI of the packets traveling along P_{ij}, as well as periodical samples of the ground noise floor when the kth node is idle. Similarly, the NM plane keeps track of the unicast traffic generated at the MAC layer and exploits this information for calculating the transmission and reception rate of each operational node. In addition, the NM plane periodically interacts with the Application layer of both the ith node for accessing the on-board temperature and humidity sensors, and for measuring the input voltage supply, as well as the jth node for calculating $\widehat{PRR}_{ij}$.

The design of the NM plane also considers the NM Relay Entity, which interacts with the Transport layer. The NM Relay entity is responsible for acquiring the information that the NM plane collects from the vertical layers, and generates the NM preamble, which is encapsulated in the data packets directed to the jth node of the P_{ij} path. The NM preamble, presented in Fig. 4(b), precedes the data payload and is comprised of three fields: (a) the PHY NM field, which contains information regarding the quality of the physical links; (b) the MAC NM field, which describes the MAC activity of each WBN node i in terms of transmission and reception rate; (c) the NWK NM field, which contains the trace of P_{ij}.

The NM Relay Entity additionally updates the contents of the NM preamble with the locally available NM information, by means of a piggy-packing process. As such, when $k \in P_{ij}$ receives a data packet corresponding to the $i \rightarrow j$ traffic, the NM Relay Entity adds its physical trace in the PHY NM field and adapts the NWK NM field of the NM preamble, in order to store its routing trace in the packet. An illustrative example is shown in Fig. 5, highlighting how the contents of the NM preamble are updated as the packet travels towards the jth node, over P_{ij}. It is worthwhile to note that the overhead that the NM preamble introduces to the length of the transmitted data packets is optimized by using encoding techniques (e.g., XOR, byte shifting operations) over the PHY, MAC, and NWK NM fields. In order to address scalability issues, when P_{ij} is excessively long, only the relay nodes that are up to l hops away from the origin node i will update the NM preamble. This essentially allows the jth node to have access to the network metrics and characteristics of the farthest relay links in P_{ij}.

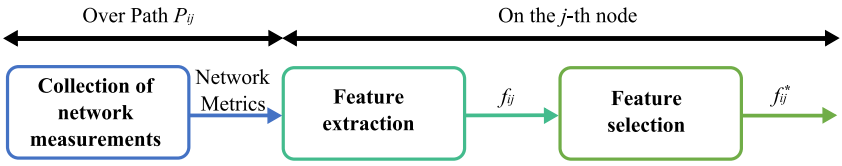

Fig. 3. The three-step procedure for enabling feature selection over multi-hop WSN topologies.

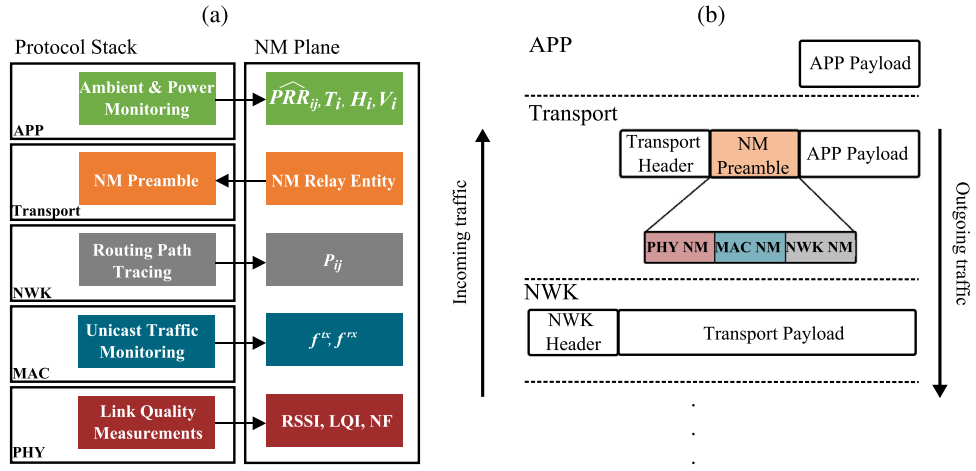

Fig. 4. (a) The NM plane and its interaction with the WSN protocol stack; (b) the NM preamble with respect to the data packet.

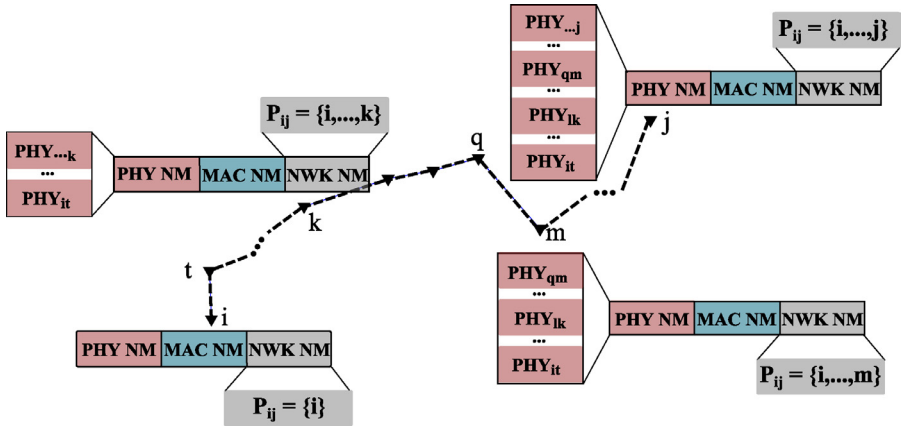

Fig. 5. A data-flow example over path P_{ij}, highlighting the piggy-packing process for updating the contents of the NM preamble.

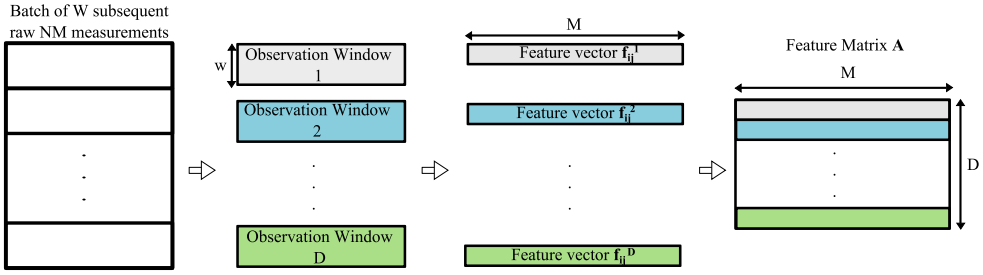

Fig. 6. The feature extraction procedure taking place at the jth node.

4.2. Feature extraction

The network monitoring plane allows each destination node to gain insights on both the status of the origin node as well as the performance of the intermediate links that are non-adjacent to it. As such, upon reception of each data packet over the $i \rightarrow j$ link, the jth node can consider the information available at the NM preamble for constructing the feature vector $\mathbf{f}_{ij}$.

Fig. 6 presents the herein proposed methodology. We consider as input W subsequent NM measurements corresponding to the $i \rightarrow j$ link. This batch of measurements is further split into smaller windows of observation with fixed length w. The NM entries in each observation window are exploited to calculate the contents of $\mathbf{f}_{ij}$. For all metrics associated to the performance over P_{ij} and the on-board conditions of the sensor nodes, the mean value $\mu(.)$ and standard deviation $\sigma(.)$ are employed for extracting the respective features. This is done for two reasons, namely: (a) such statistics can be readily extracted, even on computationally constrained sensor nodes, and (b) especially the standard deviation indicates how changeable the respective metrics are within a specific time window. The features related to f_i^{tx}, f_i^{rx}, and $\widehat{PRR}_{ij}$, are extracted by calculating the respective values for the specific observation window.

Hence, the feature vector defined in Eq. (2) becomes:

$$\begin{aligned}\mathbf{f}_{ij} = \big[&\mu(PRX_{ij}^{*}), \sigma(PRX_{ij}^{*}), \mu(LQI_{ij}^{*}), \sigma(LQI_{ij}^{*}), \mu(NF_{ij}^{*}),\\ &\sigma(NF_{ij}^{*}), f_i^{tx}, f_i^{rx}, \mu(|P_{ij}|), \sigma(|P_{ij}|),\\ &\widehat{PRR}_{ij}, \mu(T_i), \sigma(T_i), \mu(H_i), \sigma(H_i), \mu(V_i), \sigma(V_i)\big].\end{aligned} \tag{3}$$

All values in $\mathbf{f}_{ij}$ are normalized in [0, 1], using the z-score.

Applying the same procedure on all observation windows yields $D = \lceil W \setminus w \rceil$ subsequent instances $\mathbf{f}_{ij}^{d} \in [0, 1]^{1 \times M}$ of the feature

vector, where $d = 1,2,\ldots, D$. These instances formulate the features data set $\mathbf{A} \triangleq [\mathbf{f}_{ij}^1; \mathbf{f}_{ij}^2; \ldots; \mathbf{f}_{ij}^D]$, $\mathbf{A} \in [0, 1]^{D \times M}$, which is employed for calculating the dominating features for each $i \rightarrow j$ link.

4.3. Feature selection

Both the collection of the network measurements as well as the construction of the initial feature vector over P_{ij} are considered preparatory steps for addressing Problem 1, by the means of a feature selection algorithm. While the structural design of the herein presented integrated framework allows the adoption of different feature selection techniques, in this work we propose an approach that emphasizes on unsupervised learning. Driven by the aforementioned necessity to provide a lightweight, yet robust, solution on the feature selection process, our approach combines ranking and clustering techniques. This allows us to initially characterize the features based on their salience, i.e. their ability to stand out from the remaining features, and subsequently group them into clusters, based on their inter-redundancy.

Our algorithm relies on the calculation of the Representation Entropy $H_{\mathbf{A}}$ [32], which is a typical evaluation metric for measuring the amount of redundancy in the feature matrix $\mathbf{A}$. The calculation of $H_{\mathbf{A}}$ relies on the eigenvalues λ_m of the covariance matrix $M \times M$ of $\mathbf{A}$, according to $H_{\mathbf{A}} = -\sum_{m=1}^{M} \tilde{\lambda}_m \log \tilde{\lambda}_m$. Note that, $\tilde{\lambda}_m$ is the normalized value of the mth eigenvalue λ_m, with respect to the sum of all eigenvalues.

Based on the above definition, $H_{\mathbf{A}} \in \mathbb{R}^+$. It becomes minimum when all eigenvalues expect one are 0. In this case, all information is concentrated on a single, principal direction, or feature. The value of $H_{\mathbf{A}}$ becomes maximum ($\log M$) when all eigenvalues are equal ($\tilde{\lambda}_m = 1/M,\ \forall m = 1, 2, \ldots, M$), implying that the information is equally distributed among all directions.

Representation entropy expresses the amount of information compression available in a dataset; when $H_{\mathbf{A}} \rightarrow 0$, the information of the data set $\mathbf{A}$ can be compressed into the information that corresponds to the dominating feature only. By contrast, when $H_{\mathbf{A}} \rightarrow \log M$ all features are equally important and, thus, the redundancy of $\mathbf{A}$ is low. As such, $H_{\mathbf{A}}$ can characterize the ability of a set of features to efficiently reflect the classes of $\mathbf{A}$, while removing redundant information. This is considered essential for synthesizing feature selection algorithms for multi-hop WSN deployments, capable of the unsupervised and lightweight calculation of highly compressed dominant feature vectors $\mathbf{f}_{ij}^*$. Thus, we exploit $H_{\mathbf{A}}$ as a search criterion for clustering and compressing redundant features.

The proposed technique relies on a backward search in the feature space and is divided in three main steps: (a) rank the features with respect to the volume of uncertainty they carry; (b) cluster features that exhibit high redundancy with a top-ranking feature, which is appointed as the head of the cluster; (c) eliminate all members of the cluster from the feature space as redundant, expect from the cluster head.

During the first step, the procedure described in [33] is adopted, and the amount of redundancy that each feature contributes to the data set $\mathbf{A}$ is calculated as follows:

$$dH_m = H_{\mathbf{A}\setminus\mathbf{A}_m} - H_{\mathbf{A}}, \tag{4}$$

where $\mathbf{A}_m$ is the set of samples corresponding to the mth feature and $H_{\mathbf{A}\setminus\mathbf{A}_m}$ is the representation entropy of $\mathbf{A}$ when the $\mathbf{A}_m$ set of samples are not taken into account. Essentially, the value of dH_m represents the difference in the uncertainty when the mth feature is omitted; if that feature corresponds to a principal component, then the value of $H_{\mathbf{A}\setminus\mathbf{A}_m}$ will become significantly smaller than the one of $H_{\mathbf{A}}$, and the value of dH_m will, in turn, become high. On the other hand, if the mth feature describes information with limited variance or predictable behavior, then the value of $H_{\mathbf{A}\setminus\mathbf{A}_m}$ will remain similar to the one of $H_{\mathbf{A}}$. Consequently, the value of $dH_m \rightarrow 0$. The outcome of this procedure is a ranking of the feature vector, according to which the top-ranked feature m^* is the one that maximizes the value of dH_m, $m = \{1, 2, \ldots, M\}$.

During the second step of the algorithm, the search space is centered around m^*. A k-nearest neighbor technique is employed in order to cluster the features that exhibit the higher value of redundancy with the feature m^*. The value of the pairwise representation entropy $H_{\mathbf{A}_{m^*,m}}$ between feature m^* and the remaining features m is calculated, where $m^* \neq m$, $m \in \mathbf{f}_{ij}$, and $\mathbf{A}_{m^*,m} = [\mathbf{A}_{m^*}, \mathbf{A}_m] \in [0, 1]^{D\times 2}$. The resulting values of $H_{\mathbf{A}_{m^*,m}}$ are sorted in descending order. The first k features, along with m^*, form a cluster of features $\mathbf{c}_{m^*}$:

$$\mathbf{c}_{m^*} = m^* \bigcup \{m | H_{\mathbf{A}_{m^*,m}} \leq h_{m^*}(k)\}, \tag{5}$$

where $h_{m^*}(k)$ is the value of the pairwise entropy between the feature m^* and its kth nearest neighbor.

The cluster $\mathbf{c}_{m^*}$ is comprised of features that exhibit high redundancy with m^*. The latter is considered the dominating feature and, therefore appointed as the cluster head. Subsequently, during the third step of the algorithm, the cluster head m^* remains in the search space as the representative feature of cluster $\mathbf{c}_{m^*}$, while all remaining features in $\mathbf{c}_{m^*}$ are considered redundant and thus eliminated [32].

Algorithm 1 The Representation Entropy Clustering Feature Selection Algorithm (REC-FSA)

Require: (a) The initial feature vector $\mathbf{f}_{ij}$ containing M features; (b) the corresponding data set $\mathbf{A} \in [0, 1]^{D\times M}$; (c) the initial values for the $k \in [2, M-1]$ parameter and the upper entropy threshold ε.

Ensure: The reduced feature vector $\mathbf{f}_{ij}^* \subseteq \mathbf{f}_{ij}$, containing R features ($R \leq M$).

1: Initialize the reduced feature subset to the value of the original set, i.e. $\mathbf{f}_{ij}^* \leftarrow \mathbf{f}_{ij}$.
2: **if** $k \leq 1$ **then**
3: Goto Step 17.
4: **else**
5: For each mth feature $\in \mathbf{f}_{ij}^*$ calculate the value of dH_m according to Eq. (4).
6: **end if**
7: Select the top-ranked feature m^* s.t. $dH_{m^*} = \max_m dH_m$.
8: Calculate the value of the pairwise entropy $H_{\mathbf{A}_{m^*,m}}$, $\forall m \neq m^*$, $m \in \mathbf{f}_{ij}$.
9: Extract the value $h_{m^*}(k)$ of the pairwise representation entropy that corresponds to the kth nearest neighbor and create the cluster $\mathbf{c}_{m^*}$ according to Eq. (5).
10: **if** $\varepsilon \geq h_{m^*}(k)$ **then**
11: Remove all features $m \in \mathbf{c}_{m^*}$, $m \neq m^*$ from the feature vector $\mathbf{f}_{ij}^*$ and update accordingly the data set $\mathbf{A}$.
12: Update the value of the upper threshold $\varepsilon \leftarrow h_{m^*}(k)$.
13: Update the value of k: $k \leftarrow k - 1$ & go to Step 2.
14: **else**
15: $k \leftarrow k - 1$ & go to Step 2.
16: **end if**
17: Return the reduced feature vector $\mathbf{f}_{ij}^*$ and stop.

The process is repeated until either all features are clustered and discarded, or selected as dominating. The resulting algorithm, henceforth called Representation Entropy Clustering Feature Selection Algorithm (REC-FSA), is presented in Algorithm 1. The termination of the search procedure is dictated by two constants: (a) the value of the user-defined parameter k, (b) the upper threshold ε, which defines the acceptable level of redundancy between

the kth nearest neighbor and feature m^*. Specifically, while the selection of the initial value of k controls the size of the reduced set, the dynamic reduction during the execution of the algorithm allows instantaneous tuning of the search accuracy in the feature space. Consequently, the decrease of the dimension of the feature space is accompanied by a respective adjustment of the value of k, therefore accordingly adjusting the degree of detail in the search space [32]. Moreover, the introduction of ε as a termination criterion can optimize the execution time; the search procedure will be completed when the redundancy between the kth nearest neighbor and the feature m^* becomes sufficiently small.

The computational complexity of REC-FSA depends on the dimensions $D \times M$ of the data set $\mathbf{A}$ and is bounded by $O(DM^3)$, which corresponds to the calculation of dH_m. Specifically, the calculation of $H_{\mathbf{A}}$, which is based on pairwise calculations between different features is bounded by $O(DM^2)$, while the calculation of dH_m is bounded by $O(DM^3)$. In addition, the pairwise comparisons for the formation of clusters around the kth nearest neighborhood of m^* is bounded by $O(DM)$. Consequently, the overall computational complexity of the REC-FSA is $O(DM^3)$. Taking into account that for the proposed framework the value of M is relatively small (Eq. (3)), the computational complexity of the REC-FSA depends essentially on the number of samples D.

REC-FSA is expected to provide clusters with high redundancy, or equivalently low value of representation entropy. Similarly, the resulting feature vector $\mathbf{f}_{ij}^*$ is expected to have low redundancy, and thus high value of representation entropy. As presented in Section 5, this behavior yields a significant advantage of REC-FSA against competitive methods in terms of enabling the performance characterization of multi-hop WSN topologies by the means of feature selection.

5. Evaluation studies

The evaluation of the proposed framework has been made on small-scale real-life WSNs deployed both on a rural, and an industrial environment. In the first case, the WSN comprised $N=9$ nodes, and was deployed at an olive trees grid (60×15 s.m.) in user-designated locations, as part of a platform monitoring the surrounding micro-climate conditions, mainly focusing on temperature, humidity, and total solar radiation. In the second case, the WSN, comprised $N=10$ nodes, was part of an industrial smart water network deployed at a fully functional pilot desalination plant (40×12 s.m.) to monitor and control the phenomenon of biofouling process, which is related to the accumulation of unwanted bacterial matter on the surface of the reverse osmosis membranes [34]. The sensor nodes were deployed at user-designated locations, namely, the sea water intake, the pre-treatment, the security filters, and the reverse osmosis. As such, except for the metallic environment, additional RF challenges related to bulky water tanks, heavy machinery in operation, pumps and valves, and the presence of technical staff, were introduced. In both cases, the sensor nodes operate using conventional, non-rechargeable AA batteries, and thereby are characterized by limited lifetime while, the objective of each sensor node was to disseminate its data towards a plugged-in sink node. Thus, all $i \rightarrow j$ links established considered the same destination, i.e. $j=$S and $i=\{\text{A,B},\ldots,\text{I}\}$ ($i=\{\text{A,B},\ldots,\text{I,J}\}$) for the rural (industrial) deployment. The layouts for both the rural and industrial deployments are presented in Fig. 7(a) and (b) respectively. In addition, Fig. 8 illustrates snapshots of the operational environments and the sensor nodes therein deployed.

For each case of operational space, two different version of 802.15.4-based [35] WSN protocol stacks, namely Stack 1 and Stack 2, have been employed. Both stacks have been implemented in Contiki OS, a very popular real-time operating system for WSN [36], and deployed on popular embedded platforms [37],

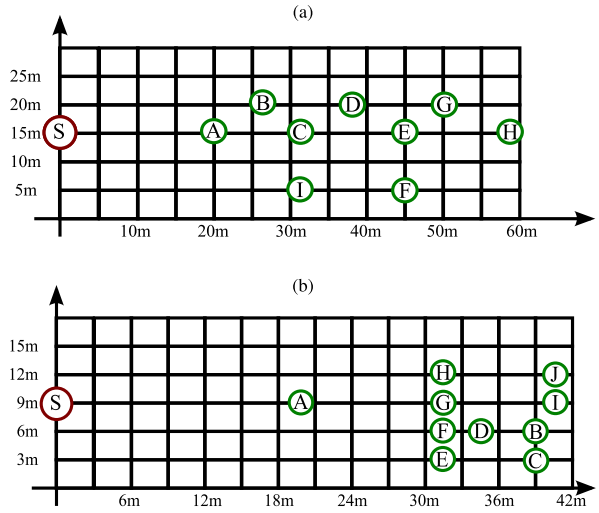

Fig. 7. The layout of the WSN deployments at (a) the rural (a), and (b) the industrial environments.

while the functionalities of the sink node have been deployed on an ARM-based single-board computer [38]. The protocol stacks do not consider a dedicated sleep mode for the sensor nodes, and feature state-of-the-art techniques in MAC and Routing protocol design for WSN [39,40]; Stack 1 employs a customized CSMA-based MAC protocol, which incorporates a physical interference model for mitigating hidden / exposed terminal problems [39]. At the routing layer, the IETF standard for Low Power and Lossy Networks (RPL) [41] is employed. Stack 2 builds upon Stack 1, by adopting a customized version of RPL [40], according to which the routing criterion is based on the probability that a packet is correctly received in each link of the route, considering the duration of the backoff period and retransmissions at the MAC layer.

It is considered important to note that both the selection of the specific protocol stacks, as well as their positioning within the operational space, is application-driven and dictated by the specific, user-level requirements whose justification is beyond the scope of this work. Nevertheless, the NM plane (Section 4.1) has been also implemented on Contiki-OS and deployed on each operational node alongside Stack 1 and Stack 2. The overhead introduced to each data packet for generating and updating the NM preamble is limited to 34 Bytes. Finally, both the construction of the initial feature vector $\mathbf{f}_{ij}$, as well as the calculation of $\mathbf{f}_{ij}^*$ is implemented at the side of the sink node j, using the Qt Framework [42], which is a C++ based cross-platform software environment, suitable for developing lightweight and time-efficient applications that can be run on various platforms, including ARM-based architectures, similar to the one used for deploying the functionalities of the sink node.

In all deployments considered, the WSN operated in a continuous unattended manner. Due to different application demands, the sampling rate and the duration of the data collection varies with respect to both the type of the environment, as well as the implemented protocol stack. More specifically, the WSN was configured to generate end-to-end traffic per 1 minute and per 0.1 minutes at the rural and industrial environment, respectively. In addition, Stack 1 (2) has been deployed for 24 (42) h at the rural environment and for 2.5 (162) hours at the industrial environment.

Fig. 9 presents the estimated Cumulative Density Function (ECDF) of PRR_{ij} for both Stack 1 and Stack 2 deployed at each case of environment, indicating that Stack 1 has the least satisfactory performance, as the probability of $PRR_{ij} \leq 0.8$ equals to 0.39, as opposed to the case of Stack 2 for which the probability of having PRR_{ij} smaller than 1 equals to 0.1781. This is due to

Fig. 8. The experimental deployments at the rural (top) and industrial environments (bottom), and the hardware of node S (bottom right).

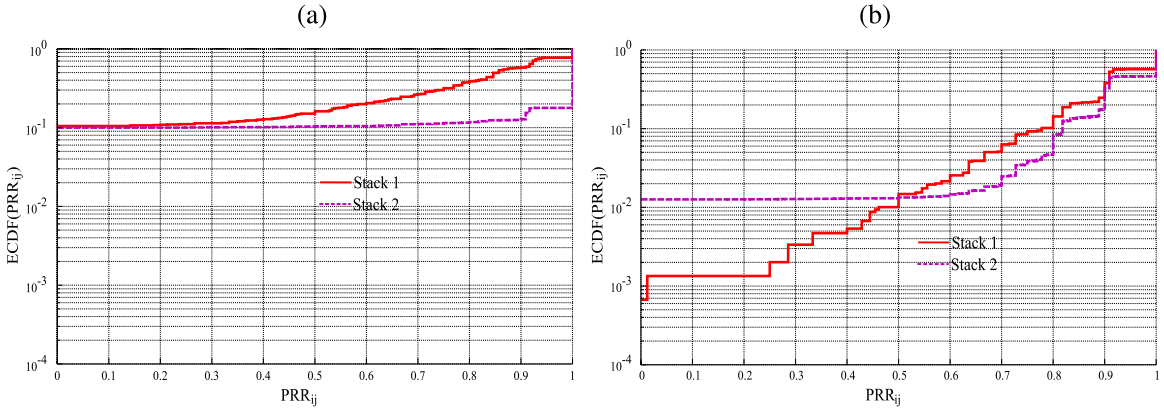

Fig. 9. The ECDF of PRR_{ij} for the deployments at (a) the rural and (b) the industrial environments.

the fact that for the case of Stack 1, the noise threshold is too optimistic and equal to -50dBm. As a result, the hidden terminal problem remains a key factor, because nodes are not taking it into account. The operational nodes will transmit, even when a concurrent transmission is present. The introduction of MAC-aware objective functions for forming the routing tree alleviates this issue, since it also considers the probability that a packet is successfully transmitted over each link of a path, within a maximum number of backoffs and retransmissions at the MAC layer. Based on the results presented in Fig. 9 and the actual PRR_{ij} values recorded during the operation of the deployed WSNs, we consider that $\alpha_E = 0.95$, $\alpha_G = 0.8$, and $\alpha_P = 0$, essentially corresponding to labels $l_{ij} = \{\text{Excellent, Good, Problematic}\}$ (Eq. (1)).

The constant sampling of the voltage supply V_i on each node i by the herein proposed framework enables the empirical calculation of the level of discharge, according to the curve fitting procedure described at Appendix A, and subsequently, the estimation of the power consumption P_i^{cons}, and lifetime τ_i. Fig. 10(a) and (b) present the kernel-based PDF estimation [43] of P_i^{cons} versus f_i^{tx} for the rural and industrial deployments of Stack 2 respectively. Similarly, Fig. 10(c) and (d) yield the kernel-based PDF estimation of τ_i with respect to measured V_i.

With regard to the power consumption, the primary observation to make is that in both cases of operational environment the power consumption has a small dependency on f_i^{tx}. Specifically, for the case of rural field studies (Fig. 10(a)) the power consumption varies within the limited range [47,52] mW, regardless of the recorded f_i^{tx} rate, which is located within the range of [17, 150] bps. The decreased network traffic is expected, as a combination of both the low traffic rate (0.0167Hz), and the increased transmission range of the operational nodes. In addition, the always-on operational mode of the transceiver of each node, alighted to the increased level of current drawn during the channel listening, yields relatively high levels of power consumption even during periods of network inactivity. Similar observations can be derived for the industrial deployments, shown at Fig. 10(b), indicating the impact of the always-on functionality of the sensor nodes, as the key source of consumption. Nevertheless, in this case we can identify a correlation on the trend of the power consumption as the network traffic f_i^{tx} increases, highlighting the limited connectivity options at the industrial environment, which in turn magnify the role of the relay nodes for enabling end-to-end connectivity between distant ith nodes and the jth node. With regard to the estimation of the nodes' lifetime we can observe a logarithmic correlation of τ_i and the operational voltage supply V_i. Considering the rural (industrial) deployment the range of the input voltage supply varies between 2.52 (2.52) and 2.72 (2.73), corresponding to τ_i which varies within the range of [0.38 ,6.445] ([0.35, 6.55]) days. In conjunc-

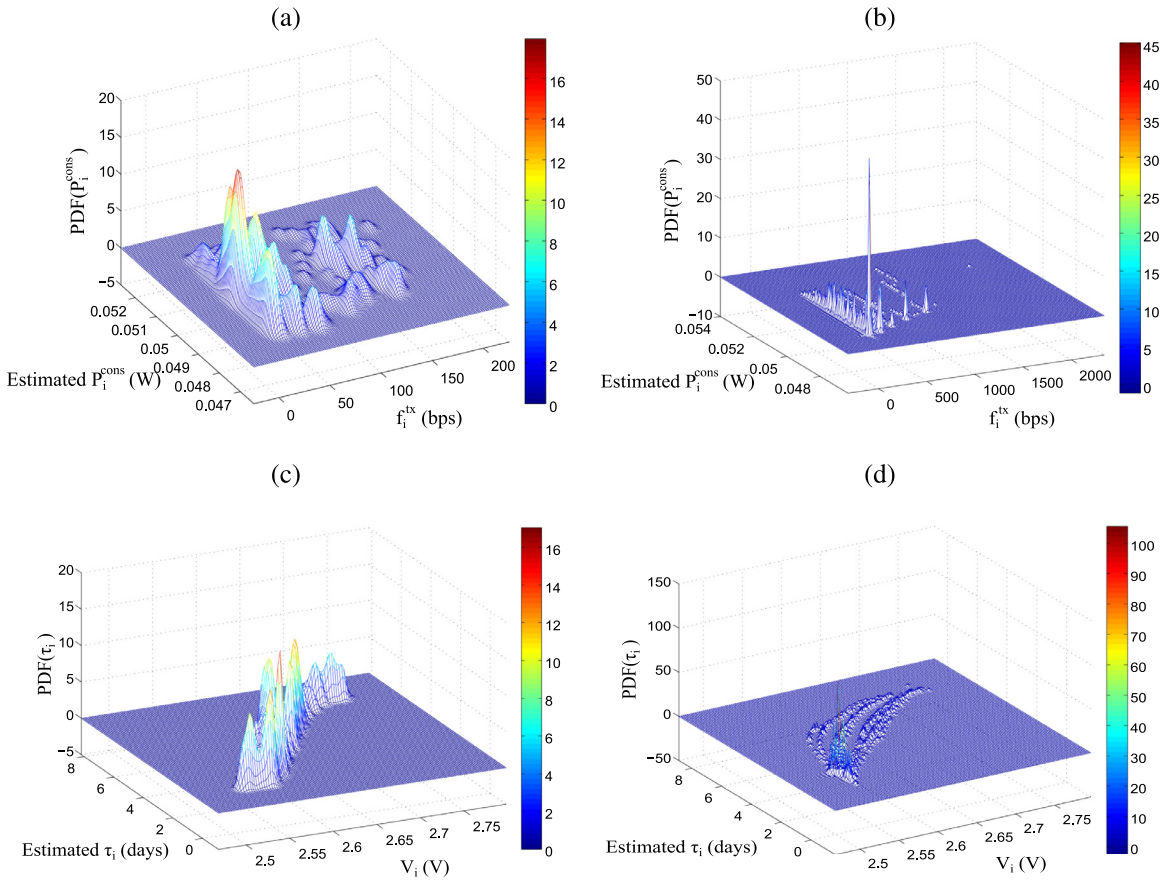

Fig. 10. The estimated kernel-based PDF of: (a) P_i^{cons} as a function of f_i^{tx} at the rural environment, (b) P_i^{cons} as a function of f_i^{tx} at the industrial environment, (c) τ_i as a function of V_i at the rural environment, and (d) τ_i as a function of V_i at the industrial environment.

tion to fact that the power consumption remains relatively stable throughout the experiments duration, with mild degradation associated to the transmission / relaying activity of the nodes, it becomes obvious that the lifetime of the network is highly depended on the discharging behavior of the battery cells. As highlighted in Section 5.1.2, the discharging behavior of the input power supply becomes a dominant factor for the performance characterization of the multi-hop WSN deployments at the industrial environment.

Despite the small scale of the WSN deployments considered, in total, 684,952 instances of raw NM traffic arrived at the sink node and employed for the construction of $\mathbf{f}_{ij}$ and the calculation of $\mathbf{f}_{ij}^{*}$. From these traces 645,032 measurements correspond to the industrial deployments and 39,920 correspond to the rural deployments. The value of w is set to 500, corresponding to observation windows of medium size. Without loss of generality, we consider that the length W of the batch of raw NM traffic for extracting $\mathbf{f}_{ij}$ equals to the total number of NM traces per $i \rightarrow j$ link available at the jth node. This is equivalent to an off-line evaluation of the proposed feature selection algorithm, focusing both on the efficacy of REC-FSA, as well as the dominating network features for each case of WSN deployment. An indicative subset of network features, which correspond to approximately 18 hours of WSN operation at the industrial environment when Stack 2 is deployed at the sensor nodes is available at [44], for testing and evaluation purposes.

5.1. Experimental results

The efficacy of REC-FSA has been evaluated against two benchmark feature selection algorithms, namely: (a) the supervised sequential forward floating selection (SFFS) [9] and (b) the feature selection based on similarity algorithm (FSSA) [32]. The selection of these two algorithms as benchmarks for evaluating the herein proposed method is based on their popularity within the machine learning community. In a nutshell, SFFS adopts a sequential forward search in the feature space with backtracing support, relying on the correlation of each feature both with the label of the class, as well as the remaining features. By contrast, FSSA is a unsupervised, clustering mechanism, and groups features based on their similarity, defined as the smallest eigenvalue of the covariance matrix of a pair of features, while it converges in quadratic time ($O(M^2)$). In fact, as explained in Section 4.3 REC-FSA adopts the philosophy of FSSA on retaining the head of the cluster $\mathbf{c}_{m^*}$ as the representative of the entire cluster and eliminating the remaining members of the $\mathbf{c}_{m^*}$ as redundant. However, in contrast to FSSA, REC-FSA proposes the entropy as the main criterion for grouping the features into clusters and calculating their pairwise similarities.

The metrics employed for evaluating the performance of the feature selection algorithms are: (a) the κ-NN cross validation accuracy (CV) [9], where the value of κ depends on D, (b) the fuzzy feature evaluation index (FFEI), and (c) the normalized value of the representation entropy for a given dataset $\overline{H}$. CV $\in [0, 1]$ is a supervised quality measure and quantifies the ability of $\mathbf{f}_{ij}^{*}$ to train a subset of the data set D and classify the remaining set using the κ-NN rule. When CV $\rightarrow$ 1 the optimal predictability is achieved. We herein consider that 20% of the data set, randomly selected, is used for training and 80% for classifying, and $\kappa = \lceil\sqrt{0.2D}\rceil$. 20 independent runs are performed and the mean value of CV is selected as the final one. The FFEI [45] associates different patterns from the data set with a distance function, which describes whether or not

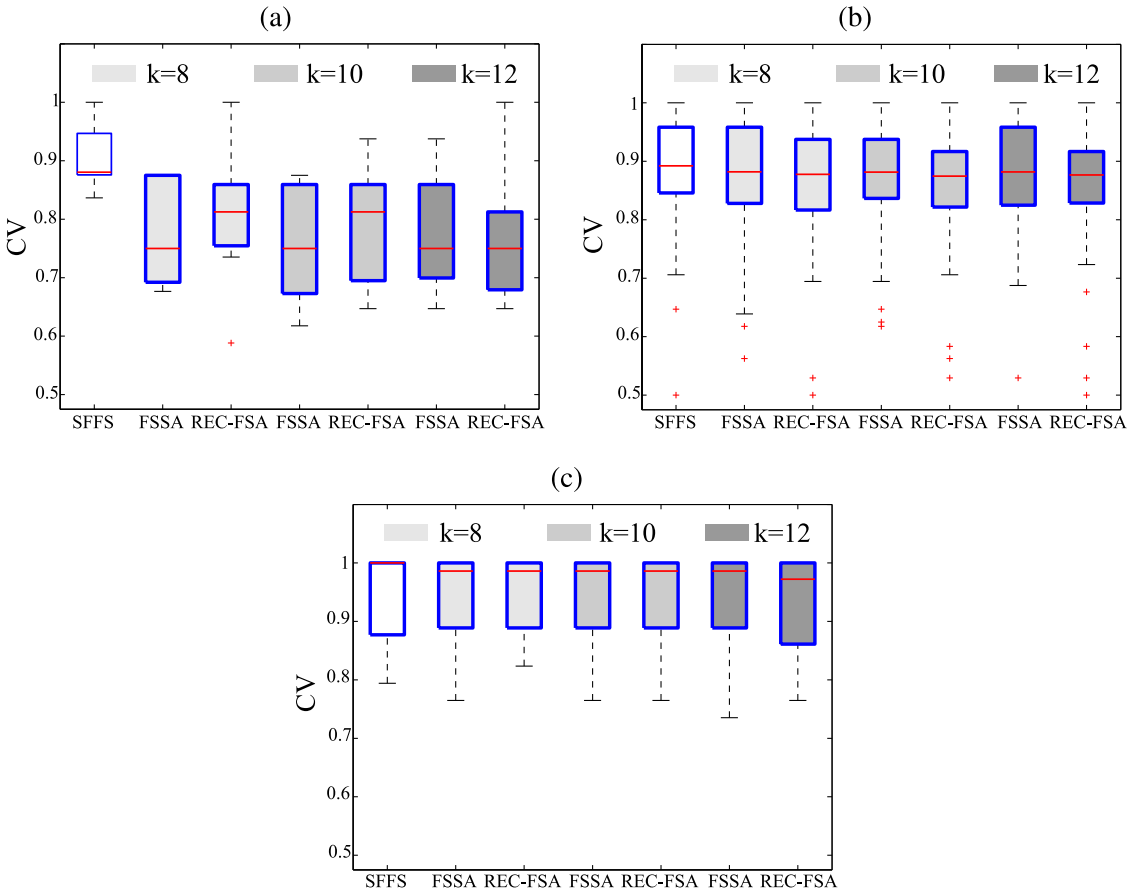

Fig. 11. The CV for all cases of WSN deployments when the link performance is: (a) problematic, (b) good), and (c) excellent.

they belong to the same cluster. FFEI takes values in [0, 1], and decreases as the distance between different clusters increases and the intra-cluster distance decreases. Thus, as the value of FFEI decreases, the structure of the clusters becomes sharper. In addition, $\overline{H}$ normalizes the value of $H_{\mathbf{A}}$ over its respective maximum value $\log M$. It is worthwhile to note that the CV metric quantifies the ability of $\mathbf{f}_{ij}^{*}$ for classification, while FFEI evaluates the quality of the formed clusters. Finally, is association to Section 4.3, $\overline{H}$ describes the uncertainty of the feature vector, or equivalently the amount of information compression that can be achieved with $\mathbf{f}_{ij}^{*}$.

5.1.1. Performance of feature selection algorithms

In the following paragraphs, the performance of the feature selection techniques will be discussed with respect to the accuracy of the prediction of the labels, the quality of clustering, in terms of compression and the redundancy that characterizes the dominant set of features.

Cross validation accuracy. The performance of REC-FSA against SFFS and FSSA in terms of cross validation accuracy for all cases of WSN deployments is presented in Fig. 11, for problematic links (Fig. 11(a)), good links (Fig. 11(b)), and excellent links (Fig. 11(c)). Especially for the FSSA and REC-FSA, the CV value for different cases of k (= 8,10,12) is also presented. As expected, the supervised algorithm (SFFS) exhibits better behavior than the unsupervised mechanisms, namely FSSA and REC-FSA. The median value of *CV* when SFFS is applied equals to 0.8805, 0.8922, and 1 for problematic, good, and excellent links respectively. This is achieved at the expense of assuming a-priori knowledge of the labels l_{ij}. Considering the unsupervised algorithms, REC-FSA exhibits better label predictability than FSSA, due to the ranking of the features with respect to the uncertainty they introduce into the data set. This is more evident for the problematic links; when FSSA is employed, the median value of *CV* accuracy equals to 0.75, as opposed to 0.8125 which is the respective value when REC-FSA is applied and $k = \{8, 10\}$. The difference observed in the CV performance of REC-FSA with respect to the increasing value of k for problematic links, is related to the search nature of the algorithm. The increase of k implies widening the search space for finding redundant information between a top-ranking feature m^{*} and the remaining features. As such, the size of the cluster increases and more features are omitted from the $\mathbf{f}_{ij}^{*}$, thereby affecting the ability of the algorithm to accurately estimate the labels of the links. The improvement on the network performance, which essentially implies an increase on the volume of NM measurements available for constructing $\mathbf{f}_{ij}$ and $\mathbf{A}$, is accompanied by an increase on the *CV* scores, for SFFS, FSSA, and REC-FSA.

Quality of clustering. Likewise, the performance of the algorithms in terms of the quality of the clusters formulated, using the FFEI score, is presented at Fig. 12, for problematic links (Fig. 12(a)), good links (Fig. 12(b)), and excellent links (Fig. 12(c)). We observe that, SFFS has the highest FFEI, or equivalently the worst clustering performance, since the elimination of features from $\mathbf{f}_{ij}^{*}$ does not consider clustering criteria. Instead, the dimensionality reduction depends on the correlating nature between individual features and the a-prior known labels l_{ij}. In all cases of link performance categorization, REC-FSA constructs clusters of better quality than those that FSSA constructs; the median value of FFEI when REC-FSA is employed remains constantly lower than 0.15, as opposed to the one when FSSA is employed, which varies between 0.07 ($PRR_{ij} \leq$

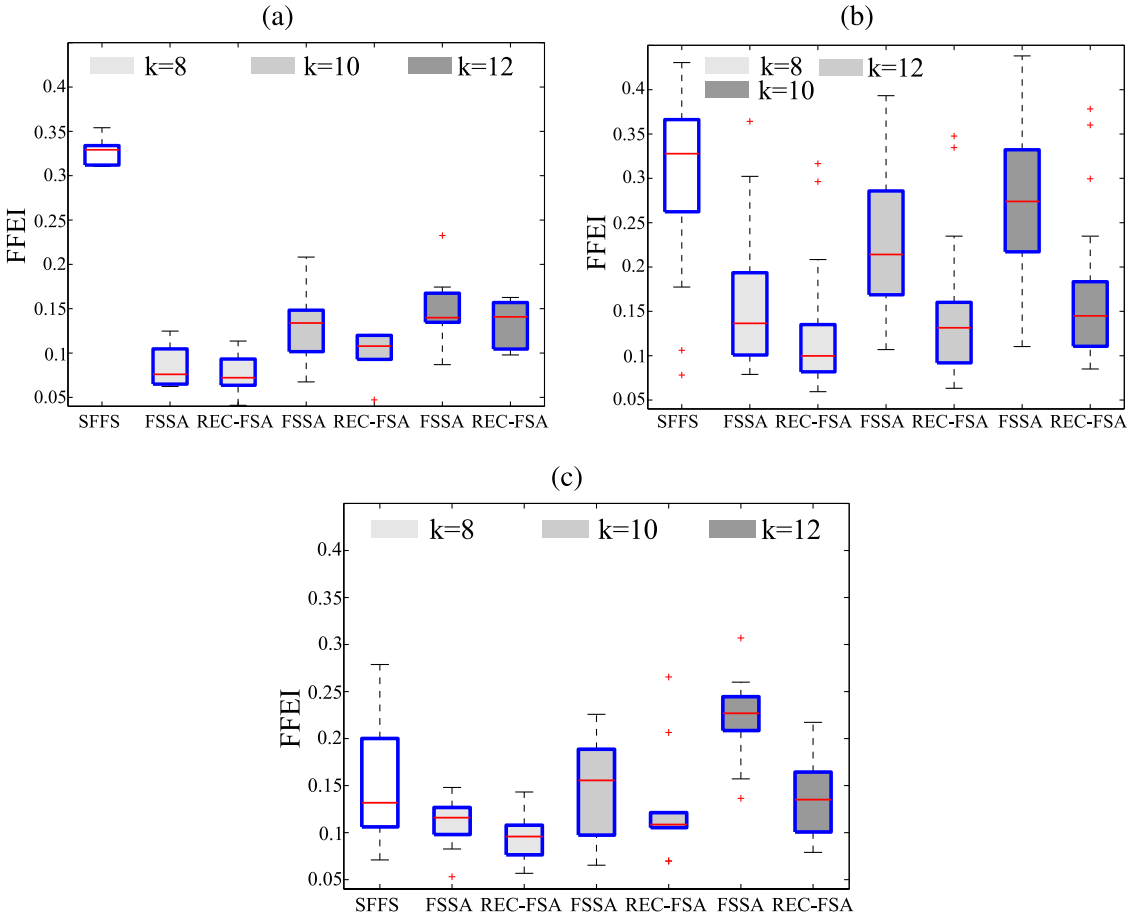

Fig. 12. The FFEI for all cases of WSN deployments when the link performance is: (a) problematic, (b) good), and (c) excellent.

0.8, $k = 8$) and 0.27 ($0.8 < PRR_{ij} \leq 0.95$, $k = 12$). This is due to the initial ranking phase of REC-FSA, which allows the construction of clusters around the one with the highest contribution to $H_{\mathbf{A}}$. With regard to the relationship between FFEI and the value of parameter k, we observe that when REC-FSA is employed, the increase in the value of k is accompanied by a small increase in the value of FFEI. Equivalently, when the value of k increases, the quality of the clusters slightly deteriorates. As explained in the previous paragraph, this is due to the increase of the search space, for constructing the clusters, which implies that less redundant features are clustered together, thereby affecting the crispness of the clusters. Nevertheless, for the same value of k, the performance of REC-FSA in terms of FFEI remains stable for all three types of network labels considered, as opposed to the one observed when FSSA is employed. This highlights the fact that, regardless of the network performance, translated to the volume D of data samples available for the calculation of $\mathbf{f}_{ij}^{*}$, REC-FSA achieves better compression of redundant features than the one that FSSA provides.

Redundancy. Another important aspect is the quality of the reduced set in terms of the redundancy it represents. Fig. 13 depicts the values of $\overline{H}$ for $\mathbf{f}_{ij}^{*}$ when the SFFS, FSSA, and REC-FSA techniques are applied on data corresponding to problematic links (Fig. 13(a)), good links (Fig. 13(b)), and excellent links (Fig. 13(c)). The results highlight the fact that REC-FSA yields the best performance among all algorithms; across all cases of l_{ij} and k, the median value of $\overline{H}$ varies in [0.79, 0.91] when REC-FSA is employed, as opposed to [0.72, 0.76] and [0.66, 0.77], which correspond to the range of values for the median of $\overline{H}$ when SFFS and FSSA are respectively applied. This behavior is expected, since the philosophy of REC-FSA is built around the benefits that the representation entropy offers during both the ranking of the features and the clustering of redundant features. Nevertheless, it is considered interesting to note that the improvements on the link performance are accompanied by corresponding enhancement of the redundancy within $\mathbf{f}_{ij}^{*}$. This is due to the combination of the pairwise characterization of redundancy and the deteriorated predictability of the feature vector for low-ranked end-to-end $i \rightarrow j$ links. In addition, the value of k affects the resulting value of $\overline{H}$, especially when REC-FSA is employed. Specifically, for links characterized by $PRR_{ij} \leq 0.8$ the paired values of the lower and higher quantiles are (0.75, 0.86), (0.78, 0.86), (0.79, 0.87) for k equal to 8, 10, 12, respectively. As we shift to links that are classified as "Excellent" ($PRR_{ij} > 0.95$) this deviation increases; the lower and higher quantiles become (0.85, 0.88), (0.86, 0.93), (0.89, 0.95) for k equal to 8, 10, 12 respectively. This occurs because a higher value of k results to more iterations of execution, which in turn increases the level of detail in the search space. Therefore, the remaining features exhibit a higher degree of uncertainty than the one that would result from a smaller value of k.

To further examine the performance of unsupervised feature selection in terms of redundancy, in Fig. 14, we present the value of $\overline{H}$ for $\mathbf{f}_{ij}^{*}$ and $\mathbf{c}_{m^*}$ for the individual nodes deployed at the rural (Fig. 14(a)-(b)), and the industrial (Fig. 14(c)-(d)) environment. Notable variations are observed for each individual node, or, equivalently, end-to-end link, when both feature selection algorithms are applied. Such deviations are more intense for the case of FSSA; the median value of $\overline{H}$ for $\mathbf{f}_{ij}^{*}$ varies from 0.69 (node ID = E) to 0.88 (node ID = A) for the rural environment, and is positioned within 0.48 (node ID = F) and 0.843 (node ID = K) for the industrial de-

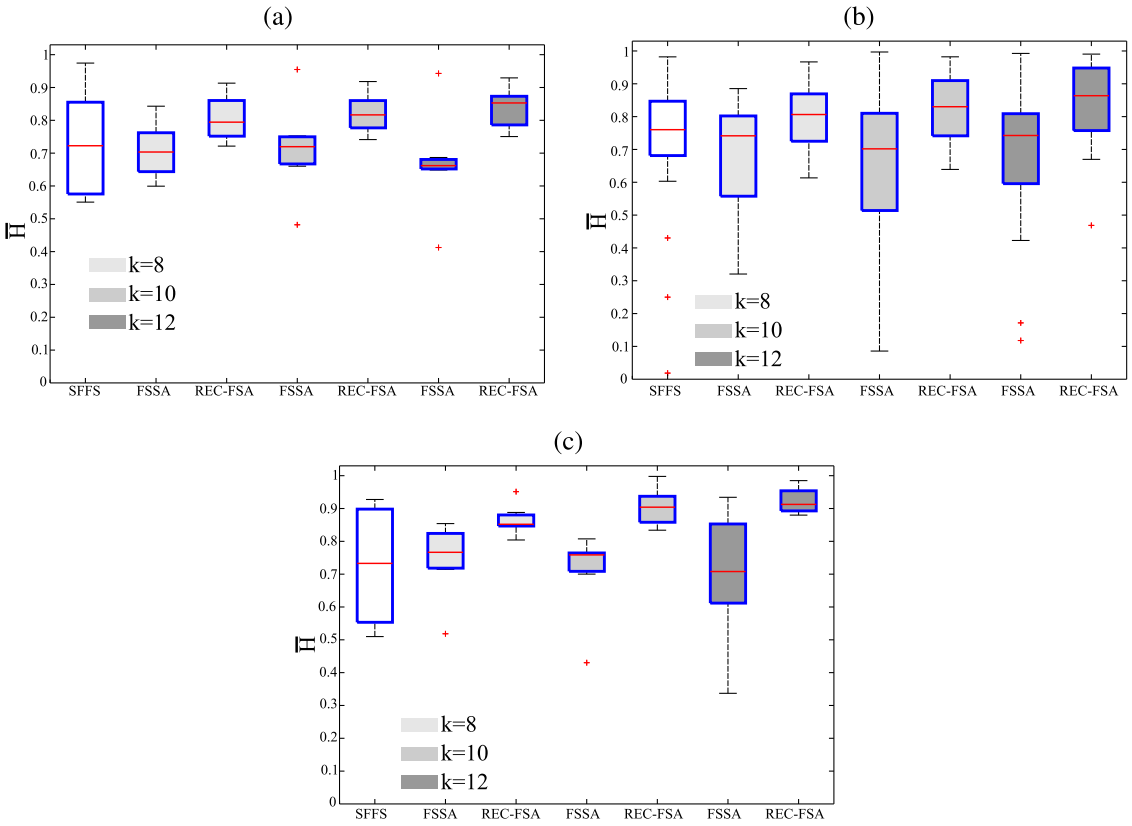

Fig. 13. The $\overline{H}$ value for the resulting reduced set $\mathbf{f}_{ij}^{*}$, for all cases of WSN deployments when the link performance: is (a) problematic, (b) good, and (c) excellent.

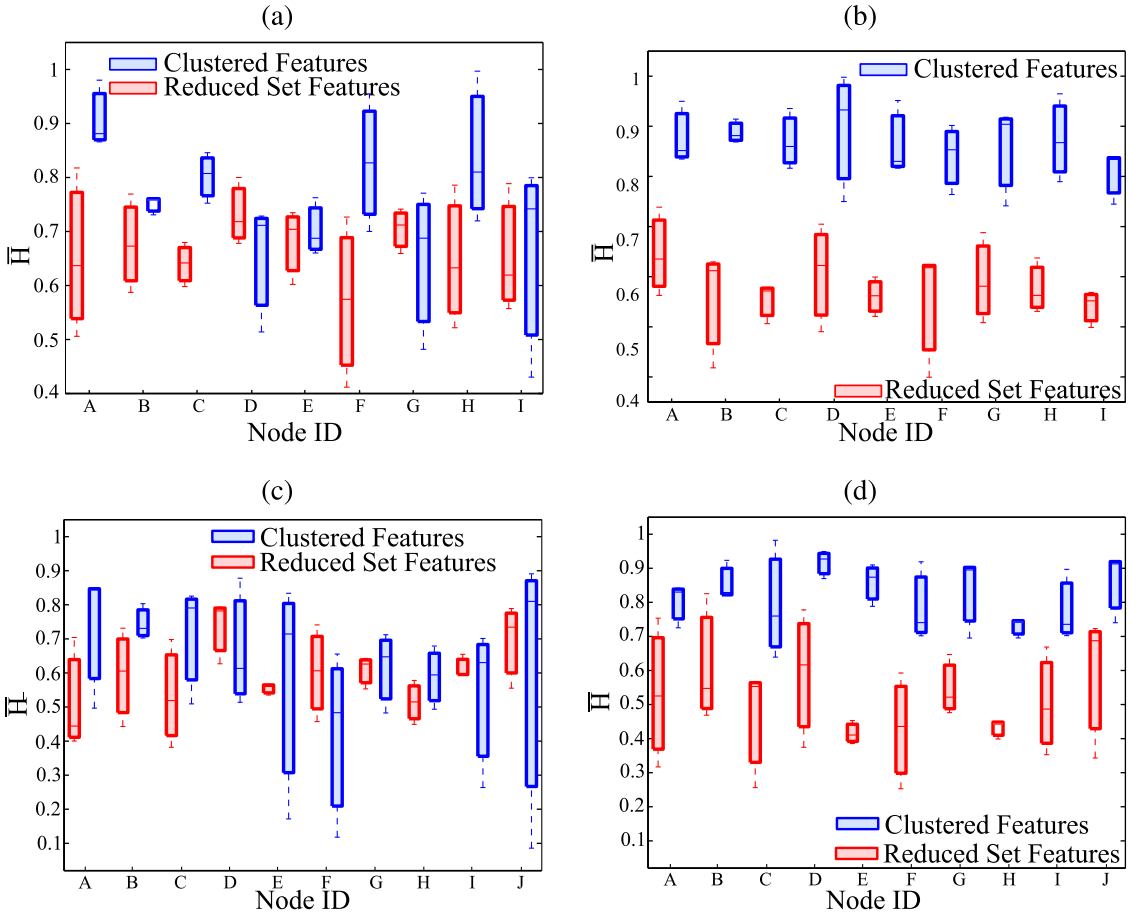

Fig. 14. The $\overline{H}$ of $\mathbf{f}_{ij}^{*}$ and the $\mathbf{c}_{m^*}$/sensor node for both cases of rural (top) and industrial (bottom) deployments (Stack 1 and Stack 2), when FSSA ((a) and (c)) and REC-FSA ((b) and (d)) are applied (k =12).

ployments. By contrast, when REC-FSA is applied, the median value of $\overline{H}$ for $\mathbf{f}_{ij}^{*}$ ranges from 0.83 (node ID = I) to 0.93 (node ID = D) for the rural environment, and from 0.73 (node ID = I) to 0.93 (node ID = D) for the industrial deployment. The difference in the performance of the feature selection with respect to the sensor node and the characteristics of the environment is expected, and highlights the asymmetry imposed by the nature of multi-hop, real-life deployments; each end-to-end link corresponds to a unique data set **A** of features and thus, the feature selection procedure cannot yield identical results. Nevertheless, the variation observed on $\overline{H}$ for $\mathbf{f}_{ij}^{*}$ across the same network when FSSA is applied equals to 0.21 (0.363) for rural (industrial) deployment. This difference is significantly higher than the one observed when REC-FSA is employed, for which the corresponding values are equal to 0.1 and 0.21 for the rural and industrial environments respectively.

Important observations additionally arise when comparing the quality of the reduced features set and the quality of the clusters formed by redundant and dominant attributes. With regard to the overall performance of FSSA in the industrial environment, the median value of $\overline{H}$ ranges within [0.48, 0.843], and [0.41, 0.78] for $\mathbf{f}_{ij}^{*}$ and $\mathbf{c}_{m^*}$, respectively. This essentially implies that the amount of uncertainty available at the reduced vector is similar to the one available at the clustered features. In addition, while in the majority of the cases the value of $\overline{H}$ for $\mathbf{f}_{ij}^{*}$ is greater than the one for $\mathbf{c}_{m^*}$, there are instances of FSSA performance where the opposite behavior is observed; for instance for node ID = D, $\overline{H}$ =0.613 for $\mathbf{f}_{ij}^{*}$, and $\overline{H}$ =0.78 for $\mathbf{c}_{m^*}$. This result indicates the failure of FSSA to construct a reduced feature vector that can efficiently grasp the network dominating characteristics. By contrast, REC-FSA outperforms in terms of $\overline{H}$ both for $\mathbf{f}_{ij}^{*}$ and $\mathbf{c}_{m^*}$. $\overline{H}$ takes median values in [0.73, 0.93], and [0.41, 0.69] for $\mathbf{f}_{ij}^{*}$ and $\mathbf{c}_{m^*}$, respectively. In addition, by employing the herein proposed scheme, we can achieve higher distance between the intra- and inter-cluster entropy, and thereby better compression than the one observed by using FSSA; for instance considering node ID = 8 the difference between intra- and inter-cluster entropy equals to 0.02 when FSSA is employed, as opposed to 0.29 when REC-FSA is employed.

Discussion. The analysis thus far highlights the following key aspects on the performance of feature selection for the characterization of end-to-end WSN links. *First, different categories of links are characterized by different performance in terms of cross validation accuracy, when both supervised as well as unsupervised feature selection techniques are employed*. This is due to the variations on the volume *D* of data available for performing feature selection, thereby highlighting the fact that the more data available over the multi-hop links, the better the accuracy on predicting the label of the link. *Second*, with regard to the quality of compression and clustering, *REC-FSA exhibits superior performance when compared to both SFFS and FSSA*, thereby highlighting the *effectiveness of the representation entropy as a criterion for compressing redundant features*. It is also worthwhile to note that the quality of compression remains essentially independent from the category of the multi-hop link, thus implying that the volume of data available on the *j*th node for calculating the set of dominant attributes does not affect the quality of clustering. *Third*, the REC-FSA achieves better combination of both maximizing the compression of information inside the cluster and preserving the necessary redundancy of the reduced set, than the one achieved by FSSA. Finally, when either REC-FSA or FSSA is employed, the increase of the value of *k*, which implies widening the search space around the most dominant feature, imposes a deterioration on the CV scores and an improvement on the value of $\overline{H}$. Therefore, the selection of the user-defined parameter *k* reveals a trade-off between the prediction accuracy that the feature selection technique yields and the redundancy that the set of dominant attributes conveys.

Taking into account the aforementioned remarks that highlight the efficacy of REC-FSA, as well as the recognized necessity of considering unsupervised feature selection for multi-hop WSN performance characterization, in the following subsection we will examine the dominant features that each unsupervised algorithm (REC-FSA, FSSA) selects for characterizing the performance of end-to-end links.

5.1.2. Dominant features for multi-hop WSN

The calculation of the set $\mathbf{f}_{ij}^{*}$ is conducted separately for each protocol stack (Stack 1 and Stack 2), and type of operational environment (rural or industrial). Our analysis considers the dominant features of each node, or equivalently, end-to-end link. The value of the parameter *k* for both algorithms is set to 12. To evaluate how representative is the set $\mathbf{f}_{ij}^{*}$ we employ the Mean Square Error (MSE) between the actual label l_{ij} of each end-to-end link and the estimated label that is extracted only when the dominant data pattern is taken into account. In addition, we present the compression ratio (CR), defined as the ratio between the length of the reduced set $\mathbf{f}_{ij}^{*}$ and the original set $\mathbf{f}_{ij}$. Finally, for the sake of simplicity we employ a numerical index for the attributes of Eq. (3):

PHY		MAC, NWK, APP		Ambient		Energy	
1	$\mu(PRX_{ij}^{*})$	7	f_i^{tx}	12	$\mu(T_i)$	16	$\mu(V_i)$
2	$\sigma(PRX_{ij}^{*})$	8	f_i^{rx}	13	$\sigma(T_i)$	17	$\sigma(V_i)$
3	$\mu(LQI_{ij}^{*})$	9	$\mu(\lvert P_{ij}\rvert)$	14	$\mu(H_i)$		
4	$\sigma(LQI_{ij}^{*})$	10	$\sigma(\lvert P_{ij}\rvert)$	15	$\sigma(H_i)$		
5	$\mu(NF_{ij}^{*})$	11	$\widehat{PRR}_{ij}$				
6	$\sigma(NF_{ij}^{*})$						

Dominant features at the rural environment. Table 3 presents the network performance (PRR_{ij}) and the results on feature selection for nodes A, D, and H, which are located 20m, 38.7m, and 58.31m away from node S (Fig. 7(a)), respectively. The initial observation to make is that there exist a slight degradation on the network performance, expressed in PRR_{ij} terms, with respect to the spatial attributes of the network. This is more evident when Stack 1 is deployed, which as mentioned above suffers from intense hidden terminal problems. This spatial correlation is also conveyed at the MSE scores, which degrade as the distance between the operational nodes and the sink node S increases. As such, when Stack 1 is deployed and the distance increases from 20 m → 38.7 m → 58.31 m, the MSE becomes 0 → 0.3787 → 0.4444 and 0 → 0.3744 → 0.4244, for the FSSA-based and the REC-FSA based calculation of $\mathbf{f}_{ij}^{*}$, respectively. With regard to Stack 2, which has a better network performance due to the cross-layer design of the routing policy, this degradation becomes smoother and the MSE score increases from 0 → 0.0519 (FSSA) and 0 → 0.0561 (REC-FSA) for node H, which has the greatest distance from node S. These results are in compliance to the observations on the dependency of the CV scores to the network performance. The increase of the MSE for networks with degraded network quality is due to the lack of sufficient volume of data at the side of the *j*th node for extracting the dominant features. However, in this case, the feature selection based on REC-FSA yields slightly improved MSE scores compared to the one that corresponds to the $\mathbf{f}_{ij}^{*}$ that is provided by FSSA.

The similarity on MSE scores does not imply that FSSA and REC-FSA classify the same types of features as important; FSSA invests on physical layer parameters, and especially on the mean value and variance of PRX_{ij}^{*}, and LQI_{ij}^{*}. By contrast, REC-FSA retains as dominant the attributes associated to the higher layers of the protocol stack, while focusing on those related to the ambient conditions

Table 3
Dominant features for selected sensor nodes w.r.t. the PRR_{ij} when FSSA and REC-FSA are employed on data collected from the rural deployments.

			FSSA			REC-FSA		
	Stack	PRR_{ij}	MSE	CR	$\mathbf{f}^*_{ij}$	MSE	CR	$\mathbf{f}^*_{ij}$
Node A (20m)	Stack 1	0.8971	0	82.3%	{1, 2, 3}	0	70.6%	{6, 7, 11, 12, 14}
	Stack 2	0.9875	0	70.6%	{2, 7, 8, 13, 17}	0	70.6%	{5, 11, 13, 14, 17}
Node D (38.7m)	Stack1	0.7803	0.3787	82.3%	{1, 9, 12}	0.3744	70.6%	{4, 5, 14, 15, 16}
	Stack 2	0.9937	0	82.3%	{2, 13, 17}	0	82.3%	{5, 7, 11}
Node H (58.31m)	Stack 1	0.7885	0.4444	70.6%	{2, 6, 9, 11, 15}	0.4244	82.3%	{5, 7, 12}
	Stack 2	0.9902	0.0519	70.6%	{1, 2, 3, 4, 13}	0.0561	70.6%	{12, 13, 14, 15, 17}

Table 4
Dominant features for selected sensor nodes w.r.t. the PRR_{ij} when FSSA and REC-FSA are employed on data collected from the industrial deployments.

			FSSA			REC-FSA		
	Stack	PRR_{ij}	MSE	CR	$\mathbf{f}^*_{ij}$	MSE	CR	$\mathbf{f}^*_{ij}$
Node A (20.1 m)	Stack 1	0.9361	0.1454	82.3%	{1, 2, 8}	0.1387	82.3%	{6, 11, 14}
	Stack 2	0.9429	0.1980	76.4%	{2, 5, 11, 16}	0.2002	76.4%	{7, 12, 13, 17}
Node D (34.63 m)	Stack 1	0.9183	0.09	82.3%	{2, 4, 13}	0.1075	82.3%	{1, 6, 17}
	Stack 2	0.9364	0.1218	82.3%	{2, 7, 8}	0.1234	82.3%	{1, 16, 17}
Node J (40.61 m)	Stack 1	0.8855	0	82.3%	{1, 2, 13}	0.0167	82.3%	{3, 6, 15}
	Stack 2	0.9379	0.1503	82.3%	{2, 6, 16}	0.1477	82.3%	{1, 8, 17}

T_i, H_i. For example, for Stack 1 and Node 1, FSSA classifies as dominant the features with index 1, 2, and 3, while REC-FSA characterizes features 6, 7, 11, 12, and 14 as the representative ones. As the duration of the network operation expands from 24 (Stack 1) to 42 hours (Stack 2), both FSSA and REC-FSA classify the changeability of the energy available V_i as a dominant feature.

Similar observations can be made for the remaining contents of Table 3, while all dominant features per node, type of protocol stack, and feature selection algorithm are highlighted in Fig. 15. Specifically, with regard to Stack 1 we observe that when FSSA is employed (Fig. 15(a)), the attributes associated with PRX^*_{ij} are common to all nodes, while significant is also the presence of LQI^*_{ij} and the length of $|P_{ij}|$. By contrast, as illustrated at Fig. 15(b), when REC-FSA is applied, the presence of features associated with the Physical layer metrics, is sparse; instead the reception rate f^{rx}_i is the most common significant feature, as it appears on four subsequent sensor nodes at the farthest side of the network. No other spatial correlation on the dominant attributes that REC-FSA selects can be derived. Shifting towards the dominant features for characterizing the performance when Stack 2 is deployed, the behavior of each node in terms of dominant features becomes more consistent across the network, when either FSSA (Fig. 15(c)) or REC-FSA (Fig. 15(d)) is applied. In fact, when FSSA is used, the received signal strength at the physical layer (PRX^*_{ij}) dominates the $\mathbf{f}^*_{ij}$ across the entire network. In addition, due to the optimal network behavior, attributes from the intermediate layers (MAC, NWK) are classified as redundant across the network, and the standard deviation of on-board temperature T_i and battery level V_i are gaining prominence. The mean value of the noise floor NF_i is classified by REC-FSA as a significant attribute for the sensor nodes that are closer to the sink node. The remaining features that are chosen by REC-FSA are related to the mean and the standard deviation of on-board temperature T_i and humidity H_i.

Dominant features at the industrial environment. Table 4 presents the value of PRR_{ij} and the results on feature selection for nodes A, D, and J, which are located 20.1m, 34.63m, and 40.61m away from node S (Fig. 7(b)), respectively.

Despite the RF-harshness of the industrial environment, both Stack 1 and Stack 2 are characterized by an optimal network performance, since $PRR_{ij} \geq 0.8855$. Moreover, both unsupervised feature selection algorithms achieve a better compression ratio than the one they deliver when applied to the datasets from the rural deployments; for the majority of the end-to-end links, 82.3% of the features are characterized as redundant, thereby classifying the remaining 17.7% in the reduced set $\mathbf{f}^*_{ij}$. However, the increased percentage of compressed features is accompanied by a degradation of the MSE scores. With regard to Stack 1, the deterioration of MSE is inversely analogous to the distance between the operational nodes and the sink node; as the distance increases from 20.1m → 34.63m → 40.61m, the MSE value decreases from 0.1454 → 0.09 → 0 and 0.1387 → 0.1075 → 0.0167, for FSSA and REC-FSA, respectively. This implies that in contrast to the homogeneity of the rural environment, for the industrial, application-driven deployments, the performance of the feature selection algorithms in terms of MSE depends on the specific characteristics of the physical space (e.g., layout, geometry, material of the objects) and the presence of heavy machinery, obstacles, people, mobility. The variation of the MSE scores increases as we shift from Stack 1 to Stack 2, and subsequently the duration of the network operation expands from 2.5 to 162 hours. In this case, as we move from node A to node D, which are located 14.53m apart from each other, the MSE value decreases from 0.1980 to 0.1218 (0.2002 to 0.1234) when FSSA (REC-FSA) is applied.

Fig. 16 presents all dominant features per node, type of protocol, and feature selection algorithm. With regard to Stack 1, when FSSA is employed (Fig. 16(a)), the attributes associated to the quality of reception at the physical layer (PRX^*_{ij}, LQI^*_{ij}) dominate in $\mathbf{f}^*_{ij}$ across the entire network. Opposed to FSSA, REC-FSA (Fig. 16(b)) selects the features related to noise floor NF^*_{ij} and humidity. Indeed, humidity becomes a primary feature for characterizing the network performance for the sensor nodes that operate closer to bulky water tanks (Nodes B, C, J, I), thereby highlighting the fact that REC-FSA can grasp the non-linear and high-level correlations of the operational environment. It is also considered important to note that for both FSSA and REC-FSA, $\sigma(|P_{ij}|)$ is always compressed by another feature, and thus, the length routing paths formulated in the industrial environment remain constant due to the limited connectivity options. With regard to Stack 2, when FSSA is applied (Fig. 16(c)), the standard deviation of PRX^*_{ij} comes up as the dom-

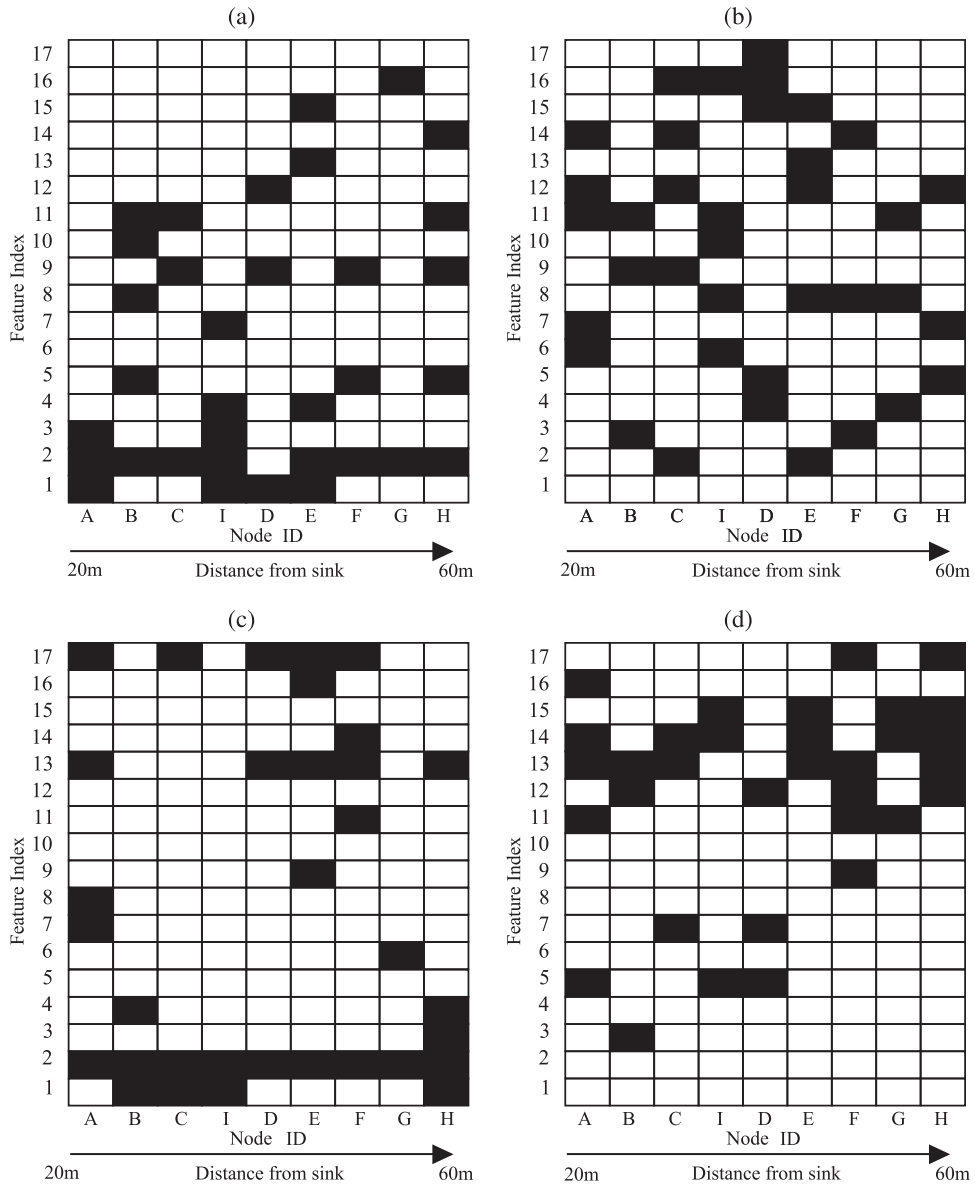

Fig. 15. Dominant features for the rural environment with respect to the distance from the sink node S: (a) FSSA on data collected over Stack 1, (b) REC-FSA on data collected over Stack 1, (c) FSSA on data collected over Stack 2, and (d) REC-FSA on data collected over Stack 2.

inant attribute and the remaining features related to the Physical layer are characterized as redundant. This result is in compliance with the respective case of rural deployment and is due to the functionality of Stack 2, which exploits the value of LQI for making routing decisions. As such, routing paths tend to retain satisfactory level of LQI and, thus the respective features are characterized by limited salience. The effect of long term operation of the network is evident on the dominant features that REC-FSA selects, as presented in Fig. 16(d). With respect to features associated to the Physical layer, the mean value of PRX_{ij}^{*} is considered the most representative attribute. In conjunction to the selection of f_i^{rx} and the mean value of $|P_{ij}|$, this highlights the changes on the routing paths selected, and accompanying variations on the MAC traffic recorded at each node, as the network topology changes due to battery depletion. The impact of the energy degradation in $\mathbf{f}_{ij}^{*}$ is also reflected by the existence of the standard deviation of V_i in $\mathbf{f}_{ij}^{*}$, thereby highlighting the fact that changeability associated to the on-board current consumption is more important than the available power-supply.

Discussion. The results presented highlight how the differences on the WSN deployments are reflected on the dominant features for characterizing the network performance. The improvement on the network performance is accompanied by a homogeneity on the dominant attributes across the entire network for both FSSA and REC-FSA. Nevertheless, especially for the industrial deployment, we can observe how the heterogeneity of the surrounding environment can affect the contents of $\mathbf{f}_{ij}^{*}$ for each individual sensor node, when Stack 1 is employed. For long-term WSN operations (Stack 2), the combination of attributes related to PRX_{ij}^{*} and the level of on-board energy V_i are classified in $\mathbf{f}_{ij}^{*}$ for all sensor nodes.

Finally, the difference on the type of features that each algorithm tends to select is considered important for the purely distributed implementation of feature selection over multi-hop WSNs; Both for the rural and the industrial environments, FSSA characterizes as dominant features that are associated with different sides of path P_{ij}, emphasizing on the received power PRX_{ij}^{*} and the link quality indicator LQI_{ij}^{*}. Notably, the selection of the physical layer

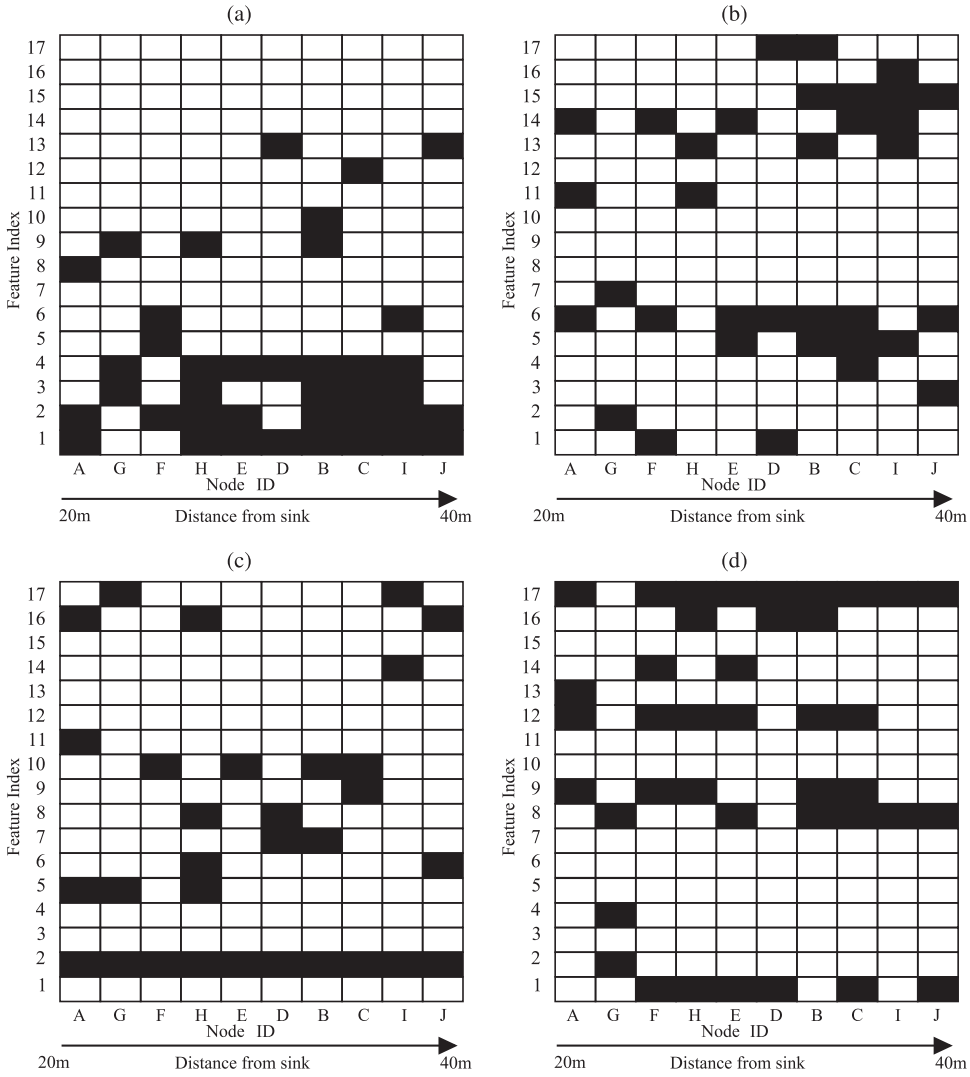

Fig. 16. Dominant features for the industrial environment with respect to the distance from the sink node: (a) FSSA on data collected over Stack 1, (b) REC-FSA on data collected over Stack 1, (c) FSSA on data collected over Stack 2, (d) REC-FSA on data collected over Stack 2.

parameters as part of the reduced feature vector $\mathbf{f}_{ij}^{*}$ is due to the spatio-temporal variations of the RSSI, and LQI parameters, which are increased as we shift from the rural to the industrial environment and while the respective metrics are not taken into consideration for routing decisions (Stack 1). As such, the dominant features provided by FSSA are in compliance to the current state of the art, which emphasizes both on the impact of multipath and scattering propagation effects for characterizing the network performance of point-to-point links (Section 2), as well as on RSSI-based design choices [46]. By contrast, REC-FSA selects features that are available on the ith node, related to both the status of the sensor node (e.g., V_i, T_i), as well as network metrics (e.g., f_i^{rx}). The results provided by REC-FSA do not contradict current state-of-the-art on the importance of link quality estimation for network performance characterization. Instead, they highlight that the similar estimation can be achieved by metrics that are available at each node i. **Thus, the respective features can be employed for promoting the inference of l_{ij} on the side of the ith node, thereby estimating the network performance of the link $i \rightarrow j$ in a purely localized manner**.

6. Conclusions and future work

In this work, we have addressed the problem of characterizing the performance of realistic WSN deployments, by proposing an integrated framework for feature selection over multi-hop, end-to-end links. The evaluation studies on two different cases of protocol stacks, deployed in both industrial and rural environments, emphasize the efficacy of our approach both in terms of the heterogeneity herein proposed for constructing the original feature vector, and in terms of salience and qualitative compression of the reduced pattern of dominant attributes.

Appropriate feature selection can improve the quality of the learned models applied during classification. The herein proposed framework for feature selection concentrates on a single aspect of network performance (PRR_{ij}), however it can serve as a design paradigm for examining additional aspects of WSN performance, such as lifetime. Moreover, the structured methodology adopted allows the scalable adaptation of the proposed scheme towards more expanded deployments. Finally, the lightweight nature of REC-FSA promotes the real-time character of the proposed feature selection scheme.

Our immediate next steps concentrate on addressing the distributed limitations of the proposed framework, and specifically the adaptive extraction of the original feature vector. In addition, while in this work we have considered pairwise characterization of redundancy and compression, our future directions link to graph-based models. We envisage that such an approach will further improve the characterization of WSN deployments, whilst yielding

the roadmap for on-node and in-network learning models in optimized processing time.

Acknowledgements

This work was supported by the FP7 EU-HYDROBIONETS project (ICT-2011-7, GA-2011-287613) and co-financed by European Union and Greek national funds through the National Strategic Reference Framework (NSRF), Research Funding Program: Cooperation-2011, Project SeNSE. We are grateful to Acciona Agua[1] for providing the premises of La Tordera's desalination plant, as well as to Professor Baltasar Beferull-Lozano and Dr Ioannis Glaropoulos for assisting with the WSN protocol stacks development, and the anonymous reviewers for their insightful comments. Finally, we would like to dedicate this work to the memory of Mr Eugenio Celada, without whom the industrial field studies would not be possible.

Appendix A. Power consumption and lifetime calculation for IEEE 802.15.4 - compatible transceivers

The lifetime τ_i of the ith node is defined as the time period within the energy available at a transceiver is sufficient both for transmitting data at a desired transmission power P_i^{tx}, as well as the receiving frames from the 1-hop neighbors. The estimations of the power consumption P_i^{cons} and the lifetime τ_i are tightly coupled to each other, and based on the widely-adopted assumption that the communication is the most energy-demanding aspect of a sensor network.

The lifetime τ_i, expressed in hours, is associated to: (a) the duty cycle of the operational transceiver, (b) P_i^{cons} (W), (c) the energy available E_i (Wh), and (d) the generated traffic at the MAC layer, which is expressed in terms of transmission f_i^{tx} and reception f_i^{rx} rate (bps). Specifically, by adopting the model of [47], the lifetime τ_i is defined as the number of duty cycles that are still available at each node, prior its communication abilities start malfunctioning due to energy depletion:

$$\tau_i = \frac{E_i}{P_i^{cons}}(h), \tag{6}$$

where P_i^{cons} depends on the voltage supply (V_i) and the mean value of current consumption I_m on the transceiver, i.e. $P_i^{cons} = V_i I_m$.

With regard to the current drawn per duty cycle, we consider the worst-case scenario, according to which an energy-efficient radio duty circle is deactivated, I_m can be considered as the sum of the current drawn for the transmission I_{tx}, the reception I_{rx}, and the stand-by I_{idle} modes, i.e., $I_m = \delta_{idle} I_{idle} + \delta_{tx} I_{tx} + \delta_{rx} I_{rx}$. Note that δ_{idle}, δ_{tx}, and δ_{rx} respectively correspond to the duration (in symbols) of channel listening, transmission and reception within a duty circle of the transceiver, and $\delta_{idle} + \delta_{tx} + \delta_{rx} = 1$. Without loss of generality, these timings can be extracted by considering the timing parameters of IEEE-802.15.4 MAC functional operations [35], related to: (a) the duration of backoff period, and the clear channel assessment period, (b) the time needed for transmitting data and acknowledgment frames as a function of the nominal bit rate and the transmission and reception rate (f_i^{tx}, f_i^{rx}), (c) the time needed for the transmitter - receiver synchronization, (d) the turn-around time of an receiver before sending an ACK frame.

In each of the three operational modes of the transceiver, the nominal current draw is provided by the transceiver's manufacturer, while especially for the transmission mode different levels of current consumption are available for different levels of transmission power. For instance, considering the case of the popular TI-CC2420 transceiver [48], when $P_i^{tx} = 0$dBm the total current draw on the ith node becomes $I_m = \delta_{idle} 18.8 + \delta_{tx} 17.4 + \delta_{rx} 18.8 (mA)$ [49].

Expect for the characteristics of the generated traffic, which result into different levels of current and power consumption, the battery-operated lifetime of the network is dictated by the operational specifications of the employed batteries, and how their capacity drops as a function of current load and time elapsed. While the respective phenomena are associated with the chemical characteristics of the batteries cells, the authors in [50], describe the procedure of extracting an empirical model of calculating the so-called State-of-Charge (SoC) of a battery, defined as the estimate on the current capacity of the battery. According to methodology therein presented, an estimation of the SoC can be derived by performing curve fitting on the normalized lifetime of the battery, based the voltage measurements. Specifically, SoC can be approximated by:

$$SoC(V_{supp}) = 1 - DoD(V_i),$$

where $DoD(V_i)$ is defined as the Depth-of-Discharge and is associated to percentage of the discharged capacity of the battery cells; the known battery voltage of a fully charged battery defines the $0\%DoD$, while the battery voltage of an empty battery defines the $100\%DoD$. Using the recorded time $t_{cuttoff}$ it takes the battery to drain, the value of DoD is empirically assigned to $DoD = \frac{t}{t_{cutoff}}$ at each time instant t. This linear association allows adopting a curve fitting approach for associating the Depth-of-Discharge to the input voltage supply:

$$DoD(V_{supp}) = \sum_{i=0}^{n} \alpha_n V_i^n,$$

with the objective to approximate the level of discharged battery as a $n-$th order polynomial function of the input voltage supply.

Based on this approach, the lifetime of each operational node can be extracted as a function of: (a) the available level of energy expressed as a function of $SoC(V_i)$ and (b) the power consumption P_i^{cons}. As such, Eq. (6) can be rewritten as follows:

$$\tau_i = \frac{SoC(V_i) \times V_{full} \times C_{init}}{P_i^{cons}}(h), \tag{7}$$

where V_{init} and C_{init} express the voltage level (V) and the capacity (Ah) of the full battery respectively.

[1] http://www.acciona-agua.com/

References

[1] M. Erol-Kantarci, H.T. Mouftah, Wireless multimedia sensor and actor networks for the next generation power grid, Ad Hoc Netw. 9 (4) (2011) 542–551.

[2] D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, T. Arn, J. Beutel, L. Thiele, Deriving high-resolution urban air pollution maps using mobile sensor nodes, Pervasive Mobile Comput. 16 (Part B (0)) (2015) 268–285. Selected Papers from the Twelfth Annual {IEEE} International Conference on Pervasive Computing and Communications (PerCom 2014)

[3] C. Poon, B. Lo, M. Yuce, A. Alomainy, Y. Hao, Body sensor networks: In the era of big data and beyond, Biomed. Eng. IEEE Rev. 8 (2015) 4–16.

[4] C. Di Martino, M. Cinque, D. Cotroneo, Automated generation of performance and dependability models for the assessment of wireless sensor networks, Comput. IEEE Trans. 61 (6) (2012) 870–884.

[5] T. Laukkarinen, J. Suhonen, T. Hamalainen, M. Hannikainen, Pilot studies of wireless sensor networks: Practical experiences, in: Design and Architectures for Signal and Image Processing (DASIP), 2011 Conference on, 2011, pp. 1–8.

[6] T. Surmacz, M. Sabicki, B. Wojciechowski, M. Nikodem, Lessons learned from the deployment of wireless sensor networks, in: A. Kwiecie, P. Gaj, P. Stera (Eds.), Computer Networks, Communications in Computer and Information Science, vol. 370, Springer, Berlin Heidelberg, 2013, pp. 76–85.

[7] Y. Liu, Y. He, M. Li, J. Wang, K. Liu, X. Li, Does wireless sensor network scale? a measurement study on greenorbs, Parallel Distributed Syst. IEEE Trans. 24 (10) (2013) 1983–1993.

[8] R. Min, M. Bhardwaj, S. Cho, N. Ickes, E. Shih, A. Sinha, A. Wang, A. Chandrakasan, Energy-centric enabling tecumologies for wireless sensor networks, Wireless Commun. IEEE 9 (4) (2002) 28–39.

[9] C.M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[10] M. Abu Alsheikh, S. Lin, D. Niyato, H.-P. Tan, Machine learning in wireless sensor networks: Algorithms, strategies, and applications, Commun. Surv. Tut. IEEE 16 (4) (2014) 1996–2018.
[11] K. Srinivasan, P. Dutta, A. Tavakoli, P. Levis, An empirical study of low-power wireless, ACM Trans. Sen. Netw. 6 (2) (2010) 16:1–16:49.
[12] N. Baccour, A. Koubâa, C.A. Boano, L. Mottola, H. Fotouhi, M. Alves, H. Youssef, M.A. Ziga, D. Puccinelli, T. Voigt, K. Römer, C. Noda, Radio Link Quality Estimation in Low-Power Wireless Networks, SpringerBriefs in Electrical and Computer Engineering, Springer, 2013.
[13] C. Boano, M. Ziga, T. Voigt, A. Willig, K. Romer, The triangle metric: Fast link quality estimation for mobile wireless sensor networks, in: Computer Communications and Networks (ICCCN), 2010 Proceedings of 19th International Conference on, 2010, pp. 1–7.
[14] Y. Wang, M. Martonosi, L.-S. Peh, Predicting link quality using supervised learning in wireless sensor networks, SIGMOBILE Mob. Comput. Commun. Rev. 11 (3) (2007) 71–83.
[15] G. Di Caro, M. Kudelski, E. Flushing, J. Nagi, I. Ahmed, L. Gambardella, Online supervised incremental learning of link quality estimates in wireless networks, in: Ad Hoc Networking Workshop (MED-HOC-NET), 2013 12th Annual Mediterranean, 2013, pp. 133–140.
[16] T. Liu, A.E. Cerpa, Data-driven link quality prediction using link features, ACM Trans. Sen. Netw. 10 (2) (2014) 37:1–37:35.
[17] T. Liu, A.E. Cerpa, Temporal adaptive link quality prediction with online learning, ACM Trans. Sen. Netw. 10 (3) (2014) 46:1–46:41.
[18] T.M. Mitchell, Machine Learning, 1st ed., McGraw-Hill, Inc., New York, NY, USA, 1997.
[19] S. Vijayakumar, A. D'souza, S. Schaal, Incremental online learning in high dimensions, Neural Comput. 17 (12) (2005) 2602–2634.
[20] G. Werner-Allen, P. Swieskowski, M. Welsh, Motelab: A wireless sensor network testbed, in: Proceedings of the 4th International Symposium on Information Processing in Sensor Networks, in: IPSN '05, IEEE Press, Piscataway, NJ, USA, 2005.
[21] M. Doddavenkatappa, M. Chan, A. Ananda, Indriya: A low-cost, 3d wireless sensor network testbed, in: T. Korakis, H. Li, P. Tran-Gia, H.-S. Park (Eds.), Testbeds and Research Infrastructure. Development of Networks and Communities, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 90, Springer, Berlin Heidelberg, 2012, pp. 302–316.
[22] N. Baccour, A. Koubía, M. Ben Jamía, D. do Rosário, H. Youssef, M. Alves, L.B. Becker, Radiale: A framework for designing and assessing link quality estimators in wireless sensor networks, Ad Hoc Netw. 9 (7) (2011) 1165–1185.
[23] V.C. Gungor, M.K. Korkmaz, Wireless link-quality estimation in smart grid environments, Int. J. Distributed Sensor Netw. 2012 (2012).
[24] S. Rekik, N. Baccour, M. Jmaiel, K. Drira, Low-power link quality estimation in smart grid environments, in: 11th International Wireless Communications & Mobile Computing Conference (IWCMC 2015), 2015.
[25] V. Gungor, B. Lu, G. Hancke, Opportunities and challenges of wireless sensor networks in smart grid, Ind. Electr. IEEE Trans. 57 (10) (2010) 3557–3564.
[26] C. Boano, K. Romer, N. Tsiftes, Mitigating the adverse effects of temperature on low-power wireless protocols, in: Mobile Ad Hoc and Sensor Systems (MASS), 2014 IEEE 11th International Conference on, 2014, pp. 336–344.
[27] F. Schmidt, M. Ceriotti, N. Hauser, K. Wehrle, If you can't take the heat: Temperature effects on low-power wireless networks and how to mitigate them, in: T. Abdelzaher, N. Pereira, E. Tovar (Eds.), Wireless Sensor Networks, Lecture Notes in Computer Science, vol. 8965, Springer International Publishing, 2015, pp. 266–273.
[28] G. Anastasi, M. Conti, M. Di Francesco, A comprehensive analysis of the mac unreliability problem in ieee 802.15.4 wireless sensor networks, Ind. Inform. IEEE Trans. 7 (1) (2011) 52–65.
[29] S. Sudevalayam, P. Kulkarni, Energy harvesting sensor nodes: Survey and implications, IEEE Commun. Surv. Tut. 13 (3) (2011) 443–461, doi:10.1109/SURV.2011.060710.00094.
[30] F. Akhtar, M.H. Rehmani, Energy replenishment using renewable and traditional energy resources for sustainable wireless sensor networks: A review, Renewable Sustain. Ener. Rev. 45 (2015) 769–784, doi:10.1016/j.rser.2015.02.021. URL http://www.sciencedirect.com/science/article/pii/S1364032115001094.
[31] A. Woo, T. Tong, D. Culler, Taming the underlying challenges of reliable multihop routing in sensor networks, in: Proceedings of the 1st International Conference on Embedded Networked Sensor Systems, in: SenSys '03, ACM, New York, NY, USA, 2003, pp. 14–27.
[32] P. Mitra, C.A. Murthy, S. Pal, Unsupervised feature selection using feature similarity, Pattern Anal. Mach. Intell. IEEE Trans. 24 (3) (2002) 301–312.
[33] V. Rao, V.N. Sastry, Unsupervised feature ranking based on representation entropy, in: Recent Advances in Information Technology (RAIT), 2012 1st International Conference on, 2012, pp. 421–425.
[34] G. Tzagkarakis, G. Tsagkatakis, D. Alonso, E. Celada, C. Asensio, A. Panousopoulou, P. Tsakalides, B. Beferull-Lozano, Signal and data processing techniques for industrial cyber-physical systems, in: D.B. Rawat, J. Rodrigues, I. Stojmenovic (Eds.), Cyber Physical Systems: From Theory to Practice, CRC Press, USA, 2015.
[35] IEEE Std 802.15.4, Part 15.4: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (LR-WPANs), 2011.
[36] A. Dunkels, B. Gronvall, T. Voigt, Contiki - a lightweight and flexible operating system for tiny networked sensors, in: Local Computer Networks, 2004. 29th Annual IEEE International Conference on, 2004, pp. 455–462.
[37] Advanticsys wireless sensor modules, 2013, http://www.advanticsys.com/wiki/).
[38] The pandaboard platform, (http://pandaboard.org/).
[39] D. Alonso-Roman, E. Celada-Funes, C. Asensio-Marco, B. Beferull-Lozano, Improving reliability and efficiency of communications in wsns under high traffic demand, in: Wireless Communications and Networking Conference (WCNC), 2013 IEEE, 2013, pp. 268–273.
[40] P. Di Marco, C. Fischione, G. Athanasiou, P.-V. Mekikis, Harmonizing mac and routing in low power and lossy networks, in: Global Communications Conference (GLOBECOM), 2013 IEEE, 2013, pp. 231–236.
[41] N. Tsiftes, J. Eriksson, A. Dunkels, Low-power wireless ipv6 routing with contikirpl, in: Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks, in: IPSN '10, ACM, New York, NY, USA, 2010, pp. 406–407.
[42] The qt framework, (http://qt-project.org/).
[43] Z.I. Botev, J.F. Grotowski, D.P. Kroese, Kernel density estimation via diffusion, The Ann. Stat. 38 (5) (2010) 2916–2957.
[44] 2015, (https://github.com/apanouso/wsn-indfeat-dataset).
[45] S. Pal, R. De, J. Basak, Unsupervised feature evaluation: a neuro-fuzzy approach, Neural Netw. IEEE Trans. 11 (2) (2000) 366–376.
[46] M.H. Rehmani, A. Rachedi, S. Lohier, T. Alves, B. Poussot, Intelligent antenna selection decision in {IEEE} 802.15.4 wireless sensor networks: An experimental analysis, Comput. Electr. Eng. 40 (2) (2014) 443–455, doi:10.1016/j.compeleceng.2013.11.021. http://www.sciencedirect.com/science/article/pii/S0045790613003030
[47] V. Mhatre, C. Rosenberg, D. Kofman, R. Mazumdar, N. Shroff, A minimum cost heterogeneous sensor network with a lifetime constraint, Mobile Comput. IEEE Trans. 4 (1) (2005) 4–15, doi:10.1109/TMC.2005.2(410)4.
[48] CC2420: Single-Chip 2.4 GHz IEEE 802.15.4 Compliant and ZigBee Ready RF Transceiver
[49] E. Casilari, J.M. Cano-GarcÈ¡a, G. Campos-Garrido, Modeling of current consumption in 802.15.4/zigbee sensor motes, Sensors 10 (6) (2010) 5443–5468, doi:10.3390/s100605443.
[50] B. Buchli, D. Aschwanden, J. Beutel, Battery state-of-charge approximation for energy harvesting embedded systems, in: P. Demeester, I. Moerman, A. Terzis (Eds.), Wireless Sensor Networks, Lecture Notes in Computer Science, vol. 7772, Springer, Berlin Heidelberg, 2013, pp. 179–196, doi:10.1007/978-3-642-36672-7_12.

Athanasia Panousopoulou received the Diploma and PhD in Electrical and Computer Engineering from the University of Patras, Greece in 2004 and 2009 respectively. She is currently a Post-doc Researcher with the Signal Processing Laboratory, the Institute of Computer Science, Foundation for Research and Technology-Hellas, Greece. Her research interest include reconfiguration techniques for wireless sensor and body-area networks, distributed network algorithms, and distributed inference.

Mikel Azkune received the M.Sc. in Telecommunications Engineering from the University of the Basque Country, Spain in 2014. His Master Thesis was on applications of Machine Learning on Wireless Sensor Networks at the Computer Science Department, University of Crete, Greece, under the supervision of Prof. Tsakalides. Currently, he is pursuing his Ph. D. in the field of micro structured polymer optical fibers for bio-sensing applications, at the Applied Photonics Group, University of the Basque Country, Spain.

Panagiotis Tsakalides received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1995. He is a Professor of Computer Science at the University of Crete, and the Head of the Signal Processing Lab at the Institute of Computer Science, Foundation for Research and Technology-Hellas, Greece. His research interests lie in the field of statistical signal processing with emphasis in non-Gaussian estimation and detection theory, and applications in sensor networks, imaging, and multimedia systems.